Martin Golubitsky
David G. Schaeffer

Applied Mathematical Sciences 51

Singularities and Groups in Bifurcation Theory

Volume I

Springer

Applied Mathematical Sciences
Volume 51

Springer Science+Business Media, LLC

Applied Mathematical Sciences

1. *John:* Partial Differential Equations, 4th ed.
2. *Sirovich:* Techniques of Asymptotic Analysis.
3. *Hale:* Theory of Functional Differential Equations, 2nd ed.
4. *Percus:* Combinatorial Methods.
5. *von Mises/Friedrichs:* Fluid Dynamics.
6. *Freiberger/Grenander:* A Short Course in Computational Probability and Statistics.
7. *Pipkin:* Lectures on Viscoelasticity Theory.
8. *Giacaglia:* Perturbation Methods in Non-linear Systems.
9. *Friedrichs:* Spectral Theory of Operators in Hilbert Space.
10. *Stroud:* Numerical Quadrature and Solution of Ordinary Differential Equations.
11. *Wolovich:* Linear Multivariable Systems.
12. *Berkovitz:* Optimal Control Theory.
13. *Bluman/Cole:* Similarity Methods for Differential Equations.
14. *Yoshizawa:* Stability Theory and the Existence of Periodic Solution and Almost Periodic Solutions.
15. *Braun:* Differential Equations and Their Applications, 3rd ed.
16. *Lefschetz:* Applications of Algebraic Topology.
17. *Collatz/Wetterling:* Optimization Problems.
18. *Grenander:* Pattern Synthesis: Lectures in Pattern Theory, Vol. I.
19. *Marsden/McCracken:* Hopf Bifurcation and Its Applications.
20. *Driver:* Ordinary and Delay Differential Equations.
21. *Courant/Friedrichs:* Supersonic Flow and Shock Waves.
22. *Rouche/Habets/Laloy:* Stability Theory by Liapunov's Direct Method.
23. *Lamperti:* Stochastic Processes: A Survey of the Mathematical Theory.
24. *Grenander:* Pattern Analysis: Lectures in Pattern Theory, Vol. II.
25. *Davies:* Integral Transforms and Their Applications, 2nd ed.
26. *Kushner/Clark:* Stochastic Approximation Methods for Constrained and Unconstrained Systems.
27. *de Boor:* A Practical Guide to Splines.
28. *Keilson:* Markov Chain Models—Rarity and Exponentiality.
29. *de Veubeke:* A Course in Elasticity.
30. *Śniatycki:* Geometric Quantization and Quantum Mechanics.
31. *Reid:* Sturmian Theory for Ordinary Differential Equations.
32. *Meis/Markowitz:* Numerical Solution of Partial Differential Equations.
33. *Grenander:* Regular Structures: Lectures in Pattern Theory, Vol. III.
34. *Kevorkian/Cole:* Perturbation Methods in Applied Mathematics.
35. *Carr:* Applications of Centre Manifold Theory.
36. *Bengtsson/Ghil/Källén:* Dynamic Meteorology: Data Assimilation Methods.
37. *Saperstone:* Semidynamical Systems in Infinite Dimensional Spaces.
38. *Lichtenberg/Lieberman:* Regular and Chaotic Dynamics, 2nd ed.
39. *Piccini/Stampacchia/Vidossich:* Ordinary Differential Equations in $\mathbf{R}^n$.
40. *Naylor/Sell:* Linear Operator Theory in Engineering and Science.
41. *Sparrow:* The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors.
42. *Guckenheimer/Holmes:* Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields.
43. *Ockendon/Taylor:* Inviscid Fluid Flows.
44. *Pazy:* Semigroups of Linear Operators and Applications to Partial Differential Equations.
45. *Glashoff/Gustafson:* Linear Operations and Approximation: An Introduction to the Theoretical Analysis and Numerical Treatment of Semi-Infinite Programs.
46. *Wilcox:* Scattering Theory for Diffraction Gratings.
47. *Hale et al:* An Introduction to Infinite Dimensional Dynamical Systems—Geometric Theory.
48. *Murray:* Asymptotic Analysis.
49. *Ladyzhenskaya:* The Boundary-Value Problems of Mathematical Physics.
50. *Wilcox:* Sound Propagation in Stratified Fluids.
51. *Golubitsky/Schaeffer:* Bifurcation and Groups in Bifurcation Theory, Vol. I.
52. *Chipot:* Variational Inequalities and Flow in Porous Media.
53. *Majda:* Compressible Fluid Flow and System of Conservation Laws in Several Space Variables.
54. *Wasow:* Linear Turning Point Theory.
55. *Yosida:* Operational Calculus: A Theory of Hyperfunctions.
56. *Chang/Howes:* Nonlinear Singular Perturbation Phenomena: Theory and Applications.
57. *Reinhardt:* Analysis of Approximation Methods for Differential and Integral Equations.
58. *Dwoyer/Hussaini/Voigt (eds):* Theoretical Approaches to Turbulence.
59. *Sanders/Verhulst:* Averaging Methods in Nonlinear Dynamical Systems.
60. *Ghil/Childress:* Topics in Geophysical Dynamics: Atmospheric Dynamics, Dynamo Theory and Climate Dynamics.

(continued following index)

Martin Golubitsky
David G. Schaeffer

Singularities and Groups in Bifurcation Theory

Volume I

With 114 Illustrations

Springer

Martin Golubitsky
Department of Mathematics
University of Houston
Houston, TX 77004
USA

David G. Schaeffer
Department of Mathematics
Duke University
Durham, NC 27706
USA

Editors

J.E. Marsden
Control and Dynamical Systems, 107-81
California Institute of Technology
Pasadena, CA 91125
USA

L. Sirovich
Division of Applied Mathematics
Brown University
Providence, RI 02912
USA

Mathematics Subject Classification: 35B32, 34D05, 34D20, 34D30, 73D05, 58F14, 34L35, 35L67, 76L05, 57R45

Library of Congress Cataloging-in-Publication Data
Golubitsky, Martin.
Singularities and groups in bifurcation theory.
(Applied mathematical sciences; 51)
Bibliography: p.
Includes index.
1. Bifurcation theory. 2. Singularity Theory.
I. Schaeffer, David G. II. Title. III. Series: Applied mathematical sciences (Springer-Verlag New York Inc.); 51.
QA374.G59 1984 515.3′53 84-1414

Printed on acid-free paper.

Originally published by Springer-Verlag New York, Inc. in 1985
Softcover reprint of the hardcover 1st edition 1985

9 8 7 6 5 4 3 2 SPIN 10782874

ISBN 978-1-4612-9533-4 ISBN 978-1-4612-5034-0 (eBook)
DOI 10.1007/978-1-4612-5034-0

To Elizabeth and Alexander

and

To Jennie

Preface

This book has been written in a frankly partisian spirit—we believe that singularity theory offers an extremely useful approach to bifurcation problems and we hope to convert the reader to this view. In this preface we will discuss what we feel are the strengths of the singularity theory approach. This discussion then leads naturally into a discussion of the contents of the book and the prerequisites for reading it.

Let us emphasize that our principal contribution in this area has been to apply pre-existing techniques from singularity theory, especially unfolding theory and classification theory, to bifurcation problems. Many of the ideas in this part of singularity theory were originally proposed by René Thom; the subject was then developed rigorously by John Mather and extended by V. I. Arnold. In applying this material to bifurcation problems, we were greatly encouraged by how well the mathematical ideas of singularity theory meshed with the questions addressed by bifurcation theory.

Concerning our title, *Singularities and Groups in Bifurcation Theory*, it should be mentioned that the present text is the first volume in a two-volume sequence. In this volume our emphasis is on singularity theory, with group theory playing a subordinate role. In Volume II the emphasis will be more balanced.

Having made these remarks, let us set the context for the discussion of the strengths of the singularity theory approach to bifurcation. As we use the term, bifurcation theory is the study of equations with multiple solutions. Specifically, by a *bifurcation* we mean a change in the number of solutions of an equation as a parameter varies. For a wide variety of equations, including many partial differential equations, problems concerning multiple solutions can be reduced to studying how the solutions x of a single scalar equation

$$g(x, \lambda) = 0 \qquad \text{(P.1)}$$

vary with the parameter λ. This simplification depends on a technique known as the Liapunov–Schmidt reduction.

The singularity theory approach deals with equations of the form (P.1); i.e., with equations *after* the Liapunov–Schmidt reduction has been performed. We shall emphasize the qualitative properties of such equations. This emphasis is sharply focused by the notion of *equivalence*, which defines precisely what it means for two such equations, and their solution sets, to be qualitatively similar.

The theory quickly leads one to generalize (P.1) to include k-parameter families of such equations; i.e., equations of the form

$$G(x, \lambda, \alpha) = 0, \tag{P.2}$$

where $\alpha = (\alpha_1, \ldots, \alpha_k)$ is a shorthand for k auxiliary parameters. We shall call G an *unfolding* of g if for $\alpha = 0$

$$G(x, \lambda, 0) = g(x, \lambda). \tag{P.3}$$

Since

$$G(x, \lambda, \alpha) = g(x, \lambda) + [G(x, \lambda, \alpha) - G(x, \lambda, 0)],$$

we may think of $G(x, \lambda, \alpha)$ as a perturbation of $g(x, \lambda)$.

In this volume we limit our discussion of (P.2) in the following four ways:

(i) we assume that the dependence of G on x, λ, and α is infinitely differentiable;
(ii) we consider primarily the case where x is a scalar (one-dimensional) unknown;
(iii) we work locally (i.e., in the neighborhood of some fixed point (x_0, λ_0)); and
(iv) we discuss dynamics only in a limited way.

Brief discussions of points (i) and (iii) occur later in this Preface. Concerning point (ii), in Volume II we will consider finite-dimensional systems of equations with several unknowns. Let us elaborate on point (iv). Typically, equations with multiple solutions arise in characterizing steady-state solutions of an evolution equation. Singularity theory methods are useful in finding the steady-state solutions and, in some instances, their stabilities. However, it does not seem to be possible with these methods to analyze essentially dynamic phenomena such as strange attractors.

One general strength of the singularity theory approach to bifurcation problems is easily stated—this approach unifies the treatment of many diverse problems in steady-state bifurcation. Such unification has the obvious advantage of elegance, but it also leads to economy of effort. Specifically, the same general methods used to study the most familiar problems in bifurcation theory continue to apply in a variety of nonstandard contexts. For example, whether or not $g(x, \lambda) = 0$ has a trivial solution and whether or not symmetries are present, the theoretical framework of the singularity theory approach is the same. Also, although in this text we consider only equations having λ as a distinguished parameter, the same techniques work equally well when all parameters are treated on the same footing.

In the next few paragraphs we discuss three specific problems in bifurcation theory that are solved by the singularity theory approach; we also discuss how this information is useful for applications. The first problem, called the *recognition problem*, is the following: Given an equation $h(x, \lambda) = 0$, when is a second equation $g(x, \lambda) = 0$ equivalent to $h(x, \lambda) = 0$? In solving this problem, singularity theory methods produce a finite list of terms in the Taylor series of g such that the question of whether equivalence obtains is determined wholly by the values of the derivatives of g on this list—all other terms may be ignored. (Of course this list depends on the given function $h(x, \lambda)$; moreover for certain pathological functions $h(x, \lambda)$, a finite list does not suffice.) Regarding applications, this list specifies precisely the calculations which must be performed to recognize an equation of a given qualitative type. As we shall illustrate in Case Study 1, this information helps organize the computations for analyzing mathematical models.

The second problem concerns perturbations of an equation $g(x, \lambda) = 0$; i.e., equations of the form

$$g(x, \lambda) + p(x, \lambda) = 0, \tag{P.4}$$

where p is appropriately small. Specifically, the problem is to enumerate all qualitatively different perturbations of a given equation. We will solve this problem by constructing and analyzing what is called a universal unfolding. By way of definition, a *universal unfolding* of g is a distinguished k-parameter family of functions, $G(x, \lambda, \alpha)$, which satisfies (P.3) and has the following crucial property: For any small perturbation p, there is a value of α such that $g + p$ is equivalent to $G(\cdot, \cdot, \alpha)$. Less formally, up to qualitative equivalence, G contains all small perturbations of g.

Let us elaborate on point (i) above, the limitation that we consider only smooth functions of x, λ, and α. In constructing a universal unfolding of g, we will show that α in the universal unfolding and p in (P.4) are related by a smooth transformation. Nonetheless, a great deal of nonsmooth behavior is contained in a universal unfolding. Specifically, it is rarely possible, even locally, to express the solution x of (P.2) as a smooth, or even continuous, function of λ and α. The spirit of our approach is to work with smooth relationships between variables for as long as possible. Thus we attempt to solve (P.2) for x only after transforming the equation to a particularly tractable form; these transformations may be performed in a purely C^∞ context.

Our work with universal unfoldings has two additional benefits for applications. First, these methods often allow one to determine quasi-global properties of a model using purely local methods (cf. point (i) above); and second, in multiparameter models, these methods impose a structure on the physical parameter space that is useful as a guide in thinking about the problem. Both these benefits are illustrated in Case Studies 1 and 2.

The third problem is to classify the qualitatively different equations $g(x, \lambda) = 0$ that may occur. This is a problem of infinite complexity in that there are infinitely many equation types and there are equation types of arbitrarily high complexity. The singularity theory notion of codimension provides a rational approach to this problem. The *codimension* of g is the number of parameters needed in a universal unfolding of g; this notion also provides a rough measure of the likelihood of an equation of a given qualitative type appearing in a mathematical model, equations with lower codimensions being more likely. Of course we do not solve the classification problem completely. In this book we list all the qualitative types of equations having codimension three or less, along with all the qualitatively different perturbations of each. (See Chapter IV.) It is possible to extend the classification to higher codimensions, but the effort required escalates rapidly.

Our list of qualitative types of equations and their perturbations includes graphs of the solution sets, which we call *bifurcation diagrams*. These diagrams may be used in applications as follows. Consider a physical problem which depends on one or more auxiliary parameters. Suppose that for various values of the parameters one can generate representative bifurcation diagrams for the problem either by experiment or by numerical solution of a model. Suppose further that comparison with our lists shows that the bifurcation diagrams so generated are many or all of the qualitatively different perturbations of one specific qualitative type of equation, say $g(x, \lambda) = 0$. Then it is natural to conjecture that for some special values of the parameters an equation equivalent to this g results. To verify such a conjecture one needs to solve a recognition problem, as was discussed above. If the conjecture is verified, then the physical parameter space inherits useful structure from the universal unfolding, as was also discussed above. Typically this sequence of events leads to a compact description of a great deal of data. Following Thom, we use the term *organizing center* to describe an equation type occurring in this way; i.e., an equation which occurs in a model for certain values of the parameters such that the universal unfolding of this equation generates many or all of the bifurcation diagrams occurring in the physical problem. Each of the case studies illustrates the use of this concept in applications.

We now outline the contents of this book, chapter by chapter. Chapters II–IV, the essential theoretical core of the book, are a unit which develops the main ideas of the theory. These three chapters deal with the three problems discussed above; i.e., Chapters II, III, and IV study the recognition problem, unfolding theory, and the classification problem, respectively.

Chapter I highlights the theory to follow and discusses how singularity theory methods are used in applications. Also in this chapter we introduce the Liapunov–Schmidt reduction in the limited context of ordinary differential equations. (As we indicated above, with this technique many problems involving multiple solutions can be reduced to a single scalar equation $g(x, \lambda) = 0$.) The style of the chapter is mainly expository, developing ideas by means of examples rather than theory.

Chapter V explores a theoretical issue that singularity theory methods raise, the subject of moduli. Moduli are currently an active topic of research in several areas of pure mathematics. Regarding applications, moduli might at first seem to be an esoteric subject, but as illustrated by Case Studies 2 and 3, we have found moduli to play an important role in the more interesting applications we have studied. (*Remark*: Chapter V considers moduli in the simplest context—one state variable with no symmetry present. In applications, including Case Studies 2 and 3, moduli usually arise in a richer context involving symmetry.)

Symmetry and its consequences are the focus of Chapter VI. The restriction to one state variable greatly simplifies the discussion of symmetry since in one variable there is only one nontrivial symmetry possible. Thus in this chapter we are able to illustrate, with a minimum of technical complications, the main issues involving symmetry. (The full complexities of symmetry will be studied in Volume II.) In particular, one point illustrated by Chapter VI is how singularity theory methods unify different contexts—this chapter uses the same methods as are used in the unsymmetric context of the preceding chapters, even though the specific results in Chapter VI are quite different from those of earlier chapters.

Chapter VII develops the Liapunov–Schmidt reduction in general, expanding on the limited treatment in Chapter I. In this chapter we also illustrate the use of this reduction in applications—specifically, in a buckling problem and in certain reaction–diffusion equations.

Chapter VIII studies Hopf bifurcation for systems of ordinary differential equations; i.e., bifurcation of a periodic solution from a steady-state solution. This dynamical problem can be formulated as a steady-state problem, thereby permitting the application of singularity theory methods. The advantage of this approach lies in the fact that these methods generalize easily to handle degenerate cases where one or more hypotheses of the classical Hopf theorem fail.

Chapters IX and X together serve as a preview of the main issues to be studied in Volume II—bifurcation problems in several state variables, especially with symmetry. The simplest bifurcation problems in two state variables are discussed in Chapter IX, and certain bifurcation problems in two state variables with symmetry are discussed in Chapter X. The treatment of these subjects is not complete; in particular, several proofs are deferred to Volume II.

The three case studies in this book form an important part of it—they illustrate how singularity theory methods are used in applications. We believe that the three problems analyzed in the case studies are of genuine scientific interest. (Other examples, of primarily pedagogical interest, have been included within various chapters. Volume II will contain several more case studies, treating technically more difficult problems.)

As to the interdependence of various parts, Chapters I–IV should be included in any serious effort to read the book. After this point there are some options. In particular, Chapters V, VI, and VII are largely independent of one another, although the latter part of Chapter VI is closely related to Chapter V. By contrast, Chapter VIII depends heavily on Chapters VI and VII. Chapter IX may be read immediately following Chapter IV. (There is some reason to do so, as Chapter IX completes a theoretical development begun in Chapter IV; viz., the classification of bifurcation problems of codimension three or less. Chapter IX eliminates the restriction to one state variable that was imposed in Chapter IV.) Chapter X draws primarily on Chapter VI. Each case study is placed immediately following the last chapter on which it depends.

In writing this book we wanted to make singularity theory methods available to applied scientists as well as to mathematicians—we have found these methods useful in studying applied bifurcation problems, especially those involving many parameters or symmetry, and we think others may too. Therefore we have tried to write the book in ways that would make it accessible to a wide audience. In particular, we have devoted much effort to explaining the underlying mathematics in relatively simple terms, and we have included many examples to illustrate important concepts and results. Several other features of the book also derive from our goal of increasing its readability. For example, each chapter and case study contains an introduction in which we summarize the issues to be addressed and the results to be derived. Likewise, in several places we have indicated material within a chapter that may be omitted without loss of continuity on a first reading, especially technically difficult material. In the same spirit, in cases where proofs are not central to the development, we have postponed these proofs, preferring first to discuss the theorems and give illustrations. Usually we have postponed proofs until the end of a section, occasionally until a later section, and in a few cases (the unfolding theorem among them) until Volume II.

The prerequisites for reading this book may seem to work against our goal of reaching a wide audience. Regarding mathematical prerequisites, the text draws on linear algebra, advanced calculus, and elementary aspects of the theory of ODE, commutative algebra, differential topology, group theory, and functional analysis. Except for linear algebra and advanced calculus, we attempt to explain the relevant ideas in the text or in the appendices. Thus we believe it is possible for a nonmathematical reader to gain an appreciation of the essentials of the theory, including how to apply it, provided he or she is comfortable with linear algebra and advanced calculus. The many examples should help greatly in this task.

Prerequisites for understanding the applications should not pose a problem for mathematical readers. Although our three case studies involve models drawn from chemical engineering, mathematical biology, and mechanics, in each case we have described the physical origins of the equations of the model and then analyzed these equations as mathematical entities. A mathematical reader could follow the analysis of the equations without understanding their origins; of course some physical intuition would thereby be sacrificed.

We are aware that many individuals whose work is not mentioned in this book have made important contributions to bifurcation theory. Consistent with our goals in writing this book, we have given references only when needed to support specific points in the text. Moreover by quoting one reference rather than another we do not mean to imply any historical precedent of one over the other—only that the quoted reference is one with which we are familiar and which establishes the point in question. The lack of a complete bibliography in this book is made less serious by the recent appearance of several monographs in bifurcation theory, for example, Carr [1981], Chow and Hale [1982], Guckenheimer and Holmes [1983], Hassard, Kazarinoff, and Wan [1981], Henry [1981], Iooss and Joseph [1981].

An amusing, personal anecdote may suggest further reasons why we have not attempted to include a complete bibliography. One of us was lecturing before an audience that included researchers in bifurcation theory. When asked to date a paper we had quoted, we guessed "around 1975." "It was in the early sixties!" came the prompt reply from someone in the audience who had been associated with the work. Like children everywhere, we find that events before our time are somewhat blurred.

There remains only the pleasant duty of thanking the many people who have contributed in one way or another to the preparation of this volume. Dave Sattinger originally suggested applying singularity theory methods to bifurcation problems. Jim Damon has been a frequent consultant on the intricacies of singularity theory; moreover Lemma 2.7 of Chapter III is due to him. Encouragement by, advice from, and lively discussion with John Guckenheimer, Jerry Marsden, and Ian Stewart have been most helpful. Joint work with our coauthors Barbara Keyfitz and Bill Langford are included in this text. The manuscript benefited greatly from suggestions made by Joe Fehribach and Ian Stewart. Barbara Keyfitz has contributed to the book in more ways than can reasonably be enumerated. To all these people, and to Giles Auchmuty, Vemuri Balakotaiah, Charlies Conley, Mike Crandall, Jack Hale, Phil Holmes, Ed Ihrig, Dan Luss, and Ed Reiss, we express our heartfelt thanks. The figures were drawn by Jim Villareal and Wendy Puntenney. While pursuing the research reported in this text, we were generously supported by the NSF and ARO, including visiting positions at the Courant Institute, the Institute for Advanced Study, the Mathematics Research Center, and the Université de Nice. Finally we are grateful to Bonnie Farrell for her most efficient typing of an illegible manuscript—we only wish that we might have written the book as quickly, accurately, and cheerfully as she typed it.

Contents

CHAPTER I

A Brief Introduction to the Central Ideas of the Theory

§0. Introduction

In this book we shall study local bifurcation problems with one state variable. Such problems may be formulated as an equation

$$g(x, \lambda) = 0 \tag{0.1}$$

for a single unknown x, the *state variable*, where the equation depends on an auxiliary parameter λ, the *bifurcation parameter*. We shall call the set of (x, λ) satisfying (0.1) the *bifurcation diagram* or *solution set* of g. The central questions about (0.1) concern multiple solutions. For each λ, let $n(\lambda)$ be the number of x's for which (x, λ) is a solution of (0.1). Our study of (0.1) will be local; thus we suppose that (0.1) may only be defined in some neighborhood of a point (x_0, λ_0) and that $n(\lambda)$ only counts solutions in this neighborhood. To avoid trivialities we assume that $g(x_0, \lambda_0) = 0$. Classically, one calls (x_0, λ_0) a *bifurcation point* if $n(\lambda)$ changes as λ varies in the neighborhood of λ_0. (*Remark*: Our theory makes liberal use of the derivatives of g; for simplicity we assume throughout that this function is infinitely differentiable.)

A surprising variety of the problems in applied mathematics which exhibit multiple steady-state solutions, even systems with infinitely many degrees of freedom, can be reduced to the form (0.1) by the so-called Liapunov–Schmidt reduction. We will illustrate this technique in §3 of this chapter and study it in earnest in Chapter VII; however, for the moment we take (0.1) as the basic datum.

The implicit function theorem (see Appendix 1) gives a simple necessary condition for (x_0, λ_0) to be a bifurcation point; namely $g_x(x_0, \lambda_0) = 0$. (Here the subscript x indicates partial differentiation.) For if $g_x(x_0, \lambda_0) \neq 0$, then (0.1) may be uniquely solved in the small for x as a function of λ; in other words, for each λ near λ_0 there is exactly one solution of (0.1) close to x_0. We shall call a point (x_0, λ_0) for which

$$g(x_0, \lambda_0) = g_x(x_0, \lambda_0) = 0 \tag{0.2}$$

a *singularity*.

Note that a singularity need not be a bifurcation point in the classical sense. For example, consider

$$x^3 + \lambda^2 = 0,$$

which has exactly one solution (viz., $x = -\lambda^{2/3}$) for any λ, but is obviously singular at the origin. (*Remark*: This example is quite important for the physical problem that we consider in §2 below.)

This chapter is divided into four sections. Section 1 is a theoretical section; in it we discuss the information that singularity theory methods provide about the pitchfork bifurcation, perhaps the most important example of bifurcation in the classical literature. In §2 we consider the application of singularity theory methods to a chemical engineering model. In §3 we introduce in a special case the Liapunov–Schmidt reduction mentioned above. Finally in §4 we analyze the relation between the Liapunov–Schmidt reduction and the stability of equilibrium solutions of an autonomous system of ordinary differential equations.

Sections 1–3 introduce the three major themes that occur throughout this volume. Section 1 leads naturally into the theoretical side of the subject, which we begin to develop in Chapters II and III. Section 2 is indicative of the applications in the Case Studies. The third section is concerned with the issue of how the study of equations as simple as (0.1) can have such wide applicability; we return to this theme in Chapter VII. By contrast, the material in §4 lies outside the mainstream of this text, although it is extremely important for bifurcation theory in general.

§1. The Pitchfork Bifurcation

In this section we discuss the pitchfork bifurcation from the singularity theory point of view. This bifurcation occurs frequently in the classical literature. It has the basic property that as λ crosses some value λ_0, the number of solutions $n(\lambda)$ jumps from one to three. The simplest equation with this behavior is

$$x^3 - \lambda x = 0, \tag{1.1}$$

where $\lambda_0 = 0$. The solution set for (1.1) is shown in Figure 1.1, which explains the nomenclature. In the figure the orientation of the coordinate axes is shown to the right.

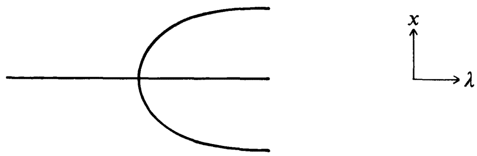


Figure 1.1. The pitchfork bifurcation $x^3 - \lambda x = 0$.

We have divided this section into five subsections. In subsection (a) we present a simple mechanical system which illustrates the pitchfork bifurcation. In subsections (b) and (c) we discuss the information that singularity theory methods provide about the pitchfork bifurcation. The fundamental ideas of the theory already occur in this simple example. In particular, Chapters II and III develop the ideas in subsections (b) and (c), respectively; i.e., they extend these ideas to a general context and supply proofs. Finally, in subsections (d) and (e) we consider certain related issues needed to understand the significance of bifurcation theory for applications. This latter material is a standard part of bifurcation theory, and not at all tied to the singularity theory approach. We present it here because it gives the subject vitality by making connections with applications.

(a) An Example of the Pitchfork Bifurcation

In Figure 1.2 we illustrate a simple physical system which exhibits a pitchfork bifurcation. (This is a finite element analogue of the Euler column which we will study in Chapter VII, §2.) The system consists of two rigid rods of unit length connected by pins which permit rotation in a plane; it is subjected to a compressive force λ which is resisted by a torsional spring of unit strength. We neglect friction. The state of the system is described by the angle x measuring the deviation of the rods from the horizontal. The potential energy of this system equals

$$V(x, \lambda) = \frac{x^2}{2} + 2\lambda(\cos x - 1),$$

the first term representing the stored energy in the torsional spring and the second, the work done by the external force. Steady states are described by an

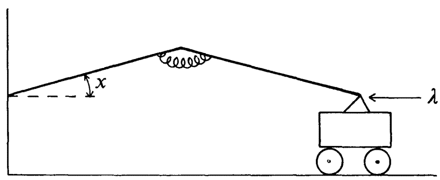


Figure 1.2. Finite element analogue of Euler buckling.

equation of the form (0.1), where

$$g(x, \lambda) = \frac{\partial V}{\partial x}(x, \lambda) = x - 2\lambda \sin x. \tag{1.2}$$

We ask the reader to check that the function (1.2) has a singularity at $(x_0, \lambda_0) = (0, \frac{1}{2})$ and that modulo higher-order terms (hot)

$$g(x, \lambda) = \frac{x^3}{6} - 2(\lambda - \tfrac{1}{2})x + \text{hot}, \tag{1.3}$$

where the neglected terms are of order x^5, $(\lambda - \frac{1}{2})^2 x$, or higher. Equation (1.3) bears a strong similarity to (1.1) but differs in three respects: first, the singularity in (1.3) is not located at the origin; second, the coefficients in (1.3) differ from unity; and third, there are higher-order terms present in (1.3). The first two differences may be absorbed by simple linear changes of coordinates; i.e., by replacing λ by $a(\lambda - \frac{1}{2})$ and x by bx for appropriate constants a and b. These two differences have no effect on the *qualitative* picture of the solution set. As for the higher-order terms, below we will describe results which show that they may be absorbed by a nonlinear change of coordinate. In other words, the third difference also has no effect on the *qualitative* behavior of the solution set *in the small.*

Let us elaborate. Formula (1.4) below provides a sufficient condition for $n(\lambda)$, the number of solutions of $g(x, \lambda) = 0$, to jump from one to three as λ crosses λ_0. We ask the reader to verify that (1.2) satisfies (1.4) at $(x, \lambda) = (0, \frac{1}{2})$. This will show that there are three equilibrium configurations of the model in Figure 1.2 when $\lambda > \frac{1}{2}$, the trivial (or undeformed) state $x = 0$ and two nontrivial (or buckled) states with $x \neq 0$. (To relate this result to the physical system, it is important to realize that for $\lambda > \frac{1}{2}$ the undeformed state $x = 0$ is unstable and hence effectively unobservable in experiments. We will discuss stability briefly in subsection (d) below, and more thoroughly in §4 of this chapter.)

(b) Finite Determinacy and the Recognition Problem

There are two basic issues on which the singularity theory approach to bifurcation focuses; the first of these concerns questions of the type just encountered; i.e., questions about the importance of higher-order terms. We shall use the singularity theory term *finite determinacy* to describe such problems, since the underlying question may be phrased: "To what extent do the low-order terms in the Taylor series expansion of a bifurcation problem $g(x, \lambda)$ determine its qualitative behavior, regardless of the higher-order terms that may be present?" For the particular case of the pitchfork, our answer, in part, is as follows. Let $g(x, \lambda)$ be a bifurcation problem such that when $(x, \lambda) = (x_0, \lambda_0)$ we have

$$g = g_x = g_{xx} = g_\lambda = 0, \qquad g_{xxx} g_{\lambda x} < 0; \tag{1.4}$$

then $n(\lambda)$, the number of solutions of $g(x, \lambda) = 0$, jumps from one to three as λ crosses λ_0. (If $g_{xxx}g_{\lambda x}$ has the opposite sign, $n(\lambda)$ jumps from three to one, while if $g_{xxx}g_{\lambda x} = 0$, more information is required.)

A direct and fairly elementary proof of this result is possible, using only the implicit function theorem; this proof is outlined in Exercise 1.1. In the singularity theory approach, however, one proves considerably more—namely, that any g satisfying (1.4) may be transformed by an appropriate change of coordinates into the standard model for the pitchfork, $x^3 - \lambda x = 0$. More precisely, if $g(x, \lambda)$ satisfies (1.4) at (x_0, λ_0), then there exist:

(i) a local diffeomorphism of $\mathbb{R}^2$ of the form $(x, \lambda) \to (X(x, \lambda), \Lambda(\lambda))$ mapping the origin to (x_0, λ_0); and
(ii) a nonzero function $S(x, \lambda)$;

such that

$$S(x, \lambda)g(X(x, \lambda), \Lambda(\lambda)) = x^3 - \lambda x \tag{1.5}$$

near the origin, where, moreover, $X_x(x, \lambda) > 0$ and $\Lambda'(\lambda) > 0$. Since the factor $S(x, \lambda)$ is nonzero, the solutions of $g(x, \lambda) = 0$ differ from those of $x^3 - \lambda x = 0$ only by the diffeomorphism (X, Λ). This is the precise sense in which the higher-order terms in (1.3) have no effect on the qualitative behavior of the model in the small—they may be transformed away entirely by a change of coordinates.

Equation (1.5) leads to the definition of equivalence, which is one of the fundamental concepts in the theory. We shall say that two bifurcation problems g and h are *equivalent* if they may be related through an equation

$$S(x, \lambda)g(X(x, \lambda), \Lambda(\lambda)) = h(x, \lambda), \tag{1.6}$$

where S is nonzero and positive and (X, Λ) is a local diffeomorphism which, as above, preserves the orientations of x and λ. Note that this definition requires that $S > 0$; as we shall explain in §4 below, this convention preserves useful information about the stability of solutions. Unfortunately this convention also leads to some nuisances regarding plus and minus signs. For example, with this convention we must decompose (1.4) into two distinct cases, depending on which factor in the product $g_{xxx}g_{\lambda x}$ is negative.

If g and h are equivalent, then the two multiplicity functions are related as follows:

$$n_g(\Lambda(\lambda)) = n_h(\lambda). \tag{1.7}$$

Indeed, (1.7) is one of the most important consequences of equivalence. It turns out that this equation is intimately related to our restriction that in the diffeomorphism (X, Λ), the second coordinate Λ may not depend on x. We explore this relationship in Exercise 1.4; here we only motivate this restriction by the remark that typically in applications λ is associated with an external force set by the experimenter, while x is associated with an internal state of the system that results from the choice of λ. In other words, λ influences x, but x does not influence λ. Coordinate transformations of the form $(X(x, \lambda), \Lambda(\lambda))$ reflect this distinction.

Our treatment of the pitchfork is representative of the general singularity theory approach to determinacy questions. Let us summarize the above discussion as a way of introducing the terminology of Chapter II. We shall call $x^3 - \lambda x = 0$ a *normal form* for the pitchfork bifurcation. Any bifurcation problem $g(x, \lambda)$ which at a specific point (x_0, λ_0) satisfies

$$g = g_x = g_{xx} = g_\lambda = 0; \qquad g_{xxx} > 0, \quad g_{\lambda x} < 0 \tag{1.8}$$

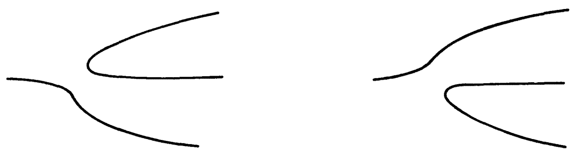
is equivalent to this normal form. We shall say that (1.8) solves the *recognition problem* for this normal form; i.e., (1.8) characterizes the bifurcation problems equivalent to $x^3 - \lambda x = 0$. Equivalent bifurcation problems have the same qualitative properties; more precisely, qualitative properties are those which are unchanged by equivalence. The object of this book is to study qualitative properties of bifurcation problems.

(c) Universal Unfoldings and Perturbed Bifurcation

The second of the two central issues in our approach to bifurcation theory arises from the study of how bifurcation problems may depend on parameters. In a bifurcation problem $g(x, \lambda)$, small variations of an auxiliary parameter usually lead to dramatic changes in the bifurcation diagram at a singularity of g. As an illustration of this phenomenon, let us consider the perturbed pitchfork

$$G(x, \lambda, \varepsilon) = x^3 - \lambda x + \varepsilon = 0. \tag{1.9}$$

The bifurcation diagrams of (1.9) with $\varepsilon \neq 0$ are shown in Figure 1.3. Complete derivations of these and other graphs are deferred until Chapter III, but the following intuitive considerations may be helpful. The unperturbed equation $G(x, \lambda, 0) = 0$ is nonsingular away from the origin; i.e., $G_x(x, \lambda, 0) = 3x^2 - \lambda$ is nonzero on both solution branches $\{x = 0\}$ and $\{\lambda = x^2\}$ except at the origin. Thus by the implicit function theorem the solution x of (1.9) depends smoothly (and hence continuously) on ε (as well as λ) away from the origin. In other words, away from the origin the diagrams in Figure 1.3 must



$\varepsilon > 0$ $\qquad\qquad$ $\varepsilon < 0$

Figure 1.3. Perturbations of the pitchfork, $x^3 - \lambda x + \varepsilon = 0$.

closely resemble the pitchfork (Figure 1.1) for which $\varepsilon = 0$. On the other hand, near the origin the quadratic term λx will be more important than the cubic term x^3. Thus near the origin the graphs of Figure 1.3 should resemble the hyperbola $\lambda x = \varepsilon$.

In the classical literature there appear to be two distinct ways in which auxiliary parameters arise in bifurcation problems. Often the original formulation of a physical model involves many auxiliary parameters, as is the case, for example, in the stirred reactor problem to be studied in §2. In other cases, however, the parameters arise from the more subtle issue of *imperfect bifurcation*. Let us elaborate. The mathematical equations which result from the choice of a model for a physical phenomenon are invariably an idealization; a more complete description would almost surely lead to a slightly perturbed set of equations. These deviations of the actual situation from the idealized one—imperfections—may be described by auxiliary parameters in the equations.

Let us illustrate how imperfect bifurcation might introduce parameters into the buckling model of Figure 1.2. One natural perturbation to consider is a small vertical force ε applied to the center pin (See Figure 1.4(a)); this force models the weight of the structure. Another such perturbation comes from imagining that the torsional spring is slightly asymmetric, exerting zero torque when $x = \delta$ rather than when $x = 0$ (See Figure 1.4(b)). The potential function with these two perturbations present is

$$V(x, \lambda, \varepsilon, \delta) = \frac{(x - \delta)^2}{2} + 2\lambda(\cos x - 1) + \varepsilon \sin x,$$

and the equilibrium equation is

$$x - \delta - 2\lambda \sin x + \varepsilon \cos x = 0. \tag{1.10}$$

Note that (1.10) is a perturbation of (1.2) depending on *two* auxiliary parameters. Near the singularity of the unperturbed problem at $x = 0$, $\lambda = \frac{1}{2}$, (1.10) has the expansion

$$\frac{x^3}{6} - 2(\lambda - \tfrac{1}{2})x + (\varepsilon - \delta) - \frac{\varepsilon}{2}x^2 + \text{hot}. \tag{1.11}$$

In the singularity theory approach, the occurrence of parameters is handled as follows. One knows that auxiliary parameters are normally an

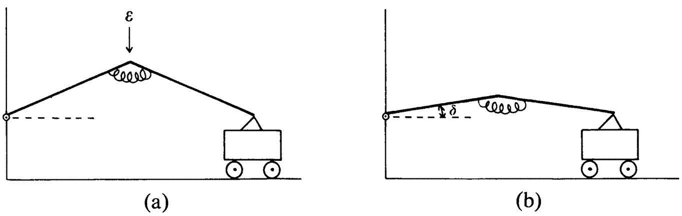


Figure 1.4. Two possible imperfections in the buckling model.

important part of a bifurcation problem, and one attempts to classify all possible behavior that can occur as a result of their presence. This problem is solved in two steps. Given a bifurcation problem g, the first step is to construct a certain distinguished family of perturbations of g. Let us elaborate. Suppose that $G(x, \lambda, \alpha_1, \ldots, \alpha_k)$ is a k-parameter family of bifurcation problems; we shall call G a *perturbation* of g if

$$G(x, \lambda, 0, \ldots, 0) = g(x, \lambda). \tag{1.12}$$

Of course, $G(x, \lambda, \alpha_1, \ldots, \alpha_k)$, which we abbreviate to $G(x, \lambda, \alpha)$, need only be defined for α close to the origin in $\mathbb{R}^k$. In the first step of solving the classification problem we seek a k-parameter family G of perturbations of g with the distinguishing property that any perturbation of g whatsoever is equivalent to $G(\cdot, \cdot, \alpha)$ for some $\alpha \in \mathbb{R}^k$ near the origin. In other words, given any perturbing term $\varepsilon p(x, \lambda, \varepsilon)$, there are parameter values $\alpha_1(\varepsilon), \ldots, \alpha_k(\varepsilon)$ such that for small ε

$$g + \varepsilon p \sim G(\cdot, \alpha(\varepsilon)),$$

where $\sim$ denotes equivalent i the sense of (1.6). We shall call such a G a *universal unfolding* of g. Not every bifurcation problem admits a universal unfolding, but, in a sense we shall clarify in Chapter III, most do. The number k of parameters required for a universal unfolding depends on the specific function g under consideration. For example, we will show in Chapter III that

$$G(x, \lambda, \alpha) = x^3 - \lambda x + \alpha_1 + \alpha_2 x^2 \tag{1.13}$$

is a universal unfolding of the pitchfork. (*Remark*: In specific physical models it is possible to relate the mathematical parameters of the universal unfolding G to physical parameters in the problem, although in realistic applications this often requires rather tedious calculations. For the beam model above,

$$\alpha_1 = -\varepsilon, \qquad \alpha_2 = \varepsilon - \delta,$$

provides such a correspondence modulo higher-order terms, as (1.11) might suggest.)

The second step in solving the classification problem is to explore the parameter space $\mathbb{R}^k$ of the universal unfolding with the goal of enumerating the various bifurcation diagrams

$$\{(x, \lambda) : G(x, \lambda, \alpha) = 0\}$$

that can occur as α varies. For the universal unfolding (1.13) of the pitchfork, there are essentially four different bifurcation diagrams which can occur as α varies; these diagrams are illustrated in Figure 1.5. This figure also indicates how the bifurcation diagram depends on α—the α_1, α_2 plane is divided into four regions by the two curves $\alpha_1 = 0$ and $\alpha_1 = \alpha_2^3/27$, and equivalent diagrams are obtained for all α within a given region. Proofs will be given in Chapter III.

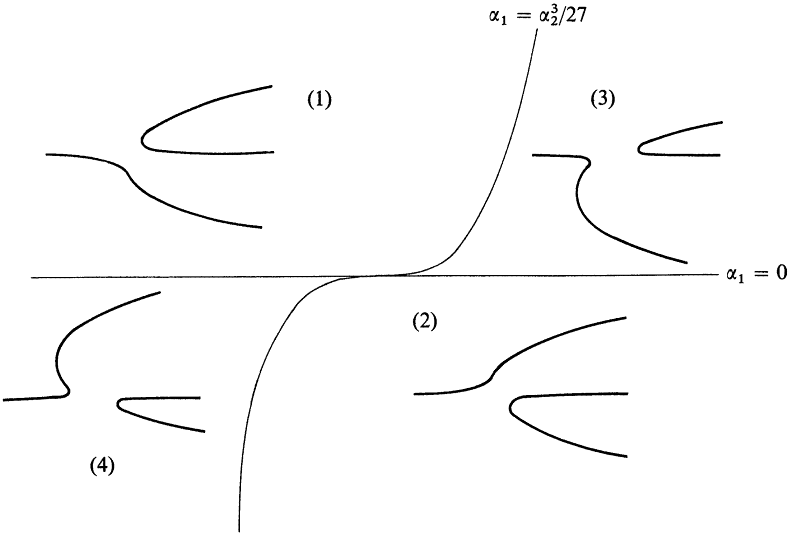


Figure 1.5. Universal unfolding of the pitchfork.

The bifurcation diagrams of regions 1 and 2 in Figure 1.5 already occurred in Figure 1.3 above. Let us discuss briefly the more complicated diagrams of regions 3 and 4. First consider (1.13) with $\alpha_1 = 0$ but $\alpha_2 > 0$. Solving the equations explicitly, we find the diagram of Figure 1.6. By choosing α_1 nonzero, we will now split apart the two crossed curves of Figure 1.6. If α_1 is chosen positive and sufficiently small (more precisely, if $0 < \alpha_1 < \alpha_2^3/27$), the primary solution branch will have a kink, as in region 3 of Figure 1.5. (By the *primary branch* we mean the solution branch which connects to the unique solution branch that exists for $\lambda \ll 0$.)

With these complications in mind, the reader may well wonder what might result from introducing three or more parameters into the model. In fact, no new behavior would occur if more parameters were introduced. This fact is a consequence of our assertion that (1.13), which contains two parameters, is a universal unfolding of the pitchfork. Singularity theory methods tell the exact number of parameters required to describe the most general perturbation of a bifurcation problem—this is one of the theory's achievements.

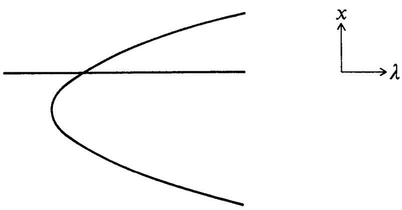


Figure 1.6. Bifurcation along the boundary between regions (1) and (2) in Figure 1.5: $\alpha_1 = 0$ and $\alpha_2 > 0$.

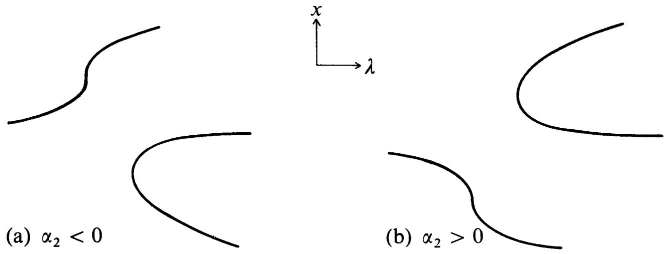


Figure 1.7. Bifurcation along the boundary between two regions in Figure 1.5: $\alpha_1 = \alpha_2^3/27$.

It is instructive to consider the dividing cases in Figure 1.5; i.e., those which occur for (α_1, α_2) along the two curves $\alpha_1 = 0$ and $\alpha_1 = \alpha_2^3/27$. Actually the first case, $\alpha_1 = 0$, has already been considered above—it is graphed in Figure 1.6 for $\alpha_2 > 0$. Note that the origin is a point of bifurcation. The bifurcation diagrams for the second case, $\alpha_1 = \alpha_2^3/27$, are illustrated in Figure 1.7. We refer to the points on these graphs with vertical tangents as *hysteresis points.* (For the origin of this term, see subsection (e) below.)

The pitchfork is an informative example because it is the simplest singularity exhibiting both bifurcation and hysteresis. There is precisely one singularity which exhibits neither bifurcation nor hysteresis, the *limit point.* As we will show in Chapter II, such a singularity is defined by the equations

$$g = g_x = 0; \qquad g_{xx} \neq 0, \qquad g_\lambda \neq 0 \quad \text{at } (0, 0),$$

and is equivalent to the normal form $\pm x^2 \pm \lambda$ for some choice of signs. The only singularities in the bifurcation diagrams in Figure 1.5 are limit points; as we shall see below, this occurrence is a special case of a general phenomenon.

(d) Stability

In applications, equations of the form $g(x, \lambda) = 0$ arise in describing the equilibria of some physical system. The notion of the stability of such equilibria lies outside the scope of a steady-state theory; stability can only be defined in a theory which follows the time evolution of the system. In this subsection we briefly discuss stability in the context of the mechanical system of Figure 1.2. (See §4 for a more general analysis of this concept.)

For this mechanical model, Newton's equations of motion are the appropriate dynamical theory. Let us write

$$M\ddot{x} = -\frac{\partial V}{\partial x}(x, \lambda) - C\dot{x}, \tag{1.14}$$

where M and C are positive constants. (*Remark*: Here we assume a dynamic frictional force proportional to $\dot{x}$, and for simplicity we make the small angle approximation in the inertial term so that the effective moment of inertia M is independent of x.) Note that a constant function $x(t) = x_0$ satisfies (1.14) iff $\partial V/\partial x(x_0, \lambda_0) = 0$; in other words, equilibrium solutions of (1.14) are characterized by $x - 2\lambda \sin x = 0$. (Cf. (1.2).)

We shall call an equilibrium solution x_0 of (1.14) *asymptotically stable* if, for all sufficiently small ε_i, the solution of (1.14) with perturbed initial data

$$x(0) = x_0 + \varepsilon_1, \qquad \dot{x}(0) = \varepsilon_2$$

decays to x_0 as $t \to \infty$. Otherwise we call x_0 *unstable*. (*Remark*: The reader should note the contrast between this subsection and subsection (c). Here we are considering perturbations of the initial data in an evolution equation. In subsection (c) we were considering perturbation of the *equation* describing equilibrium.)

There is a natural sufficient condition for stability which only involves the sign of V_{xx}. Let us define a new variable $y = \dot{x}$ and rewrite (1.14) as a first-order system

$$\begin{gathered} \dot{x} - y = 0, \\ \dot{y} + \frac{1}{M}\left\{\frac{\partial V}{\partial x}(x, \lambda) + Cy\right\} = 0. \end{gathered} \tag{1.15}$$

The Jacobian of the function in this ODE is

$$\begin{pmatrix} 0 & -1 \\ \dfrac{1}{M} V_{xx} & \dfrac{C}{M} \end{pmatrix}. \tag{1.16}$$

According to the results of §§1 and 2 in Chapter 9 of Hirsch and Smale [1974], an equilibrium solution $(x, y) = (x_0, 0)$ of (1.15) is asymptotically stable if both eigenvalues of the matrix (1.16) have positive real parts, and unstable if at least one eigenvalue has a negative real part. The eigenvalues of the Jacobian (1.16) are

$$\mu_\pm = \frac{1}{M}\left\{\pm\sqrt{\frac{C^2}{4} - MV_{xx}} + \frac{C}{2}\right\}.$$

Both eigenvalues lie in the right half plane iff

$$\frac{\partial^2 V}{\partial x^2}(x_0, \lambda_0) > 0, \tag{1.17}$$

while μ_- lies in the left half plane if V_{xx} is negative. According to the results mentioned above, x_0 is a stable equilibrium point if (1.17) holds, and unstable if the opposite sign prevails.

Note that (1.17) is precisely the condition for the potential $V(\cdot, \lambda_0)$ to have a nondegenerate local minimum at x_0. Thus x_0 is asymptotically stable if $V(\cdot, \lambda_0)$ has a nondegenerate local minimum at x_0. Similarly x_0 is unstable if $V(\cdot, \lambda_0)$ has a nondegenerate local maximum at x_0.

Let us relate the above discussion to subsection (a). Equilibrium points of (1.14) are characterized by the equation $g(x, \lambda) = 0$ where $g(x, \lambda) = V_x(x, \lambda)$. Thus an equilibrium point x_0 is asymptotically stable if $g_x(x_0, \lambda_0) > 0$ and unstable if $g_x(x_0, \lambda_0) < 0$. The borderline case where $g_x = V_{xx} = 0$ occurs precisely when g has a singularity, as defined in subsection (a). In §4 we will relate stability and singularities in a more general context.

Let us apply this criterion for stability to the model of Figure 1.2, first assuming no imperfections. We have

$$g_x(x, \lambda) = 1 - 2\lambda \cos x. \tag{1.18}$$

Now as we saw above, $g(x, \lambda) = x - 2\lambda \sin x$ has two solution branches—the trivial branch $x = 0$ and a nontrivial branch where

$$\lambda = \frac{1}{2} + \frac{x^2}{12} + \text{hot}, \tag{1.19}$$

the latter coming from (1.3). For the trivial solution $x = 0$, we have

$$g_x(x, \lambda) = 1 - 2\lambda.$$

Thus $x = 0$ is asymptotically stable for $\lambda < \frac{1}{2}$ and asymptotically unstable for $\lambda > \frac{1}{2}$. For the nontrivial solution we substitute (1.19) into (1.18) to obtain $g_x = x^2/3 + \text{hot}$. Thus the nontrivial solution is asymptotically stable for all small $x \neq 0$. There is, of course, an intimate relationship between the fact that the trivial solution loses stability as λ crosses $\frac{1}{2}$ and that stable, new solutions appear (bifurcate) at this point. Classically this phenomenon was described by the phrase *exchange of stability*.

In Figure 1.8(a) we have indicated the stability assignments for the pitchfork as just determined, using dashed lines for unstable solutions and solid lines for stable solutions. The bifurcation diagram in Figure 1.8(a) divides the $x\lambda$-plane into four regions. Since the bifurcation diagram is the zero set of g, the sign of g is constant in each of these four regions. We have indicated these signs in Figure 1.8(a). Let us show how to use these signs in determining the stability assignments when imperfections are present. We have the following rule: g_x is positive (and stability prevails) along a portion of the bifurcation diagram where a region with $g > 0$ lies above a region

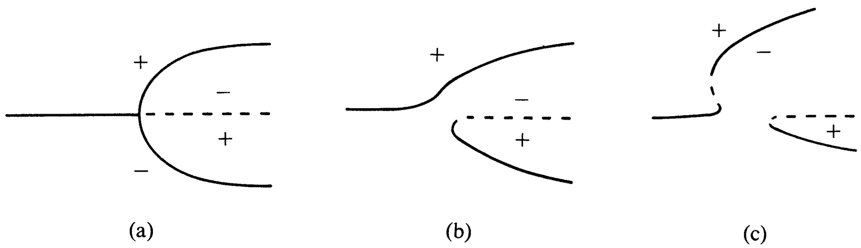


Figure 1.8. Stability assignments for the pitchfork and for typical perturbations.

where $g < 0$; g_x is negative (and instability prevails) when the situation is reversed. In Figure 1.8(b) and (c) we have indicated typical perturbations of Figure 1.8(a), corresponding to regions 2 and 4 in Figure 1.5, respectively. We ask the reader to verify the stability assignments on these two diagrams by applying the above rule.

In experiments, a solution which is unstable is effectively unobservable. This is because in an experiment there are always uncontrollable, even if small, perturbations from the idealized situation. In the unstable case these perturbations grow in time causing the system to leave any neighborhood of the equilibrium point in question. In fact, our definition of stability, which allows only a single small perturbation of the system at time zero, might be criticized as an inadequate model of the physical situation where repeated perturbations may be expected. However, questions about the long time asymptotic behavior of an autonomous system subjected to small random perturbations are very difficult indeed, and we leave such issues untouched.

(e) Quasi-Static Variation of Parameters

In this subsection we discuss a folklore interpretation of bifurcation diagrams which makes the subject much more lively. A bifurcation diagram displays the equilibrium states of a system (together with their stabilities) as a function of the bifurcation parameter λ. Under certain circumstances a bifurcation diagram can also describe the evolution in time of the system. In this subsection we discuss one such set of circumstances, called *quasi-static variation of parameters*.

For definiteness we base our discussion on the beam model of Figure 1.2 (possibly with imperfections). Imagine that the applied load is slowly varied with time; i.e., slowly compared to the relaxation time of the system. More picturesquely, imagine applying a small increase in the load, waiting until the system returns to a new equilibrium, then applying a second small increase in load and again waiting for re-equilibration, and so on. The behavior resulting from such variations in load depends crucially on whether or not the current equilibrium state of the system lies near a singularity of the bifurcation diagram.

We consider the nonsingular case first. Let x_0 and λ_0 be the current equilibrium state and load of the system, respectively, where, of course, x_0 is a stable equilibrium. In the nonsingular case there is a smooth branch of equilibrium points $x(\lambda)$ passing through (x_0, λ_0), and these are the only equilibria in the neighborhood of (x_0, λ_0). If the load is increased to $\lambda_0 + \Delta\lambda$, the system will be out of equilibrium, but its initial data (namely x_0) will lie within the basin of attraction of the equilibrium at $x(\lambda_0 + \Delta\lambda)$. Thus the system will settle into this new, close-by equilibrium. In other words, under quasi-static variation of the load, the system will simply move along regular portions of the bifurcation diagram.

At a singularity, however, quite a variety of behavior is possible. Let us consider several examples. First we consider quasi-static variation of λ in the idealized beam model (i.e., $\varepsilon = \delta = 0$). The bifurcation diagram for this system is pictured in Figure 1.8(a), with the bifurcation point at $\lambda = \frac{1}{2}$. For $\lambda < \frac{1}{2}$ the system will follow the trivial solution branch $x = 0$, but for $\lambda > \frac{1}{2}$ it will follow one of the two nontrivial solution branches. (Exactly which branch it will follow is indeterminate in the present case—this is another illustration of the importance of imperfections.) Thus for Figure 1.8(a), the derivative of the observed solution with respect to λ is discontinuous at the bifurcation point, although the solution itself is continuous.

Next we consider quasi-static variation of λ in the beam model with imperfections present (i.e., ε and δ nonzero). Let us suppose that the system is governed by the bifurcation diagram of Figure 1.8(b). Here the primary solution branch consists entirely of regular points, so the evolution under quasi-static variation of λ will be smooth. In words, the imperfections introduce a preferred direction into the system and smooth out the transition to buckled states. There are stable buckled states on the secondary solution branch, but these will never be reached by quasi-static variation of λ starting from small λ. However, these states can easily be reached by temporarily applying a large vertical force (once λ is sufficiently large), and then releasing the system in the neighborhood of the new equilibrium point. Once reached, these equilibria will endure, since they are stable. Now imagine pushing the system onto the secondary solution branch in this way and then *decreasing* λ quasi-statically. When the limit point is reached, a further decrease in λ will necessarily result in a large jump of the system, since there are no nearby equilibria. In other words, at a limit point the solution itself, not merely its derivative, will vary discontinuously with λ.

If ε, δ are such as to produce the bifurcation diagram in Figure 1.8(c), then there are limit points on the primary solution branch. Thus the solution x will undergo a jump even in the simplest experiment of increasing λ quasi-statically from zero. Note that jumps occur for different values of λ, depending on whether λ is being increased or decreased. This is similar to the hysteresis which occurs in magnetism and is the origin of our term *hysteresis point*, which describes the borderline case between the presence or absence of hysteresis.

We conclude this subsection with two remarks. The first concerns the behavior of n-dimensional systems (as opposed to 1-dimensional). What happens in several dimensions when, as in Figure 1.8(c), λ is increased beyond a limit point, causing the system to behave discontinuously? *A priori* there is no reason why such a system has to jump to another equilibrium; it is quite possible for the system to evolve to some sort of dynamic steady state such as a periodic orbit. Only by close examination of the dynamical equations can one decide this issue, although in this book we consider primarily cases where the new equilibrium is static. Second, in bifurcation theory it is customary to study the dependence of the solution on a distinguished parameter—λ in our notation. In our estimation, this practice has its origins in the interpretation of bifurcation diagrams as describing the evolution of a system under quasi-static variation of parameters. In effect the parameter λ is identified with time.

Exercises

1.1. Show that if g satisfies (1.4) at $(x_0, \lambda_0) = (0, 0)$ then there exists smooth functions $M(x, \lambda)$, $\psi(\lambda)$, and $\phi(x)$ defined on neighborhoods of the origin such that

$$g(x, \lambda) = (\lambda - \phi(x)x^2)(x - \psi(\lambda))M(x, \lambda), \tag{1.20}$$

where $\phi(0) > 0$, $\psi(0) = 0$, and $M(0, 0) \neq 0$. Conclude from (1.20) that

$$n_g(\lambda) = \begin{cases} 3 & \text{if } \lambda > 0, \\ 1 & \text{if } \lambda \leq 0. \end{cases}$$

Prove (1.20) by using the following sequence of hints.

(a) Let $s(x) = g(x, \mu x)$ for fixed μ. Using (1.4) show that $s(0) = s'(0) = 0$. Using Taylor's theorem conclude that $g(x, \mu x) = x^2 K(x, \mu)$.

(b) Show that $K(0, 0) = 0$, $K_x(0, 0) = g_{xxx}(0, 0)/6$ and $K_\mu(0, 0) = g_{\lambda x}(0, 0)$. Then use the implicit function theorem to find a smooth function $\mu(x)$ satisfying $K(x, \mu(x)) \equiv 0$, $\mu(0) = 0$, and $\mu'(0) > 0$.

(c) Use Taylor's theorem to conclude that $\mu(x) = x\phi(x)$ where $\phi(0) > 0$. Hence $g(x, x^2\phi(x)) \equiv 0$. Use Taylor's theorem again to show that

$$g(x, \lambda) = (\lambda - x^2\phi(x))L(x, \lambda).$$

(d) Show that $L(0, 0) = 0$ and $L_x(0, 0) \neq 0$. Apply the implicit function theorem to obtain a smooth function $\psi(\lambda)$ satisfying $L(\psi(\lambda), \lambda) \equiv 0$, $\psi(0) = 0$. Now apply Taylor's theorem to obtain (1.20).

Comment. Exercise 1.1 gives a "classical" proof of the pitchfork bifurcation. The basic idea in the proof is to construct the zero set of g by clever uses of the implicit function theorem and Taylor's theorem. Note that in order to apply such methods one has to know *a priori* certain qualitative information about the zero set of g; essentially one has to know that $x^3 - \lambda x$ is a good "model" for the general case.

1.2. Show that equivalence for bifurcation problems—as defined in (1.6)—is an equivalence relation. In particular, let $g(x, \lambda)$, $h(x, \lambda)$, and $k(x, \lambda)$ be bifurcation problems. Assume that g is equivalent to h and that h is equivalent to k. Then show that g is equivalent to k.

1.3. Using (1.20), prove that if $g(x, \lambda)$ satisfies (1.4) at (0, 0) then g is equivalent $x^3 - \lambda x$. Prove this fact by considering the following sequence of four equivalences.

(a) $g(x, \lambda)$ is equivalent to $h(x, \lambda) = (x - \psi(\lambda))(x^2\phi(x) - \lambda)$.

(b) $h(x, \lambda)$ is equivalent to $k(x, \lambda) = x^3\phi(x) - \lambda x\, q(x, \lambda)$ where $q(0, 0) = 1$.

(c) $k(x, \lambda)$ is equivalent to $l(x, \lambda) = x^3 p(x, \lambda) - \lambda x$ where $p(0, 0) > 0$.

(d) $l(x, \lambda)$ is equivalent to $x^3 - \lambda x$. Consider $X(x, \lambda) = x\sqrt{p(x, \lambda)}$ and evaluate $[1/\sqrt{p(x, \lambda)}](X^3(x, \lambda) - \lambda X(x, \lambda))$.

1.4. Show that the formula (1.7) is false if one allows Λ in (1.6) to depend on x by considering the following two examples.
(a) $g(x, \lambda) = x^2 - \lambda^2$, $X(x, \lambda) = x - \lambda$, and $\Lambda(x, \lambda) = \lambda + x$.
(b) $g(x, \lambda) = x^3 - \lambda x$ and $\Lambda(x, \lambda) = \lambda + x^2$.

1.5. Let $g(x, \lambda)$ and $h(x, \lambda)$ be equivalent, as in (1.6). Show that if h has a singularity at (x_0, λ_0) then g has a singularity at $(X(x_0, \lambda_0), \Lambda(\lambda_0))$.

§2. The Continuous Flow Stirred Tank Reactor (CSTR)

In this section we discuss the application of singularity theory methods to the continuous flow stirred tank reactor (CSTR), a model problem from chemical engineering that exhibits multiple solutions. This problem differs from many traditional problems of bifurcation theory in that there is no trivial solution branch. It gives an excellent illustration of how singularity theory methods may be applied in bifurcation problems; in particular, it leads to the important concept of an *organizing center*. The problem is easy to describe and has been the object of much study, but yet singularity theory methods were able to provide new insights. In this section we only attempt to survey the situation; all proofs are deferred for Chapters II and III and Case Study 1.

This section is divided into three subsections, as follows. In subsection (a) we derive the equations which govern the CSTR. In subsection (b) we show that these equations admit multiple solutions. Finally, in subsection (c) we discuss how singularity theory methods apply to this problem. The issues raised in subsection (c) are important in many applications of singularity theory methods, not just the CSTR.

(a) Formulation of the Governing Equations for the CSTR

Figure 2.1 gives a schematic diagram for a continuous stirred tank reactor (CSTR). In this problem a reactant flows into a reactor vessel of unit volume at a rate r and undergoes a single, exothermic reaction to form inert products. We suppose that the reactor is well stirred, which means that the concentration c of the reactant and the temperature T are uniform throughout the vessel. The unused reactant and the products leave the vessel at the same rate r as the input; the concentration of the reactant and the temperature in the exit stream are equal to those in the reactor itself. Heat is removed from the reactor by a coolant fluid at temperature T_c.

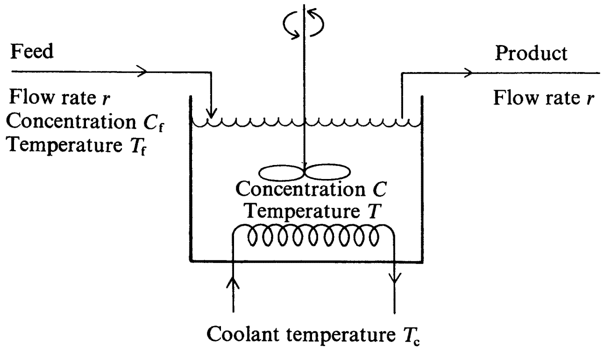


Figure 2.1. Schematic diagram for CSTR.

The concentration and temperature in the reactor are modeled by the following pair of coupled ordinary differential equations (ODE's).

$$\text{(a)} \quad \frac{dc}{dt} = r(c_f - c) - ZcA(T),$$

$$\text{(b)} \quad \frac{dT}{dt} = r(T_f - T) + k(T_c - T) + hZcA(T). \tag{2.1}$$

There are three physical processes represented in (2.1). In the first terms on the right in (2.1), c_f and T_f denote the concentration and temperature of the feed (i.e., the incoming reactant). In the absence of other terms, c and T would decay exponentially (at rate r) to the feed values c_f and T_f. The middle term in (2.1b) represents heat removed by the coolant. The parameter k is a lumped one, involving the heat transfer area, specific heats, etc. The final term in each equation is associated with the reaction. Note that the reaction depletes the concentration but increases the temperature. The factor $A(T)$, which governs the temperature dependence of the reaction rate, typically has Arrhenius form

$$A(T) = \exp\left\{\frac{T_a}{T_f} - \frac{T_a}{T}\right\}, \tag{2.2}$$

where T_a is the *activation energy* (converted to a temperature by means of the universal gas constant R). We have added the constant term T_a/T_f in the exponent in (2.2) so that Z in (2.1) represents the reaction rate at the feed temperature T_f. (Typically T_a is much larger than T_f, say $\gamma = T_a/T_f > 10$, so that the factor $\exp(T_a/T_f)$ is substantial.) Finally the parameter h is proportional to the heat released by the reaction.

We are interested in equilibrium solutions of (2.1), so we set the left-hand sides equal to zero. On solving the first equation for c and substituting into the second, we obtain the relation

$$r(T_f - T) + k(T_c - T) + \frac{\mathrm{rhc_f}}{1 + \dfrac{r}{ZA(T)}} = 0. \tag{2.3}$$

In order to nondimensionalize (2.3), we define a normalized temperature $x = (T - T_f)/T_f$. Then (2.3) may be rewritten as

$$g(x, \lambda) = (1 + \lambda)x - \eta - \frac{B\lambda}{1 + \delta\lambda\mathscr{A}(x)} = 0, \tag{2.4}$$

where $\lambda = r/k$, $\delta = k/Z$, $\eta = (T_c - T_f)/T_f$, $B = hc_f/T_f$, and

$$\mathscr{A}(x) = \exp\left\{-\frac{\gamma x}{1 + x}\right\},$$

with $\gamma = T_a/T_f$. Note that λ and δ represent comparisons of the flow rate and the reaction rate, respectively, with the rate of heat loss; η and γ compare two temperature parameters with the feed temperature; and B is a dimensionless measure of the heat of reaction. (*Remark*: We have inserted a factor -1 in passing from (2.3) to (2.4) because this facilitates applying the stability results of §4. The discrepancy in sign is a result of our convention in §4 of writing all terms in a differential equation on the left; i.e., a minus sign is needed to write (2.1) in the form (4.1).)

(b) The Occurrence of Multiple Solutions

Equation (2.4) determines the possible equilibrium temperature(s) x of a CSTR. We regard (2.4) as defining x as a (possibly multiple-valued) function of λ. As is customary, we treat λ as a distinguished parameter, since the flow rate is the quantity most readily varied in the laboratory. We are primarily interested in the bifurcation phenomena exhibited by (2.4); i.e., the occurrence of multiple solutions and their dependance on the various auxiliary parameters in the problem.

It is quite easy to see that multiple solutions of (2.4) are possible. In Figure 2.2 we have graphed the third term in (2.4) as a function of x for typical parameter values. Solutions of (2.4) correspond to intersections of this curve with the straight line $(1 + \lambda)x - \eta$. It is clear from the figure that there may be either one or three intersections, depending on the value of the parameters; three possible cases are sketched in the figure. In other words, multiplicity

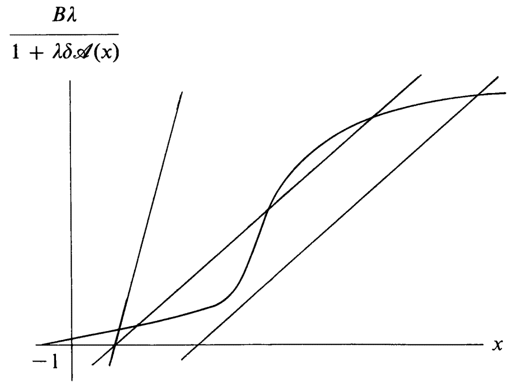


Figure 2.2. Possible intersection configurations.

results from the different balances that can be achieved between a linear term (heat removed by cooling and by heat exchange of the flow) against a non-linear one (temperature dependence of the reaction). (*Remark*: Calculation shows that the curve in Figure 2.2 has a single inflection point; for large γ the inflection point occurs at approximately $x = 2/\gamma$.)

In the limiting cases of $\lambda \to 0$ or $\lambda \to \infty$, the solution of (2.4) is unique. If $\lambda \to 0$, then insufficient new reactant is available to continue the reaction, so the system will come to equilibrium at approximately the coolant temperature T_c. If $\lambda \to \infty$, then the reactant temperature is not changed significantly by any heat produced by the reaction or absorbed from the coolant since the reactant remains in the reactor only a negligible time; thus the equilibrium temperature will be approximately T_f. Therefore multiple solutions can only occur for intermediate values of λ.

Independently, Zeldovich and Zisin [1941] and Uppal, Ray, and Poore [1976] made the surprising discovery that there may be two distinct ranges of λ where (2.4) has multiple solutions and that some of these solutions may lie on an isolated branch not connected to the unique solution of (2.4) which occurs in the limits $\lambda \to 0$ or $\lambda \to \infty$. Uppal *et al.* [1976] conducted an extensive numerical study of the problem, and we present their results in Figure 2.3. The upper half of this figure shows the (first quadrant of the) δB parameter plane divided by the curves $\mathscr{B}_1 P_i P_c P_p \mathscr{B}_2$ and $\mathscr{H}_1 P_i P_p \mathscr{H}_2$ into five regions. (Notation is discussed below.) Associated to each of these five regions, in the lower half of the figure, is a plot of the solutions x of (2.4) as a function of λ. (*Remarks*: The residence time, $\tau = 1/\lambda$, is the bifurcation parameter used in Uppal *et al.* [1976]. Qualitatively speaking, this change of parameter simply reverses the orientation of the graphs. Similarly δ^{-1} is the abscissa in the upper part of the figure. Finally, throughout Figure 2.3 the parameter η, which measures coolant temperature, is held fixed.)

Note that in graphs 2 and 5 there is an isolated solution branch, and that in graphs 3 and 5 there are two distinct ranges of λ with multiple solutions.

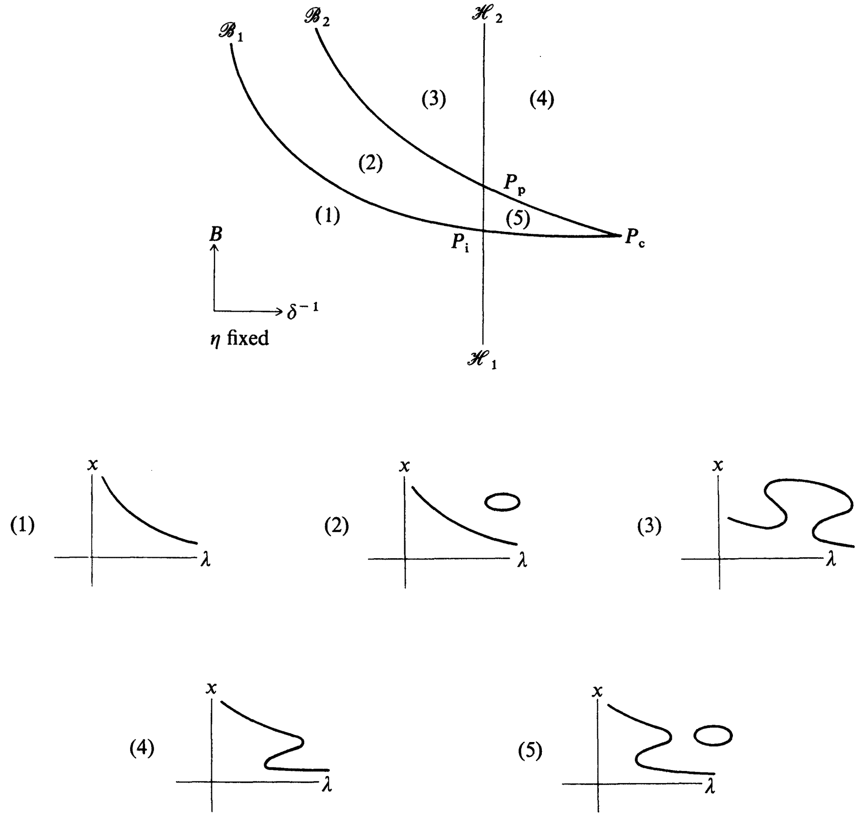


Figure 2.3. Multiple solutions in the CSTR (after Uppal, Roy, and Poore [1976]).

(c) A Primer on the Application of Singularity Theory Methods

One achievement of singularity theory is to provide a natural explanation for the data summarized in Figure 2.3. Incidentally the information gained from the theory shows that the two curves $\mathscr{B}_1 P_i P_c P_p \mathscr{B}_2$ and $\mathscr{H}_1 P_i P_p \mathscr{H}_2$ *must be tangent* at their intersection at P_p, a fact that was apparently not clear from the numerical evidence. (We have drawn Figure 2.3 in imitation of Uppal *et al.* [1976]; in Figure 2.4 the curves are correctly shown as tangent.) Another achievement is that, using singularity theory methods, we can analyze (2.4) analytically, without recourse to the computer. For the mathematician, this has the advantage of elegance and rigor. For the engineer, it led to the discovery that two additional bifurcation diagrams, not shown in Figure 2.3, could occur for certain parameter values.

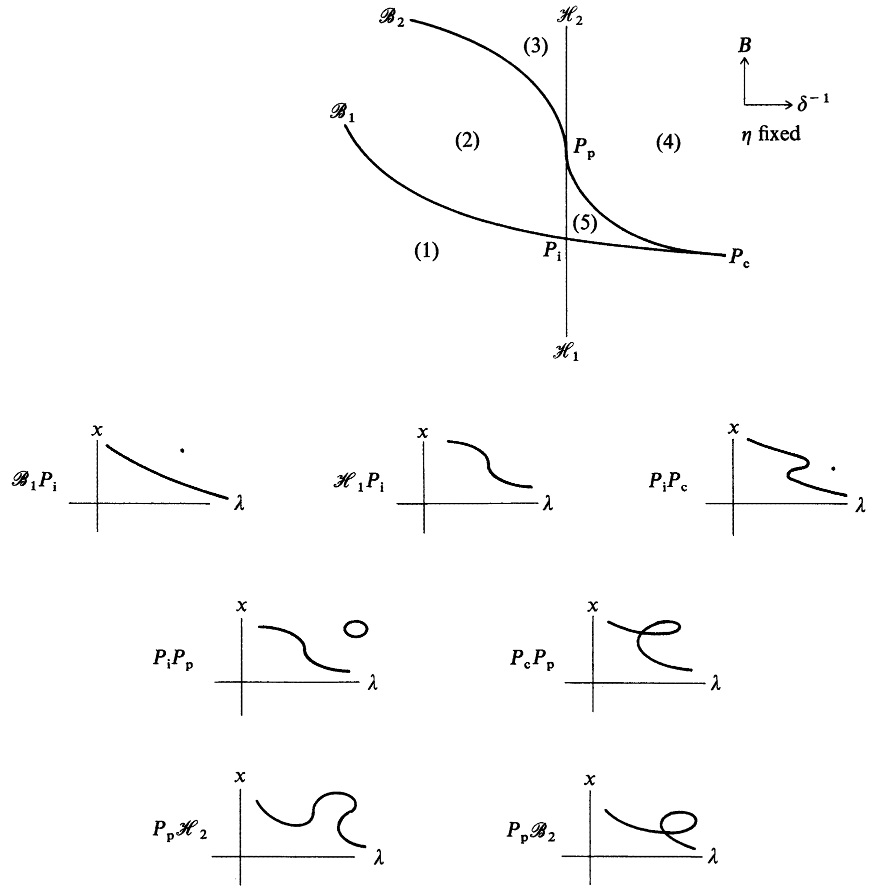


Figure 2.4. Bifurcation diagrams along the boundary between two regions in Figure 2.3.

To understand how to apply singularity theory methods in problems such as this one, it is helpful to consider the dependence of x on λ when (δ, B) lies on the boundary between two regions in Figure 2.3. There are seven distinct portions of the boundary ($\mathscr{B}_1P_i$, P_iP_c, P_cP_p, $P_p\mathscr{B}_2$, $\mathscr{H}_1P_i$, P_iP_p, $P_p\mathscr{H}_2$), and in Figure 2.4 we have shown the associated bifurcation diagrams. For example in the case of the boundary between regions 1 and 2, $\mathscr{B}_1P_i$, the diagram consists of a smooth curve $x = x(\lambda)$ and an isolated point solution (x_0, λ_0) of (2.4). As the parameters B and δ vary, this isolated solution can either disappear into the complex plane, as in region 1 of Figure 2.3, or open up into a small "circle" of solutions, as in region 2. (We will supply proofs in Case Study 1, drawing on Chapter III, §8.) Similarly, in the case of the boundary between regions 1 and 4, $\mathscr{H}_1P_i$, the curve of x as a function of λ has a vertical tangent, although it is still single-valued. (As indicated in §1, we shall call points having vertical tangents hysteresis points.) If δ or B is varied to perturb this graph, it can either pull out to give a smooth, single-valued curve, as in region 1, or twist back to give a range of λ where there are multiple solutions, as in region 4.

Let us explain the notation in Figures 2.3 and 2.4. We have labeled the ends of one boundary curve with the letters $\mathscr{B}_i$, $i = 1, 2$ because, as may be seen in Figure 2.4, traditional bifurcation phenomena occur along this curve—either the formation of new solution branches or the crossing of already existing ones. We label the ends of the other curve $\mathscr{H}_i$, $i = 1, 2$ because points on the curves are associated with the onset of possible hysteresis. For the three distinguished points on these curves, P_{p}, P_{c}, and P_{i}, the subscripts are mnemonics for pitchfork, cusp, and intersection, respectively. Understanding the significance of these points is our next task.

In the singularity theory approach, one focuses relentlessly on degenerate cases. Thus having seen the type of bifurcation diagrams which occur along the boundaries of regions in Figure 2.3, we now ask what happens at boundaries of the boundary; i.e., at the three points P_{c}, P_{p}, and P_{i}. We exhibit this behavior in Figure 2.5; again, justifications will be given in Case Study 1 and Chapter III, §8.

The following remarks on the three cases may be helpful. In case P_{c}, one may regard this diagram as the limit of the diagrams in case $P_{\mathrm{c}}P_{\mathrm{p}}$ of Figure 2.4 as the loop shrinks to diameter zero, or alternatively as the limit of case $P_{\mathrm{i}}P_{\mathrm{c}}$ as the isolated point meets the main solution branch. In case P_{p}, the crucial issue is that one of the intersecting curves has a vertical tangent—this is what separates cases $P_{\mathrm{c}}P_{\mathrm{p}}$ and $P_{\mathrm{p}}\mathscr{B}_2$ of Figure 2.4. The bifurcation of case P_{p} is equivalent to the normal form (1.1) for the pitchfork (in a neighborhood of the bifurcation point). Finally in case P_i there are two singular points in the bifurcation diagram, a bifurcation point and a hysteresis point, and they are more or less independent.

Let us carry this focusing on the "worst case" to the extreme. Specifically, we will show in Case Study 1 that it is possible to vary η, which was held fixed in Figure 2.3, so as to make the three points P_{c}, P_{p}, and P_i merge into a single, superdegenerate point. This leads to a bifurcation diagram as shown in Figure 2.6—analogous to case P_{c} in Figure 2.5 except the tangent to the cusp is vertical. Indeed, Figure 2.6 arises as a limiting case of Figure 2.6(c), since as η is increased the tangent to the cusp in Figure 2.5(a) rotates clockwise. Of

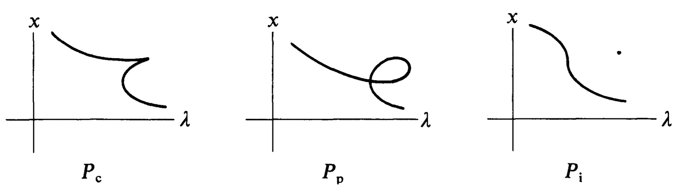


Figure 2.5. Bifurcation diagrams at the three distinguished points in Figure 2.3.

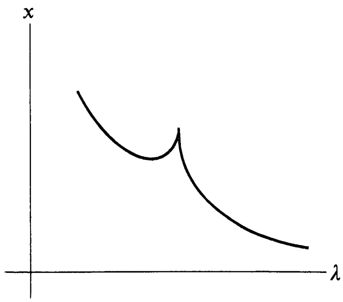


Figure 2.6. Bifurcation diagram of the organizing center for Figure 2.3.

course if η is increased beyond the critical point where the tangent is vertical, one obtains a bifurcation diagram as in Figure 2.7(a), and on perturbation of this, Figure 2.7(b), (c)—the two new diagrams whose existence was predicted by singularity theory. This analysis is due to Golubitsky and Keyfitz [1980].

We now try to summarize how we will apply singularity theory methods to the CSTR in Case Study 1, drawing on Chapter III. It turns out that

$$h(x, \lambda) = x^3 + \lambda^2 \tag{2.5}$$

is the appropriate normal form to describe the bifurcation diagram of Figure 2.6 near the singularity. (Following Golubitsky and Keyfitz [1980], we call (2.5) the *winged cusp.*) First, in Chapter II, §9 we will show that the recognition problem for (2.5) is solved by the following conditions:

$$g = g_x = g_\lambda = g_{xx} = g_{\lambda x} = 0; \qquad g_{xxx} > 0, \quad g_{\lambda x} > 0. \tag{2.6}$$

In other words, a function $g(x, \lambda)$ has a singularity equivalent to (2.5) at some point (x_0, λ_0) if and only if (2.6) holds at that point. Next, in Case Study 1 we will use (2.6) to prove that there is a unique set of values for the parameters δ, B, and η in (2.4) such that the resulting function has a singularity equivalent to (2.5). Let δ_0, B_0, and η_0 be the values which yield (2.6). Finally, we will call on results from Chapter III, §§4 and 8 concerning the universal unfolding of (2.5) in order to understand the solution set of (2.4) for δ, B, and η close to δ_0, B_0, and η_0. All the bifurcation diagrams of Figure 2.3 and the two additional diagrams of Figure 2.7(b), (c) will then emerge from perturbation of the

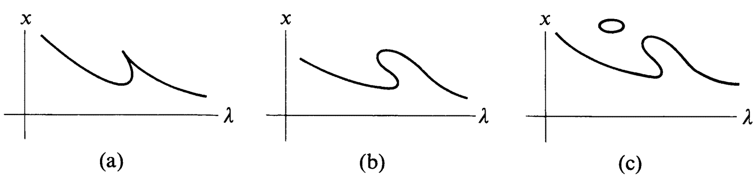


Figure 2.7. Additional bifurcation diagrams deduced from singularity theory.

distinguished values, δ_0, B_0, and η_0. In particular, the geometry of the various regions in Figure 2.3 is predicted *a priori* by the mathematics.

Although the above methods are local, Balakotaiah and Luss [1981] verified numerically that the conclusions are in fact valid *globally*. In particular, all seven diagrams predicted theoretically have been found numerically. This verification proceeded by computing numerically the transition curves in global parameter space. Balakotaiah and Luss [1981, 1982, 1983] have applied these ideas to a number of chemical reactor systems.

Following René Thom, we will refer to the bifurcation diagram in Figure 2.6 as an *organizing center* for this problem. This very suggestive, but somewhat vague, term only acquired meaning for us after we had analyzed several physical problems along the lines sketched above. Let us attempt at least a loose description of what this term means. Consider a physical problem which exhibits a variety of qualitatively different behaviors, depending on various parameters. An organizing center is associated with a distinguished set of values for the parameters such that all (or at least many) of the different behaviors occur for parameter values in a small neighborhood of the distinguished values. Typically at an organizing center the system exhibits its most singular behavior. We do not attempt a precise definition of this concept anywhere in the text, even though this idea occurs frequently. The use of an organizing center is best illustrated by the case studies. In particular, quasi-global results may often be obtained by the application of local analysis near an appropriately chosen organizing center.

In a typical application the parameters of the organizing center are distinguished because for these values the several physical effects in the problem are exactly balanced. Let us illustrate this for the CSTR. It may be seen from case 3 of Figure 2.3 that there can be hysteresis in the jump to the high-temperature solution branch on both the high λ and low λ sides. It is possible to vary two of the three parameters δ, B, and η so that hysteresis is on the verge of disappearing on both sides of the diagram, as sketched in Figure 2.8. Further, with appropriate variation of the third parameter, the two hysteresis points in Figure 2.8 approach one another and merge, resulting in the bifurcation diagram Figure 2.6. In other words, in Figure 2.6, the organizing center, the dissipative effects which lead to a unique solution as $\lambda \to 0$ or $\lambda \to \infty$

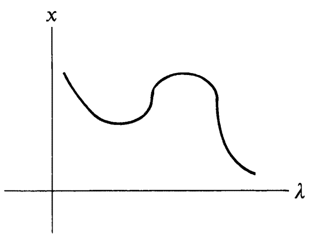


Figure 2.8. A bifurcation diagram with two hysteresis points.

are exactly balanced with the nonlinear temperature dependence of the reaction rate which pushes toward multiple solutions.

EXERCISES

2.1. Suppose that $g(x, \lambda)$ is equivalent to the winged cusp, $x^3 + \lambda^2$. Using the definition (1.6) of equivalence, show that g satisfies (2.6).

2.2. Show that the perturbation of the winged cusp $g(x, \lambda, \alpha) = x^3 + \lambda^2 + \alpha\lambda x$ has a pitchfork singularity at $(x_0, \lambda_0) = (0, 0)$ when $\alpha \neq 0$. Determine the orientation of these pitchfork bifurcations. (cf. (1.8).)

§3. A First View of the Liapunov–Schmidt Reduction

In this section we explore how it happens that so many problems in applied mathematics involving multiple solutions can be reduced to a single equation $g(x, \lambda) = 0$. The discussion centers around what is called the Liapunov–Schmidt reduction. Historically this procedure was used to reduce certain infinite-dimensional problems to one dimension. In the present section we consider the Liapunov–Schmidt procedure only in a finite-dimensional context. This reduces technicalities to a bare minimum, and we hope it will bring the essential issues into clearer focus. (We shall return to the reduction of infinite-dimensional problems in Chapter VII.)

Let us now set the context for the reduction. Consider a system of n equations

$$\Phi_i(y, \alpha) = 0, \qquad i = 1, \ldots, n, \tag{3.1}$$

where $\Phi\colon \mathbb{R}^n \times \mathbb{R}^{k+1} \to \mathbb{R}^n$ is a smooth mapping. We regard the vector $y = (y_1, \ldots, y_n)$ as the unknown to be solved for in (3.1); $\alpha = (\alpha_0, \ldots, \alpha_k)$ is a vector of parameters. (Usually we think of α_0 as a bifurcation parameter λ, which is distinguished, and $\alpha_1, \ldots, \alpha_k$ as auxiliary parameters. The reduction is already of interest when $k = 0$; i.e., when there are no auxiliary parameters. But since it does not complicate the analysis, right from the start we treat the case where auxiliary parameters may be present.) We assume that $\Phi_i(0, 0) = 0$ and we attempt to describe the solutions of this system locally near the origin. Let $(d\Phi)_{0,0}$ be the $n \times n$ Jacobian matrix $(\partial\Phi_i/\partial y_j(0, 0))$. If $\text{rank}(d\Phi)_{0,0} = n$, it follows from the implicit function theorem that (3.1) may be solved uniquely for y as a function of α; in other words, this is a nondegenerate case where no bifurcation occurs. In this section we consider the minimally degenerate case where

$$\text{rank}(d\Phi)_{0,0} = n - 1. \tag{3.2}$$

This section is divided into five subsections, which address the following issues:

(a) In subsection (a) we show that under the assumption (3.2), solutions of the full system (3.1) locally may be put in one-to-one correspondence with solutions of a single equation

$$g(x, \alpha) = 0, \tag{3.3}$$

where $g: \mathbb{R} \times \mathbb{R}^{k+1} \to \mathbb{R}$. This is the Liapunov–Schmidt reduction for (3.1). In words, (3.3) is a k-parameter family of bifurcation problems of the form $g(x, \lambda) = 0$.

(b) We summarize the essential steps of the reduction in subsection (b). (This is primarily for reference in Chapter VII.)

(c) We interpret the reduction geometrically in subsection (c).

(d) There are several arbitrary choices that must be made while performing the Liapunov–Schmidt reduction, and different choices lead to different reduced equations of the form of (3.3). In subsection (d) we present a theorem which states that different choices lead to *equivalent* reduced equations (as defined in §1), apart from some $\pm$ signs that must be inserted explicitly. This result provides further motivation for our definition of equivalence. (We prove this result in Appendix 2.)

(e) In subsection (e) we compute a few of the low-order derivatives of the reduced function (3.3) at the origin. Being able to make these calculations is important, since g is only defined implicitly—in most applications it is impossible to obtain a formula for g.

The methods of singularity theory may be applied to a bifurcation problem most readily after the Liapunov–Schmidt reduction has already been performed. The reduction meshes well with our theory. This is illustrated by items (d) and (e) above. Specifically:

(i) Although the reduced function is not uniquely determined, all possible reduced functions are equivalent (apart from possible differences of sign).
(ii) Singularity theory methods analyze the reduced function in terms of the data that is computable in applications; i.e., a finite number of the derivatives of g at the bifurcation point.

The ideas in this section will be used in §4 below, but then will not reappear until Chapter VII.

(a) Derivation of the Reduced Equations

Two arbitrary choices are required to set up the Liapunov–Schmidt reduction. As a convenient shorthand let us write $L = (d\Phi)_{0,0}$. We must choose vector space complements M and N to ker L and range L, respectively, obtaining the splittings

$$\mathbb{R}^n = \ker L \oplus M, \tag{3.4}$$

and

$$\mathbb{R}^n = N \oplus \text{range } L. \tag{3.5}$$

Observe that by assumption (3.2), dim range $L = n - 1$ and dim ker $L = 1$, so that dim $M = n - 1$ and dim $N = 1$. Let E denote the projection of $\mathbb{R}^n$ onto range L with ker $E = N$. The complementary projection $I - E$ has range equal to N and kernel equal to range L.

The following trivial observation starts the derivation: If $u \in \mathbb{R}^n$

$$u = 0 \quad \text{iff} \quad Eu = 0 \quad \text{and} \quad (I - E)u = 0. \tag{3.6}$$

Thus the system of equations (3.1) (i.e., $\Phi(y, \lambda) = 0$) may be expanded to an equivalent pair of equations

$$\begin{aligned} &\text{(a)} \qquad E\Phi(y, \alpha) = 0, \\ &\text{(b)} \quad (I - E)\Phi(y, \alpha) = 0. \end{aligned} \tag{3.7}$$

The basic idea underlying the Liapunov–Schmidt reduction is that (3.7a) may be solved for $n - 1$ of the y variables, and (3.7b) then yields an equation for the remaining unknown if values for these $n - 1$ variables are substituted into (3.7b).

Let us expand on this idea. First we apply the implicit function theorem to show that (3.7a) may be solved for $n - 1$ of the y variables. Because of the splitting (3.4), we may decompose any vector $y \in \mathbb{R}^n$ in the form $y = v + w$, where $v \in \ker L$ and $w \in M$. Let us write (3.7a) as

$$E\Phi(v + w, \alpha) = 0. \tag{3.8}$$

More abstractly, we are thinking of (3.8) as defining a map F: (ker L) $\times$ $M \times \mathbb{R}^{k+1} \to$ range L, where

$$F(v, w, \alpha) = E\Phi(v + w, \alpha).$$

By the chain rule, the differential of (3.8) with respect to the w variables at the origin is

$$E(d\Phi)_{0,0} = EL = L,$$

the first equality holding by definition and the second because E acts as the identity on range L. However, the linear map

$$L: M \to \text{range } L$$

is invertible. Thus it follows from the implicit function theorem that (3.7a) is uniquely solvable for w near the origin. Let us write this solution as $w = W(v, \alpha)$; thus W: ker $L \times \mathbb{R}^{k+1} \to M$ satisfies

$$E\Phi(v + W(v, \alpha), \alpha) \equiv 0, \qquad W(0, 0) = 0. \tag{3.9}$$

We substitute W into (3.7b) to obtain the reduced mapping $\phi\colon \ker L \times \mathbb{R}^{k+1} \to N$ where

$$\phi(v, \alpha) = (I - E)\Phi(v + W(v, \alpha), \alpha). \tag{3.10}$$

Then the zeros of $\phi(v, \alpha)$ are in one-to-one correspondence with the zeros of $\Phi(y, \alpha)$, the correspondence being given by

$$\phi(v, \alpha) = 0 \quad \text{iff} \quad \Phi(v + W(v, \alpha), \alpha) = 0.$$

The reduced function ϕ has all the information we need from the Liapunov–Schmidt reduction, but it suffers from the disadvantage that it maps between one-dimensional subspaces of $\mathbb{R}^n$; i.e., $\phi\colon \ker L \times \mathbb{R}^{k+1} \to N$. In applications it is customary to choose explicit coordinates on $\ker L$ and N and thereby obtain a reduced map $g\colon \mathbb{R} \times \mathbb{R}^{k+1} \to \mathbb{R}$. Of course, this introduces additional arbitrary choices into the method, beyond the choices of M and N in (3.4) and (3.5). We introduce coordinates as follows. Let v_0 and v_0^* be nonzero vectors in $\ker L$ and $(\text{range } L)^\perp$, respectively, where the orthogonal complement is taken with respect to the usual inner product

$$\langle y, z \rangle = \sum_{i=1}^{n} y_i z_i.$$

Any vector $v \in \ker L$ may be written uniquely in the form $v = xv_0$ where $x \in \mathbb{R}$. We define $g\colon \mathbb{R} \times \mathbb{R}^{k+1} \to \mathbb{R}$ by

$$g(x, \alpha) = \langle v_0^*, \phi(xv_0, \alpha) \rangle. \tag{3.11}$$

Since $\phi(xv_0, \alpha) \in N$, $g(x, \alpha) = 0$ iff $\phi(xv_0, \alpha) = 0$. Thus the zeros of g are also in one-to-one correspondence with solutions of $\Phi(y, \alpha) = 0$.

It is worth noting that in substituting the definition (3.10) of Φ into (3.11) the projection $(I - E)$ drops out; i.e.,

$$g(x, \phi) = \langle v_0^*, \Phi(xv_0 + W(xv_0, \alpha), \alpha) \rangle.$$

The reason for this simplification is that $v_0^* \in (\text{range } L)^\perp$, and for any vector $V \in \mathbb{R}^n$, $EV \in \text{range } L$, so $\langle v_0^*, EV \rangle = 0$. Hence

$$\langle v_0^*, (I - E)V \rangle = \langle v_0^*, V \rangle. \tag{3.12}$$

Remark 3.1. We use the phrase "reduced function" to refer to both $\phi(v, \alpha)$ and $g(x, \alpha)$. Both functions contain the same information—g is just the representation of ϕ in specific coordinates. For theoretical analysis ϕ is typically more convenient; for applications, g. We shall use whichever seems more appropriate.

(b) An Overview of the Liapunov–Schmidt Reduction

For purposes of reference in Chapter VII, we divide the above derivation of the reduced equation (3.11) into the following five steps:

Step 1. Decompose the ambient space into summands related to L. (Cf. (3.4), (3.5).)

Step 2. Transfer this decomposition to the equation. (Cf. (3.7).)

Step 3. Show that (3.7a) may be solved for all but one of the variables, using the implicit function theorem.

Step 4. Substitute the solution of (3.7a) into (3.7b) to obtain (3.10).

Step 5. Choose coordinates on $\ker L$ and $(\text{range } L)^\perp$ to obtain (3.11).

The essence of the Liapunov–Schmidt reduction is to show that the implicit function theorem is applicable in situations where its applicability may not be readily apparent. Thus Step 3 is the fundamental step in the reduction. The other steps are required to carry out Step 3. Note that in Steps 1 and 5 a choice must be made, while Steps 2 and 4 are primarily notational.

(c) A Geometric View of the Liapunov–Schmidt Reduction

It is instructive to think of the Liapunov–Schmidt reduction pictorially. In particular, this view clarifies the identification of the bifurcation diagram

$$\{(v, \alpha) \in \ker L \times \mathbb{R}^{k+1}: \phi(v, \alpha) = 0\}$$

with the solution set

$$\{(y, \alpha) \in \mathbb{R}^n \times \mathbb{R}^{k+1}: \Phi(y, \alpha) = 0\}$$

of the full equations.

We claim that the set

$$\mathscr{V} = \{(y, \alpha): E\Phi(y, \alpha) = 0\} \tag{3.13}$$

is a $(k + 2)$-dimensional submanifold of $\mathbb{R}^n \times \mathbb{R}^{k+1}$ whose tangent space is $\ker L \times \mathbb{R}^{k+1}$. (See Figure 3.1.) In fact, solving (3.9) by the implicit function theorem, we see that $\mathscr{V}$ may be parametrized by a map $\Omega: \ker L \times \mathbb{R}^{k+1} \to \mathbb{R}^n \times \mathbb{R}^{k+1}$, where

$$\Omega(v, \alpha) = (v + W(v, \alpha), \alpha). \tag{3.14}$$

In this formula v and $W(v, \alpha)$ belong to $\ker L$ and M, respectively. Since we have the decomposition $\mathbb{R}^n = \ker L \oplus M$, we could rewrite (3.14) in an equivalent notation as

$$\Omega(v, \alpha) = (v, W(v, \alpha), \alpha). \tag{3.14a}$$

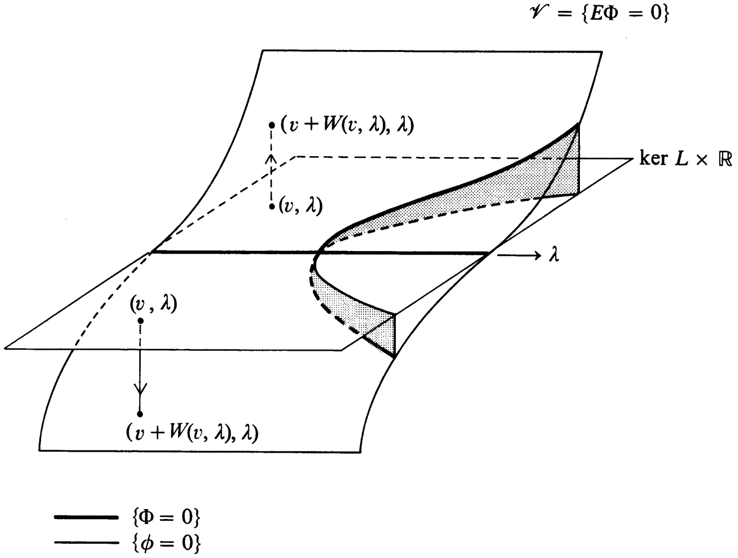


Figure 3.1. A picture of the Liapunov–Schmidt reduction: $n = 2, k = 0$.

It is clear from (3.14a) that $(d\Omega)_{0,0}$ is nonsingular; therefore $\mathscr{V}$ is a $(k + 2)$-dimensional submanifold of $\mathbb{R}^n \times \mathbb{R}^{k+1}$. We will show below by implicit differentiation that $(\partial/\partial x)W(xv_0, 0)|_{x=0} = 0$. (Cf. (3.15).) It follows that the tangent space to $\mathscr{V}$ at the origin is ker $L \times \mathbb{R}^{k+1}$, as claimed.

In Figure 3.1 we have attempted to sketch ker $L \times \mathbb{R}$, $\mathscr{V}$, and the zero sets of Φ and ϕ in a case where $n = 2$, $k = 0$, and the reduced function exhibits a pitchfork bifurcation. One can see from the figure how the bifurcation diagram $\{\phi(v, \lambda) = 0\}$, which lies in ker $L \times \mathbb{R}^{k+1}$, is identified with the zero set of Φ, which lies in $\mathscr{V}$.

(d) Relation with Equivalence

Different choices of the data needed to carry out the Liapunov–Schmidt reduction lead to reduced equations which are (essentially) equivalent. To formulate this assertion carefully, we need to set up some notation. Let $\Phi: \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ be a smooth mapping satisfying $\Phi(0, 0) = 0$ and (the minimal degeneracy condition) $\text{rank}(d\Phi)_{0,0} = n - 1$. Choose complements M_1 and M_2 to ker L as in (3.4). Choose complements N_1 and N_2 to range L as in (3.5). Choose v_1 and v_2 in ker L, v_1^* and v_2^* in (range $L)^\perp$. Let $g_1(x, \lambda)$ and $g_2(x, \lambda)$ be the reduced bifurcation equations obtained by using the four choices subscripted by 1 and 2, respectively. (We are assuming $k = 0$ here, and we write λ for α_0. There is no difficulty in extending the following theorem to positive k.)

Theorem 3.2. *Let* $\varepsilon = \operatorname{sgn}\langle v_1, v_2\rangle$ *and* $\delta = \operatorname{sgn}\langle v_1^*, v_2^*\rangle$. *Then* $g_2(x, \lambda)$ *is equivalent to* $\delta g_1(\varepsilon x, \lambda)$.

The importance of Theorem 3.2 lies in the motivation it provides for the definition of equivalence. We shall not make further use of this result in the text, so we have relegated its proof to Appendix 2. However, let us mention an issue that might not be apparent—reading this proof is a wonderful exercise for the reader who wishes to understand just what is involved in the Liapunov–Schmidt reduction.

(e) Computation of Derivatives of the Reduced Equations

In this subsection we show how to compute the derivatives of the reduced function $g(x, y)$ from derivatives of the original mapping $\Phi(y, \alpha)$. Let us summarize the calculations before performing them. We can find the derivatives of g by substitution into (3.11), $g(x, \alpha) = \langle v_0^*, \phi(xv_0, \alpha)\rangle$, if we know the derivatives of the function ϕ. To this end we rewrite the definition (3.10) of ϕ in the form in which it appears in (3.11):

$$\phi(xv_0, \alpha) = (I - E)\Phi(xv_0 + W(xv_0, \alpha), \alpha). \tag{3.10a}$$

Calculation of derivatives of (3.10a) is a straightforward application of the chain rule. However, the resulting formulas contain derivatives of W, and these must be determined by implicit differentiation of (3.9)

$$E\Phi(v + W(v, \alpha), \alpha) = 0.$$

This step is the most tedious part of the calculation, both in the present theoretical discussion and in actual applications.

It turns out that the first derivative of W with respect to x vanishes. We digress to prove this. (*Warning*: We sometimes write $W(x, \alpha)$ for $W(xv_0, \alpha)$. In this way derivatives of W with respect to x make sense. Cf. Remark 3.1.) We substitute $v = xv_0$ into (3.9) and differentiate with respect to x to obtain $E\, d\Phi \cdot (v_0 + W_x) = 0$. This becomes $EL(v_0 + W_x) = 0$ on evaluating at $(0, 0)$. However $v_0 \in \ker L$ and $EL = L$ so that we have

$$LW_x(0, 0) = 0.$$

But $W_x(0, 0) \in M$ and $L\colon M \to \text{range } L$ is invertible. Thus it follows that

$$W_x(0, 0) = 0. \tag{3.15}$$

Before actually starting the calculations we introduce an invariant notation for higher-order derivatives of functions of several variables. If $v_1, \ldots, v_k \in \mathbb{R}^n$, we define

$$(d^k\Phi)_{y,\alpha}(v_1, \ldots, v_k) = \frac{\partial}{\partial t_1}\cdots\frac{\partial}{\partial t_k}\Phi\left(y + \sum_{i=1}^{k} t_i v_i, \alpha\right)\Bigg|_{t_1 = \cdots = t_k = 0}. \tag{3.16}$$

Note that $(d^k\Phi)_{y,\alpha}$ is a symmetric, multilinear function of k arguments. If desired, we may represent (3.16) in terms of kth-order partial derivatives of Φ; for example, if $k = 2$

$$(d^2\Phi)_{y,\alpha}(v, w) = \sum_{i,j=1}^{n} \frac{\partial^2\Phi}{\partial y_i\, \partial y_j}(y, \alpha)v_i w_j.$$

In this notation the chain rule takes the following form. If the base point y and the vectors v_i, $i = 1, \ldots, k$ depend on a parameter t, then

$$\frac{\partial}{\partial t}\{(d^k\Phi)_{y,\alpha}(v_1, \ldots, v_k)\} = (d^{k+1}\Phi)_{y,\alpha}\left(\frac{\partial y}{\partial t}, v_1, \ldots, v_k\right) + \sum_{i=1}^{k} (d^k\Phi)_{y,\alpha}\left(v_1, \ldots, \frac{\partial v_i}{\partial t}, \ldots, v_k\right). \quad (3.17)$$

Here we are assuming that Φ does not depend explicitly on t; when it does, an additional term with $d^k(\partial\Phi/\partial t)$ is required on the right-hand side of (3.17).

We finally begin the calculations. Repeated application of the chain rule to (3.10a) yields the following formulas for the derivatives of ϕ.

$$\begin{aligned}
&\text{(a)} \quad \phi_x = (I - E)(d\Phi(v_0 + W_x)),\\
&\text{(b)} \quad \phi_{xx} = (I - E)(d\Phi(W_{xx}) + d^2\Phi(v_0 + W_x, v_0 + W_x)),\\
&\text{(c)} \quad \phi_{xxx} = (I - E)(d\Phi(W_{xxx}) + 3d^2\Phi(v_0 + W_x, W_{xx})\\
&\qquad\qquad + d^3\Phi(v_0 + W_x, v_0 + W_x, v_0 + W_x)), \qquad (3.18)\\
&\text{(d)} \quad \phi_{\alpha_l} = (I - E)(\Phi_{\alpha_l} + d\Phi(W_{\alpha_l})),\\
&\text{(e)} \quad \phi_{\alpha_l x} = (I - E)(d\Phi_{\alpha_l}(v_0 + W_x)\\
&\qquad\qquad + d\Phi(W_{\alpha_l x}) + d^2\Phi(v_0 + W_x, W_{\alpha_l})).
\end{aligned}$$

We evaluate at $x = 0$. $\alpha = 0$ and recall that $(I - E)L = 0$ and $W_x(0, 0) = 0$; the formulas become

$$\begin{aligned}
&\text{(a)} \quad \phi_x(0, 0) = 0,\\
&\text{(b)} \quad \phi_{xx}(0, 0) = (I - E)(d^2\Phi(v_0, v_0)),\\
&\text{(c)} \quad \phi_{xxx}(0, 0) = (I - E)(3d^2\Phi(v_0, W_{xx}(0, 0)) + d^3\Phi(v_0, v_0, v_0), \qquad (3.19)\\
&\text{(d)} \quad \phi_{\alpha_l}(0, 0) = (I - E)(\Phi_{\alpha_l}(0, 0)),\\
&\text{(e)} \quad \phi_{\alpha_l x}(0, 0) = (I - E)(d\Phi_{\alpha_l}(v_0) + d^2\Phi(v_0, W_{\alpha_l}(0, 0))).
\end{aligned}$$

Before continuing the calculation, we make two remarks. First, the fact that $\phi_x(0, 0)$ vanishes is totally expected. This means that the reduced equation has a singularity at the origin. If it did not then we could have applied the implicit function theorem to the original system $\Phi(y, \alpha) = 0$; that is, $\operatorname{rank}(d\Phi)_{0,0}$ would be n and not $n - 1$ as assumed in (3.2).

Second, one of the major problems in evaluating (3.19) lies in the computation of $W_{xx}(0, 0)$ and $W_{\alpha_l}(0, 0)$. We shall give specific formulas for these quantities below. The difficulty is that these formulas require inverting a linear operator. However the computation of $W_{xx}(0, 0)$ and $W_{\alpha_l}(0, 0)$ is not required in (3.19) if $(d^2\Phi)_{0,0}$ happens to be zero. This circumstance is not unusual in applications, as often Φ is an odd function; i.e.,

$$\Phi(-y, \alpha) = -\Phi(y, \alpha). \tag{3.20}$$

If (3.20) is satisfied, then $\Phi_{\alpha_l}(0, 0) = 0$ and $(d^2\Phi)_{0,0} = 0$. So the formulas (3.19) reduce to

$$\begin{gathered} \phi_x(0, 0) = 0, \qquad \phi_{xx}(0, 0) = 0, \qquad \phi_{\alpha_l}(0, 0) = 0, \\ \phi_{xxx}(0, 0) = (I - E)(d^3\Phi(v_0, v_0, v_0)), \\ \phi_{\alpha_l x}(0, 0) = (I - E)(d\Phi_{\alpha_l}(v_0)). \end{gathered} \tag{3.21}$$

(*Remark*: The property of Φ being odd is a specific instance of Φ possessing a certain symmetry; we shall study symmetry and its consequences in later chapters, especially in Chapter VI. One point deserves comment here. If an odd function has a singularity (i.e., if $\Phi_x(0, 0) = 0$), then automatically $\Phi_{xx}(0, 0) = 0$ and $\Phi_{\alpha_l}(0, 0) = 0$. We saw in §1 that the vanishing of these derivatives was part of the characterization of the pitchfork. Thus in the context of odd functions, the pitchfork is the minimally degenerate singularity.)

Returning to the calculation, we claim that at $x = \alpha = 0$

$$\begin{aligned} &\text{(a)} \quad W_{xx}(0, 0) = -L^{-1}E\, d^2\Phi(v_0, v_0), \\ &\text{(b)} \quad W_{\alpha_l}(0, 0) = -L^{-1}E\Phi_{\alpha_l}(0, 0), \end{aligned} \tag{3.22}$$

where L^{-1}: range $L \to M$ denotes the inverse of the linear map $L|M$. To verify (3.22a), we differentiate (3.9), $E\Phi(xv_0 + W(xv_0, \alpha), \alpha) \equiv 0$, twice with respect to x and evaluate at $x = \alpha = 0$; this yields

$$EL(W_{xx}(0, 0)) + Ed^2\Phi(v_0 + W_x(0, 0), v_0 + W_x(0, 0)) = 0.$$

Recalling that $W_x(0, 0) = 0$ and that $EL = L$ we solve this equation for $W_{xx}(0, 0)$ to obtain (3.22a). Similarly, one verifies (3.22b) by differentiating (3.9) with respect to α_l and evaluating at $x = \alpha = 0$.

To complete the calculation we will substitute (3.22) into (3.19) and then use the resulting formulas in (3.11), $g(x, \alpha) = \langle v_0^*, \phi(xv_0, \alpha)\rangle$. Carrying out the above steps and recalling that $\langle v_0^*, (I - E)v\rangle = \langle v_0^*, v\rangle$ (cf. (3.12)), we find that

$$\begin{aligned} &\text{(a)} \quad g_x = 0, \\ &\text{(b)} \quad g_{xx} = \langle v_0^*, d^2\Phi(v_0, v_0)\rangle, \\ &\text{(c)} \quad g_{xxx} = \langle v_0^*, d^3\Phi(v_0, v_0, v_0) - 3d^2\Phi(v_0, L^{-1}Ed^2\Phi(v_0, v_0))\rangle, \\ &\text{(d)} \quad g_{\alpha_l} = \langle v_0^*, \Phi_{\alpha_l}\rangle, \\ &\text{(e)} \quad g_{\alpha_l x} = \langle v_0^*, d\Phi_{\alpha_l} \cdot v_0 - d^2\Phi(v_0, L^{-1}E\Phi_{\alpha_l})\rangle. \end{aligned} \tag{3.23}$$

These are the formulas we were seeking.

The principal difficulty in computing these derivatives lies in the evaluation of the inverse of L. This difficulty is even greater when we generalize to infinite-dimensional problems in Chapter VII, where inverting L requires solving a differential equation. Thus any special circumstances which cause some of these terms to vanish are most welcome. One such circumstance was mentioned above; i.e., if $\Phi(y, \alpha)$ is odd in y. Another occurs if $y = 0$ is a solution of the equation for all values of the bifurcation parameter $\alpha_0 = \lambda$; i.e., if $\Phi(0, \alpha) = 0$ for any α of the form $(\lambda, 0, \ldots, 0)$. In the latter case $\Phi_\lambda(0, 0) = 0$, so by (3.22b), $W_\lambda(0, 0) = 0$. It follows from (3.19d) that $\phi_\lambda(0, 0) = 0$, and moreover the troublesome $d^2\Phi(v_0, W_\lambda)$ in (3.19e) drops out of this equation.

Exercises

3.1. In §2 we reduced the equilibrium equations for (2.1), a 2×2 system of ODE, to the single scalar equation (2.3). This is a particular instance of the Liapunov–Schmidt reduction. Specifically, form a 2×2 system of equations of the form (3.1) by setting the right-hand side in (2.1) equal to zero. Show that if we take

$$M = \mathbb{R}\{\tbinom{1}{0}\}, \qquad N = \mathbb{R}\{\tbinom{0}{1}\},$$

$$v_0 = \tbinom{a}{1}, \qquad v_0^* = \tbinom{b}{1},$$

where a and b are appropriate constants, then the general reduction process leads to (2.3).

3.2. Let $\Phi\colon \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ be defined by

$$\Phi(u_1, u_2, \lambda) = \begin{pmatrix} 2u_1 - 2u_2 + 2u_1^2 + 2u_2^2 - \lambda u_1 \\ u_1 - u_2 + u_1 u_2 + u_2^2 - 3\lambda u_1 \end{pmatrix} = 0.$$

Using the Liapunov–Schmidt reduction, show that $\Phi = 0$ has a pitchfork bifurcation in a neighborhood of the origin. (To check your answer, solve the second equation in $\Phi = 0$ by the implicit function theorem and use implicit differentiation to obtain

$$u_2 = u_1 + 2u_1^2 - 3u_1\lambda + 6u_1^3 + \cdots. \tag{3.24}$$

Then substitute (3.24) into the first equation.)

3.3. (*Discussion*) Consider the finite element approximation of the Euler column illustrated in Figure 3.2. It consists of three rigid rods of unit length connected by

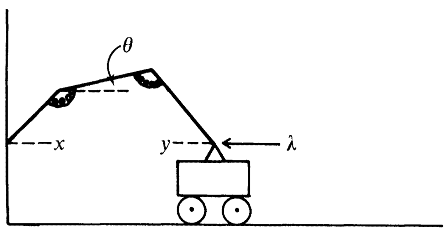


Figure 3.2. A second finite element analogue of Euler buckling.

pins which permit rotation in a plane. It is subjected to a compressive force λ which is resisted by torsional springs of unit strength at the two connecting pins. This system is quite similar to the one illustrated in Figure 1.2; however the present system has two degrees of freedom. The point of the exercise is to use the Liapunov–Schmidt reduction to show that the qualitative behavior of the present system is identical to that of Figure 1.2.

(a) Let x and y be defined as in Figure 3.2. Derive a 2×2 system of equations for x and y, say

$$\Phi_i(x, y, \lambda) = 0, \qquad i = 1, 2,$$

which characterizes equilibria of the model.

Hint: The easiest derivation is to use the potential function

$$V(x, y) = \tfrac{1}{2}(x - \theta)^2 + \tfrac{1}{2}(y + \theta)^2 + \lambda(\cos x + \cos \theta + \cos y).$$

Here θ is the angle the middle rod makes with the horizontal. Note that θ is not an independent variable—rather θ is the following function of x and y:

$$\theta = \sin^{-1}[\sin y - \sin x].$$

This relation comes from requiring that the two end pins be at the same height. Now let $\Phi_1 = \partial V/\partial x$ and $\Phi_2 = \partial V/\partial y$.

(b) Note that $x = y = 0$ is a solution of the equations in (a) for any value of λ. Let $(d\Phi)_{0,0,\lambda}$ be the 2×2 Jacobian of the equations at this solution. Show that $(d\Phi)_{0,0,\lambda}$ is invertible for $0 \leq \lambda < 1$ and singular for $\lambda = 1$. Show also that $(d\Phi)_{0,0,\lambda}$ has rank 1 when $\lambda = 1$.

(c) Find the kernel and range of $(d\Phi)_{0,0,\lambda}$ when $\lambda = 1$.

(d) Show that if the Liapunov–Schmidt reduction is applied to the equations at $x = y = 0$, $\lambda = 1$, then the reduced bifurcation equation is equivalent to $x^3 - \lambda x$.

Hint: Verify (1.8), using (3.23) to perform the calculations. The calculations are simplest if in (3.5) one takes

$$M = \mathbb{R}\binom{1}{-1}, \qquad N = \mathbb{R}\binom{1}{1}.$$

§4. Asymptotic Stability and the Liapunov–Schmidt Reduction

In this section we discuss how the stability of an equilibrium solution of an ODE is affected by the Liapunov–Schmidt reduction. Specifically, let $F\colon \mathbb{R}^n \times \mathbb{R}^{k+1} \to \mathbb{R}^n$ and consider the $n \times n$ system of ODE

$$\dot{y} + F(y, \alpha) = 0, \tag{4.1}$$

where $\alpha = (\alpha_0, \ldots, \alpha_k)$ is a vector of parameters. Equilibrium solutions of (4.1) are characterized by the equation

$$F(y, \alpha) = 0. \tag{4.2}$$

Suppose that $F(y_0, 0)$ (i.e., y_0 is a rest point of (4.1) when $\alpha = 0$) and that

$$\operatorname{rank}(dF)_{y_0, 0} = n - 1. \tag{4.3}$$

Since $(dF)_{y_0, 0}$ is singular, the equilibrium solution of (4.1) at $y = y_0$ for $\alpha = 0$ may split into several equilibrium solutions when $\alpha \neq 0$. (Let us refer to these by the term *perturbed* equilibrium solutions.) Using the Liapunov–Schmidt reduction we may associate such perturbed equilibrium solutions of (4.1) with solutions of a single scalar equation

$$g(x, \alpha) = 0, \tag{4.4}$$

where $g\colon \mathbb{R} \times \mathbb{R}^{k+1} \to \mathbb{R}$. The main conclusion of this section is that, under a slight strengthening of hypothesis (4.3), the stability or instability of these perturbed equilibrium solutions of (4.1) is determined by the sign of g_x, the derivative of the reduced function (4.4).

This section is divided into two subunits. In subsection (a) we review the theory of asymptotic stability for ODE, and in subsection (b) we formulate and prove our main result, Theorem 4.1.

In both §3 and §4 we have restricted ourselves to finite dimensions. We will generalize the results of §3 to infinite dimensions in Chapter VII. Although the results of the present section also have infinite-dimensional analogues, they are more technical, and we shall not pursue them in this text.

(a) Asymptotic Stability

In defining asymptotic stability, let us temporarily suppress parameters in the differential equation. Let $F\colon \mathbb{R}^n \to \mathbb{R}^n$, and suppose that y_0 is an equilibrium solution of the ODE

$$\dot{y} + F(y) = 0; \tag{4.5}$$

i.e., suppose that $F(y_0) = 0$. We shall call the rest point y_0 *asymptotically stable* if every solution to (4.5) with initial condition close to y_0 decays to y_0. More precisely, y_0 is asymptotically stable if there are positive constants ε and M such that for any solution $y(t)$ to (4.5) satisfying $|y(0) - y_0| < \varepsilon$ we have $|y(t) - y_0| < M$ and $\lim_{t\to\infty} y(t) = y_0$. Otherwise we call y_0 *unstable*.

There is a useful sufficient condition for asymptotic stability in terms of the eigenvalues of the Jacobian matrix $(dF)_{y_0}$. We shall call y_0 *linearly stable* if every eigenvalue of $(dF)_{y_0}$ has a positive real part, *linearly unstable* if at least one eigenvalue has a negative real part. In Chapter 9, §§1 and 2 of Hirsch and Smale [1974] it is shown that y_0 is asymptotically stable if it is linearly stable and that y_0 is unstable if it is linearly unstable. (*Remark*: Note that this apparent dichotomy is not complete: if every eigenvalue of $(dF)_{y_0}$ has a nonnegative real part but at least one real part vanishes, then y_0 is neither linearly stable or linearly unstable. In this situation there is no simple test for asymptotic stability.)

The intuition here can be obtained by looking for solutions of (4.5) of the form

$$y(t) = y_0 + \varepsilon z(t). \tag{4.6}$$

On substituting into (4.5) and neglecting terms of order ε^2 or higher, we find the equation for $z(t)$

$$\dot{z} + Lz = 0, \tag{4.7}$$

where $L = (dF)_{y_0}$. (This equation is commonly called the *linearization* of (4.5) at y_0.) Let μ_i, $i = 1, \ldots, n$ be the eigenvalues of L, and let v_i be the associated eigenvectors. Then the general solution of (4.7) has the form

$$z(t) = \sum_{i=1}^{n} c_i e^{-\mu_i t} v_i, \tag{4.8}$$

where the c_i's are constants. (Equation (4.8) holds provided the μ_i's are distinct; a slight modification is required if there are repeated eigenvalues.) If $\mu_i > 0$ for all i, then $z(t) \to 0$ as $t \to \infty$. The spirit of the theorem is that on an appropriately small neighborhood of y_0, the full equation (4.5) mimics the behavior of the linearization (4.7). (*Remark*: With our convention of writing $F(y)$ on the left in (4.5), positive eigenvalues correspond to stability. The opposite convention results from writing F on the right.)

Let us now return to (4.1); i.e., the case where the ODE depends on one or more parameters. To simplify the notation we will suppose that $y_0 = 0$; in other words, we are assuming that $y = 0$ is an equilibrium solution of (4.1) when $\alpha = 0$. (The discussion applies with trivial modifications to an arbitrary rest point y_0.) Let $L = (dF)_{0,0}$. As noted above, linearization of the equations yields no information about asymptotic stability if one eigenvalue of L is zero and all the rest are positive. Of course our assumption (4.3) states unequivocally that zero is an eigenvalue of L. Thus, the task of this section is to analyze the stability of equilibrium solutions of (4.1) in the neighborhood of a borderline case.

Let us expand on this point. Suppose that the eigenvalues $\mu_1, \ldots, \mu_n$ of L satisfy the following:

$$\mu_1 = 0, \qquad \operatorname{Re} \mu_i > 0 \quad \text{for} \quad i = 2, \ldots, n. \tag{4.9}$$

(Equation (4.3) follows from (4.9), but not conversely.) As we remarked above, the equilibrium solution of (4.1) at $y = 0$ for $\alpha = 0$ may split into several perturbed equilibrium solutions when $\alpha \neq 0$. Now an equilibrium solution (y, α) of (4.1) will be asymptotically stable if all the eigenvalues of $(dF)_{y,\alpha}$ have positive real parts and unstable if at least one eigenvalue has a negative real part. We claim that for (y, α) near $(0, 0)$ the eigenvalues of $(dF)_{y,\alpha}$ will be close to those of $(dF)_{0,0}$. To prove this, let us write $(dF)_{y,\alpha}$ as a perturbation of $(dF)_{0,0} = L$:

$$(dF)_{y,\alpha} = L + [(dF)_{y,\alpha} - L].$$

The perturbing term $(dF)_{y,\alpha} - L$ is small if (y, α) is close to zero, and the change in the eigenvalues tends to zero as the perturbation tends to zero. This proves the claim. However, the eigenvalues of L satisfy (4.9). Thus by the claim the last $n - 1$ eigenvalues of $(dF)_{y,\alpha}$ will be bounded away from the imaginary axis on an appropriately small neighborhood of (0, 0) and could not cause (y, α) to be an unstable rest point of (4.1). In constrast, the first eigenvalue (which we denote by $\mu(y, \alpha)$) will be close to zero and might cause such instability. Indeed an equilibrium solution (y, α) of (4.1) will be linearly stable or unstable according as $\mu(y, \alpha)$ is positive or negative, respectively. Our goal is to show that $g_x(x, \alpha)$, where g is the reduced function from the Liapunov–Schmidt process, has the same sign as $\mu(y, \alpha)$.

(b) Statement and Proof of the Main Result

Recall from §3 that several arbitrary choices are necessary when making a Liapunov–Schmidt reduction. Specifically, in Step 1 one must choose complements to ker L and to range L, and in Step 5 one must choose a nonzero vector v_0 in ker L and a nonzero vector v_0^* in (range $L)^\perp$. We claim that the assumption (4.9) implies that v_0 is not in range L. For suppose v_0 were in range L, say $v_0 = Lw$. Then $Lw \neq 0$ and $L^2 w = 0$. Thus both v_0 and w would belong to ker L^2, so that dim ker $L^2 \geq 2$. This would contradict the assumption (4.9) that zero is an algebraically simple eigenvalue for L.

Since v_0 does not belong to range L, $\langle v_0, v_0^* \rangle \neq 0$. In order for the reduced equation g to give the correct stability information we need to match the orientations of v_0 and v_0^* by requiring that

$$\langle v_0, v_0^* \rangle > 0. \tag{4.10}$$

A choice of vectors satisfying (4.10) is said to be a *consistent choice.*

Theorem 4.1. *Let $\dot{z} + Lz = 0$ be the linearization of $\dot{y} + F(y, \alpha) = 0$, and assume that the eigenvalues of L satisfy* (4.9). *Let $g(x, \lambda)$ be the reduced equation obtained by a Liapunov–Schmidt reduction of $F(y, \alpha) = 0$ using a consistent choice of v_0 and v_0^*. Then the rest point of $\dot{y} + F(y, \alpha) = 0$ corresponding to a solution (x, α) of $g(x, \alpha) = 0$ is asymptotically stable if $g_x(x, \alpha) > 0$ and unstable if $g_x(x, \alpha) < 0$.*

Roughly speaking, the proof of Theorem 4.1 requires showing that the quotient μ/g_x, where μ is the first eigenvalue of dF, defines a smooth function which is positive near the origin. We perform this division by means of the following proposition. Both Proposition 4.2 and the techniques involved in its proof will be used in later chapters. However, with the one exception of Chapter VIII, §4, the remainder of this text is independent of the proof of Theorem 4.1 itself. Thus the reader may omit this proof with no loss of continuity.

Proposition 4.2. *Let $\phi, \psi\colon \mathbb{R}^n \to \mathbb{R}$ be C^∞ functions defined on a neighborhood of zero which vanish at zero. Assume that*

$$\text{(a)} \quad \psi(y) = 0 \quad \textit{implies} \quad \phi(y) = 0,$$

$$\text{(b)} \quad \nabla\phi(0) \neq 0, \qquad \nabla\psi(0) \neq 0,$$

where ∇ indicates gradient. Then $a(y) = \phi(y)/\psi(y)$ is C^∞ and nonvanishing on some neighborhood of the origin. Moreover, sgn $a(0) =$ sgn$\langle\nabla\phi(0), \nabla\psi(0)\rangle$.

We derive Proposition 4.2 from a preliminary result.

Lemma 4.3. *Let $\psi\colon \mathbb{R}^n \to \mathbb{R}$ be a smooth function defined on a neighborhood of 0 such that $\psi(0) = 0$ and $\nabla\psi(0) \neq 0$. Then there exists a diffeomorphism $Y(y)$ such that $Y(0) = 0$ and $\psi(Y(y)) = y_n$, where y_n is the* nth *coordinate function of $y = (y_1, \ldots, y_n)$.*

Proof. Since $\nabla\psi(0) \neq 0$ there is an index i such that $\partial\psi/\partial y_i(0) \neq 0$. By relabeling the coordinates, if necessary, we may assume that $\partial\psi/\partial y_n(0) \neq 0$. Consider the map

$$\Psi(y) = (y_1, \ldots, y_{n-1}, \psi(y)).$$

Observe that $\det(d\Psi)_0 = \partial\psi/\partial y_n(0) \neq 0$. By the inverse function theorem there exists a smooth mapping $Y(y)$ satisfying $Y(0) = 0$ and $\Psi(Y(y)) = y$. Equating the last coordinates yields $\psi(Y(y)) = y_n$ as desired. □

Proof of proposition 4.2. The property of being C^∞ is, of course, invariant under C^∞ changes of coordinates. Thus we may verify that $a(y)$ is C^∞ in any coordinate system that simplifies the calculations. We choose the system of Lemma 4.3. In other words, we assume without loss of generality that $\psi(y) = y_n$.

The proposition is based on the fundamental theorem of calculus:

$$\phi(y', y_n) - \phi(y', 0) = \int_0^{y_n} \frac{\partial\phi}{\partial y_n}(y', s)\, ds. \tag{4.11}$$

Here $y' = (y_1, \ldots, y_{n-1})$. Now $\phi(y', 0) \equiv 0$ by hypothesis (a) of the theorem, since $\psi(y) = y_n$ vanishes on the hyperplane $\{y_n = 0\}$. We make the substitution $s = ty_n$ in (4.11) to obtain

$$\phi(y', y_n) = y_n \int_0^1 \frac{\partial\phi}{\partial y_n}(y', ty_n)\, dt. \tag{4.12}$$

Thus $\phi(y) = a(y)\psi(y)$, where $a(y)$ is the integral in (4.12). Clearly a is C^∞.

It remains to determine the sign of $a(0)$. Differentiating the relation $\phi = a\psi$ and recalling that $\psi(0) = 0$, we find that

$$\nabla\phi(0) = a(0)\nabla\psi(0).$$

Since $\nabla\phi(0) \neq 0$, we deduce that $a(0) \neq 0$ and moreover that sgn $a(0) =$ sgn$\langle\nabla\phi(0), \nabla\psi(0)\rangle$. □

PROOF OF THEOREM 4.1. In proving Theorem 4.1 we shall need some information from §3 about the mechanics of the Liapunov–Schmidt reduction of $F(y, \alpha) = 0$. We record this information here for reference in the proof. We continue to use the notation of §3 without comment. First, we rephrase (3.6): If $u \in \mathbb{R}^n$, then

$$u = 0 \quad \text{iff} \quad \langle v_0^*, u \rangle = 0 \quad \text{and} \quad Eu = 0. \tag{4.13}$$

Also, we rewrite (3.9) and (3.11) as

$$EF(\Omega(x, \alpha), \alpha) \equiv 0, \tag{4.14}$$

$$g(x, \alpha) = \langle v_0^*, F(\Omega(x, \alpha), \alpha) \rangle, \tag{4.15}$$

where $\Omega: \mathbb{R} \times \mathbb{R}^{k+1} \to \mathbb{R}^n$ is defined by

$$\Omega(x, \alpha) = xv_0 + W(xv_0, \alpha). \tag{4.16}$$

(This formula differs from (3.14) in that here we do not retain the α-coordinate on the right-hand side.)

Let $\mu(y, \alpha)$ be the first eigenvalue of $(dF)_{y,\alpha}$, as described above. We claim that μ is a smooth function of y and α. To motivate the proof, suppose that the entries of $(dF)_{y,\alpha}$ vary smoothly with y and α and that the eigenvalues of a matrix vary smoothly with its entries; composing these two smooth dependences, we see that μ is a smooth function of y and α. Unfortunately, it is not always true that the eigenvalues of a matrix vary smoothly with its entries: this property fails precisely at multiple eigenvalues. However, by (4.9) zero is a simple eigenvalue of $(dF)_{0,0} = L$, so this difficulty does not arise here. This proves the claim. Recall that the rest point of (4.1) corresponding to a solution of $g(x, \alpha) = 0$ is asymptotically stable or unstable according as $\mu(\Omega(x, \alpha), \alpha)$ is positive or negative. (We will abbreviate $\mu(\Omega(x, \alpha), \alpha)$ to $\mu(\Omega, \alpha)$.) Our task is to prove that $\mu(\Omega, \alpha)$ and $g_x(x, \alpha)$ always have the same sign. We do this by invoking Proposition 4.2 to show that $\mu(\Omega, \alpha)$ divided by $g_x(x, \alpha)$ is a positive C^∞ function. To this end, we now prove that

$$g_x(x, \alpha) = 0 \quad \text{implies} \quad \mu(\Omega, \alpha) = 0,$$

i.e., we verify condition (a) in Proposition 4.2.

Suppose that $g_x(x, \alpha) = 0$ for some $(x, \alpha) \in \mathbb{R} \times \mathbb{R}^{k+1}$. Differentiating (4.14) and (4.15) we see that

$$E \cdot dF_{\Omega(x,\alpha),\alpha} \cdot \Omega_x = 0,$$

$$\langle v_0^*, dF_{\Omega(x,\alpha),\alpha} \cdot \Omega_x \rangle = 0.$$

By (4.13), $dF_{\Omega(x,\alpha),\alpha} \cdot \Omega_x = 0$. In other words, zero is an eigenvalue of $dF_{\Omega(x,\alpha),\alpha}$ associated to the eigenvector Ω_x. Thus $\mu(\Omega, \alpha) = 0$, since all the other eigenvalues of $dF_{y,\alpha}$ are bounded away from zero.

In general it is not true that ∇g_x and $\nabla \mu(\Omega)$ are nonzero. Thus condition (b) of Proposition 4.2 is not valid, and the proposition cannot be applied directly. However, we shall use an unfolding trick which allows us to apply Proposition 4.2. Specifically, we insert an extra scalar parameter β into F by defining $\tilde{F}: \mathbb{R}^n \times \mathbb{R}^{k+1} \times \mathbb{R} \to \mathbb{R}^n$,

$$\tilde{F}(y, \alpha, \beta) = F(y, \alpha) + \beta y.$$

We then define $\tilde{\mu}(y, \alpha, \beta)$ to be the eigenvalue of $(d\tilde{F})_{y,\alpha,\beta}$ that is close to zero. Applying the Liapunov–Schmidt reduction to $\tilde{F}$, we obtain a reduced function

$$\tilde{g}(x, \alpha, \beta) = \langle v_0^*, \tilde{F}(\tilde{\Omega}(x, \alpha, \beta), \alpha, \beta)\rangle.$$

The argument above shows that

$$\tilde{g}_x(x, \alpha, \beta) = 0 \quad \text{implies that} \quad \tilde{\mu}(\tilde{\Omega}, \alpha, \beta) = 0.$$

To apply Proposition 4.2, we show that the gradients of these two functions are nonzero. First consider $\tilde{g}_x$. We have from (3.23e)

$$\tilde{g}_{x\beta}(0, 0, 0) = \langle v_0^*, d(\tilde{F}_\beta)\cdot v_0 - d^2\tilde{F}(v_0, L^{-1}E\tilde{F}_\beta)\rangle.$$

However $\tilde{F}_\beta(0, 0, 0) = 0$ and $d(\tilde{F}_\beta)\cdot v_0 = v_0$. Therefore

$$\tilde{g}_{x\beta}(0, 0, 0) = \langle v_0^*, v_0\rangle,$$

and since $\langle v_0^*, v_0\rangle > 0$ (i.e., v_0^* and v_0 were chosen consistently)

$$\tilde{g}_{x\beta}(0, 0, 0) > 0.$$

Next we turn to the β derivative of $\tilde{\mu}(\tilde{\Omega}, \alpha, \beta)$. We claim that

$$\tilde{\Omega}(0, 0, \beta) \equiv 0. \tag{4.17}$$

To see this, observe that $\tilde{F}(0, 0, \beta) \equiv 0$. Thus formula (4.17) satisfies $E\tilde{F}(\tilde{\Omega}(x, \alpha, \beta), \alpha, \beta) = 0$, the analogue of (4.14), at points of the form $(0, 0, \beta)$. But $\tilde{\Omega}(x, \alpha, \beta)$ is obtained by solving this analogue of (4.14) for Ω with the implicit function theorem; thus $\tilde{\Omega}$ is uniquely determined in the solution process. This proves (4.17).

Now

$$(d\tilde{F})_{0,0,\beta}\cdot v_0 = Lv_0 + \beta v_0 = \beta v_0,$$

since $Lv_0 = 0$. Therefore β is an eigenvalue of $(d\tilde{F})_{0,0,\beta}$, with eigenvector v_0. Since the other eigenvalues of $d\tilde{F}$ are bounded away from zero, we must have

$$\tilde{\mu}(0, 0, \beta) = \beta.$$

Combining this equation with (4.17), we see that

$$\tilde{\mu}(\tilde{\Omega}, 0, \beta) = \beta. \tag{4.18}$$

On differentiating (4.18) we conclude that

$$\frac{\partial}{\partial\beta}\tilde{\mu}(\tilde{\Omega}, 0, 0) = 1 > 0.$$

We may now apply Proposition 4.2, obtaining

$$\tilde{\mu}(\tilde{\Omega}, \alpha, \beta) = \tilde{a}(x, \alpha, \beta)\tilde{g}_x(x, \alpha, \beta), \tag{4.19}$$

where $\tilde{a}$ is a smooth function. Moreover,

$$\tilde{a}(0, 0, 0) = \frac{\dfrac{\partial}{\partial \beta}\tilde{\mu}(\tilde{\Omega}, 0, 0)}{\tilde{g}_{x\beta}(0, 0, 0)} > 0.$$

Theorem 4.1 follows on setting $\beta = 0$ in (4.19). □

Remark 4.4. From Theorem 3.2 we know that any two reduced equations g_1 and g_2 obtained by different Liapunov–Schmidt reductions are essentially equivalent. More precisely, we have that $g_2(x, \lambda)$ is equivalent to $\delta g_1(\varepsilon x, \lambda)$ where $\operatorname{sgn} \delta = \operatorname{sgn}\langle v_1^*, v_2^*\rangle$ and $\operatorname{sgn} \varepsilon = \operatorname{sgn}\langle v_1, v_2\rangle$. If v_1, v_1^* and v_2, v_2^* are both consistent choices (i.e., both satisfy (4.10)) then $\delta = \varepsilon$ and we have that $g_2(x, \lambda)$ is equivalent to $\varepsilon g_1(\varepsilon x, \lambda)$. In particular, $g_{1,x}$ and $g_{2,x}$ have the same sign, since $\varepsilon^2 = +1$. Let us show, more generally, that if g and h are equivalent then g_x and h_x have the same sign when g and h vanish. Indeed suppose

$$h(x, \lambda) = S(x, \lambda)g(X(x, \lambda), \Lambda(\lambda)),$$

where $S(0, 0) > 0$, $X_x(0, 0) > 0$. Then

$$h_x = S_x g + S g_x X_x.$$

If $g = 0$ the first term vanishes; thus h_x and g_x have the same sign. Of course, our reason for requiring that $S(x, \lambda) > 0$ and that $X_x(x, \lambda) > 0$ in the definition of equivalence is to obtain this property.

Exercise

4.1. Check the stability of the steady-state solutions to the differential equation

$$\frac{du}{dt} + F(u, \lambda) = 0,$$

where F is defined as in Exercise 3.2; i.e.,

$$F(u_1, u_2, \lambda) = \begin{pmatrix} 2u_1 - 2u_2 + 2u_1^2 + 2u_2^2 - \lambda u_1 \\ u_1 - u_2 + u_1 u_2 + u_2^2 - 3\lambda u_1 \end{pmatrix}.$$

Bibliographical Comments

The various perturbations of the pitchfork described in §1 may be found in Matkowsky and Reiss [1977]; it was proven in Golubitsky and Schaeffer [1979a] that the unfolding of the pitchfork given in Figure 1.5 is universal.

Further references to work on the CSTR are given in Balakotaiah and Luss [1981]. Alternative treatments of the Liapunov–Schmidt reduction may be found in Chow and Hale [1982], Ch. 5; Crandall and Rabinowitz [1971]; Sattinger [1979], Ch. 3; Carr [1981], Ch. 1; and Chapter VII of the present text. (Thompson and Hunt [1973] discuss the same procedure under the name "elimination of the passive coordinates".) Similarly, alternative treatments of the stability of solutions obtained from the Liapunov–Schmidt reduction may be found in the above references, in Crandall and Rabinowitz [1973], and in Kielhofer [1976]; in particular, these references study PDE, not just ODE as we have done in the text.

APPENDIX 1
The Implicit Function Theorem

(a) Finite Dimensions

The implicit function theorem in finite dimensions is concerned with a system of equations of the form

$$f_i(x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_k) = 0, \qquad i = 1, \ldots, n, \tag{A1.1}$$

depending on the k parameters α_j. Specifically, this theorem gives a sufficient condition which guarantees that the system (A1.1) may be solved locally for $x_1, \ldots, x_n$ as functions of the parameters α_j. Note that the number of equations in (A1.1) equals the number of unknowns.

We reformulate (A1.1) using vector notation. Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, $\alpha = (\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$, and $f = (f_1, \ldots, f_n) \in \mathbb{R}^n$. Thus (A1.1) defines a mapping $f: \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ which we assume is s-times differentiable, where $1 \leq s \leq \infty$. For any $(x, \alpha) \in \mathbb{R}^n \times \mathbb{R}^k$ let $(df)_{x,\alpha}$ denote the $n \times n$ Jacobian matrix

$$\left(\frac{\partial f_i}{\partial x_j}(x, \alpha)\right)_{i,j=1,\ldots,n}.$$

We shall work in a neighborhood of a fixed point $(x_0, \alpha_0) \in \mathbb{R}^n \times \mathbb{R}^k$.

Implicit Function Theorem. *Let f be as above. Suppose that $f(x_0, \alpha_0) = 0$ and that*

$$\det(df)_{x_0,\alpha_0} \neq 0.$$

Then there exist neighborhoods U of x_0 in $\mathbb{R}^n$ and V of α_0 in $\mathbb{R}^k$ and a function $X: V \to U$ such that for every $\alpha \in V$, (A1.1) *has the unique solution $x = X(\alpha)$ in U. Moreover, if f is of class C^s so is X. In symbols we have*

$$f(X(\alpha), \alpha) \equiv 0, \qquad X(\alpha_0) = x_0.$$

For the proof of the implicit function theorem, we refer, for example, to Chapter 8 of Taylor and Mann [1983].

Let us illustrate the use of the theorem with three examples. As a first example we refer the reader to Chapter I, §1. Here we consider possible bifurcation in an equation $g(x, \lambda) = 0$ where $g: \mathbb{R}^1 \times \mathbb{R}^1 \to \mathbb{R}^1$. We conclude from the implicit function theorem that $g_x(x_0, \lambda_0) = 0$ is a necessary condition for a solution (x_0, λ_0) of $g(x, \lambda) = 0$ to be a bifurcation point, for otherwise we could solve uniquely for x as a smooth function of λ.

As a second example, let us consider the special case

$$g(x, \alpha) = x^2 - \alpha x.$$

Observe that $x = 0$ is a solution to $g = 0$ for each α. Moreover, $(dg)_{0,\alpha} = -\alpha$, so that for $\alpha \neq 0$ the implicit function theorem guarantees that $x = 0$ is the only solution to $g(x, \alpha) = 0$ near $x = 0$. However, a simple glance at the set $\{(x, \lambda): g(x, \lambda) = 0\}$ shows that the neighborhood on which the implicit function theorem is valid shrinks to a point as α approaches 0.

As a third, less trivial, example we show that limit points are both isolated and persistent to small perturbations. The point (x_0, λ_0) is a *limit point* for $g(x, \lambda) = 0$ if $g(x_0, \lambda_0) = g_x(x_0, \lambda_0) = 0$ and $g_{xx}(x_0, \lambda_0) \neq 0$, $g_\lambda(x_0, \lambda_0) \neq 0$. To show that limit points are isolated, define a mapping $f: \mathbb{R} \times \mathbb{R} \to \mathbb{R}^2$, where

$$f(x, \lambda) = (g(x, \lambda), g_x(x, \lambda)).$$

At the limit point (x_0, λ_0) we have

$$\det(df)_{x_0, \lambda_0} = \det\begin{pmatrix} g_x & g_\lambda \\ g_{xx} & g_{x\lambda} \end{pmatrix}\Bigg|_{(x_0, \lambda_0)} = -g_{xx}g_\lambda|_{(x_0, \lambda_0)} \neq 0.$$

Thus the implicit function theorem, applied with $k = 0$, implies that (x_0, λ_0) is the only solution to $g = g_x = 0$ on a neighborhood of (x_0, λ_0).

We next show that limit points are persistent to small perturbations. Let

$$G(x, \lambda, \varepsilon) = g(x, \lambda) + \varepsilon p(x, \lambda),$$

where p is any perturbation term and ε is small. The limit points of $G(\cdot, \cdot, \varepsilon)$ must satisfy

$$F(x, \lambda, \varepsilon) \equiv (G(x, \lambda, \varepsilon), G_x(x, \lambda, \varepsilon)) = 0.$$

Observe that $\det(dF)_{x_0, \lambda_0, 0} = \det(df)_{x_0, \lambda_0} \neq 0$. Thus the implicit function theorem, applied with $k = 1$, guarantees that the solutions to $F = 0$ near $(x_0, \lambda_0, 0)$ have the form $(X(\varepsilon), \Lambda(\varepsilon), \varepsilon)$ where $(X(0), \Lambda(0)) = (x_0, \lambda_0)$. Thus for each ε sufficiently small, $G(\cdot, \cdot, \varepsilon)$ has a unique limit point near (x_0, λ_0).

(b) Infinite Dimensions

In Chapter VII, we will need to apply an infinite-dimensional version of the implicit function theorem which we formulate here. First we define a C^1 mapping $\Phi: \mathscr{X} \to \mathscr{Y}$ between Banach spaces. The mapping Φ is called (Fréchet) *differentiable* at a point $u \in \mathscr{X}$ if there is a bounded linear mapping $L: \mathscr{X} \to \mathscr{Y}$ such that

$$\|\Phi(u + v) - \Phi(u) - Lv\| = o(\|v\|) \tag{A1.2}$$

for v in some neighborhood of zero in $\mathscr{X}$. The linear operator in (A1.2) will be denoted $(d\Phi)_u$, the differential of Φ at u. We will say that Φ is of class C^1 if Φ is differentiable for every $u \in \mathscr{X}$ and the mapping $u \to d\Phi_u$ is continuous in the norm topologies. (*Remarks*: (i) The reader should understand that this definition and the theorem below apply equally well to a mapping into $\mathscr{Y}$ defined on an open subset of $\mathscr{X}$; we ignore this generalization, as it complicates the notation without adding insight. (ii) Mappings of class C^s are defined in Appendix A3.)

Let $\Phi: \mathscr{X} \times \mathscr{Y} \to \mathscr{Z}$ be a C^1 mapping between Banach spaces. Let $(d\Phi)_{u,v}: \mathscr{X} \to \mathscr{Z}$ denote the differential of Φ (with respect to $\mathscr{X}$ only). Consider the equation

$$\Phi(u, v) = 0 \tag{A1.3}$$

near a fixed point, say (0, 0), such that $\Phi(0, 0) = 0$.

Implicit Function Theorem for Banach Spaces. *Let Φ be as defined above and suppose that $(d\Phi)_{0,0}: \mathscr{X} \to \mathscr{Z}$ has a bounded inverse. Then* (A1.3) *may be solved locally for $u = \Psi(v)$, where $\Psi: \mathscr{Y} \to \mathscr{X}$ is a C^1 mapping.*

A comprehensive treatment of the implicit function theorem in infinite dimensions is given in Chapter 2 of Chow and Hale [1982].

In the applications of this theorem in Chapter VII, Φ is typically a differential operator. For such an operator to be bounded it is essential to allow the domain and range of Φ to be different spaces. For example, the Laplacian is bounded from $C^2(\Omega)$ to $C^0(\Omega)$ but is *not* bounded operating from any Banach space to itself.

Exercises

A1.1. A function $f: \mathbb{R}^n \to \mathbb{R}$ has a *nondegenerate singularity* at x if $(df)_x = 0$ and $\det(d^2f)_x \neq 0$ where d^2f is the Hessian matrix $(\partial^2 f/\partial x_i\, \partial x_j)$. Show that nondegenerate singularities are isolated and persistent to small perturbations of f.

A1.2. The mapping $F: \mathbb{R}^n \to \mathbb{R}^n$ has a *nondegenerate fixed point* at x if $F(x) = x$ and 1 is not an eigenvalue of the $n \times n$ Jacobian matrix $(dF)_x$. Show that nondegenerate fixed points are isolated and persistent to small perturbations of F.

APPENDIX 2
Equivalence and the Liapunov–Schmidt Reduction

In this appendix we prove Theorem I,3.2. We recall the notation of that theorem—g_1 and g_2 are two different reduced functions obtained by applying the Liapunov–Schmidt reduction to (I,3.1). Each reduced function g_i, $i = 1, 2$, depends on four arbitrary choices; viz., on subspaces M_i and N_i as in (I,3.4) and (I,3.5), and on vectors v_i and v_i^* in ker L and (range $L)^\perp$. Our task is to show that, up to the $\pm$ signs indicated in the theorem, g_1 and g_2 are equivalent. We consider in turn the four cases where three of the four choices needed for the reduction are the same for g_1 and g_2 but the fourth is different. Since equivalence is a transitive relationship, the general case follows by combining these four special cases. As it happens two of these four cases are trivial to analyze, and we deal with both of these simultaneously in Case I below.

Case I. Suppose $M_1 = M_2$, $N_1 = N_2$, but possibly $v_1 \neq v_2$ and/or $v_1^* \neq v_2^*$.

This case is easy because the reduced equation ϕ_i: ker $L \times \mathbb{R} \to N_i$, $i = 1, 2$, obtained in Step 4 of the reduction is the same for g_1 and g_2. (We will therefore simply write ϕ, omitting the subscript.) The only difference between g_1 and g_2 is in the parametrization of ker L and N. We have $v_2 = cv_1$ and $v_2^* = dv_1^*$ for some coefficients c and/or d, where $c \neq 0$ and $d \neq 0$. Indeed sgn $c = \varepsilon$ and sgn $d = \delta$. Now

$$\begin{aligned} g_2(x, \lambda) &= \langle dv_1^*, \phi(cxv_1, \lambda)\rangle \\ &= dg_1(cx, \lambda). \end{aligned}$$

This equation yields the equivalence of $g_2(x, \lambda)$ and $\delta g_1(\varepsilon x, \lambda)$; specifically we may take $S(x, \lambda) = |d|^{-1}$, $X(x, \lambda) = |c|^{-1}x$, and $\Lambda(\lambda) = \lambda$ in (I,1.6).

For the remaining two cases we recall from Chapter I, §3(c) the geometric view of the Liapunov–Schmidt reduction. Let

$$\mathscr{V}_i = \{(y, \lambda): E_i\Phi(y, \lambda) = 0\}, \qquad i = 1, 2. \tag{A2.1}$$

Each $\mathscr{V}_i$ is a two-dimensional submanifold of $\mathbb{R}^n \times \mathbb{R}$ whose tangent space at the origin is ker $L \times \mathbb{R}$. The submanifolds $\mathscr{V}_i$ may be parametrized by

$$\Omega_i(v, \lambda) = (v + W_i(v, \lambda), \lambda). \tag{A2.2}$$

(Cf. (I,3.14).) Moreover

$$g_i(x, \lambda) = \langle v_i^*, \Phi \circ \Omega_i(xv_i, \lambda)\rangle. \tag{A2.3}$$

(Cf. (I, 4.15).)

Case II. Suppose that $N_1 = N_2$, $v_1 = v_2$, $v_1^* = v_2^*$ but possibly $M_1 \neq M_2$.

The projection E_i depends only on N_i, and we have $N_1 = N_2$. Thus in (A2.1) we have $\mathscr{V}_1 = \mathscr{V}_2$. However, the *parametrization* of $\mathscr{V} = \mathscr{V}_1 = \mathscr{V}_2$ by Ω_i in (A2.2) does depend on the choice of M_i, and this point is the only difference between g_1 and g_2. This observation is the basis of our proof of equivalence in this case.

Let us elaborate. Let $\pi_1: \mathbb{R}^n \times \mathbb{R} \to \ker L \times \mathbb{R}$ be the projection with kernel $M_1 \times \{0\}$. We claim that $\Omega_1 \circ \pi_1 | \mathscr{V}$ is the identity. It may be seen from (A2.2) that for composition in the reverse order we have $\pi_1 \circ \Omega_1(v, \lambda) = (v, \lambda)$, whence $\Omega_1 \circ \pi_1 \circ \Omega_1(v, \lambda) = \Omega_1(v, \lambda)$. But Ω_1 parametrizes $\mathscr{V}$, so any point in $\mathscr{V}$ has the form $\Omega_1(v, \lambda)$ for some (v, λ). The claim follows.

Since $v_1 = v_2$ and $v_1^* = v_2^*$, we use a common subscript "zero" for these vectors. (v without a subscript is a generic element of ker L.) We insert the identity into the representation (A2.3) of g_2 as follows.

$$\begin{aligned} g_2 &= \langle v_0^*, \Phi \circ \Omega_2(xv_0, \lambda)\rangle \\ &= \langle v_0^*, \Phi \circ \Omega_1 \circ \pi_1 \circ \Omega_2(xv_0, \lambda)\rangle. \end{aligned}$$

We shall extract the required diffeomorphism from the factor $\pi_1 \circ \Omega_2$. Let us elaborate. Since π_1 maps into ker $L \times \mathbb{R}$, there exists a smooth function X such that

$$\pi_1 \circ \Omega_2(xv_0, \lambda) = (X(x, \lambda)v_0, \lambda). \tag{A2.4}$$

Thus

$$\begin{aligned} g_2(x, \lambda) &= \langle v_0^*, \Phi \circ \Omega_1(X(x, \lambda)v_0, \lambda)\rangle \\ &= g_1(X(x, \lambda), \lambda). \end{aligned}$$

Hence g_2 can be obtained from g_1 by composition with a change of coordinates, and it remains only to show that $\partial X/\partial x(0, 0) > 0$. Applying the chain rule to (A2.4) we observe that $\partial X/\partial x(0, 0)v_0$ equals the first component of $d\pi_1 \circ d\Omega_2 \cdot (v_0, 0)$, which we abbreviate to $d\pi_1 \cdot d\Omega_2 \cdot v_0$. However by (I,3.15) we have $\partial W_2/\partial x = 0$, so $d\Omega_2 \cdot v_0 = v_0$. Moreover, $d\pi_1 = \pi_1$, since π_1 is linear, and $\pi_1 v_0 = v_0$. Combining these, we conclude that $\partial X/\partial x(0, 0) = 1$.

Case III. $M_1 = M_2$, $v_1 = v_2$, $v_1^* = v_2^*$, but possibly $N_1 \neq N_2$.

The basic observation here is the following containment:

$$\{(y, \lambda): \Phi(y, \lambda) = 0\} \subset \mathscr{V}_1 \cap \mathscr{V}_2.$$

We deduce from this observation that for $v \in \ker L$, $\phi_1(v, \lambda)$ vanishes if and only if $\phi_2(v, \lambda) = 0$; thus, introducing coordinates, we see that

$$g_1(x, \lambda) = 0 \quad \text{iff} \quad g_2(x, \lambda) = 0. \tag{A2.5}$$

We now want to apply Proposition I,4.2 to show that $g_2(x, \lambda)/g_1(x, \lambda)$ is a positive C^∞ function near the origin. For this we need to show $\nabla g_i \neq 0$, $i = 1, 2$. We have from (I,3.23a) that $\partial g_i/\partial x(0, 0) = 0$, but from (I,3.23d)

$$\frac{\partial g_1}{\partial \lambda}(0, 0) = \left\langle v_0^*, \frac{\partial \Phi}{\Phi \lambda}(0, 0) \right\rangle = \frac{\partial g_2}{\partial \lambda}(0, 0).$$

Therefore, if either derivative is nonzero, then so is the other, and moreover $\langle \nabla g_1, \nabla g_2 \rangle > 0$. Thus in this case

$$g_2(x, \lambda) = S(x, \lambda) g_1(x, \lambda)$$

for some positive, C^∞ function $S(x, \lambda)$; in particular, g_2 is equivalent to g_1.

However, more commonly it will happen that $\nabla g_i(0, 0) = 0$. We cannot then apply Proposition I,4.2 directly, but must resort to a trick. (Cf. the proof of Theorem I, 4.1.) We "unfold" Φ by introducing an additional parameter β as follows. Define a function $\Phi: \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^n$,

$$\tilde{\Phi}(y, \lambda, \beta) = \Phi(y, \lambda) + \beta v_0^*.$$

Now apply the Liapunov–Schmidt reduction to $\tilde{\Phi}$, obtaining

$$\tilde{\phi}_i: \ker L \times \mathbb{R} \times \mathbb{R} \to N_i$$

as in (I,3.10) and $\tilde{g}_i(x, \lambda, \beta)$ as in (I,3.11). As before, the zeros of $\tilde{g}_i(x, \lambda, \beta)$ are in a natural one-to-one correspondence with the zeros of $\tilde{\Phi}(y, \lambda, \beta)$. Likewise define the solution manifolds $\tilde{\mathscr{V}}_i = \{(y, \lambda, \beta): E_i \tilde{\Phi}(y, \lambda, \beta) = 0\}$ and note that

$$\{\tilde{\Phi}(y, \lambda, \beta) = 0\} \subset \tilde{\mathscr{V}}_1 \cap \tilde{\mathscr{V}}_2.$$

This fact implies, as in (A2.5), that

$$\tilde{g}_1(x, \lambda, \beta) = 0 \quad \text{if and only if} \quad \tilde{g}_2(x, \lambda, \beta) = 0.$$

We claim that $(\partial/\partial\beta)\tilde{g}_i(0, 0, 0) > 0$. Indeed, applying (I,3.23d) we have

$$\frac{\partial \tilde{g}_i}{\partial \beta}(0, 0, 0) = \left\langle v_0^*, \frac{\partial \tilde{\Phi}}{\partial \beta} \right\rangle = \langle v_0^*, v_0^* \rangle > 0$$

as claimed. Thus we may apply Proposition I,4.2 to functions of the three variables x, λ, and β to conclude that $g_2(x, \lambda, \beta) = S(x, \lambda, \beta)g_1(x, \lambda, \beta)$ for some C^∞, positive function $S(x, \lambda, \beta)$. On restricting to $\beta = 0$ we obtain the desired equivalence. □

CHAPTER II

The Recognition Problem

§0. Introduction

In this chapter we consider a notion of equivalence slightly different from that of Chapter I. We say that two smooth mappings $g, h: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined near the origin are *strongly equivalent* if there exist functions $X(x, \lambda)$ and $S(x, \lambda)$ such that the relation

$$g(x, \lambda) = S(x, \lambda)h(X(x, \lambda), \lambda) \tag{0.1}$$

holds near the origin. This definition differs from our earlier version (I,1.8) in that here we do not allow the bifurcation parameter λ to be transformed. In this definition we still require that

$$X(0, 0) = 0, \qquad X_x(x, \lambda) > 0, \qquad S(x, \lambda) > 0. \tag{0.2}$$

Here and below we work in the neighborhood of the origin in $\mathbb{R}^2$; this is merely for convenience, as we could equally well work near any given point.

This chapter is entitled "The Recognition Problem," by which we understand the following. Consider a smooth mapping $h: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined near (0, 0). To solve the recognition problem for h means to characterize explicitly the mappings which are strongly equivalent to h. For example, we will show in this chapter that a mapping g is strongly equivalent to $x^3 - \lambda x$ if

$$g(0, 0) = g_x(0, 0) = g_{xx}(0, 0) = g_\lambda(0, 0) = 0 \tag{0.3a}$$

and

$$g_{xxx}(0, 0) > 0, \qquad g_{\lambda x}(0, 0) < 0 \tag{0.3b}$$

(Cf. (I,1.6).) In other words (0.3) solves the recognition problem for $x^3 - \lambda x$.

In this chapter we present algorithms which lead to a solution of the recognition problem for a rather general mapping h, and we carry out the calculations required by the algorithms in a number of interesting special cases, $x^3 - \lambda x$ among them. This information is useful in applications, in that one can test whether an equation $g(x, \lambda) = 0$ coming from a specific application is equivalent to another, presumably simpler, equation $h(x, \lambda) = 0$. However, this information is incomplete, in that it does not tell how to derive an appropriate h—throughout this chapter we assume that h is already given. We shall refer to h as a *normal form* for the bifurcation problem. Typically h will be the simplest representative from a whole equivalence class of mappings.

The issue of obtaining an appropriate normal form for a bifurcation problem is one of the most subtle aspects of our theory. At this early stage of the exposition, we limit ourselves to the following two simple remarks concerning this issue:

(i) Given a specific mapping $g(x, \lambda)$, it may be possible to compute which of its derivatives vanish at the bifurcation point. Suppose we have a list

$$\left(\frac{\partial}{\partial x}\right)^{k_i}\left(\frac{\partial}{\partial \lambda}\right)^{l_i} g(x, \lambda) = 0, \qquad i = 1, 2, \ldots, N. \tag{0.4}$$

A natural choice for h is a linear combination of the lowest-order monomials $x^k\lambda^l$ *not* on this list.

However, even in relatively simple examples, this procedure requires care in its implementation. For example, the monomials associated to the list (0.3a) are $1, x, \lambda, x^2$; the lowest order monomials not on this list would seem to be $x\lambda$, λ^2, and x^3. As (0.3b) suggests, we need only consider the x^3 and $x\lambda$ terms in h. In many cases the classification theorem (Theorem 2.1 of Chapter IV) eliminates the need for computation in carrying out this procedure—for several of the simpler bifurcation problems, this theorem gives the normal form associated to a list of vanishing derivatives such as (0.4). (Corollary 9.1 of this chapter is a less complete result in this direction.)

(ii) Often in applications one has some idea of what kind of bifurcation diagrams to expect. In the figures of Chapter IV, §4 we tabulate the bifurcation diagrams associated to the normal forms considered in Theorem IV,2.1. Scanning this table and looking for the expected behavior of the applied problem is another way to generate candidates for a normal form.

Of course these remarks are terribly sketchy. We refer the reader to the Case Studies for a more complete presentation of how we choose normal forms in specific problems.

At this point we present our solution of the recognition problem for another example. We consider the normal forms $h(x, \lambda) = \varepsilon(x^2 + \delta\lambda^2)$, where ε and δ equal ± 1. If $\delta = -1$, the zero set of h consists of two crossed lines; if $\delta = +1$, it consists of the single point $x = \lambda = 0$. Thus, the bifurcation diagrams in these two cases are quite different, although algebraically the two cases are quite similar. This example foreshadows some of the nuisance that plus and minus signs create. In Proposition 9.3 below we obtain the following solution to the recognition problem for $\varepsilon(x^2 + \delta\lambda^2)$: if a mapping g satisfies

$$g(0, 0) = g_x(0, 0) = g_\lambda(0, 0) = 0 \tag{0.5a}$$

and

$$g_{xx}(0, 0) \neq 0, \qquad \det d^2g(0, 0) \neq 0, \tag{0.5b}$$

then g is strongly equivalent to $\varepsilon(x^2 + \delta\lambda^2)$, where $\varepsilon = \operatorname{sgn} g_{xx}(0, 0)$ and $\delta = \operatorname{sgn} \det(d^2g(0, 0))$. Here d^2g stands for the Hessian matrix

$$d^2g = \begin{pmatrix} g_{xx} & g_{\lambda x} \\ g_{\lambda x} & g_{\lambda\lambda} \end{pmatrix}.$$

In the rest of §0 we attempt to give an overview of the contents of Chapter II. The discussion in the next paragraph starts this effort.

The most noteworthy feature of (0.3) and (0.5) is that these conditions only involve a small number of the derivatives of g. We say that h is *finitely determined* if we need compute only a finite number of terms in the Taylor expansion of g when deciding whether g is strongly equivalent to h. Let us explore some consequences of this fact. Consider, for example, the second case, $h = \varepsilon(x^2 + \delta\lambda^2)$. According to our solution of the recognition problem, if p is any polynomial all of whose terms are of degree 3 or greater, then $h + p$ is strongly equivalent to h. For an arbitrary monomial $x^k\lambda^l$, let us ask for what a, if any, is $h + ax^k\lambda^l$ strongly equivalent to h? This question separates monomials into three classes as follows:

(i) Low-order terms: The derivative $(\partial/\partial x)^k(\partial/\partial\lambda)^l g$ must vanish, so equivalence only obtains if $a = 0$.
(ii) Higher-order terms: The derivative $(\partial/\partial x)^k(\partial/\partial\lambda)^l g$ nowhere appears in (0.5), so equivalence obtains for all a.
(iii) Intermediate-order terms: Whatever terms are not included above.

In general, our solution of the recognition problem for a normal form h will split monomials into three classes in this fashion. The exact definition of higher-order terms is more complicated than is indicated in (ii); however, (ii) does convey the essential spirit of the correct definition. For the specific case of (0.5) the three classes may be related to degrees of homogenity as follows:

(i) Low order: degree ≤ 1.
(ii) Intermediate order: degree $= 2$.
(iii) Higher order: degree ≥ 3.

Usually the three classes will not mesh so nicely with degrees of homogeneity.

The main results of Chapter II are formulated in §8 where we describe the algorithm which solves the recognition problem for a mapping h. This algorithm has three parts, corresponding to the three kinds of monomials mentioned above. The first part enumerates the low-order terms; this is a rather simple task. The second part provides an algorithm to characterize the higher-order terms; the algorithm may be carried out even in rather complicated bifurcation problems. The third part identifies the conditions that the intermediate-order terms must satisfy for equivalence to obtain; it will appear in Chapter V that these conditions can in general be rather complicated, but in the present chapter, they will always have the form of inequalities, as in (0.3b) and (0.5b).

In §9 we apply the results of §8 to derive normal forms for several important classes of bifurcation problems. Sections 11–12 contain the proofs of results that either were too long to give earlier or were extraneous to the main sequence of ideas.

Sections 1–7 develop the material needed for an efficient presentation of our main results. In §1 we introduce the formalism of germs, which is a notational convention incorporating the fact that all our results are only valid in a sufficiently small neighborhood of some fixed point. The fundamental theoretical concept of the chapter, the restricted tangent space $RT(g)$, is defined in §2; loosely speaking, $RT(g)$ consists of those mappings p such that for t small, $h + tp$ is strongly equivalent to h, modulo error terms that are $\mathcal{O}(t^2)$. It is important to be able to compute $RT(g)$ efficiently; §§3–7 are a unit which addresses this issue. In §§3 and 6, we compute $RT(g)$ for certain simple and complicated examples, respectively. In §§4, 5, and 7 we abstract general principles from the calculations of the earlier sections; we give these their natural, algebraic formulation.

In §10, we consider the recognition problem in the context of general equivalences; i.e., those where the bifurcation parameter may be transformed. (This is to be contrasted with strong equivalences as defined in (0.1).) Actually our principal interest lies with general equivalences, but the mathematical development of the subject is simplified by considering strong equivalences first. In any case, we use essentially the same techniques in either context, and for simple normal forms the solution to the recognition problem is the same in either context. In particular, (0.3) and (0.5) solve the recognition problem for their respective normal forms in either context.

§1. Germs: A Preliminary Issue

The theory we are presenting is a local one; i.e., our results are valid only in a sufficiently small neighborhood of some fixed point. The terminology of germs provides a convenient way of formulating results in a local theory which avoids infinite repetition of the phrase "in a sufficiently small neighborhood of the origin."

To better understand what it means that our theory is only local, let us consider the function

$$g(x, \lambda) = x^3 - \lambda x + ax^4.$$

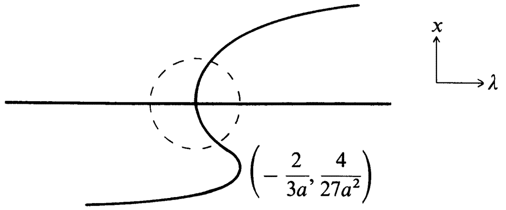


Figure 1.1. Solution set of $x^3 - \lambda x + ax^4 = 0$; $a > 0$.

Now g satisfies conditions (0.3), so by the results quoted above, g is strongly equivalent to the pitchfork $x^3 - \lambda x$. However, the zero set of g, shown in Figure 1.1 when $a > 0$, consists of the two curves $x = 0$ and $\lambda = x^2 + ax^3$. Intuitively, it is clear that $x^3 - \lambda x + ax^4$ can be equivalent to $x^3 - \lambda x$ only on a neighborhood which is not too large. Specifically, equivalence can only hold on a neighborhood which excludes the limit point of $\lambda = x^2 + ax^3$ at $(x, \lambda) = (-2/3a, 4/27a^2)$; such a neighborhood is indicated by the dotted lines in Figure 2.1. (This intuition may be supported by a rigorous argument based on counting solutions as in formula (I,1.9).) Moreover, the largest disk on which equivalence obtains in fact shrinks to the origin as $a \to \infty$. From this example we see that it is impossible to choose one fixed neighborhood of the origin on which (0.3) would imply pitchfork-like behavior.

Given suitable bounds for higher-order derivatives of g it is possible to estimate the size of the neighborhood on which an equivalence obtains. However, the situation concerning such results is similar to the situation for the implicit function theorem. This theorem (see Appendix 1) gives sufficient conditions which guarantee that an equation $F(x, y) = 0$ may be solved for x as a function of y, say $x(y)$. We can estimate the exact domain of the function $x(y)$, but for most applications this is unnecessary, and usually such estimates are hard to apply.

Moreover, to be completely explicit about the domains of functions can be rather a nuisance. For example, if $g(x, \lambda)$ is defined on U, a neighborhood of the origin in $\mathbb{R}^2$, and if $h(x, \lambda)$ is defined on V, then the sum $g + h$ or product gh is only defined on $U \cap V$. But what is the relation of the restriction $g \mid U \cap V$ to the original function g? For purposes of a local theory they are effectively indistinguishable.

We shall say that two functions (defined near the origin, possibly on different sets) are *equal as germs* if there is some neighborhood of the origin on which they coincide. This definition applies to artificial examples such as $g_1(x, \lambda) = x$ and

$$g_2(x, \lambda) = \begin{cases} x & \text{if } \lambda \leq 1, \\ x + e^{-1/(\lambda - 1)} & \text{if } \lambda > 1, \end{cases}$$

since g_1 and g_2 coincide in the disk $\{(x, \lambda): x^2 + \lambda^2 < 1\}$. More importantly, this definition speaks to the question of the preceding paragraph: If g is defined on a neighborhood of zero, then g and $g \mid U \cap V$ are equal as germs.

In general, if V_1 and V_2 are neighborhoods of the origin, then $g|U \cap V_1$ and $g|U \cap V_2$ are equal in this sense.

Let $\mathscr{E}_{x,\lambda}$ denote the space of all functions $g: \mathbb{R}^2 \to \mathbb{R}$ that are defined and C^∞ on some neighborhood of the origin. We shall identify any two functions in $\mathscr{E}_{x,\lambda}$ which are equal as germs. We call the elements of $\mathscr{E}_{x,\lambda}$ germs. (In more technical language, a germ is an equivalence class in $\mathscr{E}_{x,\lambda}$ with respect to this identification. A similar issue concerning equivalence classes arises in integration theory, where we want to identify functions that are equal almost everywhere.) We shall sometimes write $\mathscr{E}_2$ or simply $\mathscr{E}$ for this space—the subscript "2" indicates the number of variables. Similarly, we will occasionally need $\mathscr{E}_n$, germs of functions of n variables, and $\mathscr{E}_\lambda$, germs of functions of λ alone.

Germ concepts allow us to shrink the domain of a function as needed. For example, if $g \in \mathscr{E}_{x,\lambda}$ and $g(0, 0) \neq 0$, then $1/g \in \mathscr{E}_{x,\lambda}$, since $1/g$ is defined and smooth in *some* neighborhood of the origin. It does not matter if that neighborhood is smaller than the original domain of g. Similarly, in formula (0.2) it suffices to require that

$$X_x(0, 0) > 0, \qquad S(0, 0) > 0;$$

by continuity these inequalities continue to hold in some neighborhood of the origin, and what happens away from such a neighborhood is irrelevant.

In dealing with germs the reader should bear the following points in mind:

(i) Evaluation of a germ g at a fixed point (x_0, λ_0) different from (0, 0) is not compatible with germ concepts, since for any such point (x_0, λ_0) there always exist smooth functions g_1 and g_2 which are equal as germs but still satisfy $g_1(x_0, \lambda_0) \neq g_2(x_0, \lambda_0)$. On the other hand, any derivative of a germ evaluated at the origin *is* well defined (See Exercise 1.1.)

(ii) Limit processes with germs are suspect, since the domains may shrink to a point in the limit. Of course, finite processes such as addition and multiplication pose no such problem.

Exercise

1.1. (a) Let g be a germ in $\mathscr{E}_n$. Show that $g(0)$ is well defined; that is, choose two functions f_1 and f_2 representing g and show that $f_1(0) = f_2(0)$.

(b) Show that $\partial g/\partial x_j$ is a well-defined germ in $\mathscr{E}_{x,\lambda}$ by showing that $\partial f_1/\partial x_j$ and $\partial f_2/\partial x_j$ define the same germs.

(c) Show that $(\partial/\partial x_1)^{\alpha_1} \cdots (\partial/\partial x_n)^{\alpha_n} g(0)$ is well defined.

§2. The Restricted Tangent Space

In this section we define the restricted tangent space of a germ g in $\mathscr{E}_{x,\lambda}$; this definition is a fundamental theoretical concept that underlies most of Chapter II. This concept arises naturally from the following question: "Given a germ g, for what perturbations p is $g + tp$ strongly equivalent to g for all small t?" Suppose that for some perturbation p the answer is affirmative; then for all small t there exist functions $S(x, \lambda, t)$ and $X(x, \lambda, t)$ such that

$$g(x, \lambda) + tp(x, \lambda) = S(x, \lambda, t)g(X(x, \lambda, t), \lambda), \tag{2.1}$$

where

$$X(0, 0, t) \equiv 0. \tag{2.2}$$

Suppose further that S and X vary smoothly in x, λ and t and that at $t = 0$, S and X define the identity transformation on g; in symbols

$$S(x, \lambda, 0) \equiv 1, \qquad X(x, \lambda, 0) \equiv x. \tag{2.3}$$

(*Remark*: Because of (2.3) we need not assume explicitly that $X_x(0, 0, t) > 0$, $S(0, 0, t) > 0$—this follows for small t by continuity.) We differentiate (2.1) with respect to t, set $t = 0$, and use (2.3) to simplify the right-hand side; this yields

$$p(x, \lambda) = \dot{S}(x, \lambda, 0)g(x, \lambda) + g_x(x, \lambda)\dot{X}(x, \lambda, 0),$$

where dot indicates a t derivative. Note that by (2.2), $\dot{X}_0(0, 0, 0) = 0$.

The restricted tangent space $RT(g)$ is defined as the totality of functions that arise through the above construction. Let us formalize this in the following definition.

Definition 2.1. The *restricted tangent space* of a germ g, denoted by $RT(g)$, is the set of all germs p which may be written in the form

$$p(x, \lambda) = a(x, \lambda)g(x, \lambda) + b(x, \lambda)g_x(x, \lambda), \tag{2.4}$$

where $a, b \in \mathscr{E}_{x, \lambda}$ and $b(0, 0) = 0$.

We use the word "restricted" to indicate a construction associated with strong equivalence. In the next chapter we shall consider an unrestricted tangent space associated with ordinary equivalence. For our purposes here, the restricted tangent space leads to a simpler theory.

It follows from the discussion above that $p \in RT(g)$ is a necessary condition for $g + tp$ to be strongly equivalent to g when t is small. It is not a sufficient condition, but we can prove the following theorem.

Theorem 2.2. *Let g and p be germs in $\mathscr{E}_{x, \lambda}$. If*

$$RT(g + tp) = RT(g) \quad \textit{for all} \quad t \in [0, 1], \tag{2.5}$$

then $g + tp$ is strongly equivalent to g for all $t \in [0, 1]$.

This theorem is proved in §11. There we construct an appropriate S and X to show equivalence by solving certain ODE. Moreover, we will use the ideas involved in this construction later in the text.

Let us show that if condition (2.5) is satisfied then $p \in RT(g)$. Indeed, suppose

$$RT(g + tp) = RT(g) \tag{2.6}$$

for just one nonzero t. Certainly $g + tp \in RT(g + tp)$, as we may choose $a \equiv 1$, $b \equiv 0$ in (2.4). Thus, by (2.5) we see that

$$g + tp = ag + bg_x$$

for some $a, b \in \mathscr{E}_{x,\lambda}$ where $b(0, 0) = 0$. Subtracting g from both sides and dividing by t we obtain

$$p = \frac{a - 1}{t} g + \frac{b}{t} g_x,$$

which shows $p \in RT(g)$.

§3. Calculation of $RT(g)$, I: Simple Examples

Our main purpose in this section is to calculate $RT(g)$ for the following two simple examples:

$$\begin{aligned} &\text{(a)} \quad g = x^2 + \lambda \quad \text{(limit point)},\\ &\text{(b)} \quad g = x^3 - \lambda x \quad \text{(pitchfork)}. \end{aligned} \tag{3.1}$$

These calculations lead to general principles for determining $RT(g)$ that are an essential part of our theory.

(a) Preliminaries Needed for the Calculation

According to Definition 2.2, a germ $f \in \mathscr{E}_{x,\lambda}$ belongs to $RT(g)$ if it may be written in the form (2.4). In this formula the condition that $b(0, 0) = 0$ is something of a nuisance. We will use Lemma 3.1 below to obtain a more convenient formula; since we will usually prefer this reformulation of (2.4) to (2.4) itself, we incoprorate the reformulation into Lemma 3.2 for later reference.

Lemma 3.1. *Let $f(x)$ be a germ in $\mathscr{E}_n$ with $f(0) = 0$. Then there exist smooth germs $a_1, \ldots, a_n$ in $\mathscr{E}_n$ such that*

$$f(x) = x_1 a_1(x) + \cdots + x_n a_n(x),$$

where $x = (x_1, \ldots, x_n)$.

We apply the lemma to reformulate (2.4) before giving the proof.

Lemma 3.2. *Let $g \in \mathscr{E}_{x,\lambda}$. A germ f belongs to $RT(g)$ if and only if there exist germs a, b, $c \in \mathscr{E}_{x,\lambda}$ such that*

$$f = ag + (xb + \lambda c)g_x. \tag{3.2}$$

Proof of Lemma 3.2. If f has the form (3.2), then f also has the form (2.4), since the coefficient of g_x in (3.2) vanishes at the origin. Conversely, suppose f has the form (2.4). Then $b(0, 0) = 0$, so by Lemma 3.1 there exist smooth coefficients $\tilde{b}$ and $\tilde{c}$ such that

$$b(x, \lambda) = x\tilde{b}(x, \lambda) + \lambda\tilde{c}(x, \lambda).$$

We obtain (3.2) on substituting this representation for $b(x, \lambda)$ into (2.4) □

Proof of Lemma 3.1. Suppose f is defined on some ball $B_\varepsilon \subset \mathbb{R}^n$. For any fixed $x \in B_\varepsilon$ define a function of one variable $h(s) \equiv f(sx)$, where $0 \leq s \leq 1$. Note that $h(0) = f(0) = 0$. Now

$$f(x) = h(1) - h(0) = \int_0^1 \frac{dh}{ds}(s)\, ds.$$

By the chain rule

$$\frac{dh}{ds}(s) = \sum_{i=1}^n x_i \frac{\partial f}{\partial x_i}(sx).$$

If we define

$$a_i(x) = \int_0^1 \frac{\partial f}{\partial x_i}(sx)\, ds,$$

then we obtain the desired representation for f. □

It will be useful below to generalize Lemma 3.1. This generalization is just a version of Taylor's theorem; the new wrinkle in Lemma 3.3 is a specific form for the remainder of order k which shows, in particular, that this remainder is itself a smooth function. If $f \in \mathscr{E}_n$, we use the following notation for the kth-order Taylor polynomial of f at the origin (or k-jet, as it is often called):

$$j^k f(x) = \sum_{|\alpha| \leq k} \frac{1}{\alpha!} \left(\frac{\partial}{\partial x}\right)^\alpha f \bigg|_{x=0} x^\alpha.$$

Here $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a multi-index of nonnegative integers and we observe the standard conventions with multi-indices; thus

$$|\alpha| = \alpha_1 + \cdots + \alpha_n, \qquad \alpha! = (\alpha_1)! \cdots (\alpha_n)!,$$

$$x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}, \qquad \left(\frac{\partial}{\partial x}\right)^\alpha = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n}.$$

Lemma 3.3 (Taylor's Theorem). *Let f be in $\mathscr{E}_n$. For any nonnegative integer k there exists coefficients $a_\alpha \in \mathscr{E}_n$, indexed by multi-indices α with $|\alpha| = k + 1$, such that*

$$f(x) = j^k f(x) + \sum_{|\alpha|=k+1} a_\alpha(x)x^\alpha. \tag{3.3}$$

PROOF. The proof of Lemma 3.3 is by induction. Lemma 3.1 applied to $f(x) - f(0)$ starts the induction when $k = 0$. We ask the reader to supply the inductive step of the argument in Exercise 3.1. □

(b) Calculation of $RT(g)$ for the Two Examples

For (3.1a) we will show that

$$RT(x^2 + \lambda) = \{f \in \mathscr{E}_{x,\lambda}: f(0, 0) = f_x(0, 0) = 0\}. \tag{3.4}$$

To see this we argue as follows. Observe from Lemma 3.2 that $RT(x^2 + \lambda)$ consists of all germs of the form

$$a(x^2 + \lambda) + (xb + \lambda c)(2x) = (a + 2b)x^2 + (a + 2xc)\lambda, \tag{3.5}$$

where $a, b, c \in \mathscr{E}_{x,\lambda}$. We claim that $RT(x^2 + \lambda)$ may equally be characterized as all germs of the form

$$\alpha x^2 + \beta\lambda, \tag{3.6}$$

where $\alpha, \beta \in \mathscr{E}_{x,\lambda}$. Certainly every germ with the form (3.5) has the form (3.6). Conversely, given $\alpha, \beta \in \mathscr{E}_{x,\lambda}$ we set

$$a = \beta, \qquad b = (\alpha - \beta)/2, \qquad c = 0; \tag{3.7}$$

on substituting these coefficients into (3.5) we see that $\alpha x^2 + \beta\lambda$ can be expressed in the form (3.5), as claimed. We now derive (3.4) from the claim. If $f \in RT(x^2 + \lambda)$ then f may be written in the form (3.6), from which we calculate that $f(0, 0) = f_x(0, 0) = 0$. Conversely, suppose $f(0, 0) = f_x(0, 0) = 0$. Then applying Taylor's theorem (Lemma 3.3) with $k = 1$, we see that there exist coefficients a_{20}, a_{11}, and a_{02} such that

$$f(x, \lambda) = f_\lambda(0, 0)\lambda + a_{20}x^2 + a_{11}\lambda x + a_{02}\lambda^2. \tag{3.8}$$

But (3.8) has the required form (3.6)—we may set

$$\alpha = a_{20}, \qquad \beta = f_\lambda(0, 0) + a_{11}x + a_{02}\lambda.$$

This completes the verification of (3.4).

Passing to the second example, we claim that for (3.1b)

$$RT(x^3 - \lambda x) = \{f \in \mathscr{E}_{x,\lambda}: f(0, 0) = f_x(0, 0) = f_\lambda(0, 0) = f_{xx}(0, 0) = 0\}. \tag{3.9}$$

The calculation is quite similar to the preceding case. First we apply Lemma 3.2 and regroup terms to show that $RT(x^3 - \lambda x)$ consists of all germs of the form

$$(a + 3b)x^3 - (a + 2b - 3xc)\lambda x - c\lambda^2, \tag{3.10}$$

where $a, b, c \in \mathscr{E}_{x,\lambda}$. Then we argue that every germ of the form (3.10) may be written as

$$\alpha x^3 + \beta\lambda x + \gamma\lambda^2, \tag{3.11}$$

where $\alpha, \beta, \gamma \in \mathscr{E}_{x,\lambda}$, and less trivially, every germ of the form (3.11) may be written in the form (3.10); the important point here is that the linear system

$$\begin{aligned} a + 3b &= \alpha, \\ -a - 2b + 3xc &= \beta, \\ -c &= \gamma, \end{aligned}$$

is invertible. Finally we apply Taylor's theorem to show that a germ may be expressed in the form (3.11) if and only if

$$f(0, 0) = f_x(0, 0) = f_\lambda(0, 0) = f_{xx}(0, 0) = 0.$$

This proves the claim.

(c) Afterthoughts

Let us examine the calculations of this section with an eye towards identifying what is essential. We find that there are three basic ideas involved in the calculation. In discussing these ideas we use the term *generator* as follows. Consider an expression such as the right-hand side of (3.5),

$$a(x^2 + \lambda) + b(2x^2) + c(2\lambda x). \tag{3.12}$$

This formula describes a linear combination with arbitrary germs as coefficients of the three functions $x^2 + \lambda$, $2x^2$, and $2\lambda x$. We refer to $x^2 + \lambda$, $2x^2$, and $2\lambda x$ as generators in (3.12). Similarly, we call x^2 and λ generators in (3.6); x^3, λx, and λ^2 generators in (3.11).

The following three steps in the previous calculations are of general applicability.

(i) *Casting out redundant generators*. This was the first step in the calculation for (3.1a). The third generator in (3.5), $2\lambda x$, is a linear combination of the other two—specifically,

$$2\lambda x = 2x(x^2 + \lambda) - \lambda(2x^2). \tag{3.13}$$

Consequently there was no loss in (3.7) when we set the coefficient of the generator $2\lambda x$ equal to zero. By contrast, in the calculation of (3.1b) there were no redundant generators.

(ii) *Forming linear combinations of generators to simplify them.* This idea was used in both examples, to derive (3.6) from (3.5) and to derive (3.11) from (3.10). More specifically, in deriving (3.6) we formed the following two linear combinations of the generators in (3.5):

$$\begin{aligned} x^2 &= \tfrac{1}{2}(2x^2), \\ \lambda &= (x^2 + \lambda) - \tfrac{1}{2}(2x^2). \end{aligned} \tag{3.14}$$

Concerning the derivation of (3.11), in Exercise 3.2 we ask the reader to identify the specific linear combinations of the generators in (3.10) that are involved.

(iii) *Passing from generators of $RT(g)$ to a characterization of $RT(g)$.* By a characterization of $RT(g)$ we mean necessary and sufficient conditions on a function f for f to belong to $RT(g)$. These conditions involve $f(0)$ and of a finite number of derivatives of f at the origin. This step was required in both examples.

In the next section we formalize the first two of these three steps in algebraic language that is natural for the problem. In particular, this leads to a better understanding of when and why these techniques are effective. In the following section, §5, we do likewise for the third step. The first section is a fairly straightforward formalization of the remarks above. The second section is far less obvious; some of the fundamental ideas of singularity theory first appear there.

Exercises

3.1. Complete the proof of Taylor's theorem.

(a) Show by induction that if $j^k g(x) \equiv 0$ then

$$g(x) = \sum_{|\alpha| = k+1} a_\alpha(x) x^\alpha. \tag{3.15}$$

To do this observe that if $j^{k+1} g(x) \equiv 0$ than $a_\alpha(0) = 0$ in (3.15). Then apply Lemma 3.1 to each a_α.

(b) To obtain (3.3) observe that $j^k g(x) \equiv 0$ where $g = f - j^k f$.

3.2. Let $g(x, \lambda) = x^3 - \lambda x$. Find explicitly germs A_j, B_j, C_j ($j = 1, 2, 3$) such that:

$$\begin{aligned} x^3 &= A_1 g + A_2 x g_x + A_3 \lambda g_x, \\ x\lambda &= B_1 g + B_2 x g_x + B_3 \lambda g_x, \\ \lambda^2 &= C_1 g + C_2 x g_x + C_3 \lambda g_x. \end{aligned}$$

3.3. Prove the following version of Taylor's theorem. Let $f(x, y)$ be a smooth, real-valued function defined on a neighborhood of $(0, 0)$ in $\mathbb{R}^m \times \mathbb{R}^n$. Let $Y: \mathbb{R}^m \to \mathbb{R}^n$ be smooth, defined on a neighborhood of 0 in $\mathbb{R}^m$, and satisfy $Y(0) = 0$. Assume that

$f(x, Y(x)) \equiv 0$. Show that there exist smooth, real-valued functions $a_1(x, y), \ldots, a_n(x, y)$ such that on some neighborhood of (0, 0)

$$f(x, y) = \sum_{i=1} a_i(x, y)(y_i - Y_i(x)),$$

where $y = (y_1, \ldots, y_n)$ and $Y(x) = (Y_1(x), \ldots, Y_n(x))$ in coordinates. *Hint*: Adapt the proof of Lemma 3.1 to the above situation by letting $h(s) = f(x, sy + (1 - s)Y(x))$.

§4. Principles for Calculating $RT(g)$, I: Basic Algebra

(a) Terminology from Algebra

In this section we work with $\mathscr{E}_n$, germs of functions of n variables, although for our intended applications we need only consider $\mathscr{E}_2$.

The set $\mathscr{E}_n$ is a vector space, meaning that given any two elements $f, g \in \mathscr{E}_n$ we may form an arbitrary linear combination with scalar coefficients $c_1 f + c_2 g$, where $c_i \in \mathbb{R}$. It is also possible to multiply elements of $\mathscr{E}_n$. The mathematical name for a set admitting these two kinds of operation is a *ring*.

An *ideal* $\mathscr{I}$ in $\mathscr{E}_n$ is a linear subspace with the following special property:

$$\text{If } \phi \in \mathscr{E}_n \text{ and } f \in \mathscr{I}, \text{ then } \phi f \in \mathscr{I}.$$

Concerning our intended application, if $g \in \mathscr{E}_{x,\lambda}$ then $RT(g)$ is an ideal in $\mathscr{E}_{x,\lambda}$. To see this, recall the characterization of $RT(g)$ in Lemma 3.2 as the set of all linear combinations

$$ag + bxg_x + c\lambda g_x, \tag{4.1}$$

where $a, b, c \in \mathscr{E}_{x,\lambda}$. If f has the form (4.1) for some coefficients $a, b, c \in \mathscr{E}_{x,\lambda}$, then for any $\phi \in \mathscr{E}_{x,\lambda}$, ϕf also has this form, with coefficients ϕa, ϕb, and ϕc. Similarly, if f_1 and f_2 both have the form (4.1), so does $f_1 + f_2$.

The characterization (4.1) of $RT(g)$ is a typical construction of ring theory. More generally, if $p_1, \ldots, p_k$ are germs in $\mathscr{E}_n$, then the set of all linear combinations,

$$a_1 p_1 + \cdots + a_k p_k,$$

where $a_i \in \mathscr{E}_n$, is an ideal in $\mathscr{E}_n$. We denote this ideal by $\langle p_1, \ldots, p_k \rangle$, and we call $p_1, \ldots, p_k$ the *generators* of the ideal. (This is consistent with our previous use of the term.) In this notation, we may summarize Lemma 3.2 as

$$RT(g) = \langle g, xg_x, \lambda g_x \rangle. \tag{4.2}$$

An ideal such as $\langle p_1, \ldots, p_k \rangle$ which is generated by a finite number of germs is called *finitely generated*. Although there are ideals in $\mathscr{E}_n$ which are not finitely generated, all of the ideals we consider here will be finitely generated. (See Exercise 5.2 for an example of an ideal which is not finitely generated.)

(b) Principles for Calculating $RT(g)$

In §3(c) we abstracted three principles for calculating $RT(g)$ from the examples of §3(b). The following two lemmas formalize the first two of these, using the algebraic language above. We continue to work with ideals in n dimensions, although (4.2) is the case we have in mind.

Lemma 4.1. *Let* $\mathscr{I} = \langle p_1, \ldots, p_k \rangle$ *be an ideal in* $\mathscr{E}_n$ *with generators* $p_1, \ldots, p_k$. *If* $p_k = a_1 p_1 + \cdots + a_{k-1} p_{k-1}$ *for some germs* $a_i \in \mathscr{E}_n$, *then* $\mathscr{I}$ *is generated by* $p_1, \ldots, p_{k-1}$.

Lemma 4.2. *Let* $\mathscr{I} = \langle p_1, \ldots, p_k \rangle$ *be an ideal in* $\mathscr{E}_n$. *For* $i = 1, \ldots, k$ *let*

$$q_i = \sum_{j=1}^{k} a_{ij} p_j, \tag{4.3}$$

where $a_{ij} \in \mathscr{E}_n$. *If the* $k \times k$ *matrix (of scalars)*

$$\{a_{ij}(0): i, j = 1, \ldots, k\}$$

is invertible, then $\mathscr{I}$ *is also generated by* $q_1, \ldots, q_k$.

Before proving the lemmas, we apply them to rephrase the calculations of §3(b). The argument leading up to (3.6) may be written compactly as

$$RT(x^2 + \lambda) = \langle x^2 + \lambda, 2x^2, 2\lambda x \rangle = \langle x^2 + \lambda, 2x^2 \rangle = \langle x^2, \lambda \rangle. \tag{4.4}$$

The first equality here is the definition (4.2) of $RT(g)$. The second equality follows from Lemma 4.1, since, as noted in (3.13), the third generator is redundant. To derive the last equality in (4.4) from Lemma 4.2 we express one set of generators in terms of the other as follows:

$$\begin{pmatrix} x^2 + \lambda \\ 2x^2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} x^2 \\ \lambda \end{pmatrix}. \tag{4.5}$$

Since the matrix in (4.5) is invertible for $x = \lambda = 0$ (indeed for all (x, λ), since it is constant), either pair of germs generates $RT(x^2 + \lambda)$. (*Remark*: Equation (4.5) is the inverse of (3.14).)

Similarly, the derivation of (3.11) may be written

$$RT(x^3 - \lambda x) = \langle x^3 - \lambda x, 3x^3 - \lambda x, 3x^2\lambda - \lambda^2 \rangle = \langle x^3, \lambda x, \lambda^2 \rangle. \tag{4.6}$$

The first equality is (4.2). To derive the second equality, observe that

$$\begin{pmatrix} x^3 - \lambda x \\ 3x^3 - \lambda x \\ 3x^2\lambda - \lambda^2 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 3 & -1 & 0 \\ 0 & 3x & -1 \end{pmatrix} \begin{pmatrix} x^3 \\ \lambda x \\ \lambda^2 \end{pmatrix}. \tag{4.7}$$

When $x = \lambda = 0$ this matrix equals

$$\begin{pmatrix} 1 & -1 & 0 \\ 3 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

which is nonsingular. The second equality in (4.6) follows from Lemma 4.2.

The proof of Lemma 4.1 is quite simple; we ask the reader to supply this proof in Exercise 4.1.

Proof of Lemma 4.2. We must show that $\langle q_1, \ldots, q_k \rangle = \langle p_1, \ldots, p_k \rangle$. First we claim that $\langle q_1, \ldots, q_k \rangle$ is contained in $\langle p_1, \ldots, p_k \rangle$. Certainly each generator q_i belongs to $\langle p_1, \ldots, p_k \rangle$, since by (4.3) q_i is a linear combination of $p_1, \ldots, p_k$. But any element $f \in \langle q_1, \ldots, q_k \rangle$ is a linear combination of $q_1, \ldots, q_k$; since an ideal is closed under such operations, f belongs to $\langle p_1, \ldots, p_k \rangle$, as claimed.

To obtain the reverse containment we invert (4.3). Let $A(x)$ be the $k \times k$ matrix with entries $\{a_{ij}(x)\}$. We recall Cramer's rule:

$$A^{-1} = \frac{1}{\det A} \operatorname{adj}(A), \tag{4.8}$$

where adj(A) is the classical adjoint of A; i.e., the matrix whose entries are cofactors of A. By hypothesis $\det A(0) \neq 0$, and by continuity $\det A(x)$ is nonzero in some neighborhood of the origin. Thus it follows from (4.8) that the entries of $A^{-1}(x)$, like those of $A(x)$, are smooth germs.

On inverting (4.3) we obtain

$$p_i = \sum_{j=1}^{k} b_{ij} q_j,$$

where $b_{ij} \in \mathscr{E}_n$ is the i, j entry of $A^{-1}(x)$. Thus each generator p_i is a linear combination of $q_1, \ldots, q_k$. Reversing the above argument we may deduce that $\langle p_1, \ldots, p_k \rangle$ is contained in $\langle q_1, \ldots, q_k \rangle$, which proves the lemma. □

Exercise

4.1. Prove Lemma 4.1. *Hint*: Show that every linear combination of $p_1, \ldots, p_k$ may be written as a linear combination of $p_1, \ldots, p_{k-1}$.

§5. Principles for Calculating $RT(g)$, II: Finite Determinacy

In §3(c) we isolated three ideas used in the calculation of $RT(g)$ for the examples of that section. The third of these involved passing from generators of $RT(g)$ to a characterization of $RT(g)$ by conditions on a function and its derivatives at the origin. In this section we explore this third step more fully. This exploration leads us into some fundamental ideas of the singularity theory approach to bifurcation.

(a) More Terminology from Algebra

Let

$$\mathcal{M} = \{f \in \mathcal{E}_n : f(0) = 0\}. \tag{5.1}$$

(We do not put a subscript on $\mathcal{M}$ since usually the context indicates the number of independent variables.) We ask the reader to show that $\mathcal{M}$ is an ideal in Exercise 5.3. We claim that $\mathcal{M}$ is generated by $x_1, \ldots, x_n$; in symbols

$$\mathcal{M} = \langle x_1, \ldots, x_n \rangle.$$

Certainly each generator x_i belongs to $\mathcal{M}$. But by Lemma 3.1, if $f \in \mathcal{E}_n$ satisfies $f(0) = 0$, then

$$f(x) = x_1 a_1(x) + \cdots + x_n a_n(x)$$

for some $a_i \in \mathcal{E}_n$. Thus $f \in \langle x_1, \ldots, x_n \rangle$, which proves the claim. (*Remark*: We use the letter $\mathcal{M}$ because $\mathcal{M}$ is a maximal ideal, in the following sense: For any ideal $\mathcal{I} \subset \mathcal{E}_n$, either $\mathcal{I} \subset \mathcal{M}$ or $\mathcal{I} = \mathcal{E}_n$. See Exercise 5.4 for the proof of this fact.)

Given two ideals $\mathcal{J}$ and $\mathcal{I}$ in $\mathcal{E}_n$, there are standard constructions in ring theory which leads to new ideals, the sum ideal $\mathcal{I} + \mathcal{J}$ and the product ideal $\mathcal{I} \cdot \mathcal{J}$. The *sum* ideal $\mathcal{I} + \mathcal{J}$ consists of all germs of the form $f + g$ where $f \in \mathcal{I}$ and $g \in \mathcal{J}$; the *product* ideal consists of all finite sums of the form

$$f_1 g_1 + \cdots + f_m g_m,$$

where $f_i \in \mathcal{I}$ and $g_i \in \mathcal{J}$. If $\mathcal{I}$ is generated by $p_1, \ldots, p_k$ and $\mathcal{J}$ is generated by $q_1, \ldots, q_l$, then $\mathcal{I} + \mathcal{J}$ is generated by the $k + l$ germs

$$p_1, \ldots, p_k, q_1, \ldots, q_l, \tag{5.2a}$$

and $\mathcal{I} \cdot \mathcal{J}$ is generated by the $k \cdot l$ germs

$$\{p_i q_j : i = 1, \ldots, k; j = 1, \ldots, l\}. \tag{5.2b}$$

We ask the reader to verify these statements in Exercise 5.6.

Remark. These generators of $\mathscr{I} + \mathscr{J}$ and $\mathscr{I} \cdot \mathscr{J}$ may be redundant, even though there are no redundancies in either $\{p_i\}$ or in $\{q_j\}$ by themselves.

We give two examples of the product construction that are important in the discussion below. If $g \in \mathscr{E}_{x,\lambda}$, then

$$\mathscr{M} \cdot RT(g) = \langle xg, \lambda g, x^2 g_x, \lambda x g_x, \lambda^2 g_x \rangle.$$

These generators may be derived from (5.2b) on observing that $\mathscr{M} = \langle x, y \rangle$ and $RT(g) = \langle g, xg_x, \lambda g_x \rangle$. Note that the generator $\lambda x g_x$ occurs twice in the enumeration, as $\lambda(xg_x)$ and $x(\lambda g_x)$.

For the second example we return to n dimensions. Proceeding inductively we define a sequence of ideals

$$\mathscr{M}^2 = \mathscr{M} \cdot \mathscr{M}, \qquad \mathscr{M}^3 = \mathscr{M} \cdot \mathscr{M}^2, \qquad \mathscr{M}^4 = \mathscr{M} \cdot \mathscr{M}^3, \ldots .$$

It follows from (5.2b) that $\mathscr{M}^k$ is generated by all monomials of degree k,

$$\{x^\alpha : |\alpha| = k\}.$$

In Exercise 5.5 we ask the reader to derive the following, alternative characterization of $\mathscr{M}^k$:

$$\mathscr{M}^k = \left\{ f \in \mathscr{E}_n : \left(\frac{\partial}{\partial x}\right)^\alpha f(0) = 0 \text{ for } |\alpha| \le k - 1 \right\}. \tag{5.3}$$

Let $\mathscr{I} \subset \mathscr{E}_n$ be an ideal. We shall say that two germs $f, g \in \mathscr{E}_n$ are *equal modulo* $\mathscr{I}$, in symbols

$$f \equiv g \pmod{\mathscr{I}},$$

if $f - g \in \mathscr{I}$. For example, using this terminology Taylor's theorem (Lemma 3.3) may be rephrased

$$f \equiv j^k f \pmod{\mathscr{M}^{k+1}}. \tag{5.4}$$

The algebraic operations in $\mathscr{E}_n$ (i.e., addition, multiplication, etc.) preserve this notion of equality. Thus if $f_1 \equiv f_2 \pmod{\mathscr{I}}$ and $g_1 \equiv g_2 \pmod{\mathscr{I}}$, then $f_1 + g_1 \equiv f_2 + g_2 \pmod{\mathscr{I}}$ and $f_1 g_1 \equiv f_2 g_2 \pmod{\mathscr{I}}$.

(b) Conditions for Membership in $RT(g)$

Let $g \in \mathscr{E}_{x,\lambda}$. Our goal in this subsection is to show that if some power of the maximal ideal is contained in $RT(g)$, then deciding whether a given germ $f \in \mathscr{E}_{x,\lambda}$ belongs to $RT(g)$ reduces to a problem in finite-dimensional linear algebra. This conclusion follows simply from the lemma below. The purpose of this lemma is theoretical; i.e., to derive the above conclusion. We do not recommend the lemma for serious calculations, as we will introduce more efficient techniques below. In the lemma we adapt the multi-index notation from n dimensions by defining

$$D^\alpha = \left(\frac{\partial}{\partial x}\right)^{\alpha_1}\left(\frac{\partial}{\partial \lambda}\right)^{\alpha_2}.$$

Lemma 5.1. *Let $g \in \mathscr{E}_{x,\lambda}$, and suppose that for some $k \geq 0$, $\mathscr{M}^{k+1} \subset RT(g)$. A germ $f \in \mathscr{E}_{x,\lambda}$ belongs to $RT(g)$ if and only if there exist polynomials $a(x, \lambda)$, $b(x, \lambda)$, $c(x, \lambda)$ of degree k or less satisfying the following system of equations:*

$$D^\alpha[f - (ag + bxg_x + c\lambda g_\lambda)](0, 0) = 0 \quad \text{for} \quad |\alpha| \leq k. \tag{5.5}$$

Proof. If $f \in RT(g)$, there exist *germs* $\tilde{a}, \tilde{b}, \tilde{c} \in \mathscr{E}_{x,\lambda}$ such that

$$f - (\tilde{a}g + \tilde{b}xg_x + \tilde{c}\lambda g_x) = 0 \quad \text{for all } x, \lambda. \tag{5.6}$$

Let a, b, c be the kth-order Taylor polynomials of $\tilde{a}, \tilde{b}, \tilde{c}$, respectively. Then

$$D^\alpha a(0, 0) = D^\alpha \tilde{a}(0, 0) \quad \text{for} \quad |\alpha| \leq k,$$

and similar formulas hold for b and c. Equation (5.5) results from combining this observation with appropriate derivatives of (5.6).

Conversely, suppose $f \in \mathscr{E}_{x,\lambda}$ satisfies (5.5) for some polynomials a, b, c. Let

$$r = f - (ag + bxg_x + c\lambda g_x).$$

We see from (5.3) that $r \in \mathscr{M}^{k+1} \subset RT(g)$. Now the formula

$$f = (ag + bxg_x + c\lambda g_x) + r$$

displays f as the sum of two terms in $RT(g)$; thus $f \in RT(g)$. □

Lemma 5.1 reduces the question of membership in $RT(g)$ to the solvability of (5.5). We interpret (5.5) as a system of $(k + 1)(k + 2)/2$ *linear* equations for unknown coefficients in the polynomials

$$a(x, \lambda) = \sum_{|\alpha| \leq k} a_\alpha x^{\alpha_1}\lambda^{\alpha_2}, \quad \text{etc.} \tag{5.7}$$

There are

$$3k(k + 1)/2 \tag{5.8}$$

such unknown coefficients. Because of the terms $D^\alpha f(0, 0)$ in (5.5), this system is inhomogeneous. It turns out that (5.5) cannot be solved unless the inhomogeneity $D^\alpha f(0, 0)$ satisfies auxiliary conditions; this is of course a familiar situation in linear algebra.

Let us illustrate these ideas by relating them to the calculation of $RT(x^3 - \lambda x)$ in §3(b). First we claim that

$$\mathscr{M}^3 \subset RT(x^3 - \lambda x). \tag{5.9}$$

By (3.11), $RT(x^3 - \lambda x) = \langle x^3, \lambda x, \lambda^2 \rangle$. For each generator of $\mathscr{M}^3$ we have a representation

$$x^3 = 1 \cdot x^3, \qquad \lambda x^2 = x \cdot \lambda x,$$
$$\lambda^2 x = \lambda \cdot \lambda x, \qquad \lambda^3 = \lambda \cdot \lambda^2.$$

This shows that all generators of $\mathscr{M}^3$ belong to $RT(x^3 - \lambda x)$, which proves the claim.

Now we unscramble (5.5) when $g = x^3 - \lambda x$, taking $k = 2$. Observe that if $|\alpha| \leq 2$

$$D^\alpha[a(x, \lambda)(x^3 - \lambda x)](0, 0) = \begin{cases} -a(0, 0) & \text{if } \alpha = (1, 1), \\ 0 & \text{otherwise.} \end{cases}$$

In other words, only the first coefficient a_0 of $a(x, \lambda)$ in (5.7) actually contributes to (5.5), because $x^3 - \lambda x$ already vanishes to fairly high order. Similarly, for the other two unknown polynomials, only b_0 and c_0 contribute to (5.5). Equation (5.5) written out in components becomes the following inhomogeneous system of six equations in the three unknowns a_0, b_0, c_0:

$$\begin{aligned} 0 &= f(0, 0), \\ 0 &= f_x(0, 0), \\ 0 &= f_\lambda(0, 0), \\ 0 &= f_{xx}(0, 0), \\ -a_0 - b_0 &= f_{\lambda x}(0, 0), \\ -c_0 &= f_{\lambda\lambda}(0, 0). \end{aligned}$$

Clearly this system is solvable if and only if

$$f(0, 0) = f_x(0, 0) = f_\lambda(0, 0) = f_{xx}(0, 0) = 0,$$

which recovers our earlier result.

The example illustrates that, in general, formula (5.8) is a gross overestimate for the number of unknowns in (5.5). Typically g vanishes to fairly high order, and only the low-order coefficients of a, b, and c contribute to (5.5). Moreover, there is substantial overlap between contributions of bxg_x and $c\lambda g_x$ which further reduces the effective number of independent variables.

The following fact is of the utmost importance: The solvability condition for (5.5) only involves a finite number of the derivatives of f at the origin. In other words, if the derivatives of f of order k or less are such that (5.5) is solvable, then $f \in RT(g)$ no matter what the higher-order derivatives of f may be. Thus Lemma 5.1 begins to address the fundamental issue of finite determinacy.

(c) Nakayama's Lemma

Lemma 5.1 above indicates the importance of being able to ascertain whether some power of $\mathcal{M}$ is contained in $RT(g)$. In this connection, a result called "Nakayama's Lemma" is most useful—given an integer k, this result provides a simple test for whether or not $\mathcal{M}^k \subset RT(g)$. In this subsection we first state without proof Nakayama's lemma in the most relevant special case; then we use the special case of the lemma to show $\mathcal{M}^k \subset RT(g)$ in two specific examples; next we state and prove the general case of the lemma; and finally we mention a corollary of the lemma.

Lemma 5.2. *Let $g \in \mathscr{E}_{x,\lambda}$, and let k be a positive integer. If*

$$\mathcal{M}^k \subset RT(g) + \mathcal{M}^{k+1},$$

then $\mathcal{M}^k \subset RT(g)$.

As our first application, let us derive (5.9) using this lemma. By definition

$$RT(x^3 - \lambda x) = \langle x^3 - \lambda x, 3x^3 - \lambda x, 3\lambda x^2 - \lambda^2 \rangle. \tag{5.10}$$

We may express each of the generators of $\mathcal{M}^3$ in terms of the three generators in (5.10), modulo errors in $\mathcal{M}^4$, as follows:

$$\begin{aligned} x^3 &= -\tfrac{1}{2}(x^3 - \lambda x) + \tfrac{1}{2}(3x^3 - \lambda x), \\ \lambda x^2 &= -x(x^3 - \lambda x) + r_1(x, \lambda), \\ \lambda^2 x &= -\lambda(x^3 - \lambda x) + r_2(x, \lambda), \\ \lambda^3 &= -\lambda(3\lambda x^2 - \lambda^2) + r_3(x, \lambda), \end{aligned} \tag{5.11}$$

where $r_i \in \mathcal{M}^4$. We see from (5.11) that

$$\mathcal{M}^3 \subset RT(x^3 - \lambda x) + \mathcal{M}^4,$$

so (5.9) follows from Lemma 5.2.

In general Lemma 5.2 simplifies the treatment of higher-order terms in a bifurcation problem. To illustrate this, we consider the following perturbation of the pitchfork:

$$g(x, \lambda) = x^3 - \lambda x + ax^4,$$

where $a \in \mathbb{R}$. (Cf. §1.) Specifically we will show that

$$\mathcal{M}^3 \subset RT(g). \tag{5.12}$$

Indeed, let us just repeat (5.11) for the present example. Since $ax^4 \in \mathscr{M}^4$, each equation in (5.11) is merely perturbed by some element of $\mathscr{M}^4$; in symbols

$$\begin{aligned} x^3 &= -\tfrac{1}{2}g + \tfrac{1}{2}xg_\lambda + \tilde{r}_0(x, \lambda), \\ \lambda x^2 &= -xg + \tilde{r}_1(x, \lambda), \\ \lambda^2 x &= -\lambda g + \tilde{r}_2(x, \lambda), \\ \lambda^3 &= -\lambda(\lambda g_x) + \tilde{r}_3(x, \lambda), \end{aligned}$$

where $\tilde{r}_i \in \mathscr{M}^4$. Thus (5.12) follows from Lemma 5.2.

Remarks. (i) It is quite possible to derive (5.12) directly using the result in Exercise 3.2.

(ii) The proof of Nakayama's lemma has much in common with the proof of Lemma 4.2. The advantage of Nakayama's lemma is that it is not necessary to be so explicit in specifying the relationship between two sets of generators.

Lemma 5.2 is the special case of the following lemma which results from taking $\mathscr{I} = \mathscr{M}^k$, $\mathscr{J} = RT(g)$. Thus we only prove Lemma 5.3.

Lemma 5.3 (Nakayama's Lemma). *Let $\mathscr{I}$ and $\mathscr{J}$ be ideals in $\mathscr{E}_n$, and assume that $\mathscr{I} = \langle p_1, \ldots, p_l \rangle$ is finitely generated. Then $\mathscr{I} \subset \mathscr{J}$ if and only if $\mathscr{I} \subset \mathscr{J} + \mathscr{M} \cdot \mathscr{I}$.*

Proof. Since $\mathscr{J}$ is contained in $\mathscr{J} + \mathscr{M} \cdot \mathscr{I}$, the "only if" part is a triviality. Conversely, let us assume that $\mathscr{I} \subset \mathscr{J} + \mathscr{M} \cdot \mathscr{I}$. This implies that each generator p_i of $\mathscr{I}$ can be written

$$p_i = f_i + \sum_{j=1}^{l} a_{ij} p_j, \tag{5.13}$$

where $f_i \in \mathscr{J}$ and $a_{ij} \in \mathscr{E}_n$ satisfy $a_{ij}(0) = 0$. The form of the second term in (5.13) follows from the fact that $\mathscr{M} \cdot \mathscr{I}$ is generated by products of p_j with germs vanishing at the origin. Let A be the $l \times l$ matrix with entries a_{ij}; let I be the $l \times l$ identity. We rewrite (5.13) in matrix notation

$$(I - A)\begin{pmatrix} p_1 \\ \vdots \\ p_l \end{pmatrix} = \begin{pmatrix} f_1 \\ \vdots \\ f_l \end{pmatrix}. \tag{5.14}$$

Now $(I - A)(0) = I$, so $I - A$ is invertible in some neighborhood of the origin. Inverting $(I - A)$ in (5.14) and writing out components we find that

$$p_i = \sum_{j=1}^{l} b_{ij} f_j,$$

where $\{b_{ij}\}$ are the entries of $(I - A)^{-1}$. This shows that each p_i is a linear combination of the f_j's and hence belongs to $\mathscr{J}$. Therefore $\mathscr{I} \subset \mathscr{J}$. □

There are two useful consequences of Nakayama's lemma which we mention at this time.

Corollary 5.4. (a) *Let* $\mathscr{I} = \langle p_1, \ldots, p_l \rangle$ *be an ideal in* $\mathscr{E}_n$ *and suppose that* $q_1, \ldots, q_l$ *are in* $\mathscr{M} \cdot \mathscr{I}$. *Then* $\mathscr{I}$ *is also generated by* $p_1 + q_1, \ldots, p_l + q_l$.

(b) *If* g *is a germ such that* $\mathscr{M}^k \subset RT(g)$, *then* g *is strongly equivalent to its Taylor polynomial* $j^k g$.

Proof. (a) Since each $q_j \in \mathscr{M} \cdot \mathscr{I} \subset \mathscr{I}$ it follows that $p_j + q_j \in \mathscr{I}$ which in turn implies that $\langle p_1 + q_1, \ldots, p_l + q_l \rangle \subset \mathscr{I}$. Conversely, each $p_j = (p_j + q_j) - q_j \in \langle p_1 + q_1, \ldots, p_l + q_l \rangle + \mathscr{M} \cdot \mathscr{I}$, implying that

$$\mathscr{I} \subset \langle p_1 + q_1, \ldots, p_l + q_l \rangle + \mathscr{M} \cdot \mathscr{I}.$$

It follows from Nakayama's lemma that $\mathscr{I} \subset \langle p_1 + q_1, \ldots, p_l + q_l \rangle$ proving part (a).

(b) Let us write $g = j^k g - r$, where $r \in \mathscr{M}^{k+1}$. (Note the minus sign.) According to Theorem 2.2, to prove that g is strongly equivalent to $j^k g$, it suffices to show that

$$RT(g) = RT(g + tr)$$

for all real numbers t satisfying $0 \leq t \leq 1$. Now $RT(g + tr)$ is generated by $g + tr$, $x(g_x + tr_x)$, and $\lambda(g_x + tr_x)$. Each of these generators differs from the corresponding generators of $RT(g)$ by an element of $\mathscr{M}^{k+1} \subset \mathscr{M} \cdot RT(g)$. Thus by part (a) of the corollary, $RT(g) = RT(g + tr)$. □

Part (b) of the corollary plays a fundamental role in issues of finite determinacy. We shall expand greatly on this corollary in §8(b).

(d) Finite Codimension for Ideals

In this subsection we explore the concept of finite codimension. As shown by Proposition 5.7 below, this concept is intimately related to the question of whether an ideal contains some power of the maximal ideal $\mathscr{M}$. We use this concept extensively in the rest of this chapter; the proofs of this subsection, however, are less germane, and they may be skipped on a first reading.

Definition 5.5. Let $\mathscr{I} \subset \mathscr{E}_n$ be a vector subspace. If there exists a finite-dimensional vector subspace $V \subset \mathscr{E}_n$ such that

$$\mathscr{I} + V = \mathscr{E}_n, \tag{5.15}$$

we say that $\mathscr{I}$ has *finite codimension*. If no such subspace exists, we say that $\mathscr{I}$ has *infinite codimension*.

As the notation suggests, we will usually apply this concept to a vector subspace $\mathscr{I}$ which is in fact an ideal, at least in the present chapter.

We consider some examples to clarify this definition. Let us show that for any positive integer k, $\mathscr{M}^k$ has finite codimension. If $f \in \mathscr{E}_n$, we may write

$$f = [f - j^{k-1}f] + j^{k-1}f.$$

By (5.3) the first term here belongs to $\mathscr{M}^k$, and the second term is a polynomial of degree at most $k - 1$. Therefore

$$\mathscr{E}_n = \mathscr{M}^k + \mathbb{R}\{x^\alpha : |\alpha| \le k - 1\}, \tag{5.16}$$

where the second summand in (5.15) indicates all linear combinations of monomials of degree $k - 1$ or less. It follows from (5.16) that $\mathscr{M}^k$ has finite codimension for any k, as claimed.

As a second example, we claim that $RT(x^3 - \lambda x)$ has finite codimension in $\mathscr{E}_{x,\lambda}$. Recall the characterization (3.9) of $RT(x^3 - \lambda x)$. For any $f \in \mathscr{E}_{x,\lambda}$ there is a (unique) polynomial of the form

$$\pi(x) = c_1 + c_2 x + c_3 \lambda + c_4 x^2,$$

such that $f - \pi \in RT(x^3 - \lambda x)$. Thus

$$\mathscr{E}_{x,\lambda} = RT(x^3 - \lambda x) + \mathbb{R}\{1, x, \lambda, x^2\}, \tag{5.17}$$

which proves the claim.

Given a vector subspace $\mathscr{I} \subset \mathscr{E}_n$ of finite codimension, there are many choices for a complementary subspace V (i.e. a subspace satisfying (5.15)). For example, a possible modification of (5.17) for $RT(x^3 - \lambda x)$ is indicated in Exercise 5.8. It is natural to require, however, that V have as small a dimension as possible. This occurs if and only if

$$\mathscr{I} \cap V = \{0\},$$

in which case we say that $\mathscr{E}_n$ is the direct sum of $\mathscr{I}$ and V, written

$$\mathscr{E}_n = \mathscr{I} \oplus V. \tag{5.18}$$

The decompositions (5.16) and (5.17) are direct sum decompositions. (See Exercise 5.9 for the proof.) These ideas lead to the following refinement of Definition 5.5.

Definition 5.6. Let $\mathscr{I} \subset \mathscr{E}_n$ be a vector subspace of finite codimension. The codimension of $\mathscr{I}$ is the dimension of any subspace $V \subset \mathscr{E}_n$ which satisfies (5.18); in symbols

$$\operatorname{codim} \mathscr{I} = \dim V.$$

Proposition 5.7. *Let $\mathscr{I} \subset \mathscr{E}_n$ be an ideal. There is an integer k such that $\mathscr{M}^k \subset \mathscr{I}$ if and only if $\mathscr{I}$ has finite codimension.*

We prove this proposition below; first we state and apply a corollary.

Corollary 5.8. *If an ideal* $\mathscr{I} \subset \mathscr{E}_n$, *where* $n \geq 2$, *is generated by only one germ* p *such that* $p(0) = 0$, *then* $\mathscr{I}$ *has infinite codimension.*

This corollary is proved below. Before that, let us use the corollary to construct two examples of a germ g such that $RT(g)$ has infinite codimension.

Example 5.9a. We consider any germ of the form $g(x, \lambda) = \lambda q(x, \lambda)$, where $q \in \mathscr{E}_n$. Now $RT(g) = \langle g, xg_x, \lambda g_x \rangle$, and it is easily seen that λ is a factor of all three of these generators. Thus, $RT(g) \subset \langle \lambda \rangle$, and it follows from Corollary 5.8 that $RT(g)$ has infinite codimension.

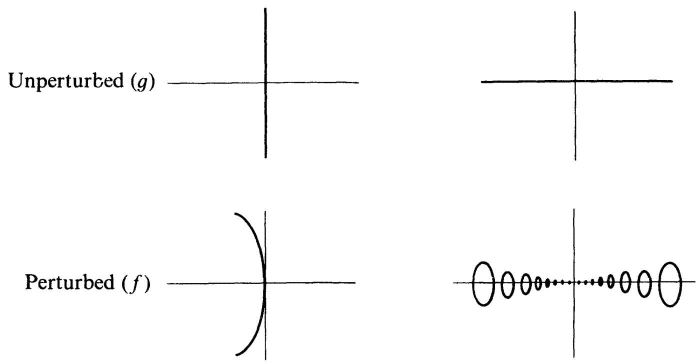
Incidentally, germs of the form $\lambda q(x, \lambda)$ do not have the finite determinancy property contained in Lemma 5.1. Indeed, the germs $g(x, \lambda) = \lambda$ and

$$f(x, \lambda) = e^{-1/x^2} + \lambda$$

have equal derivatives of all orders at the origin, but yet g and f are not equivalent. To verify the latter statement, compare the zero sets of the two germs as shown in Figure 5.1(a). Note that the number of solutions x as a function of λ is different for f and g.

Example 5.9b. We consider any germ of the form $x^2 q(x, \lambda)$. In this case x is a factor of all three generators of $RT(g)$, so $RT(g) \subset \langle x \rangle$. Thus $RT(g)$ has infinite codimension. Notice also that the germs $g(x, \lambda) = x^2$ and $f(x, \lambda) = x^2 + e^{-1/\lambda^2} \sin(1/\lambda)$ have the same Taylor expansions, but the zero sets of g and f are very different indeed. See Figure 5.1(b).

Examples 5.9 show some of the difficulties which arise when $RT(g)$ has infinite codimension. Pathological behavior is the rule rather than the exception.



(a) $g(x, \lambda) = \lambda$, $f(x, \lambda) = \lambda + e^{-1/x^2}$

(b) $g(x, \lambda) = x^2$, $f(x, \lambda) = x^2 + e^{-1/\lambda^2} \sin(1/\lambda^2)$

Figure 5.1. Perturbations of bifurcation problems of infinite codimension.

Proof of proposition 5.7. If $\mathscr{M}^k \subset \mathscr{I}$, then

$$\operatorname{codim} \mathscr{I} \leq \operatorname{codim} \mathscr{M}^k < \infty.$$

Conversely, assume that $\mathscr{I}$ has finite codimension; say $\operatorname{codim} \mathscr{I} = l - 1$. Trivially, we have the inclusions

$$\mathscr{I} \subset \mathscr{I} + \mathscr{M}^l \subset \mathscr{I} + \mathscr{M}^{l-1} \subset \cdots \subset \mathscr{I} + \mathscr{M}^2 \subset \mathscr{I} + \mathscr{M}.$$

Hence, reading back to front we have

$$1 = \operatorname{codim}(\mathscr{I} + \mathscr{M}) \leq \operatorname{codim}(\mathscr{I} + \mathscr{M}^2) \leq \cdots \leq \operatorname{codim}(\mathscr{I} + \mathscr{M}^l) \leq \operatorname{codim} \mathscr{I} = l - 1.$$

Equality must hold in at least one case here, because there are more inequalities than available integers. Thus, there exists an integer $k \leq l$ for which

$$\operatorname{codim}(\mathscr{I} + \mathscr{M}^k) = \operatorname{codim}(\mathscr{I} + \mathscr{M}^{k+1}).$$

However, $\mathscr{I} + \mathscr{M}^{k+1} \subset \mathscr{I} + \mathscr{M}^k$. These two ideals can have equal codimension only if they are equal; in symbols

$$\mathscr{I} + \mathscr{M}^k = \mathscr{I} + \mathscr{M}^{k+1}.$$

In particular, $\mathscr{M}^k \subset \mathscr{I} + \mathscr{M}^{k+1}$. It follows from Nakayama's lemma (Lemma 5.3) that $\mathscr{M}^k \subset \mathscr{I}$. The proof is complete. □

Proof of corollary 5.8. We assume that $\mathscr{I}$ has finite codimension and derive a contradiction. By Proposition 5.7 there is a k such that $\mathscr{M}^k \subset \mathscr{I} \subset \langle p \rangle$. It follows from Corollary 5.4 that p and its Taylor polynomial $j^k p$ generate the same ideal since $p = j^k p + r$ where $r \in \mathscr{M}^{k+1} \subset \mathscr{M}\langle p \rangle$. Thus in proving Corollary 5.8 we may assume without loss of generality that p is a polynomial of degree at most k—if it is not, we replace p by $j^k p$.

Since p is a polynomial, we may extend p to a function on $\mathbb{C}^n$. Moreover $p(0) = 0$. Now in $\mathbb{C}^n$, $n \geq 2$, the zero set of a polynomial is never an isolated point; necessarily in any neighborhood of the origin there are infinitely many points where p vanishes.

On the other hand, since $\mathscr{M}^k \subset \langle p \rangle$, any monomial q of degree k may be factored $q = ap$ for some $a \in \mathscr{E}_n$. Indeed, a must be a polynomial of degree k or less. The factorization $q = ap$ still holds over the complex numbers. Thus, any such monomial q vanishes on $\{p = 0\}$. However, this contradicts our remarks above, since the set of simultaneous zeros in $\mathbb{C}^n$ of all monomials of degree k is just the origin, an isolated point. □

We will need a technical result about finite codimension in §13 which we present here.

Proposition 5.10. *Let $f(x, \lambda)$, $g(x, \lambda)$ be in $\mathscr{E}_{x,\lambda}$ and assume that the ideal $\langle f, g \rangle$ has finite codimension. Suppose there exist α, β in $\mathscr{E}_{x,\lambda}$ such that*

$$\alpha(x, \lambda) f(x, \lambda) + \beta(x, \lambda) g(x, \lambda) = 0 \quad \text{for all } x, \lambda.$$

Then for each k there exists a germ $Q(x, \lambda)$ in $\mathscr{E}_{x,\lambda}$ such that

$$\alpha(x, \lambda) \equiv -Q(x, \lambda)g(x, \lambda) \quad \textit{and} \quad \beta(x, \lambda) \equiv +Q(x, \lambda)f(x, \lambda) \mod \mathscr{M}^{k+1}.$$

We shall not give a proof of this proposition in its full generality; this proof depends on ideas from algebraic geometry. See Zariski and Samuel [1960], p. 293. However, we do sketch the proof in the special case $f(x, \lambda) = x^2$.

Suppose

$$\alpha(x, \lambda)x^2 + \beta(x, \lambda)g(x, \lambda) = 0 \quad \text{for all } x, \lambda, \tag{5.19}$$

and suppose that the ideal $\langle x^2, g(x, \lambda)\rangle$ has finite codimension. As observed in the proof of Corollary 5.8, near the origin, the origin itself is the only common zero of x^2 and $g(x, \lambda)$. Evaluating equation (5.19) at $x = 0$ yields

$$\beta(0, \lambda)g(0, \lambda) = 0 \quad \text{for all } \lambda.$$

Since $g(0, \lambda) \neq 0$ if $\lambda \neq 0$ it follows by continuity that $\beta(0, \lambda) = 0$ for all λ. Now apply Taylor's theorem to see that

$$\beta(x, \lambda) = x\gamma(x, \lambda).$$

Substitution into (5.19), division by x, and appeals to continuity yield

$$\alpha(x, \lambda)x + \gamma(x, \lambda)g(x, \lambda) \equiv 0. \tag{5.20}$$

Iterating the above argument, we see that

$$\gamma(x, \lambda) = xQ(x, \lambda) \quad \text{and} \quad \beta(x, \lambda) = x^2Q(x, \lambda).$$

Substitution into (5.20) and division by x yield

$$\alpha(x, \lambda) = -Q(x, \lambda)g(x, \lambda)$$

as desired. □

Exercises

5.1. Let $\mathscr{I}$ be a finitely generated ideal. Use Nakayama's lemma to show that if $\mathscr{I} = \mathscr{M} \cdot \mathscr{I}$ then $\mathscr{I} = \{0\}$.

5.2. Let $\mathscr{M}^\infty = \bigcap_{k=1}^\infty \mathscr{M}^k$.

(a) Show that $\mathscr{M}^\infty$ is an ideal and that $\mathscr{M}^\infty \neq \{0\}$. ($\mathscr{M}^\infty$ is called the ideal of *flat* germs.)
(b) Show that $\mathscr{M}^\infty = \mathscr{M} \cdot \mathscr{M}^\infty$ and conclude using Exercise 5.1 that $\mathscr{M}^\infty$ is *not* finitely generated.

5.3. Using only the definition of $\mathscr{M}$ in (5.1), show that $\mathscr{M}$ is an ideal. (The characterization of $\mathscr{M}$ as $\langle x_1, \ldots, x_n\rangle$ using Lemma 3.1 also shows that $\mathscr{M}$ is a finitely generated ideal.)

5.4. Prove that $\mathscr{M}$ is the unique maximal ideal in $\mathscr{E}_n$. *Hint*: Suppose $\mathscr{I}$ is an ideal in $\mathscr{E}_n$ which contains a germ g such that $g(0) \neq 0$; use the fact that $1/g \in \mathscr{E}_n$ to show that $1 = g \cdot 1/g \in \mathscr{I}$.

5.5. Use Taylor's theorem to verify (5.3).

5.6. Show that (5.2a, b) give generators for $\mathscr{I} + \mathscr{J}$ and $\mathscr{I} \cdot \mathscr{J}$, respectively.

5.7. Show that every ideal of finite codimension is finitely generated. Give an example of a finitely generated ideal of infinite codimension.

5.8. Show that $\mathscr{E}_{x,\lambda} = \{e^{x\lambda}, \sin x, \lambda, 1 - \cos x\} + RT(x^3 - \lambda x)$.

5.9. Show that (5.16) and (5.17) are direct sums.

5.10. Use Exercise 3.5—with $Y(x) \equiv 0$—to prove Taylor's theorem with parameters; that is, if $f\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is smooth, then

$$f(x, y) \equiv \sum_{|\alpha| \leq k} a_\alpha(x) y^\alpha \quad (\text{mod } \mathscr{M}_y^{k+1}),$$

where $\mathscr{M}_y$ is the maximal ideal $\langle y_1, \ldots, y_m \rangle$ in the y-variables, $\alpha = (\alpha_1, \ldots, \alpha_m)$ is a multi-index and each a_α is smooth. In fact one has the formula

$$a_\alpha(x) = \frac{1}{\alpha!} \frac{\partial^{|\alpha|}}{\partial y^\alpha} f(x, 0).$$

§6. Calculation of $RT(g)$, II: A Hard Example

In this section we calculate $RT(g)$ when

$$g(x, \lambda) = x^5 + \lambda x^3 + \lambda^2. \tag{6.1}$$

This is a somewhat academic example that we chose for its pedagogical value rather than for any intended application. As in §3 we shall characterize $RT(x^5 + \lambda x^3 + \lambda^2)$ both by generators and by conditions for membership on a function.

Specifically, our characterization of $RT(g)$ by generators is

$$RT(x^5 + \lambda x^3 + \lambda^2) = \mathscr{M}^6 + \mathscr{M}^4\langle\lambda\rangle + \mathscr{M}\langle\lambda^2\rangle + \mathbb{R}\{x^5 + \lambda x^3 + \lambda^2, 5x^5 + 3\lambda x^3\}. \tag{6.2}$$

Let us compare (6.2) with (3.11), the corresponding formula for $x^3 - \lambda x$. To facilitate this comparison we rewrite (3.11) in our present notation; we claim that

$$RT(x^3 - \lambda x) = \mathscr{M}^3 + \mathscr{M}\langle\lambda\rangle. \tag{6.3}$$

We see from (3.11) that $RT(x^3 - \lambda x)$ is generated by the three monomials $x^3, \lambda x, \lambda^2$. Now $\mathscr{M}^3 + \mathscr{M}\langle\lambda\rangle$ has generators x^3, λx^2, $\lambda^2 x$, λ^3, $x\lambda$, and λ^2. We use Lemma 4.1 to discard the redundant generators λx^2, $\lambda^2 x$, λ^3 from the latter list, thereby proving (6.3). Part of the difficulty of the example (6.1) stems from the fact that (6.2) is not generated by monomials.

As our first step in deriving (6.2) we use Lemma 5.2 (the special case of Nakayama's lemma) to show that

$$\mathcal{M}^6 \subset RT(x^5 + \lambda x^3 + \lambda^2). \tag{6.4}$$

Now

$$\begin{aligned} RT(g) &= \langle g, xg_x, \lambda g_x \rangle \\ &= \langle x^5 + \lambda x^3 + \lambda^2, 5x^5 + 3\lambda x^3, 5\lambda x^4 + 3\lambda^2 x^2 \rangle \end{aligned} \tag{6.5}$$

We must show that each generator of $\mathcal{M}^6$ (i.e., $x^6, \lambda x^5, \ldots, \lambda^6$) belongs to $RT(g) + \mathcal{M}^7$. This is trivial for every generator except the first—we have

$$\begin{aligned} \lambda x^5 &= \frac{x^2}{3}(xg_x) + r_1, \\ \lambda^2 x^4 &= x^4 g + r_2, \\ \lambda^3 x^3 &= \lambda x^3 g + r_3, \\ \lambda^4 x^2 &= \lambda^2 x^2 g + r_4, \\ \lambda^5 x &= \lambda^3 xg + r_5, \\ \lambda^6 &= \lambda^4 g + r_6, \end{aligned}$$

where $r_i \in \mathcal{M}^7$. After some finagling we find for the remaining generator

$$x^6 = \tfrac{9}{25}x^2 g + (\tfrac{1}{5}x - \tfrac{3}{25}x^2)(xg_x) - \tfrac{3}{25}(\lambda g_x) + r_0, \tag{6.6}$$

where $r_0 \in \mathcal{M}^7$. Thus (6.4) follows from Lemma 5.2.

Remark. One might also use Lemma 5.1, but this is quite tedious.

As our next step in deriving (6.2) we show that

$$\mathcal{M}^4\langle \lambda \rangle \subset RT(x^5 + \lambda x^3 + \lambda^2). \tag{6.7}$$

Now $\mathcal{M}^4\langle \lambda \rangle$ is generated by $x^4\lambda, x^3\lambda^2, \ldots, \lambda^5$. For each of these generators, we have the representation

$$\begin{aligned} x^4\lambda &= x(xg_x) + r_1, \\ x^3\lambda^2 &= x^3 g + r_2, \\ x^2\lambda^3 &= x^2 g + r_3, \\ x\lambda^4 &= \lambda^2 xg + r_4, \\ \lambda^5 &= \lambda^3 g + r_5, \end{aligned} \tag{6.8}$$

were $r_i \in \mathcal{M}^6$. But $\mathcal{M}^6 \subset RT(g)$, so both terms on the right in (6.8) belong to $RT(g)$. This proves (6.7).

We argue similarly to show

$$\mathcal{M}\langle \lambda^2 \rangle \subset RT(g); \tag{6.9}$$

specifically, we have

$$x\lambda^2 = xg + r_1,$$

$$\lambda\lambda^2 = \lambda g + r_2,$$

where $r_i \in \mathcal{M}^6 + \mathcal{M}^4\langle\lambda\rangle \subset RT(g)$.

It follows by combining (6.4), (6.7), and (6.9) that

$$\mathcal{M}^6 + \mathcal{M}^4\langle\lambda\rangle + \mathcal{M}\langle\lambda^2\rangle \subset RT(g). \tag{6.10}$$

To complete the derivation of (6.2) we must specify the elements of $RT(g)$ that are not contained in the left-hand side of (6.10). Now $RT(g)$ consists of the totality of germs of the form

$$ag + b(xg_x) + c(\lambda g_x), \tag{6.11}$$

where $a, b, c \in \mathcal{E}_{x,\lambda}$. Note that λg_x already belongs to the left-hand side of (6.10); thus this generator will not contribute any new elements of $RT(g)$. Although g and xg_x do not lie in the left-hand side of (6.10), we do have the following:

$$xg,\ \lambda g,\ x(xg_x),\ \lambda(xg_x) \in \mathcal{M}^6 + \mathcal{M}^4\langle\lambda\rangle + \mathcal{M}\langle\lambda^2\rangle. \tag{6.12}$$

For any $a \in \mathcal{E}_{x,\lambda}$ we may write

$$a = a(0, 0) + \tilde{a},$$

where $\tilde{a} \in \mathcal{M}$; thus

$$ag = a(0, 0)g + \tilde{a}g,$$

and by (6.12) the term $\tilde{a}g$ belongs to the left-hand side of (6.10). Similar remarks hold for $b(xg_x)$. In conclusion, every germ of the form (6.11) may be written as

$$a(0, 0)g + b(0, 0)xg_x + r$$

where $r \in \mathcal{M}^6 + \mathcal{M}^4\langle\lambda\rangle + \mathcal{M}\langle\lambda^2\rangle$, which completes the proof of (6.2).

Let us reflect on the transition from (6.10) to (6.2). We know that $\mathcal{M}^6$ has finite codimension in $\mathcal{E}_{x,\lambda}$, so *a fortiori* the larger ideal $\mathcal{M}^6 + \mathcal{M}^4\langle\lambda\rangle + \mathcal{M}\langle\lambda^2\rangle$ has finite codimension in the smaller space $RT(g) \subset \mathcal{E}_{x,\lambda}$; i.e., there exists a finite-dimensional subspace V such that

$$RT(g) = [\mathcal{M}^6 + \mathcal{M}^4\langle\lambda\rangle + \mathcal{M}\langle\lambda^2\rangle] \oplus V.$$

In deriving (6.2) we showed that V is two dimensional, with g and xg_x as a basis.

We now derive conditions on a function which characterize membership in $RT(g)$. (*Remark*: Lemma 5.1 provides a straightforward but tiresome procedure for doing this; here we use *ad hoc* methods that we will formalize in the next section.) First we claim that $f \in RT(g)$ if and only if $j^5 f \in RT(g)$, where $j^5 f$ is the fifth-order Taylor polynomial of f. To see this we observe that

$$f = j^5f + (f - j^5f);$$

by (5.3) the second term here belongs to $\mathcal{M}^6 \subset RT(g)$, and the claim follows. Let us write

$$j^5f = \sum_{|\alpha| \le 5} a_\alpha x^{\alpha_1}\lambda^{\alpha_2}, \tag{6.13}$$

using multi-index notation. It follows from (6.7) and (6.9) that many of the coefficients in (6.13) have no bearing on whether or not $j^5f \in RT(g)$. Consider, for example, $a_{4,1}$, the coefficient of $x^4\lambda$. We observe that

$$j^5f = (j^5f - a_{4,1}x^4\lambda) + a_{4,1}x^4\lambda.$$

Because $x^4\lambda \in \mathcal{M}^4\langle\lambda\rangle \subset RT(g)$, j^5f belongs to $RT(g)$ if and only if the sum (6.13) with the $x^4\lambda$ term omitted belongs to $RT(g)$. On dropping in this way all the monomials in (6.13) that are contained in $\mathcal{M}^4\langle\lambda\rangle + \mathcal{M}\langle\lambda^2\rangle$, we find that $j^5f \in RT(g)$ if and only if

$$\begin{aligned}(a_{00} + a_{10}x + a_{20}x^2 + a_{30}x^3 + a_{40}x^4 + a_{50}x^5)&\\ + (a_{01} + a_{11}x + a_{21}x^2 + a_{31}x^3)\lambda + a_{22}\lambda^2&\end{aligned} \tag{6.14}$$

belongs to $RT(g)$. Most of the coefficients in (6.14) must vanish for f to belong to $RT(g)$. Let us see why. The three generators g, xg_x, λg_x of $RT(g)$ all satisfy

$$\left(\frac{\partial}{\partial x}\right)^j f(0, 0) = 0, \qquad j = 0, 1, 2, 3, 4,$$

$$\left(\frac{\partial}{\partial \lambda}\right)\left(\frac{\partial}{\partial x}\right)^j f(0, 0) = 0, \qquad j = 0, 1, 2,$$

and any combination of these generators as in (6.11) must also satisfy these conditions. Thus f can belong to $RT(g)$ only if

$$a_{00} = a_{10} = a_{20} = a_{30} = a_{40} = a_{01} = a_{11} = a_{21} = 0. \tag{6.15}$$

Assuming (6.15) holds, (6.14) reduces to the three terms

$$a_{50}x^5 + a_{31}x^3\lambda + a_{02}\lambda^2.$$

This polynomial belongs to $RT(g)$ if and only if it can be written as a linear combination of the basis vectors $x^5 + \lambda x^3 + \lambda^2$ and $5x^5 + 3\lambda x^3$ in (6.2); i.e., if and only if

$$3a_{50} - 5a_{31} + 2a_{02} = 0. \tag{6.16}$$

EXERCISE

6.1. Show directly that $\mathcal{M}^6 + \mathcal{M}^4\langle\lambda\rangle + \mathcal{M}\langle\lambda^2\rangle \subset RT(x^5 + \lambda x^3 + \lambda^2)$ using Nakayama's Lemma. First verify that $\mathcal{M}^6 + \mathcal{M}^4\langle\lambda\rangle + \mathcal{M}\langle\lambda^2\rangle = \langle x^6, x^4\lambda, x\lambda^2, \lambda^3\rangle$, then show that each generator $x^6, x^4\lambda, x\lambda^2, \lambda^3$ is contained in

$$RT(x^5 + \lambda x^3 + \lambda^2) + \mathcal{M}^7 + \mathcal{M}^5\langle\lambda\rangle + \mathcal{M}^2\langle\lambda^2\rangle.$$

§7. Principles for Calculating $RT(g)$, III: Intrinsic Ideals

The ideas used in the calculation of $RT(x^5 + \lambda x^3 + \lambda^2)$ in the previous section are generally applicable. In this section we formalize these ideas. We will draw on the concepts described here in stating our main results in the next section. Like §§4 and 5, the earlier sections entitled "Principles for Calculating $RT(g)$...", this section introduces some algebraic concepts. In the earlier sections these concepts were a standard part of algebraic terminology; by contrast the concepts here are quite specialized to our task, and the terminology is not at all standard. In particular, these concepts only apply to $\mathcal{E}_{x,\lambda}$, rather than $\mathcal{E}_n$ for arbitrary n, as now we deal with equivalence transformations.

(a) Basic Ideas Concerning Intrinsic Ideals

We shall call an ideal $\mathcal{I} \subset \mathcal{E}_{x,\lambda}$ *intrinsic* if the following implication is valid: For all $g, h \in \mathcal{E}_{x,\lambda}$,

$$g \in \mathcal{I} \quad \text{and} \quad h \sim g \;\Rightarrow\; h \in \mathcal{I},$$

where $h \sim g$ means h is strongly equivalent to g. Alternatively put, an ideal is intrinsic if, as a set, it is invariant under all strong equivalence transformations.

For example, in Exercise 7.1 we ask the reader to verify that $\mathcal{M}$ and $\langle\lambda\rangle$ are intrinsic ideals. Also, if $\mathcal{I}$ and $\mathcal{J}$ are intrinsic ideals, so are $\mathcal{I} + \mathcal{J}$ and $\mathcal{I} \cdot \mathcal{J}$. To see this for $\mathcal{I} + \mathcal{J}$, suppose that $g \in \mathcal{I} + \mathcal{J}$ and $h = Sg(X, \lambda)$ is strongly equivalent to g. We may write $g = g_1 + g_2$ where $g_1 \in \mathcal{I}$ and $g_2 \in \mathcal{J}$. Thus

$$h = S(g_1 + g_2)(X, \lambda) = Sg_1(X, \lambda) + Sg_2(X, \lambda).$$

But $Sg_1(X, \lambda) \in \mathcal{I}$, since $\mathcal{I}$ is intrinsic, and similarly $Sg_2(X, \lambda) \in \mathcal{J}$. Therefore $h \in \mathcal{I} + \mathcal{J}$. This shows $I + \mathcal{J}$ is intrinsic; the proof for $\mathcal{I} \cdot \mathcal{J}$ is similar. The following proposition shows that the most general intrinsic ideal can be obtained from the two basic examples, $\mathcal{M}$ and $\langle\lambda\rangle$, through these two operations.

Proposition 7.1. *Let $\mathscr{I} \subset \mathscr{E}_{x,\lambda}$ be an ideal of finite codimension. Then $\mathscr{I}$ is intrinsic if and only if it can be written*

$$\mathscr{I} = \mathscr{M}^k + \mathscr{M}^{k_1}\langle\lambda^{l_1}\rangle + \cdots + \mathscr{M}^{k_s}\langle\lambda^{l_s}\rangle \tag{7.1}$$

for some finite set of nonnegative integers k_i, l_i.

In (7.1) we use the convention that $\mathscr{M}^0 = \mathscr{E}_{x,\lambda}$. We shall normally require that

$$\begin{aligned} &\text{(a)} \quad 0 < l_1 < l_2 < \cdots < l_s, \\ &\text{(b)} \quad k > k_1 + l_1 > k_2 + l_2 > \cdots > k_s + l_s > 0, \end{aligned} \tag{7.2}$$

so that each summand in (7.1) actually contributes something to $\mathscr{I}$. We will prove this proposition in subsection (b) below.

Now let us argue that for any ideal $\mathscr{I}$ in $\mathscr{E}_{x,\lambda}$ of finite codimension, there is a largest intrinsic ideal that is contained in $\mathscr{I}$. By Proposition 5.7 there is an integer k such that $\mathscr{M}^k \subset \mathscr{I}$. It follows from Proposition 7.1 that there are only finitely many intrinsic ideals $\mathscr{J}$ such that

$$\mathscr{M}^k \subset \mathscr{J} \subset \mathscr{I}. \tag{7.3}$$

The sum of all these is an intrinsic ideal which also satisfies (7.3); thus it must be the largest intrinsic ideal contained in $\mathscr{I}$. We denote it Itr $\mathscr{I}$. Because of (7.3), Itr $\mathscr{I}$ has finite codimension.

Formula (6.2) shows that in general $RT(g)$ need not be an intrinsic ideal. However, we may interpret the calculations of §6 in terms of this concept as follows: First we found the largest intrinsic ideal contained in $RT(g)$, namely Itr $RT(g) = \mathscr{M}^6 + \mathscr{M}^4\langle\lambda\rangle + \mathscr{M}\langle\lambda^2\rangle$, and then we characterized what was left over. This is a generally applicable method; in the rest of subsection (a) we explore this method more fully.

Let $\mathscr{I} \subset \mathscr{E}_{x,\lambda}$ be an ideal of finite codimension. We use the notation $\mathscr{I}^\perp$ for the finite-dimensional vector subspace of $\mathscr{E}_{x,\lambda}$ spanned by the monomials not belonging to $\mathscr{I}$. For example $(\mathscr{M}^{k+1})^\perp$ consists of all polynomials of degree k or less. We claim that for any ideal of finite codimension

$$\mathscr{I} + \mathscr{I}^\perp = \mathscr{E}_{x,\lambda}. \tag{7.4}$$

We derive (7.4) as follows. Since $\mathscr{I}$ has finite codimension, Proposition 5.7 implies that there is an integer k such that $\mathscr{M}^{k+1} \subset \mathscr{I}$. For any $f \in \mathscr{E}_{x,\lambda}$ we may write

$$f = j^k f + (f - j^k f). \tag{7.5}$$

The second term on the right-hand side of (7.5) belongs to $\mathscr{M}^{k+1} \subset \mathscr{I}$. Thus to determine whether f is in $\mathscr{I} + \mathscr{I}^\perp$ we need only consider whether the polynomial $j^k f$ is in $\mathscr{I} + \mathscr{I}^\perp$. However, the set of monomials of degree less than $k + 1$ divide into two sets: those in $\mathscr{I}$ and those not in $\mathscr{I}$. Using this division we may write $j^k f = f_1 + f_2$ where $f_1 \in \mathscr{I}$ and $f_2 \in \mathscr{I}^\perp$. This proves the claim.

In general, it is not the case that $\mathscr{I} \cap \mathscr{I}^\perp = \{0\}$—the example $\langle x^2 + \lambda^2, \lambda x \rangle$ shows this. However, for intrinsic ideals the situation is different.

Lemma 7.2. *If $\mathscr{I}$ is intrinsic ideal of finite codimension, then*

$$\mathscr{E}_{x,\lambda} = \mathscr{I} \oplus \mathscr{I}^\perp.$$

Proof. In view of (7.4), one need only prove that $\mathscr{I} \cap \mathscr{I}^\perp = \{0\}$. This fact follows directly from the next proposition, and the latter is proved in subsection (b) below. □

Proposition 7.3. *Let $\mathscr{I}$ be an intrinsic ideal of finite codimension. A polynomial*

$$p(x, \lambda) = \sum a_\alpha x^{\alpha_1} \lambda^{\alpha_2}$$

belongs to $\mathscr{I}$ if and only if for every α such that $a_\alpha \neq 0$, the monomial $x^{\alpha_1}\lambda^{\alpha_2}$ belongs to $\mathscr{I}$.

Corollary 7.4. *Let $\mathscr{I} \subset \mathscr{E}_{x,\lambda}$ have finite codimension. Then*

$$\mathscr{I} = (\text{Itr } \mathscr{I}) \oplus V, \tag{7.6}$$

where $V = \mathscr{I} \cap (\text{Itr } \mathscr{I})^\perp$.

Proof. Since $V \subset (\text{Itr } \mathscr{I})^\perp$, it follows from Lemma 7.2 applied to Itr $\mathscr{I}$ that $(\text{Itr } \mathscr{I}) \cap V = \{0\}$. On the other hand, for any $f \in \mathscr{E}_{x,\lambda}$ we may write $f = f_1 + f_2$ where $f_1 \in \text{Itr } \mathscr{I}$ and $f_2 \in (\text{Itr } \mathscr{I})^\perp$. If $f \in \times$, then $f_2 \in \mathscr{I}$, since $f \in \text{Itr } \mathscr{I} \subset \mathscr{I}$. Thus $f_2 \in \mathscr{I} \cap (\text{Itr } \mathscr{I})^\perp = V$. This shows that $(\text{Itr } \mathscr{I}) + V = \mathscr{I}$ which completes the proof. □

Remark. The conclusion of this corollary, formula (7.6), is also valid if $\mathscr{I}$ is just a vector subspace of $\mathscr{E}_{x,\lambda}$ containing $\mathscr{M}^k$ for some integer k. See Exercise 7.3.

We noted above that the calculation of $RT(g)$ may be divided into two stages: First to determine Itr $RT(g)$ and then to characterize the remaining elements of $RT(g)$. Corollary 7.4 sets a context for this second stage of the calculation—specifically, it is required to find the subspace V in (7.6). For example, it may be seen from (6.2) that for our example $g(x, \lambda) = x^5 + \lambda x^3 + \lambda^2$ we have

$$V = \mathbb{R}\{x^5 + \lambda x^3 + \lambda^2, 5x^5 + 3\lambda x^3\};$$

in particular, dim $V = 2$. More generally, for any $g \in \mathscr{E}_{x,\lambda}$ such that $RT(g)$ has finite codimension, $(\text{Itr } RT(g))^\perp$ is a finite-dimensional subspace of $\mathscr{E}_{x,\lambda}$; indeed, since the monomials provide a distinguished basis for $(\text{Itr } RT(g))^\perp$, this subspace is *canonically* isomorphic to $\mathbb{R}^N$ for some N. Moreover, all calculations in determining V may be performed within this space. In other words, this part of the calculation of $RT(g)$ only involves linear algebra.

(b) Further Study of Intrinsic Ideals

The primary task of this subsection is to prove Propositions 7.1 and 7.3 above. These proofs are based on the following lemma.

Lemma 7.5. *Let $\mathscr{I} \subset \mathscr{E}_{x,\lambda}$ be an intrinsic ideal of finite codimension. If a germ f belongs to $\mathscr{I}$, then xf_x and λf_x also belong to $\mathscr{I}$. In particular, if $f \in \mathscr{I}$, then $RT(f) \subset \mathscr{I}$.*

Remark. The conclusion of Lemma 7.5 remains valid if f is a polynomial, even if $\mathscr{I}$ does not have finite codimension. See Exercise 7.4.

PROOF. Since $\mathscr{I}$ has finite codimension, there is a k such that $\mathscr{M}^{k+1} \subset \mathscr{I}$. By Taylor's theorem

$$f(x, \lambda) = j^k f(x, \lambda) + r(x, \lambda),$$

where $r \in \mathscr{M}^{k+1}$. Since xr_x and λr_x certainly belong to $\mathscr{M}^{k+1} \subset \mathscr{I}$, we may replace f by $j^k f$ in the lemma. Thus it suffices to prove the lemma when f is a polynomial of degree k or less; in symbols, when $f \in (\mathscr{M}^{k+1})^\perp \cap \mathscr{I}$. Observe that $f(tx, \lambda) \in (\mathscr{M}^{k+1})^\perp \cap \mathscr{I}$ for all $t > 0$, since $\mathscr{I}$ is intrinsic. It follows that

$$\rho(t) = \frac{f(tx, \lambda) - f(x, \lambda)}{t - 1}$$

is in $(\mathscr{M}^{k+1})^\perp \cap \mathscr{I}$ for each t. However, $\mathscr{I} \cap (\mathscr{M}^{k+1})^\perp$ is a closed set, being a linear subspace of the finite-dimensional space $(\mathscr{M}^{k+1})^\perp$. Hence $\lim_{t\to 1} \rho(t)$ is in $(\mathscr{M}^{k+1})^\perp \cap \mathscr{I}$; but this limit is just $xf_x(x, \lambda)$.

Similarly $f(x + t\lambda, \lambda) \in \mathscr{I}$ for all t since $\mathscr{I}$ is intrinsic. Differentiation with respect to t and evaluation at $t = 0$ yields the germ $\lambda f_x(x, \lambda)$, and this is in $\mathscr{I}$. □

PROOF OF PROPOSITION 7.3. It is trivial that $p \in \mathscr{I}$ if $a_\alpha \neq 0$ only when $x^{\alpha_1}\lambda^{\alpha_2}$ belongs to $\mathscr{I}$. To prove the converse, we consider an arbitrary multi-index $\alpha = (l, m)$ such that $a_\alpha \neq 0$, and we show that $x^l\lambda^m \in \mathscr{I}$. Choose a k such that $\mathscr{M}^{k+1} \subset \mathscr{I}$. If $l + m > k$, then the desired conclusion follows trivially; thus we assume that $l + m \leq k$. As in proving Lemma 7.5, we may reduce to the case that p is a polynomial of degree k or less. We group the terms in p according to degree in x—say

$$p(x, \lambda) = p_0(\lambda) + p_1(\lambda)x + \cdots + p_k(\lambda)x^k. \tag{7.7}$$

We claim that $p_j(\lambda)x^j$ is in $\mathscr{I}$ for $0 \le j \le k$. To prove this we observe, as in the proof of Lemma 7.5, that

$$p(tx, \lambda) = p_0(\lambda) + tp_1(\lambda)x + \cdots + t^k p_k(\lambda)x^k$$

is in $\mathscr{I}$ for every $t > 0$. Repeated differentiation with respect to t yields germs still belonging to $\mathscr{I}$. Differentiating k-times, we see that the last term, $p_k(\lambda)x^k$, is in $\mathscr{I}$. The claim follows from a simple induction argument proceeding from the last term forward.

Now we focus on $p_l(\lambda)$, the coefficient of x^l in (7.7) Let us write

$$p_l(\lambda) = b_0 + b_1\lambda + \cdots + b_{k-l}\lambda^{k-l}. \tag{7.8}$$

This polynomial cannot vanish identically, since $b_m = a_{lm} \neq 0$. Let b_μ be the first nonvanishing coefficient in (7.8); then $\mu \le m$. Thus

$$p_l(\lambda) = \lambda^\mu q(\lambda), \tag{7.9}$$

where $q(0) \neq 0$; this means that $1/q \in \mathscr{E}_{x,\lambda}$, so that (7.9) may be inverted. Therefore

$$x^l\lambda^m = \lambda^{\mu-m}(\lambda^\mu x^m) = \lambda^{\mu-m}\frac{1}{q(\lambda)}p_l(\lambda)\,x^m.$$

But $p_l(\lambda)x^m \in \mathscr{I}$ and $\mathscr{I}$ is an ideal; therefore, $x^l\lambda^m \in \mathscr{I}$. □

Proof of Proposition 7.1. Clearly every ideal of the form (7.1) is intrinsic, since sums and products of intrinsic ideals are intrinsic and $\mathscr{M}$ and $\langle\lambda\rangle$ are intrinsic.

For proving the converse we use the following corollary of Lemma 7.5.

Corollary 7.6. *If a monomial $x^l\lambda^m$ belongs to $\mathscr{I}$, where $\mathscr{I}$ is an intrinsic ideal of finite codimension, then $\mathscr{M}^l\langle\lambda^m\rangle \subset \mathscr{I}$.*

Proof of Corollary 7.6. To show this we must prove that the generators of $\mathscr{M}^l\langle\lambda^m\rangle$ belong to $\mathscr{I}$; in symbols

$$x^l\lambda^m, x^{l-1}\lambda^{m+1}, \ldots, \lambda^{l+m} \in \mathscr{I}. \tag{7.10}$$

But (7.10) follows from repeated application of $\lambda\,\partial/\partial x$ to $x^l\lambda^m$ as in Lemma 7.5. □

Proof of Proposition 7.1 (Continued). Assume that $\mathscr{I}$ is an intrinsic ideal of finite codimension. We choose the smallest integer k such that $\mathscr{M}^k \subset \mathscr{I}$. Suppose that there are nonzero elements in $\mathscr{I} \sim \mathscr{M}^k$. (*Note*: Here the symbol $\sim$ indicates the difference of sets, *not equivalence*.) We may take these to be polynomials. By Proposition 7.3 there are monomials $x^l\lambda^m$ in $\mathscr{I} \sim \mathscr{M}^k$. As noted above, if $x^l\lambda^m \in \mathscr{I}$, then $\mathscr{M}^l\langle\lambda^m\rangle \subset \mathscr{I}$. Applying these ideas to each of the finitely many monomials not contained in $\mathscr{M}^k$, we conclude that $\mathscr{I}$ is the sum of ideals of the form $\mathscr{M}^{k_i}\langle\lambda^{l_i}\rangle$, as in (7.1). On eliminating redundancies we obtain (7.2). □

Remark 7.7. We have shown in this classification that intrinsic ideals are invariant under all equivalences, not just strong equivalences. This follows because $\mathcal{M}$ and $\langle\lambda\rangle$ are invariant under changes of coordinate in λ of the form $\Lambda(\lambda)$, as well as invariant under strong equivalences. See Exercise 7.2.

In the above proof we showed that the smallest intrinsic ideal containing

$$x^k, x^{k_1}\lambda^{l_1}, \ldots, x^{k_s}\lambda^{l_s} \tag{7.11}$$

is

$$\mathcal{M}^k + \mathcal{M}^{k_1}\langle\lambda^{l_1}\rangle + \cdots + \mathcal{M}^{k_s}\langle\lambda^{l_s}\rangle. \tag{7.12}$$

We incorporate that fact in the following definition.

Definition 7.8. Assuming (7.2) holds, we call the monimals (7.11) the *intrinsic generators* of the ideal (7.12).

EXERCISES

7.1. Show that the ideals $\mathcal{M}$ and $\langle\lambda\rangle$ are intrinsic.

7.2. Show that the ideals $\mathcal{M}$ amd $\langle\lambda\rangle$ are invariant under all equivalences, including those involving λ.

7.3. Let $\mathcal{I} \subset \mathcal{E}_{x,\lambda}$ be a vector subspace containing $\mathcal{M}^k$ for some k.

(a) Show that $\mathcal{I}_1 + \mathcal{I}_2 \subset \mathcal{I}$ whenever $\mathcal{I}_1, \mathcal{I}_2$ are intrinsic ideals contained in $\mathcal{I}$. (It follows that Itr $\mathcal{I}$ is well defined.)

(b) Prove that Corollary 7.4 is valid for $\mathcal{I}$.

7.4. Verify that the proof of Lemma 7.4 is valid assuming that $f \in \mathcal{I}$ is a polynomial and that $\mathcal{I}$ is intrinsic, but not necessarily of finite codimension.

7.5. Prove the converse of Lemma 7.5. That is, show that an ideal $\mathcal{I}$ of finite codimension in $\mathcal{E}_{x,\lambda}$ is intrinsic if for every p in $\mathcal{I}$, $RT(p) \subset \mathcal{I}$.

§8. Formulation of the Main Results

As we indicated in §0, our algorithm for solving the recognition problem splits monomials into three classes: low-, intermediate-, and higher-order terms. We discuss the three classes in sequence: low-order first, then higher-order, and intermediate-order last.

(a) Low-Order Terms

In describing the low-order terms we need to speak of "the smallest intrinsic ideal containing a germ h." Let us argue that such an object exists. If $\mathcal{I}$ and $\mathcal{J}$ are two ideals in $\mathcal{E}_{x,\lambda}$, then the intersection $\mathcal{I} \cap \mathcal{J}$ is also an ideal—this statement and the ones following are easy exercises left to the reader. Moreover, if $\mathscr{I}$ and $\mathscr{J}$ are both intrinsic, so is $\mathscr{I} \cap \mathscr{J}$. Indeed, for an arbitrary collection of intrinsic ideals $\{\mathscr{I}_\alpha\}$, the intersection $\bigcap_\alpha \mathscr{I}_\alpha$ is an intrinsic ideal. Thus we may identify the smallest intrinsic ideal containing h as the intersection of all intrinsic ideals which contain h.

Definition 8.1. If $h \in \mathscr{E}_{x,\lambda}$, we denote by $\mathscr{S}(h)$ the smallest intrinsic ideal containing h.

The next proposition lists several properties of $\mathscr{S}(h)$. Note that in spite of the purely existential description of $\mathscr{S}(h)$ above, part (b) of the proposition characterizes $\mathscr{S}(h)$ explicitly.

Proposition 8.2. *Let $h \in \mathscr{E}_{x,\lambda}$ be a germ such that $RT(h)$ has finite codimension.*

(a) *$\mathscr{S}(h)$ is an intrinsic ideal of finite codimension.*

(b) $$\mathscr{S}(h) = \sum_{\alpha=(\alpha_1,\alpha_2)} \{\mathscr{M}^{\alpha_1}\langle\lambda^{\alpha_2}\rangle : D^\alpha h(0,0) \neq 0\}. \tag{8.1}$$

(c) *If g is equivalent to h, then $\mathscr{S}(g) = \mathscr{S}(h)$.*

Proof. (c) Suppose g is equivalent to h. Since we have $h \in \mathscr{S}(h)$ and since $\mathscr{S}(h)$ is intrinsic, we deduce that $g \in \mathscr{S}(h)$. (This holds even though we consider general equivalences, not just strong equivalences—see Remark 7.7.) In other words, $\mathscr{S}(h)$ is an intrinsic ideal which contains g; therefore $\mathscr{S}(g) \subset \mathscr{S}(h)$. Reversing the roles of g and h, we obtain the reverse containment.

(a) By construction $\mathscr{S}(h)$ is an intrinsic ideal; we need only show that $\mathscr{S}(h)$ has finite codimension. We reduce to the case where h is a polynomial as follows. Since $RT(h)$ has finite codimension, $\mathscr{M}^k \subset RT(h)$ for some integer k. It follows from Corollary 5.4(b) that h is equivalent to $j^k h$ and from part (c) of the present proposition that $\mathscr{S}(h) = \mathscr{S}(j^k h)$. Thus we may assume without loss of generality that h is a polynomial.

To show that $\mathscr{S}(h)$ has finite codimension, we prove that $RT(h) \subset \mathscr{S}(h)$. Indeed, each of the three generators of $RT(h)$ belongs to $\mathscr{S}(h)$—by definition $h \in \mathscr{S}(h)$, and it follows from Exercise 7.4 that xh_x, $\lambda h_x \in \mathscr{S}(h)$. Thus $\mathscr{S}(h)$ has finite codimension.

(b) As in part (a) we may reduce to the case where h is polynomial. (Thus the sum in (8.1) is effectively finite.) Since $h \in \mathscr{S}(h)$, it follows from Proposition 7.3 and Corollary 7.6 that $\mathscr{S}(h)$ contains the right-hand side of (8.1). For the reverse containment, we observe that the right-hand side of (8.1) is an intrinsic ideal to which h belongs; since $\mathscr{S}(h)$ is the *smallest* intrinsic ideal containing h, we see that $\mathscr{S}(h)$ is contained in the right-hand side of (8.1). □

We label the next result a theorem, in spite of the simplicity of its proof, because it characterizes the low-order terms in the recognition problem for h—low order terms are those which belong to $\mathscr{S}(h)^{\perp}$.

Theorem 8.3. *Let h be in $\mathscr{E}_{x,\lambda}$, and suppose $RT(h)$ has finite codimension. If g is equivalent to h, then for every monomial $x^{\alpha_1}\lambda^{\alpha_2} \in \mathscr{S}(h)^{\perp}$, we have $D^{\alpha}g(0, 0) = 0$.*

PROOF. Suppose $x^{\alpha_1}\lambda^{\alpha_2} \in \mathscr{S}(h)^{\perp}$ but $D^{\alpha}g(0, 0) \neq 0$. By Proposition 8.2(b), $x^{\alpha_1}\lambda^{\alpha_2} \in \mathscr{S}(g) = \mathscr{S}(h)$, a contradiction. □

$\mathscr{S}(h)$ also yields some information about intermediate-order terms in the recognition problem for h. Specifically, we have the following result. (We remind the reader of Definition 7.8, where intrinsic generators are introduced.)

Theorem 8.4. *Let $h \in \mathscr{E}_{x,\lambda}$ and suppose $RT(h)$ has finite codimension. If g is equivalent to h, then for every intrinsic generator $x^{\alpha_1}\lambda^{\alpha_2}$ of $\mathscr{S}(h)$, we have $D^{\alpha}g(0, 0) \neq 0$.*

PROOF. Consider an intrinsic generator $x^k\lambda^l$ of $\mathscr{S}(h)$; we ask whether $x^k\lambda^l \in \mathscr{S}(g)$. By Theorem 8.2

$$\mathscr{S}(g) = \sum\{\mathscr{M}^{\alpha_1}\langle\lambda^{\alpha_2}\rangle : D^{\alpha}g(0, 0) \neq 0\}. \tag{8.2}$$

What summands in (8.2) might contribute a term $x^k\lambda^l$ to $\mathscr{S}(g)$? By Theorem 8.3, $D^{\alpha}g(0, 0) = 0$ for all multi-indices α such that $x^{\alpha_1}\lambda^{\alpha_2} \in \mathscr{S}(h)^{\perp}$, so these terms contribute nothing. Of the terms which remain, only $\mathscr{M}^k\langle\lambda^l\rangle$ can contribute. By Proposition 8.3(c), $x^k\lambda^l \in \mathscr{S}(g)$, so this term *must* contribute in (8.2); i.e.,

$$\left(\frac{\partial}{\partial x}\right)^k \left(\frac{\partial}{\partial \lambda}\right)^l g(0, 0) \neq 0.$$ □

Let us illustrate the usefulness of these concepts on the pitchfork, $h(x, \lambda) = x^3 - \lambda x$. By Proposition 8.2(b), $\mathscr{S}(h) = \mathscr{M}^3 + \mathscr{M}\langle\lambda\rangle$. Now

$$\mathscr{S}(h)^{\perp} = \mathbb{R}\{1, x, \lambda, x^2\},$$

and the intrinsic generators of $\mathscr{S}(h)$ are x^3 and λx. From Theorems 8.3 and 8.4 we deduce that if g is equivalent to h, then at $(x, \lambda) = (0, 0)$

$$g = g_x = g_\lambda = g_{xx} = 0; \qquad g_{xxx} \neq 0, \quad g_{\lambda x} \neq 0. \tag{8.3}$$

As we saw in (0.3), this information provides essentially the complete solution of the recognition problem for the pitchfork; only the signs of g_{xxx} and $g_{\lambda x}$ are lacking. Of course, for more complicated normal forms the information given by $\mathscr{S}(h)$ will not be so complete.

(b) Higher-Order Terms

What do we mean by higher-order terms in the recognition problem for h? This concept should meet the following requirements: If p is a higher-order term, then $h + p$ is equivalent to h. However, it turns out that a more useful theory results if we strengthen this requirement, as follows: If p is a higher-order term, then for *any g equivalent to h, $g + p$* is equivalent to g. In our formal definition we make use of Theorem 2.2 to give this idea an algebraic formulation.

Definition 8.5. If $h \in \mathscr{E}_{x,\lambda}$, we define $\mathscr{P}(h)$ by the following condition: $p \in \mathscr{P}(h)$ if for every g strongly equivalent to h and for every $t \in \mathbb{R}$

$$RT(g + tp) = RT(g).$$

In words, $\mathscr{P}(h)$ is the set of higher-order terms in the recognition problem for h. (We use the letter "p" for perturbation.) The following proposition gives two properties of $\mathscr{P}(h)$.

Proposition 8.6. (a) *If $p \in \mathscr{P}(h)$ and if g is strongly equivalent to h, then $g + p$ is strongly equivalent to g.*
(b) *If $RT(h)$ has finite codimension, then $\mathscr{P}(h)$ is an intrinsic ideal of finite codimension.*

Part (a) of the lemma follows immediately from Theorem 2.2; we merely record it here for reference. Part (b) will be proved in §12 below.

The following theorem completely characterizes $\mathscr{P}(h)$ in an effectively computable way. It is the most important result of §8; indeed, of Chapter II.

Theorem 8.7. *If $RT(h)$ has finite codimension, then* $\mathscr{P}(h) = \operatorname{Itr} \mathscr{J}(h)$, *where*

$$\mathscr{J}(h) = \langle xh, \lambda h, x^2 h_x, \lambda h_x \rangle. \tag{8.4}$$

We prove Theorem 8.7 in §13.

The following is another formula for $\mathscr{J}(h)$:

$$\mathscr{J}(h) = \mathscr{M} \cdot RT(h) + \mathbb{R}\{\lambda h_x\}. \tag{8.5}$$

This result identifies the higher-order terms which cannot enter into the solution of the recognition problem. To illustrate this let us apply it to the pitchfork, $x^3 - \lambda x$. According to (6.3)

$$RT(x^3 - \lambda x) = \mathscr{M}^3 + \mathscr{M}\langle \lambda \rangle.$$

Substituting into (8.5) and computing $\lambda h_x = 3x^2\lambda - \lambda^2$ we find

$$\mathscr{J}(x^3 - \lambda x) = \mathscr{M}^4 + \mathscr{M}^2\langle \lambda \rangle + \langle \lambda^2 \rangle. \tag{8.6}$$

In this example we see that $\mathscr{J}(h)$ is already intrinsic, so $\mathscr{P}(x^3 - \lambda x)$ is also equal to the right-hand side of (8.6).

What are the consequences (8.6)? Consider, for example, the monomial $x^4 \in \mathscr{P}(x^3 + \lambda x)$. By Proposition 8.6(a), if a germ f is strongly equivalent to $x^3 - \lambda x$, then so is $f + tx^4$ for any $t \in \mathbb{R}$. By varying t we can make $(\partial/\partial x)^4 f(0, 0)$ achieve any value whatsoever. In other words, $(\partial/\partial x)^4 f(0, 0)$ *cannot enter into the solution of the recognition problem* for $x^3 - \lambda x$. Similarly, for higher-order derivatives with respect to x and for derivatives associated to monomials in $\mathscr{M}^2\langle\lambda\rangle$ or $\langle\lambda^2\rangle$; e.g., $(\partial/\partial x)^2(\partial/\partial\lambda)f(0, 0)$, $(\partial/\partial\lambda)^2 f(0, 0)$, etc. Of course, none of these derivatives appears in (0.3), our solution of the recognition problem for $x^3 - \lambda x$.

More generally, we shall describe a bifurcation problem h as *k-determined* if $h + p$ is equivalent to h for every $p \in \mathscr{M}^{k+1}$. (Thus $x^3 - \lambda x$ is 3-determined.) For h to be k-determined, it is necessary and sufficient that $\mathscr{M}^{k+1} \subset \mathscr{P}(h)$.

(c) Intermediate-Order Terms

Our treatment of intermediate-order terms is not so clean as our treatment of low- and higher-order terms, for the following reason: A concise description of what is going on at this level requires fairly sophisticated mathematical concepts from the theory of Lie groups, and some of the complexities of representation theory for Lie groups play a significant role. We don't address these issues in a serious way in this text. In this subsection we limit ourselves to the following three tasks:

(i) We complete the solution of the recognition problem for the pitchfork.
(ii) We solve the recognition problem for $x^5 + \lambda x^3 + \lambda^2$, the example considered in §6.
(iii) We sketch briefly what is required of intermediate-order terms in the solution of the recognition problem in general.

The methods that we use for items (i) and (ii) are elementary, and they suffice for all the examples we consider in §9. The discussion under item (iii) is intended more as a focus for our thinking than as a guide for computing; basically, this material is just a formalization of the methods used for the examples. In Chapter V we will indicate by example some of the complexities of the general case. There are many interesting theoretical issues needing to be investigated more fully, but we do not pursue these.

Our first task is to complete the solution of the recognition problem for the pitchfork. This is quite easy, given our results above. Let $g \in \mathscr{E}_{x,\lambda}$ be a germ strongly equivalent to $x^3 - \lambda x$. Combining the information in (8.3) and (8.6) we see that

$$g(x, \lambda) = ax^3 + b\lambda x + p(x, \lambda),$$

where $a \neq 0, b \neq 0$, and $p \in \mathscr{P}(x^3 - \lambda x)$. By Proposition 8.6, $g(x, \lambda)$ is strongly equivalent to $x^3 - \lambda x$ if and only if

$$\tilde{g}(x, \lambda) = ax^3 + b\lambda x \tag{8.7}$$

is strongly equivalent to $x^3 - \lambda x$. We may transform $\tilde{g}$ into $\varepsilon x^3 + \delta\lambda x$, where $\varepsilon = \operatorname{sgn} a$ and $\delta = \operatorname{sgn} b$, by scaling transformations; specifically we have

$$\varepsilon x^3 + \delta\lambda x = \left|\frac{a}{b^3}\right|^{1/2} \tilde{g}\left(\left|\frac{b}{a}\right|^{1/2} x, \lambda\right).$$

Because of the restrictions (0.2) on the sign of S and X_x, it is not possible to change the sign of the two coefficients in (8.7) by an equivalence transformation. Therefore, g is strongly equivalent to the pitchfork if and only if $a > 0$, $b < 0$; i.e., if (0.3) is satisfied. Our analysis also shows that reversing an inequality in (0.3b) merely changes a sign in the normal form $\pm x^3 \pm \lambda x$.

Next we solve the recognition problem for $h(x, \lambda) = x^5 + \lambda x^3 + \lambda^2$. The treatment of low- and higher-order terms here is the same as for the pitchfork; we have chosen this example for the new phenomena that appears in the intermediate-order terms. By Proposition 8.2(b),

$$\mathscr{S}(h) = \mathscr{M}^5 + \mathscr{M}^3\langle\lambda\rangle + \langle\lambda^2\rangle.$$

It follows from Theorems 8.3 and 8.4 that if g is strongly equivalent to h then

$$g(x, \lambda) = ax^5 + b\lambda x^3 + c\lambda^2 + p(x, \lambda),$$

where $a \neq 0$, $b \neq 0$, $c \neq 0$, and

$$p(x, \lambda) \in \mathscr{M}^6 + \mathscr{M}^4\langle\lambda\rangle + \mathscr{M}\langle\lambda^2\rangle. \tag{8.8}$$

Next we compute (cf. Exercise 8.2) that $\mathscr{P}(h)$ is precisely the right-hand side of (8.8), so g is strongly equivalent to h if and only if

$$\tilde{g}(x, \lambda) = ax^5 + b\lambda x^3 + c\lambda^2 \tag{8.9}$$

is strongly equivalent to h. We claim that the latter statement holds if and only if

$$a > 0, \qquad c > 0, \quad \text{and} \quad a^3c^2 = b^5. \tag{8.10}$$

(*Remark*: It follows from (8.10) that $b > 0$.) To show the sufficiency of (8.10), let us consider the effect of a pure scaling equivalence on (8.9); i.e., an equivalence of the form

$$S(x, \lambda) = \alpha, \qquad X(x, \lambda) = \beta x.$$

We find

$$\alpha\tilde{g}(\beta x, \lambda) = \alpha\beta^5 a x^5 + \alpha\beta^3 b\lambda x^3 + \alpha c\lambda^2. \tag{8.11}$$

By matching the three coefficients in (8.11) with those in h, we obtain three equations for two unknowns, α and β; these equations have a solution with

$\alpha > 0$, $\beta > 0$ if and only if (8.10) holds. In other words, (8.10) is a sufficient condition for g to be equivalent to h. We leave the proof that (8.10) is also necessary to the reader in Exercise 8.3. The basis of this part of the proof is the fact that *only* scaling equivalences make a useful contribution towards transforming (8.9)—higher-order terms in S or X only affect higher-order terms in g. Expressing the above calculation in terms of g, we find that g is strongly equivalent to h if and only if $g = g_x = g_{xx} = g_{xxx} = g_{xxxx} = g_\lambda = g_{\lambda x} = g_{\lambda xx} = 0$,

$$\left(\frac{g_{xxxxx}}{5!}\right)^3 \left(\frac{g_{\lambda\lambda}}{2}\right)^2 = \left(\frac{g_{\lambda xxx}}{3!}\right)^5; \qquad g_{xxxxx} > 0, \quad g_{\lambda\lambda} > 0. \tag{8.12}$$

It is instructive to contrast the above two examples. The solution (0.3) to the recognition problem for the pitchfork consists of the equalities (0.3a) and the inequalities (0.3b). All of the former came from consideration of low-order terms; all of the latter, from intermediate-order terms. In (8.12) most of the equalities came from the low-order terms, but one equality came from the intermediate-order terms. In complicated examples the intermediate-order terms often contribute equations as well as inequalities to the defining conditions of a singularity. This is related to the issue of moduli which we take up in Chapter V.

Let us attempt to describe the above treatment of intermediate level terms in a general context. Consider the recognition problem for a normal form h: Is a given germ g strongly equivalent to h? The essential idea in the above calculation is the following: Having reduced g modulo $\mathscr{P}(h)$ to as few terms as possible, we perform explicit changes of coordinate on g modulo $\mathscr{P}(h)$ to determine precisely when g is equivalent to h. In symbols, we attempt to find S and X such that

$$g \equiv Sh(X, \lambda) \mod \mathscr{P}(h). \tag{8.13}$$

The "mod $\mathscr{P}(h)$" in (8.13) is of the utmost importance—modulo $\mathscr{P}(h)$ equivalence transformations simplify enormously. Without the "mod $\mathscr{P}(h)$" the unknowns S and X in (8.13) would be arbitrary functions; with the "mod $\mathscr{P}(h)$" only finitely many terms in the Taylor series of X and S actually contribute to (8.13). More specifically, if $\mathscr{M}^{k+1} \subset \mathscr{P}(h)$, only terms of degree k or less in the Taylor series of S and X can contribute to (8.13), and usually only a fraction of these actually contribute. Indeed, for the calculations above, only the lowest-order terms contributed.

To conclude, our treatment of intermediate-order terms in the recognition problem is an explicit calculation involving a finite number of undetermined parameters. These calculations must be done on a case-by-case basis. The information provided by Theorem 8.4 is most useful in starting these calculations. In elementary examples the intermediate-order terms only contribute inequalities in the recognition problem; in more complicated examples they may contribute one or more equalities as well.

The natural context for these calculations is the theory of Lie groups. For the benefit of readers familiar with Lie theory we describe the calculations in these terms. Consider the action of the group of strong equivalence transformations on $\mathscr{S}(h)$; this action is a linear representation of the group. Now $\mathscr{P}(h)$ is an invariant subspace of this action, so there is an induced representation on $\mathscr{S}(h)/\mathscr{P}(h)$, a finite-dimensional space. In this induced representation, the infinite dimensional group of equivalence transformations reduces to a finite-dimensional algebraic group. The treatment of intermediate-order terms in the recognition problem for h may be summarized as follows: A germ g is equivalent to h if and only if g belongs to the orbit of h in $\mathscr{S}(h)/\mathscr{P}(h)$ under this action.

EXERCISES

8.1. Rederive Corollary 5.4(b) as a consequence of Theorem 8.7.

8.2. Compute $\mathscr{I}(x^5 + x^3\lambda + \lambda^2)$ and verify that

$$\mathscr{P}(x^5 + x^3\lambda + \lambda^2) = \mathscr{M}^6 + \mathscr{M}^4\langle\lambda\rangle + \mathscr{M}\langle\lambda^2\rangle.$$

8.3. Show that $\tilde{g}$ in (8.9) is strongly equivalent to $h(x, \lambda) = x^5 + x^3\lambda + \lambda^2$ precisely when (8.10) is valid. *Hint*: Compute the general strong equivalence of h modulo $\mathscr{P}(h)$.

8.4. We call a bifurcation problem $g(x, \lambda)$ *k-determined* if $g + p$ is equivalent to g for every $p(x, \lambda)$ in $\mathscr{M}^{k+1}$. Prove that g is k-determined if

$$\mathscr{M}^{k+1} \subset \mathscr{M} \cdot RT(g).$$

§9. Solution of the Recognition Problem for Several Examples

In this section we illustrate the use of the theorems in the previous section by solving the recognition problem for the following normal forms:

$$\begin{aligned} &\text{(a)} \quad \varepsilon x^k + \delta\lambda, \quad k \geq 2,\\ &\text{(b)} \quad \varepsilon x^k + \delta\lambda x, \quad k \geq 3,\\ &\text{(c)} \quad \varepsilon(x^2 + \delta\lambda^2),\\ &\text{(d)} \quad \varepsilon x^3 + \delta\lambda^2. \end{aligned} \tag{9.1}$$

Here ε and δ equal ± 1; i.e., we consider all possible signs in (9.1). Note that (9.1a, b) are actually infinite sequences of normal forms indexed by k. In particular, if $k = 3$ then (9.1b) yields the pitchfork, which we already analyzed in §8. Also we considered (9.1c) in the Introduction.

We consider the four normal forms (9.1) in sequence in the four propositions below. In starting these propositions we use the following convention concerning nondegeneracy conditions such as (9.2b): The equation $\varepsilon = \mathrm{sgn}(A)$ includes the requirement that $A \neq 0$. Our proofs of the last three propositions are much terser than the first, as all four proofs have much the same character. The only exception to this is that the treatment of intermediate-order terms for (9.1c) is a little more involved—in this case more than just a simple scaling is required. (Indeed, this is a good example to study in order to gain insight about the treatment of intermediate-order terms in general.)

Proposition 9.1. *A germ $g \in \mathscr{E}_{x,\lambda}$ is strongly equivalent to* (9.1a), $\varepsilon x^k + \delta\lambda$, *if and only if at $x = \lambda = 0$*

$$g = \frac{\partial}{\partial x} g = \cdots = \left(\frac{\partial}{\partial x}\right)^{k-1} g = 0, \tag{9.2a}$$

and

$$\varepsilon = \mathrm{sgn}\left(\frac{\partial}{\partial x}\right)^k g, \qquad \delta = \mathrm{sgn}\,\frac{\partial}{\partial \lambda} g. \tag{9.2b}$$

Proof. We prove the proposition in three stages, which correspond to the divisions of §8.

For brevity let us write $h(x, \lambda) = \varepsilon x^k + \delta\lambda$. First, we apply Proposition 8.2(b) to conclude that

$$\mathscr{S}(h) = \mathscr{M}^k + \langle\lambda\rangle.$$

It follows from Theorems 8.3 and 8.4 that if a germ g is strongly equivalent to h, then

$$g(x, \lambda) = ax^k + b\lambda + p(x, \lambda), \tag{9.3}$$

where $a \neq 0$, $b \neq 0$, and

$$p \in \mathscr{M}^{k+1} + \mathscr{M}\langle\lambda\rangle.$$

In particular, (9.2a) must hold.

Next, we compute that

$$RT(h) = \langle x^k, kx^{k-1}, \lambda\rangle = \mathscr{M}^k + \langle\lambda\rangle.$$

It follows from Theorem 8.7 that

$$\mathscr{P}(h) = \mathscr{M}^{k+1} + \mathscr{M}\langle\lambda\rangle.$$

In other words the remainder term p in (9.3) has no influence on whether or not g is equivalent to h.

Finally, we ask whether $\tilde{g}(x, \lambda) = ax^k + b\lambda$ is strongly equivalent to h. If (9.2b) is satisfied, then we have a simple scaling

$$\varepsilon x^k + \delta\lambda = \frac{1}{|b|}\tilde{g}\left(\left|\frac{b}{a}\right|^{1/k} x, \lambda\right).$$

If (9.2b) is not satisfied no such transformation is possible, because of conditions (0.2). □

Proposition 9.2. *A germ $g \in \mathscr{E}_{x,\lambda}$ is strongly equivalent to* (9.1b), $\varepsilon x^k + \delta\lambda x$, *if and only if at $x = \lambda = 0$*

$$g = \frac{\partial}{\partial x} g = \cdots = \left(\frac{\partial}{\partial x}\right)^{k-1} g = \frac{\partial}{\partial \lambda} g = 0, \tag{9.4a}$$

and

$$\varepsilon = \operatorname{sgn}\left(\frac{\partial}{\partial x}\right)^k g, \qquad \delta = \operatorname{sgn}\frac{\partial}{\partial \lambda}\frac{\partial}{\partial x} g. \tag{9.4b}$$

Proof. We have

$$\mathscr{S}(h) = \mathscr{M}^k + \mathscr{M}\langle\lambda\rangle$$

so that if g is strongly equivalent to h

$$g(x, \lambda) = ax^k + b\lambda x + p(x, \lambda), \tag{9.5}$$

where

$$p \in \mathscr{M}^{k+1} + \mathscr{M}^2\langle\lambda\rangle + \langle\lambda^2\rangle. \tag{9.6}$$

Now it turns out that $\mathscr{P}(h)$ is precisely the right-hand side of (9.6), so that we drop $p(x, \lambda)$ from (9.5). Finally, we may scale $\tilde{g}$ to h if and only if (9.4b) holds. □

Proposition 9.3. *A germ $g \in \mathscr{E}_{x,\lambda}$ is strongly equivalent to* (9.1c), $\varepsilon(x^2 + \delta\lambda^2)$, *if and only if at $x = \lambda = 0$*

$$g_x = g_\lambda = 0, \tag{9.7a}$$

and

$$\varepsilon = \operatorname{sgn} g_{xx}, \qquad \delta = \operatorname{sgn}\det d^2 g, \tag{9.7b}$$

where $d^2 g$ is the 2×2 Hessian matrix of the second derivatives of g.

Proof. In this case we have $\mathscr{S}(h) = \mathscr{M}^2$. From Theorems 8.3 and 8.4 we may conclude that

$$g(x, \lambda) = ax^2 + p(x, \lambda), \tag{9.8}$$

where $a \neq 0$ and $p \in \mathscr{M}^3 + \mathscr{M}\langle\lambda\rangle$. Unfortunately $\mathscr{P}(h) = \mathscr{M}^3$, so that not all possible remainders in (9.8) may be discarded. However, we may write

$$g(x, \lambda) = ax^2 + b\lambda x + c\lambda^2 + \tilde{p}(x, \lambda),$$

where $\tilde{p} \in \mathscr{M}^3$. When is $\tilde{g}(x, \lambda) = ax^2 + b\lambda x + c\lambda^2$ equivalent to h? Because of the sign restriction (0.2), an equivalence transformation cannot change $\operatorname{sgn} g_{xx}$ or $\operatorname{sgn} \det d^2g$, so (9.7b) is a necessary condition. Let us perform explicit changes of coordinates to show it is also sufficient. In this calculation it is convenient to use the fact that the composition of two equivalence transformations is also an equivalence transformation; therefore we may reduce $\tilde{g}$ to h in steps. If (9.7b) is satisfied, we first eliminate the cross term λx in $\tilde{g}$ by considering

$$\tilde{g}\left(x - \frac{b}{2a}\lambda, \lambda\right),$$

and then we reduce to h with scalings as in the preceding cases. □

Proposition 9.4. *A germ $g \in \mathscr{E}_{x,\lambda}$ is strongly equivalent to the winged cusp (9.1d), $\varepsilon x^3 + \delta\lambda^2$, if and only if at $x = \lambda = 0$*

$$g = g_x = g_\lambda = g_{xx} = g_{\lambda x} = 0,$$

and

$$\varepsilon = \operatorname{sgn} g_{xxx}, \qquad \delta = \operatorname{sgn} g_{\lambda\lambda}.$$

PROOF. Here $\mathscr{S}(h) = \mathscr{M}^3 + \langle \lambda^2 \rangle$ and $\mathscr{P}(h) = \mathscr{M}^4 + \mathscr{M}^2\langle\lambda\rangle$. After elimination of the low- and higher-order terms, a simple scaling suffices to reduce g to h. □

§10. The Recognition Problem: General Equivalences

Let $h \in \mathscr{E}_{x,\lambda}$ be a germ such that $RT(h)$ has finite codimension. In this section we address the recognition problem for h in the context of general equivalence transformations; i.e., given $g \in \mathscr{E}_{x,\lambda}$, we ask whether there is a general equivalence, possibly not a strong equivalence, which transforms g into h. In answering this question we again consider low-, higher-, and intermediate-order terms, as in §8; moreover, our treatment of low- and higher-order terms is exactly the same as in the preceding case. More precisely, low-order terms are those in $\mathscr{S}(h)^\perp$, higher-order terms are those in $\mathscr{P}(h)$, and intermediate-order terms are those which are left over. In §8(a) the important results concerning $\mathscr{S}(h)$, Theorems 8.3 and 8.4 already apply to general equivalences. In §8(b) we characterized $\mathscr{P}(h)$ (i.e., those terms which may be transformed away by a strong equivalence). Certainly these terms can be transformed away by a more general equivalence. The only difference between the strong and general equivalence contexts is in the intermediate-order terms; in the present case, there are a few extra parameters in the Taylor series of $\Lambda(\lambda)$ that may help in the solution of (8.13). Even this difference doesn't change the solution of the recognition problem for many simple examples; in particular, for all of the normal forms considered in §9, the solution of the recognition problem is the same in either context.

Let us illustrate these remarks on two examples, the pitchfork and $x^5 + \lambda x^3 + \lambda^2$. First we discuss the pitchfork; suppose g is equivalent to $x^3 - \lambda x$. As before, we deduce from Theorems 8.3 and 8.4 that

$$g(x, \lambda) = ax^3 + b\lambda x + p(x, \lambda),$$

where $a \neq 0$, $b \neq 0$, and

$$p \in \mathcal{M}^4 + \mathcal{M}^2\langle\lambda\rangle + \langle\lambda^2\rangle. \tag{10.1}$$

Since $\mathscr{P}(h)$ equals the right-hand side of (10.1), it follows that g is equivalent to h if and only if $ax^3 + b\lambda x$ is equivalent to h. The only obstacle to this is possible differences of sign. However, we require that Λ preserve orientation (in symbols $\Lambda'(\lambda) > 0$), so the additional flexibility provided by Λ does not help. In other words, (0.3) is necessary and sufficient for g to be equivalent to h.

Passing to the second example, we now suppose that g is equivalent to $h(x, \lambda) = x^5 + \lambda x^3 + \lambda^2$. As always, the question reduces to whether a polynomial $\tilde{g}(x, \lambda) = ax^5 + b\lambda x^3 + c\lambda^2$ is equivalent to h. In this case Λ provides a third scaling parameter that we may use to eliminate the complicated equality in (8.12) that came from the intermediate-order terms. More precisely, we can solve the equation

$$\alpha\tilde{g}(\beta x, \gamma\lambda) = x^5 + \lambda x^3 + \lambda^2,$$

with α, β, and γ all positive if and only if

$$a > 0, \qquad b > 0, \qquad c > 0.$$

Expressing this in terms of g, we conclude that g is equivalent to $x^5 + \lambda x^3 + \lambda^2$ if and only if

$$g = g_x = g_{xx} = g_{xxx} = g_{xxxx} = g_\lambda = g_{\lambda x} = g_{\lambda xx} = 0,$$
$$g_{xxxxx} > 0, \qquad g_{\lambda xxx} > 0, \qquad g_{\lambda\lambda} > 0.$$

Remark. In this example the complicated equality in (8.12) drops out of the solution of the recognition problem when we consider general equivalences. However, as we shall see in Chapter V, the solution to the recognition problem for complicated singularities, even in the context of general equivalences, may include such equalities.

§11. Proof of Theorem 2.2

Theorem 2.2 states that if

$$RT(g + tp) = RT(g) \tag{11.1}$$

for all $t \in [0, 1]$, then $g + tp$ is strongly equivalent to g for all $t \in [0, 1]$. It turns out that the following local version of the theorem is sufficient to derive the full result.

Proposition 11.1. *Let $g, p \in \mathscr{E}_{x,\lambda}$ be germs such that* (11.1) *is valid for t near* 0. *Then $g + tp$ is strongly equivalent to g for all t sufficiently near* 0.

Proof of theorem 2.2 (Assuming Proposition 11.1). Define t_1 and t_2 in $[0, 1]$ to be equivalent if $g + t_1 p$ is strongly equivalent to $g + t_2 p$. We claim that Proposition 11.1 implies that equivalence classes of t's in $[0, 1]$ are open. If the claim is valid, then it follows from either the compactness or connectedness of $[0, 1]$ that there is exactly one equivalence class. Hence $g + tp$ is strongly equivalent to g for all $t \in [0, 1]$.

To verify the claim let $h = g + t_0 p$ for some $t_0 \in [0, 1]$. Then

$$RT(h + sp) = RT(g + (s + t_0)p) = RT(g) = RT(h)$$

for all s sufficiently near 0. It follows from Proposition 11.1 that $h + sp$ is strongly equivalent to h for all s near 0. Thus $g + tp$ is strongly equivalent to $g + t_0 p$ for all t near t_0, and the equivalence classes of t's are open. The claim is verified. □

The main step in the proof of Proposition 11.1 is to construct the strong equivalence between $g + tp$ and g by solving certain ODE's. The following lemma specifies the precise information from hypothesis (11.1) that we need to formulate the ODE's.

Lemma 11.2. *If* (11.1) *is valid for all t near* 0, *then there exist coefficients a and $b \in \mathscr{E}_{x,\lambda,t}$ such that*

$$p(x, \lambda) = a(x, \lambda, t)G(x, \lambda, t) + b(x, \lambda, t)G_x(x, \lambda, t), \tag{11.2}$$

where $G(x, \lambda, t) = g(x, \lambda) + tp(x, \lambda)$. In addition $b(0, 0, t) \equiv 0$.

Remark. For each fixed t the validity of (11.2) follows directly from (11.1), since (11.1) implies that $p \in RT(g) = RT(g + tp)$. The point of Lemma 11.2 is that a and b can be chosen to vary smoothly in t.

Proof. Assume that (11.1) is valid for $t = t_0$ where $t_0 \neq 0$ is near 0. Assumption (11.1) implies that each generator of $RT(g + t_0 p)$ may be written as a linear combination of the generators of $RT(g)$. Let us elaborate. Recall that $RT(g)$ is generated by g, xg_x, and λg_x, and $RT(g + t_0 p)$ is generated by

$g + t_0 p$, $xg_x + t_0 xp_x$, and $\lambda g_x + t_0 \lambda p_x$. Thus there exist germs A_i, B_i, C_i ($i = 1, 2, 3$) such that

$$\begin{aligned} g + t_0 p &= A_1 g + B_1 x g_x + C_1 \lambda g_x, \\ x g_x + t_0 x p_x &= A_2 g + B_2 x g_x + C_2 \lambda g_x, \\ \lambda g_x + t_0 \lambda p_x &= A_3 g + B_3 x g_x + C_3 \lambda g_x. \end{aligned} \tag{11.3}$$

We may rearrange the terms in the system (11.3) to obtain a matrix equation

$$\begin{pmatrix} p \\ xp_x \\ \lambda p_x \end{pmatrix} = Q \begin{pmatrix} g \\ xg_x \\ \lambda g_x \end{pmatrix}, \tag{11.4}$$

where Q is a 3×3 matrix whose entries are smooth germs in $\mathscr{E}_{x,\lambda}$. Now for any germ h, we introduce the notation $v(h)$ for the column vector

$$v(h) = \begin{pmatrix} h \\ xh_x \\ \lambda h_x \end{pmatrix}.$$

Using this notation, we rewrite (11.4) in the form

$$v(p) = Qv(g). \tag{11.5}$$

Recalling that $G = g + tp$, we have

$$v(g) = v(G) - tv(p). \tag{11.6}$$

Substituting (11.6) into (11.5) and rearranging we find

$$(I + tQ)v(p) = Qv(G). \tag{11.7}$$

Observe that (11.7) is a system of equations with smooth dependence on t. Since I is invertible, it follows that for sufficiently small t, $I + tQ$ is an invertible 3×3 matrix. Thus $(I + tQ)^{-1}$ is a 3×3 matrix whose entries are smooth germs in $\mathscr{E}_{x,\lambda,t}$; in particular, these germs are smooth in t. The invertibility of $I + tQ$ and (11.7) imply

$$v(p) = (I + tQ)^{-1}Qv(G). \tag{11.8}$$

Equating the first components on each side of (11.8) yields the equation

$$p = \alpha g + \beta x G_x + \gamma \lambda G_x,$$

where α, β, and γ are in $\mathscr{E}_{x,\lambda,t}$. Finally, one obtains (11.2) by setting $a = \alpha$ and $b = x\beta + \lambda\gamma$. □

Proof of proposition 11.1. Lemma 11.2 states that (11.2) is valid for germs; hence this relation holds on some neighborhood of (0, 0, 0) in $x\lambda t$-space. Choose intervals K, L, M such that (11.2) is valid on $K \times L \times M$.

We wish to prove that $G(\cdot, \cdot, t)$ is strongly equivalent to g for each t sufficiently near 0. Specifically, we will construct mappings $X(x, \lambda, t)$ and $S(x, \lambda, t)$ varying smoothly in t and satisfying

$$\begin{aligned} &\text{(a)} \quad S(x, \lambda, t)G(X(x, \lambda, t), \lambda, t) = g(x, \lambda), \\ &\text{(b)} \quad X(0, 0, t) \equiv 0, \qquad X(x, \lambda, 0) \equiv x, \\ &\text{(c)} \quad S(x, \lambda, 0) \equiv 1. \end{aligned} \tag{11.9}$$

The functions X and S are found by solving certain ODE's. Specifically, consider

$$\begin{aligned} &\text{(a)} \quad \frac{dX}{dt}(x, \lambda, t) = -b(X(x, \lambda, t), \lambda, t), \\ &\text{(b)} \quad X(x, \lambda, 0) \ = x, \end{aligned} \tag{11.10}$$

and

$$\begin{aligned} &\text{(a)} \quad \frac{dS}{dt}(x, \lambda, t) = -a(X(x, \lambda, t), \lambda, t)S(x, \lambda, t), \\ &\text{(b)} \quad S(x, \lambda, 0) \ = 1, \end{aligned} \tag{11.11}$$

where a and b are the coefficients in (11.2).

To understand the reason for this choice of coefficients, let us assume for the moment that (11.10) and (11.11) have solutions on $K \times L \times M$ and differentiate the left-hand side of (11.9a) with respect to t. This yields:

$$\begin{aligned} \frac{d}{dt}[S(x, \lambda, t)G(X(x, \lambda, t), \lambda, t)] &= S_t(x, \lambda, t)G(X(x, \lambda, t), \lambda, t) \\ &+ S(x, \lambda, t)G_x(X(x, \lambda, t), \lambda, t)X_t(x, \lambda, t) \\ &+ S(x, \lambda, t)G_t(X(x, \lambda, t), \lambda, t). \end{aligned} \tag{11.12}$$

The right-hand side of (11.12) simplifies considerably using the facts that X and S solve the ODE's (11.10) and (11.11) and that $G = g + tp$. Setting $y = X(x, \lambda, t)$ we obtain

$$\begin{aligned} \frac{d}{dt}[S(x, \lambda, t)G(y, \lambda, t)] = S(x, \lambda, t)[&-a(y, \lambda, t)G(y, \lambda, t) \\ &-b(y, \lambda, t)G_x(y, \lambda, t) + p(y, \lambda)]. \end{aligned} \tag{11.13}$$

It follows from (11.2) that the right-hand side of (11.13) is identically zero. Hence

$$\begin{aligned} S(x, \lambda, t)G(X(x, \lambda, t), \lambda, t) &= S(x, \lambda, 0)G(X(x, \lambda, 0), \lambda, 0) \\ &= g(x, \lambda). \end{aligned}$$

In other words, (11.9a) follows if X and S satisfy (11.10) and (11.11), respectively.

We claim that the initial conditions (11.9b) and (11.9c) also follow if X and S satisfy (11.10) and (11.11). Of course, the second equation in (11.9b) is just the initial condition (11.10b), and (11.9c) is just the initial condition (11.11b). Thus we need only show that $X(0, 0, t) \equiv 0$. However, by Lemma 11.1, $b(0, 0, t) \equiv 0$, so the function $X(0, 0, t) \equiv 0$ is a solution of (11.10). By uniqueness of solutions, it is the only solution.

We end our proof by discussing why the ODE's (11.10) and (11.11) have solutions on $K \times L \times M$. We may, in fact, have to shrink K and L, but this will not disturb the argument above.

As we observed above, $X(0, 0, t) = 0$ is a solution to (11.10) for all t and thus for all t in M. The standard existence theorem for ODE's with smooth dependence on parameters states that the interval in t on which one can solve an ODE like (11.10) varies continuously with the parameters. Cf. Hirsch and Smale [1974], p. 169. Thus there exist intervals K, L containing 0 and an X defined on $K \times L \times M$ solving (11.10).

Once X has been defined, the ODE (11.11) is linear and as such has a solution for all t. □

Exercises

Exercises 11.1–11.5 form a block of material covering the basic determinacy results for elementary catastrophe theory.

11.1. Let f and g be germs in $\mathscr{E}_n$. We call f and g *right equivalent* if there exists a germ of a diffeomorphism $\phi: \mathbb{R}^n \to \mathbb{R}^n$ with $\phi(0) = 0$ such that $g(x) = f(\phi(x))$. Compute the restricted tangent space $RT_r(f)$ of the germ f *under right equivalence*. More precisely, let

$$f_t(x) = f(\phi(x, t)),$$

with $\phi(0, t) = 0$. Compute all possible tangent vectors $(d/dt)f_t(x)|_{t=0}$. *Answer*: $RT_r(f) = \mathscr{M} \cdot J(f)$ where $J(f) = \langle \partial f/\partial x_1, \ldots, \partial f/\partial x_n \rangle$. (*Note*: $J(f)$ is called the *Jacobian ideal* of f.)

11.2. Prove that if $p(x)$ is in $RT_r(f)$ and

$$RT_r(f + tp) = RT_r(f) \quad \text{for all } t,$$

then $f + p$ is right equivalent to f. *Hint*: Mimic the proof of Theorem 2.2.

11.3. The germ f in $\mathscr{E}_n$ is called *k-determined* with respect to right equivalence if $f + p$ is right equivalent to f for every $p \in \mathscr{M}^{k+1}$. Using Exercises 11.1 and 11.2, prove that if

$$\mathscr{M}^{k+1} \subset \mathscr{M}^2 \cdot J(f),$$

then f is k-determined with respect to right equivalence.

11.4. A germ f in $\mathscr{E}_n$ has a *singularity* at 0 if $\partial f/\partial x_i(0) = 0$ for $1 \leq i \leq n$. This singularity is *nondegenerate* if the hessian matrix

$$(d^2f)_0 = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(0)\right)$$

is nonsingular. Using Exercise 11.3, show that if f has a nondegenerate singularity at the origin then f is 2-determined with respect to right equivalence. (This is the classical Morse lemma.)

11.5. Show that $x^3 + xy^2$ (the *elliptic umbilic*) and $x^3 - xy^2$ (the *hyperbolic umbilic*) are 3-determined with respect to right equivalence.

§12. Proof of Proposition 8.6(b)

Proposition 8.6(b) states that for a normal form h of finite codimension the higher-order terms $\mathscr{P}(h)$ constitute an intrinsic ideal of finite codimension. Recall that $\mathscr{P}(h)$ is defined to be the set of those germs $\mathscr{P}$ for which $RT(g + tp) = RT(g)$ for all t in $\mathbb{R}$ and all g which are strongly equivalent to h. We subdivide our proof of Proposition 8.6(b) into three parts. We first prove that $\mathscr{P}(h)$ is an ideal, then that $\mathscr{P}(h)$ is intrinsic, and finally that $\mathscr{P}(h)$ has finite codimension.

Lemma 12.1. *$\mathscr{P}(h)$ is an ideal.*

Proof. To prove that $\mathscr{P}(h)$ is an ideal we must verify two points. First, if p_1 and p_2 are in $\mathscr{P}(h)$ then so is $p_1 + p_2$, and second, if $p \in \mathscr{P}(h)$ and $f \in \mathscr{E}_{x,\lambda}$ then $fp \in \mathscr{P}(h)$.

Suppose p_1, $p_2 \in \mathscr{P}(h)$ and that g is strongly equivalent to h. We must compute $RT(g + t(p_1 + p_2))$. Since p_1 in $\mathscr{P}(h)$ we know that $\tilde{g} = g + tp_1$ is strongly equivalent to g and hence to h. (Cf. Proposition 8.6(a).) Since $p_2 \in \mathscr{P}(h)$ it follows that $RT(\tilde{g} + sp_2) = RT(\tilde{g})$ for all s. Setting $s = t$ implies $RT(g + t(p_1 + p_2)) = RT(g + tp_1) = RT(g)$, the last equality following from the fact that $p_1 \in \mathscr{P}(h)$. Thus $p_1 + p_2$ is in $\mathscr{P}(h)$.

Next, we show that $fp \in \mathscr{P}(h)$ whenever $p \in \mathscr{P}(h)$ and $f \in \mathscr{E}_{x,\lambda}$. We do this in two parts. First we use Taylor's theorem to write f as

$$f(x, \lambda) = s + k(s, \lambda),$$

where $s = f(0, 0)$ and $k(0, 0) = 0$. Then, using the addition property for $\mathscr{P}(h)$ proved above, we may show that $fp \in \mathscr{P}(h)$ by proving that $sp \in \mathscr{P}(h)$ and $kp \in \mathscr{P}(h)$ separately. The first step is easy. Observe that

$$RT(g + t(sp)) = RT(g + (ts)p) = RT(g)$$

since $p \in \mathscr{P}(h)$. Thus $sp \in \mathscr{P}(h)$.

We shall prove that $kp \in \mathscr{P}(h)$ by use of Nakayama's lemma. Since $RT(g + p) = RT(g)$, it follows that the generators of $RT(g + p)$ are also in $RT(g)$. This implies, in particular, that

$$p, xp_x \quad \text{and} \quad \lambda p_x \in RT(g). \tag{12.1}$$

A direct calculation using the fact that $k \in \mathscr{M}$ shows that

$$kp, x(kp)_x \quad \text{and} \quad \lambda(kp)_x \in \mathscr{M}RT(g).$$

It follows by Corollary 5.4(a) that

$$\begin{aligned} RT(g + tkp) &= \langle g + tkp, xg_x + tx(kp)_x, \lambda g_x + t\lambda(kp)_\lambda \rangle \\ &= \langle g, xg_x, \lambda g_x \rangle \\ &= RT(g). \end{aligned}$$

Thus $kp \in \mathscr{P}(h)$. □

In the second step of the proof of Proposition 8.6(b) we show that the ideal $\mathscr{P}(h)$ is intrinsic. This proof proceeds most smoothly after the introduction of some notation and the proof of a preliminary lemma. The question we address is: How are $RT(g)$ and $RT(h)$ related when g and h are equivalent? The answer is given in Lemma 12.2 below.

Consider the change of coordinates

$$\Phi(x, \lambda) = (X(x, \lambda), \Lambda(\lambda)),$$

where $\Phi(0, 0) = (0, 0)$. Define the *pull-back* mapping $\Phi^*: \mathscr{E}_{x,\lambda} \to \mathscr{E}_{x,\lambda}$ by

$$\Phi^*(g)(x, \lambda) \equiv g(\Phi(x, \lambda)). \tag{12.2}$$

The map Φ^* has several useful properties; namely

$$\begin{aligned} &\text{(a)} \quad \Phi^*(g + h) = \Phi^*(g) + \Phi^*(h), \\ &\text{(b)} \quad \Phi^*(g \cdot h) = \Phi^*(g) \cdot \Phi^*(h). \end{aligned} \tag{12.3}$$

In words, (12.3) states that Φ^* is a ring homomorphism.

We are interested in invertible changes of coordinate, so that Φ is a local diffeomorphism. This means that Φ^* is invertible; in fact

$$(\Phi^*)^{-1} = (\Phi^{-1})^*. \tag{12.4}$$

If $\mathscr{I}$ is a vector subspace of $\mathscr{E}_{x,\lambda}$, let $\Phi^*(\mathscr{I})$ denote the vector space of all germs of the form $\Phi^*(g)$ for g in $\mathscr{I}$. Let us show that if $\mathscr{I}$ is an ideal, so is $\Phi^*(\mathscr{I})$. By (12.3a), $\Phi^*(\mathscr{I})$ is closed under sums. Suppose $g \in \mathscr{I}$ and $f \in \mathscr{E}_{x,\lambda}$. Then $(\Phi^{-1})^*(f) \cdot g \in \mathscr{I}$. Applying (12.4), we deduce that $f \cdot \Phi^*(g) \in \Phi^*(\mathscr{I})$. Thus $\Phi^*(\mathscr{I})$ is an ideal. In particular, if $\mathscr{I} = \langle p_1, \ldots, p_k \rangle$ is a finitely generated ideal, then $\Phi^*(\mathscr{I})$ is the ideal $\langle \Phi^*(p_1), \ldots, \Phi^*(p_k) \rangle$.

An equivalence transformation γ consists of a diffeomorphism $\Phi(x, \lambda) = (X(x, \lambda), \Lambda(\lambda))$ as above plus a pre-multiplying function $S(x, \lambda)$. We can think of γ as a mapping

$$\gamma\colon \mathscr{E}_{x,\lambda} \to \mathscr{E}_{x,\lambda}$$

defined by

$$\gamma(h) = S(x, \lambda)h(\Phi(x, \lambda)).$$

We make two observations about the mapping γ. First, γ is invertible, and γ^{-1} is also an equivalence. Explicitly

$$\gamma^{-1}(g) = \frac{1}{S(\Phi^{-1}(x, \lambda))} g(\Phi^{-1}(x, \lambda)).$$

Second,

$$\gamma(\mathscr{I}) = \Phi^*(\mathscr{I}),$$

whenever $\mathscr{I} \subset \mathscr{E}_{x,\lambda}$ is an ideal. This property holds even though γ is not a ring homomorphism—the analogue of (12.3b) fails because

$$\gamma(g \cdot h) = S^{-1}\gamma(g) \cdot \gamma(h).$$

We denote the identity equivalence ($S = 1$, $X = x$, $\Lambda = \lambda$) by 1.

Lemma 12.2. *Let g and h be equivalent, where the equivalence γ is given by*

$$g(x, \lambda) = S(x, \lambda)h(\Phi(x, \lambda)) = \gamma(h), \tag{12.5}$$

and $\Phi(x, \lambda) = (X(x, \lambda), \Lambda(\lambda))$. Then

$$RT(g) = \gamma(RT(h)) = \Phi^*(RT(h)). \tag{12.6}$$

Remark. This lemma may be proved either algebraically (by verifying (12.6) directly with the aid of Nakayama's lemma) or geometrically (using the fact that $RT(g)$ is a tangent space). We prefer the latter, as that proof will be useful in other contexts. A sketch of the algebraic proof is given in Exercise 12.5.

Proof of lemma 12.2. We define a smooth curve of strong equivalences δ_t to be a pair $S(x, \lambda, t)$, $X(x, \lambda, t)$, both of which depend smoothly on t. In addition, we demand that $\delta_0 = 1$. The restricted tangent space $RT(h)$ may be defined abstractly using curves of strong equivalences as follows: p is in $RT(h)$ if and only if there is a curve of strong equivalences δ_t (with $\delta_0 = 1$) such that

$$p = \frac{d}{dt}\delta_t(h)\bigg|_{t=0}. \tag{12.7}$$

We use this representation to verify (12.6). First let us show that

$$\gamma(RT(h)) \subset RT(g). \tag{12.8}$$

Suppose $p \in RT(h)$. Then from (12.7)

$$\gamma(p) = \left(\gamma \frac{d}{dt} \delta_t(h) \Big|_{t=0}\right). \tag{12.9}$$

However γ is independent of t and may be brought through the differentiation. Thus we have

$$\gamma(p) = \frac{d}{dt} \gamma\delta_t(h) \Big|_{t=0} = \frac{d}{dt} \gamma\delta_t\gamma^{-1}\, \gamma(h) \Big|_{t=0}. \tag{12.10}$$

But $\gamma(h) = g$; let us define $\tilde{\delta}_t = \gamma\delta_t\gamma^{-1}$. Rewriting (12.10), we see that

$$\gamma(p) = \frac{d}{dt} \tilde{\delta}_t(g) \Big|_{t=0}.$$

Now $\tilde{\delta}_t$ is itself a smooth curve of equivalences, so by (12.7), $\gamma(p) \in RT(g)$.
Similarly, by interchanging the roles of g and h we may show that

$$\gamma^{-1}(RT(g)) \subset RT(h). \tag{12.11}$$

The lemma follows from combining (12.8) and (12.11). □

We now complete the second part of the proof of Proposition 8.6(b).

Lemma 12.3. *The ideal $\mathscr{P}(h)$ is intrinsic.*

PROOF. Let p be in $\mathscr{P}(h)$ and let $\gamma: \mathscr{E}_{x,\lambda} \to \mathscr{E}_{x,\lambda}$ be a strong equivalence. We must show that $\gamma(p) \in \mathscr{P}(h)$. Suppose g is strongly equivalent to h. We have

$$\begin{aligned} RT(g + t\gamma(p)) &= RT(\gamma(\gamma^{-1}(g) + tp)) \\ &= \gamma RT(\gamma^{-1}(g) + tp), \end{aligned} \tag{12.12}$$

the second equality in (12.12) following from Lemma 12.2. Now $\gamma^{-1}(g)$ is strongly equivalent to g and hence to h. Since $p \in \mathscr{P}(h)$

$$RT(\gamma^{-1}(g) + tp) = RT(\gamma^{-1}(g)).$$

Combining with (12.12) we see that

$$RT(g + t\gamma(p)) = \gamma RT(\gamma^{-1}(g)) = RT(g), \tag{12.13}$$

the second equality in (12.13) following from Lemma 12.2. Equation (12.13) implies that $\gamma(p) \in \mathscr{P}(h)$, as desired. □

In the final part of the proof of Proposition 8.6(b) we show that if $RT(h)$ has finite codimension then $\mathscr{P}(h)$ is an intrinsic ideal of finite codimension. More precisely, we prove:

Lemma 12.4. Itr $\mathscr{M} \cdot RT(h) \subset \mathscr{P}(h)$ *if* codim $RT(h) < \infty$.

We claim that Lemma 12.4 completes the proof of Proposition 8.6(b). If $RT(h)$ has finite codimension, then $\mathcal{M}^k \subset RT(h)$ for some k. (Cf. Proposition 5.7.) Thus $\mathcal{M}^{k+1} \subset \text{Itr } \mathcal{M} \cdot RT(h) \subset \mathcal{P}(h)$ by Lemma 12.4, and hence $\mathcal{P}(h)$ has finite codimension.

Before proving Lemma 12.4 we state and prove the next lemma.

Lemma 12.5. *Let $\mathcal{I}$ be an intrinsic ideal. Assume that $RT(h + p) = RT(h)$ for all p in $\mathcal{I}$. Then $\mathcal{I} \subset \mathcal{P}(h)$.*

Remark. The point of lemma 12.5 is that if $\mathcal{I}$ is known to be an intrinsic ideal, we do not have to compute $RT(g + tp)$ for all g strongly equivalent to h.

PROOF. Let p be in $\mathcal{I}$, let t be in $\mathbb{R}$, and let g be strongly equivalent to h. We must show that $RT(g + tp) = RT(g)$.

Let γ be the strong equivalence satisfying $\gamma(h) = g$. Then

$$\begin{aligned} RT(g + tp) &= RT(\gamma(h + \gamma^{-1}(tp)) \\ &= \gamma RT(h + \gamma^{-1}(tp)), \end{aligned} \tag{12.14}$$

the second equality in (12.14) following from Lemma 12.2.

Observe that $tp \in \mathcal{I}$ since $\mathcal{I}$ is an ideal, and that $\gamma^{-1}(tp) \in \mathcal{I}$ since $\mathcal{I}$ is intrinsic. Thus

$$RT(h + \gamma^{-1}(tp)) = RT(h).$$

Combining with (12.14), and using Lemma 12.2, we see that

$$RT(g + tp) = RT(g).$$

□

PROOF OF LEMMA 12.4. Let $\mathcal{I} = \text{Itr } \mathcal{M} \cdot RT(h)$. By Lemma 12.5, to prove $\mathcal{I} \subset \mathcal{P}(h)$ it suffices to show for each $p \in \mathcal{I}$ that $RT(h + p) = RT(h)$. We do this by Nakayama's lemma in the form of Corollary 5.4(a).

Let p be in $\mathcal{I}$. Since $\mathcal{I}$ is intrinsic and of finite codimension, Lemma 7.5 implies that xp_x and $\lambda p_x \in \mathcal{I}$. Since $\mathcal{I} \subset \mathcal{M} \cdot RT(h)$, we see that p, xp_x, and $\lambda p_x \in \mathcal{M} \cdot RT(h)$. We now use Corollary 5.4(a) to conclude that

$$\begin{aligned} RT(h + p) &= \langle h + p, xh_x + xp_x, \lambda h_x + \lambda p_x \rangle = \langle h, xh_x, \lambda h_x \rangle \\ &= RT(h). \end{aligned}$$

□

EXERCISES

12.1. Prove: If $\mathcal{I}$ is an ideal and $\Phi(x, \lambda) = (X(x, \lambda), \Lambda(\lambda))$ is an invertible coordinate transformation, then

$$\text{Itr } \mathcal{I} = \text{Itr } \Phi^*(\mathcal{I}).$$

12.2. Prove that if $g, h \in \mathcal{E}_{x,\lambda}$ are equivalent then $\operatorname{codim} RT(g) = \operatorname{codim} RT(h)$.

12.3. Let $\mathscr{I} \subset \mathscr{E}_{x,\lambda}$ be an ideal. Prove that

$$\operatorname{Itr} \mathscr{I} = \bigcap_{\Phi} \Phi^*(\mathscr{I}),$$

where the intersection is taken over all diffeomorphisms Φ.

12.4. Let $h \in \mathscr{E}_{x,\lambda}$. Use Exercise 12.3 to prove

$$\operatorname{Itr} RT(h) = \bigcap_{g} \{RT(g)\colon g \sim h\},$$

where $\sim$ indicates strong equivalence.

12.5. Complete the algebraic proof of Lemma 12.2. (*Hint*: Show that

$$\Phi^*(RT(h)) = \langle \Phi^*h, \Phi^*(xh_x), \Phi^*(\lambda h_x)\rangle = \langle \Phi^*h, x(\Phi^*h)_x, \lambda(\Phi^*h)_x\rangle.)$$

§13. Proof of Theorem 8.7

We divide the proof of Theorem 8.7 into two parts:

$$\text{(a)}\quad \operatorname{Itr} \mathscr{J}(h) \subset \mathscr{P}(h),$$
$$\text{(b)}\quad \mathscr{P}(h) \subset \operatorname{Itr} \mathscr{J}(h).$$

The complete proof of part (b) is much more technical than the proof of part (a). Moreover, only part (a) will be used in applications. Part (b) gives a more elegant "if and only if" result, but its proof may be omitted without loss of continuity.

(a) Proof that $\operatorname{Itr} \mathscr{J}(h) \subset \mathscr{P}(h)$

The ideas needed for part (a) have already been introduced. In Lemma 12.4 we proved that $\mathscr{M} \cdot RT(h) \subset \mathscr{P}(h)$. Recall that $\mathscr{M} \cdot RT(h) = \langle xh, \lambda h, x^2h_x, \lambda xh_x, \lambda^2 h_x\rangle$ while $\mathscr{J}(h) = \langle xh, \lambda h, x^2h_x, \lambda h_x\rangle$. The only new issue here is how to deal with the term λh_x which belongs to $\mathscr{J}(h)$ but not $\mathscr{M} \cdot RT(h)$. Because of the following lemma, the same techniques in fact suffice for this term, too.

Lemma 13.1. *Suppose $h \in \mathscr{M}$ has a singularity of finite codimension at the origin. If $p \in \operatorname{Itr} \mathscr{J}(h)$, then $\lambda p_x \in \mathscr{M} \cdot RT(h)$.*

Remark. Note that the operator $\lambda(\partial/\partial x)$ is nilpotent on $\mathscr{E}_{x,\lambda}/\mathscr{M}^k$ for any k—this operator preserves the degree of homogeneity of any monomial but substitutes a power of λ for one of x. We do not use this fact in any way in our proof, but it provides motivation for why λp_x might be different from xp_x.

PROOF. First we dispense with the case where $h_\lambda(0, 0) \neq 0$. (We assume that $h_x(0, 0) = 0$—otherwise h would not have a singularity.) We claim that

$$\mathscr{J}(h) = \mathscr{M} \cdot RT(h) \quad \text{if} \quad h_\lambda(0, 0) \neq 0, \tag{13.1}$$

so that the lemma follows trivially in this case. Since $RT(h)$ has finite codimension, h is strongly equivalent to $g = \pm x^k \pm \lambda$ where $k \geq 2$. We compute that

$$\mathscr{M} \cdot RT(g) = \mathscr{M}^{k+1} + \mathscr{M}\langle\lambda\rangle.$$

Denote the strong equivalence between g and h by γ; that is, $h = \gamma(g)$. Using Lemma 12.2 and that fact $\mathscr{M}$ and $\mathscr{M} \cdot RT(g)$ are intrinsic we have

$$\mathscr{M} \cdot RT(h) = \mathscr{M} \cdot \gamma(RT(g)) = \gamma(\mathscr{M} \cdot RT(g)) = \mathscr{M} \cdot RT(g).$$

As noted above, $h_x(0, 0) = 0$. Hence $\lambda h_x \in \mathscr{M}\langle\lambda\rangle$, so $\mathscr{J}(h) = \mathscr{M} \cdot RT(h)$, as claimed.

We now prove Lemma 13.1 assuming that $h_\lambda(0, 0) = 0$. We claim that if $p \in \operatorname{Itr} \mathscr{J}(h)$, then xp_x and λp_x also belong to $\operatorname{Itr} \mathscr{J}(h)$. To see this, note that $RT(h)$ has finite codimension, so that $\mathscr{J}(h)$ also has finite codimension. Thus $\operatorname{Itr} \mathscr{J}(h)$ is an intrinsic ideal of finite codimension. Therefore the claim follows from Lemma 7.5.

Since we have

$$xp_x, \lambda p_x \in \operatorname{Itr} \mathscr{J}(h) \subset \mathscr{J}(h) = \langle xh, \lambda h, x^2 h_x, \lambda h_x\rangle,$$

there exist smooth coefficients $\alpha, \beta, \gamma, \delta$ such that

$$\begin{aligned} &\text{(a)} \quad xp_x = \alpha h + \beta h_x, \\ &\text{(b)} \quad \lambda p_x = \gamma h + \delta h_x, \end{aligned} \tag{13.2}$$

where

$$\begin{aligned} &\text{(a)} \quad \alpha(0, 0) = \beta(0, 0) = \beta_x(0, 0) = 0, \\ &\text{(b)} \quad \gamma(0, 0) = \delta(0, 0) = \delta_x(0, 0) = 0. \end{aligned}$$

We will prove further that

$$\delta_\lambda(0, 0) = 0. \tag{13.3}$$

It may then be seen from (13.2b) that $\lambda p_x \in \mathscr{M} \cdot RT(h)$.

To prove (13.3) we multiply (13.2a) by λ, (13.2b) by x, and subtract, obtaining

$$(\lambda\alpha - x\gamma)h + (\lambda\beta - x\delta)h_x = 0.$$

However, h and h_x generate an ideal of finite codimension—in particular, $RT(h) \subset \langle h, h_x\rangle$. It follows from Proposition 5.10 that there exists a smooth germ Q such that

$$\begin{aligned} &\text{(a)} \quad \lambda\alpha - x\gamma \equiv -Qh_x \quad \operatorname{mod} \mathscr{M}^k, \\ &\text{(b)} \quad \lambda\beta - x\delta \equiv Qh \qquad \operatorname{mod} \mathscr{M}^k, \end{aligned} \tag{13.4}$$

where k is as large as we like. We take the mixed partial derivative with respect to x and λ of (13.4b) and evaluate at the origin. This yields

$$\delta_\lambda(0, 0) = -Q(0, 0)h_{x\lambda}(0, 0),$$

where we have used the relations

$$\beta_x(0, 0) = h(0, 0) = h_x(0, 0) = h_\lambda(0, 0) = 0.$$

On differentiating (13.4a) with respect to λ, evaluating at the origin and eliminating terms which vanish, we see that

$$Q(0, 0)h_{x\lambda}(0, 0) = 0,$$

from which (13.3) follows. The proof of Lemma 13.1 is complete. □

PROOF THAT Itr $\mathscr{J}(h) \subset \mathscr{P}(h)$. Our proof is based on an application of Lemma 12.5. The main step is to show that for any $p \in \text{Itr } \mathscr{J}(h)$,

$$RT(h + p) = RT(h). \tag{13.5}$$

It then follows from the lemma that Itr $\mathscr{J}(h) \subset \mathscr{P}(h)$. (*Remark*: The idea we use to prove (13.5) already occurred in Lemma 11.2.)

If $p \in \text{Itr } \mathscr{J}(h)$, then by Lemma 7.5

$$xp_x, \lambda p_x \in \text{Itr } \mathscr{J}(h) \subset \mathscr{J}(h).$$

Thus there exist smooth coefficients such that

$$\begin{aligned} p &= A_1 h + B_1 x h_x + C_1 \lambda h_x, \\ xp_x &= A_2 h + B_2 x h_x + C_2 \lambda h_x, \\ \lambda p_x &= A_3 h + B_3 x h_x + C_3 \lambda h_x, \end{aligned} \tag{13.6}$$

where

$$A_i(0, 0) = B_i(0, 0) = 0, \qquad i = 1, 2, 3.$$

Moreover by Lemma 13.1, $C_3(0, 0) = 0$. We may write (13.6) in a matrix notation

$$\begin{pmatrix} p \\ xp_x \\ \lambda p_x \end{pmatrix} = Q \begin{pmatrix} h \\ xh_x \\ \lambda h_x \end{pmatrix}, \tag{13.7}$$

where $Q(0, 0)$ is strictly upper triangular; i.e., upper triangular with zeros along the diagonal. Adding h to both sides of (13.7) we have

$$\begin{pmatrix} h + p \\ x(h + p)_x \\ \lambda(h + p)_x \end{pmatrix} = (I + Q) \begin{pmatrix} h \\ xh_x \\ \lambda h_x \end{pmatrix}.$$

But $I + Q$ is invertible in some neighborhood of the origin, since $Q(0, 0)$ is upper triangular. Thus the generators of $RT(h)$ and $RT(h + p)$ are related by an invertible linear transformation; in other words, (13.5) follows by Lemma 4.2. □

(b) Proof that $\mathscr{P}(h) \subset \operatorname{Itr} \mathscr{J}(h)$

The idea behind our proof is the calculation in §2 with which we motivated the definition of $RT(h)$. (This calculation occurs several times above, but we repeat it here.) Suppose $p \in \mathscr{P}(h)$. By Proposition 8.6(a), for any t, $h + tp$ is strongly equivalent to h. Moreover, the equivalence transformation varies smoothly with t and equals the identity when $t = 0$. Thus we have

$$h(x, \lambda) + tp(x, \lambda) = S(x, \lambda, t)h(X(x, \lambda, t), \lambda), \tag{13.8}$$

where

$$X(0, 0, t) = 0. \tag{13.9}$$

On differentiating (13.8) with respect to t and setting $t = 0$ we find

$$p(x, \lambda) = \dot{S}(x, \lambda, 0)h(x, \lambda) + h_x(x, \lambda)\dot{X}_x(x, \lambda, 0). \tag{13.10}$$

The crux of the present proof is to show that

$$\dot{S}(0, 0, 0) = \dot{X}_x(0, 0, 0) = 0. \tag{13.11}$$

It will then follow from (13.10) that $p \in \mathscr{J}(h)$. In other words, verifying (13.11) will show that $\mathscr{P}(h) \subset \mathscr{J}(h)$; since $\mathscr{P}(h)$ is intrinsic, this will show that $\mathscr{P}(h) \subset \operatorname{Itr} \mathscr{J}(h)$, as desired.

In verifying (13.11) we shall in fact prove that

$$S(0, 0, t) \equiv 1, \qquad X_x(0, 0, t) \equiv 1. \tag{13.12}$$

(Henceforth we shall suppress the dependence of S and X on t.) The intuition behind our analysis is as follows. Equation (13.8) states that the equivalence transformation (S, X) applied to h may change the higher-order terms (represented by p), but *only higher-order terms are affected.* Our strategy is to isolate two "lower-order terms" in h and to extract (13.12) from the fact that S, X does not change these lower-order terms. One of these terms is easy to identify; the following simple lemma is useful in this task.

Lemma 13.2. (a) $\mathscr{P}(h) \subset \mathscr{S}(h)$.
(b) *The intrinsic generators of $\mathscr{S}(h)$ do not belong to $\mathscr{P}(h)$.*

Proof. We already know that

$$\mathscr{P}(h) \subset RT(h) \subset \mathscr{S}(h),$$

which verifies part (a). For part (b), suppose $x^k\lambda^l$ is an intrinsic generator of $\mathscr{S}(h)$ which also belongs to $\mathscr{P}(h)$. By Proposition 8.6(a), $h + tx^k\lambda^l$ is equivalent to h for any $t \in \mathbb{R}$. Yet there is a choice of t which makes the derivative

$$\left(\frac{\partial}{\partial x}\right)^k \left(\frac{\partial}{\partial \lambda}\right)^l (h + tx^k\lambda^l)$$

vanish. By Theorem 8.4, $h + tp$ for this choice of t is not equivalent to h. This contradiction proves the lemma. □

We identify the first of the "low-order terms" in h as follows. Since h has finite codimension, there is an integer k such that

$$h(x, 0) \equiv ax^k \mod \mathscr{M}^{k+1},$$

where $a \neq 0$. Then x^k is an intrinsic generator of $\mathscr{S}(h)$. However, according to Lemma 13.2, $x^k \notin \mathscr{P}(h)$. In other words, if $p \in \mathscr{P}(h)$, then $p(x, 0) \in \mathscr{M}^{k+1}$. Let us compute the coefficient of x^k on the left and the right in (13.8). On the left we have

$$\text{LHS}(x, 0) \equiv ax^k \mod \mathscr{M}^{k+1},$$

while

$$\text{RHS}(x, 0) \equiv aS(x, 0)X^k(x, 0) \mod \mathscr{M}^{k+1}.$$

But

$$X(x, 0) \equiv X_x(0, 0)x \mod \mathscr{M}^2,$$

so that $X^{k+1} \in \mathscr{M}^{k+1}$ and

$$X^k(x, 0) \equiv X_x^k(0, 0)x^k \mod \mathscr{M}^{k+1}.$$

Therefore matching coefficients of x^k in (13.8) yields the relation

$$S(0, 0)X_x^k(0, 0) = 1. \tag{13.13}$$

We need to identify a second lower-order term in h and to extract a relation analogous to (13.13). This requires very little effort if $\mathscr{S}(h)$ has at least two intrinsic generators; i.e., if the decomposition

$$\mathscr{S}(h) = \mathscr{M}^k + \mathscr{M}^{k_1}\langle\lambda\rangle^{l_1} + \cdots + \mathscr{M}^{k_s}\langle\lambda\rangle^{l_s} \tag{13.14}$$

according to Proposition 7.1 contains at least two distinct terms. Thus at this juncture we split the proof into two cases.

Case I. In (13.14), $s > 0$.

Case II. $\mathscr{S}(h) = \mathscr{M}^k$.

Case I. Let $x^{k_1}\lambda^{l_1}$ be an intrinsic generator of $\mathscr{S}(h)$ as indicated in (13.14). By matching the coefficients of $x^{k_1}\lambda^{l_1}$ in (13.8) we shall show that

$$S(0, 0)X_x^{k_1}(0, 0) = 1. \tag{13.15}$$

Since $k_1 \neq k$, (13.12) then follows from (13.13) and (13.15).

In deriving (13.15) we introduce an appropriate notion of higher-order terms. Let $\mathscr{H}$ be the ideal

$$\mathscr{H} = \mathscr{M}^{k_1+l_1+1} + \langle\lambda\rangle^{l_1+1}; \tag{13.16}$$

this is the largest intrinsic ideal that does not contain $x^{k_1}\lambda^{l_1}$. In particular, the other intrinsic generators of $\mathscr{S}(h)$ all belong to $\mathscr{H}$, and we may deduce from Lemma 13.2 that $\mathscr{P}(h) \subset \mathscr{H}$. Thus we have in (13.8)

$$\text{LHS} \equiv bx^{k_1}\lambda^{l_1} \pmod{\mathscr{H}},$$

where $b \neq 0$. Moreover, since $\mathscr{H}$ is intrinsic we also have

$$\text{RHS} \equiv bSX^{k_1}\lambda^{l_1} \pmod{\mathscr{H}},$$

and

$$X^{k_1}\lambda^{l_1} \equiv X_x^{k_1}(0, 0)x^{k_1}\lambda^{l_1} \pmod{\mathscr{H}}.$$

Therefore (13.15) follows by matching the coefficients of $x^{k_1}\lambda^{l_1}$ in (13.8). This completes the analysis of Case I.

Case II. In the present case where $\mathscr{S}(h) = \mathscr{M}^k$, it does not seem to be possible to prove that $\mathscr{P}(h) \subset \text{Itr } \mathscr{J}(h)$ by working with h directly; rather we work with a carefully constructed germ g that is equivalent to h. Specifically, we will show for this g that $\mathscr{P}(g) \subset \text{Itr } \mathscr{J}(g)$. To obtain the desired conclusion that $\mathscr{P}(h) \subset \text{Itr } \mathscr{J}(h)$, we need to know that $\mathscr{P}(h) = \mathscr{P}(g)$ and

$$\text{Itr } \mathscr{J}(h) = \text{Itr } \mathscr{J}(g).$$

The first equality is obvious from the definition of $\mathscr{P}$. The second equality we state here as a lemma; the proof is given at the end of this section.

Lemma 13.3. *If $g, h \in \mathscr{E}_{x,\lambda}$ are equivalent, then*

$$\text{Itr } \mathscr{J}(g) = \text{Itr } \mathscr{J}(h).$$

We now show how to construct g from h, and then we prove that

$$\mathscr{P}(g) \subset \text{Itr } \mathscr{J}(g)$$

for this g.

Order the monomials in $\mathscr{M}^k$ by

$$x^k, x^{k-1}\lambda, \ldots, \lambda^k, x^{k+1}, x^k\lambda, \ldots, \lambda^{k+1}, x^{k+2}, \ldots .$$

If g is any germ equivalent to h, then we may write

$$g = ax^k + bx^{k_1}\lambda^{l_1} + \cdots, \tag{13.17}$$

where $a \neq 0, b \neq 0$, and $\cdots$ refers to later terms relative to the above ordering of terms. Let us justify (13.17). Since $\mathscr{S}(g) = \mathscr{S}(h) = \mathscr{M}^k$, all monomials appearing in the Taylor series of g must be of degree at least k, and the coefficient of x^k cannot vanish. Moreover, there must be at least one more monomial with a nonvanishing coefficient; otherwise g, and hence h, would have infinite codimension. Indeed we may refine the argument to obtain an upper bound on $k_1 + l_1$ as follows. Suppose $\mathscr{M}^K \subset \mathscr{M} \cdot \mathrm{RT}(h)$. Then $\mathscr{M}^K \subset \mathrm{Itr}\, \mathscr{M} \cdot \mathrm{RT}(h)$, so by Lemma 12.4

$$\mathscr{M}^K \subset \mathscr{P}(h) = \mathscr{P}(g).$$

Now if g were equal to $ax^k + p$ where $p \in \mathscr{M}^K$, then g would be equivalent to ax^k, contradicting the hypothesis of finite codimension. In other words in (13.17) we must have $k_1 + l_1 < K$. The existence of this *a priori* bound on $k_1 + l_1$ makes the following construction possible: among all g's equivalent to h, choose g so that the first nonzero term after x^k in (13.17) has maximal degree with respect to the above ordering.

The verification for the g just defined that

$$\mathscr{P}(g) \subset \mathrm{Itr}\, \mathscr{J}(g) \tag{13.18}$$

proceeds in much the same way as the calculations for Case I. More precisely, if $p \in \mathscr{P}(g)$, differentiate the relation (13.8) expressing the equivalence of $g + tp$ with g. Our task is to derive (13.11). In fact we prove that (13.12) is valid.

We will verify (13.12) by matching low-order terms in (13.8), replacing h by g. By low-order terms we mean terms not belonging to the intrinsic ideal $\mathscr{H}$ defined by (13.16). For reference below we note that

$$x^\mu \lambda^\nu \notin \mathscr{H} \quad \text{iff} \quad \mu + \nu < k_1 + l_1 + 1 \quad \text{and} \quad \nu < l_1 + 1. \tag{13.19}$$

We claim that $\mathscr{P}(g) \subset \mathscr{H}$. We prove this by showing that

$$x^{k_1}\lambda^{l_1} \notin \mathscr{P}(g) \tag{13.20}$$

and recalling that $\mathscr{H}$ is the largest intrinsic ideal not containing $x^{k_1}\lambda^{l_1}$. To prove (13.20) we argue by contradiction. If $x^{k_1}\lambda^{l_1} \in \mathscr{P}(g)$, then $g - bx^{k_1}\lambda^{l_1}$ would be equivalent to g, and its first nonzero term after x^k would occur further along in the ordering than $x^{k_1}\lambda^{l_1}$. This contradicts the construction of g, thereby proving (13.20).

Let us begin to match low-order terms in (13.8). Since $p \in \mathscr{P}(g) \subset \mathscr{H}$ we have

$$\mathrm{LHS} \equiv ax^k + bx^{k_1}\lambda^{l_1} \mod \mathscr{H}; \tag{13.21}$$

because $\mathscr{H}$ is intrinsic, $x^\mu \lambda \in \mathscr{H}$ iff $X^\mu \lambda^\nu \in \mathscr{H}$, so

$$\mathrm{RHS} \equiv aSX^k + bSX^{k_1}\lambda^{l_1} \mod \mathscr{H}. \tag{13.22}$$

It is easy to match the coefficients of x^k by restricting to $\lambda = 0$ and computing modulo $\mathscr{M}^{k+1}$. Indeed this argument was carried out above; it yields the conclusion (13.13); i.e.,

$$S(0, 0)X_x^k(0, 0) = 1. \tag{13.23}$$

It is tempting to differentiate (13.21) and (13.22) l_1 times with respect to λ and match coefficients of x^{k_1}. It seems this would lead to the relation

$$S(0, 0)X_x^{k_1}(0, 0) = 1, \tag{13.24}$$

which could be combined with (13.23) to yield (13.12). In fact, (13.24) is a valid equation and the proof of Theorem 8.7 does indeed emerge from these considerations as indicated, but the justification of (13.24) is considerably more subtle. The difficulty is that the naive argument above overlooks the possibility that X^k might contribute to the coefficient of $x^{k_1}\lambda^{l_1}$. For example, if $k_1 + l_1 = k$ (i.e., if both terms in (13.21) have the same degree) and if $X(x, \lambda) = x + c\lambda$, then the expansion of X^k includes a term in $x^{k_1}\lambda^{l_1}$. As it turns out, this possibility does not actually occur, for the following reason: for *all* monomials not belonging to $\mathscr{H}$, the coefficients in (13.21) and (13.22) must match, not just for x^k and $x^{k_1}\lambda^{l_1}$. However this must be shown; we do so in the lemma below. (Remark: It is shown in the proof of this lemma that $k_1 \neq k$ so that (13.23) and (13.24) are actually independent relations.)

Lemma 13.4. *The coefficient of* $x^{k_1}\lambda^{l_1}$ *in* (13.22) *equals*

$$bS(0, 0)X_x^{k_1}(0, 0). \tag{13.25}$$

Proof. It is clear that (13.25) represents the contribution of the second term in (13.22) to the indicated coefficient. We must show that the first term does not contribute.

First, we claim that $k_1 \leq k - 2$. For suppose otherwise; i.e., suppose $k_1 \geq k - 1$. If, in fact, $k_1 \geq k$, we may transform away the term $x^{k_1}\lambda^{l_1}$ by an equivalence transformation on the range,

$$\left\{1 - \frac{b}{a} x^{k_1 - k}\lambda^{l_1}\right\}^{-1} g \equiv ax^k \mod \mathscr{H},$$

which contradicts the construction of g. If $k_1 = k - 1$ we may transform $x^{k_1}\lambda^{l_1}$ away by an equivalence transform on the domain,

$$g\left(x - \frac{b}{ka}\lambda^{l_1}, \lambda\right) \equiv ax^k \mod \mathscr{H},$$

which is again a contradiction. Thus

$$k_1 \leq k - 2. \tag{13.26}$$

In the proof we will need the following assertion: Let $\phi_1, \phi_2 \in \mathscr{M}$ be two germs such that $\phi_1 \equiv \phi_2 \bmod \mathscr{M}^\mu$ where $\mu \geq 1$; then

$$\phi_1^k \equiv \phi_2^k \mod \mathscr{M}^{\mu+k-1}.$$

This assertion is proved by writing $\phi_1 = \phi_2 + r$ where $r \in \mathscr{M}^\mu$ and expanding $(\phi_2 + r)^k$ by the binomial theorem. We leave this to the reader.

The main task of the proof is to show that there is a germ Y such that

$$X(x, \lambda) \equiv xY(x, \lambda) \mod \mathscr{M}^{k_1+l_1-k+2}. \tag{13.27}$$

Given (13.27), it follows from the above assertion that

$$X^k \equiv x^k Y^k \mod \mathscr{M}^{k_1+l_1+1},$$

so that in (13.22), X^k cannot contribute to the coefficient of $x^{k_1}\lambda^{l_1}$.

We prove (13.27) by induction. Suppose that

$$X \equiv xY \quad \text{mod } \mathcal{M}^j \tag{13.28}$$

for some j satisfying

$$j < k_1 + l_1 - k + 2. \tag{13.29}$$

(We may trivially start the induction with $j = 1$, since $X(0, 0) = 0$.) Modifying Y if necessary, we deduce from (13.28) that

$$X \equiv xY + c\lambda^j \quad \text{mod } \mathcal{M}^{j+1} \tag{13.30}$$

for some $c \in \mathbb{R}$. By the assertion above

$$X^k \equiv (xY + c\lambda^j)^k \quad \text{mod } \mathcal{M}^{k+j},$$

and, moreover,

$$(xY + c\lambda^j)^k \equiv (xY)^k + k(xY)^{k-1}(c\lambda^j) \quad \text{mod}\langle\lambda^{2j}\rangle$$

We may deduce that $x^{k-1}\lambda^j \notin \mathcal{H}$ by operating on (13.29), using (13.26), to obtain the criterion (13.19). Thus the coefficient of $x^{k-1}\lambda^j$ in (13.22) equals

$$aS(0, 0)X_x^k(0, 0)c.$$

Equating this to zero, the value of the corresponding coefficient in (13.21), we see that $c = 0$. Thus, comparing with (13.30) we see that the induction continues. This completes the proof of Lemma 13.4 and of Theorem 8.7. □

We end this section by proving Lemma 13.3. We begin by observing that $\mathcal{J}(h)$ is a "tangent space" under a restricted form of strong equivalence.

Definition 13.5. A $\mathcal{J}$-*equivalence* is a strong equivalence defined by a pair of functions $S(x, \lambda)$ and $X(x, \lambda)$ satisfying

$$S(0, 0) = 1, \qquad X_x(0, 0) = 1. \tag{13.31}$$

It is an easy exercise to show that if δ_t is a curve of $\mathcal{J}$-equivalences satisfying $\delta_0 = 1$ then

$$p = \frac{d}{dt}\,\delta_t(h)\bigg|_{t=0} \tag{13.32}$$

is in $\mathcal{J}(h)$. In fact $\mathcal{J}(h)$ consists of all germs defined according to (13.32). Thus $\mathcal{J}(h)$ is the tangent space to h obtained by considering $\mathcal{J}$-equivalences.

Remark. The essence of the above proof that $\mathcal{P}(h) \subset \text{Itr}\,\mathcal{J}(h)$ was to show that if $p \in \mathcal{P}(h)$ then the equivalence between $h + tp$ and h must be a $\mathcal{J}$-equivalence for each t. Differentiation with respect to t then shows that $p \in \mathcal{J}(h)$.

Lemma 13.6. *Let γ be an equivalence (not necessarily a $\mathscr{I}$-equivalence) operating on $\mathscr{E}_{x,\lambda}$ and let $\gamma(h) = g$. Then*

$$\mathscr{I}(g) = \gamma\mathscr{I}(h).$$

The proof of Lemma 13.6 is identical to the proof of Lemma 12.2, except that one uses a curve of $\mathscr{I}$-equivalences rather than a curve of strong equivalences. The details are left to the reader.

PROOF OF LEMMA 13.3. Let $g = \gamma h$. By Lemma 13.6 we have

$$\mathscr{I}(g) = \gamma(\mathscr{I}(h)).$$

Therefore

$$\text{Itr } \mathscr{I}(g) = \text{Itr } \gamma\mathscr{I}(h) = \text{Itr } \mathscr{I}(h)$$

since intrinsic parts are invariants of equivalences. □

BIBLIOGRAPHICAL COMMENTS

As we stated in the Preface, singularity theory is largely the creation of John Mather—our contribution was to adapt his work to the context of bifurcation problems. In particular, the papers of Mather relevant to the present chapter include Mather [1968], [1969b]. There are several books describing local singularity theory including Martinet [1982], Arnold [1981], Gibson [1979], and Brocker [1975]. The global theory may be found in Golubitsky and Guillemin [1973].

CHAPTER III
Unfolding Theory

§0. Introduction

As we saw in Chapter I, bifurcation diagrams may change their form dramatically when the defining equation is subjected to a small perturbation. The study of such changes is often termed "imperfect bifurcation". In this chapter, we address the general problem of imperfect bifurcation, using the theory of universal unfoldings as the main tool. The construction of universal unfoldings is now a standard procedure in singularity theory; we adapt this method to the specific context of bifurcation theory.

Our goal is to present an algorithm allowing us to enumerate, up to equivalence, all perturbations of a given bifurcation problem $g(x, \lambda) = 0$. This algorithm divides neatly into two parts, an analytic part and a geometric part. Both of these have been described, by example, in Chapter I. We recall the following two points from Chapter I, §1 for reference below:

(A) Equation (I,1.13) gives a formula for a universal unfolding of the pitchfork.
(B) Perturbations of the pitchfork are enumerated in Figure I,1.5. This includes both the diagrams shown in the open regions of Figure I,1.5 and the transition diagrams (i.e., those which occur between regions in Figure I,1.5) that are sketched in Figures I,1.6 and I,1.7.

These two points correspond to the division of this chapter into an analytic and a geometric part, as mentioned above. Before continuing, let us attempt a definition of "universal unfolding." Roughly speaking, a universal unfolding of g is a parametrized family of mappings $G(x, \lambda, \alpha)$, where α lies in a parameter space $\mathbb{R}^k$, satisfying the following two conditions:

(a) $G(x, \lambda, 0) = g(x, \lambda)$.

(b) Any sufficiently small perturbation of g is equivalent to $G(\cdot, \cdot, \alpha)$ for some α near 0. (0.1)

(*Remark*: We call any parametrized family satisfying (0.1a) an unfolding of g; this is our description for perturbations of g.)

As we have indicated above, the algorithm given in this chapter divides into two parts. These extend points (A) and (B) above to a general normal form. The first four sections of the chapter present the analytical part of the algorithm; the last six present the geometric part. The principal result from the first part is the Universal Unfolding Theorem, Theorem 2.3; the principal result from the second part is Theorem 6.1 which shows that perturbed bifurcation diagrams can be enumerated by certain open regions in parameter space. The mathematical techniques presented in these two parts are quite different. In the first part we continue to use the algebraic constructions of Chapter II; in the second part we use methods from differential topology in deforming bifurcation diagrams.

We now summarize the contents of this chapter section by section. Sections 1–4 together show how to find and work with universal unfoldings. In §1, we give precise definitions of unfoldings and universal unfoldings. In §2, we state the Universal Unfolding Theorem, which gives a necessary and sufficient condition for an unfolding to be universal. Interestingly, the necessity of this condition is found by considering only traditional one-parameter perturbations. This derivation leads to a basic concept in our theory, the tangent space to g, denoted $T(g)$. We defer the proof of sufficiency until Volume II, as it is quite technical. One important aspect of this theorem is that it characterizes the precise number of unfolding parameters needed to capture all possible perturbed behavior. This number is called the codimension of the germ g. There are several different ways of defining the codimension of g; the equivalence of these definitions is shown in Corollary 2.4. In §3, we apply the theorem to compute unfoldings for several normal forms; the guidance provided by the theorem makes the calculations elementary. In §4, we extend the universal unfolding theorem in a way that is useful for applications. Specifically, let G be an unfolding of a germ g, where g is equivalent to a normal form h; we show how to decide when G is universal. We call this the recognition problem for universal unfoldings. As with the recognition problem for a normal form in Chapter II, our solution only depends on G and finitely many of its derivatives at the origin. This is important for applications, where an explicit formula for G is often unavailable or unwieldy.

Before describing the second half of Chapter III, we make two comments about the tangent space $T(g)$ and the universal unfolding theorem.

(i) Consider a normal form h such that $RT(h)$ has finite codimension. Once $RT(h)$ has been computed, only straightforward linear algebra is required to compute $T(h)$. In fact, for any such germ h there is an integer l such

$$T(h) = RT(h) \oplus \mathbb{R}\{h_x, h_\lambda, \lambda h_\lambda, \ldots, \lambda^l h_\lambda\}.$$

See §2(b).

(ii) The universal unfolding theorem states that finding a universal unfolding of h is equivalent to finding a basis for a complementary subspace of $T(h)$ in $\mathscr{E}_{x,\lambda}$. It is a pleasant fact that the unfolding theorem is relatively easy to apply, even though it is difficult to prove. The hardest aspect of applying the theorem is determining $RT(g)$, and we dealt with this issue in Chapter II.

The second half of Chapter III deals with the geometric part of our algorithm. The main idea is to show that certain bifurcation diagrams are unchanged, up to equivalence, by small perturbations; we call such diagrams *persistent*. More precisely, we show that a diagram is persistent if its only singularities are limit points and if no two limit points have the same λ-coordinate. This result leads to an enumeration analogous to Figure I,1.5 of the qualitatively different perturbed bifurcation diagrams. The highlights of this derivation are as follows. First, we classify the kinds of nonpersistent behavior. Then we show that the set of parameter values in $\mathbb{R}^k$ for which nonpersistent obtains is a finite union of (possibly singular) hypersurfaces in $\mathbb{R}^k$. For example, in Figure I,1.5 nonpersistent behavior occurs along the two curves $\alpha_1 = 0$ and $\alpha_1 = \alpha_2^3/27$; the associated bifurcation diagrams are shown in Figures I,1.6 and I,1.7. In the general case, as in Figure I,1.5, these hypersurfaces divide $\mathbb{R}^k$ into finitely many regions. Any two choices of parameter from within one such region give equivalent bifurcation diagrams—the diagrams *must* be equivalent, since one parameter may be deformed into the other without encountering any nonpersistent behavior. Thus these regions in $\mathbb{R}^k$ enumerate the different perturbed diagrams.

Section by section, the second half of the chapter breaks down as follows. Nonpersistent behavior is classified in §§5 and 10; the latter section deals with the case where singularities may cross the boundary of the neighborhood under consideration. The main theorem is stated in §6. In §§7 and 8, we apply the algorithm of §6 to a simple example and a complicated example; the pitchfork and the winged cusp, respectively. In §9, we sketch a proof of the main theorem from §6. This section is primarily to convey the ideas involved; in a careful treatment we prove the main theorem of §6 as a corollary of a more global theorem, which we state in §10 and prove in §11. The proof sections, §§9 and 11, use techniques from differential topology.

Finally, in §12, we describe the path formulation for bifurcation problems. This pictorial formulation connects bifurcation theory with elementary catastrophe theory. It has proven useful in finding organizing centers in certain applications and in analyzing certain specific bifurcation problems.

We have frequently emphasized that our theory is a local one. Nevertheless, often seemingly global properties of bifurcation diagrams can be derived by studying degenerate local bifurcation problems and their perturbations. Already in the model studied in Chapter I, §2 we obtained such apparently global information. Understanding what the definition of unfoldings means sheds light on this paradox. We address this point at some length in §1.

§1. Unfoldings and Universal Unfoldings

In this section we define the basic concepts of unfolding theory, and we discuss how these definitions relate to germ concepts. The intent of this discussion is to clarify how a local theory can lead to information about global behavior.

First we define "unfolding." This concept is the notion of perturbation with which we work. Let g be in $\mathscr{E}_{x,\lambda}$; a *k-parameter unfolding* of g is a germ $G \in \mathscr{E}_{x,\lambda,\alpha}$, where $\alpha = (\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$, such that for $\alpha = 0$

$$G(x, \lambda, 0) = g(x, \lambda). \tag{1.1}$$

Here G is a germ in all the variables: $x, \lambda, \alpha_1, \ldots, \alpha_k$. Thus G is defined and C^∞ on a neighborhood of zero in $\mathbb{R}^{k+2}$. The restriction to $\alpha = 0$ in (1.1) is compatible with germ concepts.

Let $G(x, \lambda, \alpha)$, $\alpha \in \mathbb{R}^k$ and $H(x, \lambda, \beta)$, $\beta \in \mathbb{R}^l$ be unfoldings of a germ g, where l and k need not be equal. Suppose that for each $\beta \in \mathbb{R}^l$, $H(\cdot, \cdot, \beta)$ is equivalent to some member of the unfolding G; in symbols,

$$H(\cdot, \cdot, \beta) \sim G(\cdot, \cdot, A(\beta)), \tag{1.2}$$

where $A: \mathbb{R}^l \to \mathbb{R}^k$. In such a case we would say that all the perturbations in H are contained in G. Let us formalize this in the following definition.

Definition 1.1. Let $G(x, \lambda, \alpha)$ and $H(x, \lambda, \beta)$ be unfoldings of a germ g. We say that *H factors through G* if there exist smooth mappings S, X, Λ, and A such that

$$H(x, \lambda, \beta) = S(x, \lambda, \beta)G(X(x, \lambda, \beta), \Lambda(\lambda, \beta), A(\beta)), \tag{1.3}$$

where for $\beta = 0$ the following hold: $S(x, \lambda, 0) \equiv 1$, $X(x, \lambda, 0) \equiv x$, $\Lambda(\lambda,0) \equiv \lambda$, and $A(0) = 0$.

We make two comments about this definition.

Remarks 1.2. (a) Since G and H are both unfoldings of the same germ g, it is natural to require that when $\beta = 0$ the equivalence of g with itself induced by (1.3) be the identity equivalence. This is the reason for the conditions on S, X, Λ, and A when $\beta = 0$.

(b) We do not require that $(X(0, 0, \beta), \Lambda(0, \beta)) = (0, 0)$; i.e., when β is nonzero, the equivalence need not preserve the origin. We shall amplify this point considerably in our discussion below.

An unfolding H may factor through another unfolding which contains fewer parameters. For example, consider the one-parameter unfolding

$$H(x, \lambda, \beta) = x^3 - \lambda x + \beta x \tag{1.4}$$

of the pitchfork. Observe that H factors through the zero-parameter unfolding of the pitchfork $G(x, \lambda) = x^3 - \lambda x$; indeed we may let $\Lambda(\lambda, \beta) = \lambda - \beta$. The point at which the pitchfork bifurcation occurs moves from $(x, \lambda) = (0, 0)$ to $(x, \lambda) = (0, \beta)$. From a qualitative point of view, this change is insignificant.

Rather remarkably, for most germs g there are special unfoldings G of g which contain all perturbations of g, up to equivalence. More formally, such an unfolding G has the following property: Every unfolding of g may be factored through G. This property is by far the most important property of a universal unfolding, but is is not quite the definition—we reserve the term "universal" for an unfolding with this property such that there is no redundancy in the parameters. The following definition formalizes this most important concept.

Definition 1.3. An unfolding G of g is *versal* if every other unfolding of g factors through G. A versal unfolding of g depending on the minimum number of parameters possible is called *universal.* That minimum number is called the *codimension of* g.

We augment this definition with the following convention: If g does not possess a versal unfolding we say that g has *codimension infinity.*

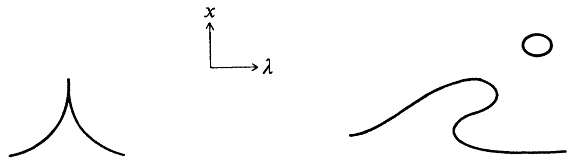
The following formula gives a versal unfolding of the pitchfork which is not universal because there is a redundant unfolding parameter. (Cf. (I, 1.13).)

$$G(x, \lambda) = x^3 - \lambda x + \alpha_1 + \alpha_2 x + \alpha_3 x^2.$$

As with (1.4), the term $\alpha_2 x$ may be absorbed into a change of the λ-coordinate and is therefore redundant.

In the next sections we consider how to find universal unfoldings. The rest of this section is a theoretical discussion expanding on Remark 1.2(b) above; viz., in (1.2) when $\beta \neq 0$ the equivalence (X, Λ) need not preserve the origin in $\mathbb{R}^2$. We have already seen a simple illustration of why this is appropriate in the unfolding (1.4) of the pitchfork. However, the issues here are far more important than this; in particular, they relate to the possibility of obtaining apparently global behavior from a local theory.

First let us discuss how such global information may emerge from local considerations. A single degenerate singularity may split, upon perturbation, into several less degenerate singularities. For a specific perturbed bifurcation



(a) $x^3 + \lambda^2 = 0$ (b) Perturbation of $x^3 + \lambda^2 = 0$

Figure 1.1. The winged cusp and a perturbation.

diagram the exact way in which these less degenerate singularities are connected yields global information about that bifurcation diagram. For example, we recall from Chapter I, §2 that perturbation of the winged cusp $x^3 + \lambda^2$ can lead to the bifurcation diagram shown in Figure 1.1(b). Note that this perturbation of the winged cusp contains four limit points, two of which are connected to form an isola (i.e., an isolated solution branch). It would indeed be impossible to understand this behavior merely from a local study of one point on the *perturbed* diagram in Figure 1.1(b), but this behavior follows naturally in a local analysis of the *degenerate* diagram *and* its perturbations. The point is that as the size of the perturbation tends to zero, all four limit points in Figure 1.1(b) collapse into a single degenerate singularity; in other words, for sufficiently small perturbations, the interesting portion of the bifurcation diagram will be completely inside any given neighborhood of the degenerate singularity.

Let us relate these ideas to germ concepts. One important aspect of germs is that they have base points; i.e., a germ is defined locally in the neighborhood of some given point. (For convenience we have set the base point of germs in $\mathscr{E}_{x,\lambda}$ at the origin.) In the unperturbed bifurcation diagram of Figure 1.1(a) there is indeed a distinguished point which may serve as the base point of a germ. However, in the perturbed diagram Figure 1.1(b) there are *several*, and vital information would be lost by focusing on one to the exclusion of the others. In other words, for $\alpha \neq 0$ in an unfolding, it is of the greatest importance *not* to have a base point. But how is this compatible with germ concepts? The difficulties here are resolved by a careful definition of unfolding. The simplicity of this definition is deceptive. We defined an unfolding as a germ in $\mathbb{R}^{k+2}$; thus the point $x = \lambda = \alpha_1 = \cdots = \alpha_k = 0$ is distinguished, and no other. In conclusion, it might seem that the two notions

$$\text{(i)} \quad G \in \mathscr{E}_{x,\lambda,\alpha}$$

and

$$\text{(ii)} \quad (\forall\alpha) G(\cdot, \cdot, \alpha) \in \mathscr{E}_{x,\lambda}$$

are virtually indistinguishable, but this is far from true; because of problems with base points, the second notion would be wholly unsatisfactory for our purposes. In fact, historically, this distinction was of importance in the development of singularity theory.

§2. A Characterization of Universal Unfoldings

The main objective of this section is to formulate a necessary and sufficient condition which characterizes precisely when an unfolding is universal. This result, Theorem 2.3, is stated in subsection (b). In subsection (a), we present a motivating discussion that in fact leads to a proof of necessity. In subsection (b), besides stating the main result, we also derive three corollaries from Theorem 2.3. Subsection (c) is concerned with how to use the theorem to derive a universal unfolding of a germ g. Specifically, we show that only linear algebra is needed to apply the thorem, once $RT(g)$ has been determined.

(a) Motivation of the Theorem

Suppose that $G(x, \lambda, \alpha)$, $\alpha \in \mathbb{R}^k$, is a universal unfolding of a germ $g \in \mathscr{E}_{x,\lambda}$. This means, in particular, that all one-parameter unfoldings of g may be factored through G. In this subsection, we explore the implications of this factorization.

For any $q \in \mathscr{E}_{x,\lambda}$, consider the one-parameter unfolding of g

$$H(x, \lambda, \varepsilon) = g(x, \lambda) + \varepsilon q(x, \lambda).$$

Since G is universal, H factors through G. Thus we may write

$$H(x, \lambda, \varepsilon) = S(x, \lambda, \varepsilon) \cdot G(X(x, \lambda, \varepsilon), \Lambda(\lambda, \varepsilon), A(\varepsilon)), \tag{2.1}$$

where

$$S(x, \lambda, 0) = 1, \qquad X(x, \lambda, 0) = x, \qquad \Lambda(\lambda, 0) = \lambda, \qquad A(0) = 0. \tag{2.2}$$

On differentiating (2.1) with respect to ε and evaluating at $\varepsilon = 0$, we find

$$\begin{aligned} q(x, \lambda) = {} & \frac{d}{d\varepsilon}[S(x, \lambda, \varepsilon)g(X(x, \lambda, \varepsilon), \Lambda(\lambda, \varepsilon))]\Bigg|_{\varepsilon=0} \\ & + \sum_{i=1}^{k} \dot{A}_i(0)\frac{\partial G}{\partial \alpha_i}(x, \lambda, 0), \end{aligned} \tag{2.3}$$

where $A(\varepsilon) = (A_1(\varepsilon), \ldots, A_k(\varepsilon))$ in coordinates and dot indicates a derivative with respect to ε. The first term in (2.3) is strongly reminiscent of what occurred in the derivation of $RT(g)$ in Chapter II, §2. However, there are the following two important differences:

(i) Before we had $\Lambda(\lambda, \varepsilon) \equiv \lambda$; here Λ can be any smooth germ;
(ii) Before we had $X(0, 0, \varepsilon) = 0$ for all ε; this is not required here.

Our treatment of this term is similar to that of Chapter II, §2. We apply the chain rule to the first term in (2.3) and use (2.2) to obtain

$$\dot{S}(x, \lambda, 0)g(x, \lambda) + g_x(x, \lambda)\dot{X}(x, \lambda, 0) + g_\lambda(x, \lambda)\dot{\Lambda}(\lambda, 0). \tag{2.4}$$

In the following definition we define the tangent space to g as the set of all germs that can arise from this construction.

Definition 2.1. The *tangent space to a germ g* in $\mathscr{E}_{x,\lambda}$ denoted by $T(g)$, consists of all germs of the form

$$ag + bg_x + cg_\lambda,$$

where $a, b, \in \mathscr{E}_{x,\lambda}$ and $c \in \mathscr{E}_\lambda$.

Remark 2.2. Unlike the restricted tangent space, $T(g)$ is *not* an ideal. The difficulty lies with the term $c(\lambda)g_\lambda(x, \lambda)$—multiplication of this term by an arbitrary germ in $\mathscr{E}_{x,\lambda}$ does not preserve its form. This fact is a consequence of our assumption that changes of coordinates in λ are independent of x. In subsection (c), we discuss how this difficulty affects computations with $T(g)$.

We now derive a necessary condition for G to be a universal unfolding of g. We showed above that any germ $q \in \mathscr{E}_{x,\lambda}$ admits the representation (2.3). The first term here is just an element of $T(g)$. The second term is an element of the vector subspace of $\mathscr{E}_{x,\lambda}$ spanned by the k germs

$$\frac{\partial G}{\partial \alpha_1}(x, \lambda, 0), \ldots, \frac{\partial G}{\partial \alpha_k}(x, \lambda, 0)$$

since the coefficients $\dot{A}_i(0)$ are scalars. Thus if G is a universal unfolding of g, then

$$\mathscr{E}_{x,\lambda} = T(g) + \mathbb{R}\left\{\frac{\partial G}{\partial \alpha_1}(x, \lambda, 0), \ldots, \frac{\partial G}{\partial \alpha_k}(x, \lambda, 0)\right\}. \tag{2.5}$$

(b) Statement of the Results

The main theorem in this subject states that the necessary condition (2.5) is also sufficient.

Theorem 2.3. (Universal Unfolding Theorem). *Let g be a germ in $\mathscr{E}_{x,\lambda}$, and let G be a k-parameter unfolding of g. Then G is a versal unfolding of g if and only if*

$$\mathscr{E}_{x,\lambda} = T(g) + \mathbb{R}\left\{\frac{\partial G}{\partial \alpha_1}(x, \lambda, 0), \ldots, \frac{\partial G}{\partial \alpha_k}(x, \lambda, 0)\right\}. \tag{2.6}$$

We defer the proof of sufficiency in this theorem until Volume II.

According to Definition 1.3, a versal unfolding G of a germ g is universal if G contains the minimum number of parameters. The above theorem leads to a convenient characterization of this minimum number. Specifically, the minimum number of parameters in a versal unfolding of g is the minimum number for which equation (2.6) can hold. But the latter number is precisely the codimension of $T(g)$, as defined in Definition II,5.6. We record this fact in the following corollary.

Corollary 2.4. *A versal unfolding G of a germ g is universal if and only if the number of parameters in G equals the codimension of $T(g)$.*

The following special case of the above results is the version we will most commonly apply.

Corollary 2.5. *Let g be a germ in $\mathscr{E}_{x,\lambda}$ of codimension k, and suppose there exist k germs $p_1, \ldots, p_k \in \mathscr{E}_{x,\lambda}$ such that*

$$\mathscr{E}_{x,\lambda} = T(g) \oplus \mathbb{R}\{p_1, \ldots, p_k\}. \tag{2.7}$$

Then

$$G(x, \lambda, \alpha) = g(x, \lambda) + \sum_{j=1}^{k} \alpha_j p_j(x, \lambda) \tag{2.8}$$

is a universal unfolding of g.

Above we observed that codim g, as given in Definition 1.3, equals codim $T(g)$, as given by Definition II,5.6. In the last corollary of this subsection we show that both these notions of codimension are equal to a third notion of codimension that we now introduce.

Let $g \in \mathscr{E}_{x,\lambda}$ be a germ of finite codimension. By the *orbit* of g we mean the set of all germs $f \in \mathscr{E}_{x,\lambda}$ that are equivalent to g; in symbols,

$$\mathcal{O}_g = \{f \in \mathscr{E}_{x,\lambda} : f \sim g\}.$$

We think of $\mathcal{O}_g$ as a "submanifold" of $\mathscr{E}_{x,\lambda}$. Suppose we apply the methods of Chapter II, §10 to characterize germs equivalent to g by a set of defining conditions

$$P_i(f, Df, \ldots, D^m f) = 0 \quad \text{at} \quad x = \lambda = 0; \qquad i = 1, \ldots, K, \tag{2.9}$$

and a set of nondegeneracy conditions (i.e., inequalities). Our third notion of codimension is $K - 2$, where K is the number of equations in (2.9). This definition is in analogy with the definition of codimension of submanifolds in finite dimensions. In a N-dimensional space a system of K equations (typically) defines a manifold of dimension $N - K$, or codimension K. Here we regard $\mathcal{O}_g$ as the solution set in the infinite dimensional space $\mathscr{E}_{x,\lambda}$ of the K equations (2.9). The minus two arises from the fact that in Chapter II we considered equivalence of germs with a fixed base point, whereas in the present context we allow translations in x and λ. (Alternatively, given the fact that every singularity satisfies $g = g_x = 0$, we may regard $K - 2$ as the number of defining conditions beyond these basic two.) In the following proof it will turn out that because of finite determinacy, the infinite-dimensional set $\mathcal{O}_g$ can in fact be analyzed with finite-dimensional techniques.

Corollary 2.6. *If g is a germ of finite codimension in $\mathcal{E}_{x,\lambda}$, the following three integers are equal*:

(i) codim g (*Definition* 1.3).
(ii) codim $T(g)$ (*Definition* II,5.6).
(iii) (*the number of defining conditions for* g) $-$ 2.

PROOF. As noted above, the equality of the first two integers is a direct consequence of Theorem 2.3. To prove equality of the last two we apply some techniques from Lie groups. The reader not familiar with these concepts may skip the proof without loss of continuity. Let Γ be the group of all equivalences acting on $\mathcal{E}_{x,\lambda}/\mathcal{P}(g)$. Since Γ is an algebraic group, the orbit of g in $\mathcal{E}_{x,\lambda}/\mathcal{P}(g)$ is a smooth submanifold. The tangent space to this orbit at g is just $(RT(g) + \mathcal{E}_\lambda\{\lambda g_\lambda\})/\mathcal{P}(g))$; that is, those vectors in $T(g)$ generated by curves of diffeomorphisms $(X(x, \lambda), \Lambda(\lambda))$ which fix the origin. Thus the codimension of the orbit of g is codim $T(g) + 2$. The number of defining conditions is just the number of equations which specify the orbit of g; this number is the codimension of the orbit of g. This proves the second equality.

□

(c) Computation of Universal Unfoldings with Linear Algebra

In Remark 2.2 above, we noted that $T(g)$ is generally not an ideal, in contrast to $RT(g)$. This means that many of the algebraic techniques of Chapter II are not directly applicable to $T(g)$. However, in this subsection we show how these techniques may be adapted to the present context. More specifically, suppose we wish to find a universal unfolding of a germ $g \in \mathcal{E}_{x,\lambda}$ and that we have already determined $RT(g)$. In this subsection we show that the calculations needed to apply the universal unfolding theorem may be divided into the following three stages, each of which requires only linear algebra:

(i) Determine an integer l such that

$$T(g) = RT(g) \oplus \mathbb{R}\{g_x, g_\lambda, \lambda g_\lambda, \ldots, \lambda^l g_\lambda\}. \tag{2.10}$$

(ii) Decompose $T(g)$ in the form

$$T(g) = [\operatorname{Itr} T(g)] \oplus V_g, \tag{2.11}$$

where $V_g = T(g) \cap [\operatorname{Itr} T(g)]^\perp$.

(iii) Find a basis for a complement to V_g in the finite-dimensional space $[\text{Itr } T(g)]^\perp$; in symbols, find linearly independent germs $p_1, \ldots, p_k$ in $[\text{Itr } T(g)]^\perp$ such that

$$[\text{Itr } T(g)]^\perp = V_g \oplus \mathbb{R}\{p_1, \ldots, p_k\}. \tag{2.12}$$

Then (2.8) provides a universal unfolding for g. In §3 below we compute universal unfoldings for several examples following these steps. In the present subsection we prove in general that the various constructions above are possible and lead to the desired goal; i.e., a universal unfolding for g.

The next lemma, due to J. Damon [1980] is the first step in this program.

Lemma 2.7. *$RT(g)$ has finite codimension if and only if $T(g)$ has finite codimension.*

Since $RT(g) \subset T(g)$, one direction of the implication is automatic. We sketch the reverse implication at the end of this subsection.

Let $g \in \mathscr{E}_{x,\lambda}$ be a germ with finite codimension. Let us show that there is an integer l such that (2.10) holds. We recall that a typical germ in $T(g)$ has the form

$$a(x, \lambda)g + b(x, \lambda)g_x + c(\lambda)g_\lambda.$$

Such a germ is in $RT(g)$ if both $b(0, 0) = 0$ and $c(\lambda) \equiv 0$. Thus

$$T(g) = RT(g) + \mathbb{R}\{g_x\} + \mathscr{E}_\lambda\{g_\lambda\}. \tag{2.13}$$

By Lemma 2.7, $RT(g)$ has finite codimension, which implies that $\lambda^s \in RT(g)$ for all sufficiently large s. Hence $\lambda^s g_\lambda \in RT(g)$. Since $RT(g)$ is an ideal there is a unique l satisfying

$$\lambda^l g_\lambda \notin RT(g) \quad \text{and} \quad \lambda^{l+1} g_\lambda \in RT(g). \tag{2.14}$$

Equation (2.10) follows from (2.13) and (2.14). (*Remark*: In many simple examples (2.10) is satisfied with $l = 0$. In particular, this is always true for quasi-homogeneous polynomials—see Exercise 2.1 for a definition and further exploration of this topic.)

Concerning (2.11), we repeat the construction of Chapter II, §7 to show that $T(g)$ has a well-defined intrinsic part. By Lemma 2.7, $RT(g)$ has finite codimension, so by Proposition II,5.7, $RT(g)$ contains $\mathscr{M}^k$ for some k. $T(g)$, being larger, also contains $\mathscr{M}^k$. It follows from Proposition II,7.1 that there are only finitely many intrinsic ideals $\mathscr{I}$ such that

$$\mathscr{M}^k \subset \mathscr{I} \subset \mathrm{T}(g).$$

The sum of all these is the largest intrinsic ideal contained in $T(g)$, denoted Itr $T(g)$.

In Corollary II,7.4 we showed that a decomposition of the form (2.11) is possible for any ideal of finite codimension, and Exercise II,7.3 extended this result to subspaces which contain $\mathscr{M}^k$ for some k. Thus this earlier work shows that the decomposition (2.11) is possible.

Let $p_1, \ldots, p_k \in [\text{Itr } T(g)]^\perp$ be chosen as in (2.12), where $k = \text{codim } T(g)$. It follows from (2.11) that condition (2.7) holds; thus by Corollary 2.4, (2.8) is a versal unfolding of g. Since $p_1, \ldots, p_k$ are linearly independent, (2.8) is, in fact, a universal unfolding.

SKETCH OF PROOF OF LEMMA 2.7. We assume that $T(g)$ has finite codimension and show that $RT(g)$ also has finite codimension, reasoning by contradiction.

The first step in Damon's [1980] proof is to reduce the case where g is a polynomial. This allows us to consider the equations

$$g = g_x = 0 \tag{2.15}$$

over the complex numbers; i.e., as two equations for two unknown complex scalars.

Suppose that the ideal $\langle g, g_x \rangle$ has infinite codimension. This is equivalent to assuming that $RT(g)$ has infinite codimension since

$$\langle g, g_x \rangle \supset RT(g) \supset \mathscr{M} \cdot \langle g, g_x \rangle.$$

If $RT(g)$ has infinite codimension then the solutions of (2.15) are not isolated; indeed (2.15) defines a nontrivial algebraic variety in $\mathbb{C}^2$. The curve selection lemma (Milnor [1968], p. 25) allows us to quantify this "nonisolatedness." This result states that the solution set of (2.15) contains a nonconstant smooth curve $X(t)$, $\Lambda(t)$, where t is a real parameter, such that $X(0) = \Lambda(0) = 0$; in symbols,

$$\begin{aligned} &\text{(a)} \quad g(X(t), \Lambda(t)) \equiv 0, \\ &\text{(b)} \quad g_x(X(t), \Lambda(t)) \equiv 0. \end{aligned} \tag{2.16}$$

Differentiating (2.16a) with respect to t and applying (2.16b) yields

$$g_\lambda(X(t), \Lambda(t)) \cdot \Lambda'(t) \equiv 0. \tag{2.17}$$

By continuity, either $g\,(X(t), \Lambda(t)) \equiv 0$ or $\Lambda'(t) \equiv 0$. The first case coupled with (2.16) shows that the ideal $\langle g, g_x, g_\lambda \rangle$ has infinite codimension. Since this ideal contains $T(g)$ we have a contradiction. Hence $\Lambda'(t) \equiv 0$.

Since $\Lambda(0) = 0$, we see that $\Lambda(t) \equiv 0$. Thus (2.16) implies

$$g(X(t), 0) \equiv 0, \qquad g_x(X(t), 0) \equiv 0. \tag{2.18}$$

Now consider the ideal $\mathscr{I}$ spanned by $T(g)$ and λ. It is easy to compute that

$$\mathscr{I} = \langle g(x, 0), g_x(x, 0), \lambda \rangle. \tag{2.19}$$

Since $\mathscr{I} \supset T(g)$, $\mathscr{I}$ must have finite codimension. However, it follows from (2.19) that $\mathscr{I}$ has finite codimension precisely when the ideal $\langle g(x, 0), g_x(x, 0) \rangle$ has finite codimension in $\mathscr{E}_x$. Thus $x = 0$ is the only common zero of $g(x, 0)$

$= g_x(x, 0) = 0$. This means that in (2.18) we must have $X(t) \equiv 0$, contradicting the choice of $(X(t), \Lambda(t))$. Thus the ideal $\langle g, g_x \rangle$ must have finite codimension.

EXERCISES

2.1. We say that g is *quasi-homogeneous* if there exist positive integers α, β, γ such that

$$g(t^\alpha x, t^\beta \lambda) = t^\gamma g(x, \lambda). \tag{2.20}$$

Let $g(x, \lambda) \in \mathscr{E}_{x, \lambda}$ be quasi-homogeneous. Show that $\lambda g_\lambda \in RT(g)$ and that $T(g) = RT(g) + \mathbb{R}\{g_x, g_\lambda\}$.

Hint: Differentiate (2.20) with respect to t and evaluate at $t = 1$.)

2.2. (*Discussion*) Consider bifurcation problems $g(x, \lambda)$ which are constrained to have a trivial solution; that is, bifurcation problems such that $g(0, \lambda) \equiv 0$. We call two such bifurcation problems g and h *t-equivalent* if g and h are equivalent and the equivalence preserves the trivial solution; more precisely, if

$$g(x, \lambda) = S(x, \lambda)h(X(x, \lambda), \Lambda(\lambda)),$$

where (X, Λ, S) is an equivalence, then we require $X(0, \lambda) \equiv 0$.

By Taylor's theorem, a bifurcation problem with a trivial solution may be written in the form

$$g(x, \lambda) = xf(x, \lambda). \tag{2.21}$$

Compute $T_t(g)$, the formal tangent space to g under t-equivalence. *Answer*: If g has the form (2.21), then

$$T_t(g) = [\langle f, xf_x \rangle + \mathscr{E}_\lambda\{f_\lambda\}]\{x\}. \tag{2.22}$$

2.3. (*Discussion*) Let $g(x, \lambda)$ be a bifurcation problem such that $g(0, \lambda) \equiv 0$. If there is a finite dimensional subspace V of $\mathscr{E}_{x, \lambda}\{x\}$ such that

$$\mathscr{E}_{x, \lambda}\{x\} = T_t(g) \oplus V,$$

we say that g has finite t-codimension, and we define $\operatorname{codim}_t g = \dim V$. It can be proved that a t-universal unfolding of g may be constructed from a basis for V. (Cf. Theorem 2.3.)

(a) Show that $x^2 - \lambda x$ has t-codimension zero. (*Remark*: This singularity is persistent to perturbations preserving the trivial solution.)

(b) Show that $x^3 - \lambda x$ has t-codimension one and that $x^3 - \lambda x + \alpha x^2$ is a t-universal unfolding. Graph the resulting bifurcation diagrams.

§3. Examples of Universal Unfoldings

In subsection (a) we compute universal unfoldings for two simple examples: the pitchfork and limit point singularities. For these examples we perform explicitly the steps described in §2(c) above. In subsection (b) we list universal unfoldings for several examples considered in Chapter II, along with some highlights of the computation. Finally in subsection (c) we study further the three singularities of codimension one that occur in subsection (b). (In fact these are the only singularities of codimension one.)

(a) Two Simple Examples

We begin this subsection by showing that both

$$\begin{aligned} &\text{(a)} \quad G(x, \lambda, \alpha, \beta) = x^3 - \lambda x + \alpha + \beta x^2, \quad \text{and} \\ &\text{(b)} \quad H(x, \lambda, \alpha, \beta) = x^3 - \lambda x + \alpha + \beta\lambda \end{aligned} \tag{3.1}$$

are universal unfoldings of the pitchfork, $h(x, \lambda) = x^3 - \lambda x$. The first step is to compute $T(x^3 - \lambda x)$. Recall from (II,6.3) that

$$RT(x^3 - \lambda x) = \mathcal{M}^3 + \mathcal{M}\langle\lambda\rangle.$$

Now observe that

$$\lambda h_\lambda = x\lambda \in \mathcal{M}\langle\lambda\rangle \subset RT(x^3 - \lambda x).$$

It therefore follows from (2.10) that

$$T(x^3 - \lambda x) = (\mathcal{M}^3 + \mathcal{M}\langle\lambda\rangle) \oplus \mathbb{R}\{3x^2 - \lambda, x\},$$

which is already in the form (2.11), Itr $T(x^3 - \lambda x) \oplus V_g$. Since

$$[\text{Itr } T(x^3 - \lambda x)]^\perp = (\mathcal{M}^3 + \mathcal{M}\langle\lambda\rangle)^\perp = \mathbb{R}\{1, x, \lambda, x^2\},$$

we need only find a basis for a complementary subspace to

$$V_g = \mathbb{R}\{3x^2 - \lambda, x\} \quad \text{in} \quad \mathbb{R}\{1, x, \lambda, x^2\}.$$

It is easy to see that either $\{1, x^2\}$ or $\{1, \lambda\}$ form such a basis. Applying Corollary 2.4, we see that both unfoldings in (3.1) are universal and that the codimension of the pitchfork is two.

Next we consider the simplest singularity $h(x, \lambda) = x^2 + \lambda$. Recall from (II,3.4) that

$$RT(x^2 + \lambda) = \mathcal{M}^2 + \langle\lambda\rangle.$$

A short calculation using (2.10) shows that

$$T(x^2 + \lambda) = \mathscr{E}_{x,\lambda}.$$

Hence the codimension of the limit point is zero and the limit point is its own universal unfolding. A consequence of this fact is that any small perturbation of the limit point is equivalent to the same normal form. This is the property of *persistence* which we will study below; indeed we will show that the only persistent singularity is the limit point. (*Remark*: Using the unfolding theorem to prove that limit points are persistent is a wasteful use of mathematical power. In Appendix 1 we give a much simpler proof based on the implicit function theorem.)

(b) A Tabulation of Some Simple Universal Unfoldings

We list the universal unfoldings and codimensions of several germs in Table 3.1.

The calculations involved in completing Table 3.1 are now all elementary. The main computations are summarized in Table 3.2.

We have shown above that once $RT(g)$ has been computed, the computation of universal unfoldings uses only linear algebra. We illustrate this by performing the calculations for one of the cases in the tables; viz., our academic example $g(x, \lambda) = x^5 + x^3\lambda + \lambda^2$. Recall from (II,6.2) that $RT(x^5 + x^3\lambda + \lambda^2)$ is the entry given in Table 3.2. Next observe that

$$\lambda^2 g_\lambda(x, \lambda) = x^3\lambda^2 + 2\lambda^3 \in RT(g).$$

Thus by formula (2.10)

$$T(x^5 + x^3\lambda + \lambda^2) = (\mathcal{M}^6 + \mathcal{M}^4\langle\lambda\rangle + \mathcal{M}\langle\lambda^2\rangle) \oplus \mathbb{R}\{x^5 + x^3\lambda + \lambda^2, 5x^5 + 3x^3\lambda, g_x, g_\lambda, \lambda g_\lambda\}.$$

Table 3.1. Universal Unfoldings for Several Examples.

g	codim g	Basis for V
(1) $x^n + \lambda$ $(n \geq 2)$	$n - 2$	$x, x^2, \ldots, x^{n-2}$
(2) $x^n + \lambda x$ $(n \geq 3)$	$n - 1$	$1, x^2, x^3, \ldots, x^{n-1}$
(3) $x^2 \pm \lambda^2$	1	1
(4) $x^3 + \lambda^2$	3	$1, x, x\lambda$
(5) $x^5 + x^3\lambda + \lambda^2$	6	$1, x, x^2, \lambda, \lambda x, \lambda x^2$

Table 3.2. Summary of Computations.

g	$RT(g)$	$T(g)$
(1) $x^n + \lambda$ $(n \geq 2)$	$\mathcal{M}^n + \langle\lambda\rangle$	$(\mathcal{M}^{n-1} + \langle\lambda\rangle) \oplus \mathbb{R}\{1\}$
(2) $x^n + \lambda x$ $(n \geq 3)$	$\mathcal{M}^n + \mathcal{M}\langle\lambda\rangle$	$(\mathcal{M}^n + \mathcal{M}\langle\lambda\rangle) \oplus \mathbb{R}\{nx^{n-1} + \lambda, x\}$
(3) $x^2 \pm \lambda^2$	$\mathcal{M}^2$	$\mathcal{M}$
(4) $x^3 + \lambda^2$	$\mathcal{M}^3 + \langle\lambda^2\rangle$	$(\mathcal{M}^3 + \langle\lambda^2\rangle) \oplus \mathbb{R}\{x^2, \lambda\}$
(5) $x^5 + x^3\lambda + \lambda^2$	$\mathcal{M}^6 + \mathcal{M}^4\langle\lambda\rangle + \mathcal{M}\langle\lambda^2\rangle + \mathbb{R}\{x^5 + x^3\lambda + \lambda^2, 5x^5 + 3x^3\lambda\}$	$\mathcal{M}^5 + \mathcal{M}^3\langle\lambda\rangle + \langle\lambda^2\rangle + \mathbb{R}\{5x^4 + 3x^2\lambda, x^3 + 2\lambda\}$

Observe that the vector space spanned by

$$x^5 + x^3\lambda + \lambda^2,\ 5x^5 + 3x^3\lambda, \text{ and } \lambda g_\lambda = x^3\lambda + 2\lambda^2$$

is the same as the vector space spanned by

$$x^5, x^3\lambda, \text{ and } \lambda^2.$$

These monimials when added to $\mathcal{M}^6 + \mathcal{M}^4\langle\lambda\rangle + \mathcal{M}\langle\lambda^2\rangle$ generate the intrinsic ideal $\mathcal{M}^5 + \mathcal{M}^3\langle\lambda\rangle + \langle\lambda^2\rangle$. The formula for $T(x^5 + x^3\lambda + \lambda^2)$ in Table 3.2 follows from the comment that $g_x = 5x^4 + 3x^2\lambda$ and $g_\lambda = x^3 + 2\lambda$.

To complete our calculation of the universal unfolding of $x^5 + x^3\lambda + \lambda^2$ we note that

$$(\mathcal{M}^5 + \mathcal{M}^3\langle\lambda\rangle + \langle\lambda^2\rangle)^\perp = \mathbb{R}\{1, x, x^2, x^3, x^4, \lambda, \lambda x, \lambda x^2\} \qquad (3.2)$$

whose dimension is 8. To find a universal unfolding for $x^5 + x^3\lambda + \lambda^2$ we need only find a basis for a subspace of (3.2) which is complementary to $\mathbb{R}\{g_x, g_\lambda\}$. Such a complement is six dimensional, and we have chosen a particular basis in Table 3.1(5).

(c) Singularities of Codimension One

We end this section with a discussion of the universal unfoldings of the three types of codimension one singularities which appear in Table 3.1; namely, $x^2 - \lambda^2$, $x^2 + \lambda^2$, and $x^3 + \lambda$. In Chapter IV, we shall prove that these are, in fact, the only codimension one singularities. These singularities of codimension one are important in the second half of this chapter. Unlike the limit point considered above, they definitely are not persistent—a small perturbation yields a diagram with different qualitative behavior, as can be seen from the figures.

We call the normal form $x^2 - \lambda^2$ *simple bifurcation* as it is the normal form for the simplest bifurcation problem (in the sense of lowest codimension) in which bifurcation in the classical sense occurs. The bifurcation diagrams contained in the universal unfolding $x^2 - \lambda^2 + \alpha$ are given in Figure 3.1.

The bifurcation problem $x^2 + \lambda^2$ is called an *isola center*. This bifurcation problem was first brought to our attention by E. L. Reiss who observed that such bifurcation problems appear frequently in the chemical engineering

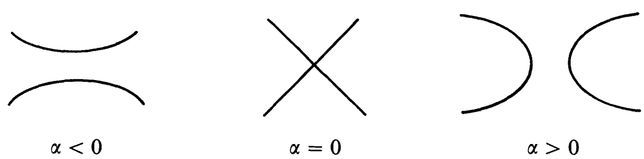


Figure 3.1. Simple bifurcation $x^2 - \lambda^2 + \alpha = 0$.

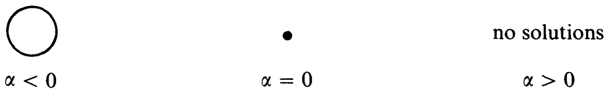


Figure 3.2. The isola center $x^2 + \lambda^2 + \alpha = 0$.

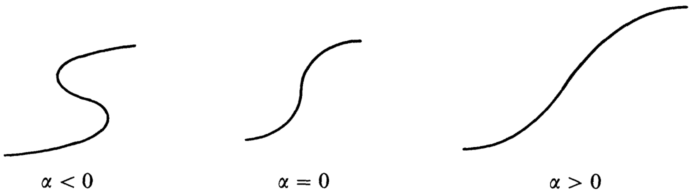


Figure 3.3. The hysteresis point $x^3 - \lambda + \alpha x = 0$.

literature. Our discussion of the CSTR in Chapter I supports this observation. The bifurcation diagrams contained in the universal unfolding of the isola center $x^2 + \lambda^2 + \alpha$ appear in Figure 3.2.

We call the bifurcation problem $x^3 - \lambda$ a *hysteresis point*. The bifurcation diagrams contained in the universal unfolding $x^3 - \lambda + \alpha x = 0$ are presented in Figure 3.3. The justification for this terminology was given in our discussion of the pitchfork in Chapter I, §1(e). We note that hysteresis points yield (when $\alpha < 0$) the S-curve frequently observed in combustion theory.

§4. The Recognition Problem for Universal Unfoldings

In this section we consider the following situation which often arises in applications: Let $G(x, \lambda, \alpha)$ be an unfolding of a germ g, where g is equivalent to some normal form h. Is G a universal unfolding of g? We call this the recognition problem for universal unfoldings. Theorem 2.3 provides a way to answer this question, but an attempt to apply this theorem directly often leads to unwieldy calculations. In this section we show how to reduce the calculations, taking advantage of the simplicity of the normal form h, to one very specific task. More precisely, we show that G is a universal unfolding of g if and only if a certain $m \times m$ determinant is nonzero, where m is the codimension of Itr $T(h)$. The entries of this determinant are various derivatives of G. For example, if G is a one-parameter unfolding of a singularity g which is equivalent to the hysteresis point $x^3 + \lambda$, then G is a universal unfolding of g if and only if

$$\det\begin{pmatrix} g_\lambda & g_{\lambda x} \\ G_\alpha & G_{\alpha x} \end{pmatrix} \neq 0, \tag{4.1}$$

when $x = \lambda = \alpha = 0$.

This section is divided into three subsections. In subsection (a) we present some theoretical facts needed to justify our method. We analyze three explicit normal forms in subsection (b): hysteresis points (mentioned above), the pitchfork, and winged cusp. For each of these we obtain a characterization of universal unfoldings analogous to (4.1). In the last subsection we briefly summarize the method in general.

(a) Theoretical Basis of the Method

As above, let G be a k-parameter unfolding of g, where g is equivalent to a normal form h. Combining Theorem 2.3 and formula (2.11) we see that G is a universal unfolding of g provided

$$\mathscr{E}_{x,\lambda} = \operatorname{Itr} T(g) \oplus V_g \oplus \mathbb{R}\left\{\frac{\partial G}{\partial \alpha_1}(x, \lambda, 0), \ldots, \frac{\partial G}{\partial \alpha_k}(x, \lambda, 0)\right\}. \tag{4.2}$$

The following lemma provides the first simplification of the calculations. (Cf. Lemma II,12.2 concerning the analogous result for $RT(g)$.)

Lemma 4.1. *If g and h are equivalent germs in $\mathscr{E}_{x,\lambda}$,*

$$\operatorname{Itr} T(g) = \operatorname{Itr} T(h).$$

We prove this lemma at the end of subsection (a).

We will show below that the validity of (4.2) can be tested by calculations performed in the finite-dimensional space $[\operatorname{Itr} T(h)]^\perp$. Now by Lemma 4.1, V_g is already contained in $[\operatorname{Itr} T(h)]^\perp$. However, the third summand in (4.2) is not contained in $[\operatorname{Itr} T(h)]^\perp$. To deal with this difficulty let us construct explicitly the projection

$$J\colon \mathscr{E}_{x,\lambda} \to [\operatorname{Itr} T(h)]^\perp \tag{4.3}$$

associated to the decomposition

$$\mathscr{E}_{x,\lambda} = \operatorname{Itr} T(h) \oplus [\operatorname{Itr} T(h)]^\perp. \tag{4.4}$$

We claim that for any $f \in \mathscr{E}_{x,\lambda}$

$$Jf = \sum_\alpha{}' \frac{1}{\alpha!} D^\alpha f(0, 0) x^{\alpha_1}\lambda^{\alpha_2}, \tag{4.5}$$

where $\sum'$ indicates the (finite) sum over monomials $x^{\alpha_1}\lambda^{\alpha_2}$ not belonging to $\operatorname{Itr} T(h)$. Certainly $Jf \in [\operatorname{Itr} T(h)]^\perp$ and $(f - Jf) \in \operatorname{Itr} T(h)$, so the claim follows. (*Remark*: If by chance $\operatorname{Itr} T(h) = \mathscr{M}^{k+1}$, then $Jf = j^k f$. This is the reason for using the letter J in (4.3).)

Lemma 4.2. *Let G be a k-parameter unfolding of a germ g of codimension k. Formula* (4.2) *is valid if and only if*

$$[\operatorname{Itr} T(h)]^\perp = V_g + \mathbb{R}\left\{J\frac{\partial G}{\partial \alpha_1}(x, \lambda, 0), \ldots, J\frac{\partial G}{\partial \alpha_k}(x, \lambda, 0)\right\}. \tag{4.6}$$

Remark. It follows by counting dimensions that if (4.6) holds, the sum (4.6) is in fact a direct sum.

Lemma 4.2 provides the foundation for our solution of the recognition problem for universal unfoldings with linear algebra calculations; i.e., to prove that G is a universal unfolding of g it suffices to verify (4.6). Now (4.6) can be verified by computing that a certain determinant is nonzero. However, to show this, it is necessary to choose a basis for V_g, and this can only be done on a case by case basis. Let us elaborate. V_g may be characterized as the image of $T(g)$ in $[\text{Itr } T(g)]^\perp$ under J. It follows from Definition 2.1, the definition of $T(g)$, that V_g is spanned (as a vector space) by the germs

$$J(x^r\lambda^s g),\ J(x^r\lambda^s g_x),\ J(\lambda^s g_\lambda), \tag{4.7}$$

where $r, s \geq 0$. Of course, by finite determinancy only finitely many of the terms in (4.7) are nonzero. Even so, there remains the problem of which ones to select to obtain a basis for V_g. In attacking this problem, we use information about the recognition problem for normal forms to make this selection in a way that requires less computation. However, we prefer to discuss this issue by example first; thus we consider three specific examples in the next subsection, and in subsection (c) we return to a theoretical discussion of this method.

Proof of Lemma 4.1. Since g and h are equivalent, there exist S, X, and Λ satisfying

$$h(x, \lambda) = S(x, \lambda) \cdot g(X(x, \lambda), \Lambda(\lambda)). \tag{4.8}$$

Moreover, we can think of the triple (S, X, Λ) as being a fixed equivalence γ and define the action $\gamma(g)$ of γ on g by the right-hand side of (4.8). Since γ is an equivalence, there is an inverse equivalence, which we denote by γ^{-1}. In particular,

$$\gamma^{-1}(h) = \frac{1}{S(\Phi^{-1}(x, \lambda))} \cdot h(\Phi^{-1}(x, \lambda)),$$

where $\Phi(x, \lambda) = (X(x, \lambda), \Lambda(\lambda))$.

Now suppose p is in Itr $T(g)$. It follows that $\gamma^{-1}(p)$ is in Itr $T(g) \subset T(g)$. Upon recalling the definition of the tangent space $T(g)$, we conclude that there is a curve of bifurcation problems $g_t(x, \lambda)$ with $g_0(x, \lambda) = g(x, \lambda)$ satisfying

$$\left.\frac{d}{dt} g_t\right|_{t=0} = \gamma^{-1}(p). \tag{4.9}$$

(*Remark*: Here subscript "t" merely indicates another variable on which g_t depends, *not* a partial derivative.) Applying the equivalence γ, which is independent of t, to (4.9) yields

$$p = \gamma\left(\left.\frac{d}{dt} g_t\right|_{t=0}\right) = \left.\frac{d}{dt} \gamma(g_t)\right|_{t=0}.$$

Since $\gamma(g_0) = \gamma(g) = h$, it follows that $\gamma(g_t)$ is a curve of bifurcation problems based at h and that p is in $T(h)$. Hence Itr $T(g) \subset$ Itr $T(h)$. Interchanging the roles of g and h yields equality. □

PROOF OF LEMMA 4.2. This proof proceeds most naturally by a quotient space argument. Let

$$\pi\colon \mathscr{E}_{x,\lambda} \to \mathscr{E}_{x,\lambda}/\mathrm{Itr}\ T(h)$$

be the standard projection. Then

$$\pi \circ J = \pi. \tag{4.10}$$

We introduce π for the following reason: A subspace W is a complementary subspace (to Itr $T(h)$ in $\mathscr{E}_{x,\lambda}$) if and only if $\pi(W) = \mathscr{E}_{x,\lambda}/\mathrm{Itr}\ T(h)$.

Now let $W = V_g + \mathbb{R}\{G_{\alpha_1}, \ldots, G_{\alpha_k}\}$. In this notation the lemma states that W is a complementary subspace if and only if $J(W) = \mathrm{Itr}\ T(h)^\perp$. However, we see from (4.10) that W is a complementary subspace if and only if $J(W)$ is a complementary subspace. Since $J(W) \subset \mathrm{Itr}\ T(h)^\perp$ and

$$\mathscr{E}_{x,\lambda} = \mathrm{Itr}\ T(h) \oplus [\mathrm{Itr}\ T(h)]^\perp$$

it follows that $J(W)$ is a complementary subspace if and only if $J(W) = [\mathrm{Itr}\ T(h)]^\perp$. □

(b) Three Examples

In this subsection we solve the recognition problem for hysteresis points, the pitchfork, and the winged cusp, in that order. We treat the first case in some detail; since all three calculations are rather similar we are somewhat briefer with the last two.

Proposition 4.3. *Suppose g is equivalent to $h(x, \lambda) = \pm x^3 \pm \lambda$, and let G be a one-parameter unfolding of g. Then G is a universal unfolding of g if and only if*

$$\det\begin{pmatrix} g_\lambda & g_{\lambda x} \\ G_\alpha & G_{\alpha x} \end{pmatrix} \neq 0 \tag{4.11}$$

at $x = \lambda = \alpha = 0$.

Remark. Since for $\alpha = 0$, we have $G(x, \lambda, 0) \equiv g(x, \lambda)$, in the first row of (4.11) we could replace g by G.

PROOF. To avoid cumbersome notations, we display only the normal form $h(x, \lambda) = x^3 + \lambda$; consideration of the other possible signs is no different.

We carry out the steps outlined in subsection (a) above. From Table 3.2 we see that

$$T(x^3 + \lambda) = (\mathcal{M}^2 + \langle\lambda\rangle) \oplus \mathbb{R}\{1\}.$$

Therefore

$$[\text{Itr } T(g)]^\perp = [\text{Itr } T(h)]^\perp = \mathbb{R}\{1, x\}.$$

The projection J in (4.5) reduces to

$$Jf = f(0, 0) + f_x(0, 0)x, \tag{4.12a}$$

or in components

$$Jf = (f(0, 0), f_x(0, 0)). \tag{4.12b}$$

According to Lemma 4.2, G is a universal unfolding of g if and only if

$$\mathbb{R}\{1, x\} = V_g + \mathbb{R}\left\{J\frac{\partial G}{\partial \alpha}\right\}. \tag{4.13}$$

Now we must choose a basis for the (one-dimensional) space V_g from the list (4.7). At this stage we use information about the recognition problem for germs. Specifically, since g is equivalent to $x^3 + \lambda$ we know that

$$g = g_x = g_{xx} = 0 \quad \text{at} \quad x = \lambda = 0.$$

Substituting into (4.12) we see that

$$Jg = 0, \qquad Jg_x = 0, \qquad J(\lambda g_\lambda) = 0.$$

In other words, on the list (4.7) only the term Jg_λ is nonzero. Thus, Jg_λ is a basis for V_g, and we may rewrite (4.13)

$$\mathbb{R}\{1, x\} = \mathbb{R}\{Jg_\lambda, JG_\alpha\}. \tag{4.14}$$

To conclude, G is a universal unfolding of g if and only if (4.14) holds. Writing these two vectors in terms of components as in (4.12b) leads immediately to (4.11). □

Remarks. (i) Note that (4.11) contains the derivative $g_{\lambda x}$ which does *not* enter into the solution of the recognition problem for the normal form $x^3 + \lambda$. This is typical—terms which are higher-order in the recognition problem for normal forms may not be higher order in the recognition problem for universal unfoldings.

(ii) Since $x^3 + \lambda + \alpha x$ is a universal unfolding of $x^3 + \lambda$, it is tempting to think that G is a universal unfolding of g if $G_{\alpha x} \neq 0$. However, we see from (4.6) that this statement is valid only if $g_{\lambda x} = 0$. Although $g_{\lambda x} = 0$ for the normal form $x^3 + \lambda$, $g_{\lambda x}$ is not zero for every g equivalent to $x^3 + \lambda$.

Proposition 4.4. *Let $G(x, \lambda, \alpha, \beta)$ be a two-parameter unfolding of a germ g equivalent to $h(x, \lambda) = \pm x^3 \pm \lambda x$. Then G is a universal unfolding of g if and only if*

$$\det\begin{pmatrix} 0 & 0 & g_{x\lambda} & g_{xxx} \\ 0 & g_{\lambda x} & g_{\lambda\lambda} & g_{\lambda xx} \\ G_\alpha & G_{\alpha x} & G_{\alpha\lambda} & G_{\alpha xx} \\ G_\beta & G_{\beta x} & G_{\beta\lambda} & G_{\beta xx} \end{pmatrix} \neq 0 \tag{4.15}$$

at $x = \lambda = \alpha = \beta = 0$.

Proof. We display only the case $h(x, \lambda) = +x^3 - \lambda x$. From Table 3.2 we see that

$$T(x^3 - \lambda x) = (\mathcal{M}^3 + \mathcal{M}\langle\lambda\rangle) \oplus \mathbb{R}\{3x^2 - \lambda, x\}.$$

Thus

$$Jf = f(0, 0) + f_x(0, 0)x + f_\lambda(0, 0)\lambda + \tfrac{1}{2}f_{xx}(0, 0)x^2. \tag{4.16}$$

According to Lemma 4.2, G is a universal unfolding of g if and only if

$$\mathbb{R}\{1, x, \lambda, x^2\} = V_g + \mathbb{R}\{JG_\alpha, JG_\beta\}. \tag{4.17}$$

To choose a basis for V_g from the list (4.7), we recall that if g is equivalent to the pitchfork, then

$$g = g_x = g_\lambda = g_{xx} = 0 \tag{4.18}$$

at $x = \lambda = 0$. Therefore

$$Jg = 0, \qquad J(xg_x) = 0, \qquad J(\lambda g_x) = 0, \qquad J(\lambda g_\lambda) = 0.$$

In other words, only Jg_x and Jg_λ are nonzero in (4.7). We rewrite (4.17) as

$$\mathbb{R}\{1, x, \lambda, x^2\} = \mathbb{R}\{Jg_x, Jg_\lambda, JG_\alpha, JG_\beta\}. \tag{4.19}$$

We obtain (4.15) on writing (4.19) in components and using (4.18) to eliminate some terms which are zero. □

Proposition 4.5. *Let $G(x, \lambda, \alpha, \beta, \gamma)$ be a three-parameter unfolding of a germ g equivalent to $h(x, \lambda) = \pm x^3 \pm \lambda^2$. Then G is a universal unfolding of g if and only if*

$$\det\begin{pmatrix} 0 & 0 & 0 & g_{xxx} & g_{xx\lambda} \\ 0 & 0 & g_{\lambda\lambda} & g_{\lambda xx} & g_{\lambda x\lambda} \\ G_\alpha & G_{\alpha x} & G_{\alpha\lambda} & G_{\alpha xx} & G_{\alpha x\lambda} \\ G_\beta & G_{\beta x} & G_{\beta\lambda} & G_{\beta xx} & G_{\beta x\lambda} \\ G_\gamma & G_{\gamma x} & G_{\gamma\lambda} & G_{\gamma xx} & G_{\gamma x\lambda} \end{pmatrix} \neq 0 \tag{4.20}$$

at $x = \lambda = \alpha = \beta = \gamma = 0$.

PROOF. We display only the case $h(x, \lambda) = +x^3 + \lambda^2$. From Table 3.2 we see that

$$T(x^3 + \lambda^2) = (\mathcal{M}^3 + \langle \lambda^2 \rangle) \oplus \mathbb{R}\{x^3, \lambda\}.$$

Thus

$$Jf = f(0, 0) + f_x(0, 0)x + f_\lambda(0, 0)\lambda + \tfrac{1}{2}f_{xx}(0, 0)x^2 + f_{\lambda x}(0, 0)\lambda x.$$

G is a universal unfolding of g if and only if

$$\mathbb{R}\{1, x, \lambda, x^2, \lambda x\} = V_g + \mathbb{R}\{JG_\alpha, JG_\beta, JG_\gamma\}. \tag{4.21}$$

To choose a basis for V_g from the list (4.7), we recall that if g is equivalent to the winged cusp, then

$$g = g_x = g_\lambda = g_{xx} = g_{\lambda x} = 0$$

at $x = \lambda = 0$. Therefore

$$Jg = 0, \qquad J(xg_x) = 0, \qquad J(\lambda g_x) = 0, \qquad J(\lambda g_\lambda) = 0.$$

We rewrite (4.21) as

$$\mathbb{R}\{1, x, \lambda, x^2, \lambda x\} = \mathbb{R}\{Jg_x, Jg_\lambda, JG_\alpha, JG_\beta, JG_\gamma\},$$

from which (4.20) follows. □

(c) Summary

Let us summarize the above method for solving the recognition problem for universal unfoldings. Let G be a k-parameter unfolding of a germ g, where g is equivalent to some normal form h. We assume that h has codimension k. The recognition problem for h must be solved before applying the present method; in particular, we regard Itr $T(h)$ as known. The method leads to a $m \times m$ determinant characterizing the universality of G, where m is the codimension of Itr $T(g)$.

We isolate the following three conceptual steps in applying the method; only the second requires actual computations:

(i) Given Itr $T(h)$, construct the projection J as in (4.5).
(ii) Use the recognition problem for h to eliminate linearly dependent germs in the list (4.7); this leads to a basis for V_g, say $Jp_1, \ldots, Jp_l$ where $l = m - k$.
(iii) Expand the m vectors

$$Jp_1, \ldots, Jp_l, JG_{\alpha_1}, \ldots, JG_{\alpha_k}$$

in the monomonial basis for $[\text{Itr } T(h)]^\perp$ to obtain the desired $m \times m$ determinant. Often some of the elements of this matrix are zero, because of defining conditions in the recognition problem.

§5. Nonpersistent Bifurcation Diagrams

With this section we begin the second major theme of Chapter III. Sections 1–4 dealt with the first theme; viz., how to find or recognize universal unfoldings. The second theme is how to enumerate perturbed bifurcation diagrams, given a universal unfolding.

In carrying out the enumeration we focus on the following question: Which of the perturbed bifurcation diagrams in the universal unfolding of g would remain unchanged (in the qualitative sense of equivalence) if subjected to an additional small perturbation? We call such diagrams *persistent*. Actually, we focus on nonpersistent diagrams. We will show that there are three sources of nonpersistence; namely, bifurcation, hysteresis, and double limit points. (See Remark 5.2(i) below concerning the isola center.) In Figure 5.1 we sketch these three phenomena, along with small perturbations which demonstrate their nonpersistence. In each case, note that the indicated perturbation changes the number of solutions x as a function of λ.

More formally, let $G: \mathbb{R} \times \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}$ be a universal unfolding of a germ $g: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. We isolate the above nonpersistence phenomena in the following definition.

Definition 5.1.

(a) $\mathscr{B} = \{\alpha \in \mathbb{R}^k: \exists (x, \lambda) \in \mathbb{R} \times \mathbb{R}$ such that $G = G_x = G_\lambda = 0$ at $(x, \lambda, \alpha)\}$.
(b) $\mathscr{H} = \{\alpha \in \mathbb{R}^k: \exists (x, \lambda) \in \mathbb{R} \times \mathbb{R}$ such that $G = G_x = G_{xx} = 0$ at $(x, \lambda, \alpha)\}$.
(c) $\mathscr{D} = \{\alpha \in \mathbb{R}^k: \exists (x_1, x_2, \lambda) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}, x_1 \neq x_2$ such that
$G = G_x = 0$ at $(x_i, \lambda, \alpha), i = 1, 2\}$.
(d) $\Sigma = \mathscr{B} \cup \mathscr{H} \cup \mathscr{D} =$ *transition set.*

The following remarks relate this definition to the codimension one bifurcation problems studied in §3(c).

Remarks 5.2. (i) We have grouped both simple bifurcation and the isola center into the set $\mathscr{B}$, since the recognition problem for each involves exactly

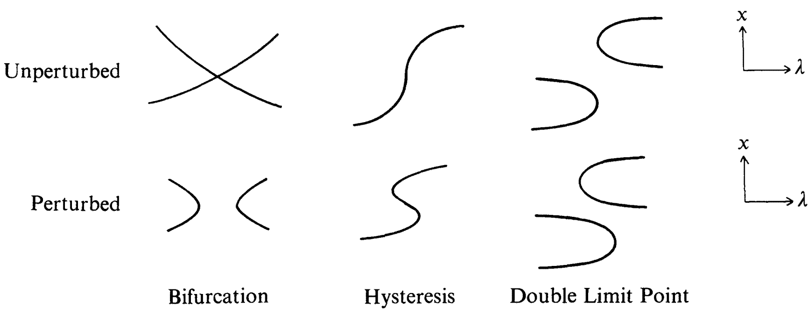


Figure 5.1. Nonpersistent phenomena.

the same equations; viz.,

$$g = g_x = g_\lambda = 0. \tag{5.1}$$

Only the nondegeneracy conditions are different for these two singularities, and we have ignored nondegeneracy conditions throughout in Definition 5.1.

(ii) Double limit points did not appear in §3(c) because they are quasi-global in the following sense: two distinct points are involved in the conditions defining $\mathscr{D}$. (By contrast, $\mathscr{B}$ and $\mathscr{H}$ only involve one point.) However, double limit points can occur in arbitrarily small perturbations of a degenerate singularity; in this sense this phenomenon is actually a local one. Consider the example $h(x, \lambda) = x^4 - \lambda - \varepsilon x^2$. If $\varepsilon = 0$, h has a generalized hysteresis point at the origin, while if $\varepsilon > 0$, h has a double limit point at $\lambda_0 = -3\varepsilon^2/16$. Thus, the global notion of double limit points is necessary in any local theory which handles small perturbations.

We now argue, at least heuristically, that each of the three sets in Definition 5.1 is described by a single scalar equation $\psi_i(\alpha_1, \ldots, \alpha_k) = 0$, where the subscript i equals $\mathscr{B}$, $\mathscr{H}$, or $\mathscr{D}$. Consider, for example, the defining equations for $\mathscr{B}$:

$$G(x, \lambda, \alpha) = G_x(x, \lambda, \alpha) = G_\lambda(x, \lambda, \alpha) = 0. \tag{5.2}$$

If we solve two of the equations in (5.2) for x and λ as functions of α and then substitute the result into the third equation, we obtain a single equation for α as claimed. Similar considerations apply to $\mathscr{H}$ and $\mathscr{D}$.

This analysis is subject to three caveats. First, it seems conceivable that the three scalar equations in (5.2) are not indpendent, which would of course spoil the argument. However, as it turns out, the hypothesis that G is a universal unfolding preculudes this possibility, at least in a suitably small neighborhood of the singularity. We shall not prove this assertion in general; we will however derive it explicitly for each of the specific examples we consider.

The second caveat is that the elimination process may introduce singularities into the defining equation ψ_i. For example, the bifurcation set of the unfolded singularity

$$x^2 - \lambda^3 + \alpha_1 + \alpha_2 \lambda$$

is the cusped curve

$$\left(\frac{\alpha_1}{2}\right)^2 = \left(\frac{\alpha_2}{3}\right)^3.$$

(See Exercise 5.1.)

The third caveat comes from the fact that we are working over the real numbers; it may happen that in eliminating x and λ certain inequalities among the α's appear. For example, consider the following hypothetical system where $k = 2$:

$$x^2 - \alpha_1 = 0, \qquad \lambda = 0, \qquad \alpha_2 = 0;$$

in eliminating x and λ from this system one obtains

$$\alpha_2 = 0, \qquad \alpha_1 \geq 0,$$

a half line in the plane. A second, less academic, example comes from the double limit point set of $x^4 - \lambda$. (This was hinted at above; see Chapter IV, §3 for more details.)

These ideas have the following simple but noteworthy consequence: for almost every $\alpha \in \mathbb{R}^k$ the bifurcation diagram

$$\{(x, \lambda): G(x, \lambda, \alpha) = 0\} \tag{5.3}$$

consists of nonintersecting regular curves whose only singularities are limit points. This may be seen by combining the following three facts:

(i) The bifurcation diagram (5.3) consists of nonintersecting regular curves if $\alpha \notin \mathscr{B}$.
(ii) The only singularities which appear on (5.3) are limit points if $\alpha \notin \mathscr{B} \cup \mathscr{H}$.
(iii) $\mathscr{B}$ and $\mathscr{H}$ are (possibly singular) surfaces in $\mathbb{R}^k$ of dimension $k - 1$.

We summarize the above discussion in the next theorem. Let us change coordinates so that G is a polynomial in all its arguments; this is possible because g has finite codimension. In the theorem, we refer to the following concepts from algebraic geometry (i.e., the study of polynomials and their zeros). An *algebraic variety* S in $\mathbb{R}^k$ is a set which can be expressed as the simultaneous zeros of a finite number of polynomial equations:

$$S = \{\alpha \in \mathbb{R}^k: P_i(\alpha_1, \ldots, \alpha_k) = 0, i = 1, \ldots, I\}.$$

Loosely speaking, the *codimension* of S is the smallest number of equations which may be used naturally in such a representation of S. In particular, a variety of *codimension one* is a hypersurface; however this hypersurface may have certain types of singularities such as self-intersections or cusps. A *semi-algebraic variety* is a subset of an algebraic variety which verifies certain additional polynomial inequalities, say

$$S' = \{\alpha \in S: Q_j(\alpha) \geq 0, j = 1, \ldots, J\}.$$

Theorem 5.3. *Let g: $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a polynomial map of finite codimension, and let $G(x, \lambda, \alpha)$ be a universal unfolding, also a polynomial. The set Σ of Definition* 5.1 *is a semialgebraic variety in $\mathbb{R}^k$ of codimension* 1.

It follows from this theorem that $\mathbb{R}^k \sim \Sigma$ has finitely many connected components. We don't prove either the theorem or this corollary in general, although we derive both in all specific cases we study. We do remark, however, that the equation describing the union $\Sigma = \mathscr{B} \cup \mathscr{H} \cup \mathscr{D}$ comes from the product

$$\psi_{\mathscr{B}}(\alpha)\psi_{\mathscr{H}}(\alpha)\psi_{\mathscr{D}}(\alpha) = 0,$$

where $\psi_i = 0$, $i = \mathscr{B}, \mathscr{H}$, or $\mathscr{D}$, describes the surfaces $\mathscr{B}, \mathscr{H}$, or $\mathscr{D}$, respectively. In addition, several inequalities may be needed to characterize Σ.

Exercises

5.1. Find the transition set for the universal unfolding $x^2 - \lambda^3 + \alpha_1 + \alpha_2 \lambda$.

5.2. Find the transition set for the universal unfolding $x^3 - \lambda x + \alpha_1 + \alpha_2 \lambda$.

5.3. Let the l-parameter unfolding $H(x, \lambda, \beta)$ factor through the k-parameter unfolding $G(x, \lambda, \alpha)$. That is, let

$$H(x, \lambda, \beta) = S(x, \lambda, \beta)G(X(x, \lambda, \beta), \Lambda(\lambda, \beta), A(\beta)),$$

where $S > 0$, $X_x > 0$, $\Lambda_\lambda > 0$. Show that $A: \mathbb{R}^l \to \mathbb{R}^k$ satisfies $A(\mathscr{B}_H) \subset \mathscr{B}_G$, $A(\mathscr{H}_H) \subset \mathscr{H}_G$, and $A(\mathscr{D}_H) \subset \mathscr{D}_G$, where $\mathscr{B}, \mathscr{H}, \mathscr{D}$ indicate the bifurcation, hysteresis, and double limit sets of Definition 5.1, subscripted by the appropriate unfolding.

5.4. Let $G: \mathbb{R} \times \mathbb{R} \times \mathbb{R}^k$ be a universal unfolding of some singularity. Show that for each $\alpha \in \Sigma$, the qualitative type of the bifurcation diagram $\{(x, \lambda): G(x, \lambda, \alpha) = 0\}$ may be changed by an arbitrarily small perturbation.
(*Discussion.*) Suppose $\alpha \in \mathscr{B}$. If $G(\cdot, \cdot, \alpha)$ is equivalent to $\pm x^2 \pm \lambda^2$, then the result is trivial to prove—we know from §3(c) that $\pm x^2 \pm \lambda^2$ has the stated property, and we may use the Universal Unfolding Theorem to deduce it for $G(\cdot, \cdot, \alpha)$. However $G(\cdot, \cdot, \alpha)$ might have a very degenerate singularity whose behavior is very difficult to analyze; how do we know that G still has this property? To show this, we ask the reader first to prove that if $\alpha \in \mathscr{B}$, then for small $\eta \neq 0$ the function

$$G(x, \lambda, \alpha) + \eta(x^2 + \lambda^2)$$

has a singularity equivalent to $\pm x^2 \pm \lambda^2$, and then to perturb $G + \eta(x^2 + \lambda^2)$ to obtain the desired result. The analysis for $\alpha \in \mathscr{H}$ and $\alpha \in \mathscr{D}$ is similar and is also left to the reader. (*Remark*: The conclusion of this exercise still holds even if we restrict to perturbations in the given unfolding.)

§6. Statement of the Main Geometric Theorem

For the most part, in this book we have formulated our results in terms of germs. With this formalism one may often avoid specifying explicit neighborhoods on which a theorem is valid; this can simplify the statement of results. The formalism works out best in cases where all calculations are performed in the neighborhood of some fixed point (x_0, λ_0); for example, this is the case with results about $\mathscr{P}(g)$, the higher-order terms associated with a given germ g. In such cases we have considered in our equivalences only diffeomorphisms (X, Λ) which fix (x_0, λ_0). However, in cases where we must consider diffeomorphisms which leave no point fixed, the formalism of germs is less satisfactory. Such transformations occur in Definition 1.1, the definition of one unfolding factoring through another; nonetheless, germ concepts are still adequate for that topic. In the present section, however, the germ formalism must be temporarily abandoned.

Let us briefly discuss what goes wrong with germs in the present context. Suppose $g(x, \lambda)$ is a germ of finite codimension with universal unfolding $G(x, \lambda, \alpha)$, $\alpha \in \mathbb{R}^k$. Define Σ as in Definition 5.1, and let W be an appropriate neighborhood of zero in $\mathbb{R}^k$. In loose terms, the main result of this section states that if α_1, α_2 belong to the same connected component of $W \sim \Sigma$, then $G(\cdot, \cdot, \alpha_1)$ and $G(\cdot, \cdot, \alpha_2)$ are equivalent to one another. In other words, for each pair α_1, α_2 in a given component of $W \sim \Sigma$, there is an appropriate equivalence transformation linking $G(\cdot, \cdot, \alpha_1)$ and $G(\cdot, \cdot, \alpha_2)$. Since none of these diffeomorphisms need leave any points fixed, it seems impossible to pin down the situation adequately with germ concepts.

Nevertheless, the result we seek does follow from local analysis. (We surely could not hope to prove the global equivalence of $G(\cdot, \cdot, \alpha_1)$ and $G(\cdot, \cdot, \alpha_2)$ on $\mathbb{R} \times \mathbb{R}$, even for α_1 and α_2 near the origin in $\mathbb{R}^k$.) We will choose a carefully constructed neighborhood V of the origin in $\mathbb{R} \times \mathbb{R}$, and then we will prove equivalence of $G(\cdot, \cdot, \alpha_1)$ and $G(\cdot, \cdot, \alpha_2)$ on V (for suitable α_1, $\alpha_2 \in W$, as above.) Because we consider diffeomorphisms (X, Λ) where Λ does not depend on x, it is most convenient to take V to be a *rectangular* neighborhood of (0, 0); i.e., a neighborhood of the form $V = U \times L$ where U and L are *closed* intervals. We will say a function $g\colon U \times L \to \mathbb{R}$ is C^∞ if g admits a C^∞ extension to an open neighborhood of $U \times L$.

We now begin the presentation of our main result. Let $g(x, \lambda)$ have a singularity of finite codimension at the origin. We first choose an appropriate neighborhood $U \times L$ of the origin on which to formulate this result. Specifically we will choose $U \times L$ such that

(a) g and g_x do not vanish simultaneously on $\partial(U \times L)$.
(b) g does not vanish on $(\partial U) \times L$ (the top and bottom faces). (6.1)

Why is this possible? First, since the ideal $\langle g, g_x \rangle$ has finite codimension, we may conclude from Corollary II,5.10 that the origin is an isolated solution to the pair of equations

$$g(x, \lambda) = g_x(x, \lambda) = 0.$$

Thus any sufficiently small rectangle containing the origin will satisfy (6.1a). To verify (6.1b) we observe that by finite codimension

$$g(x, 0) = ax^l + \mathcal{O}(x^{l+1})$$

for some integer l and some $a \neq 0$. In other words, the origin is an isolated zero of $g(x, 0)$. Let U be a small (closed) interval containing zero such that no other zero of $g(x, 0)$ belongs to U. Then g is nonzero on $\partial U \times \{0\}$; by continuity we may choose an interval L so small that g is nonzero on $(\partial U) \times L$, thus verifying (6.1b).

Let $G(x, \lambda, \alpha)$, $\alpha \in \mathbb{C}^k$ be a universal unfolding of g. Now $G(\cdot, \cdot, 0) = g$, which satisfies conditions (6.1a, b) above. By continuity we may choose a neighborhood of zero $W \subset \mathbb{R}^k$ such that for any $\alpha \in W$

(a) G and G_x do not vanish simultaneously on $\partial(U \times L)$.

(b) G does not vanish on $(\partial U) \times L$. (6.2)

In constructing the bifurcation, hysteresis, and double limit point sets associated to G, we modify Definition 5.1 by restricting (x, λ) to $U \times L$; for example we take

$$\mathscr{B} = \{\alpha \in W : \exists (x, \lambda) \in U \times L \text{ such that } G = G_x = G_\lambda = 0 \text{ at } (x, \lambda, \alpha)\}, \tag{6.3}$$

with similar changes for $\mathscr{H}$ and $\mathscr{D}$. Recall that the transition set Σ is $\mathscr{B} \cup \mathscr{H} \cup \mathscr{D}$.

Theorem 6.1. *Let g, G, U, L, W, and Σ be chosen as above. If α_1, α_2 belong to the same connected component of $W \sim \Sigma$, then there is a diffeomorphism $(X(x, \lambda), \Lambda(\lambda))$ mapping $U \times L$ onto itself and a positive function $S(x, \lambda)$ such that*

$$G(x, \lambda, \alpha_2) = S(x, \lambda)G(X(x, \lambda), \Lambda(\lambda), \alpha_1).$$

The diffeomorphism maps each edge of $U \times L$ onto itself.

In other words, the persistent bifurcation diagrams in the unfolding G are enumerated by the components of $W \sim \Sigma$. *A priori* it is possible that equivalent diagrams could occur in two different components of $W \sim \Sigma$, but we have not found any examples where this actually occurs.

When k is large, say $k \geq 4$, the determination of Σ is a nontrivial task. Even when k is small, the actual construction of the bifurcation diagrams associated to G, we modify Definition 5.1 by restricting (x, λ) to $U \times L$; for see in our discusson of the unfolding of the winged cusp in §8, this construction often proceeds more smoothly by first considering well-chosen transition diagrams corresponding to α in Σ.

We will prove Theorem 6.1 in §10 as a corollary of a more general result. A more direct proof is sketched in §9.

§7. A Simple Example: The Pitchfork

In this section we apply Theorem 6.1 to derive the information contained in Figure 1.5, Chapter I, §1 about perturbations of the pitchfork. (We will make a similar, but technically more complicated, application to the winged cusp in §8 below.)

Let $G(x, \lambda, \alpha, \beta) = x^3 - \lambda x + \alpha + \beta x^2$ be the universal unfolding of the pitchfork, and let $U, L \subset \mathbb{R}$ and $W \subset \mathbb{R}^2$ be neighborhoods verifying the conditions of Theorem 6.1. (In Exercise 7.1 below we guide the reader through one possible explicit choice for these neighborhoods.) Note that the only singularity of the unperturbed problem (i.e., solution of $g = g_x = 0$) occurs at the origin, and of course the choice of U, L, and W guarantees that for $\alpha \in W$ no singularity can escape across $\partial(U \times L)$. Thus the modification (6.1) of Definition 5.1 has no effect here.

We claim that there are no double limit points for the case at hand. We will use the following lemma to verify this.

Lemma 7.1. *If $h(x)$ is a polynomial of degree 3 or less such that $h = h_x = 0$ at two distinct points x_1 and x_2, then $h \equiv 0$.*

Proof. By performing a translation of axes, $x \to x - x_1$, we may assume that $x_1 = 0$. Then $h(0) = h_x(0) = 0$, which implies that $h(x)$ has the form $h(x) = ax^3 + bx^2$ for some coefficients $a, b \in \mathbb{R}$. Now the two equations $h(x_2) = h_x(x_2) = 0$ may be written as a matrix equation

$$\begin{pmatrix} x_2^3 & x_2^2 \\ 3x_2^2 & 2x_2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 0.$$

But the determinant of this matrix equals $-x_2^4 \neq 0$, which implies that $a = b = 0$. □

It follows from the lemma that it is not possible to satisfy $G = G_x = 0$ at two points x_1 and x_2 for fixed α, β, and λ, since G has degree 3 in x. This proves the claim. In symbols, $\mathscr{D} = \varnothing$.

In order to compute $\mathscr{B}$ and $\mathscr{H}$ we first calculate

$$\begin{aligned} G &= x^3 + \beta x^2 - \lambda x + \alpha, \\ G_x &= 3x^2 + 2\beta x - \lambda, \\ G_\lambda &= -x, \\ G_{xx} &= 6x + 2\beta. \end{aligned}$$

To determine $\mathscr{B}$, note that $G_\lambda = G_x = 0$ implies that $x = \lambda = 0$. From $G = 0$ we obtain the equation $\alpha = 0$. Hence

$$\mathscr{B} = \{(\alpha, \beta) \in W\colon \ \alpha = 0\}.$$

To determine $\mathscr{H}$, note that $G = G_x = G_{xx} = 0$ yields

$$\begin{aligned} x &= -\beta/3, \\ \lambda &= -\beta^2/3, \\ \alpha &= \beta^3/27. \end{aligned}$$

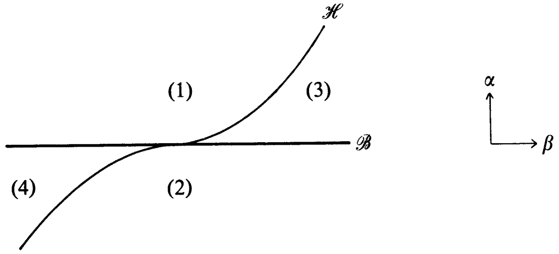


Figure 7.1. Nonpersistence set for the pitchfork $x^3 - \lambda x$.

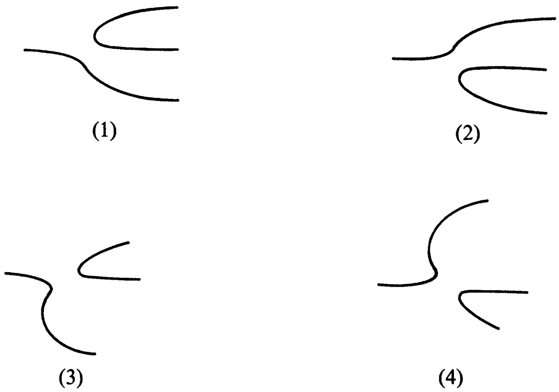


Figure 7.2. Persistent perturbations of the pitchfork. (Numbers refer to Figure 7.1.)

It follows that

$$\mathscr{H} = \{(\alpha, \beta) \in W\colon \ \alpha = \beta^3/27\}.$$

As shown in Figure 7.1, $W \sim \Sigma$ has four connected components.

According to Theorem 6.1, any two choices of parameters in the same component of $W \sim \Sigma$ give equivalent bifurcation diagrams. Thus to obtain the associated bifurcation diagram we could graph the solution set

$$\{(x, \lambda)\colon G(x, \lambda, \alpha, \beta) = 0\}$$

for one choice of parameters from each region. In practice such calculations can be shortened considerably by considering the dividing cases in Figure 7.1 (i.e., parameter values on $\mathscr{B}$ or $\mathscr{H}$) as was done in Chapter I, §1. In Figure 7.2 we indicate bifurcation diagrams for each region of Figure 7.1.

Exercise

7.1. (a) Show that $g(x, \lambda) = x^3 - \lambda x$ satisfies (6.1) when $U = [-1, 1]$ and $L = [-1, 1]$.
(b) Show that the universal unfolding of g, $G(x, \lambda, \alpha, \beta) = x^3 - \lambda x + \alpha + \beta x^2$, satisfies (6.2) for U, L as in (a) and $W = \{(\alpha, \beta) \in \mathbb{R}^2 : \alpha^2 + \beta^2 < 1\}$.

§8. A Complicated Example: The Winged Cusp

In this section we study in detail the universal unfolding

$$G(x, \lambda, \alpha, \beta, \gamma) = x^3 + \lambda^2 + \gamma\lambda x + \beta x + \alpha \tag{8.1}$$

of the winged cusp singularity. We use Theorem 6.1 to list the various persistent bifurcation diagrams which can be found in the universal unfolding G. The details of many calculations will be left to the reader.

Our procedure involves computing the subvariety Σ of $\mathbb{R}^3$ given in Definition 5.1. For ease of exposition we shall work globally in $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^3$, noting that technically our results are valid only on the neighborhoods U, L, and W of Theorem 6.1.

We begin by observing that the double limit point variety $\mathcal{D}$ is empty for G. This fact follows directly from Lemma 7.1, since G is a cubic polynomial in x. To find the varieties $\mathcal{B}$ and $\mathcal{H}$ note that

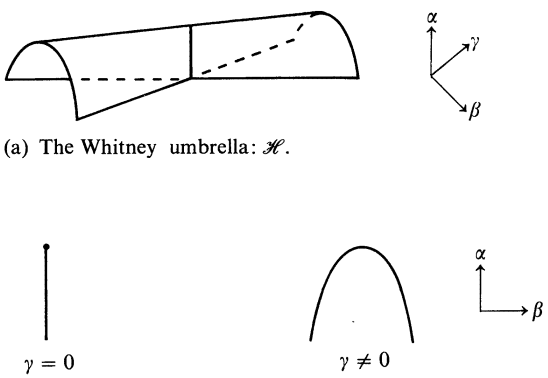
$$\begin{aligned} G_x &= 3x^2 + \gamma\lambda + \beta, \\ G_\lambda &= 2\lambda + \gamma x, \\ G_{xx} &= 6x. \end{aligned} \tag{8.2}$$

Our plan is first to calculate $\mathcal{B}$ and $\mathcal{H}$ separately, second to construct $\Sigma = \mathcal{B} \cup \mathcal{H}$, then to list the connected components of $\mathbb{R}^3 \sim \Sigma$ and finally to determine the associated bifurcation diagrams.

Recall that the hysteresis variety $\mathcal{H}$ is defined by $G = G_x = G_{xx} = 0$. It follows from (8.1) and (8.2) that

$$\mathcal{H} = \{\alpha\gamma^2 = -\beta^2; \alpha \leq 0\}. \tag{8.3}$$

$\mathcal{H}$ is the so-called "Whitney Umbrella" and is pictured in Figure 8.1(a). It is also convenient to graph slices of constant γ; this is done in Figure 8.1(b).



(a) The Whitney umbrella: $\mathcal{H}$.

(b) Slices of $\mathcal{H}$ for γ constant.

Figure 8.1. The hysteresis variety for the winged cusp.

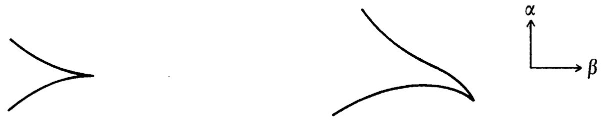


Figure 8.2. Slices of $\mathscr{B}$ for γ constant (the winged cusp).

The bifurcation variety $\mathscr{B}$ is determined by solving the equations $G = G_x = G_\lambda = 0$ simultaneously. For this it is easiest to give $\mathscr{B}$ in parametric form—parametrized by x and γ—as:

$$\begin{aligned} \beta &= -3x^2 + \gamma^2(x/2), \\ \alpha &= 2x^3 - \gamma^2(x^2/4). \end{aligned} \tag{8.4}$$

We can now graph slices of $\mathscr{B}$ given by γ constant. When $\gamma = 0$, we obtain the cusp $(\beta/3)^3 + (\alpha/2)^2 = 0$. When $\gamma \neq 0$, we claim that one obtains a cusp curve which is tilted down as indicated in Figure 8.2. Observe that the cusp point is defined by $d\alpha/dx = d\beta/dx = 0$ and occurs at $x = \gamma^2/12$. To see that the cusp tilts down, compute $d\alpha/d\beta = -x < 0$ at the cusp point. To complete the picture, show that the two nappes of the cusp never intersect.

We now discuss $\Sigma = \mathscr{B} \cup \mathscr{H}$. From the above discussion it seems most natural to graph slices of Σ for γ constant. This is done in Figure 8.3. Although the slices of Σ for γ and $-\gamma$ are identical we have given both in Figure 8.3 as both copies are necessary in order to determine the number of connected components in the complement of Σ. These connected components are enumerated in this figure. Let us discuss how to obtain this figure.

The graph of Σ for $\gamma = 0$ is easily recovered from the discussions of $\mathscr{B}$ and $\mathscr{H}$ above. For $\gamma = 0$, the picture of Σ is plausible, given that $\mathscr{H}$ is a parabola opening downward and $\mathscr{B}$ is a cusp curve which tilts down. In order to confirm this picture we check two points. First, we show that for each nonzero γ, $\mathscr{B} \cap \mathscr{H}$ consists of two points, one on each nappe; and second, we show that at the intersection of the left-hand nappe of $\mathscr{B}$ with $\mathscr{H}$, these two surfaces are tangent and cross one another.

The first point is to compute $\mathscr{B} \cap \mathscr{H}$. We do this by substituting the parametrization (8.4) for $\mathscr{B}$ into the equation (8.3) for $\mathscr{H}$, thus obtaining the equation for x

$$x^3(9x - \gamma^2) = 0. \tag{8.5}$$

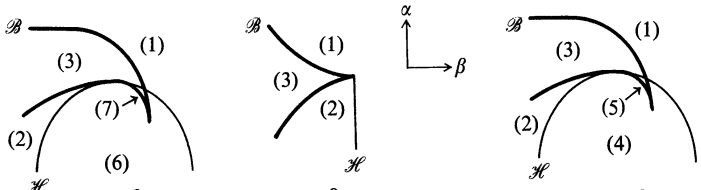


Figure 8.3. Slices of Σ for γ constant (the winged cusp).

The two solutions of (8.5) are the two intersections. One intersection occurs on each nappe of $\mathscr{B}$, since the cusp point occurs at $x = \gamma^2/12$ and this point is between the two solutions of (8.5). To verify the second item, consider the solution $x = 0$ in (8.5); this lies in the lower nappe of $\mathscr{B}$. By (8.4), $\alpha = \beta = 0$ at this point. We claim that at this point the unfolding

$$G(x, \lambda, 0, 0, \gamma) = x^3 + \lambda^2 + \gamma\lambda x$$

has a pitchfork singularity at (0, 0). This is easily checked using the solution to the recognition problem for the pitchfork (Proposition II,9.2). (See Exercise 8.1.) Moreover, the two-parameter unfolding $G(x, \lambda, \alpha, \beta, \gamma)$ of $G(x, \lambda, 0, 0, \gamma)$ is a universal unfolding; this follows from Proposition 3.4. Recalling our discussion of the pitchfork in §7, especially Figure 7.1, we see that $\mathscr{B}$ and $\mathscr{H}$ are tangent and cross one another at $\alpha = \beta = 0$.

Given the pictures in Figure 8.3 it is easy to enumerate the seven connected components of $\mathbb{R}^3 \sim \Sigma$ as is indicated on the figure. Note that regions 4 and 6 and that regions 5 and 7 are *not* connected in $\mathbb{R}^3 \subset \Sigma$.

To complete our discussion of the winged cusp, it remains to find the seven persistent perturbed bifurcation diagrams predicted by the analysis of Σ above. These seven diagrams are given in Figure 8.4, with numbers corresponding to regions in Figure 8.3. We shall use the existence of the pitchfork at $\alpha = \beta = 0, \gamma \neq 0$ heavily in our derivation of these figures.

We begin by graphing a particular diagram in region 1. Let $\beta = \gamma = 0$, $\alpha > 0$. The equation

$$G(x, \lambda, \alpha, 0, 0) = x^3 + \lambda^2 + \alpha = 0$$

can be solved easily, yielding

$$x(\lambda) = -(\lambda^2 + \alpha)^{1/3}. \tag{8.6}$$

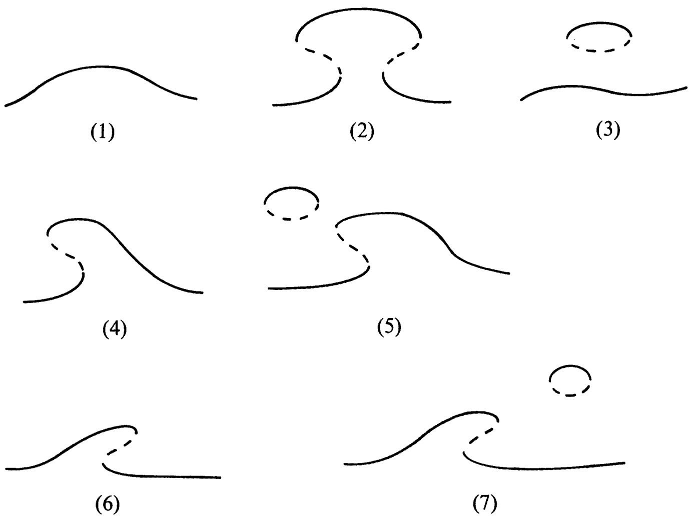


Figure 8.4. Persistent perturbations of the winged cusp. (Numbers refer to Figure 8.3.)

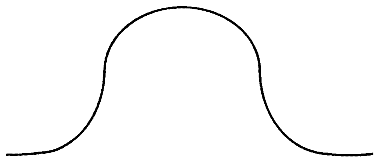


Figure 8.5. $x^3 + \lambda^2 + \alpha = 0, \alpha < 0$.

Note that $x(\lambda)$ is C^∞, since $\alpha > 0$. Thus one sees that the bifurcation diagrams corresponding to region 1 are typified by graph number 1 in Figure 8.4.

Next we show that if $\beta = \gamma = 0$, $\alpha < 0$ (i.e., the half line opposite the case above) then the bifurcation diagram is the one shown in Figure 8.5; this has two hysteresis points. Note that (8.6) is still valid, so there is one solution x for each λ. However, $G = G_x = G_{xx} = 0$ at $(x, \lambda) = (0, \pm\sqrt{|\alpha|})$ while $G_{xxx} \cdot G_\lambda \neq 0$ at these points. Using Proposition II,9.1 we see that $(0, \pm\sqrt{|\alpha|})$ are both hysteresis points for $G(x, \lambda, \alpha, 0, 0)$. (*Remark*: The points we are considering lie on the line of self-intersections of $\mathscr{H}$.) By Proposition 3.3, the one-parameter unfolding $G(x, \lambda, \alpha, \beta, 0)$ of $G(x, \lambda, \alpha, 0, 0)$ is a universal unfolding for either of these hysteresis points. Thus for $\beta \neq 0$, there will be either 2 or 0 limit points near each hysteresis point, depending on the sign of β. Since the bifurcation diagrams corresponding to region 1 have no limit points, it follows that those in region 2 must have four limit points and that the corresponding diagrams look like the mushroom of case 2 in Figure 8.4.

Now recall that $G(x, \lambda, 0, 0\ \gamma)$ has a pitchfork bifurcation at (0, 0) for $\gamma \neq 0$. The bifurcation diagrams for $G(x, \lambda, 0, 0, \gamma) = 0$ are given in Figure 8.6. Let us fix $\gamma < 0$ for further discussion. As we saw above, $G(x, \lambda, \alpha, \beta, \gamma)$ is a universal unfolding of the pitchfork $G(x, \lambda, 0, 0, \gamma)$.

Perturbation of the pitchfork with $\gamma < 0$ in Figure 8.6 leads to the four bifurcation diagrams shown in Figure 8.7; this may be derived from our discussion of the universal unfolding of the pitchfork in §7, especially Figure 7.2. We claim that the four diagrams in Figure 8.7 may be identified with regions in Figure 8.3 as follows:

Figure 8.7	Figure 8.3
(a)	(3)
(b)	(6)
(c)	(7)
(d)	(2)

We have already made the identification of (d) with region 2 in discussing Figure 8.5 above. Since one crosses $\mathscr{B}$ when moving from region (2) to region

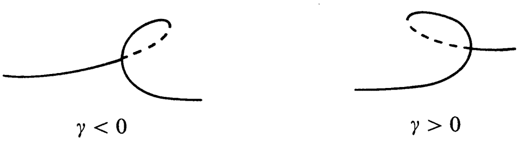


Figure 8.6. $x^3 + \lambda^2 + \gamma x\lambda = 0$.

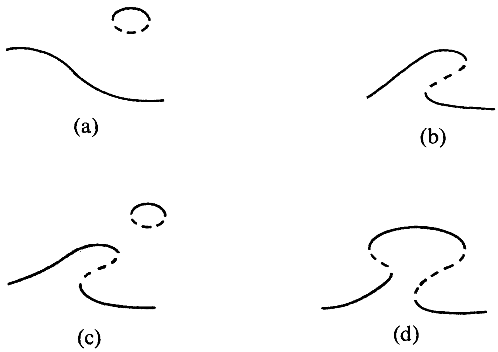


Figure 8.7. $G = 0$ for α, β small compared with $|\gamma|$, $\gamma < 0$. (Perturbation of Figure 8.6.)

(3) in Figure 8.3, the identification of (a) with region (3) follows. On the other hand, diagrams (a) and (b) correspond to the fat regions in Figure 7.1 (the unfolding of the pitchfork) while diagrams (c) and (d) correspond to the thin regions in that figure. This leads to the remaining two identifications: (b) with (6) and (c) with (7). A similar analysis with $\gamma > 0$ allows us to complete Figure 8.4.

EXERCISE

8.1. Show that $h = x^3 + \lambda^2 + \gamma\lambda x$ has a pitchfork singularity at $(x, \lambda) = (0, 0)$ when $\gamma \neq 0$. Show that $x^3 + \lambda^2 + \gamma\lambda x + \alpha + \beta x$ is a universal unfolding of h.

§9. A Sketch of the Proof of Theorem 6.1

We must show that if α and β are in the same component of $W \sim \Sigma$ then $G(\cdot, \cdot, \alpha)$ and $G(\cdot, \cdot, \beta)$ are equivalent on $U \times L$. The main task in this proof is to show that $G(\cdot, \cdot, \alpha)$ and $G(\cdot, \cdot, \beta)$ are equivalent *when α and β are sufficiently close* to one another and belong to the same component of $W \sim \Sigma$. The general case may be deduced from this special case by a simple connectivity argument, and in this sketch we concentrate only on the special case.

Suppose that in fact

$$G(x, \lambda, \alpha) = S(x, \lambda)G(X(x, \lambda), \Lambda(\lambda), \beta), \tag{9.1}$$

where S, X, Λ satisfy the usual restrictions. Then $G(x, \lambda, \alpha) = 0$ if and only if $G(X(x, \lambda), \Lambda(\lambda), \beta) = 0$. In other words, the diffeomorphism (X, Λ) maps the zero set of $G(\cdot, \cdot, \alpha)$ onto that of $G(\cdot, \cdot, \beta)$. It turns out that by far the most difficult and most informative step in the proof is to construct the diffeomorphism linking the two zero sets. We begin with this step.

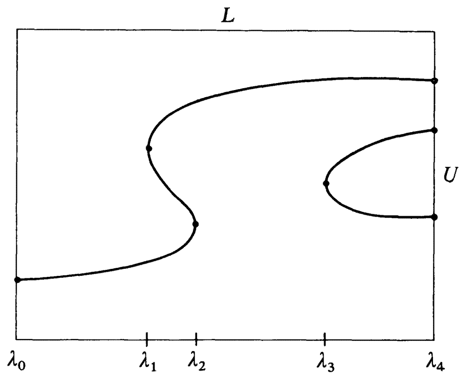


Figure 9.1. A possible zero set for $G(x, \lambda, \alpha)$.

In Figure 9.1, we indicate a possible zero set for $G(\cdot, \cdot, \alpha)$ when $\alpha \notin \Sigma$. Note that the only singularities in this figure are limit points with different λ coordinates, none of which lie on the boundary, and that no solution branch crosses $(\partial U) \times L$. We describe the construction of (X, Λ) in terms of this figure. Let us establish some notation. The three limit points in the figure are at $\lambda_1, \lambda_2, \lambda_3$; let λ_0 and λ_4 be the endpoints of the interval L. On any open interval $(\lambda_i, \lambda_{i+1})$ the number of solutions $x(\lambda)$ is constant, say J_i. For $\lambda_i < \lambda < \lambda_{i+1}$ define continuous solution branches $x_j(\lambda), j = 1, 2, \ldots, J_i$, where, for example, the solutions are enumerated in order of increasing x.

Now imagine that $G(\cdot, \cdot, \alpha)$, and hence Figure 9.1, is perturbed slightly. Our task is to map Figure 9.1 diffeomorphically onto the perturbed diagram. If the perturbation is small enough, then there will be exactly three limit points on the perturbed diagram, although at slightly different locations. Suppose these occur at μ_1, μ_2, μ_3. Also, on each interval (μ_i, μ_{i+1}) there will be the same number J_i of solutions on the perturbed diagram as solutions on $(\lambda_i, \lambda_{i+1})$ of the unperturbed diagram. (By convention we take $\mu_0 = \lambda_0$, $\mu_4 = \lambda_4$.) For $\mu_i < \lambda < \mu_{i+1}$ we denote the solutions on the perturbed diagram by $y_j(\lambda), j = 1, \ldots, J_i$.

Now let us consider what is required for a diffeomorphism $(X(x, \lambda), \Lambda(\lambda))$ to map Figure 9.1 onto the perturbed zero set. Recall that such diffeomorphisms preserve slices of constant λ. It is clear that the slices in Figure 9.1 which contain a limit point must be mapped into slices of the perturbed diagram which also contain a limit point. This gives rise to our first requirement, namely

$$\Lambda(\lambda_i) = \mu_i; \qquad i = 0, 1, 2, 3, 4. \tag{9.2}$$

To satisfy (9.2) is a simple interpolation problem, although the condition $\Lambda'(\lambda) > 0$ must be borne in mind. We do not dwell on this—suppose that a Λ satisfying (9.2) has been chosen.

Next we turn to the choice of X. To see what is involved, consider a λ slice *not* containing any of the limit points; more specifically, suppose

$\lambda \in (\lambda_i, \lambda_{i+1})$. For (X, Λ) to map Figure 9.1 onto the perturbed diagram, we must have

$$X(x_j(\lambda), \lambda) = y_j(\Lambda(\lambda)), \qquad j = 1, \ldots, J_i. \tag{9.3}$$

This is again an interpolation problem, subject to the auxiliary requirement $X_x(x, \lambda) > 0$, but there is an additional complication here—*X must vary smoothly with* λ. Actually it is easy to satisfy (9.3) with an X that varies smoothly in λ, provided λ stays away from the limit points. However, special care is required near a limit point, as at such points two of the branches $x_i(\lambda)$ meet one another. To deal with this difficulty, we first construct X near each of the limit points, and then we extend X to the intervals between limit points using well-established techniques from differential topology.

Let us elaborate on the first part of this argument—we construct X near a limit point λ_i as follows. Let Φ_i be a diffeomorphism mapping a portion of Figure 9.1 near this limit point into the zero set of the normal form for a limit point, namely $\pm x^2 \pm \lambda = 0$; and let Ψ_i do likewise for the perturbed diagram near μ_i. We obtain X near λ_i from $\Psi_i^{-1} \circ \Phi_i$.

The complete construction of (X, Λ) turns out to be fairly technical because there are quite a few details to be kept straight. However, we have presented all the ideas involved. Once we have the diffeomorphism from the zero set of $G(\cdot, \cdot, \alpha)$ onto that of $G(\cdot, \cdot, \beta)$, it is quite simple to argue that

$$G(x, \lambda, \alpha)/G(X(x, \lambda), \Lambda(\lambda), \beta)$$

is a smooth, nonvanishing function on $U \times L$, which yields the function S required in (9.1). (In fact we need only apply Proposition I,4.2—since α and β are not in $\mathscr{B}$, the functions G, G_x, and G_λ cannot vanish simultaneously.) This completes our sketch of the proof of Theorem 6.1. □

§10. Persistence on a Bounded Domain in a Parametrized Family

In this section we formulate a result which extends Theorem 6.1 in the following two respects:

(i) We consider an arbitrary parametrized family of bifurcation problems,

$$F(x, \lambda, \alpha) = 0, \tag{10.1}$$

where $\alpha \in \mathbb{R}^k$. By contrast, in §6 we considered only the universal unfolding of one singularity.

(ii) We discuss the solution set of (10.1) on a bounded domain in $\mathbb{R} \times \mathbb{R}$, with no assumption that this domain is small. By contrast, in §6 we worked in a carefully constructed neighborhood of one point.

The second of these two generalizations is by far the more significant. In particular, it suggests numerical procedures for exploring parameter space in applications. For example, Balakotaiah and Luss [1981, 1984] have conducted a global study of the CSTR using these methods.

First we establish the notation. Let $U, L \subset \mathbb{R}$ be closed intervals, and let $W \subset \mathbb{R}^k$ be a closed disk. Let $F: U \times L \times W \to \mathbb{R}$ be C^∞ in the sense that F may be extended to a C^∞ function on some open set in $\mathbb{R}^{k+2}$ containing $U \times L \times W$. We ask for what $\alpha, \beta \in W$ are $F(\cdot, \cdot, \alpha)$ and $F(\cdot, \cdot, \beta)$ globally equivalent on $U \times L$, where we define this term as follows. We call two bifurcation problems $f, g: U \times L \to \mathbb{R}$ *globally equivalent* on $U \times L$ if there exists a diffeomorphism $\Phi: U \times L \to U \times L$ of the form $\Phi(x, \lambda) = (X(x, \lambda), \Lambda(\lambda))$ and a positive function $S(x, \lambda)$ such that

$$g(x, \lambda) \coloneqq S(x, \lambda) f(X(x, \lambda), \Lambda(\lambda)), \tag{10.2}$$

where $X_x > 0$ on $U \times L$, $\Lambda' > 0$ on L, and Φ maps each face of $\partial(U \times L)$ onto itself. We summarize the latter condition as

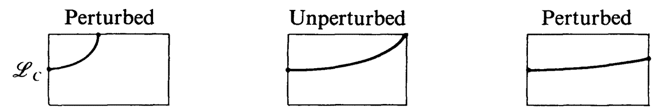
$$\Phi(\partial U \times L) = \partial U \times L, \qquad \Phi(U \times \partial L) = U \times \partial L. \tag{10.3}$$

As in §6, we proceed by considering persistent bifurcation diagrams; i.e., diagrams which remain qualitatively unchanged by small perturbations. In this section also, we list the possible sources of nonpersistence. Of course, the three sources listed in §6 (viz., bifurcation, hysteresis points, double limit points) are possible sources of nonpersistence in the present context. However, there is now a new source of nonpersistence, which is related to how bifurcation diagrams meet the boundary, $\partial(U \times L)$. In fact, it turns out that there are many possibilities (at first bewildering) for nonpersistence on the boundary. Thse are illustrated in Figures 10.1 and 10.2. We use the following concept in dividing the cases between the two figures: A phenomenon is *local* if only one point in $U \times L$ is involved, *global* otherwise. Already in §6 both categories occurred; bifurcation and hysteresis points are local, while double limits are global. (*Remark*: This use of global has no relation to "global" in global equivalence.)

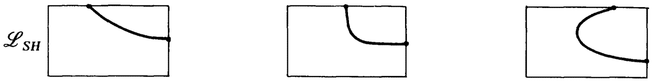
Let us discuss the motivation for these figures. We focus temporarily on the local case, which is a little more straightforward. Recall that the limit point is the only persitent singularity. The solution to the recognition problem for the limit point has two defining equations; viz.,

$$g = g_x = 0.$$

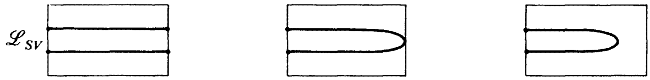
Any local phenomenon which has three or more defining conditions necessarily is nonpersistent. We saw this in §6, and the same analysis applies here. The phenomena in Figures 10.1 all have three defining conditions. Indeed this list is a complete enumeration of the local phenomena on the boundary which involve exactly three defining conditions. (*Remark*: We have already shown that equivalences preserve singularities, i.e., if g is singular, so is



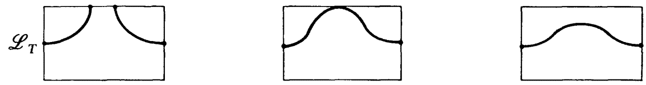
(a) Diagram meets corner.



(b) Limit point on horizontal face.



(c) Limit point on vertical face.



(d) Diagram tangent to boundary.

Figure 10.1. Local nonpersistence on the boundary.

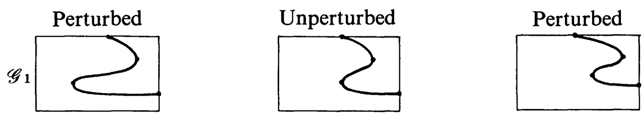
$Sg(X, \Lambda)$. Similarly, global equivalences preserve the boundary phenomena of Figures 10.1 and 10.2.)

Before deriving equations that characterize the phenomena in Figure 10.1, we note that the relation

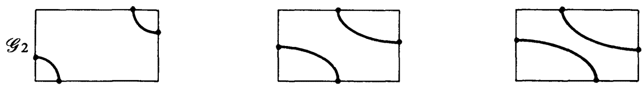
$$(x, \lambda) \in \partial(U \times L) \tag{10.4}$$

amounts to one scalar equation—(10.4) can hold only if

$$x \in \partial U \quad \text{or} \quad \lambda \in \partial L, \tag{10.5}$$



(a) Limit point with same λ-value as boundary point.



(b) Two boundary points with same λ-values.

Figure 10.2. Global nonpersistence on the boundary.

and either condition specifies a value of one scalar quantity. For both relations (10.5) to hold (i.e., for (x, λ), to be in a corner), two scalar equations must be satsified.

We now derive equations for Figure 10.1. Specifically $F(\cdot, \cdot, \alpha)$ exhibits the phenomena of Figures 10.1 (a, b, c, d) if and only if α belongs to the set

$$\mathscr{L}_C = \{\alpha \in W: \exists (x, \lambda) \in \partial U \times \partial L \text{ such that } F(x, \lambda, \alpha) = 0\}, \tag{10.6a}$$

$$\mathscr{L}_{SH} = \{\alpha \in W: \exists (x, \lambda) \in (\partial U) \times L \text{ such that } F = F_x = 0 \text{ at } (x, \lambda, \alpha)\}, \tag{10.6b}$$

$$\mathscr{L}_{SV} = \{\alpha \in W: \exists (x, \lambda) \in U \times (\partial L) \text{ such that } F = F_x = 0 \text{ at } (x, \lambda, \alpha)\}, \tag{10.6c}$$

$$\mathscr{L}_T = \{\alpha \in W: \exists (x, \lambda) \in (\partial U) \times L \text{ such that } F = F_\lambda = 0 \text{ at } (x, \lambda, \alpha)\}, \tag{10.6d}$$

respectively. Note that each of these sets involves exactly three defining conditions. For example, with $\mathscr{L}_{SV}$, one equation comes from $\lambda \in \partial L$ and two come from $F = F_x = 0$.

For the global case, $F(\cdot, \cdot, \alpha)$ exhibits the phenomena of Figure 10.2(a, b) if and only if α belongs to the set

$$\mathscr{G}_1 = \{\alpha \in W: \exists (x, \lambda) \in U \times L \text{ and } (x_0, \lambda) \in (\partial U) \times L \text{ with } x_0 \neq x \text{ such that } F(x_0, \lambda, \alpha) = 0 \text{ and } F = F_x = 0 \text{ at } (x, \lambda, \alpha)\}, \tag{10.6e}$$

$$\mathscr{G}_2 = \{\alpha \in W: \exists (x_1, \lambda), (x_2, \lambda) \in (\partial U) \times L \text{ with } x_1 \neq x_2 \text{ such that } F = 0 \text{ at } (x_i, \lambda, \alpha)\}, \tag{10.6f}$$

respectively. Here each set involves exactly four defining conditions. Four equations lead to nonpersistence for a global phenomenon involving two points (x_1, λ) and (x_2, λ) with the same λ-coordinate (Cf. §6). Figure 10.2 enumerates the global boundary phenomena with exactly four defining conditions.

For the reader's convenience we repeat the definition of the bifurcation, hysteresis point, and double limit point sets.

$$\mathscr{L}_{\mathscr{B}} = \{\alpha \in W: \exists (x, \lambda) \in U \times L \text{ such that } F = F_x = F_\lambda = 0\}, \tag{10.7a}$$

$$\mathscr{L}_{\mathscr{H}} = \{\alpha \in W: \exists (x, \lambda) \in U \times L \text{ such that } F = F_x = F_{xx} = 0\}, \tag{10.7b}$$

$$\mathscr{G}_{\mathscr{D}} = \{\alpha \in W: \exists (x_1, \lambda), (x_2, \lambda) \in U \times L \text{ such that } F = F_x = 0 \text{ at } (x_i, \lambda),\ i = 1, 2\}. \tag{10.7c}$$

We define

$$\begin{aligned} \mathscr{L} &= \mathscr{L}_{\mathscr{B}} \cup \mathscr{L}_{\mathscr{H}} \cup \mathscr{L}_C \cup \mathscr{L}_{SH} \cup \mathscr{L}_{SV} \cup \mathscr{L}_T, \\ \mathscr{G} &= \mathscr{G}_{\mathscr{D}} \cup \mathscr{G}_1 \cup \mathscr{G}_2, \\ \Sigma &= \mathscr{L} \cup \mathscr{G}. \end{aligned} \tag{10.8}$$

The main theorem in this subject states that if α and β are in the same connected component of $W \sim \Sigma$ then $F(\cdot, \cdot, \alpha)$ and $F(\cdot, \cdot, \beta)$ are globally equivalent on $U \times L$. The proof of this theorem is not particularly deep, but it is complicated. In order to reduce this complication, we only prove a simpler result; this is sufficient for our purposes. This simpler result still retains the essential flavor of the complete theorem. For the simpler result we assume that the family F has no zeros on $(\partial U) \times L$. (This situation occurs naturally if, for example, there is an *a priori* estimate which guarantees that for each fixed λ, all solutions must lie inside some bounded set.) Under this assumption five of the nine forms of nonpersistence cannot occur; specifically

$$\mathscr{L} = \mathscr{L}_{\mathscr{B}} \cup \mathscr{L}_{\mathscr{H}} \cup \mathscr{L}_{SV},$$

$$\mathscr{G} = \mathscr{G}_{\mathscr{D}}.$$

We shall prove the following result in §11.

Theorem 10.1. *Let $F\colon U \times L \times W \to \mathbb{R}$ be a family of bifurcation problems satisfying*

$$F(x, \lambda, \alpha) \neq 0 \quad \text{for } \forall x \in \partial U, \forall \lambda \in L, \forall \alpha \in W. \tag{10.9}$$

Let α and β be in the same connected component of $W \sim \Sigma$. Then $F(\cdot, \cdot, \alpha)$ and $F(\cdot, \cdot, \beta)$ are globally equivalent on $U \times L$.

Remark. Theorem 10.1 applies to an arbitrary parametrized family of bifurcation problems, $F\colon U \times L \times W \to \mathbb{R}$. In this generality there is no guarantee analogous to Theorem 5.1 that Σ is a hypersurface in $\mathbb{R}^k$ of codimension one. For example, one could imagine a degenerate situation in which F was independent of α so that $\Sigma = W$; in such a case, $\alpha \in W \sim \Sigma$ would never be satisfied. Of course, such artificial examples are unlikely in applications.

Our main use of Theorem 10.1 is to prove the local result, Theorem 6.1. This we now do.

PROOF OF THEOREM 6.1. Recall that in §6 we constructed W so that no limit points of G occur near the boundary of $U \times L$ and no zeros of G are found in $(\partial U) \times L$. It follows that (10.9) is satisfied and that $\mathscr{L}_{SV}$ is empty. Thus Theorem 6.1 is a corollary of Theorem 10.1. □

As we mentioned above, Theorem 10.1 and the more general result with nine nonpersistence sets (that we never actually formulated) suggest a numerical procedure for exploring parameter space by computing the various nonpersistence sets directly. Such a procedure is sometimes easier to implement than a direct numerical search for the bifurcation diagrams associated with F.

§11. The Proof of Theorem 10.1

We divide the proof of Theorem 10.1 into two parts. In the first part, Proposition 11.4, we show that the bifurcation diagram

$$\{(x, \lambda) \mid F(x, \lambda, \alpha) = 0\} \tag{11.1}$$

is the union of a finite number of branches when $\alpha \notin \Sigma$. Roughly speaking, a *branch* is a curve $(B(\lambda), \lambda)$ in (11.1) which connects a boundary or limit point of (11.1) with another such point. In this proposition, we also show that when α_0 and α_1 are in the same connected component of $W \sim \Sigma$ then the bifurcation diagrams corresponding to α_0 and α_1 decompose into the same number of branches which, moreover, may be identified in a natural way. This identification allows one to reconstruct, the bifurcation diagrams corresponding to α_0 and α_1, at least in a qualitative fashion. In fact, for most applications of Theorem 10.1, the information contained in Proposition 11.4 suffices.

The second part of the proof of Theorem 10.1, summarized by Proposition 11.5, is much more technical. Here we must construct a diffeomorphism $(X(x, \lambda), \Lambda(\lambda))$ of $U \times L$ which maps the bifurcation diagram corresponding to α_1 to the one corresponding to α_0. This construction requires certain interpolation and extension lemmas from differential topology. Here we sketch the structure of the proof, indicating where the technical points occur; we will not attempt to prove the needed lemmas. (Cf. Golubitsky and Guillemin [1973], p. 132 concerning proofs of such lemmas.)

In order to state Proposition 11.4 in a coherent manner, we need to make several definitions. Let $f: U \times L \to \mathbb{R}$ be a smooth mapping.

Definition 11.1. A *branch* of the bifurcation diagram $f(x, \lambda) = 0$ is a continuous function

$$C: [\Lambda_1, \Lambda_2] \to U,$$

which is smooth on the open interval (Λ_1, Λ_2) and satisfies:

(a) $f(C(\lambda), \lambda) \equiv 0$ in $[\Lambda_1, \Lambda_2]$; and
(b) either $\Lambda_i \in \partial L$ or $(C(\Lambda_i), \Lambda_i)$ is a limit point of f, $i = 1, 2$.

Definition 11.2. The smooth mapping f is *combinatorially regular* if:

(a) The only singularities of f are limit points and no limit point occurs in $\partial(U \times L)$.
(b) There are a finite number of limit points

$$P_i = (x_i, \lambda_i), \qquad i = 1, \ldots, s \quad \text{with} \quad \lambda_1 < \cdots < \lambda_s.$$

(c) Let $L = [\lambda_0, \lambda_{s+1}]$. There are a finite number of left-hand boundary points $Y_i = (y_i, \lambda_0)$, $i = 1, \ldots, m$ with $y_1 < \cdots < y_m$ and a finite number of right-hand boundary points $Z_j = (z_j, \lambda_{s+1})$, $j = 1, \ldots, n$ with $z_1 < \cdots < z_n$ comprising the solutions to $f = 0$ on ∂L.

(d) The bifurcation diagram $f = 0$ is the union of $(m + n + 2s)/2$ branches.

Remarks. (1) Part (d) of Definition 11.2 just reflects the fact that there are no solutions to $f = 0$ on $(\partial U) \times L$, the upper and lower boundaries of $U \times L$. Solution curves must begin and end at limit points or boundary points.

(2) The enumeration of branches follows trivially from the observations that each branch has a beginning and an end, each limit point lies on exactly two branches and each boundary point lies on precisely one branch.

(3) There are analogous definitions in the general case when Σ consists of nine sets.

If f is combinatorially regular then each branch of $f = 0$ is uniquely specified by its two endpoints, with one exception. The exception concerns isolas, as illustrated in Figure 11.1 If an isola contains just two limit points, say Q_1 and Q_2, then there are two distinct solution branches which have endpoints Q_1 and Q_2. In this case we distinguish between the two branches by referring to "upper" and "lower" branches.

In the next definition we describe what it means for two bifurcation diagrams to have the same combinatorial scheme. Loosely speaking, two bifurcation diagrams are combinatorially equivalent if their branches are in one-to-one correspondence. Now branches are specified by their endpoints, which are either limit points or boundary points. Thus to show that two bifurcation diagrams are combinatorially equivalent, we must first establish a correspondence between their respective limit points and their respective boundary points. Concerning boundary points, note that there is a natural order for enumeration of the left-hand boundary points on a bifurcation diagram—the point Y_1 with the smallest x-coordinate comes first, and so forth. Thus if two bifurcation diagrams have the same number of left-hand boundary points, this enumeration provides a natural correspondence between their respective left-hand boundary points. Similarly, if two bifurcation diagrams have the same number of right-hand boundary points, there is a natural correspondence between these points. It is perhaps less obvious that if two bifurcation diagrams have the same number of limit

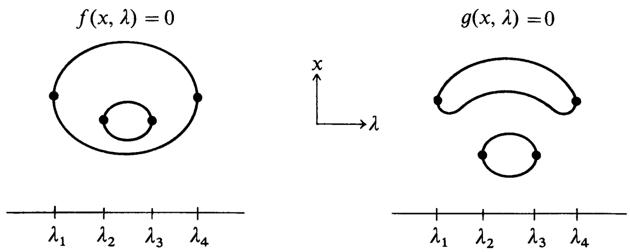


Figure 11.1. Bifurcation diagrams illustrating the need for part (c) in Definition 11.3.

points, there is again a natural correspondence between them, *provided* both diagrams are combinatorially regular. The idea here is that there are no double limit points in a combinatorially regular bifurcation diagram, so that its limit points may be uniquely enumerated in order of increasing λ-coordinate. This enumeration provides the desired correspondence.

Suppose that f and g define combinatorially regular bifurcation diagrams in $U \times L$ which have equal numbers of left-hand and right-hand boundary points and of limit points. In the following definition, if K^f is a boundary or limit point of f, we write K^g for the corresponding boundary or limit point of g.

Definition 11.3. Let $f, g: U \times L \to \mathbb{R}$ be smooth combinatorially regular mappings. We say that f and g are *combinatorially equivalent* if:

(a) The mappings f and g have the same number of limit points, say s; of left-hand boundary points, say m; and of right-hand boundary points, say n.
(b) The natural correspondence of boundary and limit points which exists as a consequence of (a) induces a bijection between branches of f and branches of g in the following sense. If there is a branch (K_1^f, K_2^f) then there is a branch (K_1^g, K_2^g) and conversely. If K_1^f and K_2^f are limit points spanning an isola then so are K_1^g and K_2^g.
(c) The bijection of branches of f to branches of g preserves the ordering of branches.

Remark. Let us elaborate on condition (c) of this definition. Let λ_{i-1} and λ_i be the λ-coordinates of consecutive limit points of f. Since branches of f cannot intersect one another in the interval $(\lambda_{i-1}, \lambda_i)$, we may order the branches of f on this interval according to increasing x-coordinates. Figure 11.1 shows a case when the bijection does not respect this ordering.

Proposition 11.4. (i) *If α is in $W \sim \Sigma$, then $F(\cdot, \cdot, \alpha)$ is combinatorially regular.*
(ii) *If α_0 and α_1 are in the same connected component of $W \sim \Sigma$, then $F(\cdot, \cdot, \alpha_0)$ and $F(\cdot, \cdot, \alpha_1)$ are combinatorially equivalent.*

It is an easy exercise to show that if f and g are globally equivalent combinatorially regular mappings, then f and g are combinatorially equivalent. In fact, the converse is also true.

Proposition 11.5. *Let $f, g: U \times L \to \mathbb{R}$ be combinatorially regular mappings such that $f \cdot g > 0$ near $(\partial U) \times L$. Then f and g are globally equivalent if and only if f and g are combinatorially equivalent.*

Remark. The proofs of Propositions 11.4 and 11.5 together consitute a proof of Theorem 10.1. Let us begin these proofs.

PROOF OF PROPOSITION 11.4. (i) Let α be in $W \sim \Sigma$ and let $f(x, \lambda) = F(x, \lambda, \alpha)$. To show that f is combinatorially regular we must show that (a)–(d) of Definition 11.2 are satisfied. Using Remark (1) after that definition and the assumption in Theorem 10.1 that F does not vanish on $(\partial U) \times L$, we see that (d) is satisfied. Since $\alpha \notin \mathscr{L}_{\mathscr{H}} \cup \mathscr{L}_{\mathscr{B}} \subset \Sigma$ it follows that the only singularities of f are limit points. Since $\alpha \notin \mathscr{L}_{SV}$ these limit points are not on the boundary and thus (a) is valid.

Since limit points are isolated (consider the normal form $\pm x^2 \pm \lambda$), and since $U \times L$ is compact, there can be at most a finite number of limit points for f; thus (b) is satisfied.

To show that (c) is satisfied we argue by contradiction. Suppose that there is an infinite number of boundary points; suppose for definiteness there are infinitely many on the left-hand boundary. Denote these points by $Y_i = (y_i, \lambda_0)$ $i = 1, 2, \ldots$. Since U is compact the y_i's have a convergent subsequence converging to y_∞. By continuity, $f(y_\infty, \lambda_0) = 0$. Since f is assumed not to vanish on $(\partial U) \times L$ we deduce that y_∞ is in the interior of U. The mean value theorem coupled with the fact that $f(y_i, \lambda) = 0$ for all i implies that $f_x(y_\infty, \lambda_0) = 0$. But this information contradicts the fact that $\alpha \notin \mathscr{L}_{SV}$; i.e., that there are no singularities on the vertical boundary.

(ii) We use a homotopy argument. Let $\alpha(t)$ be a curve in $W \sim \Sigma$ connecting $\alpha(0) = \alpha_0$ to $\alpha(1) = \alpha_1$. We will show that $F(\cdot, \cdot, \alpha_0)$ is combinatorially equivalent to $F(x, \lambda, \alpha(t))$ for every t. Of course, setting $t = 1$ yields the proposition.

Let $C: [\Lambda_1, \Lambda_2] \to U$ be a branch of $F(\cdot, \cdot, \alpha_0)$. We will prove there is unique continuous extension

$$B(t): [\Lambda_1(t), \Lambda_2(t)] \to U \tag{11.2}$$

of C where for each t, $B(t)$ is a branch of $F(\cdot, \cdot, \alpha(t))$. In particular, $\Lambda_1(t)$ and $\Lambda_2(t)$ are limit or boundary points of $F(\cdot, \cdot, \alpha(t))$ depending smoothly on t. This construction will also show that all limit and boundary points of $F(\cdot, \cdot, \alpha(t))$ are obtained by these functions $\Lambda_i(t)$ starting from limit and boundary points of $F(\cdot, \cdot, \alpha_0)$. It follows that for each t, $F(\cdot, \cdot, \alpha(t))$ is combinatorially equivalent to $F(\cdot, \cdot, \alpha_0)$, for the following reason: The natural identification of limit and boundary points is given by $\Lambda_i \to \Lambda_i(t)$, and if a branch (or two branches) connect Λ_1 with Λ_2 then there is a branch (or two branches) connecting $\Lambda_1(t)$ with $\Lambda_2(t)$. Moreover the x-coordinates along the branches cannot be interchanged by this construction.

The construction of $B(t)$ is made locally in t. In particular, we start with the branch $B(t_0)$ and find the unique extension for all t sufficiently close to t_0. The compactness of the interval $[0, 1]$ guarantees that we can patch together the local extensions and define $B(t)$ for all $t \in [0, 1]$.

First we prove that if $\Lambda_i(t_0)$ is a limit point then we can extend Λ_i to t near t_0 by the implicit function theorem. Define

$$\Phi(x, \lambda, t) = (F(x, \lambda, \alpha(t)), F_x(x, \lambda, \alpha(t))).$$

Then $\Phi(x, \lambda, t) = 0$ if and only if $F(\cdot, \cdot, \alpha(t))$ has a singularity. Moreover, since $F(\cdot, \cdot, \alpha(t))$ is combinatorially regular, its only singularities are limit points; thus $F_{xx} \cdot F_\lambda \neq 0$ at zeros of Φ. Now observe that the determinant of the Jacobian of Φ with respect to x, λ is $-F_{xx} \cdot F_\lambda \neq 0$. Thus the implicit function theorem implies the existence of a smooth curve $\Lambda_i(t)$ for t near t_0 such that $\Lambda_i(t_0)$ is the initial limit point and $\Lambda_i(t)$ is a limit point for $F(\cdot, \cdot, \alpha(t))$.

There is a similar construction for left- and right-hand boundary points. Let

$$\Psi(x, t) = F(x, \lambda_0, \alpha(t)).$$

Zeros of Ψ are left-hand boundary points of $F(\cdot, \cdot, \alpha(t))$. Moreover, $\Psi_x = F_x \neq 0$ at boundary points, since the combinatorial regularity of $F(\cdot, \cdot, \alpha(t))$ implies that there are no singularities on the boundary. Using the implicit function theorem we may construct a smooth function $y_i(t)$, with $y_i(0)$ any given left-hand boundary point of $F(\cdot, \cdot, \alpha_0)$, such that $(y_i(t), \lambda_0)$ is a left-hand boundary point for $F(\cdot, \cdot, \alpha(t))$.

We claim that these smooth functions are then globally defined on the interval $[0, 1]$ and, moreover, they induce a one-to-one correspondence between boundary and limit points of $F(\cdot, \cdot, \alpha_0)$ with those of $F(\cdot, \cdot, \alpha(t))$. To see this, observe that each such point for F at α_0 is connected to precisely one such point at $\alpha(t)$; this uses the uniqueness part of the implicit function theorem. Moreover, this correspondence of boundary and limit points is the natural, order-preserving one. For boundary points this is obvious, since a crossing of two such curves would contradict uniqueness. For limit points two such implicitly defined curves could, in principle, have their λ-values cross without their intersecting. However, this would imply the existence of a double limit point in F for some value of $\alpha(t)$, which we have ruled out by hypothesis.

Having defined the curve $\Lambda_i(t)$, we now construct the branch $B(t)$ whose existence we asserted in (11.2). As mentioned above, it suffices to construct $B(t)$ locally near each $t \in [0, 1]$ and then patch together. Consider such a point, say t_0. Recall that the branch $B(t_0)$ satisfies the equation

$$F(B(t_0)(\lambda), \lambda, \alpha(t_0)) \equiv 0$$

for $\lambda \in [\Lambda_1(t_0), \Lambda_2(t_0)]$. Note that $F_x \neq 0$ for all $\lambda \in (\Lambda_1(t_0), \Lambda_2(t_0))$ since branches do not go through singularities. Thus, for any fixed $\lambda \in (\Lambda_1(t_0), \Lambda_2(t_0))$ the equation

$$F(x, \lambda, \alpha(t)) = 0$$

may be solved uniquely from the initial condition $X = B(t_0)(\lambda)$, $t = t_0$. (The implicit function theorem applies, since $F_x \neq 0$.) This solution is valid for all t within some fixed ε of t_0; however, ε does depend on λ. The same construction works at the endpoints $\Lambda_i(t_0)$ if the endpoint is a boundary point since no singularities occur on the boundary. If $\Lambda_i(t_0)$ is a limit point, however, the fact that the width ε of the interval of solvability tends to zero near a limit point requires a different argument, as follows.

Suppose $B(t_0)$ has a limit point $\Lambda_i(t_0)$ as an endpoint. Limit points have two branches attached to them; let us assume, for definiteness, that $B(t_0)$ is the upper branch. Now at the limit point $F(\cdot, \cdot, \alpha(t_0))$ is equivalent to the normal form $\pm x^2 \pm \lambda$; for definiteness let us consider the signs $x^2 - \lambda$. The limit point is its own universal unfolding. Since $F(\cdot, \cdot, \alpha(t))$ is a one-parameter unfolding of $F(\cdot, \cdot, \alpha(t_0))$, there are equivalences such that

$$F(x, \lambda, \alpha(t)) = S(x, \lambda, t)\{X^2(x, \lambda, t) - \Lambda(\lambda, t)\} \tag{11.3}$$

near the limit point at t_0. Because $X_x > 0$ we may solve (11.3) for x on the upper branch; i.e., solve

$$X(x, \lambda, t) = \sqrt{\Lambda(\lambda, t)}$$

for x. Thus we obtain $B(t)$ near the limit point. For λ near $\Lambda_i(t)$ this construction of $B(t)$ matches with the construction of $B(t)$ by the implicit function theorem given in the last paragraph, because of uniqueness of solutions. By compactness of $[\Lambda_1(t_0), \Lambda_2(t_0)]$ there exists an ε, independent of λ, such that if t is ε close to t_0 then the branch $B(t)$ exists. □

Sketch of proof of proposition 11.5. We assume that $f, g: U \times L \to \mathbb{R}$ are both combinatorially regular and combinatorially equivalent. We show that f and g are globally equivalent on $U \times L$. Suppose there exists a diffeomorphism $(X(x, \lambda), \Lambda(\lambda))$ on $U \times L$ mapping the zero set of f to the zero set of g. Then we claim that

$$S(x, \lambda) = f(x, \lambda)/g(X(x, \lambda), \Lambda(\lambda))$$

is defined and C^∞ on $U \times L$ and bounded away from zero. This follows from Chapter I, Proposition 3.2 and the observation that combinatorial regularity implies that $\nabla f \neq 0$ on $f = 0$ and similarly for g. Moreover, $S > 0$ since $f \cdot g > 0$ near $(\partial U) \times L$.

We sketch the construction of (X, Λ) in several steps. Let $\lambda_1 < \cdots < \lambda_s$ be the limit points of f and let $\mu_1 < \cdots < \mu_s$ be the limit points for g. Let $L = [\lambda_0, \lambda_{s+1}]$. The first step is the construction of a diffeomorphism $\Lambda: L \to L$ satisfying $\Lambda(\lambda_i) = \mu_i, i = 1, \ldots, s$, and $\Lambda(\partial L) = \partial L$. Then $g(x, \Lambda(\lambda))$ has limit points at the same values of λ as f. Henceforth we assume that the limits points of f and g have the same λ-coordinates—because equivalence is a transitive relation, there is no loss of generality in this assumption.

Let the limit points for f and g with λ-coordinate λ_i be (x_i, λ_i) and $(\tilde{x}_i, \lambda_i)$, respectively. Now construct an orientation preserving diffeomorphism $X(\cdot, \lambda_i): U \to U$ for which $X(x_i, \lambda_i) = \tilde{x}_i$ and $X(\partial U, \lambda_i) = \partial U$. One may extend X to $U \times L$ such that $X_x > 0, X(\partial(U \times L)) \subset \partial U, X(x, \lambda) \equiv x$ when $|\lambda - \lambda_i| > \varepsilon$ where $\varepsilon < \min(\lambda_i - \lambda_{i-1}, \lambda_{i+1} - \lambda_i)/2$. Thus $f(x, \lambda)$ and $g(X(x, \lambda), \lambda)$ both have limit points at (x_i, λ_i). Making s such constructions allows us to assume, without loss of generality, that f and g have the same limit points.

We next make a similar construction moving the boundary points of f to those of g; here we require that $X(x, \lambda) \equiv x$ when $\lambda_0 + \varepsilon < \lambda < \lambda_s - \varepsilon$. Thus, we may assume that f and g have the same boundary and limit points.

The penultimate step is to use the normal form theorem for limit points (which is proved by finding a strong equivalence) to construct locally about a given limit point a mapping X such that $(X(x, \lambda), \lambda)$ maps the zero set of f to that of g near the limit point. We may extend X to be the identity on some slightly larger neighborhood of the limit point. Making $s + 2$ such constructions allows us to assume that f and g have the same limit and boundary points and the same zero sets near the limit points. Here we use the combinatorial equivalence of f with g to conclude that the limit points of f and g both point in the same directions, subcritical or subcritical.

Finally, we construct a diffeomorphism (X, λ) mapping a fixed branch of f to the corresponding branch of g. This is possible since (X, λ) may be chosen to be the identity near limit points. We must perform this construction once for each branch, taking care not to move apart branches that have already been identified. Since branches do not intersect except at limit points, this construction is possible. □

§12. The Path Formulation

The following thesis is the underlying theme of this section: Every bifurcation problem $g(x, \lambda)$ in $\mathscr{E}_{x,\lambda}$ may be viewed as a path in the universal unfolding space of one of the cuspoids of elementary catastrophe theory. Making the correspondence between bifurcation theory and catastrophe theory gives us a different view of a bifurcation diagram and its perturbations which is useful in certain contexts. In particular, this path formulation is helpful in finding organizing centers.

We divide this section into three parts:

(a) A comparison of catastrophe theory, singularity theory, and bifurcation theory.
(b) The path formulation.
(c) Elementary bifurcations and the cusp catastrophe.

As part of subsection (a) we discuss the similarities and differences between elementary catastrophe theory, singularity theory and steady-state bifurcation theory.

(a) A Comparison of Catastrophe Theory, Singularity Theory, and Bifurcation Theory

As a mathematical topic, catastrophe theory is the study of the local structure of critical points of real-valued functions f in $\mathscr{E}_n$. In this section we use the term "singularity theory" to refer to the study of the local structure of the zeros of mappings $h: \mathbb{R}^n \to \mathbb{R}^n$. In this book we have used the term "bifurcation theory" to refer to the study of the local structure of the zeros of mappings $g: \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ depending on a distinguished parameter.

These three subjects are intimately related; there are common ideas and techniques in their study. This is not surprising since the critical points of f are found by looking for zeros of the mapping $h = \nabla f$, and bifurcation problems g are just one-parameter families of mappings h. The path formulation for bifurcation problems is related to this last observation. There are some differences, however. The natural equivalence relation in catastrophe theory is right equivalence. Two functions $f_1, f_2 \in \mathscr{E}_x$ are *right-equivalent* if there exists a diffeomorphism germ $X(x)$ with $X(0) = 0$ and a constant K satisfying

$$f_1(x) = f_2(X(x)) + K.$$

The natural equivalence relation in singularity theory is John Mather's notion of contact equivalence. Two mappings h_1 and h_2 are *contact equivalent* if there exists a diffeomorphism germ $X(x)$ with $X(0) = 0$ and a nonsingular $n \times n$ matrix $S(x)$ depending smoothly on x such that

$$h_1(x) = s(x) \cdot h_2(X(x)).$$

This is the most general set of equivalences of mappings h which preserve the structure of the zeros of h. Our notion of equivalence for bifurcation problems is a one-parameter version of contact equivalence specialized to the case $n = 1$. In addition, because we are interested in the linearized stability of solutions (cf. Chapter I, §4), we have restricted $S(0)$ and $X_x(0)$ to be positive; but these are minor points.

In each category we can analyze the recognition problem and find universal unfoldings in much the same way as we have described in Chapters II and III here. John Mather [1969a, 1969b] first proved the unfolding theorem for contact equivalence and then for right equivalence. Martinet [1977] reworked Mather's proofs into a nice geometric form, and our proof of the unfolding theorem for bifurcation problems (to be given in Volume II) is an adaptation of Martinet's proof. The important point here is that unfolding theory is much the same in each category; given our purpose, we have described universal unfoldings from the point of view of bifurcation theory.

When $n = 1$ the clear separation discussed above between catastrophe theory and singularity theory becomes blurred. In both cases the functions $f: \mathbb{R} \to \mathbb{R}$ and the mappings $h: \mathbb{R} \to \mathbb{R}$ are elements of $\mathscr{E}_x$. However, the equivalence relations in the two categories are different. Nevertheless, it turns out that the unfolding theory in the two categories have identical structure. In the first place, each function $f(x)$ of finite codimension is right-equivalent to $\pm x^m$ for some m. Similarly, each mapping $h(x)$ of finite codimension is contact equivalent to $\pm x^m$ for some m. (The normal forms $\pm x^m$ are called *cuspoids* in elementary catastrophe theory.) In the second place, the forms of the universal unfolding of x^m in the two categories are related in a way which we now describe.

In catastrophe theory, the cuspoid x^{m+1} has codimension $m - 1$; a universal unfolding of x^{m+1} with respect to right-equivalence is:

$$F(x, \alpha) = x^{m+1} + \alpha_{m-1}x^{m-1} + \cdots + \alpha_1 x.$$

In singularity theory, the cuspoid x^m (one lower degree) has the same codimension $m - 1$; a universal unfolding of x^m with respect to contact equivalence is:

$$H(x, \beta) = x^m + \beta_{m-2}x^{m-2} + \cdots + \beta_1 x + \beta_0.$$

Upon differentiation of the first unfolding we obtain, apart from some trivial rescaling of parameters, the second unfolding.

In our discussion below of the path formulation for bifurcation problems we shall use the universal unfoldings of the cuspids with respect to contact equivalence. Nevertheless, the pictures of catastrophe theory, which are generated by solving $\partial f/\partial x = 0$, are precisely the same in the singularity theory category. The only difference is that in the latter case, we solve the equation $H = 0$.

The following remark is an aside for those readers interested in the controversies surrounding catastrope theory. One criticism of various applications of catastrophe theory is that for many applications, there did not exist a potential function. However, as long as the reduction to $n = 1$ is appropriate, we believe this criticism is a red herring. We could just as well use contact equivalence and obtain the same set of pictures. Moreover, when $n = 1$ the correspondence between catastrophe theory and singularity theory can be made in either direction, through differentiation or integration, as appropriate. In other words, for $n = 1$, potential functions always may be constructed.

(b) The Path Formulation

We now relate bifurcation problems g in one state variable ($n = 1$) to a path through the universal unfolding of a cuspoid. Suppose $g(x, \lambda)$ has finite codimension. Then $g(x, 0) = ax^{m+1} + \cdots$ for some m where $a \neq 0$. (This fact is not hard to prove. For the interested reader it will be proved in Lemma IV,2.4(a).) We assume for convenience that $a > 0$. Now $g(x, \lambda)$ may be viewed as a one-parameter unfolding (singularity theory context) of the mapping $g(x, 0)$. Thus $g(x, \lambda)$ may be factored through the universal unfolding (singularity theory context) for the cuspoid x^{m+1}; in symbols

$$\begin{aligned} &g(x, \lambda) \\ &= S(x, \lambda)[X(x, \lambda)^{m+1} + A_{m-1}(\lambda)X(x, \lambda)^{m-1} + \cdots + A_1(\lambda)X(x, \lambda) + A_0(\lambda)] \end{aligned} \tag{12.1}$$

for some functions S, X, and A_j. Using the mappings S and X we see that the bifurcation problem $g(x, \lambda)$ is equivalent to the normal form

$$h(x, \lambda) = x^{m+1} + A_{m-1}(\lambda)x^{m-1} + \cdots + A_1(\lambda)x + A_0(\lambda). \tag{12.2}$$

This normal form has the special property that all the λ-dependence is isolated in the coefficients of low powers of x.

In (12.2) we have identified the bifurcation problem g with an integer m and a path in the m-dimensional parameter space of the universal unfolding of x^{m+1}; namely

$$\lambda \to (A_0(\lambda), A_1(\lambda), \ldots, A_{m-1}(\lambda)). \tag{12.3}$$

We refer to this identification as the *path formulation* of bifurcation problems.

Let us show that this identification can be extended to perturbations. Specifically, we identify universal unfoldings of bifurcation problems with families of paths in the unfolding space of the cuspoids.

Let $h(x, \lambda)$ be the bifurcation problem (12.2). Let $G(x, \lambda, \alpha)$ be a universal unfolding (as a bifurcation problem) of h, depending on k parameters. Let us regard G as a $(k + 1)$-parameter unfolding of $G(x, 0, 0)$. Then the unfolding theorem (singularity theory context) for the cuspoid states that

$$\begin{aligned} G(x, \lambda, \alpha) = S(x, \lambda, \alpha)[&X(x, \lambda, \alpha)^{m+1} + A_{m-1}(\lambda, \alpha)X(x, \lambda, \alpha)^{m-1} \\ &+ \cdots + A_0(\lambda, \alpha)], \end{aligned} \tag{12.4}$$

where $X(x, \lambda, 0) \equiv x$, $S(x, \lambda, 0) \equiv 1$, and $A_j(\lambda, 0) = A_j(\lambda)$.

From (12.4) we see that the universal unfolding G is equivalent to the following universal unfolding (bifurcation theory context) of h:

$$H(x, \lambda, \alpha) = x^{m+1} + A_{m-1}(\lambda, \alpha)x^{m-1} + \cdots + A_0(\lambda, \alpha). \tag{12.5}$$

As with (12.2), we extract from (12.5) the k-parameter family of paths

$$\lambda \to (A_0(\lambda, \alpha), A_1(\lambda, \alpha), \ldots, A_{m-1}(\lambda, \alpha))$$

through the universal unfolding of the x^{m+1}. In other words, the universal unfolding of any bifurcation problem (of finite codimension, in one variable) may be identified with a parametrized family of paths through the universal unfolding of the cuspoid.

(c) Elementary Bifurcations and the Cusp Catastrophe

By the *cusp catastrophe* we mean the universal unfolding (singularity theory context) of the cuspoid x^3. We write this unfolding as

$$x^3 - Bx + A = 0. \tag{12.6}$$

The universal unfoldings of all bifurcation problems $h(x, \lambda)$ which satisfy

$$h(x, 0) = x^3$$

Table 12.1. Elementary Paths Through Cusp.

Normal Form	Universal Unfolding	$A(\lambda)$	$B(\lambda)$
Hysteresis	$x^3 + \lambda - \alpha x$	λ	α
Pitchfork	$x^3 - \lambda x + \alpha_1 + \alpha_2\lambda$	$\alpha_1 + \alpha_2\lambda$	λ
Winged Cusp	$x^3 + \lambda^2 + \alpha + \beta x + \gamma x\lambda$	$\lambda^2 + \alpha$	$-(\beta + \gamma\lambda)$

may be written as parametrized families of paths through the cusp. These bifurcations include the hysteresis point, the pitchfork, and the winged cusp. In Table 12.1 we present these paths explicitly. (Also see the cubic bifurcation problems of Chapter V.)

For the examples in Table 12.1, we give the pictures associated with these paths and another method for seeing the perturbed bifurcation. We begin by describing the geometry of (12.6). See Figure 12.1.

In Figure 12.1, we have indicated the projection of the cusp surface defined by (12.6) onto the unfolding space, the AB-plane. In the AB-plane there is a separation given by the cusp curve

$$\left(\frac{A}{2}\right)^2 = \left(\frac{B}{3}\right)^3.$$

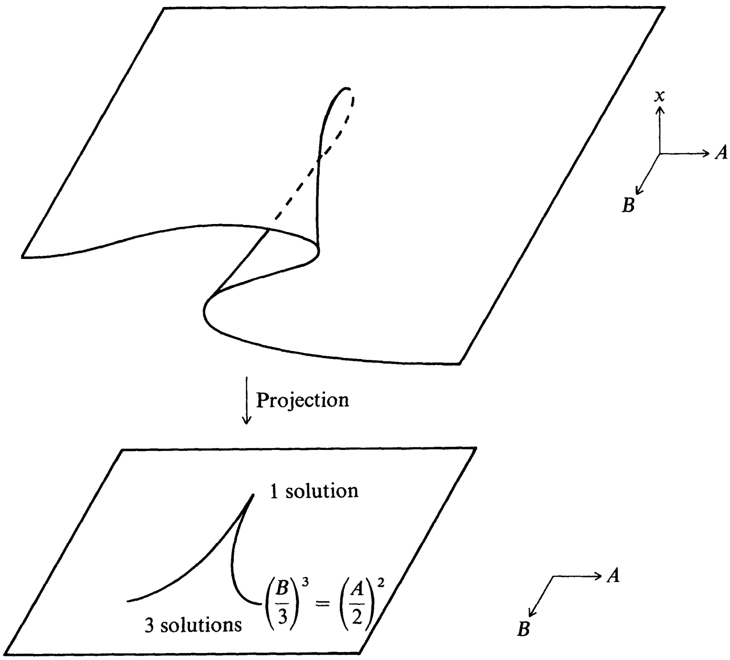


Figure 12.1. Geometry of the cusp catastrophe.

For points (A, B) inside this curve, the associated cubic has three distinct real roots; this fact is indicated in Figure 12.1 by the three points on the cusp surface lying above (A, B). For points outside the cusp, the associated cubic has one real root and the cusp surface has one point lying above (A, B).

The paths for each of the unperturbed bifurcation problems of Table 12.1 are given in Figure 12.2.

There is a simple way of recovering a bifurcation diagram from the path through the cusp given by lifting the path to the cusp surface. One has to be careful to remember that the path is parametrized by λ and that the path may traverse the same image in the AB-plane more than once, cf. the winged cusp. In Figure 12.3 we have given the liftings for the paths in Figure 12.2.

From the topologist's point of view, the only time a bifurcation diagram which corresponds to a path through the cusp can have a singularity is when the path intersects the cusp curve $(B/3)^3 = (A/2)^2$. So we should try to understand the perturbed paths listed in Table 12.1 in terms of these intersections. Typical perturbed paths are given in Figure 12.4. For the hysteresis point the perturbations contained in the universal unfolding are obtained by translating the line in the B-direction. For the pitchfork we can both translate (path (a) in the figure) and rotate (path (b)) the path. The universal unfolding theorem states that no new phenomena can be obtained (up to equivalence) by any other perturbation. This gives a geometric explanation of why the pitchfork has codimension 2. (In Exercise 12.1 we ask the reader to show that the path in Figure 12.5 leads to the bifurcation diagram in that figure.)

The perturbations of the winged cusp, given in Figure 12.4, provide some insight into the structure of the universal unfolding of the winged cusp. First note from Figure 12.3 that the perturbed paths are (possibly degenerate) parabolas whose axes of symmetry are parallel to the A-axis. Since such parabolas are defined by three parameters—two for the vertex and one for the latus rectum—one has geometric confirmation of the fact that the winged cusp has codimension 3. The reader should reconstruct the bifurcation diagrams corresponding to the parabolas listed in Figure 12.4 and convince himself that these diagrams correspond to five of the seven persistent perturbations of the winged cusp which are given in Figure 8.4.

We end this section with two remarks about the path formulation—one positive and one negative. The negative point stems from the fact that the diagrams associated with regions 5 and 7 of Figure 8.4 are missing in the paths of Figure 12.4. These two persistent diagrams correspond to a parabola

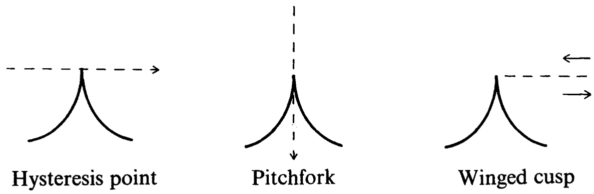


Figure 12.2. Pictures of paths through the cusp.

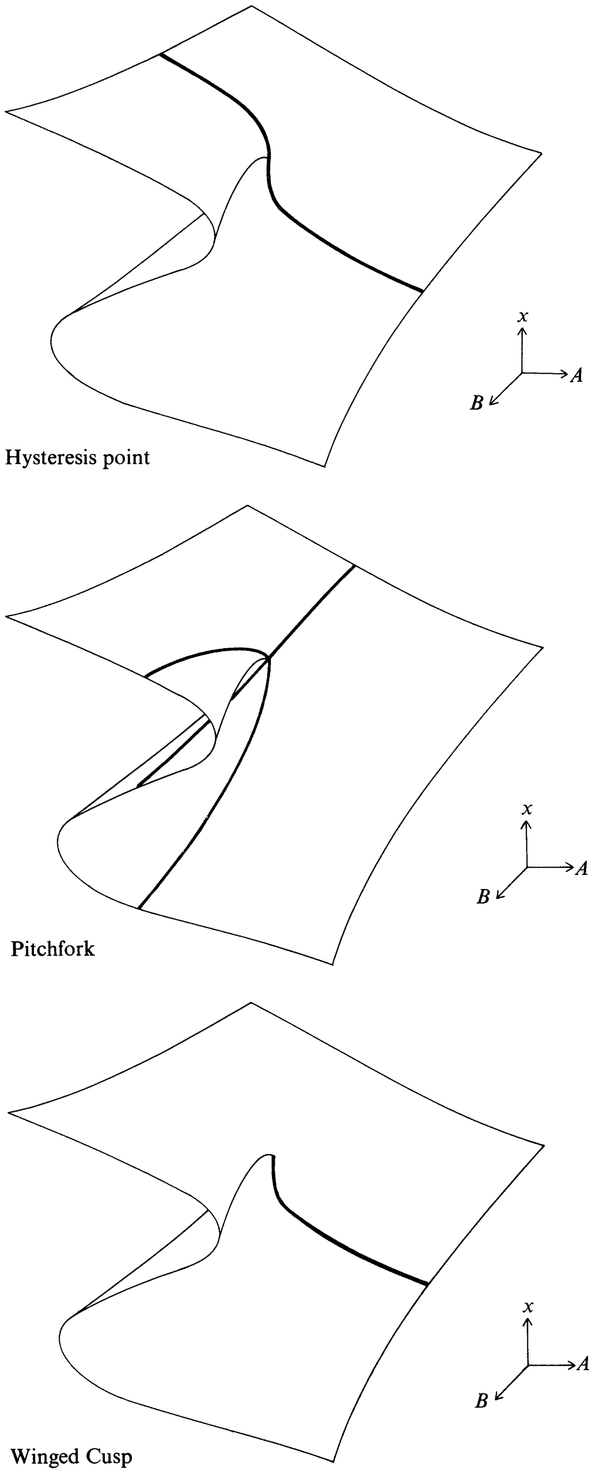


Figure 12.3. Liftings of paths through the cusp.

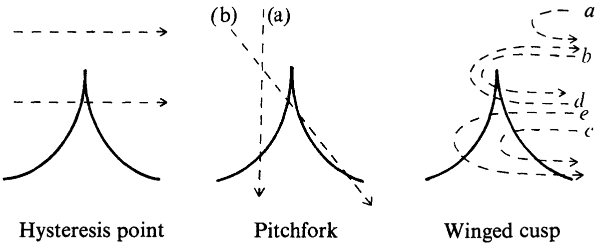


Figure 12.4. Perturbations of paths through the cusp.

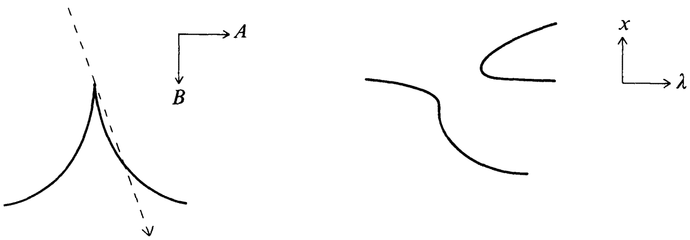


Figure 12.5. Perturbations of the pitchfork.

like that in Figure 12.6, transversed in either direction. It is probably not obvious to the reader that a parabola whose axis is parallel to the A-axis can intersect the cusp curve four times—once on the left-hand nappe and three times on the right-hand nappe. Yet this is exactly what our analysis of §8, using the bifurcation and hysteresis varieties, proves. In other words, the the path formulations can be misleading unless we are careful.

The positive remark is that the path formulation is often helpful in finding organizing centers. For example, suppose we have a family of bifurcation diagrams each of which contains at most three solutions x for each λ. Then the bifurcation diagrams correspond to paths through the cusp. By staring at these paths we can sometimes pick out one path which has all of the paths in the family as small perturbations. This (presumably degenerate) path is a good candidate for an organizing center for the problem. Indeed this was the method by which Golubitsky and Keyfitz [1980] originally found the organizing center for the CSTR (Cf. Chapter I, §2).

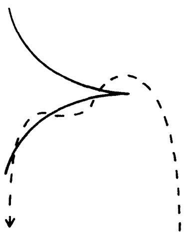


Figure 12.6. "Difficult" perturbation of the winged cusp.

Exercise

12.1. Consider the path $(A(\lambda), B(\lambda)) = (\lambda, \lambda)$ in the universal unfolding of the cusp singularity $x^3 - Bx + A$. Show that this path and the associated bifurcation diagram

$$x^3 - \lambda x + \lambda = 0$$

are the ones pictured in Figure 12.5. In particular, note that there is a hysteresis point at the origin.

Bibliographical Comments

The unfolding theorem was conjectured by Thom and proved by Mather [1969a]. Our results in §§1–4 are primarily an adaptation of this work to the context of bifurcation problems. Our proof of the unfolding theorem, see Golubitsky and Schaeffer [1979a], is modeled on the proof of Mather's theorem in Martinet [1982]. A more general version of the unfolding theorem will be proved in Volume II. The methods discussed in §§5–12 were first presented in Golubitsky and Schaeffer [1979a].

CASE STUDY 1
The CSTR

In Chapter I, §2 we outlined a program of analysis for the continuous flow stirred tank chemical reactor; in this Case Study we carry out that program. Let us recall that (after scaling) equilibria of the CSTR are described by the equation

$$G(x, \lambda; B, \delta, \lambda) = (1 + \lambda)y - \eta - \frac{B\lambda}{1 + \delta\lambda\mathscr{A}(x)} = 0, \qquad \text{(C1.1)}$$

where

$$\mathscr{A}(x) = \exp\left\{-\frac{\gamma x}{1 + x}\right\}.$$

(*Remark*: We do not include γ as an argument in (C1.1), since we do not vary this parameter.) Our goal is to analyze (C1.1) in terms of the winged cusp singularity; i.e.,

$$h(x, \lambda) = x^3 + \lambda^2. \qquad \text{(C1.2)}$$

Specifically, the task of this Case Study is to prove the two theorems below.

Let $\Omega \subset \mathbb{R}^5$ be the set of physically acceptable parameter values; viz.,

$$\Omega = \{(x, \lambda, B, \delta, \eta): B, \lambda, \delta > 0 \text{ and } x, \eta > -1\}, \qquad \text{(C1.3)}$$

By way of explanation of (C1.3), we recall that x and η are temperature parameters and that absolute zero has been scaled to -1; and that λ, δ are transfer coefficients which must be positive. It will turn out for the problem we consider that $B > 0$; i.e., the reaction is exothermic. We also recall that γ is large, typically $\gamma > 10$.

Theorem C1.1. *For any* $\gamma > \frac{8}{3}$, *there exists a unique point*

$$Z_0 = (x_0, \lambda_0, B_0, \delta_0, \eta_0) \in \Omega$$

such that $G(x, \lambda, B_0, \delta_0, \eta_0)$ *is equivalent to the winged cusp* (C1.2) *in a neighborhood of* (x_0, λ_0). *The following asymptotic expressions for* Z_0 *hold when* γ *is large*:

$$x_0 \sim \frac{1}{\gamma}, \qquad \lambda_0 \sim \frac{\gamma}{2}, \qquad B_0 \sim \frac{4}{\gamma}, \qquad \delta_0 \sim \frac{2e}{\gamma}, \qquad \eta_0 \sim -\tfrac{1}{2}. \quad \text{(C1.4)}$$

Theorem C1.2. *The three-parameter unfolding* $G(x, \lambda; B, \delta, \eta)$ *in* (C1.1) *is a universal unfolding of* $G(x, \lambda; B_0, \delta_0, \eta_0)$ *near* (x_0, λ_0).

In Chapter III, §8 we studied the universal unfolding

$$H(x, \lambda; \alpha, \beta, \gamma) = x^3 + \lambda^2 + \alpha + \beta x + \gamma\lambda x$$

of (C1.2). Universal unfoldings of equivalent singularities only differ by a change of coordinates. Thus it follows from Theorems C1.1 and C1.2 that as B, δ, η vary near B_0, δ_0, η_0, (C1.1) exhibits exactly the seven persistent perturbations listed in Figure III,8.4. Moreover, the regions in parameter space where the various perturbations occur in the two unfoldings are diffeomorphic, but we shall not determine this diffeomorphism.

Let us interpret the parameter values (C1.4) of the winged cusp point in terms of the original, unscaled variables of the CSTR. The flow rate r and the reaction rate Z are of the same order, with r larger by a factor of e. The rate constant k for cooling is much smaller than either r or Z, being reduced by a factor of γ^{-1}. The coolant temperature T_c is much lower than the feed temperature T_f—on the absolute scale $T_c = T_f/2$. The reaction is only mildly exothermic, and the steady-state temperature in the reaction vessel is only slightly elevated from the feed temperature T_f.

These results were first presented in Golubitsky and Keyfitz [1980]. The results there are a little stronger than here. Specifically, to obtain the winged cusp, one need only assume that the reaction rate term $\mathscr{A}$ is a C^∞-function which is C^3 close to Arrhenius form. Moreover, all other singularities which occur in the family G, for *any* parameter values, are singularities already found in the universal unfolding of the winged cusp. This fact along with the numerical work of Uppal *et al.* [1976] suggested that these local results obtained by singularity theory techniques were probably valid globally. In fact, Balakotaiah and Luss [1982] performed numerical calculations verifying that the local description of the bifurcation diagrams given by unfolding the winged cusp is indeed global. Their method involved following numerically the bifurcation, hysteresis, and double limit point varieties away from the winged cusp point.

The theoretical framework for proving Theorems C1.1 and C1.2 is clear. Both theorems correspond to recognition problems—for normal forms in

Theorem C1.1, for universal unfoldings in Theorem C1.2. Thus to prove these two theorems we must carry out the calculations specified in Proposition II,9.4 and Proposition III,4.5, respectively. Below we shall present the full details of these calculations. (We won't do so for the other two Case Studies. It seems important to see such calculations in their entirety for at least one example; we have chosen this one because it is the simplest technically.)

PROOF OF THEOREM C1.1. According to Proposition II,9.4, we must show that for each fixed γ there is a unique solution $(x_0, \lambda_0, B_0, \delta_0, \eta_0) = Z_0$ in Ω to the system of five equations

$$G = G_x = G_{xx} = G_\lambda = G_{x\lambda} = 0 \tag{C1.5a}$$

in the five unknowns $x, \lambda, B, \delta, \eta$ and that for this solution

$$G_{xxx} > 0 \quad \text{and} \quad G_{\lambda\lambda} > 0 \quad \text{at} \quad Z_0. \tag{C1.5b}$$

(*Remark*: In Theorem C1.1, the restriction $\gamma > \frac{8}{3}$ arises from the requirement that Z_0 lie in the physical region Ω.)

We begin by computing the derivatives in (C1.5a). To simplify the notation, we write

$$\Delta = 1 + \delta\lambda\mathscr{A}(x)$$

for the denominator in (C1.1). Then we have

$$\begin{aligned}
&\text{(a)} \quad G = (1 + \lambda)y - \eta - B\lambda/\Delta,\\
&\text{(b)} \quad G_\lambda = x - B/\Delta^2,\\
&\text{(c)} \quad G_x = (1 + \lambda) + B\delta\lambda^2\mathscr{A}'/\Delta^2,\\
&\text{(d)} \quad G_{x\lambda} = 1 + 2B\delta\lambda\mathscr{A}'/\Delta^3,\\
&\text{(e)} \quad G_{xx} = B\delta\lambda^2[\Delta\mathscr{A}'' - 2\delta\lambda(\mathscr{A}')^2]/\Delta^3.
\end{aligned} \tag{C1.6}$$

We also record the derivatives of $\mathscr{A}(x)$,

$$\begin{aligned}
&\text{(a)} \quad \mathscr{A} = \exp\left(\frac{-\gamma x}{1 + x}\right) > 0,\\
&\text{(b)} \quad \mathscr{A}' = \frac{-\gamma}{(1 + x)^2}\mathscr{A}(x) < 0,\\
&\text{(c)} \quad \mathscr{A}'' = \frac{\gamma^2 + 2\gamma(1 + x)}{(1 + x)^4}\mathscr{A}(x) > 0,
\end{aligned} \tag{C1.7}$$

where the inequalities hold at least for $x > -1$.

The main difficulty in solving (C1.5a) lies in the fact that the dependence on x of the derivatives in (C1.6) is so horribly nonlinear. We shall deal with this difficulty by manipulating (C1.5a) to extract a simple equation which depends only on x. Having solved this equation for x, it will be relatively easy to solve for the remaining unknowns.

Specifically, we proceed as follows. Observe from (C1.6b) that

$$B = \Delta^2 x. \tag{C1.8}$$

Starting from the two equations $G_{x\lambda} = G_{xx} = 0$ in (C1.5a), we process these by substituting the value (C1.8) for B in (C1.6d), by canceling the nonzero factors $B\delta\lambda^2\Delta^{-3}$ in (C1.6e), and by writing out the remaining factors of Δ; we obtain

$$\begin{aligned} &\text{(a)} \qquad \delta\lambda(\mathscr{A} + 2x\mathscr{A}') = -1, \\ &\text{(b)} \quad \delta\lambda[2(\mathscr{A}')^2 - \mathscr{A}\mathscr{A}''] = \mathscr{A}''. \end{aligned} \tag{C1.9}$$

Next we multiply (C1.9a) by $\mathscr{A}''$, add it to (C1.9b) and divide the result by $2\delta\lambda\mathscr{A}'$ to deduce

$$\mathscr{A}' + x\mathscr{A}'' = 0. \tag{C1.10}$$

Because (C1.10) is homogeneous, the exponential cancels; specifically substituting (C1.7) into (C1.10) yields

$$x^2 + \gamma x - 1 = 0. \tag{C1.11}$$

Equation (C1.11) has a unique solution x_0 satisfying $x_0 > -1$. (Note that in fact $x_0 > 0$, since $\gamma > 0$.)

Now we solve for δ and λ. Substituting the value (C1.8) for B into the equation $G_x = 0$, we deduce

$$\delta\lambda x\mathscr{A}'(x) = -\frac{1 + \lambda}{\lambda}. \tag{C1.12}$$

But (C1.9a) gives a value for the product $\delta\lambda$ on the left in (C1.12), depending only on x_0. Substitution of (C1.9a) into (C1.12) yields

$$\lambda_0 = -\frac{\mathscr{A} + 2x\mathscr{A}'}{\mathscr{A} + x\mathscr{A}'}\bigg|_{x=x_0}, \tag{C1.13}$$

and substitution in turn of (C1.13) into (C1.9) yields

$$\delta_0 = +\frac{\mathscr{A} + x\mathscr{A}'}{(\mathscr{A} + 2x\mathscr{A}')^2}\bigg|_{x=x_0}. \tag{C1.14}$$

Finally, we have from (C1.8) that

$$B_0 = x_0\Delta_0^2, \tag{C1.15}$$

where

$$\Delta_0 = 1 + \delta_0\lambda_0\mathscr{A}(x_0),$$

and setting (C1.6a) equal to zero yields

$$\eta_0 = (1 + \lambda_0)x_0 - B_0\lambda_0/\Delta_0. \tag{C1.16}$$

This solves (C1.5a); it remains to check that Z_0 lies in the physical region, to verify (C1.5b), and to obtain the asymptotic formulas (C1.4). We leave the latter task to the reader.

First we shall prove that Z_0 lies in the physical region. In this direction, we need to determine the signs of the numerator and denominator in (C1.13), the formula for λ_0. We claim that

$$\begin{aligned} &\text{(a)} \quad \mathscr{A}(x_0) + x_0\mathscr{A}'(x_0) > 0, \\ &\text{(b)} \quad \mathscr{A}(x_0) + 2x_0\mathscr{A}'(x_0) < 0, \end{aligned} \tag{C1.17}$$

the second inequality holding if $\gamma > 8/3$. To derive (C1.17a), we substitute the value $x_0 = -\mathscr{A}'(x_0)/\mathscr{A}''(x_0)$ from (C1.10) into (C1.17a) to obtain

$$\mathscr{A} + x\mathscr{A}'\bigg|_{x=x_0} = \frac{\mathscr{A}\mathscr{A}'' - (\mathscr{A}')^2}{\mathscr{A}''}\bigg|_{x=x_0}.$$

By (C1.7), $\mathscr{A}''$ is positive and $\mathscr{A}$ is nonzero; since we are only interested in signs, we may replace the factor $\mathscr{A}''$ in the denominator by $\mathscr{A}^2$ to conclude

$$\operatorname{sgn}(\mathscr{A} + x\mathscr{A}')\bigg|_{x=x_0} = \operatorname{sgn}\left(\frac{\mathscr{A}\mathscr{A}'' - (\mathscr{A}')^2}{\mathscr{A}^2}\right)\bigg|_{x=x_0}.$$

But the right-hand side of this equation is simply the second derivative of $\log \mathscr{A}$, and we compute that

$$(\log \mathscr{A})'' = \frac{2\gamma}{(1+x)^3} > 0.$$

This proves (C1.17a).

For (C1.17b), we observe from (C1.9a) and (C1.9b) that

$$(\mathscr{A} + 2x\mathscr{A}')\bigg|_{x=x_0} = -\frac{1}{\delta_0\lambda_0} = -\frac{2(\mathscr{A}')^2 - \mathscr{A}\mathscr{A}''}{\mathscr{A}''}\bigg|_{x-x_0}.$$

As above, we invoke the signs in (C1.7) to conclude that

$$\operatorname{sgn}(\mathscr{A} + 2x\mathscr{A}')\bigg|_{x=x_0} = -\operatorname{sgn}\left(\frac{2(\mathscr{A}')^2 - \mathscr{A}\mathscr{A}''}{\mathscr{A}^2}\right)\bigg|_{x=x_0}.$$

Using (C1.7) to evaluate these derivatives, we see that

$$\frac{2(\mathscr{A}')^2 - \mathscr{A}\mathscr{A}''}{\mathscr{A}^2} = \gamma\frac{\gamma - 2(1+x)}{(1+x)^4},$$

which is positive precisely when

$$x < \frac{\gamma}{2} - 1. \tag{C1.18}$$

But x_0 defined by (C1.11) satisfies (C1.18) if and only if $\gamma > 8/3$. This proves (C1.17b).

We now show that Z_0 lies in Ω. In solving (C1.11) we chose $x_0 > -1$. It is immediate from (C1.17) that $\varepsilon_0 > 0$ and $\delta_0 > 0$. We need only show that $\eta_0 > -1$. To do this we manipulate (C1.16) to express η_0 in terms of x_0; we claim that

$$\eta_0 = \left. \frac{x^2}{x + \dfrac{\mathscr{A}}{\mathscr{A}'}} \right|_{x=x_0} . \tag{C1.19}$$

To show this, we substitute (C1.8) for B into (C1.16), obtaining

$$\eta_0 = (1 + \lambda_0)x_0 - x_0\lambda_0\Delta_0. \tag{C1.20}$$

On the other hand, combining (C1.6c) and (C1.6d) we conclude that

$$\lambda_0\Delta_0 = 2(1 + \lambda_0),$$

which when substituted into (C1.20) yields

$$\eta_0 = -(1 + \lambda_0)x_0. \tag{C1.21}$$

The claim (C1.19) follows from substituting (C1.13) for λ_0 into (C1.21) and rearranging. Note from ((C1.7b) and (C1.17a)) that the denominator in (C1.19) is negative; thus to show that $\eta_0 > -1$ we must show that

$$x_0^2 < -\left[x_0 + \frac{\mathscr{A}}{\mathscr{A}'}(x_0)\right]. \tag{C1.22}$$

On rearranging terms and taking $\mathscr{A}/\mathscr{A}'$ from (C1.7b), we see that (C1.22) is equivalent to

$$x_0^2 + x_0 - \frac{(1 + x_0)^2}{\gamma} < 0, \tag{C1.23}$$

and (C1.23) may be derived from (C1.11). This proves $\eta_0 > -1$.

Finally, it remains to prove (C1.5b). We first compute that

$$\begin{aligned} &\text{(a)} \quad G_{\lambda\lambda} = 2B\delta\mathscr{A}/\Delta^3, \\ &\text{(b)} \quad G_{xxx} = B\delta\lambda^2\{\Delta\mathscr{A}''' - 3\delta\lambda\mathscr{A}'\mathscr{A}''\}/\Delta^3; \end{aligned} \tag{C1.24}$$

the second formula holding *only if* $G_{xx} = 0$. To prove $G_{\lambda\lambda}(Z_0) > 0$, note that $\Delta_0 = 1 + \delta_0\lambda_0\mathscr{A}(x_0)$ is positive since δ_0, λ_0, and $\mathscr{A}$ are positive. Thus all factors in (C1.24a) are positive, and $G_{\lambda\lambda}(Z_0) > 0$. Let us turn to $G_{xxx}(Z_0)$. We ignore all the factors outside the brackets in (C1.24), as they are positive. We substitute the value

$$\Delta_0 = 2\delta_0\lambda_0\frac{(\mathscr{A}')^2}{\mathscr{A}''}$$

obtained from $G_{xx} = 0$ (i.e., (C1.6e)) into (C1.24). Thus

$$\operatorname{sgn} G_{xxx}(Z_0) = \operatorname{sgn}\left.\left\{2\frac{(\mathscr{A}')^2\mathscr{A}'''}{\mathscr{A}''} - 3\mathscr{A}'\mathscr{A}''\right\}\right|_{x=x_0} .$$

We multiply by $\mathscr{A}''/(\mathscr{A}')^3$, which by (C1.7) is negative; this yields

$$\operatorname{sgn} G_{xxx}(Z_0) = \operatorname{sgn}\left\{3\left(\frac{\mathscr{A}''}{\mathscr{A}'}\right)^2 - 2\frac{\mathscr{A}'''}{\mathscr{A}''}\right\}\Bigg|_{x=x_0}. \tag{C1.25}$$

We compute that the right-hand side of (C1.25) equals $\gamma/(1 + x_0)^4$, so $G_{xxx}(Z_0) > 0$. (*Remark*: Apart from sign, the right-hand side of (C1.25) is the Schwarzian derivative of $\mathscr{A}$.) □

Proof of theorem C1.2. According to Proposition III,4.5, we must show that the determinant of the matrix

$$\begin{pmatrix} 0 & 0 & 0 & G_{xxx} & G_{xx\lambda} \\ 0 & 0 & G_{\lambda\lambda} & G_{\lambda xx} & G_{\lambda\lambda x} \\ G_B & G_{Bx} & G_{B\lambda} & G_{Bxx} & G_{Bx\lambda} \\ G_\delta & G_{\delta x} & G_{\delta\lambda} & G_{\delta xx} & G_{\delta x\lambda} \\ G_\eta & G_{\eta x} & G_{\eta\lambda} & G_{\eta xx} & G_{\eta x\lambda} \end{pmatrix}$$

evaluated at the winged cusp point Z_0 is nonzero. Observe from (C1.1) that $G_\eta \equiv 1$ identically; thus the 5, 1-entry is the only nonzero entry in the bottom row of this matrix. Expanding in minors about the fifth row, we reduce to the determinant of

$$\begin{pmatrix} 0 & 0 & G_{xxx} & G_{xx\lambda} \\ 0 & G_{\lambda\lambda} & G_{\lambda xx} & G_{\lambda\lambda x} \\ G_{Bx} & G_{B\lambda} & G_{Bxx} & G_{Bx\lambda} \\ G_{\delta x} & G_{\delta\lambda} & G_{\delta xx} & G_{\delta x\lambda} \end{pmatrix}. \tag{C1.26}$$

We simplify this determinant by operating on the second and fourth rows. We claim that at Z_0

$$\frac{G_{\delta\lambda}}{G_{\lambda\lambda}} = \frac{G_{\delta xx}}{G_{\lambda xx}} = \frac{G_{\delta x\lambda}}{G_{\lambda x\lambda}} = \frac{\lambda}{\delta}. \tag{C1.27}$$

Thus we may annihilate the last three columns of the fourth row of (C1.26) by subtracting λ/δ times the second row from the fourth row; this shows that the determinant of (C1.26) equals

$$-G_{\delta x} \det\begin{pmatrix} 0 & G_{xxx} & G_{xx\lambda} \\ G_{\lambda\lambda} & G_{\lambda xx} & G_{\lambda x\lambda} \\ G_{B\lambda} & G_{Bxx} & G_{Bx\lambda} \end{pmatrix}. \tag{C1.28}$$

Now we verify (C1.27). Observe that $\Delta = 1 + \delta\lambda\mathscr{A}(x)$ depends on δ and λ only through the product $\delta\lambda$. It is immediate from (C1.6b, d) that G_λ and $G_{\lambda x}$ also have this property. But if $f(\lambda, \delta) = h(\delta\lambda)$, then $f_\delta/f_\lambda = \lambda/\delta$. This verifies the first and third equalities in (C1.27), and it remains to consider $G_{\delta xx}/G_{\lambda xx}$. For this we rewrite (C1.6e) as $G_{xx} = B\delta\lambda^2 Q/\Delta^3$, where

$$Q = \Delta\mathscr{A}'' - 2\delta\lambda(\mathscr{A}')^2.$$

Since $Q = 0$ at Z_0, we have $G_{\delta xx}/G_{\lambda xx} = Q_\delta/Q_\lambda$ at Z_0. However Q depends only on $\delta\lambda$, so $Q_\delta/Q_\lambda = \lambda/\delta$. This proves (C1.27).

To complete the proof we will show that both factors in (C1.28), $G_{\delta x}$ and the determinant, are nonzero. Regarding the first, we differentiate (C1.6c) with respect to δ to obtain

$$G_{\delta x} = B\lambda^2\mathscr{A}'(1 - \delta\lambda\mathscr{A})/\Delta^3.$$

Substituting from (C1.9a) we see that

$$1 - \delta_0\lambda_0\mathscr{A}(x_0) = 2\frac{\mathscr{A} + x\mathscr{A}'}{\mathscr{A} + 2x\mathscr{A}'}\bigg|_{x=x_0},$$

which is nonzero by (C1.17). Thus $G_{\delta x} \neq 0$.

Finally, we show that the determinant in (C1.28) is nonzero. First note from (C1.6e) that

$$G_{Bxx} = \frac{1}{B}G_{xx} = 0 \quad \text{at } Z_0.$$

Thus expanding the determinant we obtain

$$G_{xxx}(G_{\lambda x\lambda}G_{B\lambda} - G_{\lambda\lambda}G_{Bx\lambda}) - G_{xx\lambda}^2 G_{B\lambda}. \tag{C1.29}$$

We compute the following derivatives in (C1.29) from (C1.6).

(a) $G_{B\lambda} = -1/\Delta^2$,

(b) $G_{Bx\lambda} = 2\delta\lambda\mathscr{A}'/\Delta^3$,

(c) $G_{\lambda x\lambda} = 2B\delta\mathscr{A}'[1 - 2\delta\lambda\mathscr{A}]/\Delta^4$.

We claim that (C1.29) is positive. We see from (1.30a) that $G_{B\lambda} < 0$, so the second term in (C1.29) is certainly nonnegative. Concerning the first, we already know from (C1.5b) that $G_{xxx} > 0$. Substituting from (C1.30) (and from (C1.24a) for $G_{\lambda\lambda}$) we find

$$G_{\lambda x\lambda}G_{B\lambda} - G_{\lambda\lambda}G_{Bx\lambda} = -2B\delta\mathscr{A}'/\Delta^6 > 0,$$

the inequality coming from (C1.7b). Thus the first term in (C1.29) is positive, and the claim follows. □

EXERCISE C1.1 (Assumes Exercise I, 3.1).

Consider the Liapunov–Schmidt reduction of the ODE (I,2.1) for the CSTR using the data M, N, v_0, and v_0^* described in Exercise I,3.1. Show that for the asymptotic parameter values (C1.4) we have

$$\langle v_0, v_0^* \rangle \sim \tfrac{1}{2}.$$

In other words, show that the choice of data in Exercise I,3.1 is consistent with the requirement (I,4.10), needed to make correct stability predictions. (*Warning*: Recall that a minus sign is needed to bring (I,2.1) into the standard form (I.4.1).)

CHAPTER IV
Classification by Codimension

§0. Introduction

The main purpose of this chapter is to classify all bifurcation problems (in one state variable) of codimension three or less. We find that there are eleven such singularities, which we call the *elementary bifurcation problems*. In the course of the chapter, we tabulate the following data for each of these eleven singularities:

(i) Normal form (Table 2.1).
(ii) Algebraic data (i.e., $\mathscr{S}(h)$, RT(h), $\mathscr{P}(h)$, $T(h)$, a complement to $T(h)$, codimension) (Table 2.2).
(iii) Solution of the recognition problem for normal forms (Table 2.3).
(iv) Universal unfolding (Table 3.1).
(v) Solution of the recognition problem for universal unfoldings (Table 3.2).
(vi) Equations for the bifurcation, hysteresis, and double limit point varieties (Table 4.1).
(vii) Graphs of the persistent perturbed bifurcation diagrams (Figures 4.1–4.3).

Thus the chapter should also be useful as a compact reference.

A complete classification of all singularities seems to be an impossible task. Hence we only attempt to classify singularities of low codimension. Of course, the degeneracy of a singularity increases with its codimension (cf. Corollary III,2.6). Thus we are classifying the least degenerate singularities; i.e., the ones most likely to occur in applications. Indeed, in §1 we attempt to quantify the idea that singularities of low codimension are more likely in applications by relating codimension to the number of (non-dimensionalized) auxiliary parameters in a mathematical model.

One may of course ask, "Why stop at codimension three?" Our answer is that perturbed bifurcation diagrams with a singularity of codimension k are enumerated by regions in $\mathbb{R}^k$. If $k > 3$, visualization of these regions becomes very difficult.

The chapter is divided into four sections. In §1 we discuss the kind of information about mathematical models that can be derived from codimension. The main classification theorem, identifying all singularities of codimension three or less, is stated and proved in §2. In §§3 and 4, we consider universal unfoldings and persistent perturbations of the elementary singularities, respectively.

§1. Philosophical Remarks Concerning Codimension

Consider a k-parameter family of bifurcation problems,

$$G(x, \lambda, \alpha) = 0, \tag{1.1}$$

which arises from a mathematical model for some physical problem. We imagine that (1.1) has already been nondimensionalized, so that the k parameters in (1.1) are essential. Suppose that for $\alpha = \alpha_0$, (1.1) has a singularity at $x = x_0$, $\lambda = \lambda_0$ whose *codimension exceeds the number of auxiliary parameters*; in symbols

$$\operatorname{codim} G(\cdot, \cdot, \alpha_0) > k. \tag{1.2}$$

In this section we argue the following thesis: *Such a mathematical model should be approached with caution.* We believe this thesis is a rather important principle of wide applicability in applied mathematics. It derives from René Thom's work on catastrophe theory. The ultimate mathematical basis for this thesis is the simple fact that typically an overdetermined system of algebraic equations (i.e., more equations than unknowns) has no solutions.

Actually (1.2) is just a readily applicable test addressing the following, more fundamental question: Is $G(\cdot, \cdot, \alpha)$ a universal unfolding for $G(\cdot, \cdot, \alpha_0)$, or even a versal unfolding? Surely not, under the assumption (1.2), since there are not enough unfolding variables. Of course, $G(\cdot, \cdot, \alpha)$ might fail to be a versal unfolding even when (1.2) is violated. However, in a given model detailed computations are required to decide whether this latter possibility actually occurs. By contrast (1.2) can be verified with little effort. In this section, we are interested in the information that can be derived from singularity theory *without* performing detailed computations, so we work with (1.2).

We divide the section into two subunits:

(a) Why (1.2) is usually grounds for caution, and
(b) Special circumstances which modify codimension in (1.2).

(a) Why (1.2) is Usually Grounds for Caution

Let us attempt to discuss the issues here in general terms. A mathematical model for a physical problem always neglects some physical effects, many of which are beyond the control of the experimenter. Thus (1.1) merely describes the system of the mathematical model; a more accurate description of the physical system would lead to an equation

$$G(x, \lambda, \alpha) + p(x, \lambda, \alpha) = 0, \tag{1.3}$$

where p is a small perturbation that represents what the model neglects. We regard p as *unknowable*—a more accurate model might specify p partially, but no matter how accurate the model, some things will have been neglected.

What is the effect of p in (1.3)? If $G(\cdot, \cdot, \alpha)$ is a versal unfolding of $G(\cdot, \cdot, \alpha_0)$, the effect of p is minimal. In mathematical terms (1.3) may be factored through the versal unfolding $G(\cdot, \cdot, \alpha)$. In physical terms, exactly the same bifurcation phenomena occur for (1.3) as for (1.1), just at slightly different values of the parameter α.

However suppose (1.2) holds. Equation (1.1) is an unfolding of $G(\cdot, \cdot, \alpha_0)$, so it may be factored through a universal unfolding of $G(\cdot, \cdot, \alpha_0)$. For definiteness let $H(\cdot, \cdot, \beta)$ be such a universal unfolding; suppose H requires l parameters, where $l > k$. Because (1.1) factors through H, all of the phenomena in (1.1) occur on a k-dimensional subset of unfolding space $\mathbb{R}^l$. (*Remark*: We may describe this k-dimensional subset more explicitly as follows. By the unfolding theorem there is a map between parameter spaces $A: \mathbb{R}^k \to \mathbb{R}^l$ such that

$$G(\cdot, \cdot, \alpha) \sim H(\cdot, \cdot, A(\alpha)),$$

where $\sim$ indicates equivalence. Only points on

$$\{\beta \in \mathbb{R}^l: \beta = A(\alpha) \text{ for some } \alpha \in \mathbb{R}^k\} \tag{1.4}$$

are associated with phenomena of (1.1).) Similarly, provided p is small enough, all of the phenomena in (1.3) occur on a different k-dimensional subset of $\mathbb{R}^l$. However, these two subsets need not intersect one another; in this case, *the perturbed family* (1.3) *will not even exhibit the original singularity.*

In conclusion, predictions of a mathematical model where (1.2) obtains are likely to be erratic. Experiments (described by (1.3)) will have new qualitative features not predicted by theory (described by (1.1)). Moreover, in different experiments the perturbation p in (1.3) will probably be slightly different, so experiments may not be completely reproducible.

In the remainder of subsection (a) we illustrate this general discussion with a specific example; viz., the experiments of Roorda [1965, 1968] on imperfection sensitivity in the collapse of the shallow arch. (These experiments are described on pp. 75–6 of Thompson and Hunt [1973].) The shallow arch is an infinite dimensional system; although it may be reduced to a one dimensional system with the Liapunov–Schmidt technique, we have not yet introduced the necessary theory in this text. Therefore we shall

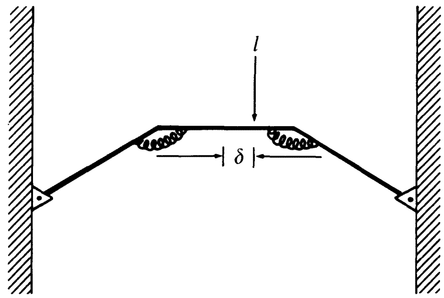


Figure 1.1. Finite element analogue of the shallow arch.

only consider a finite element analogue of the shallow arch. Even with this simplified system, the calculations are rather messy; a number of these calculations are left for the reader in Exercises 1.1–1.5.

The (finite element analogue of the) shallow arch is illustrated in Figure 1.1. This system consists of three rigid struts connected to one another and to rigid walls by pins which permit rotation in a plane. At the two interior pins there are torsional springs which resist rotation. The separation between the two outer pins is slightly less than the combined length of the three struts, so that in unstressed equilibrium the system bulges upward or downward. The system is stressed by a vertical load l which is applied at a distance δ from the center of the middle strut.

We are interested in the behavior of the upward bulbing equilibrium configuration as l is increased. It turns out that when l becomes too large, the arch collapses and snaps through to the downward bulging equilibrium configuration. Let us write $l_c(\delta)$ for the load at which the arch collapses if the load is applied a distance δ off center. The dependence of $l_c(\delta)$ on the parameter δ is our main concern. This issue has engineering significance beyond the problem studied here. Specifically, $l_c(\delta)$ represents the load-carrying capacity of the shallow arch. The load-carrying capacity is greatest if $\delta = 0$. Normally one would design such a structure so that the load is applied at $\delta = 0$ in order to take advantage of this maximum strength. However, real structures will inevitably differ from ideal structures through various imperfections. Applying the load off center on the ideal structure is a convenient way to model imperfections.

In our treatment of the shallow arch we will

(i) derive theoretically an estimate for $l_c(\delta)$ when δ is small; specifically,

$$l_c(\delta) = l_0 - C\delta^{2/3} + O(\delta^{4/3}), \tag{1.5}$$

where $l_0 = l_c(0)$ and C is a positive constant;

(ii) discuss the inadequacy of (1.5) for describing the experimental data; and

(iii) relate the discrepancy between theory and experiment to (1.2).

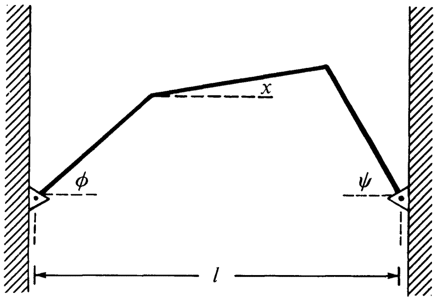


Figure 1.2. Notation to describe the shallow arch.

Remark. The well-known 2/3-power in (1.5), originally derived by Koiter [1945], indicates that a small change in δ can decrease the load-carrying capacity significantly.

(i) *A Theoretical Estimate for* $l_c(\delta)$

As a preliminary theoretical point, we show that the shallow arch of Figure 1.1 has only one degree of freedom. After scaling, we may assume that the three struts have unit length. Let L be the separation of the end pins, and let ϕ, ψ, and x be the angles defined in Figure 1.2. The fact that the two end pins are at the same height and are separated by a horizontal distance L leads to the two relations

$$\begin{aligned} &\text{(a)} \quad \cos\phi + \cos x + \cos\psi = L, \\ &\text{(b)} \quad \sin\phi + \sin x - \sin\psi = 0. \end{aligned} \tag{1.6}$$

These equations determine ϕ and ψ implicitly as functions of x. Therefore we may parametrize states of the shallow arch by the single variable x. (Strictly speaking, for each x there are two solution pairs (ϕ, ψ) of (1.6), corresponding to upward bulging and downward bulging states. However, our analysis below is only local; i.e., restricted to a neighborhood of the upward bulging equilibrium. In such a neighborhood the solution of (1.6) is unique, so that x uniquely parametrizes states of the system.)

The goal of this subunit is to derive (1.5), an estimate for the dependence on δ of the load at which collapse occurs. To begin, let us discuss how to express this collapse load as a function of δ. In Exercise 1.1 we ask the reader to derive an equation

$$G(x, l, \delta) = 0 \tag{1.7}$$

which characterizes the equilibria of the shallow arch, both stable and unstable. We regard (1.7) as a one-parameter family of bifurcation problems in which x is the state variable, l the bifurcation parameter, and δ an auxiliary parameter. Now the shallow arch collapses when, as l is increased, a stable solution x of (1.7) loses its stability. In a system with one degree of freedom, such a loss of stability can occur only at a singularity of G; i.e., at a point where

$$G(x, l, \delta) = G_x(x, l, \delta) = 0. \tag{1.8}$$

The function $l_c(\delta)$, as well as the value of x at collapse, may be obtained by solving this 2×2 system for x and l as functions of δ.

It would be rather difficult to obtain the estimate (1.5) by solving (1.8) directly. Therefore we shall use singularity theory methods to put the equations in a particular tractable form before solving (1.8). Specifically we shall prove that when $\delta = 0$, collapse of the shallow arch results from a pitchfork bifurcation of (1.7); more precisely, near the bifurcation point, G is equivalent to $-x^3 - \lambda x$. Then we shall derive (1.5) by using the universal unfolding of the pitchfork to handle small, nonzero values of δ.

First, we set δ equal to zero and look for a pitchfork in (1.7). In Exercise 1.1 we ask the reader to show that $G(x, l, 0)$ is an odd function of x. (This property is a consequence of the symmetry of the problem (when $\delta = 0$) under reflection about the vertical axis.) Since $G(x, l, 0)$ is odd, $x = 0$ is a solution of $G(x, l, 0) = 0$ for any l. For small l this solution is stable; as l increases, this solution can lose stability only at a point where (1.8) is satisfied. In other words, when $\delta = 0$ collapse occurs at a load l_0 such that

$$G(0, l_0, 0) = G_x(0, l_0, 0) = 0.$$

However, since $G(x, l, 0)$ is odd in x, it follows that

$$G_{xx}(0, l_0, 0) = G_l(0, l_0, 0) = 0.$$

Combining these four equations, we see that near $x = 0$, $l = l_0$

$$G(x, l, 0) = C_1 x^3 + C_2 \lambda x + \text{hot}, \tag{1.9}$$

where $\lambda = l - l_0$. Generically C_1 and C_2 will be nonzero; if they are nonzero, then $G(x, l, 0)$ is equivalent to a pitchfork $\pm x^3 + \lambda x$.

In Exercise 1.1 we also ask the reader to show by direct calculation that $C_1 < 0$ and $C_2 < 0$. Here we offer a heuristic proof concerning these two signs based on the following information:

(a) The solution $x = 0$ of $G(x, l, 0) = 0$ is stable if $l < l_0$.
(b) The arch collapses when $l > l_0$.

We deduce that $C_2 < 0$ from point (a) by using the sign of G_x to test for stability. (We remind the reader that $\lambda < 0$ if and only if $l < l_0$.) Concerning C_1, if $C_1 < 0$ then $G(x, l, 0)$ has a subcritical pitchfork bifurcation as illustrated in Figure 1.3. Note that for $l > l_0$ there are no stable equilibria in Figure 1.3. Hence this figure predicts discontinuous behavior near $l = l_0$, consistent with point (b) above. On the other hand, if $C_2 > 0$ then the bifurcation is supercritical. In this case the nontrivial solution branch lies in the half plane $\{(x, l): l > l_0\}$ and, by exchange of stability, these solutions are stable.

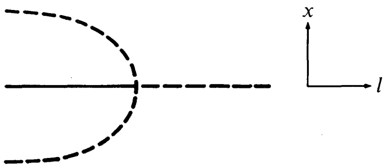


Figure 1.3 Pitchfork bifurcation $-x^3 - \lambda x = 0$ which occurs when $\delta = 0$.

This case leads to behavior contradicting point (b)—for $l > l_0$ the system would move along the nontrivial branch, never suffering a jump. We conclude that $C_2 < 0$. Thus near $x = 0$, $l = l_0$, the function $G(x, l, 0)$ is equivalent to $-x^3 - \lambda x$, where $\lambda = l - l_0$.

For any value of δ, zero or nonzero, equilibria of the shallow arch are described by the equation $G(x, l, \delta) = 0$. We know that $G(x, l, 0)$ is equivalent to $-x^3 - \lambda x$. By the universal unfolding theorem, $G(x, l, \delta)$ can be factored through the universal unfolding of $-x^3 - \lambda x$. In other words, for small δ, $G(x, l, \delta)$ is equivalent to

$$H(x, \lambda, \delta) = -x^3 - \lambda x + \alpha(\delta) + \beta(\delta)x^2, \tag{1.10}$$

where $\lambda = l - l_0$ and $\alpha(\delta)$, $\beta(\delta)$ are smooth functions of δ such that $\alpha(0) = \beta(0) = 0$. In Exercise 1.2 we ask the reader to show that $\alpha'(0) < 0$; in particular $\alpha'(0) \neq 0$. Let

$$\tilde{H}(x, \lambda, \delta) = x^3 - \lambda x + \alpha_1\delta, \tag{1.11}$$

where $\alpha_1 = \alpha'(0)$; this isolates the dominant terms in H. Note that

$$H(x, \lambda, \delta) = \tilde{H}(x, \lambda, \delta) + O(\delta^2, \delta x^2). \tag{1.12}$$

Let us consider substituting $\tilde{H}(x, \lambda, \delta)$ for $G(x, \lambda, \delta)$ in (1.8); this yields the system

$$\tilde{H}(x, \lambda, \delta) = \tilde{H}_x(x, \lambda, \delta) = 0 \tag{1.13}$$

to be solved for x and λ as functions of δ. Writing out (1.13), we have

$$\begin{aligned} -x^3 - \lambda x + \alpha_1\delta &= 0 \\ -3x^2 - \lambda &= 0. \end{aligned} \tag{1.14}$$

On eliminating x from (1.14) and setting $\lambda = l - l_0$, we find

$$l = l_0 - 3\left(\frac{\alpha_1}{2}\right)^{2/3}\delta^{2/3}.$$

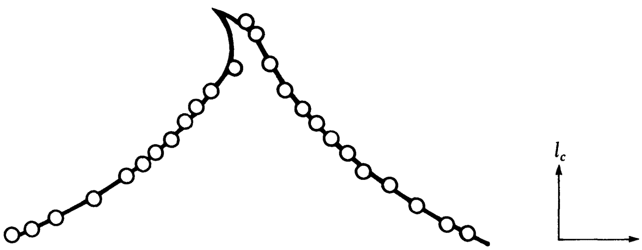
In other words, if $l_c(\delta)$ were defined by solving (1.13) rather than (1.8), then the estimate (1.5) would be exact, with no error term necessary.

In Exercise 1.3 we ask the reader to derive (1.5) by carefully analyzing the differences between (1.8) and (1.13). Let us summarize the issues. The comparison of (1.8) with (1.13) proceeds in two steps, first comparing (1.8) with

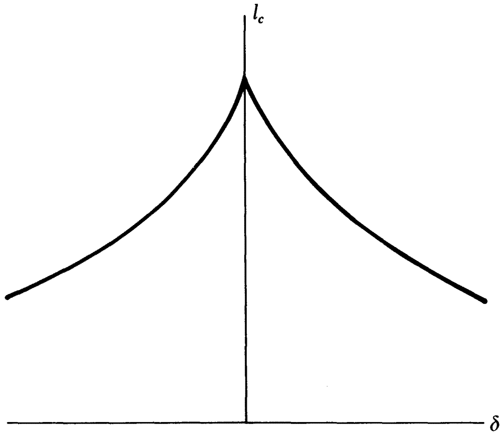
$$H(x, \lambda, \delta) = H_x(x, \lambda, \delta) = 0 \tag{1.15}$$

and then comparing (1.15) with (1.13). Let X, Λ be the diffeomorphism in the equivalence transformation which relates G to H; in the first step one must show that this diffeomorphism affects (1.5) only through a possible change in the constant C. In the second step one must show that, modulo the $O(\delta^{4/3})$ error in (1.5), the higher-order terms in (1.10) do not contribute to $l_c(\delta)$.

(ii) *Comparison of* (1.5) *with Experiment*
In Figure 1.4(a) we show the experimental values for $l_c(\delta)$ from Roorda [1965]; for comparison, in Figure 1.4(b) we have graphed the theoretical estimate (1.5). Note that the experimental values (shown by circles) seem to lie on a collapse load vs. δ curve which is a slightly tilted cusp. The reader may question our drawing a cusped curve to fit these data—the cusp lies



(a) Experiment. Data points shown schematically by circles. Cf. Roorda [1965].



(b) Theory: Cf. (1.5).

Figure 1.4. Snap through load as a function of δ.

Figure 1.5. Discontinuity in l_c as a function of δ predicted by tilted cusp. Compare with experimental data given in Figure 1.4(a).

in a region where there are no data points. It would seem from this information that $l_c(\delta)$ is simply a discontinuous function of δ such as sketched in Figure 1.5. To understand this issue, it is necessary to realize that the data in Figure 1.4(a) (taken from Roorda [1965]) were obtained by, for each value of δ, increasing the load quasi-statically until collapse occurred. Thus there was no possibility of observing the upper branches in Figure 1.4(a). In other words, if the collapse load vs. δ curve is a tilted cusp, then in an experiment of this kind, the tilting will manifest itself simply as a discontinuity in the observed values for $l_c(\delta)$.

Incidentally, the data from a subsequent experiment, which used a more sophisticated apparatus capable of detecting these upper branches, show the tilted cusp more clearly (cf. Roorda [1968]). Moreover, in the next unit we shall predict such tilting on theoretical grounds.

(iii) *Relation to* (1.2)

In deriving the estimate (1.5), we saw that $G(x, l, \delta)$ exhibits a pitchfork bifurcation when $\delta = 0$. Note that the pitchfork has codimension two but G has only one auxiliary parameter, δ. Thus (1.2) is satisfied here.

In this subunit we show that the discrepancy between theory and experiment is a natural consequence of (1.2). Let us elaborate. As we discussed above, the effects of imperfections in real structures can be modeled by subjecting the governing equations to a small, random perturbation. Thus we replace (1.7) by

$$G(x, l, \delta) + p(x, l, \delta) = 0; \tag{1.16}$$

this perturbation in turn modifies (1.8) to read

$$G + p = G_x + p_x = 0. \tag{1.17}$$

To determine the perturbed collapse load vs. δ curve, we must solve (1.17) for x and l as functions of δ. In this subunit we show that for a generic perturbation p, the collapse load vs. δ curve is a cusped curve that is slightly tilted with respect to the λ-axis. As we discussed above, such tilting accounts for the principal discrepancy between theory and experiment; viz., the discontinuity in $l_c(\delta)$ measured experimentally.

In the text we prove only that tilting occurs for the model problem

$$K(x, \lambda, \delta, \varepsilon) = \tilde{H}(x, \lambda, \delta) + \varepsilon x^2, \tag{1.18}$$

where $\tilde{H}$ is given by (1.11) and ε is a small parameter. (In Exercise 1.4 we ask the reader to reduce the general case to (1.18). The important point—both for this reduction and for the choice of (1.18) as a model—is that the right-hand side of (1.18) is a universal unfolding of the pitchfork.) Let us substitute (1.18) into (1.17) and rewrite (1.17) as

$$K - xK_x = K_x = 0;$$

this yields

$$\begin{aligned} \delta &= \frac{1}{\alpha_1}(-2x^3 + \varepsilon x^2), \\ \lambda &= -3x^2 + 2\varepsilon x. \end{aligned} \tag{1.19}$$

We would like to eliminate x from (1.19) and thereby express λ as a function of δ. Although it is not possible analytically to do so, nonetheless (1.19) gives a parametric representation of the collapse load vs. δ curve in the $\delta\lambda$-plane. Moreover, as desired, this curve has a cusp which, if $\varepsilon \neq 0$, is slightly tilted with respect to the λ-axis. The cusp arises from the common zero of $d\delta/dx$ and $d\lambda/dx$ at $x = \varepsilon/3$. The axis of the cusp points in the direction

$$\left.\left(\frac{d^2\delta}{dx^2}, \frac{d^2\lambda}{dx^2}\right)\right|_{x=\varepsilon/3} = \left(-\frac{2\varepsilon}{\alpha_1}, -6\right).$$

This completes the discussion relating the discrepancy between theory and experiment to (1.2).

Additional insight can be gained by viewing these issues geometrically. As we saw in (1.10), the ideal problem $G(x, \lambda, \delta)$ is equivalent to

$$-x^3 - \lambda x + \alpha(\delta) + \beta(\delta)x^2.$$

The coefficients $\alpha(\delta)$, $\beta(\delta)$ define a curve Γ_0 in the $\alpha\beta$-plane; i.e., in the parameter space of the universal unfolding of the pitchfork,

$$-x^3 - \lambda x + \alpha + \beta x^2. \tag{1.20}$$

Figure 1.6 shows such a curve, along with the transition variety for (1.20). Note that Γ_0 passes through the origin since $\alpha(0) = \beta(0)$. (Also note that $\alpha'(0) \neq 0$.) Similarly, a perturbed curve Γ may be associated to the perturbed problem (1.16), provided p is sufficiently small. However, a *generic perturbation of* Γ_0 *will not pass through the origin.* There is an intimate relation between the facts that generically Γ does not pass through the origin and that generically, for a small range of δ, the collapse load vs. δ curve is a triple-valued function of δ. We ask the reader to explore this further in Exercise 1.5.

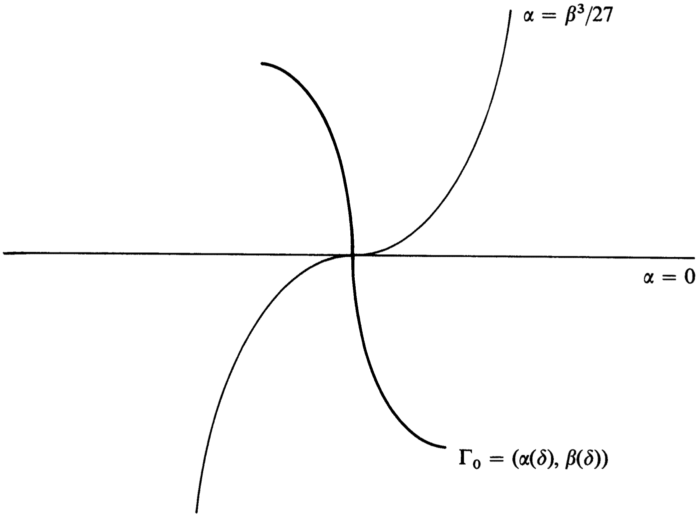


Figure 1.6. Curve Γ_0 in unfolding space of pitchfork.

(b) Special Circumstances Which Modify Codimension in (1.2)

Above we formulated the thesis that one should be cautious with a mathematical model which contains a singularity satisfying (1.2). However, this warning is predicated on the assumption that all perturbations of $G(\cdot, \cdot, \alpha_0)$ should be included in the mathematical model $G(\cdot, \cdot, \alpha)$, at least up to our notion of equivalence. There are at least two special circumstances where this assumption should be questioned. These special circumstances will serve to change the notion of codimension used in (1.2) from the one given in Chapter III; they will not, however, invalidate the warning associated with (1.2).

We discuss two special circumstances which call for a modification of the definition of codimension:

(i) The occurrence of moduli.
(ii) Special mathematical contexts: Symmetry.

In Chapters V and VI, we shall analyze in detail specific examples of these two circumstances. Here we describe them briefly, in order that these terms may have some meaning for the reader at this stage. Indeed, the only reason to mention them now is to illustrate that the thesis of this section must be applied with sensitivity for the problem under study.

(i) *Moduli*

A universal unfolding $G(x, \lambda, \alpha)$, where $\alpha \in \mathbb{R}^k$, of a singularity g is characterized by the following property: for any perturbation p, there is a map

$A\colon \mathbb{R} \to \mathbb{R}^k$ such that

$$g + \varepsilon p = SG(X, \Lambda, A(\varepsilon)) \tag{1.21}$$

for some S, X, and Λ (depending on ε). In (1.21), we have required that A, X, and Λ be C^∞. Let us consider weakening this hypothesis by only requiring these functions to be finitely differentiable—this would not change the most important information in (1.21); i.e., the fact that both sides of the equation have the same number of solutions x as a function of λ. It turns out that to obtain the most natural modification of codimension it is necessary to weaken the C^∞-hypothesis drastically, as described in the following definition.

Definition 1.1. Two germs $g, h \in \mathscr{E}_{x,\lambda}$ are *topologically equivalent* if there is a map of the form $(X(x, \lambda), \Lambda(\lambda))$ satisfying the following:

(i) $(X(x, \lambda), \Lambda(\lambda))$ and its inverse are continuous maps near the origin in $\mathbb{R}^2$.
(ii) $X(x, 0)$ and $\Lambda(\lambda)$ are monotone increasing functions of x and λ, respectively.
(iii) (X, Λ) maps the zero set of g onto that of h, locally near the origin.

In some examples, more complicated than any we have studied so far, fewer parameters are required to characterize perturbations of g up to topological equivalence than up to C^∞-equivalence. In other words, the perturbations of g associated with certain parameters in the universal unfolding of g are topologically equivalent to g itself, although not C^∞-equivalent; loosely speaking, we call parameters with this property *modal parameters* or simply *moduli.* We tentatively define the *topological codimension* of a singularity as its codimension *minus* the number of modal parameters. (This definition is subject to certain qualifications which we explore in Chapter V.)

In applying the thesis of this section we should use topological codimension on the left in (1.2). In most simple examples, the topological codimension equals the codimension as defined in Chapter III. A case where the two notions differ is explored fully in Chapter V.

The above discussion may cause the reader to wonder whether topological equivalence is the fundamental notion that we should take as the basis of our theory. Indeed, several authors, Buchner *et al.* [1983] and Percell and Brown [1984], have worked on this. However, in our opinion, the following fact is a decisive reason for staying with C^∞-equivalence: Two germs may be topologically equivalent but behave *very* differently under perturbation. The germs x and x^3 provide a gross example of this. More subtle examples will occur in Chapter V.

(ii) *Special Mathematical Contexts: Symmetry*

Symmetry is by far the most important special mathematical context which can change the notion of codimension that is appropriate in (1.2). We have already seen an example of this in the finite element analogue of the buckling of an Euler column, considered in Chapter I, §1. That model contains *no* auxiliary parameters, yet the analysis leads to a codimension two bifurcation problem; viz. the pitchfork. In other words, (1.2) holds here, running against the thesis of this section. However, there is a reflectional symmetry of the problem. Specifically, the potential energy in the buckled-up state is the same as the potential energy in the buckled-down state; it follows that the bifurcation equation is an odd function of the state variable x. We suggest that this problem should be analyzed in a different mathematical context; i.e., in the class of germs which possess the same symmetry as the original problem. Indeed, we shall show in Chapter VI that in this context the pitchfork $x^3 - \lambda x$ has codimension zero, thereby eliminating a case where (1.2) appears to be satisfied. This example is typical of the role of symmetry.

Another example of a special context which changes codimension appears in the work of Dangelmayr and Stewart [1984]. These authors consider a restricted set of equivalences which are the natural changes of coordinates in describing certain sequential chemical reaction.

We may summarize the above discussion by making the thesis of this section more specific: If a mathematical model leads to a singularity satisfying (1.2), we should question whether the problem is formulated in the right context; especially, are there any symmetries that have not been included?

(iii) *Concluding Remarks*

Both of the above issues arise from questioning whether a mathematical model $G(\cdot, \cdot, \alpha)$ should include all perturbations of $G(\cdot, \cdot, \alpha_0)$ up to C^∞-equivalence. Nonetheless, the two circumstances have different origins and different consequences. With moduli we ask whether the notion of equivalence is too strong; with symmetry we ask whether to consider fewer perturbations (i.e., only those which preserve the symmetry of the unperturbed mathematical model). With moduli the primary issues involve mathematical theory; with symmetry the primary focus is on the process of forming mathematical models of a physical situation. Moduli do not call for any particular response by the person applying the theory; by contrast, with symmetry, we have to decide which mathematical idealization best represents what we are trying to describe.

Of course, it is perfectly possible to have both effects operating simultaneously. As we shall see in Chapter VI, moduli in the symmetric case do occur; more than that, they occur in very low dimension. These issues will be important in both Case Studies 2 and 3.

Exercises

(*Note*: Exercises 1.1–1.5 are a block pertaining to the shallow arch. Some of the exercises require fairly messy calculations.)

1.1. (Discussion) The equilibria of the finite element analogue of the shallow arch may be found using the potential

$$V = \tfrac{1}{2}(\phi - x)^2 + \tfrac{1}{2}(\psi + x)^2 + l(\sin\phi + (\tfrac{1}{2} + \delta)\sin x).$$

Here we have scaled the spring constants to unity, and ϕ and ψ are the functions of x obtained by solving (implicitly) the equations (1.6). Now the function G in (1.7) equals $\partial V/\partial x$, so

$$G(x, l, \delta) = (\phi - x)(\phi' - 1) + (\psi + x)(\psi' + 1) + l(\cos\phi\,\phi' + (\tfrac{1}{2} + \delta)\cos x). \tag{1.22}$$

(a) Show that $G(-x, l, 0) = -G(x, l, 0)$. (*Hint*: Show that substituting $(-x, \psi(-x), \phi(-x))$ for $(x, \phi(x), \psi(x))$ also solves the system (1.6). Conclude using uniqueness of solutions in the implicit function theorem that $\psi(x) = \phi(-x)$. Use this fact, along with (1.6b), to show that V is even in x.)

(b) For $\delta = 0$, find the load l_0 at which the upward bulging solution ($x = 0$) becomes unstable.

(c) Show that for $l = l_0$ the governing bifurcation equation $G(x, l, 0)$ is equivalent to the normal form $-x^3 - \lambda x$; that is, show that C_1 and C_2 in (1.9) are both negative.

1.2. Prove that in (1.10) we have $\alpha'(0) < 0$. (*Hint*: Show that $\operatorname{sgn}\alpha'(0) = \operatorname{sgn}(\partial G/\partial\delta)$ at $(0, l_0, 0)$, and show that the latter derivative is negative by direct computation.)

1.3. (*Note*: In this exercise we ask the reader to complete the verification of (1.5) by comparing systems (1.8) and (1.13).)

(a) Let $\lambda_c(\delta)$ be defined by solving (1.15). Show that

$$\lambda_c(\delta) = -3\left(\frac{\alpha_1}{2}\right)^{2/3}\delta^{2/3} + O(\delta^{4/3}),$$

where, as in (1.11), $\alpha_1 = \alpha'(0)$. (*Hint*: The system (1.15) differs from (1.13), for which we have such an estimate, only by terms that are $O(\delta^2, \delta x^2)$.)

(b) Let $\lambda_c(\delta)$ and $l_c(\delta)$ be defined by solving (1.15) and (1.8), respectively. Let $(X(x, \lambda, \delta), \Lambda(\lambda, \delta))$ be the diffeomorphism in the equivalence transformation which relates $(G(x, l, \delta)$ and $H(x, \lambda, \delta)$. Show that

$$l_c(\delta) = \Lambda(\lambda_c(\delta), \delta)$$

and that

$$\Lambda(0, 0) = l_0.$$

(c) Combine (a) and (b) to obtain (1.5), where

$$C = 3\left(\frac{\alpha_1}{2}\right)^{2/3}\Lambda_\lambda(0, 0).$$

1.4. Let

$$F(x, \lambda, \delta, \varepsilon) = G(x, \lambda, \delta) + \varepsilon p(x, \lambda, \delta)$$

be a small perturbation of the shallow arch equations, as in (1.16). Assume that F is a universal unfolding of $G(x, \lambda, 0)$. Show that for ε fixed, small, and nonzero, the graph of singularities of F in the $\delta\lambda$-plane is a tilted cusp. (*Hint*: Use the uniqueness of universal unfoldings to factor F through the normal form universal unfolding, as in (1.18).)

1.5. Find the bifurcation diagrams for (1.20) associated to each of the four regions in Figure 1.6. (Cf. Chapter III, §7 for the pitchfork with the opposite sign, $+x^3 - \lambda x$.) (*Remarks*: It turns out that
(a) in regions 1 and 2, the associated diagrams have only one singularity, and this is a limit point; and
(b) in regions 3 and 4, the associated diagrams have exactly three singularities, and these are limit points.

From these two facts it may be deduced that
(a) when $\delta \neq 0$ in (1.7), collapse occurs at a limit point singularity; and
(b) if perturbing the curve Γ_0 in Figure 1.6 leads to a curve Γ which crosses either region 3 or 4, then the collapse load vs. δ curve is triple valued for a range of δ,
respectively.)

§2. The Classification Theorem

Theorem 2.1. *Let* $g(x, \lambda)$ *be a germ in* $\mathscr{E}_{x,\lambda}$ *satisfying* $g = g_x = 0$ at $(0, 0)$. *If* codim $g \leq 3$, *then* g is *equivalent to one of the bifurcation problems listed in Table* 2.1.

Remark. In association with the names listed in Table 2.1, let us introduce the following terminology. We shall apply the term *supercritical* to bifurcation problems such as $x^3 - \lambda x$ where the nontrivial solutions (i.e., $x \neq 0$) lie entirely to the right of the bifurcation point at $\lambda = 0$. Similarly, *subcritical*

Table 2.1. Normal Forms for Singularities of codim ≤ 3.

Normal Form		Codim	Nomenclature
(1)	$\varepsilon x^2 + \delta\lambda$	0	Limit point
(2)	$\varepsilon(x^2 - \lambda^2)$	1	Simple bifurcation
(3)	$\varepsilon(x^2 + \lambda^2)$	1	Isola center
(4)	$\varepsilon x^3 + \delta\lambda$	1	Hysteresis
(5)	$\varepsilon x^2 + \delta\lambda^3$	2	Asymmetric cusp
(6)	$\varepsilon x^3 + \delta\lambda x$	2	Pitchfork
(7)	$\varepsilon x^4 + \delta\lambda$	2	Quartic fold
(8)	$\varepsilon x^2 + \delta\lambda^4$	3	
(9)	$\varepsilon x^3 + \delta\lambda^2$	3	Winged cusp
(10)	$\varepsilon x^4 + \delta\lambda x$	3	
(11)	$\varepsilon x^5 + \delta\lambda$	3	

Note: ε and δ are either $+1$ or -1.

refers to cases such as $x^3 + \lambda x$ where the nontrivial solutions lie to the left of the bifurcation point, and transcritical to cases such as $x^2 - \lambda x$ where the nontrivial solutions lie on both sides of the bifurcation point.

We shall prove Theorem 2.1 later in §2, after some preliminary discussion of the theorem. The proof is based on a careful examination of the solution to the recognition problem. In preparation for this proof we have tabulated data for each of the normal forms in Table 2.1, as follows:

(i) Table 2.2: Algebraic data; i.e., $\mathscr{S}(h)$, RT(h), $\mathscr{P}(h)$, $T(h)$, a complement to $T(h)$, codimension.
(ii) Table 2.3: Solution to the recognition problem.

In making these tables, we took advantage of the fact that the eleven singularities in Table 2.1 may be divided into three families

$$\varepsilon x^k + \delta\lambda \ (k \geq 2)\text{: Numbers 1, 4, 7, 11,}$$
$$\varepsilon x^k + \delta\lambda x \ (k \geq 3)\text{: Numbers 6, 10,}$$
$$\varepsilon x^2 + \delta\lambda^k \ (k \geq 2)\text{: Numbers 2, 3, 5, 8,}$$

plus one singleton, the winged cusp: Number 9. We have already determined most of the data in the Tables 2.2 and 2.3 as follows:

$\varepsilon x^k + \delta\lambda \ (k \geq 2)$: Proposition II,9.1,
$\varepsilon x^k + \delta\lambda x \ (k \geq 3)$: Proposition II,9.2,
$\varepsilon x^2 + \delta\lambda^2$ (i.e., $k = 2$): Proposition II,9.3.
The winged cusp: Proposition II,9.4.

Table 2.2. Algebraic Data for Singularities of Codimension ≤ 3.

Normal Form	$\varepsilon x^k + \delta\lambda \ (k \geq 2)$	$\varepsilon x^k + \delta\lambda x \ (k \geq 3)$
$\mathscr{S}(h)$	$\mathscr{M}^k + \langle\lambda\rangle$	$\mathscr{M}^k + \mathscr{M}\langle\lambda\rangle$
$RT(h)$	$\mathscr{M}^k + \langle\lambda\rangle$	$\mathscr{M}^k + \mathscr{M}\langle\lambda\rangle$
$\mathscr{P}(h)$	$\mathscr{M}^{k+1} + \mathscr{M}\langle\lambda\rangle$	$\mathscr{M}^{k+1} + \mathscr{M}^2\langle\lambda\rangle + \langle\lambda^2\rangle$
$T(h)$	$\mathscr{M}^{k-1} + \langle\lambda\rangle + \mathbb{R}\{1\}$	$\mathscr{M}^k + \mathscr{M}\langle\lambda\rangle + \mathbb{R}\{x, k\varepsilon x^{k-1} + \delta\lambda\}$
Complement to $T(h)$	$\mathbb{R}\{x, x^2, \ldots, x^{k-2}\}$	$\mathbb{R}\{1, x^2, \ldots, x^{k-1}\}$ or $\mathbb{R}\{1, \lambda, x^2, \ldots, x^{k-2}\}$
codim h	$k - 2$	$k - 1$
Normal Form	$\varepsilon x^2 + \delta\lambda^k \ (k \geq 2)$	$\varepsilon x^3 + \delta\lambda^2$
$\mathscr{S}(h)$	$\mathscr{M}^2$	$\mathscr{M}^3 + \langle\lambda^2\rangle$
$RT(h)$	$\mathscr{M}^k + \mathscr{M}\langle x\rangle$	$\mathscr{M}^3 + \langle\lambda^2\rangle$
$\mathscr{P}(h)$	$\mathscr{M}^{k+1}$	$\mathscr{M}^4 + \mathscr{M}^2\langle\lambda\rangle$
$T(h)$	$\mathscr{M}^{k-1} + \langle x\rangle$	$\mathscr{M}^3 + \langle\lambda^2\rangle + \mathbb{R}\{\lambda, x^2\}$
Complement to $T(h)$	$\mathbb{R}\{1, \lambda, \ldots, \lambda^{k-2}\}$	$\mathbb{R}\{1, x, x\lambda\}$
codim h	$k - 2$	3

Table 2.3. Solution of the Recognition Problem for Singularities of Codimension ≤ 3.

Normal Form	Defining Conditions*	Nondegeneracy Conditions†
$\varepsilon x^k + \delta\lambda \; (k \geq 2)$	$g_{xx} = \cdots = \dfrac{\partial^{k-1} g}{\partial x^{k-1}} = 0$	$\varepsilon = \mathrm{sgn}\left(\dfrac{\partial^k g}{\partial x^k}\right)$, $\delta = \mathrm{sgn}(g_\lambda)$
$\varepsilon x^k + \delta\lambda x \; (k \geq 3)$	$g_{xx} = \cdots = \dfrac{\partial^{k-1} g}{\partial x^{k-1}} = g_\lambda = 0$	$\varepsilon = \mathrm{sgn}\left(\dfrac{\partial^k g}{\partial x^k}\right)$, $\delta = \mathrm{sgn}(g_{x\lambda})$
$\varepsilon(x^2 + \delta\lambda^2)$	$g_\lambda = 0$	$\varepsilon = \mathrm{sgn}(g_{xx})$, $\delta = \mathrm{sgn}(\det d^2 g)$
$\varepsilon x^2 + \delta\lambda^3$	‡$g_\lambda = \det(d^2 g) = 0$ choose $v \neq 0$ such that $g_{vv} = 0$	$\varepsilon = \mathrm{sgn}(g_{xx})$, $\delta = \mathrm{sgn}(g_{vvv})$
$\varepsilon x^2 + \delta\lambda^4$	‡$g_\lambda = \det(d^2 g) = g_{vvv} = 0$ choose $v \neq 0$ such that $g_{vv} = 0$	$\varepsilon = \mathrm{sgn}(g_{xx})$, $\delta = \mathrm{sgn}(q)$ where $q = q_{vvvv} \cdot g_{xx} - 3g_{vvx}^2$
$\varepsilon x^3 + \delta\lambda^2$	$g_\lambda = g_{xx} = g_{x\lambda} = 0$	$\varepsilon = \mathrm{sgn}(g_{xxx})$, $\delta = \mathrm{sgn}(g_{\lambda\lambda})$

* Defining conditions always include $g = g_x = 0$.

† We make the convention that expressions like $\delta = \mathrm{sgn}(g_\lambda)$ mean that $g_\lambda \neq 0$ and δ is equal to the sign of g_λ.

‡ The subscript v indicates a directional derivative in the direction v. (See text.)

In other words, the only missing information concerns the third family, $\varepsilon x^2 + \delta\lambda^k$, when $k \geq 3$. The algebraic data in Table 2.2 for the third family, $k \geq 3$, is easily computed; we leave this as an exercise for the reader.

Let us explain the notation g_{vvv} that occurs in Table 2.3 in connection with the third family when $k = 3$ or 4. In both these cases $\det d^2 g = 0$ is a defining condition for the singularity. In other words, zero is an eigenvalue of the Hessian of g. By $\partial/\partial v$ we mean a directional derivative along the eigenvector associated with the eigenvalue zero. For example, if

$$g(x, \lambda) = a(x + b\lambda)^2 + p,$$

where $a \neq 0$ and $p \in \mathscr{M}^3$, we may take the directional derivative $\partial/\partial v$ to be

$$\frac{\partial}{\partial v} = b\frac{\partial}{\partial x} - \frac{\partial}{\partial \lambda}.$$

There is an interesting observation concerning the normal forms $\varepsilon x^2 + \delta\lambda^k$ when k is large; namely, there are many intermediate-order terms. Specifically, $\mathscr{S} = \mathscr{M}^2$ and $\mathscr{P} = \mathscr{M}^{k+1}$, and all terms in between contribute to the solution of the recognition problem. Thus solving the recognition problem for $\varepsilon x^2 + \delta\lambda^k$ for all $k \geq 3$ would be a difficult task. (Unlike for the other two families, in Table 2.3 the recognition problem is solved for $\varepsilon x^2 + \delta\lambda^k$ only when $k = 2$, 3, and 4.)

In the following lemma we describe the general method for solving the recognition problem for the third family.

Lemma 2.2. *Let g be a germ in $\mathscr{E}_{x,\lambda}$ satisfying $g = g_x = g_\lambda = 0$ and*

$$\varepsilon = \operatorname{sgn}(g_{xx}) \neq 0.$$

Then there exist polynomial expressions $Q_2, Q_3, Q_4, \ldots$ in the derivatives of g evaluated at the origin, where Q_k depends only on $j^k g$, such that $Q_2 = \cdots = Q_{k-1} = 0$ and $\delta = \operatorname{sgn}(Q_k) \neq 0$ if and only if g is strongly equivalent to $\varepsilon x^2 + \delta\lambda^k$.

In other words, the recognition problem for $\varepsilon x^2 + \delta\lambda^k$ is solved by

$$g = g_x = g_\lambda = Q_2 = \cdots = Q_{k-1} = 0,$$
$$\varepsilon = \operatorname{sgn}(g_{xx}) \neq 0, \qquad \delta = \operatorname{sgn}(Q_k) \neq 0.$$

Remarks 2.3. (a) One consequence of this lemma bears on the proof of Theorem 2.1. Suppose that $g = g_x = g_\lambda = 0$, $g_{xx} \neq 0$, and codim $g = k$. Then g is strongly equivalent to $\varepsilon x^2 + \delta\lambda^{k+1}$. To see this, apply Lemma 2.2(b) inductively. If $Q_2 \neq 0$ then g is equivalent to $\varepsilon x^2 + \delta\lambda^2$ which has codimension one (since there is one defining condition beyond $g = g_x = 0$). If $Q_2 = 0$ and $Q_3 \neq 0$, then g is equivalent to $\varepsilon x^2 + \delta\lambda^3$, which has codimension two, etc.

(b) In Proposition II,9.3 we showed that

$$Q_2 = \varepsilon \det(d^2 g).$$

We claim that

$$Q_3 = g_{vvv} \quad \text{and} \quad Q_4 \equiv g_{vvvv}g_{xx} - 3g_{vvx}^2.$$

The explicit calculations verifying this claim are left as an exercise.

(c) Suppose that $g = a(x + b\lambda)^2 + p$ where $a \neq 0$ and $p \in \mathscr{M}^3$. As noted above, we may take $\partial/\partial v$ to be $b\,\partial/\partial x - \partial/\partial\lambda$. Thus, in general, all third-order derivatives of g contribute to $Q_3 = g_{vvv}$.

Proof of Lemma 2.2. Let $Q_2 = \varepsilon \det d^2 g$. Recall from Proposition II,9.3 that if $Q_2 \neq 0$ then g is strongly equivalent to $\varepsilon x^2 + \delta\lambda^2$ where $\delta = \operatorname{sgn} Q_2$. If $Q_2 = 0$, then

$$g \equiv a(x + b\lambda)^2 \pmod{\mathscr{M}^3}.$$

Letting $X(x, \lambda) = x + b\lambda$ and $S(x, \lambda) = |a|^{-1}$, we may change coordinates to deduce that g is strongly equivalent to

$$g_2 = \varepsilon x^2 + r_3, \tag{2.1}$$

where $r_3 \in \mathscr{M}^3$.

We now assume inductively that $Q_2, \ldots, Q_k$ have been defined with the property that if $Q_2 = \cdots = Q_k = 0$, then g is strongly equivalent to

$$g_k(x, \lambda) = \varepsilon x^2 + r_{k+1}, \tag{2.2}$$

where $r_{k+1} \in \mathcal{M}^{k+1}$. We write r_{k+1} in the form

$$r_{k+1} = A_k(x, \lambda)x + Q_{k+1}\lambda^k + s_{k+2},$$

where A_k is a homogeneous polynomial of degree k, Q_{k+1} is a scalar, and $s_{k+2} \in \mathcal{M}^{k+2}$. Now let

$$g_{k+1}(x, \lambda) = g_k\left(x - \frac{\varepsilon}{2} A_k(x, \lambda), \lambda\right)$$
$$= \varepsilon x^2 + Q_{k+1}\lambda^{k+1} + r_{k+2},$$

where $r_{k+2} \in \mathcal{M}^{k+2}$. If $Q_{k+1} \neq 0$, then g is strongly equivalent to $\varepsilon x^2 + \delta\lambda^{k+1}$ where $\delta = \operatorname{sgn} Q_{k+1}$; here we use the fact from Table 2.2 that

$$\mathcal{M}^{k+2} \subset \mathcal{P}(\varepsilon x^2 + \delta\lambda^{k+1}).$$

If $Q_{k+1} = 0$, then comparing with (2.2), we see that the induction continues.

It remains to remark that the exact value of Q_k at each stage depends only on sums and products of the terms in the Taylor expansion of g through order k; i.e., Q_k is a polynomial depending only on $j^k g$, as claimed. □

PROOF OF THEOREM 2.1. We proceed by enumerating all possible combinations of derivatives of g which could vanish, consistent with the requirement that codimension g be three or less. This enumeration is carried out in the flow chart of Table 2.4. The various arrows in this figure represent the solution to one of the recognition problems given in Table 2.3, except for the arrows leading to a box "codim $g \geq 4$." The latter arrows represent new information that must be derived here. The box in the first row follows directly from Remark 2.3(a). The remaining boxes are obtained using the algebraic estimates on $T(g)$ contained in Lemma 2.4. (In other words the arguments in this lemma constitute the only "new" idea in the proof of the classification theorem.)

Lemma 2.4. *Let g be a germ in $\mathcal{E}_{x,\lambda}$ with a singularity and let $g(x, 0) = x^l a(x)$. Then the following restrictions on $T(g)$ are valid:*

(a) $T(g) \subset \langle x^l, \lambda\rangle + \mathbb{R}\{g_x, g_\lambda\}$.
(b) *If $l = 5$ and $g_\lambda = 0$, then* $T(g) \subset \langle x^5, x\lambda, \lambda^2\rangle + \mathbb{R}\{g_x, g_\lambda\}$.
(c) *If $l = 4$ and $g_\lambda = g_{x\lambda} = 0$ then* $T(g) \subset \langle x^4, x^2\lambda, \lambda^2\rangle + \mathbb{R}\{g_x, g_\lambda\}$.
(d) *If $l = 3$ and $g_\lambda = g_{x\lambda} = g_{\lambda\lambda} = 0$ then* $T(g) \subset \mathcal{M}^3 + \mathbb{R}\{g_x, g_\lambda\}$.

Moreover, it follows in case (a) *that* codim $g \geq l - 2$ *and in cases* (b), (c), *and* (d) *that* codim $g \geq 4$.

Table 2.4. Flow Chart for Recognition Problem of Singularities of Codimension ≤ 3: A Proof of Theorem 2.1.

g_{xx}	$\xrightarrow{\neq 0} g_\lambda$	$\xrightarrow{=0} \det d^2 g$	$\xrightarrow{=0} d^3 g(v, v, v)$	$\xrightarrow{=0} q$	$\xrightarrow{=0}$ codim $g \geq 4$
$\downarrow =0$	$\downarrow \neq 0$: $\pm x^2 \pm \lambda$ (0)	$\downarrow \neq 0$: $\pm x^2 \pm \lambda^2$ (1)	$\downarrow \neq 0$: $\pm x^2 \pm \lambda^3$ (2)	$\downarrow \neq 0$: $\pm x^2 \pm \lambda^4$ (3)	
g_{xxx}	$\xrightarrow{\neq 0} g_\lambda$	$\xrightarrow{=0} g_{x\lambda}$	$\xrightarrow{=0} g_{\lambda\lambda}$	$\xrightarrow{=0}$ codim $g \geq 4$	
$\downarrow =0$	$\downarrow \neq 0$: $\pm x^3 \pm \lambda$ (1)	$\downarrow \neq 0$: $\pm x^3 \pm \lambda x$ (2)	$\downarrow \neq 0$: $\pm x^3 \pm \lambda^2$ (3)		
g_{xxxx}	$\xrightarrow{\neq 0} g_\lambda$	$\xrightarrow{=0} g_{x\lambda}$	$\xrightarrow{=0}$ codim $g \geq 4$		
$\downarrow =0$	$\downarrow \neq 0$: $\pm x^4 \pm \lambda$ (2)	$\downarrow \neq 0$: $\pm x^4 \pm \lambda x$ (3)			
g_{xxxxx}	$\xrightarrow{\neq 0} g_\lambda$	$\xrightarrow{=0}$ codim $g \geq 4$			
$\downarrow =0$	$\downarrow \neq 0$: $\pm x^5 \pm \lambda$ (3)				
codim $g \geq 4$					

* We assume $g = g_x = 0$.

† Numbers in parentheses after normal forms are codimensions.

PROOF. It is an easy task to verify the statements on codimension from the statement on containments. For example, consider (a). The codimension of $\langle x^l, \lambda\rangle$ is l. Since $T(g) \subset \langle x^l, \lambda\rangle + \mathbb{R}\{g_x, g_\lambda)$ it follows that

$$\operatorname{codim} T(g) \geq l - 2.$$

The containment statements are also easy to verify. Using Taylor's theorem, we may write

$$g(x, \lambda) = x^l a(x) + \lambda q(x, \lambda). \tag{2.3}$$

Thus

$$\begin{aligned} &\text{(a)} \quad xg_x = x^l \tilde{a}(x) + x\lambda q_x(x, \lambda),\\ &\text{(b)} \quad \lambda g_x = \lambda x^{l-1}\tilde{a}(x) + \lambda^2 q_x(x, \lambda),\\ &\text{(c)} \quad \lambda g_\lambda = \lambda q(x, \lambda) + \lambda^2 q_\lambda(x, \lambda). \end{aligned} \tag{2.4}$$

Recall from equation (III,2.10) that

$$T(g) = \langle g, xg_x, \lambda g_x\rangle + \mathbb{R}\{g_x, g_\lambda\} + \mathscr{E}_\lambda\{\lambda g_\lambda\}.$$

Thus the containment statements are verified if we can show that g, xg_x, λg_x, and λg_λ belong to the indicated ideals. For example, it follows from (2.3) and (2.4) that these germs all lie in $\langle x^l, \lambda\rangle$, thus proving (a). Using the stated hypotheses we may verify statements (b), (c) and (d) in a similar fashion. □

EXERCISE

2.1. Complete the classification of bifurcation problems with one state variable of codimension less than or equal to four. *Solution*: There are three new singularities of codimension four; $\varepsilon x^5 + \delta\lambda$, $\varepsilon x^5 + \delta\lambda x$, $\varepsilon x^2 + \delta\lambda^5$.

§3. Universal Unfoldings of the Elementary Bifurcations

It is a straightforward exercise using the data in Table 2.2 and the universal unfolding theorem (in the form of Corollary III,2.4) to determine universal unfoldings for the singularities listed in Table 2.1. We give these universal

Table 3.1. Universal Unfoldings for Elementary Bifurcations.

Universal Unfolding	Unperturbed Bifurcation Diagrams ($\varepsilon = 1$) $\delta = -1$	$\delta = +1$
(1) $\varepsilon x^2 + \delta\lambda$		
(2, 3) $\varepsilon(x^2 + \delta\lambda^2 + \alpha)$		
(4) $\varepsilon x^3 + \delta\lambda + \alpha x$		
(5) $\varepsilon x^2 + \delta\lambda^3 + \alpha + \beta\lambda$		
(6) $\varepsilon x^3 + \delta\lambda x + \alpha + \beta x^2$		
(7) $\varepsilon x^4 + \delta\lambda + \alpha x + \beta x^2$		
(8) $\varepsilon x^2 + \delta\lambda^4 + \alpha + \beta\lambda + \gamma\lambda^2$		
(9) $\varepsilon x^3 + \delta\lambda^2 + \alpha + \beta x + \gamma\lambda x$		
(10) $\varepsilon x^4 + \delta\lambda x + \alpha + \beta\lambda + \gamma x^2$		
(11) $\varepsilon x^5 + \delta\lambda + \alpha x + \beta x^2 + \gamma x^3$		

* Unstable solutions ($g_x < 0$) are indicated by dotted lines. Stable and unstable may be interchanged if $\varepsilon = -1$.

unfoldings in Table 3.1 along with pictures of the unperturbed bifurcation diagrams for $\varepsilon = +1$. Note that unstable solutions in these diagrams are indicated by dotted lines.

The solutions for the recognition problems for universal unfoldings of the elementary bifurcations are given in Table 3.2. We recall the statement of the recognition problem for universal unfoldings: Let h be a normal form of codimension k. Let g be equivalent to h and let G be a k-parameter unfolding of g. When is G a universal unfolding of g? As we showed in Chapter III, §4 the answer to this question takes the form of determining whether the determinant of a certain matrix (whose entries are derivatives of G) is nonzero. The exact form of the determinants for normal forms (4), (6), and (9) were computed in Chapter III, §4. We leave the computations of the remaining entries in Table 3.2 as exercises.

Table 3.2. The Recognition Problem for Universal Unfoldings of Singularities with Codimension ≤ 3.

	Normal Form	Matrix
(1)	$\varepsilon x^2 + \delta\lambda$	—
(2, 3)	$\varepsilon(x^2 + \delta\lambda^2)$	G
(4)	$\varepsilon x^3 + \delta\lambda$	$\begin{pmatrix} g_\lambda & g_{\lambda x} \\ G_\alpha & G_{\alpha x} \end{pmatrix}$
(5)	$\varepsilon x^2 + \delta\lambda^3$	$\begin{pmatrix} 0 & g_{xx} & g_{x\lambda} \\ G_\alpha & G_{\alpha x} & G_{\alpha\lambda} \\ G_\beta & G_{\beta x} & G_{\beta\lambda} \end{pmatrix}$
(6)	$\varepsilon x^3 + \delta\lambda x$	$\begin{pmatrix} 0 & 0 & g_{x\lambda} & g_{xxx} \\ 0 & g_{\lambda x} & g_{\lambda\lambda} & g_{\lambda xx} \\ G_\alpha & G_{\alpha x} & G_{\alpha\lambda} & G_{\alpha xx} \\ G_\beta & G_{\beta x} & G_{\beta\lambda} & G_{\beta xx} \end{pmatrix}$
(7)	$\varepsilon x^4 + \delta\lambda$	$\begin{pmatrix} g_\lambda & g_{\lambda x} & g_{\lambda xx} \\ G_\alpha & G_{\alpha x} & G_{\alpha xx} \\ G_\beta & G_{\beta x} & G_{\beta xx} \end{pmatrix}$
(8)	$\varepsilon x^2 + \delta\lambda^4$	$\begin{vmatrix} 0 & 0 & 0 & g_{xx} & g_{x\lambda} & g_{\lambda\lambda} \\ 0 & g_{xx} & g_{x\lambda} & g_{xxx} & g_{xx\lambda} & g_{x\lambda\lambda} \\ 0 & 0 & 0 & 0 & g_{xx} & 2g_{x\lambda} \\ G_\alpha & G_{\alpha x} & G_{\alpha\lambda} & G_{\alpha xx} & G_{\alpha x\lambda} & G_{\alpha\lambda\lambda} \\ G_\beta & G_{\beta x} & G_{\beta\lambda} & G_{\beta xx} & G_{\beta x\lambda} & G_{\beta\lambda\lambda} \\ G_\gamma & G_{\gamma x} & G_{\gamma\lambda} & G_{\gamma xx} & G_{\gamma x\lambda} & G_{\gamma\lambda\lambda} \end{vmatrix}$
(9)	$\varepsilon x^3 + \delta\lambda^2$	$\begin{pmatrix} 0 & 0 & g_{x\lambda} & g_{xxx} \\ 0 & g_{\lambda x} & g_{\lambda\lambda} & g_{\lambda xx} \\ G_\alpha & G_{\alpha x} & G_{\alpha\lambda} & G_{\alpha xx} \\ G_\beta & G_{\beta x} & G_{\beta\lambda} & G_{\beta xx} \end{pmatrix}$
(10)	$\varepsilon x^4 + \delta\lambda x$	$\begin{pmatrix} 0 & 0 & g_{x\lambda} & 0 & g_{xxxx} \\ 0 & g_{\lambda x} & g_{\lambda\lambda} & g_{\lambda xx} & g_{\lambda xxx} \\ G_\alpha & G_{\alpha x} & G_{\alpha\lambda} & G_{\alpha xx} & G_{\alpha xxx} \\ G_\beta & G_{\beta x} & G_{\beta\lambda} & G_{\beta xx} & G_{\beta xxx} \\ G_\gamma & G_{\gamma x} & G_{\gamma\lambda} & G_{\gamma xx} & G_{\gamma xxx} \end{pmatrix}$
(11)	$\varepsilon x^5 + \delta\lambda$	$\begin{pmatrix} g_\lambda & g_{\lambda x} & g_{\lambda xx} & g_{\lambda xxx} \\ G_\alpha & G_{\alpha x} & G_{\alpha xx} & G_{\alpha xxx} \\ G_\beta & G_{\beta x} & G_{\beta xx} & G_{\beta xxx} \\ G_\gamma & G_{\gamma x} & G_{\gamma xx} & G_{\gamma xxx} \end{pmatrix}$

§4. Transition Varieties and Persistent Diagrams

In this section we present the catalog of pictures associated with the universal unfoldings of the elementary bifurcation problems. Specifically, we include the following items:

(i) Table 4.1: Equations for the bifurcation, hysteresis, and double limit points,
(ii) Figures 4.1–4.3: Graphs of the transition varieties and graphs of the persistent perturbed bifurcation diagrams.

The formulas in Table 4.1 should serve as a guide for readers wishing to reproduce the pictures. We do not present the supporting calculations. (Some of the examples are worked out in detail in Chapter III.) In one case (viz., $\mathscr{D}$ for normal form (11)) we do not give a formula, since we did not determine $\mathscr{D}$ by explicitly solving the equations. However, it is not hard to determine the persistent diagrams associated with this normal form. See Golubitsky and

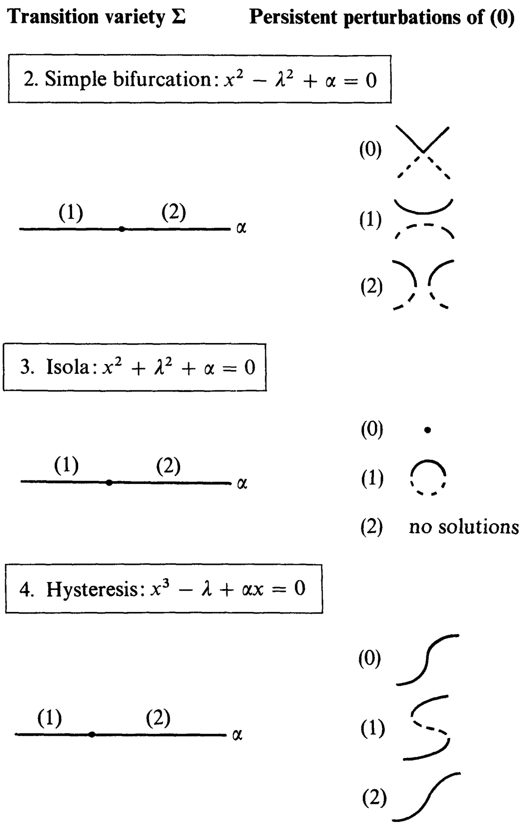


Figure 4.1. Persistent perturbations in codimension one.

Table 4.1. Equations for the Transition Variety for Singularities of Codimensions Two and Three.

	Normal Form	$\mathcal{B}$	$\mathcal{H}$	$\mathcal{D}$
(5)	$x^2 - \lambda^3 + \alpha + \beta\lambda$	$\left(\frac{\beta}{3}\right)^3 = \left(\frac{\alpha}{2}\right)^2$	$\varnothing$	$\varnothing$
(6)	$x^3 - \lambda x + \alpha + \beta x^2$	$\alpha = 0$	$\alpha = \beta^3/27$	$\varnothing$
(7)	$x^4 - \lambda + \alpha x + \beta x^2$	$\varnothing$	$\left(\frac{\alpha}{8}\right)^2 = -\left(\frac{\beta}{6}\right)^3$	$\alpha = 0, \beta \leq 0$
$(8)^+$	$x^2 + \lambda^4 + \alpha + \beta\lambda + \gamma\lambda^2$	$\alpha = 3\lambda^4 + \gamma\lambda^2, \beta = -4\lambda^3 - 2\gamma\lambda$	$\varnothing$	$\varnothing$
$(8)^-$	$x^2 - \lambda^4 + \alpha + \beta\lambda + \gamma\lambda^2$	$\alpha = -3\lambda^4 + \gamma\lambda^2, \beta = 4\lambda^3 - 2\gamma\lambda$	$\varnothing$	$\varnothing$
(9)	$x^3 + \lambda^2 + \alpha + \beta x + \gamma\lambda x$	$\alpha = 2x^3 - \frac{\gamma^2}{4}x^2, \beta = -3x^2 + \frac{\gamma^2}{2}x$	$\alpha\gamma^2 + \beta^2 = 0, \alpha \leq 0$	$\varnothing$
(10)	$x^4 - \lambda x + \alpha + \beta\lambda + \gamma x^2$	$\alpha + \gamma\beta^2 + \beta^4 = 0$	$\alpha + \frac{\gamma^2}{12} + \frac{8}{27}\gamma^3\beta^2 = 0$	$4\alpha = \gamma^2, \gamma \leq 0$
(11)	$x^5 - \lambda + \alpha x + \beta x^2 + \gamma x^3$	$\varnothing$	$\alpha = 15x^4 + 3\gamma x^2, \beta = -10x^3 - 3\gamma x$	*

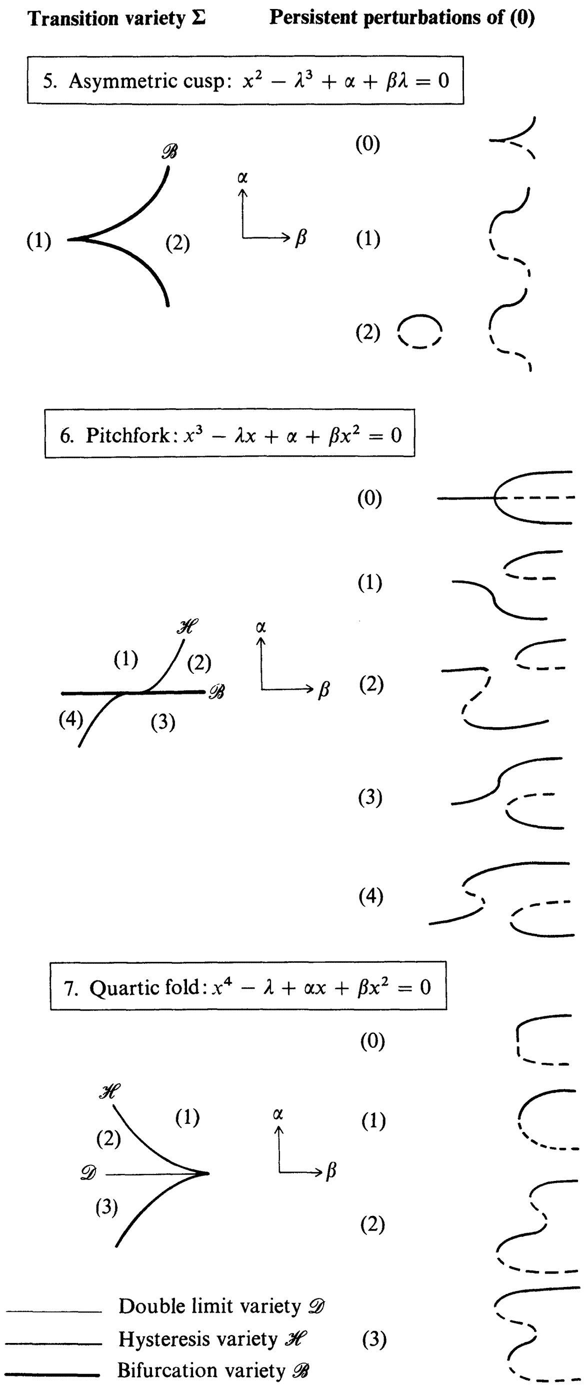
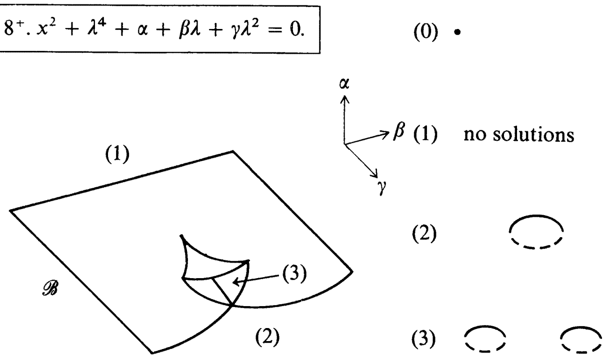


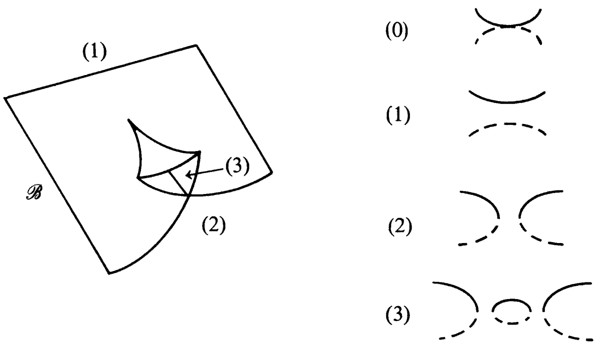
Figure 4.2. Persistent perturbations in codimension two.

Transition variety Σ **Persistent perturbations of (0)**

$8^+.\ x^2 + \lambda^4 + \alpha + \beta\lambda + \gamma\lambda^2 = 0.$



$8^-.\ x^2 - \lambda^4 + \alpha + \beta\lambda + \gamma\lambda^2 = 0$



9. The winged cusp: $x^3 + \lambda^2 + \alpha + \beta\lambda + \gamma x\lambda = 0$

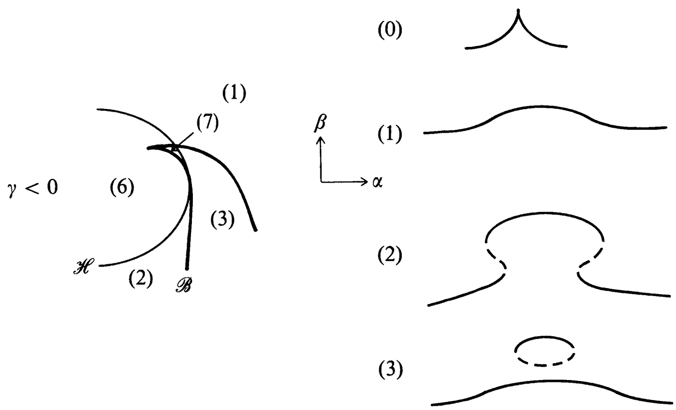


Figure 4.3. Persistent perturbations in codimension three.

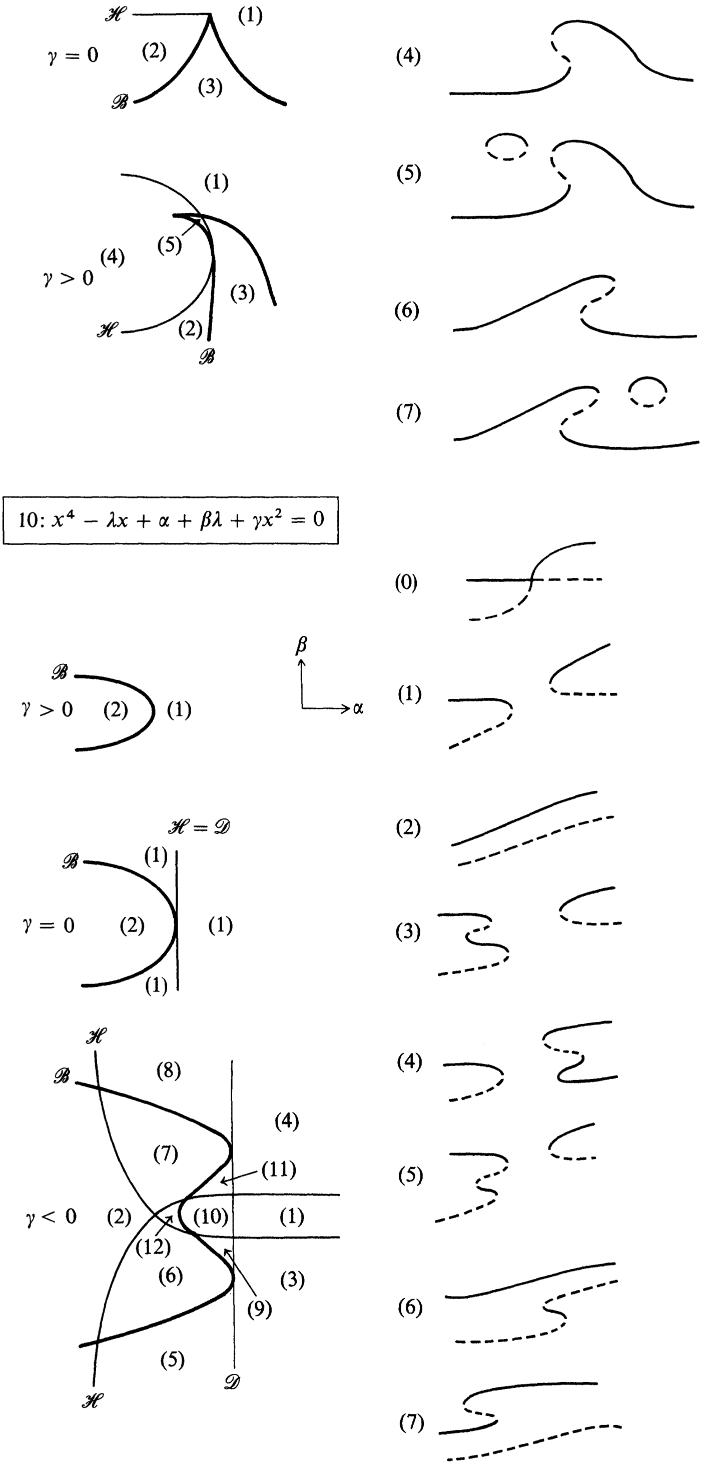
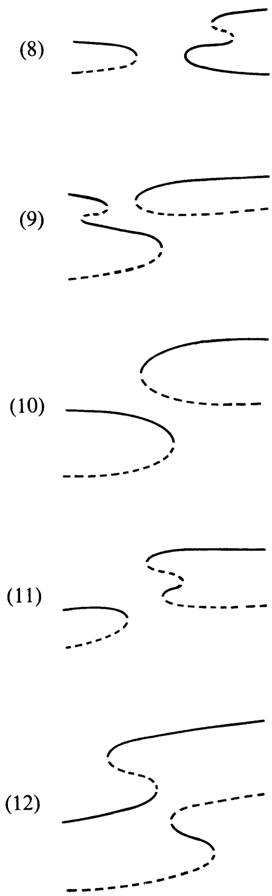


Figure 4.3 (*continued*)



11: $x^5 - \lambda + \alpha x + \beta x^2 + \gamma x^3 = 0$

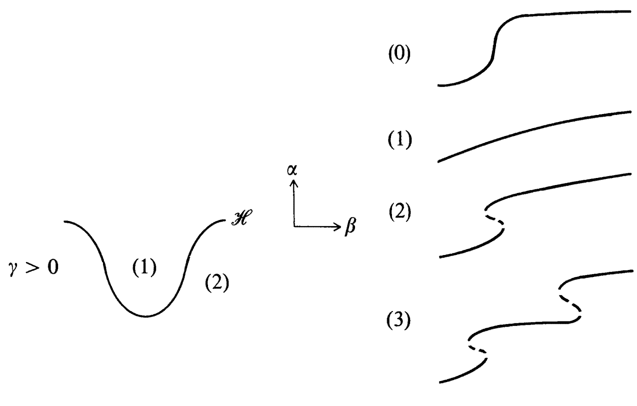


Figure 4.3 (*continued*)

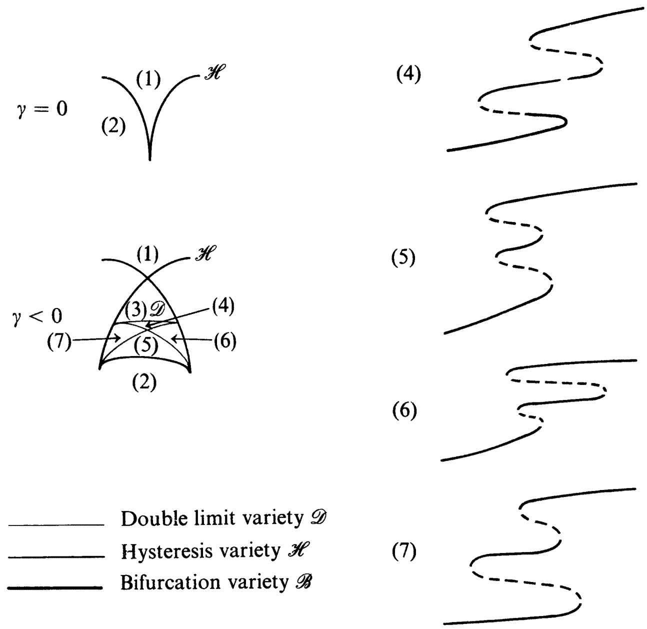


Figure 4.3 (*continued*)

Schaeffer [1979a], p. 52. From this information, we may piece together a description of $\mathscr{D}$.

In Figure 4.1, for each normal form, we assume that $\varepsilon = 1$ and, with one exception, make one choice of sign for δ. In every case except the normal form (8), $\varepsilon x^2 + \delta\lambda^4$, diagrams for the other choice of δ may be obtained by considering orientation reversing changes of coordinate; i.e., $x \to -x$ and/or $\lambda \to -\lambda$. For (8), we consider both $\delta = +1$ and $\delta = -1$. Similarly, the case $\varepsilon = -1$ can be derived by multiplying the equations by $S(x, \lambda) \equiv -1$, although this interchanges stable and unstable solution branches. In this way we avoid needless duplication of essentially identical figures. For the reader's convenience we show the unperturbed bifurcation diagram under the label "0."

Bibliographical Comments

René Thom [1972] was the first to emphasize the importance of codimension in classifying singularities. In particular, in the context of potential functions, the notorious list of seven elementary catastrophes arises from consideration of codimension—there are precisely seven catastrophes of codimension four or less. Cf. Mather [1969b] and Zeeman and Trotman [1975]. (This list has been extended to higher codimension, primarily by the Russian school. See Arnold [1976].) Thom selected the cut off of codimension four because he wanted points in unfolding space to be identified with points in physical space-time.

The classification of singularities changes if distinctions are made between various unfolding parameters. (Such distinctions first arose from identifying unfolding parameters with physical space-time and treating time differently from spatial coordinates.) Singularities of potential functions with a distinguished parameter are classified by Wassermann [1975]. Our classification of bifurcation problems is similar in that we consider a distinguished parameter but different in that we do not assume the existence of a potential function. All of this work derives from Mather's [1971] fundamental paper which considers mappings that have no distinguished parameter and that are not derivable from a potential function.

CHAPTER V

An Example of Moduli

§0. Introduction

In this chapter we analyze in detail bifurcation problems of the form

$$g(x, \lambda) = Ax^3 + B\lambda x^2 + C\lambda^2 x + D\lambda^3 + p(x, \lambda), \tag{0.1}$$

where $A, B, C, D \in \mathbb{R}$ and $p \in \mathscr{M}^4$. Such problems provide the first occurrence (i.e., lowest codimension) of moduli, and that is the reason we study (0.1). However, problems of the form (0.1) have codimension 5 or greater, so their analysis, especially exploration of parameter space, is not a simple matter. (*Remark:* In different mathematical contexts (e.g., bifurcation problems with symmetry) moduli occur in much lower codimension. For example, we analyze cases with codimension 3 in Chapter VI and in Case Study 3; in Volume II we shall encounter an example with codimension one where the one unfolding parameter is a modal parameter!)

In §1 of this chapter we motivate the occurrence of moduli in (0.1). Sections 2–4 contain the main analysis of (0.1); these sections consider in sequence the recognition problem for (0.1), universal unfolding of (0.1), and persistent perturbations of (0.1). As part of §3 we give a careful definition of moduli. In §5 we explore the moduli space associated to (0.1) more fully. In §6, we summarize the lessons we wish to draw from this chapter. Finally in §7, we briefly consider a mathematical model from the chemical engineering literature which leads to a singularity of the form (0.1). (In other words, in spite of the high codimension, moduli occur in real applications; this point is made more emphatically by Case Studies 2 and 3.)

§1. The Problem of Moduli: Smooth Versus Topological Equivalence

We recall Definition IV,1.1 which states, in essence, that two germs are topologically equivalent if there is a continuous change of coordinates mapping the zero set of one germ onto the zero set of the other. In Chapter IV, §1(b) we observed that moduli are associated with perturbations of a germ g that are topologically equivalent to g but not C^∞-equivalent. In this connection let us consider the one-parameter family of bifurcation problems

$$g_m(x, \lambda) = x(x + \lambda)(x - m\lambda), \tag{1.1}$$

where $m \in \mathbb{R}$; for simplicity we suppose $m > 0$. (Note that (1.1) is a special case of (0.1).) The zero set of (1.1) consists of three straight lines which intersect at the origin; viz.

$$\begin{aligned} &L_1(m)\colon \{x = 0\},\\ &L_2(m)\colon \{x = -\lambda\},\\ &L_3(m)\colon \{x = m\lambda\}. \end{aligned} \tag{1.2}$$

As we have indicated, this zero set depends on m, but from a qualitative point of view the dependence on m seems very mild indeed. More to the point, in the first lemma below we show that all the germs g_m in (1.1) for $m > 0$ are topologically equivalent. However, in the subsequent lemma we show that these bifurcation problems are not C^∞-equivalent. In other words, the parameter m in (1.1) is a modal parameter.

Lemma 1.1. *For any positive numbers m and n, the bifunction problems g_m and g_n are topologically equivalent.*

PROOF. Since equivalence is a transitive relationship, it suffices to show that for any $m > 0$, g_m is topologically equivalent to g_1. The following transformation maps the zero set of g_m onto that of g_1:

$$X(x, \lambda) = \begin{cases} \dfrac{1}{m}x & \text{if } x\lambda \geq 0,\\ x & \text{if } x\lambda \leq 0, \end{cases}$$

and $\Lambda(\lambda) = \lambda$. □

Remark. Lemma 1.1 may be extended as follows: Let $\{L_1, \ldots, L_k\}$ and $\{M_1, \ldots, M_k\}$ be two sets of k-distinct lines in $\mathbb{R}^2$ containing the origin. Then there exists a continuous, invertible change of coordinates $\Phi\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $\Phi(L_i) = M_i$, $i = 1, \ldots, k$. (Such a theorem may be proved by considering polar coordinates and interpolating on the unit circle.) In

Lemma 1.1 we took k to be three, but we also demanded that $\Phi = (X, \Lambda)$ have the special form where Λ does not depend on x.

Lemma 1.2. *Let m and n be positive numbers. The bifurcation problems, g_m and g_n, of the form* (1.1) *are C^∞-equivalent if and only if $m = n$.*

Proof. If two such germs are C^∞-equivalent, then there is a diffeomorphism $\Phi(x, \lambda) = (X(x, \lambda), \Lambda(\lambda))$ which maps the zero set of g_m onto that of g_n. Let $A = (d\Phi)_{0,0}$ be the differential of this map at the origin. We claim that A also maps one zero set onto the other. To see this, choose vectors $v_i \in L_i(m)$, $i = 1, 2, 3$. If $t \in \mathbb{R}$, then for each i there is some index j such that

$$\Phi(tv_i) \in L_j(n). \tag{1.3}$$

On differentiating (1.3) with respect to t at $t = 0$ we find

$$Av_i \in L_j(n), \tag{1.4}$$

which proves the claim.

In applying (1.4) it will be useful to have specific vectors in $L_i(m)$; thus we define

$$v_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} m \\ 1 \end{pmatrix}.$$

Next observe that A has an upper triangular form

$$A = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix}, \tag{1.5}$$

because $\Lambda(\lambda)$ does not depend on x. Moreover, the fact that equivalences are orientation preserving means that $a, c > 0$.

At this juncture we split the proof into three cases according to whether Av_1 belongs to $L_1(n)$, $L_2(n)$, or $L_3(n)$. In the text we consider only the first case; in Exercise 1.1 we ask the reader to analyze the other two cases. Given that $Av_1 \in L_1(n)$, we deduce that $b = 0$ in (1.5). Now let us ask whether Av_2 belongs to $L_2(n)$ or $L_3(n)$. (Since A is invertible, $Av_2 \in L_1(n)$ is impossible.) If $Av_2 \in L_3(n)$, then it follows that $n = -a/c$; since a and c are positive, this contradicts the hypothesis that $n > 0$. Thus $Av_2 \in L_2(n)$, which in turn implies that $a = c$ in (1.5); in other words, A is a multiple of the identity. To conclude, we use the fact that $Av_3 \in L_3(n)$. Since A is a multiple of the identity, we see that $v_3 \in L_3(n)$. However, $v_3 \in L_3(n)$ if and only if $m = n$. □

Remarks. (1) The C^∞-hypothesis in Lemma 1.2 is much stronger than needed. Indeed, the proof shows that it is impossible to find a map that is even once differentiable which sends the zero set of g_m to that of g_n unless $m = n$.

(2) The proof of Lemma 1.2 is related to the following geometric fact concerning linear maps: Let $\{L_1, \ldots, L_k\}$ and $\{M_1, \ldots, M_k\}$ be two sets of k

distinct lines in $\mathbb{R}^2$. Then there exists a linear mapping $A: \mathbb{R}^2 \to \mathbb{R}^2$ satisfying $A(L_i) = M_i$, $i = 1, \ldots, k$ if $k \leq 3$ but not, in general, if $k \geq 4$. See Exercises 1.2 and 1.3. In Lemma 1.2 the linear map A must conjugate *four* lines. Besides the three obvious lines associated with $g_m = 0$ and $g_n = 0$, A must also map the λ-axis into itself. This follows from our basic assumption that $\Lambda(\lambda)$ is independent of x.

(3) There is a well-known geometric invariant of four lines in the plane, called the cross ratio, with the following property: There exists a linear map sending one set of four lines to another such set precisely when the associated cross ratios are equal. In the context of (1.1), the number m is the cross ratio. Thus, in this case, we can give the modal parameter a geometric interpretation. A discussion of the cross ratio may be found in Ahlfors [1976], p. 78. See Exercise 1.3.

EXERCISES

1.1. Complete the proof of Lemma 1.2 by considering the possibilities $Av_1 \in L_j(n)$, $j = 2, 3$.

1.2. Let $\{L_1, L_2, L_3\}$, $\{M_1, M_2, M_3\}$ be two sets of three distinct lines in the plane. Show that there exists an invertible linear map $A: \mathbb{R}^2 \to \mathbb{R}^2$ satisfying $A(L_j) = M_j$, $j = 1, 2, 3$. *Hint:* Choose vectors $v_j \in L_j$, $w_j \in M_j$ ($j = 1, 2, 3$) such that $v_3 = v_1 + v_2$, $w_3 = w_1 + w_2$. Define $A(v_j) = w_j$, $j = 1, 2$.

1.3. Let $\mathscr{L} = \{L_1, L_2, L_3, L_4\}$ be an ordered set of four distinct lines in $\mathbb{R}^2$. Let l_j be the slope of L_j. Define the cross ratio of $\mathscr{L}$ to be

$$CR(\mathscr{L}) = \frac{(l_1 - l_2)(l_3 - l_4)}{(l_1 - l_3)(l_2 - l_4)}. \tag{1.6}$$

(a) Let $A: \mathbb{R}^2 \to \mathbb{R}^2$ be an invertible linear mapping. Define $A(\mathscr{L})$ to be the set $\{A(L_1), A(L_2), A(L_3), A(L_4)\}$. Verify that

$$CR(A(\mathscr{L})) = CR(\mathscr{L}). \tag{1.7}$$

(b) Use (1.7) to conclude that, in general, there does not exist a linear transformation mapping one set of four lines in the plane onto another.

§2. The Recognition Problem for Nondegenerate Cubics

Proposition 2.2 below is the main result of §2. This proposition draws on the following concept.

Definition 2.1. Let $\phi(x, \lambda) = Ax^3 + B\lambda x^2 + C\lambda^2 x + D\lambda^3$ be a homogeneous cubic. We call ϕ *nondegenerate* if the polynomial

$$\phi(x, 1) = Ax^3 + Bx^2 + Cx + D \tag{2.1}$$

has three distinct roots (not necessarily real).

If (2.1) has three roots, then, in particular, $A \neq 0$. Assuming $A \neq 0$, ϕ is nondegenerate unless (2.1) has a multiple root; i.e., unless for some x

$$\phi(x, 1) = \frac{\partial \phi}{\partial x}(x, 1) = 0. \tag{2.2}$$

Let $\phi(x, \lambda)$ be a homogeneous cubic such that $A \neq 0$. To better understand nondegeneracy, we make a preliminary change of coordinate to eliminate the $\lambda^2 x$ term in $\phi(x, \lambda)$ and to reduce the coefficient of x^3 to ± 1. Specifically let

$$\psi(x, \lambda) = \frac{1}{|A|}\phi\left(x - \frac{B}{3A}\lambda, \lambda\right); \tag{2.3}$$

then

$$\psi(x, \lambda) = \varepsilon(x^3 + 3c\lambda^2 x + 2d\lambda^3), \tag{2.4}$$

where $\varepsilon = \operatorname{sgn} A$,

$$c = \tfrac{1}{3}\left\{\frac{C}{A} - \frac{B^2}{3A^2}\right\}, \qquad d = \tfrac{1}{2}\left\{\frac{D}{A} - \frac{BC}{3A^2} + \frac{2B^3}{27A^3}\right\}. \tag{2.5}$$

Clearly ψ is nondegenerate if and only if ϕ is nondegenerate. Applying (2.2) to ψ we conclude that ψ is nondegenerate if and only if

$$c^3 + d^2 \neq 0. \tag{2.6}$$

Note that (2.3) defines an equivalence transformation. Thus for any bifurcation problem g of the form (0.1) with $A \neq 0$, we may apply (2.3) to simplify the calculations. After such a transformation the cubic terms in the Taylor series of g will be given by (2.4); in particular, in the following theorem ε, c, and d are computed from the coefficients of $j^3 g$ according to (2.5).

Proposition 2.2. *Let g be a bifurcation problem of the form* (0.1) *such that $j^3 g$ is nondegenerate.*

(i) *If $d = 0$, then g is equivalent to*

$$h(x, \lambda) = \varepsilon(x^3 + \delta\lambda^2 x), \tag{2.7}$$

where $\delta = \operatorname{sgn} c$.

(ii) *If $d \neq 0$, then g is equivalent to*

$$h_m(x, \lambda) = \varepsilon(x^3 - 3m\lambda^2 x + 2\delta\lambda^3), \tag{2.8}$$

where $\delta = \operatorname{sgn} d$ and

$$m = -\frac{c}{d^{2/3}}. \tag{2.9}$$

Remarks. (i) By (2.6), $m \neq 1$.

(ii) Formula (2.7) corresponds to the limit $m \to \pm\infty$ in (2.8); that is, $d \to 0$.

Proposition 2.3. *No two distinct bifurcation problems of the form* (2.7) *or* (2.8) *are equivalent. In other words:*

(i) (2.7) *is never equivalent to* (2.8).
(ii) *Two bifurcation problems of the form* (2.7) *are equivalent if and only if* $\delta_1 = \delta_2$.
(iii) *Two bifurcation problems of the form* (2.8) *are equivalent if and only if* $m_1 = m_2$ *and* $\delta_1 = \delta_2$.

PROOF OF PROPOSITION 2.2. We have already dealt with the lower-order terms by imposing the form (0.1) on g. Concerning the higher-order terms, below we shall prove the following: if $\phi(x, \lambda)$ is a nondegenerate homogeneous cubic then

$$\mathscr{M}^4 = \mathscr{P}(\phi). \tag{2.10}$$

Let us complete the proof given (2.10). Since $g = \phi + p$ where $p \in \mathscr{M}^4$, it follows from (2.10) that g is equivalent to ϕ. To handle ϕ, the intermediate order terms of g, we first apply (2.3) to ϕ, leading to (2.4). If $d = 0$, we reduce (2.4) to (2.7) by taking $\Lambda(\lambda) = |c|^{1/2}\lambda$; if $d \neq 0$, we reduce (2.4) to (2.8) by taking $\Lambda(\lambda) = |d|^{1/3}\lambda$.

It remains to prove (2.10). Let us show that

$$\mathscr{M} \cdot RT(\phi) = \mathscr{M}^4. \tag{2.11}$$

To begin we apply (2.3) to transform ϕ to ψ. Now $\mathscr{M} \cdot RT(\psi)$ is generated by the five functions

$$x\psi,\ \lambda\psi,\ x^2\psi_x,\ \lambda x\psi_x,\ \lambda^2\psi_x,$$

each of which is a homogeneous quartic. These generators can be written as linear combinations of the five generators $x^4, \lambda x^3, \lambda^2 x^2, \lambda^3 x, \lambda^4$ of $\mathscr{M}^4$ as follows:

	x^4	$x^3\lambda$	$x^2\lambda^2$	$x\lambda^3$	λ^4
$x\psi$	1	0	$3c$	$2d$	0
$\lambda\psi$	0	1	0	$3c$	$2d$
$\frac{1}{3}x^2\psi_x$	1	0	c	0	0
$\frac{1}{3}\lambda x\psi_x$	0	1	0	c	0
$\frac{1}{3}\lambda^2\psi_x$	0	0	1	0	c

The determinant of this 5×5 matrix is $4(c^3 + d^2)$, which is nonzero by (2.6). Thus (2.11) follows from Lemma II,4.2.

Recall that

$$\mathscr{I}(\psi) = \mathscr{M} \cdot RT(\psi) + \mathbb{R}\{\lambda\psi_x\}.$$

From (2.11) we have

$$\operatorname{Itr} \mathscr{I}(\psi) = \mathscr{M}^4 = \mathscr{P}(\psi)$$

and hence

$$\mathscr{P}(\phi) = \mathscr{M}^4$$

which proves (2.10). □

The proof of Proposition 2.3 is based on examining when two cubic polynomials are equivalent modulo higher-order terms. In this case the set of higher-order terms, $\mathscr{P}$, is $\mathscr{M}^4$, and only linear terms in the equivalences need be considered. We ask the reader to supply the details of the proof in Exercise 2.1.

EXERCISE

2.1. Complete the proof of Proposition 2.3.

§3. Universal Unfolding; Relation to Moduli

For brevity we only consider the normal form (2.8). In Exercise 3.1 we ask the reader to derive the universal unfolding of (2.7).

Proposition 3.1. *A universal unfolding for* (2.8) *when* $m \neq 1$ *is given by*

$$H(x, \lambda, \alpha_1, \alpha_2, \alpha_3, \alpha_4, n) = \varepsilon\{(x^3 - 3n\lambda^2 x + 2\delta\lambda^3) + \alpha_1 + \alpha_2\lambda + \alpha_3 x + \alpha_4\lambda x\}. \tag{3.1}$$

In particular, h has codimension 5.

Remarks. (i) We use a different notation for the fifth unfolding parameter in (3.1) because, as we show below, n is a modal parameter.

(ii) Since (3.1) is to be a small perturbation of (2.8), we suppose that $\alpha_i \approx 0$ and $n \approx m$.

PROOF OF PROPOSITION 3.1. Since $\mathscr{M}^4 \subset RT(h)$, it follows that

$$T(h) = \mathscr{M}^4 \oplus \mathbb{R}\{x^3 - 3m\lambda^2 x + 2\delta\lambda^3, x^3 - m\lambda^2 x, x^2\lambda - m\lambda^3, x^2 - m\lambda^2, m\lambda x - \delta\lambda^2\}.$$

It is easily seen that $\{x\lambda^2, \lambda x, x, \lambda, 1\}$ spans a complementary subspace to $T(h)$ in $\mathscr{E}_{x,\lambda}$. Thus codim $h = 5$ and (3.1) provides a universal unfolding of h. □

We now discuss moduli more carefully. Let $G(x, \lambda, \alpha)$, where $\alpha \in \mathbb{R}^k$, be a universal unfolding of a germ g of finite codimension. (In contrast to (3.1), here we include all unfolding parameters in the vector α, even possible modal parameters.) For many $\alpha \in \mathbb{R}^k$ different from zero, there will be points (x, λ) where $G(\cdot, \cdot, \alpha)$ has a singularity with positive codimension; indeed, this occurs precisely when α belongs to the transition set Σ introduced in Definition III,5.1. In our examples before the present chapter, all singularities for $\alpha \neq 0$ have had lower codimension than the fundamental singularity at $\alpha = 0$ that is being unfolded. In this connection the winged cusp of Chapter III, §8 is a good example to consider. Moduli are associated with the breakdown of such behavior, i.e., there are singularities for $\alpha \neq 0$ of the same codimension as for $\alpha = 0$. This idea is the basis of the following definition.

Definition 3.2. Let $G(x, \lambda, \alpha)$ be a universal unfolding of g. The *codimension constant variety* $\mathscr{C}$ is given by the formula

$$\mathscr{C} = \{\alpha \in \mathbb{R}^k : \exists (x_0, \lambda_0) \text{ near } (0, 0) \text{ such that } \operatorname{codim} G(x + x_0, \lambda + \lambda_0, \alpha) = \operatorname{codim} g\}.$$

Remarks. (i) In this definition $G(x + x_0, \lambda + \lambda_0, \alpha)$ denotes the germ

$$(x, \lambda) \mapsto G(x + x_0, \lambda + \lambda_0, \alpha),$$

where (x, λ) is close to $(0, 0)$. This awkward notation results from our convention that germs are defined on neighborhoods of the origin rather than on neighborhoods of a more general base point. Basically in this definition we want to restrict $G(\cdot, \cdot, \alpha)$ to a small neighborhood of a possible singularity at (x_0, λ_0).

(ii) Of course $\mathscr{C}$ is contained in Σ, the transition variety.

In the following proposition we apply this definition to the universal unfolding (3.1). (Exercise 3.2 contains a simpler, although highly academic, illustration of how to work with Definition 3.2.) In the proposition we return to the notational convention of letting $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ refer only to nonmodal parameters.

Proposition 3.3. *For the universal unfolding* (3.1)

$$\mathscr{C} = \{(\alpha_1, \alpha_2, \alpha_3, \alpha_4, n) \in \mathbb{R}^5 : \alpha_i = 0, i = 1, 2, 3, 4\}.$$

PROOF. It is easily seen that if $\alpha = 0$ then (3.1) has a singularity of codimension 5 at the origin. Indeed, if $\alpha = 0$ then (3.1) simply reproduces the normal form (2.8) with a different value of m. In other words, points (α, n) where $\alpha = 0$ belong to $\mathscr{C}$. The proof that $\mathscr{C}$ contains only points of this form is more involved. This proof is based on the recognition problems whose solutions were summarized in Table IV,2.3 while proving the classification theorem.

Suppose that (x_0, λ_0) is a singular point of $p(x, \lambda) = H(x, \lambda, \alpha, n)$. Then $p = p_x = 0$ at (x_0, λ_0). Observe that $p_{xxx}(x_0, \lambda_0) \neq 0$. If $p_\lambda(x_0, \lambda_0) \neq 0$, then $\operatorname{codim} p \leq 1$ since p is equivalent to either $\pm x^2 \pm \lambda$ or $\pm x^3 \pm \lambda$ near (x_0, λ_0). So we assume that $p_\lambda(x_0, \lambda_0) = 0$.

Now suppose that $p_{xx}(x_0, \lambda_0) \neq 0$. If $\det d^2p \neq 0$ at (x_0, λ_0) then p is equivalent to $\pm x^2 \pm \lambda^2$ and $\operatorname{codim} p = 1$. Thus, we assume that det $d^2p = 0$ at (x_0, λ_0). The following trick will be of great help in these calculations. Make the change of coordinates $y = x - x_0$, $\mu = \lambda - \lambda_0$. It follows that $p = p_y = p_\mu = \det d^2p = 0$ at $(0, 0)$ and $p_{yy}(0, 0) \neq 0$. This implies that the quadratic terms in $p(y, \mu)$ have the form $l(y + q\lambda)^2$. Letting $z = y + q\lambda$ we see that

$$p(z, \mu) = z^3 + rz^2\mu + sz\mu^2 + t\mu^3 + lz^2$$

with $l \neq 0$. Note that the cubic terms in $p(x, \mu)$ are still nondegenerate as we have only made linear changes of coordinates in going from (x, λ) to (z, μ). We now use the solution to the recognition problems for $\pm x^2 \pm \lambda^3$ and $\pm x^2 \pm \lambda^4$ given in Table IV,2.3. In particular, if $t \neq 0$, then p is equivalent to $\pm x^2 \pm \lambda^3$ and has codimension 2. If $t = 0$, then $s \neq 0$, since the cubic terms in p are nondegenerate; in this case it may be shown that p is equivalent to $\pm(x^2 - \lambda^4)$ and p has codimension 3. Thus all the singularities which occur when $p_{xx}(x_0, \lambda_0) \neq 0$ have codimension ≤ 3.

Finally, we consider the case when $p_{xx}(x_0, \lambda_0) = 0$. In fact, we have now assumed that $p = p_x = p_\lambda = p_{xx} = 0$ at (x_0, λ_0). If $p_{x\lambda}(x_0, \lambda_0) \neq 0$, then p is the pitchfork and has codimension 2. If $p_{x\lambda}(x_0, \lambda_0) = 0$ and $p_{\lambda\lambda}(x_0, \lambda_0) \neq 0$, then p is equivalent to the winged cusp and has codimension 3. If $p = p_x = p_\lambda = p_{xx} = p_x = p_{\lambda\lambda} = 0$ at (x_0, λ_0) then $(x_0, \lambda_0) = (0, 0)$ and $p(x, \lambda) = h_n(x, \lambda)$ which is the singularity of codimension 5 considered at the beginning of this discussion. Thus $\mathscr{C}$ is the n-axis as claimed. □

In most of the examples in this book the codimension constant variety is a smooth submanifold of $\mathbb{R}^k$; i.e., the unfolding parameters may be chosen such that for some $l < k$,

$$\mathscr{C} = \{\alpha \in \mathbb{R}^k : \alpha_i = 0, i = 1, \ldots, l\}. \tag{3.2}$$

If (3.2) obtains we shall say that the singularity g has modality $k - l$, and we shall refer to $\alpha_{l+1}, \ldots, \alpha_k$ as either moduli or modal parameters. If $\mathscr{C}$ is *not* a manifold (so that no formula such as (3.2) exists), we must interpret these concents using ideas from algebraic geometry. See Exercise 5.2 for an example where $\mathscr{C}$ is not a manifold.

Exercises

3.1. Find the universal unfolding of (2.7), $h(x, \lambda) = \varepsilon(x^3 + \delta\lambda^2 x)$.

3.2. Let

$$F(x, \lambda, \alpha) = x^3 - \lambda x + \alpha_1 + \alpha_2 x + \alpha_3 x^2,$$

which is a versal unfolding of the pitchfork $x^3 - \lambda x$. Note that F depends on three parameters but the pitchfork only has codimension two. In analogy with Definition 3.2, let

$$\mathscr{C} = \{\alpha \in \mathbb{R}^3 : \exists(x_0, \lambda_0) \text{ near } (0, 0) \text{ such that codim } F(x + x_0, \lambda + \lambda_0, \alpha) = 2\}.$$

(This is only an analogy because in Definition 3.2 we have required that G be a *universal* unfolding; i.e., that G contain only the minimum number of unfolding parameters.)

Show that

$$\mathscr{C} = \{\alpha \in \mathbb{R}^3 : \alpha_1 = \alpha_3 = 0\}.$$

(*Remark*: In this simple example, the remaining parameter, α_2, is *not* a modal parameter because the various bifurcation problems $F(\cdot, \cdot, \alpha)$ obtained by varying α_2 are all equivalent.)

§4. Persistent Perturbed Diagrams

In this section we tabulate the persistent bifurcation diagrams that occur in (3.1), the universal unfolding of (2.8). The first, rather obvious, point to make is that the bifurcation diagram of (2.8) consists of one or three straight lines according as $m < 1$ or $m > 1$, respectively. *A fortiori*, the perturbed bifurcations are different for $m < 1$ and $m > 1$. The two cases are illustrated for $\delta > 0$ in Figures 4.1 and 4.2, respectively. It is far less obvious that the perturbed diagrams for $m < 1$ depend on whether $m < 0$ or $0 < m < 1$. Figure 4.1 covers both cases—diagrams common to these cases are shown at the top of the figure, above those which are different for the two cases. (Although we have illustrated only the case $\delta = +1$, the diagrams with $\delta = -1$ are simply mirror images; this may be seen from the fact that the substitution $\lambda \to -\lambda$ reduces one normal form to the other.)

The determination of all these perturbed diagrams is a rather tedious calculation which we omit here. We refer to Golubitsky *et al.* [1981] and to Stewart [1981], for the case $m < 1$; to Stewart and Woodcock [1982, 1983] for the case $m > 1$. However, it is useful to explain briefly the method of calculation; in particular, the use of the letters L and R in Figures 4.1 and 4.2 is related to this method of calculation. Moreover, this construction will allow us to prove that the sets of perturbed bifurcation diagrams for the cases $m < 0$ and $0 < m < 1$ are different. This fact is perhaps the most important item in this section.

The calculation of perturbed diagrams is based on the path formulation discussed in Chapter III, §12. In particular, we recall Figure III,12.1 which graphs the cusp surface in three-dimensional (x, A, B) space defined by

$$x^3 - Bx + A = 0. \tag{4.1}$$

To relate (4.1) to (3.1), we rewrite (3.1) as

$$H(x, \lambda, \alpha, n) = \varepsilon\{x^3 + (\alpha_3 + \alpha_4\lambda - 3n\lambda^2)x + (\alpha_1 + \alpha_2\lambda + 2\delta\lambda^3)\}. \tag{4.2}$$

For given values of the unfolding parameters α, n, we regard (4.2) as a one-parameter family of equations (4.1) depending on the one parameter λ, where

$$A(\lambda) = \alpha_1 + \alpha_2\lambda + 2\delta\lambda^3, \qquad B(\lambda) = -\alpha_3 - \alpha_4\lambda + 3n\lambda^2. \tag{4.3}$$

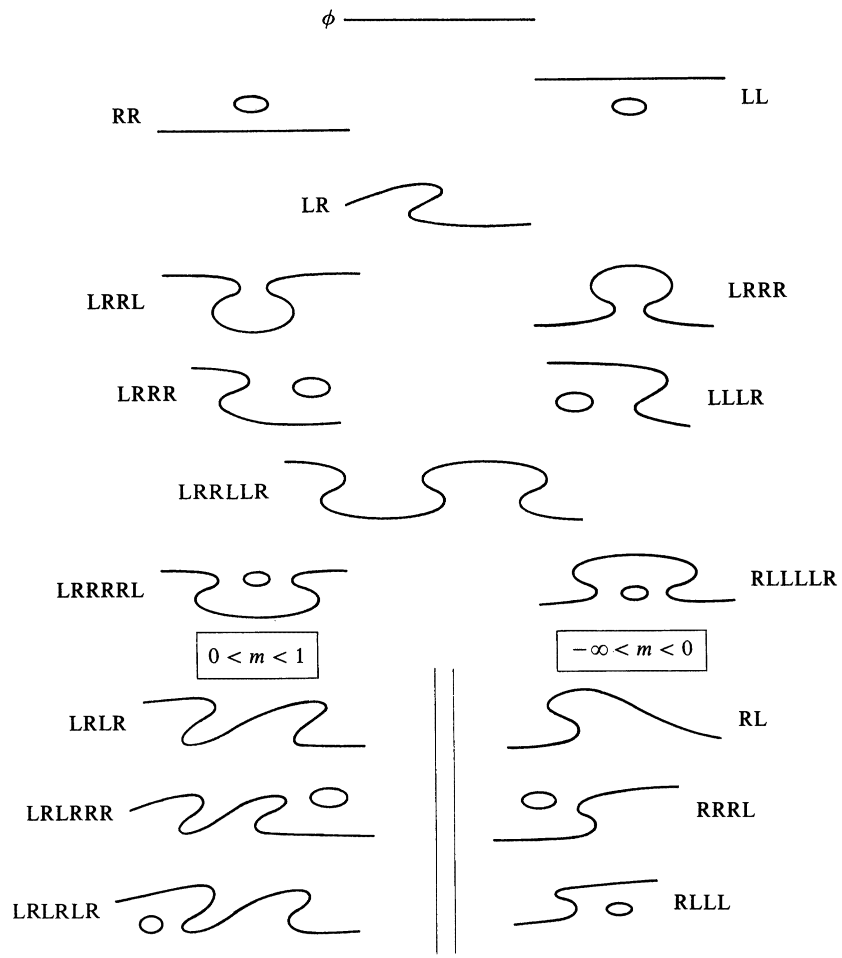


Figure 4.1. Persistent perturbations of $x^3 - 3mx\lambda^2 + 2\lambda^3$, $m < 1$.

In Figure 4.3, we have drawn the path (4.3) when $\alpha = 0$, $n = m$ as a dashed line in the A, B-plane for three cases: (i) $m < 0$, (ii) $0 < m < 1$, and (iii) $m > 1$. (*Remark:* This figure indicates why the cases $m < 0$ and $0 < m < 1$ might be different.) The solid line in the figure represents the singular curve in the projection of (4.1); i.e., the cusp

$$\left(\frac{A}{2}\right)^2 = \left(\frac{B}{3}\right)^3. \tag{4.4}$$

Let us show how to construct the bifurcation diagram associated to a path (4.2) given the sequence of intersections of the path with the left- and right-hand nappes of the cusp (4.4). We denote such intersections by a sequence of L and R's. For example, the sequence LRLR is a shorthand for

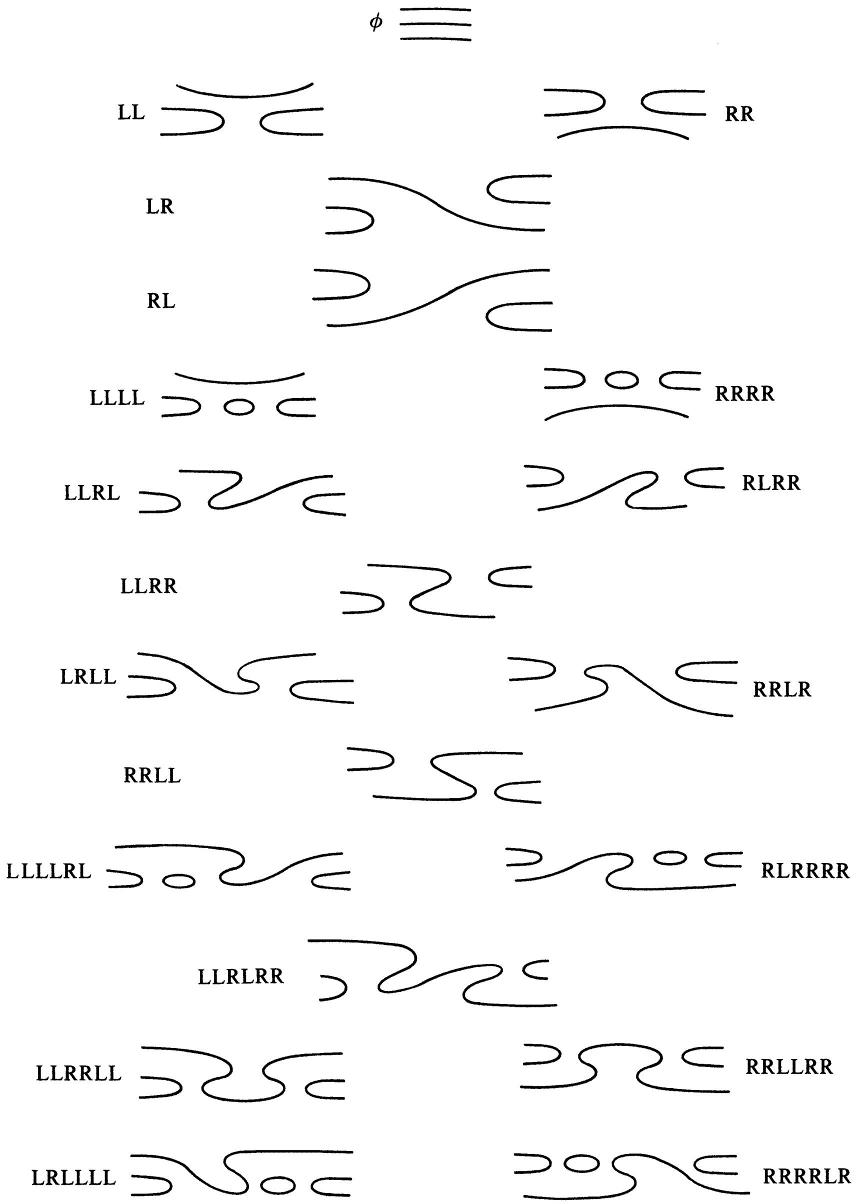


Figure 4.2. Persistent perturbations of $x^3 - 3mx\lambda^2 + 2\lambda^3$, $m > 1$.

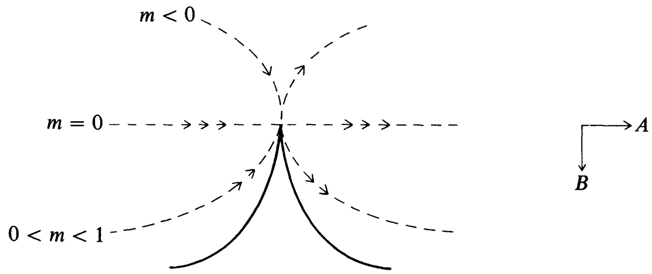


Figure 4.3. Paths representing the cross ratio $m < 1$.

the path shown in Figure 4.4. The bifurcation diagram associated to this path is also shown in the figure. The following three principles suffice for constructing bifurcation diagrams from the path (4.3):

(i) For a given λ, there are one or three points x on the bifurcation diagram according as $(A(\lambda), B(\lambda))$ is outside or inside the cusp (4.4).
(ii) If a path crosses (4.4), two new solutions appear or disappear on the bifurcation diagram; in other words, limit points are associated with crossings of (4.4).
(iii) Such limit points are associated with the bottom two solution branches for crossings of the left nappe; the top two, the right nappe.

We now use this information to prove that persistent perturbations of h_m for $m < 0$ are different from those of h_m for $0 < m < 1$. We begin by observing that $B(\lambda)$ in (4.3) is a quadratic polynomial in λ; hence $B'(\lambda) = 0$ is satisfied at precisely one point λ_0. We claim that it is impossible to obtain the sequence LRLR from (4.3) when $m < 0$. Such a path would have to start in the second quadrant and end in the first quadrant (see Figure 4.3) and look like the path in Figure 4.4 in the middle. For this to happen, B' would have to vanish at three or more values of λ. On the other hand, if $0 < m < 1$, then the perturbation

$$A(\lambda) = 2(\lambda^3 - t\lambda); \qquad B(\lambda) = 3(m\lambda^2 + t)$$

has the intersection sequence LRLR for every $t > 0$.

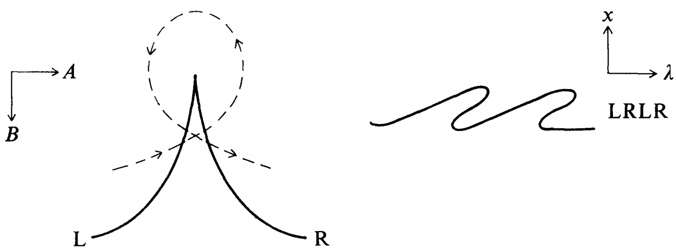


Figure 4.4. A perturbed path and the associated bifurcation diagram.

§5. A Picture of the Moduli Family

In this section we describe some of the complicated behavior that is associated with moduli. We hope that our abstractions will help the reader to synthesize the information contained in a moduli family. We do not attempt to prove that our abstractions are valid for all moduli families, even those of modality one.

In Proposition 2.2 we showed that each nondegenerate cubic is equivalent to precisely one of the two normal forms (2.7) and (2.8) which we repeat here:

$$\begin{aligned} &\text{(a)} \quad h(x, \lambda) = \varepsilon(x^3 + \sigma\lambda^2 x), \\ &\text{(b)} \quad h_m(x, \lambda) = \varepsilon(x^3 - 3m\lambda^2 x + 2\delta\lambda^3), \qquad m \neq 1. \end{aligned} \tag{5.1}$$

We recall that ε, σ, and δ represent choices of signs and m is a real parameter.

This section is divided into five parts; we summarize the information contained in each part.

(a) The moduli family consists of two circles, one for $\varepsilon = +1$ and one for $\varepsilon = -1$. These circles are pictured in Figure 5.1. The right half of each circle corresponds to $\delta = +1$; the left half to $\delta = -1$. On each circle the right and left halves are joined at the points where $m = +\infty$ and $m = -\infty$; the points at which $m = +\infty$ correspond to (5.1a) with $\sigma = -1$, and the points at which $m = -\infty$ correspond to (5.1a) with $\sigma = +1$.

(b) The moduli circles are divided into disjoint arcs by distinguished points. For any two points on an arc between a pair of distinguished points, the two associated germs are topologically equivalent and exactly the same persistent perturbations occur in the universal unfoldings of both germs. We describe such behavior by saying "the universal unfolding is topologically trivial". By *distinguished points* we mean points in the moduli family where topological triviality of the universal unfolding fails.

We have already discussed in §§3 and 4 why the points where m is 0 or 1 are distinguished. In subsection (b) we will show that the points where $m = \pm\infty$ are also distinguished.

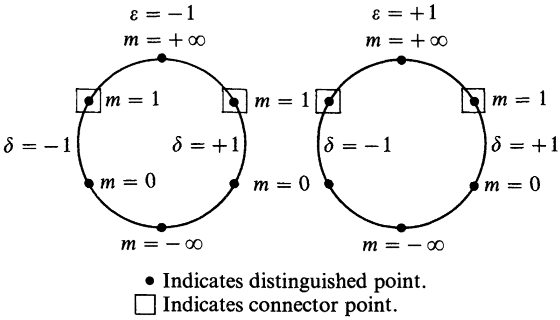


Figure 5.1. Circles of moduli with distinguished points.

As to the remaining points, it follows from the analysis of Stewart [1981] that along the arcs in Figure 5.1 the universal unfoldings are topologically trivial. His method of proof is to show that the transition set Σ in the universal unfolding of h_m in (5.1) remains constant, up to a continuous change of coordinates, along any interval in m not containing one of the distinguished points. (Cf. §6(a) for further discussion of topological triviality.)

(c) There are two types of distinguished points: regular distinguished points and connector points. In example (5.1), the distinguished points where m is 0 or $\pm\infty$ are regular. To understand what this means, let us refer to Proposition 2.2. In that proposition we showed the following: if g_1 and g_2 are germs whose linear and quadratic terms vanish and whose cubic terms yield identical values of ε, δ, and m, then g_1 and g_2 are equivalent, as long as $m \neq 1$. In particular, any germ that yields $m = 0$ is equivalent to h_0 in (5.1b) with the appropriate signs for ε and δ, and similarly for $m = \pm\infty$ and (5.1a). In general, we call a distinguished point *regular* if the modal value uniquely determines a singularity (more precisely, an equivalence class of singularities).

By contrast, a *connector point* corresponds to a value of the modal parameter having at least two inequivalent singularities which share that value. In subsection (c) we show that the points on the moduli circles where $m = 1$ are connector points. This fact is indicated by the boxes in Figure 5.1.

(d) In subsection (d) we discuss the connector points where $m = 1$ more fully. Let us summarize the issues. Suppose that g is a singularity corresponding to one of these connector points and that G is a universal unfolding of g. Consider the codimension constant variety of G; this will be contained in the set of points in the moduli family near the connector point. It turns out that the codimension constant variety of G consists of exactly two arcs of nondistinguished points. In this situation we say that these two arcs of nondistinguished points are *connected* by the singularity g. For example, we shall show that the two arcs

$$\varepsilon = +1, \qquad \delta = -1, \qquad 1 < m < \infty,$$

and

$$\varepsilon = +1, \qquad \delta = -1, \qquad 0 < m < 1,$$

are connected by the (codimension 5) singularity

$$(x^3 - 3\lambda^2 x - 2\lambda^3) + \lambda^4.$$

For another singularity $\bar{g}$, also corresponding to $m = 1$, a different pair of arcs may be connected in this way. For example, we shall show that the arcs

$$\varepsilon = -1, \qquad \delta = +1, \qquad 1 < m < \infty,$$

and

$$\varepsilon = +1, \qquad \delta = -1, \qquad 1 < m < \infty,$$

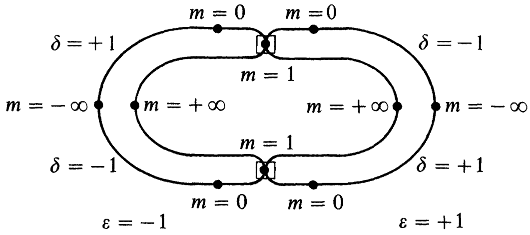


Figure 5.2. The moduli bracelet.

are connected by the (codimension 5) singularity

$$(-x^2\lambda + \lambda^3) + x^4.$$

Note that this singularity connects *arcs on different circles in Figure* 5.1. In other words, it is natural to identify the points $\varepsilon = +1$, $\delta = -1$, $m = 1$ and $\varepsilon = -1$, $\delta = +1$, $m = 1$ in Figure 5.1. A similar identification holds for $\varepsilon = +1$, $\delta = +1$, $m = 1$ and $\varepsilon = -1$, $\delta = -1$, $m = 1$. The resulting topological space is indicated in Figure 5.2. We refer to this space as a "bracelet."

At a connector point P, for some pairs of arcs abutting P there is a singularity g which connects this pair; for others, there is not. We call the set of all possible connections through P the *connector complex* of P. The connector complexes for the two connector points in Figure 5.2 are shown in Figure 5.3 below.

(e) It seems that connectors are related to interesting global properties of the bifurcation diagrams associated with the moduli family. This point will be discussed in subsection (e).

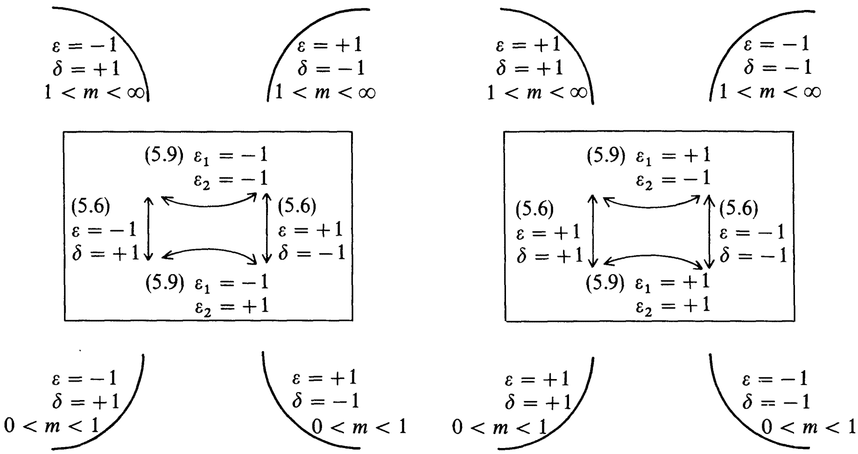


Figure 5.3. The connector complexes at $m = 1$.

(a) Circles in the Moduli Family

Recall that the normal forms (5.1) were obtained from the general cubic by considering certain equivalence transformations. The first step was to put the cubic in the form (2.4) which we repeat here:

$$\psi(x, \lambda) = \varepsilon(x^3 + 3c\lambda^2 x + 2d\lambda^3). \tag{5.2}$$

The next step was to scale d to ± 1 if $d \neq 0$; and if $d = 0$ to scale $3c$ to ± 1. This led to normal form (5.1) with $m = -c/d^{2/3}$. This was convenient for the discussion in the previous sections.

For the present section we rescale (5.2) so that $c^2 + d^2 = 1$. In particular, we let $X(x, \lambda) = ax$ and $S(x, \lambda) = a^{-3}$. Then

$$S(x, \lambda)\psi(X(x, \lambda), \lambda) = \varepsilon\left(x^3 + \frac{3c}{a^2}\lambda^2 x + \frac{2d}{a^3}\lambda^3\right).$$

We leave it to the reader to verify there is a unique positive scalar a which satisfies the equation

$$\left(\frac{c}{a^2}\right)^2 + \left(\frac{d}{a^3}\right)^2 = 1.$$

Assuming that $c^2 + d^2 = 1$, we set $c = -\sin\theta$ and $d = \cos\theta$. This yields the normal form

$$\text{(a)} \quad \psi(x, \lambda, \theta) = \varepsilon(x^3 - 3\sin\theta\,\lambda^2 x + 2\cos\theta\,\lambda^3),$$

where (5.3)

$$\text{(b)} \quad m = \sin\theta/(\cos\theta)^{2/3}.$$

Note that $\delta = +1$ when $\cos\theta > 0$ which is valid for θ in the right half circle, as indicated on Figure 5.1. The values $\theta = \pi/2$ and $3\pi/2$ correspond to $m = -\infty$ and $m = +\infty$ in (5.1a), respectively.

(b) The Distinguished Points

There are two standard ways in which a value of the modal parameter may be distinguished. In the first case, the topological type of the unperturbed diagrams changes at the point in question; in the second case the persistent perturbations undergo change, although the *unperturbed* diagrams are not affected. (*Remark:* At least for our example, the first case occurs at connector points, the second at regular distinguished points.) In earlier sections we have seen both instances of these changes occur. We review our results.

In Lemma 1.1 we showed that all the bifurcation problems h_m with $m > 1$ are topologically equivalent; also, all of the bifurcation problems with $m < 1$ are topologically equivalent. Indeed, the bifurcation diagrams associated with $m > 1$ consist of three straight lines, while the bifurcation diagrams associated with $m < 1$ contain just one single line. In particular, *there is a change in topological type* of h_m at $m = 1$. For this reason alone, the points on the moduli circle where $m = 1$ are distinguished. (As we have indicated above, there are other, more complicated, reasons why $m = 1$ is distinguished. We return to this point below.)

In §4 we showed that the persistent perturbations of h_m when $m < 1$ are different, depending on whether $m < 0$ or $0 < m < 1$. Thus the points on the moduli circle where $m = 0$ are distinguished by a change in the persistent perturbations of h_m.

A similar change in the persistent perturbations occurs when $m = +\infty$ or $m = -\infty$. We describe this change for $m = -\infty$; the case $m = +\infty$ is similar. It may be seen from Figure 5.1 that for either sign of ε, the point $m = -\infty$ separates the two arcs

$$\delta = -1, \qquad m < 0$$

and

$$\delta = +1, \qquad m < 0.$$

Let us show that the set of persistent perturbations of h_m, $m < 0$ and $\delta = +1$ is different from the set with $m < 0$ and $\delta = -1$. (At $m = -\infty$, the set of persistent perturbations of $h_{-\infty}$ is the union of those for h_m, $m < 0$, with either sign of δ.) First observe that the orientation reversing change of coordinates, $\lambda \to -\lambda$, sends h_m with δ of one sign to h_m with δ of the other sign; that is,

$$x^3 - 3mx(-\lambda)^2 + 2\delta(-\lambda)^3 = x^3 - 3mx\lambda^2 + 2(-\delta)\lambda^3.$$

Thus, we can find the persistent perturbations of h_m with $\delta = -1$ by looking at the persistent perturbations of h_m with $\delta = +1$ and reading from right to left. According to Figure 4.1 the bifurcation diagram LR occurs as a perturbation of h_m when $m < 0$ and $\delta = +1$, but RL is not a perturbation of that germ. For $\delta = -1$ the result is reversed, with RL appearing as a perturbation of h_m but not LR. This proves that the two sets of perturbations are different, and hence that $m = -\infty$ is a distinguished point.

(c) Proof That There Are Connector Points at $m = 1$

We now return to our discussion of the points where $m = 1$. The first remark is that these points do not seem to belong on the moduli circle, at least not according to the analysis of h_1 given heretofore. More precisely, the normal form

$$h_1(x, \lambda) = x^3 - 3\lambda^2 x + 2\delta\lambda^3 = (x - \delta\lambda)^2(x + 2\delta\lambda) \tag{5.4}$$

is a degenerate cubic since it has a double root. Indeed, the tangent space to h_1 has *infinite codimension*! (This can be verified directly from (5.4), as in Example II,5.9b.)

The resolution of this difficulty stems from the fact that higher-order terms affect the structure of h_1. Indeed, there are singularities which satisfy

$$j^3 g = h_1,$$

but still have codimension 5; i.e., *the same codimension as h_m for $m \neq 1$*. For example, if we take $\sigma = \pm 1$ and let

$$g = \varepsilon(x^3 - 3x\lambda^2 + 2\delta\lambda^3) + \sigma\lambda^4, \tag{5.5}$$

where $\varepsilon = \pm 1$, then g has codimension 5. (See Exercise 5.1.) Note that the two choices of sign for σ in (5.5) give *inequivalent* bifurcation problems. Thus our original moduli family can be extended in two distinct ways at $m = 1$ to give one-parameter families of codimension five singularities; namely, for $\sigma = \pm 1$ let

$$g_m = \varepsilon(x^3 - 3mx\lambda^2 + 2\delta\lambda^3) + \sigma\lambda^4. \tag{5.6}$$

Recall that we have defined a *connector point* to be a modal parameter value which has at least two inequivalent singularities sharing that value. The two functions in (5.5) show that the points in the modal family where $m = 1$ are connector points. We have indicated this fact in Figure 5.1 by putting boxes around the points where $m = 1$.

However, the situation in the moduli family at points where $m = 1$ is yet more complicated. We continue our discussion of these points in subsections (d) and (e) below.

(d) The Connector Complex at $m = 1$

In this subsection we are concerned with two related questions: (i) Which pairs of arcs of nondistinguished points in the moduli family are connected at the connector points and (ii) which connector points should be identified? For example, if we take $\varepsilon = 1$, $\delta = -1$ in (5.6), then g_m provides a connection of the arc $\varepsilon = +1$, $\delta = -1$, $0 < m < 1$ with the arc $\varepsilon = +1$, $\delta = -1$, $1 < m < \infty$ through the singularity g of (5.5). Our first task in this section is to show that the arc $\varepsilon = -1$, $\delta = +1$, $1 < m < \infty$ may also be connected with the arc $\varepsilon = +1$, $\delta = -1$, $1 < m < \infty$ using a different codimension five singularity. Then we address the more systematic questions above.

Our discussion of the nondegenerate cubic

$$\phi(x, \lambda) = Ax^3 + Bx^2\lambda + Cx\lambda^2 + D\lambda^3$$

began with the hypothesis $A \neq 0$. We now describe the simplest (i.e., lowest codimension) singularity where $A = 0$, as this provides the connection we are seeking. We assume that $B \neq 0$ and that the resulting quadratic $\phi(x, 1)$ has distinct roots. We claim that under these hypotheses, ϕ is strongly equivalent to

$$\psi(x, \lambda) = \varepsilon_1 x^2 \lambda + \varepsilon_2 \lambda^3, \tag{5.7}$$

where $\varepsilon_1 = \pm 1$, $\varepsilon_2 = \pm 1$. To verify this claim we make a linear change of coordinates in x so that C' will vanish, where C' is the coefficient of $x\lambda^2$ after the transformation. Then D', the new coefficient of λ^3, must be nonzero; otherwise, $x = 0$ would be a double root, contradicting our hypothesis above. Now scale to arrive at (5.7).

The homogeneous cubic ψ in (5.7) has infinite codimension. Thus, we must consider higher-order terms in order to define a singularity of finite codimension.

In Exercise 5.2 we ask the reader to verify that

$$k(x, \lambda) = \varepsilon_3 x^4 + \varepsilon_1 x^2 \lambda + \varepsilon_2 \lambda^3, \tag{5.8}$$

where $\varepsilon_3 = \pm 1$, has codimension 5 and that k has modality one. Moreover, we claim that this singularity belongs in the modal family at $m = 1$. In order to verify this claim we consider the one-parameter unfolding

$$k_a(x, \lambda) = \varepsilon_3 x^4 + a x^3 + \varepsilon_1 x^2 \lambda + \varepsilon_2 \lambda^3 \tag{5.9}$$

when $a \neq 0$. The singularity of k_a at the origin is a nondegenerate cubic of the form we studied in Proposition 2.2. Thus k_a is equivalent to h_m in (5.1) for some values of the parameters. We ask the reader to substitute into (2.5) and (2.9) to verify that

$$m = \tfrac{1}{9}\left(\frac{2}{\dfrac{2\varepsilon_1}{27} + \varepsilon_2 a^2}\right)^{2/3}, \qquad \delta = \varepsilon_1 \operatorname{sgn}(a), \qquad \varepsilon = \operatorname{sgn}(a). \tag{5.10}$$

From (5.10) we see that $\lim_{a \to 0} m = 1$. Thus the singularity k should be inserted into the moduli family at $m = 1$.

We now use example (5.10) to argue that the two points $(m, \varepsilon, \delta) = (1, -1, +1)$ and $(m, \varepsilon, \delta) = (1, +1, -1)$ should be identified. Consider the unfolding (5.9) in the $\varepsilon_1 = -1$ and $\varepsilon_2 = +1$. From (5.10), we see that $m > 1$ when a is nonzero (but small). In addition, we see that if $a > 0$ then $(\varepsilon, \delta) = (-1, +1)$, and if $a < 0$ then $(\delta, \varepsilon) = (+1, -1)$. Thus the family

$$k(x, \lambda, a) = \varepsilon_3 x^4 + a x^3 + x^2 \lambda + \lambda^3 \tag{5.11}$$

provides a connection between modal parameter values on the circle $\varepsilon = +1$ with those on the circle $\varepsilon = -1$. The only way this connection could be made continuously (and we expect continuity, since a varies continuously) is to identify the two distinct points listed above. Similar considerations suggest identifying $(m, \varepsilon, \delta) = (1, +1, +1)$ and $(m, \varepsilon, \delta) = (1, -1, -1)$.

In the moduli bracelet, Figure 5.2, there are after identification, two connector points. In Figure 5.3 we present the connector complexes of each of these points. Observe that there are four branches of the moduli family which abut each connector point. The singularities (5.5) and (5.8), or more properly their one-parameter unfoldings (5.6) and (5.9), generate four connections in each connector complex. These connections and the singularities which generate them are shown in Figure 5.3. Note that each connection is, in fact, generated by two singularities; the choice of σ in (5.5) and ε_3 in (5.8) does not affect which branches are connected.

A complete description of this codimension 5 unimodal moduli family may be found in Keyfitz [1984]. The situation is still more complicated than we have indicated here. (See Exercise 5.2.)

(e) Perturbations Near $m = 1$: Global Considerations

As we have stated before, we look for organizing centers in order to analyze global behavior using local techniques. This kind of analysis must be applied with some care. For example suppose we encounter a bifurcation problem of the form

$$0.001x^2 - \lambda x + p(x, \lambda) = 0, \tag{5.12}$$

where $p \in \mathcal{M}^3$. According to our solution of the recognition problem, in some neighborhood of the origin, (5.12) is equivalent to the normal form $x^2 - \lambda x$. In this neighborhood, (5.12) exhibits transcritical bifurcation. However, the point of this example is that in (5.12) the coefficient of x^2 is extremely small. Thus if the higher-order terms in (5.12) contain a term ax^3 with even a moderately large coefficient a, the neighborhood on which this analysis is valid will be extremely small. In such a situation it makes more sense to view (5.12) as a perturbation of the pitchfork $x^3 - \lambda x$. This point of view leads to a modification of the transcritical behavior of (5.12) as sketched in Figure 5.4; in particular, this point of view suggests that the bifurcation diagram of (5.12) may have a limit point close to the origin.

The present subsection is concerned with near degeneracies analogous to (5.12) but involving a modal parameter. To make this more definite, suppose we are studying a bifurcation problem with a cubic singularity such that the analysis of §2 leads to the normal form (5.1b) with $\varepsilon = \delta = +1$ and $m = 1.001$. We know that the associated bifurcation diagram consists of three

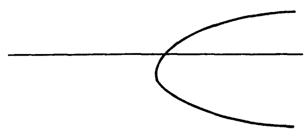


Figure 5.4. A "barely transcritical" bifurcation.

nonsingular curves intersecting at the origin. However, because the modal parameter is so close to the distinguished point $m = 1$, we also know that this analysis is valid only on a small neighborhood of the origin in the $x\lambda$-plane.

As we illustrated in subsections (c) and (d), there are two types of degeneracies at $m = 1$ connected to $\varepsilon = \delta = +1$ and $m > 1$. First there is (5.5), which corresponds to the cubic terms

$$x^3 - 3x\lambda^2 + 2\lambda^3; \tag{5.13a}$$

and second there is (5.8), which corresponds to the cubic terms

$$x^2\lambda - \lambda^3. \tag{5.13b}$$

(Cf. Figure 5.3.)

How does this apply to our hypothetical bifurcation problem above where $m = 1.001$? If in that problem the coefficient of x^3 is large (in absolute value), then we would guess that the original bifurcation problem is close to (5.13a). If, on the other hand, the coefficient of x^3 is approximately 0, then we would guess that the original bifurcation problem is close to (5.13b). Unfortunately, it is not easy to decide which case applies in a strict singularity theory context. The problem is that the notion of whether a coefficient is large or small is *not invariant* under scalings of x and λ.

There are several possible approaches in such a situation. One approach would be to measure the size of the coefficient of x^3 relative to a distinguished scaling of x and λ; for example, physical considerations in a given problem might indicate a natural nondimensionalization of these variables. This approach suggests further investigation on the connection between singularity theory and applied mathematics, but so far there is little specific information available. A second approach would be to introduce an additional parameter into the problem and to attempt to vary this parameter so that the modal parameter equals one exactly; the singularity which occurs when $m = 1$ would presumably be equivalent to either (5.13a) or (5.13b), and this would indicate which singularity was relevant. A third approach relates to the global behavior of the bifurcation diagram; we discuss the third approach more fully.

Let us consider what information the choice between (5.13a) and (5.13b) provides about the global character of the bifurcation diagrams. In Figure 5.5 we present the bifurcation diagrams associated with the unfoldings

$$\begin{aligned} &\text{(a)}\quad G(x, \lambda, a) = x^3 - 3(1 + a)x\lambda^2 + 2\lambda^3 + \sigma\lambda^4, \\ &\text{(b)}\quad k(x, \lambda, a) = x^2\lambda - \lambda^3 + \varepsilon_3 x^4 + ax^3, \end{aligned} \tag{5.14}$$

where $a \approx 0$, $\varepsilon_3 = \pm 1$ and $\sigma = \pm 1$. (Cf. (5.6) and (5.9).) The global features we find in (5.14a) are the bow which occurs for $a > 0$ and the nearby solution branch which occurs for $a < 0$. We call (5.14a) the *bowtie*

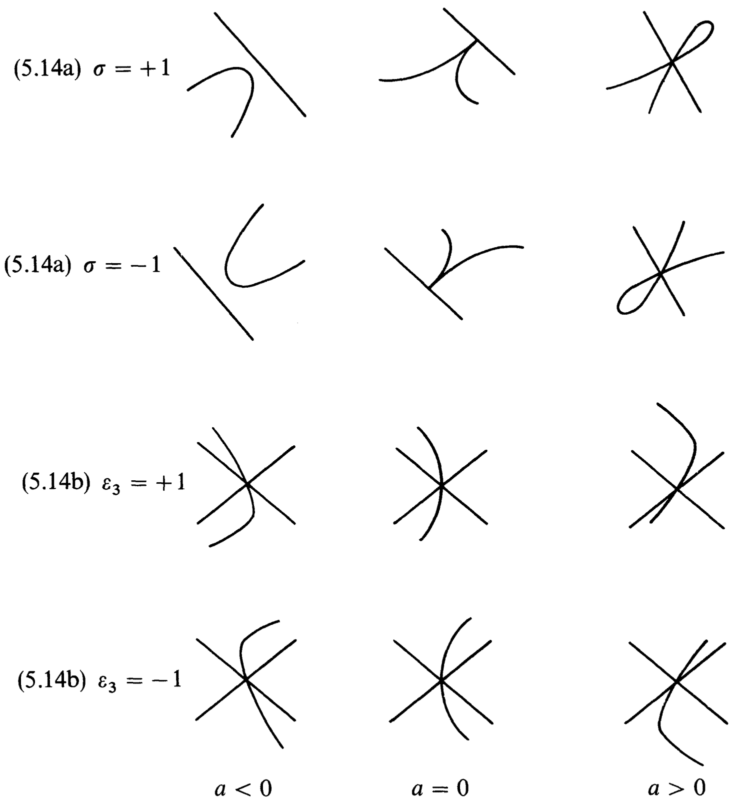


Figure 5.5. Global information near a connector complex.

bifurcation. The global feature we find in (5.14b) is the turn around of one of the three intersecting branches. Observe the different character of the bifurcation diagrams of (5.14a) and (5.14b). In analyzing a given physical problem, one could use information about its global behavior to make conjectures about whether (5.13a) or (5.13b) is the governing singularity.

Remark. Part of the difficulty of the present example stems from the fact that its codimension is so large. In §§5–8 of Chapter VI we consider a singularity which exhibits similar behavior but has codimension three. In this case it is possible to see (literally) the changes in global behavior as the unfolding parameters vary.

Exercises

5.1. Show that (5.5), namely, $g = \varepsilon(x^3 - 3x\lambda^2 + 2\delta\lambda^3) + \sigma\lambda^4$ has codimension 5. Find a universal unfolding of g which includes m as in (5.6) as one of the unfolding parameters.

5.2. (a) Show that $k(x, \lambda) = x^4 + x^2\lambda + \lambda^3$ has codimension 5 and that

$$K = x^4 + x^2\lambda + \lambda^3 + \alpha + \beta x + \gamma\lambda + \rho x^3 + \tau\lambda^2 \tag{5.15}$$

is a universal unfolding of k.

(b) Show that the codimension constant variety $\mathscr{C}$ for (5.15) consists of the two lines in the plane $\alpha = \beta = \gamma = 0$ defined by the equations $\rho = 0$ and $\tau = 0$. It follows that $\mathscr{C}$ is *not* a submanifold of $\mathbb{R}^5$ for this example. (See the classification theorem for singularities of codimension ≤ 7 in Keyfitz [1984] for more information.)

§6. Discussion of Moduli and Topological Codimension

In this section we summarize the lessons we wish to draw concerning moduli and topological codimension. The section is divided into four subsections which address the following points:

(a) The modification of the definition of topological codimension to include topological equivalence of the perturbed bifurcation diagrams.
(b) The relation of the present example to the thesis of Chapter IV, §1.
(c) The observation that moduli families typically reduce under topological equivalence to a *finite* number of inequivalent bifurcation problems.
(d) A formal justification of the word "codimension" in the term topological codimension.

(a) Qualifications in the Definition of Topological Codimension

In Chapter IV, §1, we defined the topological codimension of a germ to be equal to its C^∞-codimension less the modality. In symbols, if g is in $\mathscr{E}_{x,\lambda}$, then

$$\text{top–codim } g = (C^\infty\text{–codim } g) - \text{modality } (g). \tag{6.1}$$

However, there is an important restriction we place on g in order for this definition to apply: We require that the universal unfolding G of g be topologically trivial. This restriction has the important consequence that the structure of the set of persistent perturbations in G does not change as the modal parameters are varied. Let us define the term "topologically trivial".

Let g be a germ in $\mathscr{E}_{x,\lambda}$ of codimension k, and let $G(x, \lambda, \alpha)$ be a universal unfolding of g. Let $\mathscr{C}$ be the codimension constant variety of G (cf.

Definition 3.2). We assume that $\mathscr{C}$ has dimension $k - l$, and we choose G so that $\mathscr{C}$ has the form

$$\mathscr{C} = \{\alpha \in \mathbb{R}^k : \alpha_i = 0, i = 1, \ldots, l\}.$$

Let us write $\alpha = (\beta, m)$ where $\beta \in \mathbb{R}^l$ and $m \in \mathbb{R}^{k-l}$; this indicates the modal parameters explicitly. Let Σ be the transition variety of G in $\mathbb{R}^k$, and let

$$\Sigma_0 = \{\beta \in \mathbb{R}^l : (\beta, 0) \in \Sigma\}.$$

In words, Σ_0 is the intersection of Σ with the l-dimensional subspace of nonmodal parameters.

Definition 6.1. The universal unfolding G is *topologically trivial* if Σ is (locally) homeomorphic to $\mathscr{C} \times \Sigma_0$.

It follows from Theorem III, 10.1 that if G is topologically trivial, then for every (small) m the persistent perturbations of $G(x, \lambda, 0, m)$ are identical to those of $g(x, \lambda) = G(x, \lambda, 0, 0)$. To see this, for any m let $\Sigma_m \subset \mathbb{R}^l$ be the transition set of the l-parameter unfolding $G(\cdot, \cdot, \beta, m)$ of $G(\cdot, \cdot, 0, m)$. The connected components of $\mathbb{R}^k \sim \Sigma_0$ and $\mathbb{R}^l \sim \Sigma_m$ are in one-to-one correspondence with one another.

We have already seen that for our example (5.1), the universal unfolding is not topologically trivial if $m = 0, 1, \pm\infty$. As we mentioned in §5, the arguments in Stewart [1981] show that the universal unfoldings are topologically trivial when $m \neq 0, 1, \pm\infty$. Thus, we may say that h_m has topological codimension 4 provided $m \neq 0, 1, \pm\infty$. We do not attempt to define topological codimension at the exceptional values.

(b) Relation with Chapter IV, §1

Let us relate these concepts to the thesis of Chapter IV, §1; specifically, to the issue of whether to use C^∞-codimension or topological codimension in (IV, 1.2). Apart from the exceptional cases $m = 0, 1, \pm\infty$, nondegenerate cubics have C^∞-codimension 5 and topological codimension 4. Let us use these cubics to construct an example of a singularity with C^∞-codimension 5 in a four-parameter family F of bifurcation problems such that all small perturbations of F also contain a singularity of C^∞-codimension 5. This example illustrates that one should use topological codimension in (IV, 1.2).

To construct our example, we fix a value m_0 of the modal parameter, $m_0 \neq 1$, and we define $F: \mathbb{R} \times \mathbb{R} \times \mathbb{R}^4 \to \mathbb{R}$ by

$$F(x, \lambda, \beta_1, \beta_2, \beta_3, \beta_4) = H(x, \lambda, \beta_1, \beta_2, \beta_3, \beta_4, m_0),$$

where H is the universal unfolding (3.1). Consider an arbitrary perturbation of F, say $F + \varepsilon p$. By the universal unfolding theorem $F + \varepsilon p$ may be factored through H; i.e., there exists a mapping $(A(\beta, \varepsilon), N(\beta, \varepsilon))$ from $\mathbb{R}^5$ into $\mathbb{R}^5$ such that

$$F(\cdot, \cdot, \beta) + \varepsilon p \sim H(\cdot, \cdot, A(\beta, \varepsilon), N(\beta, \varepsilon)).$$

When $\varepsilon = 0$

$$\frac{\partial A_i}{\partial B_j}(0, 0) = \delta_{ij}; \qquad i, j = 1, 2, 3, 4.$$

For small ε, the Jacobian $\partial A_i/\partial \beta_j$ is still nonsingular. Thus by the implicit function theorem, for all small ε there is a (unique) value of β such that

$$A_i(\beta, \varepsilon) = 0, \qquad i = 1, 2, 3, 4.$$

In other words, for this value of β,

$$F(\cdot, \cdot, \beta) + \varepsilon p$$

is equivalent to $G(\cdot, \cdot, 0, N(\beta, \varepsilon))$; the latter is a nondegenerate cubic and therefore has C^∞-codimension 5. This example shows how a cubic singularity of C^∞-codimension 5 can occur stably in a four-parameter family of bifurcation problems. In general terms, the reason for this is the following: A cubic singularity of C^∞-codimension 5 in an unfolding F occurs stably provided F intersects the codimension constant variety $\mathscr{C}$ transversely; since $\mathscr{C}$ is a one-dimensional manifold, only four parameters are needed for such a transverse intersection.

Moreover, as regards applications it seems that topological codimension is generally more appropriate than C^∞-codimension. Consider a k-parameter family of bifurcation problems, $F(x, \lambda, \alpha)$, such that for some α_0, $F(\cdot, \cdot, \alpha_0)$ is equivalent to (2.8); in symbols

$$F(\cdot, \cdot, \alpha_0) \sim h_m. \tag{6.2}$$

If one wants (6.2) to hold for a precise value of m *specified in advance*, then F must depend on at least five parameters in order to avoid the erratic behavior of over-determined systems. However, if one merely wants (6.2) to hold for *some* value of m, the exact value not being important, then four parameters suffice to avoid such erratic behavior. Even if one insists that m lie in some *range* such as $1 < m < \infty$, four parameters are still sufficient. In this chapter we showed that the general properties of the bifurcation diagram of h_m and its perturbations only depend on whether $m > 1$, $0 < m < 1$, or $m < 0$. These considerations also suggest using topological codimension in (IV, 1.2).

(c) An Important Observation Concerning Moduli

The picture of moduli we developed above is typical in that the moduli space divides into finitely many regions on which the unfolding is topologically trivial. This is an important observation. We began our discussion of moduli in §1 by showing that there exist continuous families of (C^∞) inequivalent germs, and we end our discussion (of the particular example) by showing that there are just a finite number (in our example, six) of regions on which all germs are (C^0) equivalent. So, in this sense, a moduli family gives rise to a *finite number* of distinct bifurcation problems.

(d) On the Interpretation of Topological Codimension

In Corollary II,2.6 we showed that the C^∞-codimension of a germ g could be interpreted as the codimension of its orbit under equivalence. Let us discuss a similar interpretation for topological codimension. Let us recall the C^∞-situation more carefully. Let S be the space of singularities in $\mathscr{E}_{x,\lambda}$; that is,

$$S = \{g \in \mathscr{E}_{x,\lambda}: g = g_x = 0\} = \mathscr{M}^2 + \langle \lambda \rangle.$$

Let g be a germ in S and let

$$\mathcal{O}_g = \{h \in \mathscr{E}_{x,\lambda}: h \text{ is equivalent to } g\}.$$

We showed in Corollary II,2.6 that the codimension of $\mathcal{O}_g$ in S is equal to both the C^∞-codimension of g and the number of defining conditions for g less two.

Now suppose $g \in S$ is a germ of finite codimension with positive modality. Let $\mathscr{C}$ be the codimension constant variety. Define

$$\mathcal{O}_{\mathscr{C}} = \{h \in \mathscr{E}_{x,\lambda}: h \text{ is equivalent to some germ in } \mathscr{C}\}. \qquad (6.3)$$

Since the C^∞-codimension of germs in the codimension constant variety is (by definition) constant we see that locally

$$\mathcal{O}_{\mathscr{C}} \cong \mathcal{O}_g \times \mathscr{C}.$$

Thus the codimension of $\mathcal{O}_{\mathscr{C}}$ in S is equal to the codimension of $\mathcal{O}_g$ in S minus the modality of g. Hence the right-hand side of (6.1) is the codimension of $\mathcal{O}_{\mathscr{C}}$ in S.

§7. The Thermal-Chainbranching Model

The explosion peninsula is an interesting and perhaps even surprising phenomenon common to many oxidation reactions in chemical combustion theory. We discuss this phenomenon in part to demonstrate that there are reasonable mathematical models with many parameters whose analysis can be assisted by the use of singularity theory techniques. Moreover, in this particular example the moduli family discussed in this chapter appears, thus showing that even such seemingly esoteric examples may actually occur in applications.

For ease of exposition we describe the explosion peninsula in terms of the oxidation of hydrogen to make water. This reaction proceeds by the overall reaction

$$2H_2 + O_2 \longrightarrow 2H_2O, \tag{7.1}$$

although it is well accepted that a complete decomposition of (7.1) into elementary reactions would require over twenty reactions involving at least seven intermediate radicals.

The main experimental fact of relevance here is that the reaction (7.1) can proceed either at a slow speed or explosively fast, depending on initial conditions. Consider a mixture of hydrogen and oxygen in a closed container, initially at pressure P_0 and temperature T_0 and immersed in a bath of the same temperature. Experimental data concerning whether the reaction to produce water proceeds at a slow speed or explosively fast are summarized in Figure 7.1; let us interpret the figure. If the initial condition (T_0, P_0) lies to the left of the curve in Figure 7.1 then the reaction proceeds at a slow speed, while if it lies to the right an explosion occurs. The noteworthy feature in this figure is the region of slow reactions between the curves labeled "second limit" and "third limit." To see what is interesting here, suppose that T_0 is fixed between $\underline{T}_0$ and $\overline{T}_0$ and consider a sequence of experiments beginning with (P_0, T_0) in this region. Increasing the pressure P_0 is equivalent to increasing the fuel for the reaction, and it is not surprising that there should be a critical pressure at which the reaction proceeds explosively fast. However, it is indeed surprising, at least at first glance, that an explosive initial condition should be reached by *decreasing* P_0, as this decreases the available fuel. Yet this is what happens.

The properties of the explosion peninsula have been much studied by chemists, for the following reason. If (7.1) were modeled by a single reaction obeying Arrhenius kinetics, there would be no explosion peninsula—the separating curve in Figure 7.1 would be monotonic in T_0. Thus information about the explosion peninsula yields information about the individual elementary reactions in (7.1).

Before introducing a specific model for (7.1), let us try to imagine bifurcation diagrams associated to this experiment. We assume that a

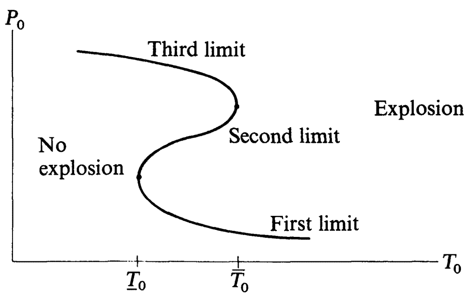


Figure 7.1. The explosion peninsula.

steady-state approximation is sufficiently accurate to describe the experiments. The state variable in the bifurcation diagrams is the rate at which the reaction proceeds. We regard P_0 as the bifurcation parameter and T_0 as an auxiliary parameter. Suppose we fix T_0 between $\underline{T}_0$ and $\overline{T}_0$. In order to relate the bifurcation diagram to experiment, there must be a sequence of low-temperature (i.e., slow reaction) and high-temperature states, depending on P_0, such as illustrated in Figure 7.2(a). In Figure 7.2(b) we have drawn possible bifurcation diagrams which yield the behavior required for Figure 7.2(a). (*Remark:* There is experimental evidence that the transition from a low-temperature equilibrium to a high-temperature one at the first limit is given by an "*S*"-shaped curve. No such information exists near the other limits. Thus the simplest possibilities are those of Figure 7.2(b).)

Note that these bifurcation diagrams occur in the list of Figure 4.1. (Mirror images should be included in Figure 4.1, to include both cases $\delta = \pm 1$ in (2.8).) It is therefore natural to conjecture (2.8) as a possible organizing center in a singularity theory analysis of this problem.

Golubitsky *et al.* [1981] verified this conjecture, in a sense that we now describe. To verify such a conjecture it is necessary to introduce a specific mathematical model for the physical problem under study. In that paper we chose the thermal-chain branching model of Gray and Yang [1965, 1967], as this is far more tractible than the full set of reactions but still yields an explosion peninsula. In this model, after nondimensionalization steady states may be described by an equation

$$G(x, \lambda, \alpha) = 0, \tag{7.2}$$

where $G: \mathbb{R} \times \mathbb{R} \times \mathbb{R}^9 \to \mathbb{R}$. Here the 9 parameters represent reaction rates, activation energies, etc. in the various reactions. There is disagreement in the literature concerning the exact values for these parameters, and we did not attempt to fit the experimental data. Rather we asked whether for any values of the parameters, (7.2) exhibits a singularity equivalent to (2.8). We

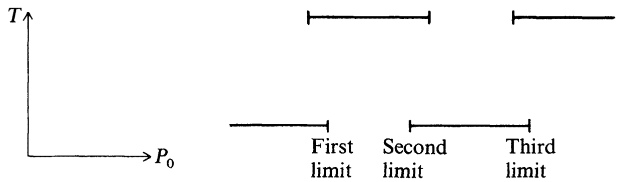


Figure 7.2(a). Necessary sequence of steady states in the explosion peninsula.

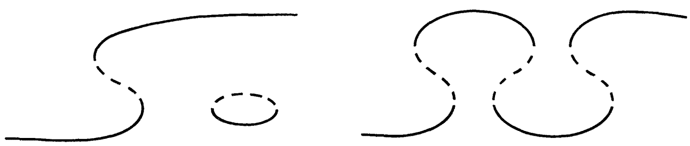


Figure 7.2(b). Possible bifurcation diagrams to effect the sequence in Figure 7.2(a).

found that such singularities did indeed occur and that the physical parameters provided a universal unfolding of the singularity. Interestingly, both cases in Figure 4.1 (i.e., $m < 0$ and $0 < m < 1$) could occur, depending on the exact values of the parameters. We refer the reader to Golubitsky *et al.* for more detail.

Bibliographical Comments

The investigation of moduli is currently an active topic of research in both algebraic geometry and singularity theory. The first instance of moduli in algebraic geometry is the cross ratio (or, more generally, the J-invariant). Cf. Mumford [1965].

In singularity theory, moduli appeared in two separate ways, both involving classifications. We discuss these in sequence. A smooth mapping $g: \mathbb{R}^n \to \mathbb{R}^m$ is called *stable* if every mapping near g is equivalent to g. (Roughly speaking, stable is the global version of "codimension zero".) Mather proved that for a range of dimensions (n, m), called the *nice* dimensions, the stable mappings form a dense subset of the space of all smooth mappings. (Much of Morse theory and Whitney embedding theory follow from the fact that $(n, 1)$ and $(n, 2n + 1)$ are nice dimensions. Cf. Golubitsky and Guillemin [1973].) The obstruction to a pair of dimensions (n, m) being nice is the existence of moduli in low codimension. See Mather [1971].

The second situation where moduli appear in singularity theory is in Thom's classification of elementary catastrophes. Up to codimension six there is a *finite* list of singularities of potential functions, but this finite enumeration breaks down in codimension seven, where moduli first exist. (This breakdown led to some of the mystical numerology that has enshrouded catastrophe theory.) Arnold has suggested counting moduli families as one entity since this makes the list finite once again. See Arnold [1976]. This point of view, which we have adopted, is now generally accepted. However, there is some difficulty concerning the precise definition of a moduli family. We have used "codimension constant" in our definition; but there are other possibilities. Cf. Wall [1983].

CHAPTER VI
Bifurcation with $\mathbf{Z}_2$-Symmetry

§0. Introduction

If $g \in \mathscr{E}_{x,\lambda}$, we say that g has $\mathbf{Z}_2$-*symmetry* if g is an odd function of x; in symbols, if

$$g(-x, \lambda) = -g(x, \lambda). \tag{0.1}$$

We use this terminology because we think of a two-element group $\mathbf{Z}_2 = \{I, R\}$ acting on the real line, where I is the identity and $Rx = -x$; equation (0.1) asserts that g commutes with the action of this group. In this chapter we study bifurcation problems with $\mathbf{Z}_2$-symmetry. Bifurcation problems with this symmetry arise often in applications. For example, the buckling model of Chapter I, §1 was $\mathbf{Z}_2$-symmetric; in that case the physical representation of the symmetry was reflection across the horizontal axis. Moreover, bifurcation problems of the form (0.1) play a central role in our treatment of the Hopf bifurcation in Chapter VIII.

The importance of symmetry already appeared in Chapter IV, §1. There we presented the thesis that we should be cautious with a mathematical model that has a singularity whose codimension is greater than the number of parameters in that model. (Cf. (IV, 1.2).) However, we also said that we must interpret codimension in (IV, 1.2) within the class of functions that possess the symmetry appropriate to the given problem. Our main reason for studying $\mathbf{Z}_2$-symmetric bifurcation problems here is to illustrate the profound effect that symmetry has on codimension, even so simple a symmetry group as $\mathbf{Z}_2$. However, the theory is interesting in its own right. Moreover, symmetry will play a fundamental role in Volume II, and the present chapter foreshadows the directions we will take there.

This chapter is divided into eight sections. In §1, we present a simple physical example which illustrates how symmetry facilitates the occurrence of bifurcation problems which would have high codimension in the absence of symmetry. Sections 2–5 are a unit which develops the general theory of Chapters II–IV for bifurcation problems with $\mathbf{Z}_2$-symmetry; specifically, we consider the recognition problem in §2, universal unfoldings in §3, the theoretical basis for enumerating perturbed bifurcation diagrams in §4, and a classification theorem by $\mathbf{Z}_2$-codimension in §5. (In other words, §2 corresponds to Chapter II, §§3 and 4, to Chapter III, and §5 to Chapter IV.) In §6 we apply the theory to enumerate the persistent perturbed bifurcation diagrams of all $\mathbf{Z}_2$-symmetric bifurcation problems of $\mathbf{Z}_2$-codimension three or less which do not have modal parameters. In §§7 and 8 we consider the one family of bifurcation problems of $\mathbf{Z}_2$-codimension three or less which does have a modal parameter. Section 7 includes the persistent perturbations of this family near the nonconnector modal parameter values. Persistent perturbations of the connector points are given in §8.

It will appear in §§2–5 that, almost without exception, there is a one-to-one correspondence between the theorems of Chapters II–IV and theorems in the present $\mathbf{Z}_2$-symmetric context. (For clarity we shall sometimes use the term "nonsymmetric context" to describe bifurcation problems without symmetry.) Nevertheless, differences appear at unexpected places. For example, there are new sources of nonpersistence in the $\mathbf{Z}_2$-symmetric case. Even for general compact groups, most of the theoretical results of Chapters II–IV extend to the group context, but new calculations are still required for each new group. (Indeed, this is one reason why it is difficult to apply results from the singularity theory approach to bifurcation problems without actually understanding the theory.)

One subtle difference between the symmetric and nonsymmetric context deserves particular emphasis. In the nonsymmetric context we localized germs around $x = 0$, but this was purely a matter of convenience—any point would do equally well. In the symmetric context, we again localize around $x = 0$, but now this is extremely important—considering any other point would lead to very different results. The reason is that $x = 0$ is a fixed point of the group under study, the only fixed point. In more physical terms, working near $x = 0$ means that we are studying the bifurcation of solutions which break the $\mathbf{Z}_2$-symmetry from a trivial solution that is $\mathbf{Z}_2$-symmetric.

§1. A Simple Physical Example with $\mathbf{Z}_2$-Symmetry

In this section we discuss a simple mechanical system with $\mathbf{Z}_2$-symmetry which was first brought to our attention in Poston and Stewart [1979]. This example demonstrates clearly how the existence of a single reflectional symmetry can cause problems of surprisingly high codimension to appear.

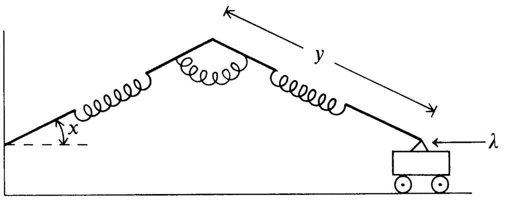


Figure 1.1. Finite-element analogue for buckling strut with compressible links.

We modify the finite-element analogue of a buckling beam considered in Chapter 1, §1 by allowing both connecting links to be compressible. (See Figure 1.1.) Specifically, suppose that the connecting links are linear springs with equal spring constants k and uncompressed length unity. For simplicity, let us assume that there is a supporting frame (not shown in the figure) that forces the two springs to have equal compression or extension; this does not change the basic conclusion, but it does simplify the analysis by eliminating an inessential degree of freedom. As in Chapter I, §1, we assume the torsional spring connecting the two links has unit strength.

We choose as coordinates x, the angle the links make with the horizontal, and y, the common lengths of the springs. Observe that there is a reflectional symmetry in the system given by $x \to -x$. More precisely, the potential energy V, given by

$$V = k(y-1)^2 + x^2/2 + 2\lambda y \cos x,$$

is even in x. On differentiating V, we obtain the equations for equilibrium:

$$\frac{\partial V}{\partial x} = x - 2\lambda y \sin x = 0,$$

$$\frac{\partial V}{\partial y} = 2k(y-1) + 2\lambda \cos x = 0.$$

If the second equation is used to eliminate y from this system, we are left with the single equation

$$g(x, \lambda) = 2\lambda\left(1 - \frac{\lambda}{k}\cos x\right)\sin x - x = 0. \tag{1.1}$$

Observe that g is an odd function in x, a result of the reflectional symmetry. In particular,

$$\left(\frac{\partial}{\partial x}\right)^j\left(\frac{\partial}{\partial \lambda}\right) g(0, \lambda) = 0 \quad \text{if } j \text{ is even.} \tag{1.2}$$

We are interested in the bifurcation from the trivial solution $x = 0$ that may occur in (1.1). Let us expand (1.1) in a Taylor series as follows:

$$g(x, \lambda) = \left(-1 + 2\lambda - \frac{2\lambda^2}{k}\right)x + \left(\frac{4\lambda^2}{3k} - \frac{\lambda}{3}\right)x^3 + \left(\frac{\lambda}{60} - \frac{4\lambda^2}{15k}\right)x^5 + O(x^7). \tag{1.3}$$

Now bifurcation occurs only if

$$g_x(0, \lambda) = -1 + 2\lambda - \frac{2\lambda^2}{k} = 0. \tag{1.4}$$

Solving (1.4) for λ we find

$$\lambda_c^{\pm}(k) = \tfrac{1}{2}\{k \pm \sqrt{k^2 - 2}\}. \tag{1.5}$$

Thus for $k > 2$, there are two distinct bifurcations but for $k < 2$, none.

What happens when $k = 2$. Note that $\lambda_c^+(2) = \lambda_c^-(2) = 1$. On inspection of (1.3) we find that near $x = 0$, $\lambda = 1$

$$g(x, \lambda) = \tfrac{1}{3}x^3 - (\lambda - 1)^2 x + O(x^5). \tag{1.6}$$

It follows from Proposition V,2.2 that g is equivalent to the normal form (V,2.7), $\varepsilon x^3 + \delta\lambda^2 x$, with $\varepsilon = +1$, $\delta = -1$. In other words, g exhibits a nondegenerate cubic singularity in this case; i.e., a singularity of codimension five (neglecting symmetry).

Now suppose $k > 2$. Then typically the two bifurcations in (1.5) are pitchforks. To see this we apply Proposition II,9.2 which solves the recognition problem for the pitchfork. Note that $g = g_\lambda = g_{xx} = 0$ by (1.2) and $g_x = 0$ if $\lambda = \lambda_c^{\pm}(k)$. Now if $k > 2$, then $g_{\lambda x}$ is nonzero at the bifurcation point. For g_{xxx}, we have

$$g_{xxx}(0, \lambda) = 6\left(\frac{4\lambda^2}{3k} - \frac{\lambda}{3}\right); \tag{1.7}$$

this is always nonzero at $\lambda = \lambda_c^+(k)$ and is nonzero at $\lambda_c^-(k)$ unless $k = \frac{8}{3}$. Thus (1.1) has a pitchfork bifurcation at $x = 0$, $\lambda = \lambda_c^{\pm}(k)$, with the one exception.

Let us consider the case $k = \frac{8}{3}$. Of course, $(\partial/\partial x)^4 g$ vanishes at the bifurcation point, by (1.2). We compute from (1.3) that $(\partial/\partial x)^5 g < 0$ at the bifurcation point. By Proposition II,9.2, g is equivalent to $-x^5 + \lambda x$ at $\lambda_c^-(\frac{8}{3}) = \frac{2}{3}$. This singularity has codimension four (neglecting symmetry).

The full bifurcation diagrams for (1.1) are shown in Figure 1.2.

In this model there is only one auxiliary parameter; viz., the spring constant k. However, when $k = 2$ or $\frac{8}{3}$ the model exhibits singularities of codimension five and four, respectively. Thus, without the symmetry this model would appear to present a clear violation of the thesis of Chapter IV, §1. However, we show in the present chapter that as $\mathbf{Z}_2$-symmetric bifurcation problems, both the above singularities have codimension one. In other words, there is no violation of the thesis of Chapter IV, §1 provided one uses the right notion of codimension in (IV,1.2).

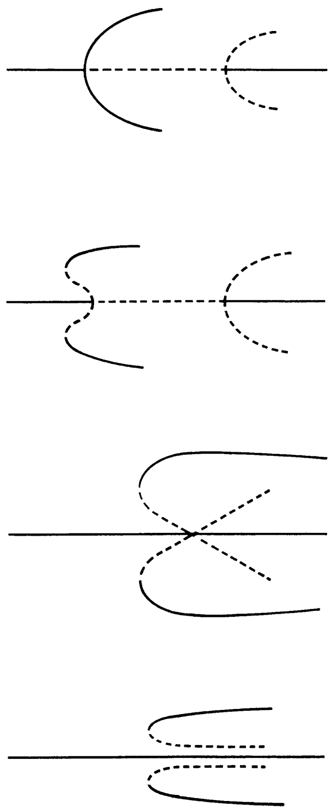


Figure 1.2. Bifurcation diagrams associated to a buckling strut with compressible links.

§2. The Recognition Problem

In this section we develop methods to solve the recognition problem for bifurcation problems with $\mathbf{Z}_2$-symmetry. As in Chapter II, we focus primarily on strong equivalence, as it is mathematically convenient to do so.

The section is divided into five subsections. In the first subsection, we address an algebraic difficulty that arises in extending our mathematical techniques to the symmetric case. The remaining four subsections follow Chapter II fairly closely—in subsections (b), (c), (d), and (e), we define the restricted tangent space in the symmetric context, study the appropriate notion of "intrinsic," state the main results, and work out two classes of examples, respectively. (Cf. Chapter II, §§2, 7, 8, 9.) We are fairly brief with proofs that are similar to their counterparts from Chapter II.

(a) Resolution of an Algebraic Difficulty

Let $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ be the set of all germs in $\mathscr{E}_{x,\lambda}$ that are odd in x; in symbols

$$\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2) = \{g \in \mathscr{E}_{x,\lambda}: g(-x, \lambda) = -g(x, \lambda)\}. \tag{2.1}$$

(*Remark:* The arrow anticipates notation from Volume II.) The following difficulty hampers the generalization of our mathematical techniques to the present context: *The set $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ is not closed under multiplication.* Indeed the product of two *odd* functions is *even*. This is potentially quite a serious problem, in that our theory relies heavily on algebraic operations for a compact description of $RT(g)$ and for efficient computation with $RT(g)$. However, the product of any element of $\mathscr{E}_{x,\lambda}(\mathbf{Z}_2)$ by a function of x and λ which is *even* in x again belongs to $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$. Moreover, the set of functions which are even in x is a ring. (In mathematical language we may say that $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ is a *module* over the ring of even functions.) It turns out that this mathematical structure is an adequate foundation for our techniques. Let us develop this structure.

Lemma 2.1. *If $f \in \mathscr{E}_{x,\lambda}$ is even in x, then f may be expressed as a smooth function of x^2 and λ; in symbols*

$$f(x, \lambda) = a(x^2, \lambda). \tag{2.2}$$

Remark. If f is analytic (i.e., if f is given by a convergent power series), then a function a satisfying (2.2) may be obtained by substituting x^j for x^{2j} in the power series of f. In the C^∞ case, Lemma 2.1 was first proved by Whitney [1943]. We give a slightly different proof which is self-contained except for one reference to the literature. This proof illustrates the issues involved in the lemma.

Proof of lemma 2.1. We use the notation $u = x^2$. If (2.2) holds, then $a(u, \lambda)$ must satisfy

$$a(u, \lambda) = f(\sqrt{u}, \lambda) \quad \text{for} \quad u \geq 0. \tag{2.3}$$

Let us define $a(u, \lambda)$ for u positive by (2.3). It is clear that a is C^∞ on $\{u > 0\}$. We claim that all derivatives of a remain bounded as $u \to 0^+$. If we attempt to prove this claim directly by simply differentiating the right-hand side of (2.3) with respect to u, we encounter powers of u in the denominator which make the limit $u \to 0^+$ appear problematic. Rather we proceed as follows. For any integer k we use Taylor's theorem to write

$$f(x, \lambda) = \sum_{j=0}^{k-1} a_j(\lambda)x^{2j} + x^{2k}g(x, \lambda) \tag{2.4}$$

for some smooth function g; only even terms appear in the sum in (2.4), since f is even in x. Combining (2.3) and (2.4) we see that for $u > 0$

$$a(u, \lambda) = \sum_{j=0}^{k-1} a_j(\lambda)u^j + u^k g(\sqrt{u}, \lambda). \tag{2.5}$$

Clearly the first term in (2.5) is C^∞ for all u, being a polynomial. The second term is C^∞ for positive u, and *because of the factor* u^k, derivatives of order k or less of the second term remain bounded as $u \to 0^+$. However, we may apply the splitting (2.5) for any value of k. Thus all derivatives of $a(u, \lambda)$ remain bounded as $u \to 0^+$. This proves the claim.

It remains to define $a(u, \lambda)$ for $u < 0$. It follows from the claim above that a C^∞-extension of a exists. There are several proofs of this in the literature; we recommend Seeley [1964]. □

Remark. The proof shows that $a(u, \lambda)$ in (2.2) is not unique. More precisely, if $\phi(u, \lambda)$ is any C^∞-function such that $\phi(u, \lambda) \equiv 0$ for $u \geq 0$, then

$$f(x, \lambda) = a(x^2, \lambda) + \phi(x^2, \lambda)$$

is another representation of the form (2.2). However, the derivatives of $a(u, \lambda)$ to all orders at zero are determined. Typically the singularities we consider are finitely determined, and the lack of uniqueness in (2.2) will be of no consequence for us.

Corollary 2.2. *If* $g \in \vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$, *there is a smooth coefficient* $a(u, \lambda)$ *such that*

$$g(x, \lambda) = a(x^2, \lambda)x. \tag{2.6}$$

Proof. Since $g(x, \lambda)$ is odd in x, $g(0, \lambda) = 0$. By Taylor's theorem, there is a smooth function $f(x, \lambda)$ such that

$$g(x, \lambda) = f(x, \lambda)x.$$

Moreover, $f(x, \lambda)$ is even in x; (2.6) follows from applying Lemma 2.1 to $f(x, \lambda)$. □

Corollary 2.2 leads to a representation for $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ that is the basis of our study of $\mathbf{Z}_2$-symmetric bifurcation problems; viz.,

$$\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2) = \mathscr{E}_{u,\lambda} \cdot \{x\}, \tag{2.7}$$

where $u = x^2$ and $\mathscr{E}_{u,\lambda}$ denotes the ring of *all* germs of smooth functions of u and λ. Observe that there are no symmetry restrictions on elements of $\mathscr{E}_{u,\lambda}$. The representation (2.7) is especially convenient precisely because it eliminates symmetry restrictions that are awkward to work with. (In Volume II such a representation is more a necessity than a convenience.)

Let us extend the representation (2.7) to (appropriate) subsets of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$. We shall call a vector subspace $\vec{\mathscr{J}}$ of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ a *submodule* of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ if for every $g \in \vec{\mathscr{J}}$ and every $a \in \mathscr{E}_{u,\lambda}$, we have $a(x^2, \lambda)g \in \vec{\mathscr{J}}$. For example, let $\mathscr{M} = \langle u, \lambda \rangle$ be the maximal ideal in $\mathscr{E}_{u,\lambda}$; then for any k, $\mathscr{M}^k \cdot \{x\}$ is a submodule of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$. In the next lemma we show that every submodule of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ arises from an ideal in $\mathscr{E}_{u,\lambda}$ in this way.

Lemma 2.3. *For every submodule $\vec{\mathscr{J}}$ of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ there is an ideal $\mathscr{J} \subset \mathscr{E}_{u,\lambda}$ such that*

$$\vec{\mathscr{J}} = \mathscr{J} \cdot \{x\}. \tag{2.8}$$

Conversely, for every ideal $\mathscr{J}$ in $\mathscr{E}_{u,\lambda}$, (2.8) defines a submodule of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$.

Proof. If $\mathscr{J}$ is an ideal in $\mathscr{E}_{u,\lambda}$, clearly $\mathscr{J} \cdot \{x\}$ is a submodule of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$. Conversely, if $\vec{\mathscr{J}}$ is a submodule of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$, let

$$\mathscr{J} = \{a \in \mathscr{E}_{u,\lambda} : a(x^2, \lambda)x \in \vec{\mathscr{J}}\}.$$

Then $\mathscr{J}$ is an ideal in $\mathscr{E}_{u,\lambda}$. By Corollary 2.2, every $g \in \vec{\mathscr{J}}$ admits the representation $g(x, \lambda) = a(x^2, \lambda)x$, so we see that $\vec{\mathscr{J}} = \mathscr{J} \cdot \{x\}$. □

We may use Lemma 2.3 to generalize various concepts involving ideals to submodules of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$. Specifically we shall use the following three concepts:

(i) If $\mathscr{J} = \langle a_1, \ldots, a_k \rangle$ is a finitely generated ideal in $\mathscr{E}_{u,\lambda}$ and if

$$\vec{\mathscr{J}} = \langle a_1, \ldots, a_k \rangle \cdot \{x\},$$

we shall say that $\vec{\mathscr{J}}$ is *generated* by $a_1(x^2, \lambda)x, \ldots, a_k(x^2, \lambda)x$.

(ii) If $\vec{\mathscr{I}}$ and $\vec{\mathscr{J}}$ are submodules of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$, then the set of all sums from $\vec{\mathscr{I}}$ and $\vec{\mathscr{J}}$ is also a submodule, denoted $\vec{\mathscr{I}} + \vec{\mathscr{J}}$. In the notation of Lemma 2.3,

$$\vec{\mathscr{I}} + \vec{\mathscr{J}} = (\mathscr{I} + \mathscr{J}) \cdot \{x\}.$$

(iii) If there is a k-dimensional subspace V of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ such that

$$\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2) = \vec{\mathscr{J}} \oplus V,$$

we say that $\vec{\mathscr{J}}$ has codimension k. In the notation of Lemma 2.3,

$$\operatorname{codim} \vec{\mathscr{J}} = \operatorname{codim} \mathscr{J},$$

where the latter codimension is computed in $\mathscr{E}_{u,\lambda}$.

We shall use freely the following four facts concerning submodules of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$. Proofs rely on the representation (2.8) for submodules; we ask the reader to supply the details. The corresponding facts concerning ideals from Chapter II are flagged in parentheses. In facts (iii) and (iv), $\mathscr{M}$ refers to the maximal ideal in $\mathscr{E}_{u,\lambda}$.

Facts 2.4. (i) If $\vec{\mathscr{J}}$ is generated by $p_1, \ldots, p_k$ and if $p_k = a_1 p_1 + \cdots + a_{k-1} p_{k-1}$, then $\vec{\mathscr{J}}$ is generated by $p_1, \ldots, p_{k-1}$. (Cf. Lemma II,4.1.)

(ii) If $\vec{\mathscr{J}}$ is generated by $p_1, \ldots, p_k$ and if

$$q_i = \sum_{i=1}^{k} a_{ij} p_j, \qquad i = 1, \ldots, k,$$

where $a_{ij} \in \mathscr{E}_{u,\lambda}$ and $\{a_{ij}(0,0)\}$ is an invertible $k \times k$ matrix, then $\vec{\mathscr{J}}$ is also generated by $q_1, \ldots, q_k$. (Cf. Lemma II, 4.2.)

(iii) Let $\vec{\mathscr{I}}$ and $\vec{\mathscr{J}}$ be submodules, with $\vec{\mathscr{I}}$ finitely generated. Then $\vec{\mathscr{I}} \subset \vec{\mathscr{J}}$ if and only if $\vec{\mathscr{I}} \subset \vec{\mathscr{J}} + \mathscr{M} \cdot \vec{\mathscr{I}}$. (Cf. Nakayama's lemma, Lemma II,5.3.)

(iv) A submodule $\vec{\mathscr{J}}$ has finite codimension if and only if $\mathscr{M}^k \cdot \{x\} \subset \vec{\mathscr{J}}$ for some k. (Cf. Proposition II,5.7.)

Remarks. (i) In the subsections below we define submodules of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ associated to the restricted tangent space, higher-order terms, etc. We will not encumber the notation by putting an arrow above these spaces.

(ii) We are aware that we have introduced a fair amount of algebraic terminology to deal with a problem that could be analyzed with "bare hands" techniques. We believe that in the long run the presentation will be clearer because of this terminology. There is also another consideration. The present terminology is a necessity in the situations we consider in Volume II; i.e., more complicated symmetry groups or bifurcation problems (even without symmetry) in several variables. In our estimation, seeing the terminology first in a relatively elementary application will simplify the reader's task in Volume II.

(b) The Restricted Tangent Space

Let us define equivalence in the symmetric context.

Definition 2.5. Let $g(x, \lambda)$ and $h(x, \lambda)$ be bifurcation problems with $\mathbf{Z}_2$-symmetry. We say that g and h are *$\mathbf{Z}_2$-equivalent* if

$$h(x, \lambda) = S(x, \lambda) \cdot g(X(x, \lambda), \Lambda(\lambda)),$$

where the triple (S, X, Λ) is an equivalence transformation such that X is odd in x and S is even in x. If this relation holds with $\Lambda(\lambda) \equiv \lambda$, we say that g and h are *strongly $\mathbf{Z}_2$-equivalent*.

Remark. The only difference between equivalence and $\mathbf{Z}_2$-equivalence is that the change of coordinates respects the $\mathbf{Z}_2$-symmetry. More precisely, if S is even in x and X is odd in x then

$$S(x, \lambda) \cdot g(X(x, \lambda), \Lambda(\lambda))$$

is odd in x for every g which is odd in x. Conversely, any equivalence for which this statement is true must be a $\mathbf{Z}_2$-equivalence. (See Exercise 2.1.)

We now turn to the restricted tangent space in the symmetric context, denoted $RT(g, \mathbf{Z}_2)$. We motivate our definition of $RT(g, \mathbf{Z}_2)$ below by a calculation that is identical in form to the calculation in Chapter II, §2 which motivates the definition of $RT(g)$; the only difference is that here we use strong equivalences which preserve the $\mathbf{Z}_2$-symmetry. Let $g \in \vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$; consider a one-parameter family G of germs strongly $\mathbf{Z}_2$-equivalent to g, say

$$G(x, \lambda, t) = S(x, \lambda, t)g(X(x, \lambda, t), \lambda),$$

where $S(x, \lambda, 0) = 1$, $X(x, \lambda, 0) = x$, S is even in x, and X is odd in x. Then a typical element of $RT(g, \mathbf{Z}_2)$ is given by

$$\left.\frac{\partial}{\partial t} G(x, \lambda, t)\right|_{t=0} = \dot{S}(x, \lambda, 0)g(x, \lambda) + g_x(x, \lambda)\dot{X}(x, \lambda, 0), \tag{2.9}$$

where dot indicates a derivative with respect to t. In other words, $RT(g, \mathbf{Z}_2)$ is the totality of germs that can arise in (2.9) through the above construction.

Let us work (2.9) into a form more suited to our algebraic concepts above. By Corollary 2.2 there is a smooth germ $r(u, \lambda)$ such that $g(x, \lambda) = r(x^2, \lambda)x$. Similarly, $\dot{S}$ is even in x and $\dot{X}$ is odd, so we may write $\dot{S}(x, \lambda, 0) = a(x^2, \lambda)$, $\dot{X}(x, \lambda, 0) = b(x^2, \lambda)x$. Substituting into (2.9) we find

$$\left.\frac{\partial}{\partial t} G(x, \lambda, t)\right|_{t=0} = a(u, \lambda)r(u, \lambda)x$$
$$+ b(u, \lambda)[r(u, \lambda) + 2ur_u(u, \lambda)]x, \tag{2.10}$$

where $u = x^2$. Now a and b are arbitrary elements of $\mathscr{E}_{u,\lambda}$. Thus $RT(g, \mathbf{Z}_2)$ is the module over $\mathscr{E}_{u,\lambda}$ generated by the two elements

$$r(u, \lambda)x, \qquad [r(u, \lambda) + 2ur_u(u, \lambda)]x.$$

We formalize this in the following definition, taking advantage of Fact 2.4(ii) to simplify the above generators.

Definition 2.6. Let $g(x, \lambda)$ be in $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$, and assume $g(x, \lambda) = r(x^2, \lambda)x$. We define

$$RT(g, \mathbf{Z}_2) = \langle r, ur_u \rangle \cdot \{x\},$$

where $u = x^2$.

For example, $RT(x^3 - \lambda x, \mathbf{Z}_2) = \langle u, \lambda \rangle \cdot \{x\}$ and $RT(x^3 - \lambda^2 x, \mathbf{Z}_2) = \langle u, \lambda^2 \rangle \cdot \{x\}$. (Note that $x^3 - \lambda x$ is the pitchfork, while $x^3 - \lambda^2 x$ is the nondegenerate cubic considered in (V,2.7).)

Definition 2.6 shows that the module $RT(g, \mathbf{Z}_2)$ is generated by just two elements. By contrast, in the nonsymmetric context $RT(g)$ requires three generators; viz., g, xg_x, and λg_x. The origin of this difference lies in the following fact: In the nonsymmetric context, we must require explicitly that

$X(0, 0) = 0$, but in the symmetric context this follows naturally from the definitions; indeed, since X is odd in x,

$$X(0, \lambda) \equiv 0. \tag{2.11}$$

The following theorem, analogous to Theorem II,2.2, provides the main sufficient condition for strong $\mathbf{Z}_2$-equivalence. We do not prove this result, as the proof requires only minor variations from the proof of Theorem II,2.2. (See Exercise 2.2.)

Theorem 2.7. *Let $h(x, \lambda)$ and $p(x, \lambda)$ be bifurcation problems with $\mathbf{Z}_2$-symmetry. Assume that*

$$RT(h + tp, \mathbf{Z}_2) = RT(h, \mathbf{Z}_2) \quad \textit{for all} \quad t \in [0, 1].$$

Then $h + tp$ is strongly $\mathbf{Z}_2$-equivalent to h for all $t \in [0, 1]$.

(c) Intrinsic Submodules

We shall call a submodule $\vec{\mathscr{J}} \subset \vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ *intrinsic* if for all $g, h \in \vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$,

$$g \in \vec{\mathscr{J}} \quad \text{and} \quad h \sim g \quad \Rightarrow \quad h \in \vec{\mathscr{J}},$$

where $h \sim g$ means h is strongly equivalent to g. In the following proposition we characterize intrinsic submodules. This result is similar in spirit to Proposition II,7.1, but the actual conclusions are quite different. For example, $\langle u \rangle \cdot \{x\}$ is an intrinsic submodule of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$, even though in the nonsymmetric context $\langle x^3 \rangle$ is definitely not an intrinsic ideal in $\mathscr{E}_{x,\lambda}$. This difference has its origins in (2.11)—in the symmetric context the λ-axis (i.e., the set $\{x = 0\}$) is mapped into itself by $\mathbf{Z}_2$-equivalences. The following analogy may be helpful in understanding this: In either context the x-axis (i.e., $\{\lambda = 0\}$) is mapped into itself by equivalences; because of this fact $\langle \lambda \rangle$ is an intrinsic ideal (in the nonsymmetric context) and $\langle \lambda \rangle \cdot \{x\}$ is an intrinsic submodule (in the symmetric context).

Proposition 2.8. *Let $\vec{\mathscr{J}}$ be a submodule of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ of finite codimension. Then $\vec{\mathscr{J}}$ is intrinsic if and only if it can be written in the form*

$$\vec{\mathscr{J}} = \langle u^{k_1}\lambda^{l_1}, \ldots, u^{k_s}\lambda^{l_s} \rangle \cdot \{x\}; \tag{2.12}$$

i.e., if and only if $\vec{\mathscr{J}}$ is generated by monomials.

Remarks. Usually in (2.12) we will require that

$$\begin{aligned} &\text{(a)} \quad k_1 > k_2 > \cdots > k_s = 0, \\ &\text{(b)} \quad 0 = l_1 < l_2 < \cdots < l_s. \end{aligned} \tag{2.13}$$

These inequalities are to avoid redundancies among the generators—if we had two generators, say $u^{k_1}\lambda^{l_1}x$ and $u^{k_2}\lambda^{l_2}x$, where $k_2 \geq k_1$ and $l_2 \geq l_1$, we could use Fact 2.4(i) to discard $u^{k_2}\lambda^{l_2}x$. We justify the equality $k_s = 0$ as follows. Let $\vec{\mathscr{J}}$ be given by (2.12) where $k_i > 0$ for $i = 1, \ldots, s$; then for every positive integer k, $\lambda^k x \notin \vec{\mathscr{J}}$. By Fact 2.4(iv), $\vec{\mathscr{J}}$ would have infinite codimension. The justification of $l_1 = 0$ is similar.

Proof of Proposition 2.8. Let $g(x, \lambda) = r(x^2, \lambda)x$ be in $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$. Let us apply a strong $\mathbf{Z}_2$-equivalence (S, X) to g. We may write $S(x, \lambda) = a(u, \lambda)$ and $X(x, \lambda) = b(u, \lambda)x$, where $a(0, 0) > 0$ and $b(0, 0) > 0$. Thus

$$S(x, \lambda)g(X(x, \lambda), \lambda) = a(u, \lambda)r(U, \lambda)b(u, \lambda)x,$$

where $U = X(x, \lambda)^2 = b(u, \lambda)^2 u$. In particular, under a $\mathbf{Z}_2$-equivalence, $u^k\lambda^l x$ is mapped into

$$a(u, \lambda)b^{2k+1}(u, \lambda)u^k\lambda^l x.$$

Thus the submodule $\langle u^k\lambda^l\rangle \cdot \{x\}$ is intrinsic. Since sums of intrinsic submodules are intrinsic, we see that (2.12) defines an intrinsic submodule.

Conversely, it follows from Proposition 2.9 below that any intrinsic submodule of finite codimension may be written in the form (2.12). (Cf. the proof of Proposition II, 7.1.) □

Proposition 2.9. *Let $\vec{\mathscr{J}}$ be an intrinsic submodule of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ of finite codimension, and let $p(u, \lambda)$ be a polynomial, say*

$$p(u, \lambda) = \sum_\alpha a_\alpha u^{\alpha_1}\lambda^{\alpha_2}.$$

Then $p(u, \lambda)x$ belongs to $\mathscr{J}$ if and only if for every α such that $a_\alpha \neq 0$, the monomial $u^{\alpha_1}\lambda^{\alpha_2}x$ belongs to $\mathscr{J}$.

We leave the proof of this proposition for the reader, as it is rather similar to the proof of Proposition II, 7.3, the counterpart of Proposition 2.9 in Chapter II.

Remark. We see from Proposition 2.8 that intrinsic submodules are in fact invariant under general (i.e., not necessarily strong) $\mathbf{Z}_2$-equivalences. (Cf. Remark II, 7.7.)

Definition 2.10. Assuming (2.13) holds, we call the monomials in (2.12) the *intrinsic generators* of $\vec{\mathscr{J}}$.

If $\vec{\mathscr{J}}$ is a submodule of finite codimension, let $\mathscr{J}^\perp$ be the finite-dimensional vector subspace of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ spanned by the monomials $u^k\lambda^l x$ *not* in $\mathscr{J}$. As in Chapter II, we may decompose such submodules as

$$\vec{\mathscr{J}} = \operatorname{Itr} \vec{\mathscr{J}} \oplus V, \tag{2.14}$$

where Itr $\vec{\mathscr{J}}$ is the intrinsic part of $\vec{\mathscr{J}}$ and $V = (\text{Itr } \vec{\mathscr{J}})^{\perp} \cap \vec{\mathscr{J}}$. In terms of the representation (2.7) of submodules, we have

$$\text{Itr } \vec{\mathscr{J}} = (\text{Itr } \mathscr{J}) \cdot \{x\},$$

where Itr $\mathscr{J}$ is determined in $\mathscr{E}_{u,\lambda}$.

(d) Statement of the Main Result

For the duration of this subsection, let $h(x, \lambda) = s(x^2, \lambda)x$ be a germ in $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ such that $RT(h, \mathbf{Z}_2)$ has finite codimension. We shall use this convention even when it is not made explicit. As in Chapter II, §8, we discuss in sequence low-order, higher-order, and intermediate-order terms for the recognition problem for h. No proofs are given in this subsection, as they are similar to their counterparts in Chapter II.

Let $\mathscr{S}(h, \mathbf{Z}_2)$ be the smallest intrinsic submodule of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ that contains h. $\mathscr{S}(h, \mathbf{Z}_2)$ is an intrinsic submodule of finite codimension. In the next two results, the multi-index notation D^α means $(\partial/\partial u)^{\alpha_1}(\partial/\partial\lambda)^{\alpha_2}$.

Proposition 2.11.

(a)
$$\mathscr{S}(h, \mathbf{Z}_2) = \left(\sum_\alpha \langle u^{\alpha_1}\lambda^{\alpha_2}\rangle\right) \cdot \{x\},$$

where the sum extends over all multi-indices α such that $D^\alpha s(0, 0) \neq 0$.

(b) *If g is $\mathbf{Z}_2$-equivalent to h, then $\mathscr{S}(g, \mathbf{Z}_2) = \mathscr{S}(h, \mathbf{Z}_2)$.*

Theorem 2.12. *Let $g(x, \lambda) = r(x^2, \lambda)x$ be equivalent to h.*

(a) *For every monomial $u^{\alpha_1}\lambda^{\alpha_2}x \in \mathscr{S}(h, \mathbf{Z}_2)^{\perp}$, we have $D^\alpha r(0, 0) = 0$.*

(b) *For each intrinsic generator of $\mathscr{S}(h, \mathbf{Z}_2)$ we have $D^\alpha r(0, 0) \neq 0$.*

We define $\mathscr{P}(h, \mathbf{Z}_2)$, the *higher-order terms* associated to h, by the following condition: $p \in \mathscr{P}(h, \mathbf{Z}_2)$ if for every g strongly $\mathbf{Z}_2$-equivalent to h and for every $t \in \mathbb{R}$

$$RT(g + tp, \mathbf{Z}_2) = RT(g, \mathbf{Z}_2).$$

It follows immediately from Theorem 2.7 that if $p \in \mathscr{P}(h, \mathbf{Z}_2)$ and if g is strongly $\mathbf{Z}_2$-equivalent to h, then $g + p$ is strongly $\mathbf{Z}_2$-equivalent to g. Also, $\mathscr{P}(h, \mathbf{Z}_2)$ is an intrinsic submodule of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ of finite codimension. The following theorem characterizes $\mathscr{P}(h, \mathbf{Z}_2)$ explicitly.

Theorem 2.13.

$$\mathscr{P}(h, \mathbf{Z}_2) = \text{Itr}\{\mathscr{M} \cdot RT(h, \mathbf{Z}_2)\}. \tag{2.15}$$

Comparison of (2.15) with (II, 8.5) shows that there is an extra term (i.e., λh_x) which contributes to $\mathscr{P}$ in the nonsymmetric context. This difference is a consequence of (2.11). The proof that $\mathscr{P}(h, \mathbf{Z}_2) \supset \operatorname{Itr}\{\mathscr{M} \cdot RT(h, \mathbf{Z}_2)\}$ follows from Nakayama's lemma as in Lemma II, 5.3. The proof of the reverse containment is much simpler than the corresponding proof in the nonsymmetric cases. (Cf. Chapter II, §13.) The reason is that intrinsic submodules must have at least two intrinsic generators, unlike the nonsymmetric case. (Cf. Exercise 2.3.)

The treatment of intermediate-order terms in the symmetric context involves the same issues as in the nonsymmetric context; i.e., having reduced a germ g modulo $\mathscr{P}(h, \mathbf{Z}_2)$, we perform explicit changes of coordinate on g to determine precisely when g is equivalent to h.

(e) Two Simple Examples

In this subsection we apply the above results to solve the recognition problem for the following two classes of normal forms:

$$\begin{aligned} &\text{(a)} \quad (\varepsilon u^k + \delta\lambda)x, \\ &\text{(b)} \quad (\varepsilon u + \delta\lambda^k)x. \end{aligned} \tag{2.16}$$

As usual, ε and δ equal ± 1; we suppose $k \geq 1$.

Proposition 2.14. *Let $g(x, \lambda) = r(u, \lambda)x$ be in $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$. Then g is strongly $\mathbf{Z}_2$-equivalent to $(\varepsilon u^k + \delta\lambda)x$ if and only if*

$$r = \frac{\partial r}{\partial u} = \cdots = \left(\frac{\partial}{\partial u}\right)^{k-1} r = 0 \tag{2.17a}$$

at $u = \lambda = 0$ and

$$\operatorname{sgn}\left(\frac{\partial}{\partial u}\right)^{k} r(0, 0) = \varepsilon, \qquad \operatorname{sgn} r_\lambda(0, 0) = \delta. \tag{2.17b}$$

Proof. If h is the normal form (2.16a), then $RT(h, \mathbf{Z}_2) = \langle u^k, \lambda\rangle \cdot \{x\}$. The defining conditions (2.17a) follow from Theorem 2.12(a). By Theorem 2.12(b) we may write

$$g(x, \lambda) = (Au^k + B\lambda)x + p(x, \lambda), \tag{2.18}$$

where $A \neq 0$, $B \neq 0$, and

$$p \in \langle u^{k+1}, \lambda u^k, \lambda^2\rangle \cdot \{x\}. \tag{2.19}$$

We know that $RT(h, \mathbf{Z}_2) = \langle u^k, \lambda\rangle \cdot \{x\}$, so that $\mathscr{M} \cdot RT(h, \mathbf{Z}_2)$ is precisely the right-hand side of (2.19), which is already intrinsic. Therefore, by Theorem 2.13 the term p in (2.18) does not affect whether g is equivalent to h. Finally, we may scale the coefficients A and B in (2.18) to ε and δ, respectively, if and only if (2.17b) holds. □

Proposition 2.15. *Let $g(x, \lambda) = r(x, \lambda)x$ be in $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$. Then g is strongly $\mathbf{Z}_2$-equivalent to $(\varepsilon u + \delta\lambda^k)x$ if and only if*

$$r = \frac{\partial r}{\partial \lambda} = \cdots = \left(\frac{\partial}{\partial \lambda}\right)^{k-1} r = 0 \tag{2.20a}$$

at $u = \lambda = 0$ and

$$\operatorname{sgn} r_u(0, 0) = \varepsilon, \qquad \operatorname{sgn}\left(\frac{\partial}{\partial \lambda}\right)^k r(0, 0) = \delta. \tag{2.20b}$$

The proof of this proposition is left as an exercise for the reader.

Remark. If $k = 1$ in either (2.16a) or (2.16b), we obtain the pitchfork. In this case there is only one defining condition in (2.17a). Of course, this is quite different from the nonsymmetric context, where the pitchfork has four defining conditions. (There is, however, a hidden defining condition in Proposition 2.14 in the implicit assumption that the singularity occurs on the axis of symmetry, $x = 0$.)

EXERCISES

2.1. Fix an equivalence X, Λ, S and suppose

$$h(x, \lambda) = S(x, \lambda)g(X(x, \lambda), \Lambda(\lambda))$$

is an odd function in x for every g which is odd in x. Show that S is even in x and X is odd in x so that the equivalence is a $\mathbf{Z}_2$-equivalence.

2.2. Prove Theorem 2.7. *Hint:* Follow the proof of Theorem II,2.2 in Chapter II, §11. To prove that the resulting strong equivalence is, in fact, a strong $\mathbf{Z}_2$-equivalence, use the uniqueness of solutions to the initial value problem for ODE's.

2.3. Prove Proposition 2.9.

2.4. (a) Show that an intrinsic submodule of finite $\mathbf{Z}_2$-codimension must have at least two intrinsic generators.
(b) Prove Theorem 2.13, using the proofs in Chapter II, §13 as a guide.

2.5. Prove Proposition 2.15.

§3. Universal Unfoldings

This section is a fairly straightforward extension of the theory of universal unfoldings to the $\mathbf{Z}_2$-symmetric context. There are three subsections, which correspond roughly to §§1–3 of Chapter III. In subsection (a) we give the basic definitions, in subsection (b), we state the universal unfolding theorem in the symmetric context, and in subsection (c), we use the theorem to compute universal unfoldings for the germs (2.16) considered above.

(a) Basic Definitions

Let $g(x, \lambda)$ be a bifurcation problem with $\mathbf{Z}_2$-symmetry. We shall call $G(x, \lambda, \alpha)$ a k-parameter *$\mathbf{Z}_2$-unfolding* of $g(x, \lambda)$ if $G: \mathbb{R} \times \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}$ satisfies $G(x, \lambda, 0) = g(x, \lambda)$ and $G(-x, \lambda, \alpha) = -G(x, \lambda, \alpha)$.

Definition 3.1. (i) Let $G(x, \lambda, \alpha)$ and $H(x, \lambda, \beta)$ be $\mathbf{Z}_2$-unfoldings of a germ g in $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$. The unfolding *H factors through* the unfolding G if

$$H(x, \lambda, \beta) = S(x, \lambda, \beta) \cdot G(X(x, \lambda, \beta), \Lambda(\lambda, \beta), A(\beta)),$$

where $S(x, \lambda, 0) = 1$, $X(x, \lambda, 0) = x$, $\Lambda(\lambda, 0) = \lambda$, $A(0) = 0$, $S(-x, \lambda, \beta) = S(x, \lambda, \beta)$, and $X(-x, \lambda, \beta) = -X(x, \lambda, \beta)$.

(ii) The $\mathbf{Z}_2$-unfolding G of g is a *$\mathbf{Z}_2$-versal unfolding* if every $\mathbf{Z}_2$-unfolding H of g factors through G.

(iii) A versal unfolding of g is *universal* if it has the minimum number of parameters in a versal unfolding of g. We call this minimum number the *$\mathbf{Z}_2$-codimension* of g; in symbols, $\operatorname{codim}_{\mathbf{Z}_2} g$.

Definition 3.2. If $g \in \vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$, the tangent space of g is the following subspace of $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$:

$$T(g, \mathbf{Z}_2) = RT(g, \mathbf{Z}_2) + \mathscr{E}_\lambda\{g_\lambda\}. \tag{3.1}$$

In Chapter III, §2(a) we motivated the corresponding definition in the nonsymmetric context by showing the following: $T(g)$ arises as the set of all possible derivatives

$$\frac{\partial}{\partial t} S(x, \lambda, t) g(X(x, \lambda, t), \Lambda(\lambda, t))\Big|_{t=0},$$

where (S, X, Λ) is a one-parameter family of equivalence transformations such that (S, X, Λ) is the identity equivalence when $t = 0$. In the symmetric context, $T(g, \mathbf{Z}_2)$ has a similar motivation; we ask the reader to carry out the required calculation in Exercise 3.1. (Cf. §2(b) above.)

Formula (3.1) differs slightly from its analogue in the nonsymmetric context; viz.,

$$T(g) = RT(g) + \mathbb{R}\{g_x\} + \mathscr{E}_\lambda\{g_\lambda\}.$$

This difference derives from the fact that in the symmetric context necessarily $X(0, 0) = 0$.

Suppose that $g(x, \lambda) = r(x^2, \lambda)x$. Recalling the definition of $RT(g, \mathbf{Z}_2)$, we may rewrite (3.1) as

$$T(g, \mathbf{Z}_2) = (\langle r, ur_u\rangle + \mathscr{E}_\lambda\{r_\lambda\}) \cdot \{x\}. \tag{3.2}$$

(b) The Universal Unfolding Theorem

Theorem 3.3. *Let* $g(x, \lambda) = r(u, \lambda)x$ *be in* $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$, *and let* $G(x, \lambda, \alpha) = R(u, \lambda, \alpha)x$ *be a k-parameter* $\mathbf{Z}_2$*-unfolding of* g. *Then* G *is a versal* $\mathbf{Z}_2$*-unfolding if and only if the codimension of* $T(g, \mathbf{Z}_2)$ *in* $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ *equals* k *and*

$$\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2) = T(g, \mathbf{Z}_2) + \mathbb{R}\left\{\frac{\partial R}{\partial \alpha_1}(u, \lambda, 0), \ldots, \frac{\partial R}{\partial \alpha_k}(u, \lambda, 0)\right\} \cdot \{x\}. \tag{3.3}$$

Just as in Chapter III, §2, the necessity of (3.3) can be derived by considering one-parameter unfoldings of g. (See Exercise 3.2.) We defer the proof of sufficiency for Volume II.

It follows from Theorem 3.3 that $\operatorname{codim}_{\mathbf{Z}_2} g$, as given in Definition 3.1, equals the codimension of $T(g, \mathbf{Z}_2)$ in $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$. As in Chapter III, §2, this number may be related to the number of defining conditions of the singularity. The formulas are slightly different in the symmetric context, because the implicit assumption $x = 0$ represents an effective defining condition. We do not pursue this here.

(c) Simple Examples

First let us consider the pitchfork. We find that

$$T(x^3 - \lambda x, \mathbf{Z}_2) = \mathscr{E}_{u,\lambda} \cdot \{x\}.$$

Hence $x^3 - \lambda x$ has codimension zero in the symmetric context and is its own universal unfolding.

In the following proposition we give universal unfoldings for the singularities (2.16).

Proposition 3.4. (a) *The bifurcation problem* $(\varepsilon u^k + \delta\lambda)x$, *where* $\varepsilon = \pm 1$ *and* $\delta = \pm 1$, *has* $\mathbf{Z}_2$*-codimension* $k - 1$; *a universal unfolding is provided by*

$$(\varepsilon u^k + \delta\lambda + \alpha_1 u + \cdots + \alpha_{k-1}u^{k-1})x. \tag{3.4}$$

(b) *The bifurcation problem* $(\varepsilon u + \delta\lambda^k)x$, *where* $\varepsilon = \pm 1$ *and* $\delta = \pm 1$, *has* $\mathbf{Z}_2$*-codimension* $k - 1$ *and universal unfolding*

$$(\varepsilon u + \delta\lambda^k + \alpha_0 + \alpha_1 + \cdots + \alpha_{k-2}\lambda^{k-2})x. \tag{3.5}$$

Let us consider (3.4) and (3.5) when $k = 2$, so that these singularities have codimension one. In Figures 3.1–3.3 we have drawn bifurcation diagrams for these universal unfoldings. (Here and below we simplify $\mathbf{Z}_2$-symmetric bifurcation diagrams by only drawing the portion for which $x \geq 0$. The part for which $x < 0$ is obtained by reflection.) These diagrams will play an important role in our analysis of persistent bifurcation diagrams in the next section. We shall see in §5 that these are the only bifurcation problems of codimension one in the symmetric context.

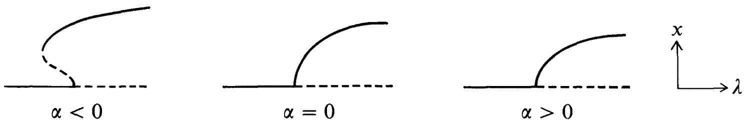


Figure 3.1. Bifurcation in the universal unfolding $(u^2 - \lambda + \alpha u)x$.

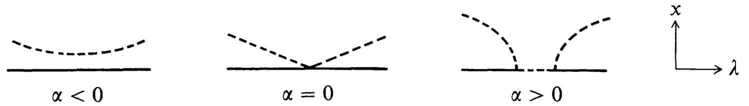


Figure 3.2. Bifurcation in the universal unfolding $(u - \lambda^2 + \alpha)x = 0$.

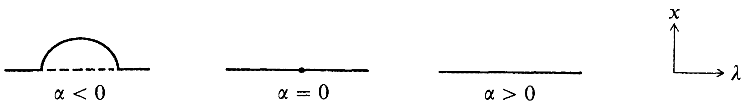


Figure 3.3. Bifurcation in the universal unfolding $(u + \lambda^2 + \alpha)x = 0$.

EXERCISES

3.1. Carry out the required differentiation to show that $T(g, \mathbf{Z}_2)$ should be defined as in (3.1).

3.2. Prove the necessity of condition (3.3) in the unfolding theorem, Theorem 3.3, using one-parameter unfoldings of g as in Chapter III, §2.

§4. Persistent Perturbations

In this section we discuss the theoretical basis for enumerating all perturbed bifurcation diagrams (up to $\mathbf{Z}_2$-equivalence) arising from a singularity. This section generalizes the material in Chapter III, §§5 and 6 to the $\mathbf{Z}_2$-symmetric context. On the whole, our methods here are rather similar to those of Chapter III. Specifically, we begin by enumerating the sources of nonpersistence; then, given a k-parameter unfolding $G(x, \lambda, \alpha)$, we identify a hypersurface Σ in the parameter space $\mathbb{R}^k$ such that $G(\cdot, \cdot, \alpha)$ exhibits nonpersistence when $\alpha \in \Sigma$; finally, we obtain an enumeration of the perturbed diagrams from the connected components of $\mathbb{R}^k \sim \Sigma$. However, the details of the construction are rather different, for the following reason. As we noted above, in the $\mathbf{Z}_2$-symmetric context, the point $x = 0$ is different from all others, being the only fixed point of the group. Different bifurcation phenomena occur for $x = 0$ and $x \neq 0$. Moreover, in unfolding a degenerate singularity at $x = 0$, we must consider possible sources of nonpersistence from both $x = 0$ and $x \neq 0$. In other words, there are more sources of nonpersistence in the symmetric context.

Let us begin to enumerate the sources of nonpersistence. Throughout the section, we consider a k-parameter, $\mathbf{Z}_2$-symmetric universal unfolding $G(x, \lambda, \alpha)$ of a germ $g \in \vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$. By Lemma 2.2 we may write G in the form

$$G(x, \lambda, \alpha) = R(u, \lambda, \alpha)x, \tag{4.1}$$

where $u = x^2$. We compile the following derivatives of G for use below:

$$\begin{aligned} &\text{(a)} \quad G_x = R + 2uR_u, \\ &\text{(b)} \quad G_\lambda = R_\lambda x, \\ &\text{(c)} \quad G_{xx} = (6R_u + 4uR_{uu})x. \end{aligned} \tag{4.2}$$

First, we consider the case $x \neq 0$ as this is more familiar. Indeed for $x \neq 0$ the only real difference between the symmetric and nonsymmetric context is that in the former context, all phenomena occur in pairs—whatever occurs at (x, λ, α) is mirrored at $(-x, \lambda, \alpha)$. The three sources of nonpersistence in the nonsymmetric context (i.e., bifurcation, hysteresis, and double limit points) carry over to $x \neq 0$ in the symmetric context essentially without change. Equations for these phenomena in the nonsymmetric context are given in Definition III,5.1; one may express these equations in terms of $R(u, \lambda, \alpha)$ in (4.1). For example bifurcation is associated with a value of $\alpha \in \mathbb{R}^k$ such that for some x, λ

$$G = G_x = G_\lambda = 0 \quad \text{at } (x, \lambda, \alpha). \tag{4.3}$$

Using (4.2) and the fact that $x \neq 0$, we see that (4.3) is equivalent to the equations

$$R = R_u = R_\lambda = 0 \quad \text{at } (u, \lambda, \alpha), \tag{4.4}$$

where $u > 0$. Thus we define the bifurcation variety

$$\mathscr{B}_1(\mathbf{Z}_2) = \{\alpha \in \mathbb{R}^k\colon \exists (u, \lambda), u > 0 \text{ such that } R = R_u = R_\lambda = 0 \text{ at } (u, \lambda, \alpha)\}. \tag{4.5}$$

(The subscript "1" indicates a phenomenon occurring for $x \neq 0$; this is to be distinguished from $\mathscr{B}_0(\mathbf{Z}_2)$, bifurcation phenomena occurring for $x = 0$.)

We have listed equations for $\mathscr{B}_1(\mathbf{Z}_2)$, for $\mathscr{H}_1(\mathbf{Z})$ (i.e., hysteresis associated to $x \neq 0$), and for $\mathscr{D}(\mathbf{Z}_2)$ (i.e., double limit points) in Table 4.1. Let us comment on the definition of $\mathscr{D}(\mathbf{Z}_2)$; viz.,

$$\begin{aligned} \mathscr{D}(\mathbf{Z}_2) = \{\alpha \in \mathbb{R}^k\colon \exists (u_1, u_2, \lambda), u_1 \neq u_2 \text{ and } u_i \geq 0 \text{ such that} \\ R = uR_u = 0 \text{ at } (u_i, \lambda, \alpha), i = 1, 2\}. \end{aligned} \tag{4.6}$$

In the $\mathbf{Z}_2$-symmetric context, double limit points can occur in a persistent way—if there is a limit point at (x, λ, α), there is also a limit point at $(-x, \lambda, \alpha)$. By working with $u = x^2$ in (4.6), we exclude such pairs from $\mathscr{D}(\mathbf{Z}_2)$. Also note that in (4.6) we only require $u_i \geq 0$, not strict inequality.

Table 4.1. Sources of Nonpersistence in the Symmetric Context.

$$\mathscr{B}_1(\mathbf{Z}_2) = \{\alpha \in \mathbb{R}^k \mid \exists (u, \lambda), u > 0 \text{ such that } R = R_u = R_\lambda = 0 \text{ at } (u, \lambda, \alpha)\}.$$
$$\mathscr{B}_0(\mathbf{Z}_2) = \{\alpha \in \mathbb{R}^k \mid \exists \lambda \text{ such that } R = R_\lambda = 0 \text{ at } (0, \lambda, \alpha)\}$$
$$\mathscr{H}_1(\mathbf{Z}_2) = \{\alpha \in \mathbb{R}^k \mid \exists (u, \lambda), u > 0 \text{ such that } R = R_u = R_{uu} = 0 \text{ at } (u, \lambda, \alpha)\}.$$
$$\mathscr{H}_0(\mathbf{Z}_2) = \{\alpha \in \mathbb{R}^k \mid \exists \lambda \text{ such that } R = R_u = 0 \text{ at } (0, \lambda, \alpha)\}.$$
$$\mathscr{D}(\mathbf{Z}_2) = \{\alpha \in \mathbb{R}^k \mid \exists \lambda, u_1 u_2 (u_1 \neq u_2; u_1, u_2 \geq 0) \text{ such that } R = uR_u = 0 \text{ at } (u_1, \lambda, \alpha) \text{ and } (u_2, \lambda, \alpha)\}$$
$$\Sigma(\mathbf{Z}_2) = \mathscr{B}_0(\mathbf{Z}_2) \cup \mathscr{B}_1(\mathbf{Z}_2) \cup \mathscr{H}_0(\mathbf{Z}_2) \cup \mathscr{H}_1(\mathbf{Z}_2) \cup \mathscr{D}(\mathbf{Z}_2)$$

In other words, the definition of $\mathscr{D}(\mathbf{Z}_2)$ mixes phenomena where $x \neq 0$ with phenomena where $x = 0$. To understand this better, suppose that $u_1 = 0$ in (4.6). Then the equation $uR_u = 0$ is satisfied automatically there; however, the equation $R = 0$ imposes a nontrivial restriction there. Indeed, by Proposition 2.14 (with $k = 1$), $G(\cdot, \cdot, \alpha)$ exhibits a pitchfork at (u_1, λ), provided $R_u \neq 0$, $R_\lambda \neq 0$. Thus when $\alpha \in \mathscr{D}(\mathbf{Z}_2)$ and $u_1 = 0$, a pitchfork bifurcation at $x = 0$ lies in the same λ-plane as a limit point at $x \neq 0$.

When $x = 0$ the nature of persistence changes dramatically. In particular, a pitchfork at $x = 0$ is persistent with respect to $\mathbf{Z}_2$-perturbations—this fact is the essential content of the statement that the pitchfork has $\mathbf{Z}_2$-codimension equal to 0. (In other words, a pitchfork at $x = 0$ in the symmetric context behaves like a limit point in the nonsymmetric context; cf. the discussion of $\mathscr{D}(\mathbf{Z}_2)$ above.) The solution to the recognition problem for the pitchfork (Proposition 2.14) is given by:

$$R = 0, \qquad R_u \neq 0, \qquad R_\lambda \neq 0 \quad \text{at } (0, 0).$$

Thus we have nonpersistent behavior if either

$$\begin{aligned} &\text{(a)} \quad R = R_\lambda = 0 \quad \text{at} \quad u = 0, \\ \text{or} \qquad &\text{(b)} \quad R = R_u = 0 \quad \text{at} \quad u = 0. \end{aligned} \tag{4.7}$$

These two possibilities lead to the sets $\mathscr{B}_0(\mathbf{Z}_2)$ and $\mathscr{H}_0(\mathbf{Z}_2)$ in Table 4.1, respectively. We justify the nomenclature "bifurcation" and "hysteresis" for the sets as follows. According to Propositions 2.14 and 2.15, the simplest $\mathbf{Z}_2$-symmetric bifurcation problems satisfying (4.7) are

$$\begin{aligned} &\text{(a)} \quad \pm(u \pm \lambda^2)x = 0, \\ &\text{(b)} \quad \pm(u^2 \pm \lambda)x = 0, \end{aligned} \tag{4.8}$$

respectively. If in (4.8) we exclude the zero solution $x = 0$ (this solution is forced by symmetry), then what remains in (4.8a) or (4.8b) is rather analogous to bifurcation and hysteresis in the nonsymmetric context. For example, $\{u = \lambda^2\}$ consists of two crossed lines in x, λ space, and $\{u^2 = \lambda\}$ consists of a single curve that makes a high order of contact with the line $\lambda = 0$. The bifurcation diagrams of the universal unfoldings of (4.8) are shown in Figures 3.1–3.3, and these graphs further support the nomenclature.

Although we do not prove it here, the above five sets enumerate all the sources of nonpersistence in the $\mathbf{Z}_2$-symmetric context. Thus we define

$$\Sigma(\mathbf{Z}_2) = \mathscr{B}_1(\mathbf{Z}_2) \cup \mathscr{B}_0(\mathbf{Z}_2) \cup \mathscr{H}_1(\mathbf{Z}_2) \cup \mathscr{H}_0(\mathbf{Z}_2) \cup \mathscr{D}(\mathbf{Z}_2).$$

If G is a versal unfolding of g, then Σ is a hypersurface in $\mathbb{R}^k$; i.e., has codimension one.

The following result provides a method for enumerating persistent bifurcation diagrams. Let G be a k-parameter, $\mathbf{Z}_2$-symmetric universal unfolding of a germ $g \in \vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ of finite codimension. In this theorem U, L, and W are appropriately small neighborhoods of zero in $\mathbb{R}$, $\mathbb{R}$, and $\mathbb{R}^k$, respectively. (We do not list the explicit requirements on these neighborhoods.)

Theorem 4.1. *If α_1, α_2 belong to the same connected component of $W \sim \Sigma(\mathbf{Z}_2)$, then there is a diffeomorphism (X, Λ) commuting with $\mathbf{Z}_2$, mapping $U \times L$ onto itself and a positive, even function S such that*

$$G(x, \lambda, \alpha_2) = S(x, \lambda) G(X(x, \lambda), \Lambda(\lambda), \alpha_1).$$

§5. The $\mathbf{Z}_2$-Classification Theorem

Theorem 5.1. *Let $g(x, \lambda) = r(u, \lambda)x$ be a germ in $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ satisfying $r(0, 0) = 0$. If $\operatorname{codim}_{\mathbf{Z}_2} g \leq 3$ then g is equivalent to one of the bifurcation problems listed in Table 5.1.*

Table 5.1. Normal Forms for Singularities of $\operatorname{codim}_{\mathbf{Z}_2} \leq 3$.

Normal Form	$\mathbf{Z}_2$-Codimension	Codimension
(1) $\varepsilon x^3 + \delta\lambda x$	0	2
(2) $\varepsilon x^3 + \delta\lambda^2 x$	1	5
(3) $\varepsilon x^5 + \delta\lambda x$	1	4
(4) $\varepsilon x^3 + \delta\lambda^3 x$	2	8
(5) $\varepsilon x^7 + \delta\lambda x$	2	6
(6) $\varepsilon x^3 + \delta\lambda^4 x$	3	11
(7) $\varepsilon x^9 + \delta\lambda x$	3	8
(8) $\varepsilon x^5 + 2m\lambda x^3 + \delta\lambda^2 x$, $m^2 \neq \varepsilon\delta$	3*	9
(9) $\phi x^7 + \varepsilon x^5 + 2\sigma\lambda x^3 + \varepsilon\lambda^2 x$	3	9
(10) $\varepsilon x^5 + \sigma\lambda x^3 + \phi\lambda^3 x$	3	11
(11) $\phi x^7 + \sigma\lambda x^3 + \varepsilon\lambda^2 x$	3	11

where $\varepsilon = \pm 1, \delta = \pm 1, \phi = \pm 1, \sigma = \pm 1$.

Table 5.2. Algebraic Data for Singularities of $\mathbf{Z}_2$-Codimension ≤ 3.

Normal form h	$\mathscr{S}(h, \mathbf{Z}_2)$*	$RT(h, \mathbf{Z}_2)$*	$\mathscr{P}(h, \mathbf{Z}_2)$*	$T(h, \mathbf{Z}_2)$*
$(\varepsilon u^k + \delta\lambda)x\ (k \geq 1)$	$\langle u^k, \lambda\rangle$	$\langle u^k, \lambda\rangle$	$\langle u^{k+1}, u\lambda, \lambda^2\rangle$	$\langle u^k, \lambda\rangle \oplus \mathbb{R}\{1\}$
$(\varepsilon u + \delta\lambda^k)x\ (k \geq 1)$	$\langle u, \lambda^k\rangle$	$\langle u, \lambda^k\rangle$	$\langle u^2, u\lambda, \lambda^{k+1}\rangle$	$\langle u, \lambda^{k-1}\rangle$
$(\varepsilon u^2 + 2m\lambda u + \delta\lambda^2)x$ $m^2 \neq \varepsilon\delta$	$\mathscr{M}^2$ if $m \neq 0$ $\langle u^2, \lambda^2\rangle$ if $m = 0$	$\mathscr{M}^3 \oplus \mathbb{R}\{\varepsilon u^2 - \delta\lambda^2, \varepsilon u^2 + m\lambda u\}$	$\mathscr{M}^3$	$RT(h, \mathbf{Z}_2) \oplus \mathbb{R}\{mu + \delta\lambda\}$
$(\phi u^3 + \varepsilon u^2 + 2\delta\lambda u + \varepsilon\lambda^2)x$	$\mathscr{M}^2$	$\mathscr{M}^4 \oplus \mathbb{R}\{\varepsilon u^2 + 4\delta u\lambda + 3\varepsilon\lambda^2,$ $3\phi u^3 + 2\varepsilon u^2 + 2\delta\lambda u,$ $\delta u^2\lambda + \varepsilon u\lambda^2, \delta u\lambda^2 + \varepsilon\lambda^3,$ $\varepsilon u^3 + \delta u^2\lambda\}$	$\mathscr{M}^4$	$RT(h, \mathbf{Z}_2) \oplus \mathbb{R}\{\delta u + \varepsilon\lambda, \delta u\lambda + \varepsilon\lambda^2\}$
$(\varepsilon u^2 + 2\delta\lambda u + \phi\lambda^3)x$	$\langle u^2, \lambda u, \lambda^3\rangle$	$\langle u^3, u^2\lambda, u\lambda^2, \lambda^4\rangle$ $\oplus \mathbb{R}\{\varepsilon u^2 - \phi\lambda^3, \varepsilon u^2 + \delta u\lambda\}$	$\langle u^3, u^2\lambda, u\lambda^2, \lambda^4\rangle$	$\langle u^2, u\lambda, \lambda^3\rangle \oplus \mathbb{R}\{2\delta u + 3\phi\lambda^2\}$
$(\varepsilon u^3 + \delta\lambda u + \phi\lambda^2)x$	$\langle u^3, \lambda u, \lambda^2\rangle$	$\langle u^4, u^2\lambda, u\lambda^2, \lambda^3\rangle$ $\oplus \mathbb{R}\{2\varepsilon u^3 - \phi\lambda^2, 3\varepsilon u^3 + \delta\lambda u\}$	$\langle u^4, u^2\lambda, u\lambda^2, \lambda^3\rangle$	$\langle u^3, u\lambda, \lambda^2\rangle \oplus \mathbb{R}\{\delta u + 2\phi\lambda\}$

* Each vector subspace and submodule in $\vec{\mathscr{E}}_{x,\lambda}(\mathbf{Z}_2)$ is of the form $V \cdot \{x\}$ in $\mathscr{E}_{u,\lambda} \cdot \{x\}$, respectively. In this table we have written only the subspace V and have omitted $\cdot \{x\}$.

Remarks 5.2. (i) The normal forms in Table 5.1 fit into three families. We have already studied the families (1), (2), (4), (6) and (1), (3), (5), (7) in Proposition 3.4. The new family consists of the moduli family (8) with distinguished points (9), (10), (11), all with $\mathbf{Z}_2$-codimension three. We shall analyze this moduli family in detail in §§7 and 8.

(ii) Normal form (8) has modality one and topological $\mathbf{Z}_2$-codimension two. We have indicated this fact in Table 5.1 by the $*$ next to the $\mathbf{Z}_2$-codimension 3.

(iii) We have included the regular codimension of those germs in Table 5.1 to emphasize the degree of complexity of these singularities should one break the $\mathbf{Z}_2$-symmetry.

In Table 5.2 we list the algebraic data associated with each of the normal forms in Table 5.1. Also in Table 5.3 we tabulate the solutions to the recognition problems for the various normal forms. The flow chart for the

Table 5.3. The Recognition Problem for Normal Forms of $\text{codim}_{\mathbf{Z}_2} \leq 3$.

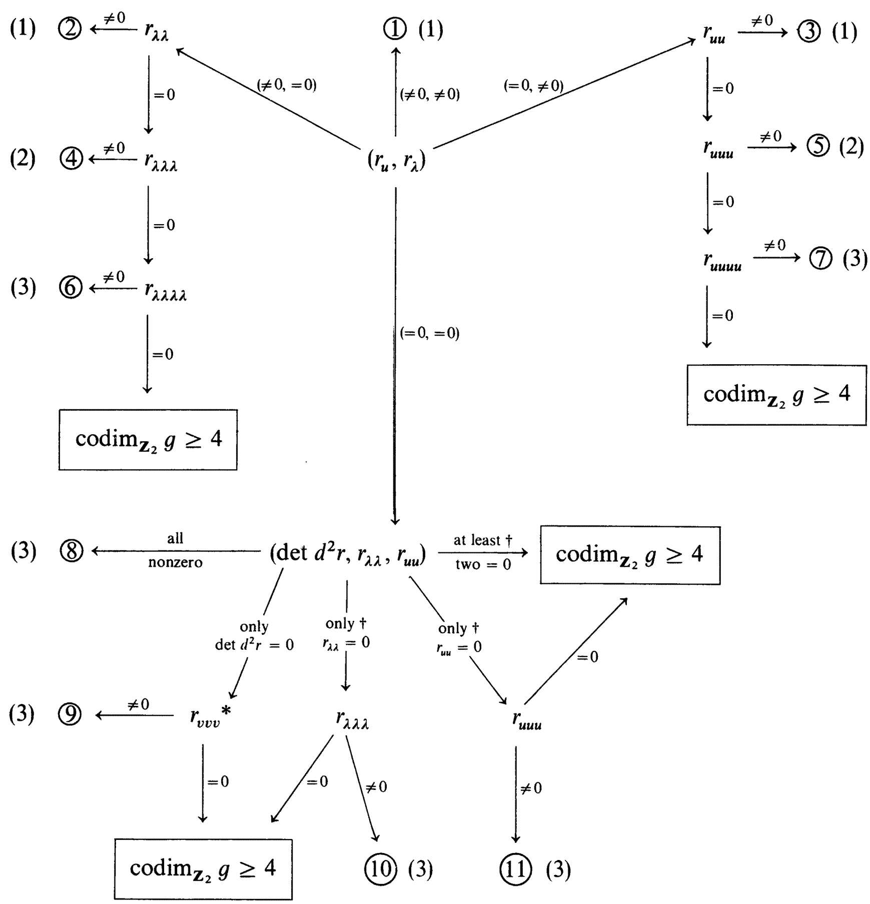
Normal Form	Defining Condition†	Nondegeneracy Conditions*
$(\varepsilon u^k + \delta\lambda)x$ $(k > 1)$	$\frac{\partial r}{\partial u} = \cdots = \left(\frac{\partial}{\partial u}\right)^{k-1} r = 0$	$\varepsilon = \text{sgn}\left(\left(\frac{\partial}{\partial u}\right)^k r\right)$ $\delta = \text{sgn}(r_\lambda)$
$(\varepsilon u + \delta\lambda^k)x$ $(k > 1)$	$\frac{\partial r}{\partial \lambda} = \cdots = \left(\frac{\partial}{\partial \lambda}\right)^{k-1} r = 0$	$\varepsilon = \text{sgn}(r_u)$ $\delta = \text{sgn}\left(\left(\frac{\partial}{\partial \lambda}\right)^k r\right)$
$(\varepsilon u^2 + 2m\lambda u + \delta\lambda^2)x$ $m^2 \neq \varepsilon\delta$	$r_u = r_\lambda = 0$	$\varepsilon = \text{sgn}(r_{uu})$ $\delta = \text{sgn}(r_{\lambda\lambda})$ $m^2 \neq \varepsilon\delta$ where $m = r_{u\lambda}/\sqrt{\lvert r_{uu} \cdot r_{\lambda\lambda}\rvert}$ *Note:* $m^2 \neq \varepsilon\delta \Leftrightarrow$ $\det d^2 r \neq 0$
$(\phi u^3 + \varepsilon u^2 + 2\delta\lambda u + \varepsilon\lambda^2)x$ where $\delta = \text{sgn}(m)$	$r_u = r_\lambda = 0$, $\det d^2 r = 0$ Choose $v \neq 0$ such that $r_{vv} = 0$	$\varepsilon = \text{sgn}(r_{uu})$, $\delta = \text{sgn}(r_{u\lambda})$ $\phi = \text{sgn}(r_{vvv})$
$(\varepsilon u^2 + 2\delta\lambda u + \phi\lambda^3)x$	$r_u = r_\lambda = r_{\lambda\lambda} = 0$	$\varepsilon = \text{sgn}(r_{uu})$, $\delta = \text{sgn}(r_{u\lambda})$, $\phi = \text{sgn}(r_{\lambda\lambda\lambda})$
$(\varepsilon u^3 + \delta\lambda u + \phi\lambda^2)x$	$r_\lambda = r_u = r_{uu} = 0$	$\varepsilon = \text{sgn}(r_{uuu})$, $\delta = \text{sgn}(r_{u\lambda})$, $\phi = \text{sgn}(r_{\lambda\lambda})$

* We use the convention $\text{sgn}(A) = \varepsilon$ means that $A \neq 0$.

† We assume $r = 0$ always.

proof of the classification theorem, Theorem 5.1, is given in Table 5.4. The flow chart is applicable to any $\mathbf{Z}_2$-symmetric germ $g(x, \lambda) = r(u, \lambda)x$ satisfying $r(0, 0) = 0$. To use the chart, compute r_u and r_λ and then follow the instructions starting with the pair (r_u, r_λ) in the upper center of the table. All details are left to the reader. They are similar in spirit to the calculations needed to prove the regular classification theorem, Theorem IV,2.1.

Table 5.4. Flow Chart for the Proof of Theorem 5.1.



* v is chosen as in Table 5.3 so that $r_{vv} = 0$.

† In these cases determining whether $\det d^2r = 0$ is equivalent to determining whether $r_{u\lambda} = 0$.

Circled numbers refer to normal forms in Table 5.1.

Numbers in parentheses indicate codimension.

Remark 5.3. From Table 5.4 we can see why normal forms (9), (10), and (11) correspond to distinguished members of the moduli family, normal form (8). In this family we look for "nondegenerate quadratics" satisfying $r_{uu} \neq 0$, $r_{\lambda\lambda} \neq 0$, $\det d^2r \neq 0$. For certain distinguished values of the second-order terms, precisely one of these conditions fails. At these values we must look at an appropriate third-order term to see that finite determinacy is maintained.

§6. Persistent Perturbations of the Nonmodal Bifurcations

There are six singularities of $\mathbf{Z}_2$-codimension one, two, or three which do not have modal parameters. (Normal forms (2)–(7) in Table 5.1.) They fit into two families: $\varepsilon x^{2k+1} + \delta\lambda x$ and $\varepsilon x^3 + \delta\lambda^k x$ with $k = 2, 3, 4$. The transition varieties and persistent perturbation, for each of these singularities is given in Figures 6.1–6.3. Analytic expressions for the transition varieties are given in Table 6.1. These expressions are derived using the formulas for the various components of the transition variety presented in Table 4.1.

To reduce the number of cases, we consider only those examples for which the trivial solution is stable when $\lambda < 0$. In addition, we do not enumerate cases $g(x, \lambda)$ which can be identified by the coordinates changes $\pm g(x, \pm\lambda)$.

Table 6.1. Transition Varieties for the Simple $\mathbf{Z}_2$-Singularities.

Universal Unfolding	$\mathscr{B}_0(\mathbf{Z}_2)$	$\mathscr{B}_1(\mathbf{Z}_2)$	$\mathscr{H}_0(\mathbf{Z}_2)$	$\mathscr{H}_1(\mathbf{Z}_2)$	$\mathscr{D}(\mathbf{Z}_2)$
$x^5 - \lambda x + \alpha x^3$	$\varnothing$	$\varnothing$	$\alpha = 0$	$\varnothing$	$\varnothing$
$\varepsilon x^3 + \lambda^2 x + \alpha x$	$\alpha = 0$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
$x^7 - \lambda x + \alpha x^3 + \beta x^5$	$\varnothing$	$\varnothing$	$\alpha = 0$	$\alpha = \beta^2/3$ $\beta \leq 0$	$\alpha = \beta^2/4$ $\beta \leq 0$
$x^3 - \lambda^3 x + (\alpha + \beta\lambda)x$	$(\beta/3)^3 = (\alpha/2)^2$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
$x^9 - \lambda x + \alpha x^3 + \beta x^5 + \gamma x^7$	$\varnothing$	$\varnothing$	$\alpha = 0$	$\alpha = 3\gamma u^2 + 8u^3$, $\beta = -3\gamma u - 6u^2$ $u > 0$	*
$\varepsilon x^3 + \lambda^4 x + (\alpha + \beta\lambda + \gamma\lambda^2)x$	$\alpha = \gamma\lambda^2 + 3\lambda^4$, $\beta = -2\gamma\lambda - 4\lambda^3$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

* Picture of $\mathscr{D}(\mathbf{Z}_2)$ in Figure 6.3 obtained geometrically.

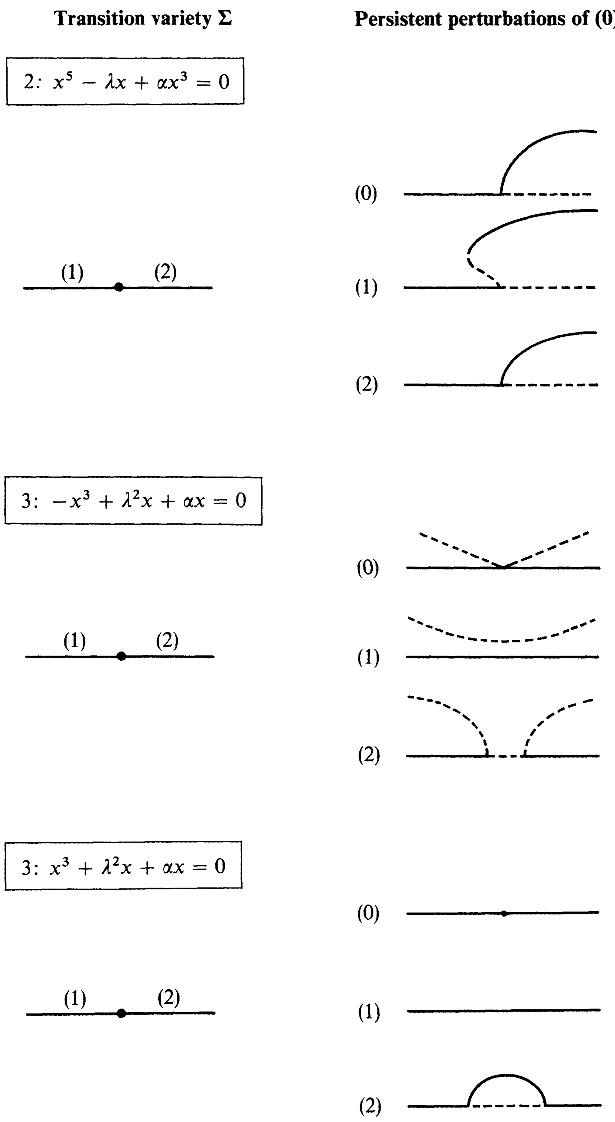
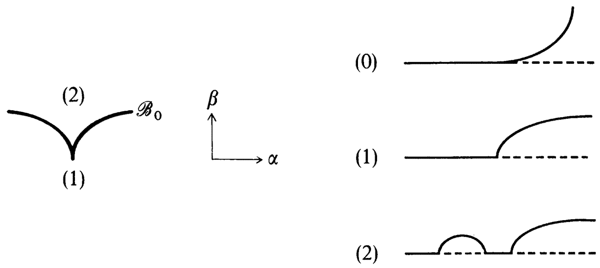


Figure 6.1. Persistent perturbations of $\mathbf{Z}_2$-codimension 1.

Transition variety Σ | **Persistent perturbations of (0)**

$$4\colon\ x^3 - \lambda^3 x + \alpha x + \beta\lambda x = 0.$$



$$5\colon\ x^7 - \lambda x + \alpha x^3 + \beta x^5 = 0$$

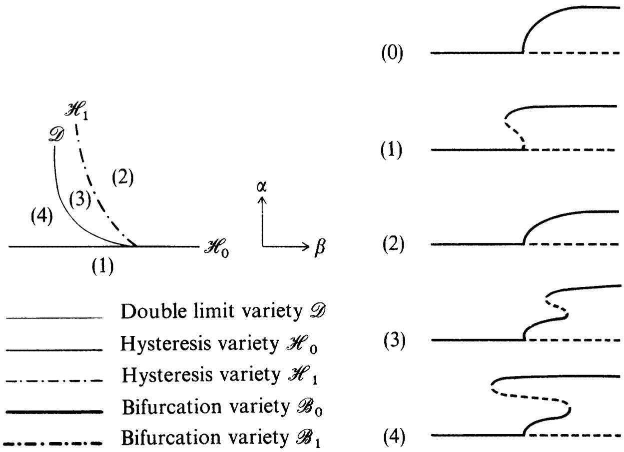


Figure 6.2. Persistent perturbations of $\mathbf{Z}_2$-codimension 2.

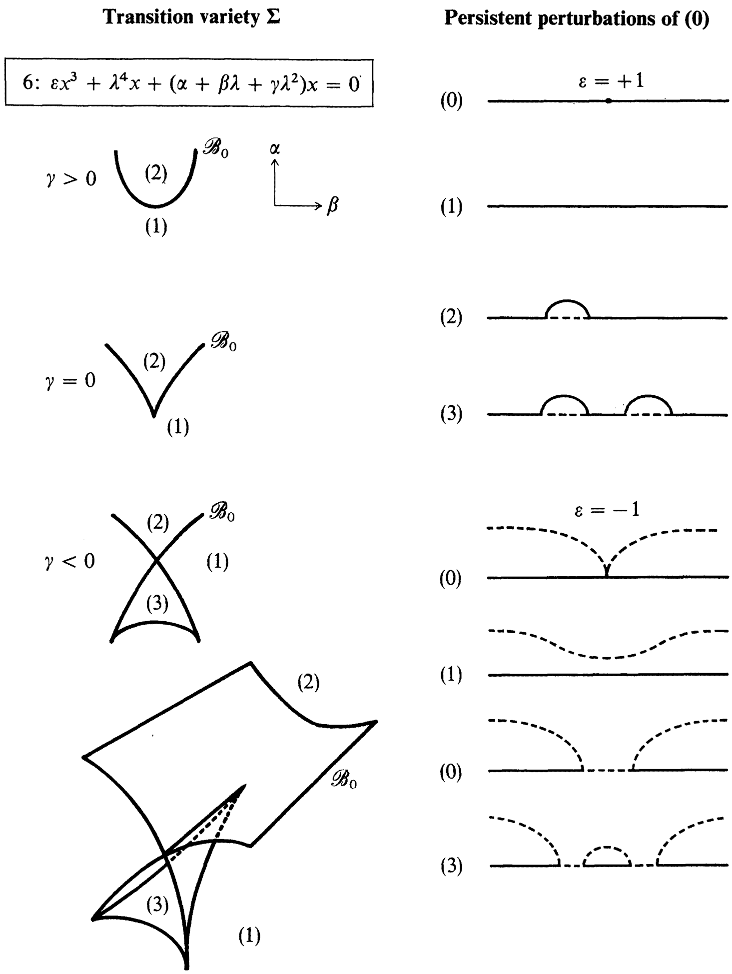


Figure 6.3. Persistent perturbations of $\mathbf{Z}_2$-codimension 3, modality zero.

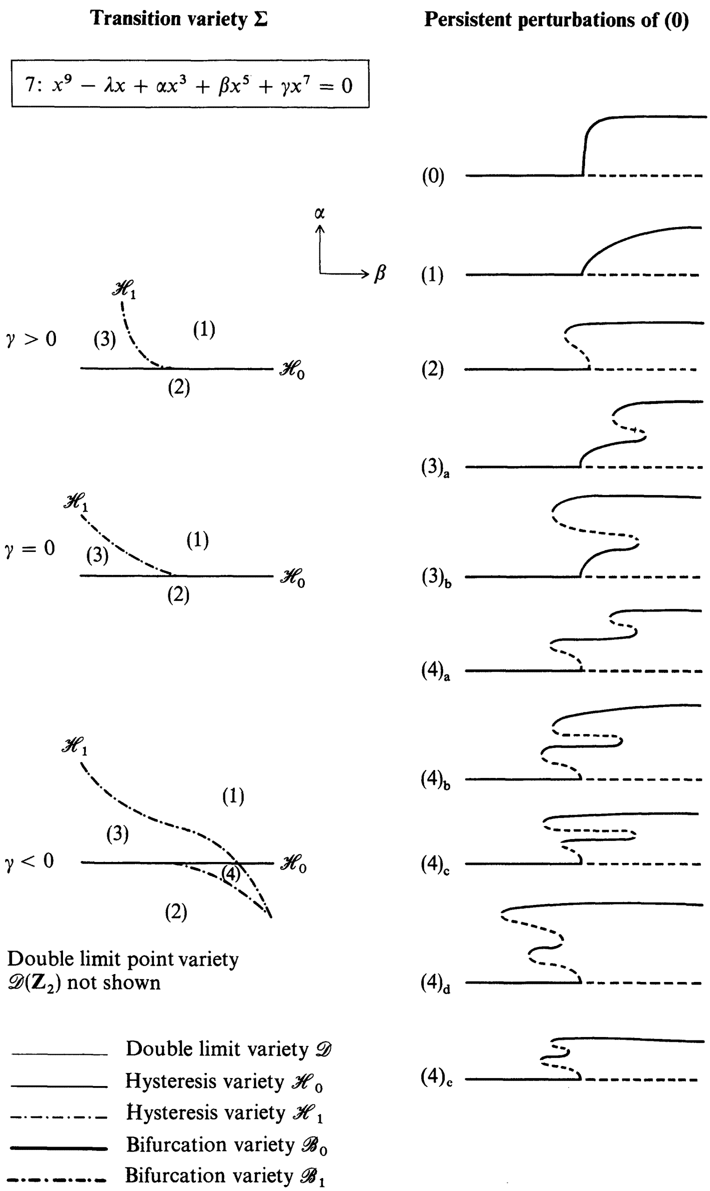


Figure 6.3 (*continued*)

§7. The Unimodal Family of Codimension Three

In this section we discuss the normal form (8) in Table 5.1, which we repeat here for convenient reference

$$h(x, \lambda) = \varepsilon x^5 + 2m\lambda x^3 + \delta\lambda^2 x, \tag{7.1}$$

where $\varepsilon = \pm 1$, $\delta = \pm 1$, and $m^2 \neq \varepsilon\delta$. We divide this section into three subsections, in which we establish the following points:

(a) For any value of m such that $m^2 \neq \varepsilon\delta$, the normal form (7.1) has codimension three. The parameter m in (7.1) is a modal parameter; thus one of the three parameters in a universal unfolding involves changes in m. This family of bifurcation problems may be viewed as a bracelet, analogous to the moduli family studied in Chapter V, §5. The family (7.1) has connector points where $m^2 = \varepsilon\delta$ or $m = \pm\infty$. The associated connector complexes give rise to the normal forms (9)–(11) of Table 5.1.

(b) There are ordinary distinguished points when $m = 0$. We determine this fact by analyzing the structure of the persistent perturbations of (7.1) at nondistinguished points and showing that there is a change in topological type of the $\mathbf{Z}_2$-universal unfolding at $m = 0$.

(c) The solution set of (7.1) near the values of m associated with connector points (i.e., $m^2 = \varepsilon\delta$ and $m = \pm\infty$) indicates some elementary global properties of the bifurcation diagram. (In subsection (c), we consider only the unperturbed bifurcation diagrams; the persistent perturbations in the connector complexes will be presented in §8 below. In terms of *content* the material in subsection (c) would fit more naturally into §8. However, we have divided the material on the basis of *level of difficulty*—the calculations of §8 are exceedingly technical, while the present section, including subsection (c), is relatively elementary.)

We have two reasons for discussing this moduli family in detail. First, since the singularity has codimension three, the transition variety $\Sigma(\mathbf{Z}_2)$ is a subset of $\mathbb{R}^3$; thus pictures of $\Sigma(\mathbf{Z}_2)$ should help the reader understand how topological triviality fails at distinguished points. (Cf. subsection (b) below.) Second, this moduli family arises in our discussion of the clamped Hodgkin–Huxley equations in Case Study 2.

(a) The Moduli Bracelet

We begin our discussion by recalling how the modal parameter m in (7.1) is defined. Consider the $\mathbf{Z}_2$-bifurcation problem

$$g(x, \lambda) = (Au^2 + Bu\lambda + C\lambda^2 + \cdots)x, \tag{7.2}$$

where the higher-order terms in (7.2) are in $\mathscr{M}^3_{u,\lambda}$. We see from Table 5.4 that if $A \neq 0$, $C \neq 0$ and $B^2 - 4AC \neq 0$, then (7.2) is $\mathbf{Z}_2$-equivalent to the normal form (7.1). Moreover, one obtains the form (7.1) by scaling A and C to be ± 1. This leads to the formulas

$$\begin{aligned} &\text{(a)} \quad \varepsilon = \operatorname{sgn}(A), \qquad \delta = \operatorname{sgn}(C), \\ &\text{(b)} \quad m = B/(2\sqrt{|AC|}). \end{aligned} \tag{7.3}$$

In the scaled variables the nondegeneracy condition $B^2 - 4AC \neq 0$ becomes $m^2 \neq \varepsilon\delta$.

Alternatively, one can scale A, B, C so that

$$|A| = 1, \qquad B^4 + C^2 = 1.$$

Since $B^4 + C^2 = 1$ is (topologically) a circle we see that the family (7.1) may be thought of as two circles, one corresponding to $\varepsilon = +1$ and one corresponding to $\varepsilon = -1$. These circles are drawn in Figure 7.1. In that figure $\delta = +1$ represents the upper semicircles and $\delta = -1$ represents the lower semicircles. Also, in anticipation of results below, we have made the circles into a bracelet by joining the points at $m = +\infty$ and $m = -\infty$.

In order to verify the information in Figure 7.1 we will now show that the points where $m^2 = \varepsilon\delta$ or $m = \pm\infty$ are connector points. (Recall that a connector point is a point on the moduli family corresponding to (at least) two distinct singularities.) First we consider the case $m^2 = \varepsilon\delta$. Referring to Table 5.4, we see that $m^2 = \varepsilon\delta$ is equivalent to det $d^2r = 0$ in that table. It follows that generically (i.e., when $r_{vvv} \neq 0$) the addition of higher-order terms leads to the normal form (Table 5.1(9))

$$\phi x^7 + \varepsilon x^5 + 2\sigma\lambda x^3 + \varepsilon\lambda^2 x, \tag{7.4}$$

where $m = \sigma = \pm 1$ and $\phi = \pm 1$. (Note that when $m^2 = \varepsilon\delta$ we must have $\varepsilon = \delta$, as indicated in (7.4).) Now for each choice of ε and σ there are *two* singularities listed in (7.4) with $\mathbf{Z}_2$-codimension equal to three; namely, $\phi = +1$ and $\phi = -1$. Thus, points where $m^2 = \varepsilon\delta$ are connector points.

We next prove that the points where $m = \pm\infty$ are connector points. We see from (7.3b) that m may approach $\pm\infty$ in two distinct ways; either

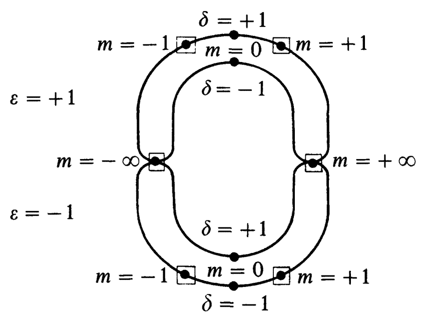


Figure 7.1. The moduli bracelet in $\mathbf{Z}_2$-codimension three.

$A \to 0$ or $C \to 0$. From Tables 5.4 and 5.1, we see that these two routes to $m = \pm\infty$ lead (generically) to the singularities

$$\begin{aligned} &\text{(a)} \quad \phi x^7 + \sigma\lambda x^3 + \delta\lambda^2 x, \\ &\text{(b)} \quad \varepsilon x^5 + \sigma\lambda x^3 + \phi\lambda^3 x. \end{aligned} \tag{7.5}$$

We show below that in (7.5) $\sigma = +1$ corresponds to $m = +\infty$ and $\sigma = -1$ corresponds to $m = -\infty$. Thus, (7.5a) and (7.5b) each have two singularities ($\phi = \pm 1$) corresponding to a point on the moduli bracelet where $m = \pm\infty$.

To establish the correspondence between $\sigma = \pm 1$ in (7.5) and $m = \pm\infty$ in (7.1), we consider the following one-parameter unfoldings of (7.5):

$$\begin{aligned} &\text{(a)} \quad \phi x^7 + a x^5 + \sigma\lambda x^3 + \delta\lambda^2 x, \\ &\text{(b)} \quad \varepsilon x^5 + \sigma\lambda x^3 + b\lambda^2 x + \phi\lambda^3 x. \end{aligned} \tag{7.6}$$

It may be seen from (7.3b) that for $a \neq 0$, (7.6a) is equivalent to (7.1) with $2m = \sigma/\sqrt{|a|}$; similarly, for (7.6b), with $2m = \sigma/\sqrt{|b|}$. Letting $a \to 0$ and $b \to 0$, we deduce that $\sigma = +1$ in (7.5) corresponds to $m = +\infty$ in (7.1) and $\sigma = -1$ to $m = -\infty$, as claimed.

Moreover, we can compute from (7.3a) the signs ε and δ in the normal form (7.1) when $a \neq 0$ and $b \neq 0$. Specifically we obtain $\varepsilon = \operatorname{sgn}(a)$ in (7.6a) and $\delta = \operatorname{sgn}(b)$ in (7.6b). Thus for example, if $\sigma = +1$ and $\delta = +1$, the singularities in (7.5a) connect the branch $\varepsilon = +1, \delta = +1, 1 < m < \infty$ with the branch $\varepsilon = -1, \delta = +1, 0 < m < \infty$. This example shows why we have identified the points $m = +\infty$ on the two moduli circles $\varepsilon = +1$ and $\varepsilon = -1$. Similarly, for $m = -\infty$. Enumerating all possible choices of signs in (7.6a, b) leads to the connector complex pictured in Figure 7.2.

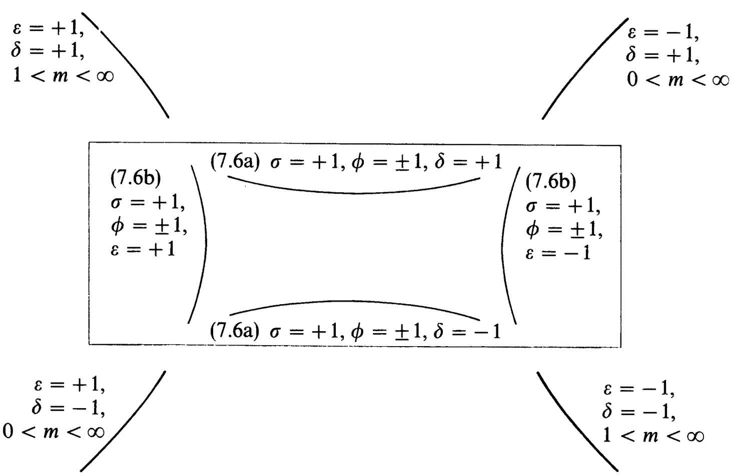


Figure 7.2. Connector complex at $m = +\infty$. The connector complex at $m = -\infty$ is obtained by replacing $\sigma = +1$ with $\sigma = -1$.

To complete our description of the moduli bracelet in Figure 7.1 we need to show that points on the moduli circle where $m = 0$ are distinguished and that all other points are nondistinguished. This information will be derived in the next subsection.

(b) Persistent Perturbations of Nonconnector Points

In this subsection we picture the transition set Σ for the $\mathbf{Z}_2$-universal unfolding of the singularities in (7.1); namely,

$$H(x, \lambda, \alpha, \beta, m) = \varepsilon x^5 + 2m\lambda x^3 + \delta\lambda^2 x + \alpha x + 2\beta x^3. \tag{7.7}$$

(Cf. Figures 7.3 and 7.4.) It will be apparent from these pictures that Σ is topologically trivial when $m \neq 0$ and that the topological type of Σ changes at $m = 0$.

We present here only the results of calculations, some of which are lengthy. We hope that the interested reader will be able to reproduce these calculations, if desired.

The formulas for the various components of Σ are:

$$\begin{aligned}
&\text{(a)}\quad \mathscr{H}_0\colon m^2\alpha = -\delta\beta^2, \quad \delta\alpha \le 0,\\
&\text{(b)}\quad \mathscr{B}_0\colon \alpha = 0,\\
&\text{(c)}\quad \mathscr{B}_1\colon \alpha = -\delta\beta^2/(m^2 - \varepsilon\delta), \quad \operatorname{sgn}\beta = \delta\operatorname{sgn}(m^2 - \varepsilon\delta),\\
&\text{(d)}\quad \mathscr{H}_1 = \mathscr{D} = \varnothing.
\end{aligned} \tag{7.8}$$

In Figures 7.3 and 7.4 we consider the cases $\varepsilon = +1$, $\delta = -1$ and $\varepsilon = +1$, $\delta = +1$, respectively. In these figures we first picture the transition variety Σ for m fixed and nonzero; then we picture Σ on a neighborhood of $m = 0$. We also picture the persistent perturbations in (7.7). It follows from these pictures that $m = 0$ is a regular distinguished point.

(c) Global Properties of Bifurcation Diagrams Near Connector Points

Above we have shown that there are two types of connector points in the modal family (7.1), corresponding to the normal forms (7.4) and (7.5). (Normal form (7.4) is associated to $m^2 = \varepsilon\delta$, and (7.5), to $m = \pm\infty$.) We also know that in the universal unfoldings at all of these normal forms, the codimension constant variety is a line parametrized by the modal parameter. In this subsection, we study the bifurcation diagrams occurring along this line. We do this to illustrate the global properties in the bifurcation diagrams which are associated with connector points. In the next section we will describe the persistent perturbations of each of these singularities.

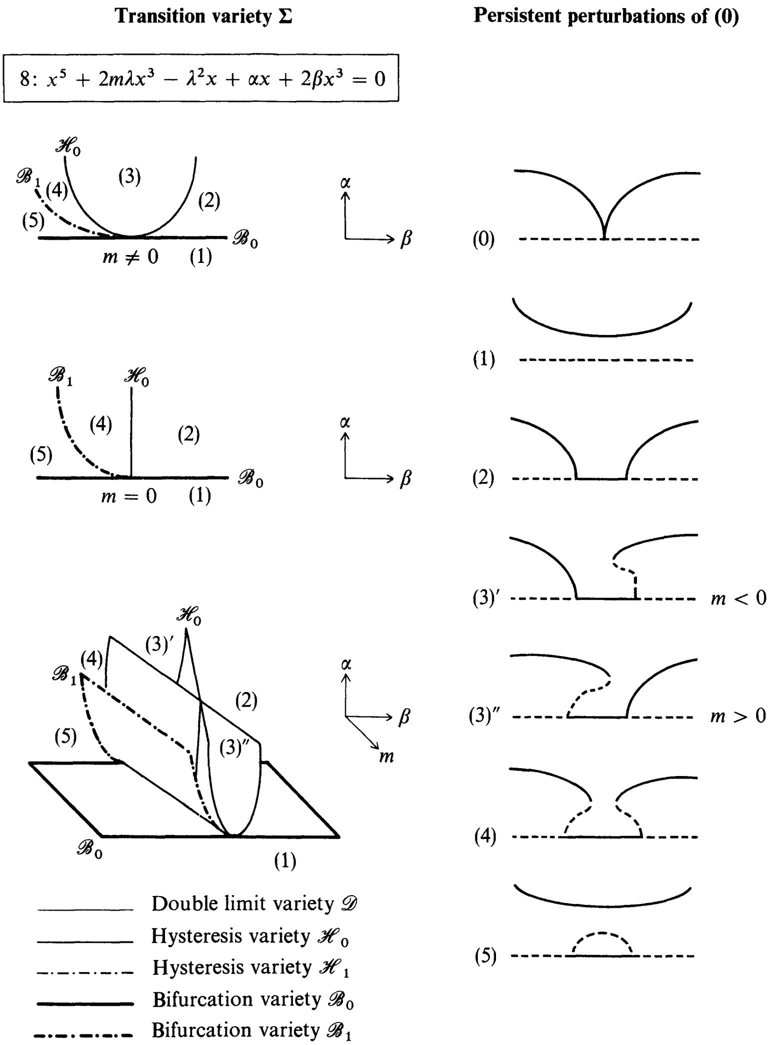


Figure 7.3. Persistent perturbations of (7.2a).

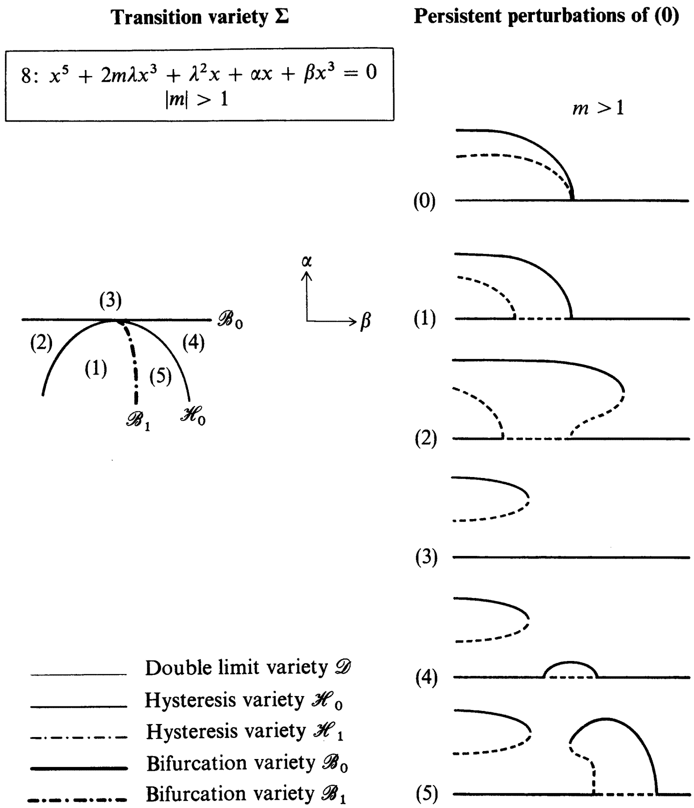


Figure 7.4. Persistent perturbations of (7.2b).

Transition variety Σ **Persistent perturbations of (0)**

$$8:\ x^5 + 2m\lambda x^3 + \lambda^2 x + \alpha x + \beta x^3 = 0$$
$$|m| < 1$$

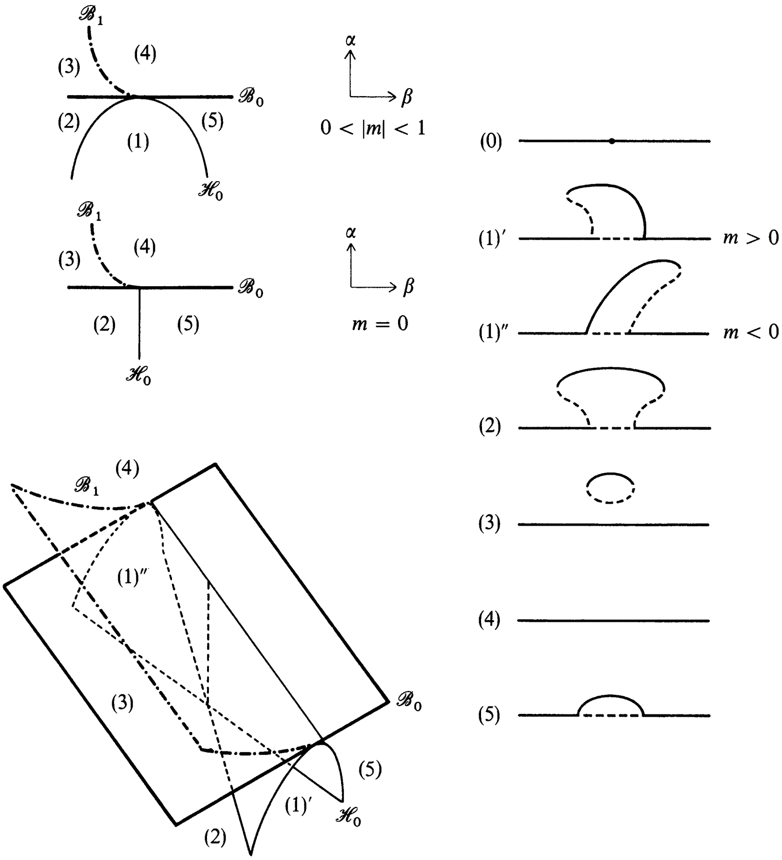


Figure 7.4 (*continued*)

To reduce the number of cases, we shall not consider all possible signs in (7.4) and (7.5). Specifically, we shall analyze only the normal forms

$$\begin{aligned}
&\text{(a)}\quad \phi x^7 + x^5 + 2m\lambda x^3 + \lambda^2 x,\\
&\text{(b)}\quad \phi x^7 + \lambda x^3 + \lambda^2 x + a x^5,\\
&\text{(c)}\quad x^5 + \sigma\lambda x^3 - \lambda^3 x - b\lambda^2 x.
\end{aligned} \tag{7.9}$$

Here $\phi = \pm 1$ and $\sigma = \pm 1$, while m, a, and b are modal parameters with $m \approx 1$, $a \approx 0$, and $b \approx 0$. Formula (7.9a) derives from (7.4) by choosing $\varepsilon = +1$ and allowing the modal parameter to vary near the distinguished value $m = 1$; similarly, (7.9b, c) derive from the one-parameter unfoldings (7.5a, b). The remaining cases in (7.4) and (7.5) may be obtained from (7.9) by using the orientation reversing coordinate changes $\lambda \to -\lambda$ and $g(x, \lambda) \to -g(x, \lambda)$.

The bifurcation diagrams associated to the normal forms in (7.9) are illustrated in Figure 7.5.

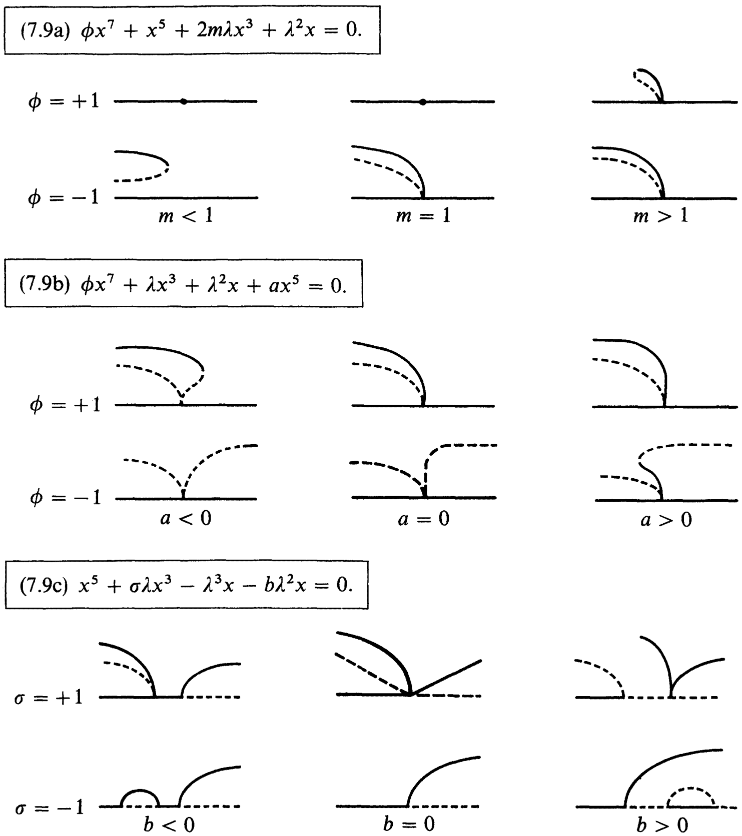


Figure 7.5. Global implications of connector points.

§8. Perturbations at the Connector Points

In the two subsections below, we describe the persistent perturbations of the singularities of (7.1) at the connector points where $m^2 = \varepsilon\delta$ and $m = \pm\infty$, respectively.

(a) The Persistent Perturbations when $m^2 = \varepsilon\delta$

The normal form for singularities where $m^2 = \varepsilon\delta$ is:

$$\phi x^7 + \varepsilon x^5 + 2\sigma\lambda x^3 + \varepsilon\lambda^2 x, \tag{8.1}$$

where $\phi = \pm 1$, $\varepsilon = \pm 1$, $\sigma = \pm 1$. (Cf. (7.4).) In order to reduce redundancy in our calculations we assume (as in §7(c)) that $\sigma = +1$ and $\varepsilon = +1$. Observe that the change of coordinates $\lambda \to -\lambda$ changes the sign of σ in (8.1); thus, the bifurcation diagrams for $\sigma = -1$ may be obtained by reading the diagrams below (for $\sigma = +1$) from right to left. The choice $\varepsilon = +1$ implies that the stationary solution is asymptotically stable (when $\lambda \neq 0$).

A universal unfolding of (8.1) is given by

$$H(x, \lambda, m, \alpha, \beta) = \phi x^7 + x^5 + 2m\lambda x^3 + \lambda^2 x + \alpha x + 2\beta x^3, \tag{8.2}$$

where $\phi = \pm 1$, $m \approx 1$, $\alpha \approx 0$, and $\beta \approx 0$. The various components of the transition set for the unfolding (8.2) are described by the following equations

$$\begin{aligned} &\text{(a)} \quad \mathscr{B}_0: \alpha = 0,\\ &\text{(b)} \quad \mathscr{H}_0: \alpha = -(\beta/2m)^2,\\ &\text{(c)} \quad \mathscr{B}_1: \alpha = -(m^2 - 1)u^2 + 2\phi u^3; \quad \beta = 2(m^2 - 1)u - 3\phi u^2,\\ &\text{(d)} \quad \mathscr{H}_1 = \mathscr{D} = \varnothing. \end{aligned} \tag{8.3}$$

Pictures of this transition set are given in Figure 8.1. This figure includes both two-dimensional cross-sections of the transition set obtained by fixing m and a full three-dimensional sketch. In addition, the perturbed bifurcation diagrams are given on this figure.

(b) The Persistent Perturbations when $m = \pm\infty$

We recall from (7.5) that there are two sets of normal forms corresponding to $m = \pm\infty$; they are, in slightly altered notation,

$$\begin{aligned} &\text{(a)} \quad \phi x^7 + 2\sigma\lambda x^3 + \delta\lambda^2 x,\\ &\text{(b)} \quad \phi x^5 + 2\sigma\lambda x^3 - \sigma\lambda^3 x, \end{aligned} \tag{8.4}$$

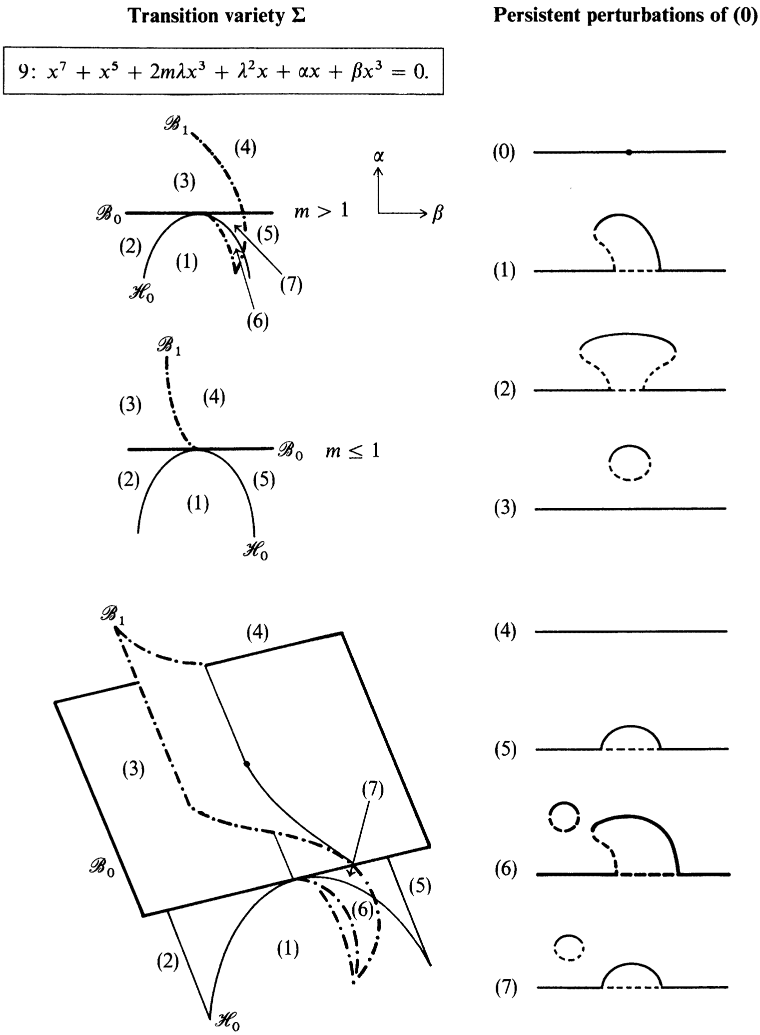
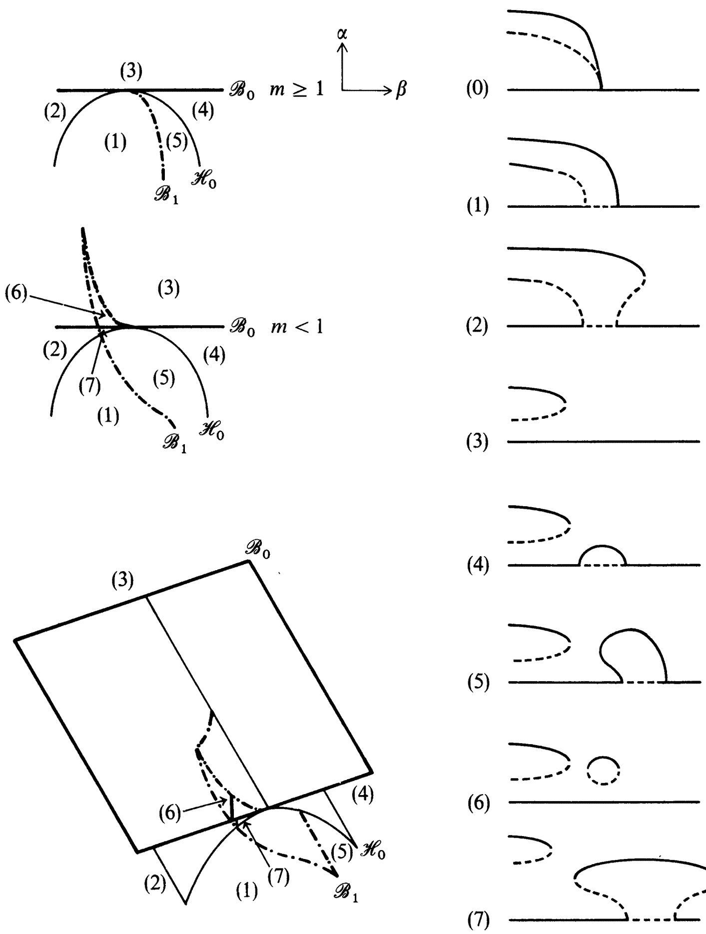


Figure 8.1. Transition set and persistent perturbations for (8.2).

Transition variety Σ **Persistent perturbations of (0)**

$9\colon\ -x^7 + x^5 + 2m\lambda x^3 + \lambda^2 x + \alpha x + \beta x^3 = 0.$



Double limit variety $\mathscr{D}$
Hysteresis variety $\mathscr{H}_0$
Hysteresis variety $\mathscr{H}_1$
Bifurcation variety $\mathscr{B}_0$
Bifurcation variety $\mathscr{B}_1$

Figure 8.1 (*continued*)

$$11: x^7 + \lambda x^3 + \lambda^2 x + \alpha x + \beta x^3 + a x^5 = 0.$$

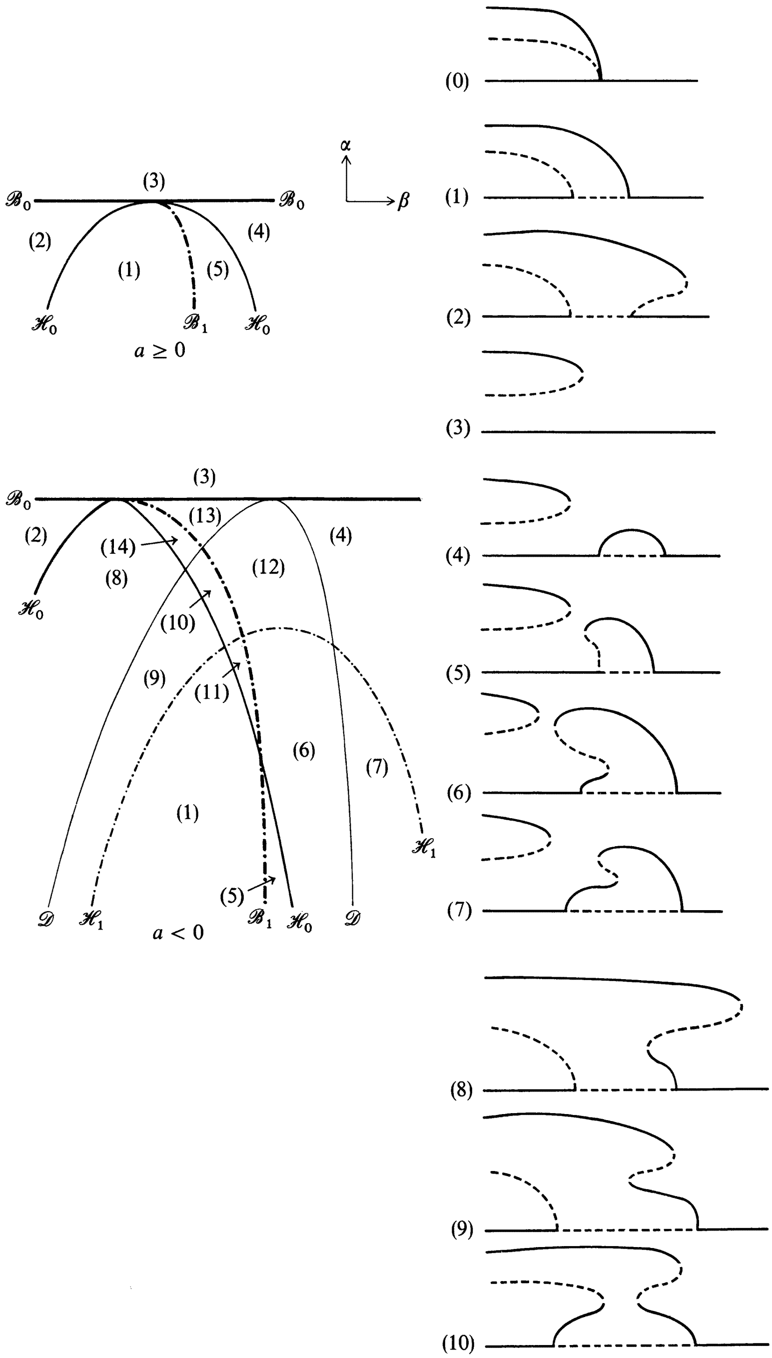
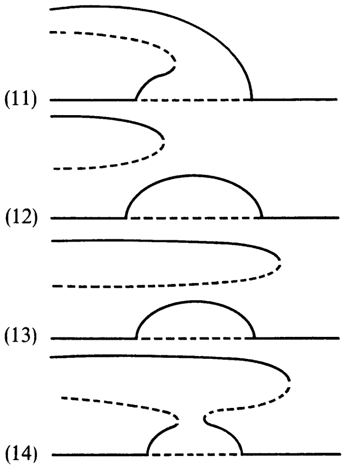
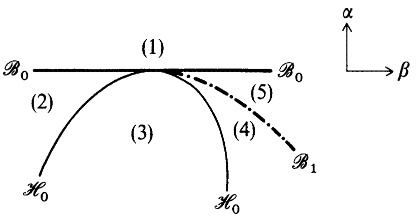
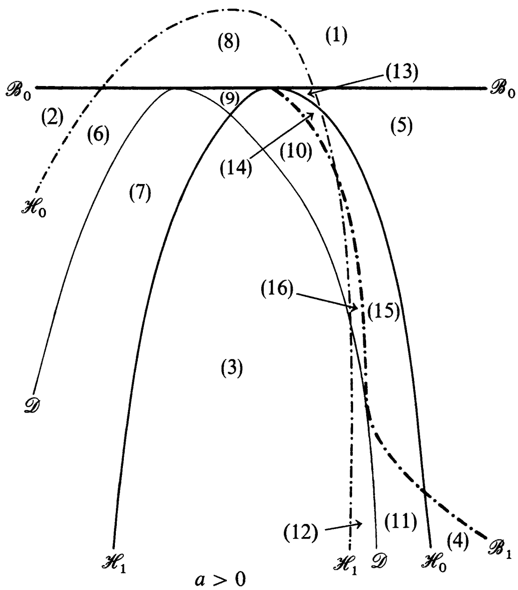


Figure 8.2. Persistent perturbations of (8.4a).



$$11: \; -x^7 + \lambda x^3 + \lambda^2 x + \alpha x + \beta x^3 + a x^5 = 0.$$





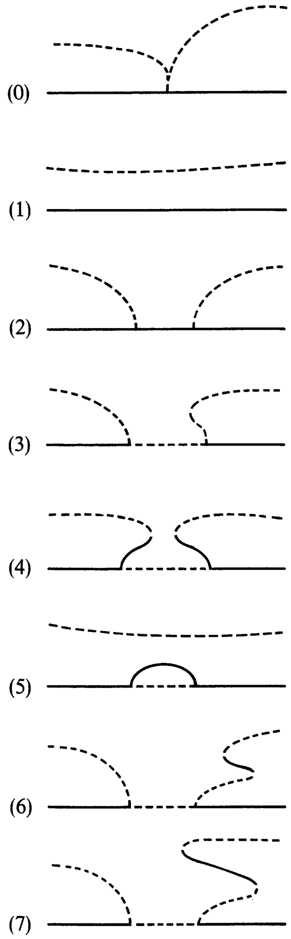


Figure 8.2 (*continued*)

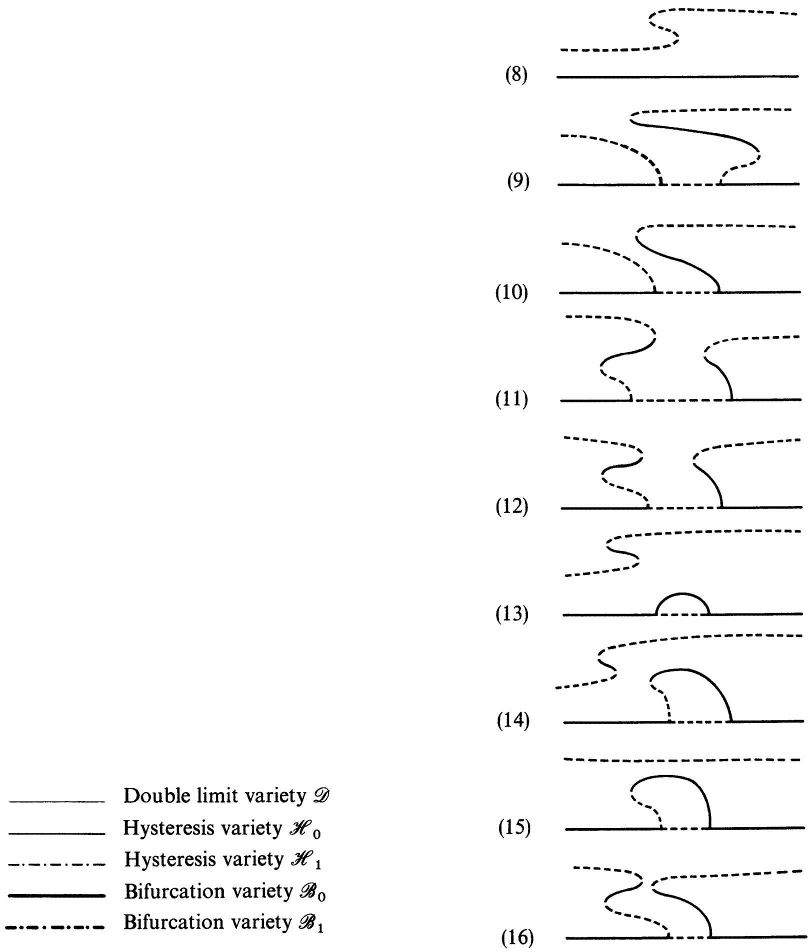


Figure 8.2 (*continued*)

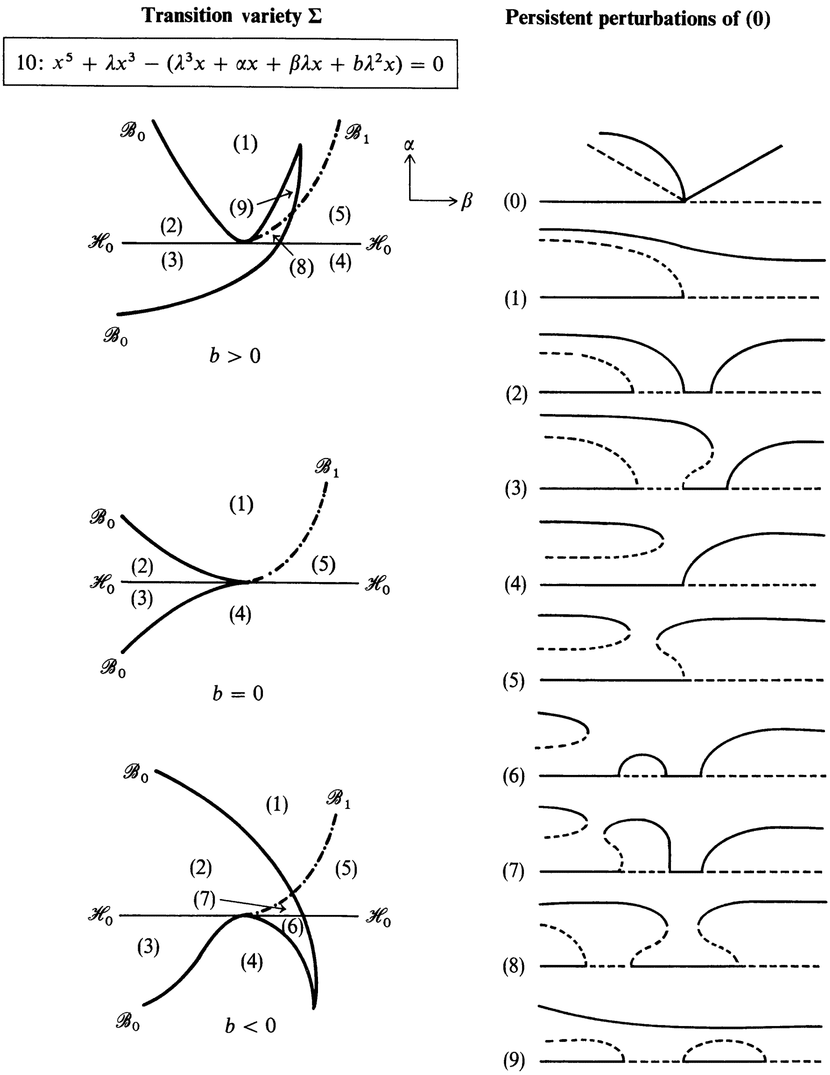


Figure 8.3. Persistent perturbations of (8.4b).

$10: x^5 - \lambda x^3 - (\lambda^3 x + \alpha x + \beta\lambda x + b\lambda^2 x) = 0$

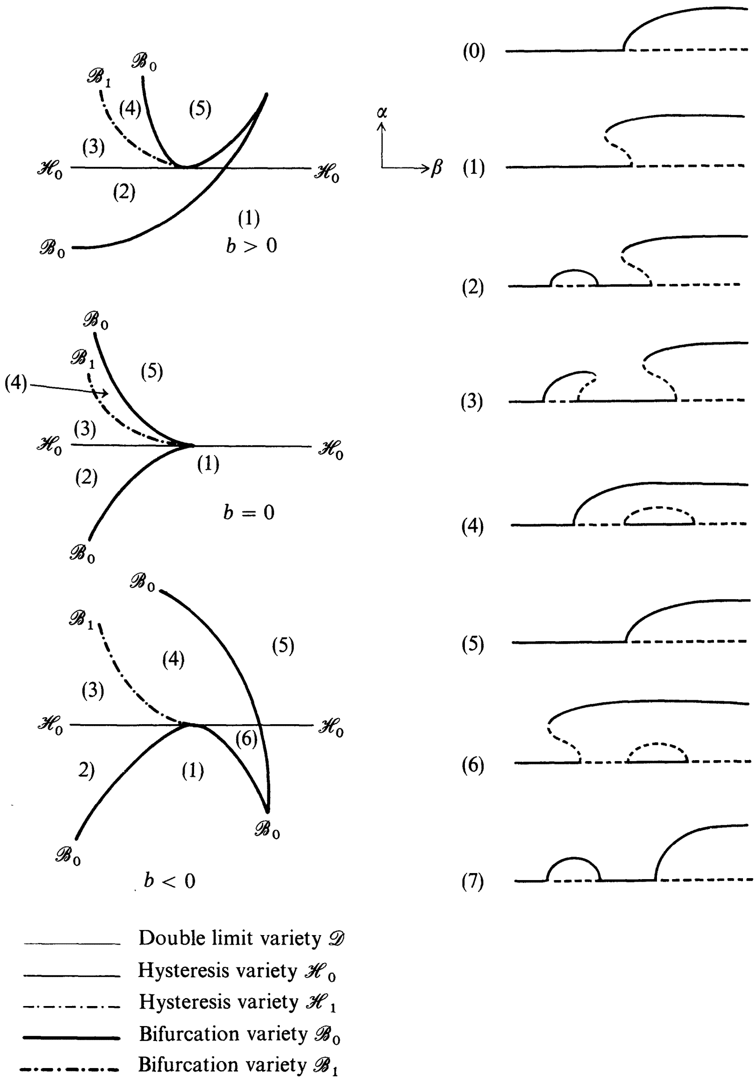


Figure 8.3 (*continued*)

Table 8.1. Equations for Transition Varieties of (8.5).

	H	K
$\mathscr{B}_0(\mathbf{Z}_2)$:	$\alpha = 0$	$\alpha = b\lambda^2 + 2\lambda^3$ $\beta = -2b\lambda - 3\lambda^2$
$\mathscr{B}_1(\mathbf{Z}_2)$:	$\alpha = (a - \frac{1}{4})u^2 + 2\phi u^3, u > 0,$ $\beta = -2(a - \frac{1}{4})u - 3\phi u^2$	$\alpha = (\frac{1}{4} + b)\lambda^2 + 2\lambda^3, \operatorname{sgn}(\lambda) = -\sigma,$ $\beta = -2(\frac{1}{4} + b)\lambda - 3\lambda^2, \operatorname{sgn} \beta = \sigma$
$\mathscr{H}_0(\mathbf{Z}_2)$:	$\alpha = -\beta^2$	$\alpha = 0$
$\mathscr{H}_1(\mathbf{Z}_2)$:	$\alpha - \dfrac{a^3}{27} = -\left(\beta - \dfrac{\phi}{3}a^2\right)^2, \operatorname{sgn}(a) = -\phi$	$\varnothing$
$\mathscr{D}(\mathbf{Z}_2)$:	$\alpha = -\left(\beta - \dfrac{\phi a^2}{4}\right)^2, \operatorname{sgn}(a) = -\phi$	$\varnothing$

where $\phi = \pm 1$, $\sigma = \pm 1$, and $\delta = \pm 1$. In order to reduce the number of cases, we suppose that $\sigma = +1$ and $\delta = +1$. (Recall that the choice $\sigma = +1$ corresponds to $m = +\infty$.) The other cases in (7.5) may be obtained from (8.4) using the transformations $\lambda \to -\lambda$ and $g(x, \lambda) \to -g(x, \lambda)$.

The universal unfoldings we analyze are:

$$\begin{aligned}
&\text{(a)} \quad H(x, \lambda, a, \alpha, \beta) = \phi x^7 + \lambda x^3 + \lambda^2 x + \alpha x + \beta x^3 + a x^5, \\
&\text{(b)} \quad K(x, \lambda, b, \alpha, \beta) = x^5 + \sigma\lambda x^3 - \lambda^3 x - \alpha x - \beta\lambda x - b\lambda^2 x,
\end{aligned} \tag{8.5}$$

where $\phi = \pm 1$, $\sigma = \pm 1$, and a, b, α, β are near zero. For each of the unfoldings (8.5), equations for the various components of transition set are given in Table 8.1. Pictures of the transition set and the persistent perturbations are given in Figures 8.2 and 8.3.

Bibliographical Comments

Although many authors have recognized the importance of symmetry in bifurcation, for us the pioneering work of Sattinger [1978, 1979, 1983] was most important. See also Thompson and Hunt [1973, 1977], Michel [1972], Marsden and McCracken [1976], Golubitsky and Schaeffer [1979b], Dancer [1980] and Vanderbauwhede [1982]. The role of symmetry in bifurcation is still an active area of research; in particular, much of Volume II will be devoted to this topic.

CHAPTER VII

The Liapunov–Schmidt Reduction

§0. Introduction

In Chapter I, §3 we described the Liapunov–Schmidt reduction in rather special circumstances. In this chapter we generalize the method in three distinct ways, as follows:

(i) We consider infinite-dimensional systems.
(ii) We allow the linearized operator to have a multidimensional kernel.
(iii) We perform the reduction when the operator commutes with a compact group of symmetries.

This chapter is divided into six sections. The above three extensions of Chapter I, §3 are discussed theoretically in §§1 and 3; the first two extensions in §1, the third in §3. The remaining four sections illustrate the use of the method in applications. Specifically, we analyze the elastica (a classical buckling model) in §2 and reaction–diffusion equations in §§4–6. The last three sections break down as follows: §4, scalar equations; §5, general description of the *Brusselator* (a specific equation intended to model the Belusov–Zhabotinsky reaction); §6, Liapunov–Schmidt reduction of the Brusselator. Sections 2 and 4 illustrate the theory of §1; §6 illustrates the theory of both §§1 and 3. (Chapter VIII, on the Hopf bifurcation, and Case Study 3 contain genuine applications, as opposed to illustrations, of the theory of §3.)

Incidentally, §§4–6 contain some rather interesting phenomena concerning reaction–diffusion equations; specifically how diffusion can affect the stability of an equilibrium solution of an ODE and how rudimentary patterns can form from the competition between the reaction and the diffusion terms.

In our discussion of the Liapunov–Schmidt reduction in the finite-dimensional case, we showed that the (linear) stability or instability of bifurcating solutions was determined by the sign of g_x, the derivative of the reduced equation. Similar results are available in many infinite-dimensional settings. However, we do not discuss these here, as we believe the prerequisites from functional analysis are too high for the goals of this text.

§1. The Liapunov–Schmidt Reduction Without Symmetry

In this section, we treat the infinite-dimensional Liapunov–Schmidt reduction at a multiple eigenvalue with auxiliary parameters; i.e., we consider generalizations (i) and (ii) above. We base our discussion on the five step summary of the Liapunov–Schmidt reduction given in Chapter I, §3(b). The principal difficulty in extending the method occurs in Step 1; i.e., forming complements in an infinite-dimensional space. In this connection we introduce some preliminary concepts.

(a) Fredholm Operators of Index Zero

Definition 1.1. Let $\mathscr{X}$ and $\mathscr{Y}$ be Banach spaces. A bounded linear operator $L: \mathscr{X} \to \mathscr{Y}$ is called *Fredholm* if the following two conditions hold.

(i) Ker L is a finite-dimensional subspace of $\mathscr{X}$.
(ii) Range L is a closed subspace of $\mathscr{Y}$ of finite codimension.

Definition 1.2. If L is Fredholm, the *index* of L is the integer

$$i(L) = \dim \ker L - \operatorname{codim} \operatorname{range} L.$$

The following result contains the main information we will need concerning Fredholm operators.

Proposition 1.3. *If $L: \mathscr{X} \to \mathscr{Y}$ is Fredholm, then there exist closed subspaces M and N of $\mathscr{X}$ and $\mathscr{Y}$, respectively, such that*

$$\text{(a)} \quad \mathscr{X} = \ker L \oplus M, \qquad \text{(b)} \quad \mathscr{Y} = N \oplus \operatorname{range} L. \tag{1.1}$$

This proposition is proved in Berger [1977], §1.3F.

Remark 1.4. In all applications in this book L will be Fredholm with index zero. For such operators, in (1.1b) we have dim ker $L =$ dim N. In particular, if ker $L = \{0\}$, then L is onto and hence, by the closed graph theorem, invertible. Thus, we have the following implication for Fredholm operators of index zero:

$$\text{If } \ker L = \{0\}, \quad \text{then } L \text{ is invertible.}$$

For differential operators, $\mathscr{X}$ and $\mathscr{Y}$ typically are subspaces of the Hilbert space $L^2(\Omega)$, where Ω is a bounded domain in $\mathbb{R}^N$. This space has the standard inner product

$$\langle u, v \rangle = \int_\Omega u(\xi)v(\xi)\, d\xi. \tag{1.2}$$

Let us discuss the use of orthogonal complements in (1.1); i.e., setting

$$\begin{aligned} &\text{(a)} \quad M = (\ker L)^\perp, \\ &\text{(b)} \quad N = (\text{range } L)^\perp, \end{aligned} \tag{1.3}$$

where for a subspace $S \subset \mathscr{Y}$ we define

$$S^\perp = \{u \in \mathscr{Y}: \langle u, v \rangle = 0 \text{ for all } v \in S\}.$$

Usually $\mathscr{X}$ and $\mathscr{Y}$ are not complete with respect to the inner product (1.2). For example, $\mathscr{X}$ might be $C^k(\Omega)$ and $\mathscr{Y}$ might be $C(\Omega)$; i.e., spaces of differentiable and continuous functions, respectively. In general, for an infinite-dimensional subspace $S \subset \mathscr{Y}$, it is not true that $\mathscr{Y} = S \oplus S^\perp$. Although $S \cap S^\perp = \{0\}$, the sum need not equal $\mathscr{Y}$; i.e., there may be too few elements in $S^\perp$. Speaking heuristically, the problem in such cases is that the missing elements lie in the dual space $\mathscr{Y}^*$, rather than $\mathscr{Y}$ itself, as required by our definition of $S^\perp$; this is caused by a mismatch between (1.2) and the natural norm on $\mathscr{Y}$. However, the decomposition $\mathscr{Y} = S \oplus S^\perp$ is valid in the following two special cases, which justify (1.3):

Case (a). S is finite dimensional.

Case (b). S is the range of an elliptic differential operator.

In Case (a), when $\dim S < \infty$, we may derive the decomposition by the Gram–Schmidt orthogonalization process. Let us summarize the issues concerning Case (b). (Cf. Appendix 4.) The discussion revolves around the *Fredholm alternative*,

$$(\text{range } L)^\perp = \ker L^*, \tag{1.4}$$

where L^* is the adjoint of L. Formula (1.4) is generally valid for linear operators, *provided* the orthogonal complement is taken in $\mathscr{Y}^*$ and the adjoint is defined as an operator $L^*: \mathscr{Y}^* \to \mathscr{X}^*$. For the cases we consider, $\mathscr{Y}^*$ is a space of generalized functions (i.e., distributions). The fundamental point in justifying Case (b) is that solutions of elliptic differential equations are regular. In particular, for such operators $\ker L^* \subset \mathscr{Y}$, rather than merely $\ker L^* \subset \mathscr{Y}^*$; in other words, $\ker L^*$ consists of *functions*, rather than merely distributions. (Cf. Appendix A4(c).) In consequence, the difficulties mentioned above do not arise, and the decomposition

$$\mathscr{Y} = (\text{range } L) \oplus (\text{range } L)^{\perp}$$

does hold.

Remarks. (i) Formula (1.4) provides a particular choice for N in (1.3b) that is often more convenient in applications.

(ii) When L is an elliptic differential operator, the codimension of range L equals the dimension of the kernel of L^*. Thus for such operators we have an alternative formula for the index:

$$i(L) = \dim \ker L - \dim \ker L^*.$$

(b) Mechanics of the Liapunov–Schmidt Reduction

Let

$$\Phi\colon \mathscr{X} \times \mathbb{R}^{k+1} \to \mathscr{Y}, \qquad \Phi(0, 0) = 0$$

be a smooth mapping between Banach spaces. We want to use the Liapunov–Schmidt reduction to solve the equation

$$\Phi(u, \alpha) = 0 \tag{1.5}$$

for u as a function of α near $(0, 0)$. Let L be the differential of Φ at the origin; in symbols

$$Lu = \lim_{h \to 0} \frac{\Phi(hu, 0) - \Phi(0, 0)}{h}.$$

We assume that L is Fredholm of index zero.

Remarks. (i) To simplify the notation we combine the bifurcation parameter λ and the k auxiliary parameters $\alpha_1, \ldots, \alpha_k$ into a single vector $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_k)$, where $\alpha_0 = \lambda$.

(ii) Smooth mappings between Banach spaces are discussed in Appendix 3. However, given our invariant notation of Chapter I, §3, the changes from the finite-dimensional case are minimal.

Let us recall the five steps summarized in Chapter I, §3(b), adapted to the present context.

Step 1. Decompose $\mathscr{X}$ and $\mathscr{Y}$,

$$\begin{aligned} &\text{(a)} \quad \mathscr{X} = \ker L \oplus M, \\ &\text{(b)} \quad \mathscr{Y} = N \oplus \text{range } L. \end{aligned} \tag{1.6}$$

Step 2. Split (1.5) into an equivalent pair of equations,

$$\begin{aligned} &\text{(a)} \qquad E\Phi(u, \alpha) = 0, \\ &\text{(b)} \quad (I - E)\Phi(u, \alpha) = 0, \end{aligned} \tag{1.7}$$

where $E: \mathscr{Y} \to \text{range } L$ is the projection associated to the splitting (1.6b).

Step 3. Use (1.6a) to write $u = v + w$, where $v \in \ker L$ and $w \in M$. Apply the implicit function theorem to solve (1.7a) for w as a function of v and α. This leads to a function $W: \ker L \times \mathbb{R}^{k+1} \to M$ such that

$$E\Phi(v + W(v, \alpha), \alpha) \equiv 0. \tag{1.8}$$

Step 4. Define $\phi: \ker L \times \mathbb{R}^{k+1} \to N$ by

$$\phi(v, \alpha) = (I - E)\Phi(v + W(v, \alpha), \alpha). \tag{1.9}$$

Step 5. Choose a basis $v_1, \ldots, v_n$ for $\ker L$ and a basis $v_1^*, \ldots, v_n^*$ for $(\text{range } L)^\perp$. Define $g: \mathbb{R}^n \times \mathbb{R}^{k+1} \to \mathbb{R}^n$ by

$$g_i(x, \alpha) = \langle v_i^*, \phi(x_1 v_1 + \cdots + x_n v_n, \alpha) \rangle. \tag{1.10}$$

We discuss how these five steps apply here.

Step 1. The hypothesis that L is Fredholm guarantees that the splittings (1.6) are possible. Moreover, $\ker L$ and N are finite dimensional.

Step 2. This is primarily notational and requires no comment.

Step 3. We want to show that the implicit function theorem is applicable. Let us mimic Chapter I, §3. We extract a map $F: \ker L \times M \times \mathbb{R}^{k+1} \to \text{range } L$ from (1.7a); i.e.,

$$F(v, w, \alpha) = E\Phi(v + w, \alpha). \tag{1.11}$$

The differential of F with respect to w at the origin is

$$EL = L.$$

Now we argue that

$$L: M \to \text{range } L \tag{1.12}$$

is invertible. In the finite-dimensional case this follows because L restricted to M is one-to-one and onto its range. In the Banach space case, (1.12) is still one-to-one and onto, but we need an additional, technical hypothesis to conclude that (1.12) is invertible; viz., that range L is closed. However, L is assumed Fredholm, so range L is indeed closed; thus (1.12) is invertible. Therefore, the implicit function theorem guarantees that (1.7a) may be solved for $w = W(v, \alpha)$. (Note that the solution obtained from the implicit function theorem depends smoothly on the parameters $\alpha_0, \ldots, \alpha_k$.)

Step 4. This is primarily notational and requires no comment.

Step 5. In writing $(\text{range } L)^\perp$ we are using (for the first time) the fact that $\mathscr{Y}$ is equipped with the inner product (1.2). Since L is Fredholm with index zero,

$$\dim \ker L = \dim(\text{range } L)^\perp$$

and both dimensions are finite. Thus the bases for $\ker L$ and $(\text{range } L)^\perp$ contain the same number of vectors.

Let us summarize the outcome of the Liapunov–Schmidt reduction.

Proposition 1.5. *If the linearization of* (1.5) *is a Fredholm operator of index zero, then solutions of* (1.5) *are* (*locally*) *in one-to-one correspondence with solutions of the finite system*

$$g_i(x, \alpha) = 0, \qquad i = 1, \ldots, n. \tag{1.13}$$

where g_i is defined by (1.10).

Remark 1.6. The use of the Liapunov–Schmidt reduction in different applications is facilitated by the fact that three of the above steps are totally independent of specific data in a particular application. More precisely, choices are involved in Steps 1 and 5—complementary subspaces in the former, bases in the latter—but Steps 2, 3, 4 adapt to any application without modification. Moreover, when $\ker L$ is one dimensional, the choice in Step 5 only amounts to a trivial scaling of the reduced equation (1.13).

(c) Relation of the Liapunov–Schmidt Reduction with Universal Unfoldings

Let us relate the Liapunov–Schmidt reduction to the concepts of singularity theory. For this we rewrite (1.5) with the bifurcation parameter $\lambda = \alpha_0$ explicitly displayed,

$$\Phi(u, \lambda, \alpha) = 0. \tag{1.5a}$$

In general (1.13), the reduced equation associated to (1.5a), consists of a *system* of n equations for n unknowns $x_1, \ldots, x_n$, depending on the parameters $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_k)$. In this volume we have developed singularity theory methods for the case of a single equation; i.e., $n = 1$. This case occurs when $\dim \ker L = 1$. For the most part we reserve the theory of systems for Volume II, although we will consider some examples on a case-by-case basis in this volume. Therefore, in the present discussion we assume that $n = 1$. We also rewrite (1.13), displaying the bifurcation parameter explicitly

$$g(x, \lambda, \alpha) = 0. \tag{1.13a}$$

Suppose we wish to study (1.5a) near some distinguished value of the parameter α, say $\alpha = 0$. (This could occur, for example, because α represents imperfections in a mathematical idealization of a physical problem or because $\alpha = 0$ is an organizing center (cf. Case Study 1).) We regard $g(x, \lambda, \alpha)$ as an unfolding of $g(x, \lambda, 0)$. Let us rephrase Proposition 1.5 in this language: For α near zero, the solutions of (1.5a) are locally in one-to-one correspondence with zeros of the unfolding $g(x, \lambda, \alpha)$. Singularity theory methods apply as follows. Suppose $g(x, \lambda, 0)$ is equivalent to a normal form $h(x, \lambda)$ and $h(x, \lambda)$ has a universal unfolding $H(x, \lambda, \beta)$; then there is a mapping A of parameter spaces such that

$$g(\cdot, \cdot, \alpha) \sim H(\cdot, \cdot, A(\alpha)),$$

where $\sim$ indicates equivalence. In other words, any perturbed bifurcation diagrams associated to (1.5a) can be found in the universal unfolding of h. Moreover, we may apply the techniques of Chapter III, §4 to test whether the given unfolding $g(x, \lambda, \alpha)$ of $g(x, \lambda, 0)$ is universal. If so, *every* perturbed bifurcation diagram in H will occur in (1.5a) for some value of the parameter α. If not, the cautions of Chapter IV, §1 may be applicable.

In realistic applications it is never possible to derive an explicit formula for $g(x, \lambda, \alpha)$. However it is quite possible to compute various derivatives of g at the bifurcation point—the next subsection gives formulas for this task. Thus the fact that the solution of the recognition problem (for normal forms and for universal unfoldings) depends only on a finite number of the derivatives of g at the bifurcation point is of the greatest importance.

(d) Calculation of the Derivatives of g

The calculations of derivatives of Chapter I, §3(e) apply to the present context, essentially without change. We quote the results and leave the details to the reader.

$$\text{(a)} \qquad \frac{\partial g_i}{\partial x_i} = 0,$$

$$\text{(b)} \qquad \frac{\partial^2 g_i}{\partial x_j \, \partial x_k} = \langle v_i^*, d^2\Phi(v_j, v_k)\rangle,$$

$$\text{(c)} \qquad \frac{\partial^3 g_i}{\partial x_j \, \partial x_k \, \partial x_l} = \langle v_i^*, V\rangle, \tag{1.14}$$

where

$$V = d^3\Phi(v_j, v_k, v_l) - d^2\Phi(v_j, w_{lk}) - d^2\Phi(v_k, w_{lj}) - d^2\Phi(v_l, w_{kj}), \quad \text{and}$$

$$w_{st} = L^{-1}E \, d^2\Phi(v_s, v_t).$$

$$\text{(d)} \qquad \frac{\partial g_i}{\partial \alpha_l} = \langle v_i^* \, \Phi_{\alpha_l}\rangle,$$

$$\text{(e)} \qquad \frac{\partial^2 g_i}{\partial x_j \, \partial \alpha_l} = \langle v_i^*, (d\Phi_{\alpha_l}) \cdot v_j - d^2\Phi(v_j, L^{-1}E\Phi_{\alpha_l}).$$

§2. The Elastica: An Example in Infinite Dimensions

In this section we perform the Liapunov–Schmidt reduction for the celebrated elastica, the model for a beam buckling under compression considered by Euler in 1744. (The buckling model of Chapter I, §1 is a simple, finite-element approximation of the elastica.) Euler found explicit solutions of the equations globally, using elliptic functions. We do less, in that our analysis is local, but our methods are more generally applicable.

The section is divided into four parts, as follows:

(a) Description of the problem.
(b) Analysis of the range $0 < \lambda < 1$.
(c) Setting up the reduction at $\lambda = 1$.
(d) Calculation of the derivatives of the reduced function.

(a) Description of the Problem

The configuration of the beam, assumed planar, is most conveniently described by $u(\xi)$, the angle the beam makes with the horizontal, as a function of arc length ξ. (See Figure 2.1.) Let us normalize the rod to have length π. The displacement $(x(\xi), y(\xi))$ may be calculated from the formulas

$$x(\xi) = \int_0^\xi \cos u(\xi')\, d\xi', \qquad y(\xi) = \int_0^\xi \sin u(\xi')\, d\xi'.$$

Equilibria of the beam are characterized by the two-point boundary problem

$$-\frac{d^2u}{d\xi^2} - \lambda \sin u = 0; \qquad u'(0) = u'(\pi) = 0 \tag{2.1}$$

where λ is the compressive force applied to the beam. This equation is just the first variation of a minimization problem with constraints; it is derived in Reiss [1969] under the following two assumptions.

(A) The beam is incompressible but capable of bending, the stored energy function being proportional to $\int_0^\pi \kappa^2(\xi)\, d\xi$ where κ is the curvature.
(B) The ends of the rod are hinged, permitting rotation freely, but are constrained to lie on a line.

Our goal in this section is to show that: (i) the zero solution of (2.1) is isolated and nonsingular for $0 < \lambda < 1$; and (ii) at $\lambda = 1$, the equation (2.1) is singular; elimination of the passive coordinates in (2.1) via the Liapunov–Schmidt reduction leads to a single scalar equation $g(x, \lambda) = 0$ which at the bifurcation point $x = 0, \lambda = 1$ satisfies

$$g = g_x = g_{xx} = g_\lambda = 0; \qquad g_{xxx} > 0, \quad g_{\lambda x} < 0. \tag{2.2}$$

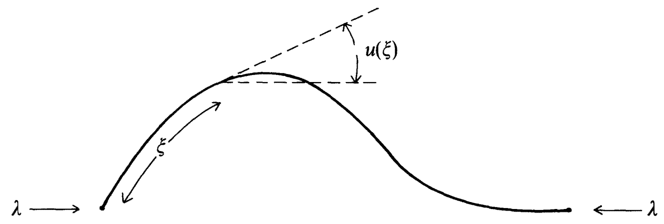


Figure 2.1. Coordinates on the beam.

This will show that the zero solution of (2.1) undergoes a supercritical pitchfork bifurcation at $\lambda = 1$. Part (ii) of this program divides naturally into two halves: first, showing that the reduction to one dimension is possible, and second, calculating the derivatives in (2.2). (*Remark*: In carrying out this program, at no time will we derive an explicit formula for $g(x, \lambda)$. As we have stated before, it is as difficult to obtain an explicit formula for $g(x, \lambda)$ as to solve the original problem. Rather we compute the derivatives (2.2) using (1.14).)

(b) Analysis of the Range $0 < \lambda < 1$

We shall write (2.1) in an abstract form

$$\Phi(u, \lambda) = 0, \tag{2.3}$$

where $\Phi: \mathscr{X} \times \mathbb{R} \to \mathscr{Y}$ is a mapping between Banach spaces defined as follows. The domain is

$$\mathscr{X} = \{u \in C^2(0, \pi): u'(0) = u'(\pi) = 0\},$$

where $C^2(0, \pi)$ is the space of real-valued, twice continuously differentiable functions, and the range is $\mathscr{Y} = C^0(0, \pi)$. Of course

$$\Phi(u, \lambda) = -u'' - \lambda \sin u. \tag{2.4}$$

Observe that $\Phi(0, \lambda) = 0$ for all λ; in other words, the undeformed configuration satisfies the equilibrium equations for any load λ. To investigate possible multiplicity of solutions, we introduce the linearization of Φ,

$$(d\Phi)_{u,\lambda} \cdot v = \lim_{h \to 0} \frac{\Phi(u + hv, \lambda) - \Phi(u)}{h}. \tag{2.5}$$

The linearization of Φ at $(0, \lambda)$ is readily computed to be

$$(d\Phi)_{0,\lambda} \cdot v = -v'' - \lambda v. \tag{2.6}$$

(*Remark:* Only the second term in (2.4) is nonlinear; its linearization amounts to the small angle approximation.)

Lemma 2.1. *The linearization $(d\Phi)_{0,\lambda}$ is invertible unless λ is one of the eigenvalues $\mu_k = k^2$, $k = 0, 1, 2, \ldots$.*

Proof. We claim that $v \in \ker(d\Phi)_{0,\lambda}$ if and only if v satisfies the Sturm–Liouville problem

$$v'' + \lambda v = 0; \qquad v'(0) = v'(\pi) = 0. \tag{2.7}$$

The differential equation comes from (2.6); the boundary conditions, from the fact that $v \in \mathscr{X}$. It is readily computed from (2.7) that $\ker(d\Phi)_{0,\lambda}$ is one dimensional when λ equals μ_k for some k and zero dimensional otherwise. According to Appendix 4, (2.7) defines a Fredholm operator of index zero. Thus by Remark 1.4, $(d\Phi)_{0,\lambda}$ is invertible unless λ equals μ_k.

Therefore, by the implicit function theorem, $u = 0$ is the only solution of (2.3) near zero for $0 < \lambda < 1$. (*Remark:* It turns out that, even globally, $u = 0$ is the only solution of (2.7) for $0 < \lambda < 1$.)

(c) Setting up the Reduction at $\lambda = 1$

We now begin the Liapunov–Schmidt reduction to study the multiplicity of solutions of (2.3) near $u = 0$, $\lambda = 1$. For brevity we define $L = (d\Phi)_{0,1}$. Note that L has a one-dimensional kernel spanned by $\cos \xi$. We split the domain of Φ into active and passive subspaces by writing

$$\mathscr{X} = \mathbb{R}\{\cos\} \oplus \mathscr{M}, \tag{2.8}$$

where $\mathscr{M} = \{u \in \mathscr{X}: \int_0^\pi \cos(\xi) u(\xi)\, d\xi = 0\}$; in words, $\mathscr{M}$ is the orthogonal complement of $\mathbb{R}\{\cos\}$ in $\mathscr{X}$ with respect to the inner product

$$\langle u, v\rangle = \int_0^\pi u(\xi) v(\xi)\, d\xi. \tag{2.9}$$

Similarly, we split the range space

$$\mathscr{Y} = N \oplus \text{range } L, \tag{2.10}$$

where $N = (\text{range } L)^\perp$. As noted in (1.4)

$$(\text{range } L)^\perp = \ker L^*.$$

However, L is self-adjoint, so

$$N = \ker L^* = \ker L = \mathbb{R}\{\cos\}. \tag{2.11}$$

(*Remark:* The following is an alternative derivation of (2.11). Since $\dim \ker L = 1$ and L is Fredholm of index zero, $(\text{range } L)^\perp$ is one dimensional. It is easily checked using integration by parts that cos is orthogonal to range L. Thus cos spans $(\text{range } L)^\perp$.)

This completes Step 1 of the reduction. We observed in Remark 1.6 that Steps 2, 3, and 4 required no specific data from the problem under study. For Step 5 we choose

$$v_1 = v_1^* = \cos.$$

All the data needed for the Liapunov–Schmidt reduction of (2.1) are now specified. Therefore solutions of (2.1) near $u = 0$, $\lambda = 1$ are in one-to-one correspondence with solutions of a single scalar equation $g(x, \lambda) = 0$, where g is given by (1.10).

(d) Calculation of the Derivatives of the Reduced Function

We obtain the derivatives of g from formulas (1.14). Moreover, in the present case Φ is odd with respect to u; i.e.,

$$\Phi(-u, \lambda) = -\Phi(u, \lambda). \tag{2.12}$$

Therefore, when $u = 0$ we have

$$(d^2\Phi)_{0,\lambda} = 0, \qquad \Phi_\lambda = 0;$$

in other words, the troublesome terms containing L^{-1} in (1.14) vanish by symmetry. Thus we have at the bifurcation point $x = 0$, $\lambda = 1$

$$\begin{aligned} &\text{(a)} \quad g = g_x = g_{xx} = g_\lambda = 0, \\ &\text{(b)} \quad g_{xxx} = \langle \cos, d^3\Phi(\cos, \cos, \cos) \rangle, \\ &\text{(c)} \quad g_{\lambda x} = \langle \cos, d\Phi_\lambda \cdot \cos \rangle. \end{aligned} \tag{2.13}$$

Now we evaluate the remaining derivatives in (2.13), showing that $g_{xxx} > 0$ and $g_{\lambda x} < 0$. First considering g_{xxx}, we claim that at $u = 0$, $\lambda = 1$

$$d^3\Phi(v_1, v_2, v_3) = v_1 v_2 v_3. \tag{2.14}$$

Indeed, from formula (A3.2) we have

$$\begin{aligned} (d^3\Phi)_{0,1}(v_1, v_2, v_3) &= -\frac{\partial^3}{\partial t_1\, \partial t_2\, \partial t_3} [(t_1 v_1'' + t_2 v_2'' + t_3 v_3'') \\ &\qquad + \sin(t_1 v_1 + t_2 v_2 + t_3 v_3)]_{t_1 = t_2 = t_3 = 0} \\ &= v_1 v_2 v_3 \cos(0) \\ &= v_1 v_2 v_3 \end{aligned}$$

as claimed. Substituting (2.14) into (2.13b) yields

$$g_{xxx} = \langle \cos, \cos^3 \rangle = \int_0^\pi \cos^4 \xi \, d\xi = \frac{3\pi}{8} > 0.$$

Similarly $\Phi_\lambda(u) = -\sin u$, so that $(d\Phi_\lambda)_{0,1} \cdot v = -v$; thus

$$g_{\lambda x} = \langle \cos, -\cos \rangle = -\pi/2 < 0.$$

We have therefore shown that the Euler strut undergoes a supercritical pitchfork bifurcation at $u = 0$, $\lambda = 1$. Note that, as in classical perturbation theory, it is possible to evaluate all numerical parameters which characterize this bifurcation.

We do not consider imperfections in the beam. It was shown in Golubitsky and Schaeffer [1979a], using a slightly more complicated formulation of the problem, that the two imperfections of a center load and a small, uniform curvature in the unstressed state provide a universal unfolding of this bifurcation problem. The interested reader is referred to that paper for further details.

We conclude §2 by showing that there is an alternative derivation of (2.13a) which does not involve any calculation. Of course, $g_x = 0$ must be satisfied at any singularity of g. We claim that the reduced function $g(x, \lambda)$ inherits the symmetry (2.12), so that

$$g(-x, \lambda) = -g(x, \lambda). \tag{2.15}$$

It follows from (2.15) that $g = g_{xx} = g_\lambda = 0$ whenever $x = 0$. To prove the claim we re-examine Step 3 in the Liapunov–Schmidt reduction. Specifically, we show that $w = -W(-v, \lambda)$ satisfies (1.8) as well as the original solution $w = W(v, \lambda)$. Let us substitute $w = -W(-v, \lambda)$ into (1.8). We find

$$\begin{aligned} E\Phi(v - W(-v, \lambda), \lambda) &= E\Phi(-[-v + W(-v, \lambda)], \lambda) \\ &= -E\Phi(-v + W(-v, \lambda), \lambda) = 0; \end{aligned}$$

the first equality is a trivial rearrangement of terms, the second comes from (2.12), and the third is (1.8) with v replaced by $-v$. Since by the implicit function theorem the function W in (1.8) is unique, we have that

$$-W(-v, \lambda) = W(v, \lambda). \tag{2.16}$$

Finally, we may prove the claim (2.15) by combining (2.16) with (1.9) and (1.10), making use of (1.12). Equation (2.15) is an example, in miniature, of the effect that the existence of a group of symmetries has on the form of the reduced equations. We consider this subject in earnest in the next section.

§3. The Liapunov–Schmidt Reduction with Symmetry

In this section we discuss the Liapunov–Schmidt reduction when the operator in equation (1.5) commutes with a compact group of symmetries. (Part of our task is to define these terms.) The main conclusion is that the reduced equations inherit the symmetry of the full equation, provided the choices made in Steps 1 and 5 of the Liapunov–Schmidt reduction respect the symmetry. (The derivation of (2.15) above is a special case of our analysis.)

Let us attempt to summarize, in the simplest terms possible, the practical implications of symmetry for applications. Frequently, as a result of symmetry, certain low-order derivatives of the reduced function g are forced to vanish. Thus to determine the qualitative behavior of the bifurcation, higher-order derivatives of g must be calculated. Sometimes this makes the Liapunov–Schmidt reduction considerably more difficult. At other times, however, symmetry itself provides a simplication, so that the calculations are no more difficult than usual.

We divide this section into three parts. In subsection (a) we both define and give examples of group actions and mappings commuting with these actions. Subsection (b) is concerned with restrictions on the choices in the reduction process, especially the construction of invariant complements. The requirement that the group of symmetries be compact enters here. In subsection (c), we prove the main result that symmetry is inherited by the reduced mapping.

Our result is an abstract one; as such, we assume that the reader has some familiarity with the basics of group theory. However, it is possible to understand the issues by considering several elementary examples. For the reader whose background in group theory is weak, we have tried to present enough examples to explain the concepts; especially, we have concentrated on those examples which will appear in later sections.

(a) Basic Definitions and Examples

Let Γ be a group of symmetries. We say that Γ *acts* on the Banach space $\mathscr{Y}$ if for each $\gamma \in \Gamma$ there is an associated invertible linear map $R_\gamma: \mathscr{Y} \to \mathscr{Y}$ with the property that for all $\gamma, \delta \in \Gamma$

$$R_{\gamma\delta} = R_\gamma \circ R_\delta. \tag{3.1}$$

(*Remark:* It follows from (3.1) that for the identity element 1 in Γ, we have R_1 is the identity on $\mathscr{Y}$. We argue as follows. By (3.3), $R_1^2 = R_1$; i.e., R_1 is a projection. But R_1 is invertible, which implies that R_1 is the identity.)

The simplest symmetries arising in applications are reflectional symmetries. We call a linear operator R on $\mathscr{Y}$ a *reflection* if $R^2 = I$. Each reflection may be identified with an action of the two-element group $\mathbf{Z}_2 = \{\pm 1\}$ by setting $R_1 = I$, $R_{-1} = R$. Note that the only nontrivial relation in (3.1) is

$$R_{-1} \circ R_{-1} = R_1,$$

which is another way of writing $R^2 = I$.

Basically, there are two ways in which such reflectional symmetries act. Typically, we are working with a space $\mathscr{Y}$ of functions $u: \Omega \to \mathbb{R}^l$, where

$\Omega \subset \mathbb{R}^k$ is some domain. The symmetry operation R may operate either on the domain or the range of functions in Y. The most common symmetry operation of the range is

$$Ru = -u. \tag{3.2}$$

In other words, $R = -I$. In the other case, R operates on the domain of functions in Y. A typical situation is that Ω is invariant under a reflection r on $\mathbb{R}^k$, say

$$r(\xi_1, \xi_2, \ldots, \xi_k) = (-\xi_1, \xi_2, \ldots, \xi_k),$$

and

$$Ru = u \circ r. \tag{3.3}$$

Let us also describe an action of the circle group $\Gamma = S^1$ on $C_{2\pi}$, the space of continuous, 2π-periodic functions $u: \mathbb{R} \to \mathbb{R}^k$. We may associate angles θ in S^1 with numbers in $[0, 2\pi)$. Thus we define

$$(R_\theta u)(s) = u(s - \theta) \tag{3.4}$$

for $\theta \in S^1$, $u \in C_{2\pi}$. In other words, S^1 acts on $C_{2\pi}$ by change of phase.

We now consider what it means for a mapping Φ to commute with the action of some group. This concept is expressed by the equation

$$\Phi(R_\gamma u) = R_\gamma \Phi(u), \tag{3.5}$$

but some discussion is required to make sense of (3.5). Suppose $\Phi: \mathscr{X} \to \mathscr{Y}$ is a mapping between Banach spaces and that Γ acts on $\mathscr{Y}$; this gives meaning to the right-hand side of (3.5). To give meaning to the left-hand side, we shall always suppose that $\mathscr{X}$ is a subspace of $\mathscr{Y}$ (usually with a different norm) such that for all $\gamma \in \Gamma$

$$u \in \mathscr{X} \Rightarrow R_\gamma u \in \mathscr{X}. \tag{3.6}$$

(We call a subspace satisfying (3.6) *invariant*.) To conclude, we shall say that a mapping $\Phi: \mathscr{X} \to \mathscr{Y}$ *commutes* with the action of Γ on $\mathscr{Y}$ if $\mathscr{X}$ is an invariant subspace of $\mathscr{Y}$ and (3.5) holds for all $\gamma \in \Gamma$, $u \in \mathscr{X}$. (More generally, one might want to define a notion of commuting which did not require $\mathscr{X} \subset \mathscr{Y}$. This can easily be done, but it requires defining a distinct action of Γ on $\mathscr{X}$. The definition we have given seems to cover the applications.)

To better understand symmetries in bifurcation problems, let us consider particular cases. For example, Φ commutes with the reflection (3.2) if and only if Φ is odd; i.e., if and only if

$$\Phi(-u) = -\Phi(u). \tag{3.7}$$

This symmetry appeared in the elastica above. The elastica also commutes with a symmetry of the form (3.3), where $r(\xi) = \pi - \xi$. Less formally, this simply means that

$$\Phi(u(\pi - \xi)) = \Phi(u)(\pi - \xi). \tag{3.8}$$

(A reflection of the form (3.3) will play an important role in our discussion of the Brusselator in §§5 and 6. Both types of reflections will contribute in Case Study 3.)

Remark 3.1. The symmetry (3.7) figured prominently in the bifurcation analysis of the elastica; by contrast, (3.8) did not matter at all. This illustrates an important principle: For symmetry to make a difference in a bifurcation problem, it must be *broken*. Let us elaborate. Symmetry is important when a trivial solution u which is symmetric (i.e., $Ru = u$) bifurcates into nontrivial solutions which are asymmetric (i.e., $Ru \neq u$). However, if the trivial solution and all bifurcating solutions are symmetric, then the presence of symmetry is irrelevant. For the elastica, (3.7) is broken but (3.8) is not.

The following example, which involves the action (3.4) of circle group, will be fundamental in Chapter VIII. Let $\mathscr{Y} = C_{2\pi}$, and let $\mathscr{X} = C^1_{2\pi}$, the set of $u \in C_{2\pi}$ which are continuously differentiable. Consider an autonomous $k \times k$ system of ODE's

$$\frac{du}{ds} + f(u) = 0.$$

To such a system we associate the operator

$$\Phi\colon C^1_{2\pi} \to C_{2\pi}$$

given by

$$\Phi(u)(s) = \frac{du}{ds}(s) - f(u(s)).$$

It is easy to check that this operator commutes with the action (3.4) of S^1. The important point here is that f does not depend explicitly on s, only implicitly through the dependence of u on s.

For the rest of this book we will simplify the notation of group actions by writing

$$\gamma \cdot u \quad \text{for} \quad R_\gamma u,$$

when $\gamma \in \Gamma$, $u \in \mathscr{Y}$.

(b) On the Construction of Invariant Complements

In Step 1 of the Liapunov–Schmidt reduction we must choose complements to certain subspaces of a Banach space. When symmetry is present, it is important to choose these complements so that they are invariant subspaces. (This concept is defined by (3.6). There is an abstract theorem, called the Peter–Weyl Theorem (Cf. Adams [1969]) which guarantees that invariant complements exist under rather general circumstances, provided Γ is compact. However, we prefer not to quote this general result. Rather, we use the fact that the situation we are dealing with has extra structure. Specifically, we have the following data:

$$\begin{aligned}&\text{(a)}\quad \Gamma \text{ acts on } \mathscr{Y};\\&\text{(b)}\quad \Phi: \mathscr{X} \to \mathscr{Y} \text{ commutes with } \Gamma;\\&\text{(c)}\quad L = (d\Phi)_0 \text{ is an elliptic differential operator.}\end{aligned} \tag{3.9}$$

Moreover, we only need to find two invariant complements M and N, as follows:

$$\mathscr{X} = \ker L \oplus M, \qquad \mathscr{Y} = N \oplus \text{range } L. \tag{3.10}$$

Because of this extra structure we can construct invariant complements with much less sophisticated arguments. To do so is the task of this subsection.

We shall assume that Γ is *compact*. In the applications of Volume I the only groups we consider are finite groups and the circle S^1; of course, all these are compact.

In the following lemma we begin to develop the above structure. (This lemma does not use the ellipticity of L.)

Lemma 3.2. *Let* Γ, Φ, $\mathscr{X}$, $\mathscr{Y}$, *and* L *be as in* (3.9). *Then*

(a) L *commutes with* Γ;

(b) $\ker L$ *is an invariant subspace of* $\mathscr{X}$;

(c) range L *is an invariant subspace of* $\mathscr{Y}$.

Proof. We use the chain rule to differentiate the identity $\Phi(\gamma \cdot u) = \gamma \cdot \Phi(u)$, and we evaluate at $u = 0$ to obtain $L\gamma = \gamma L$, thus proving (a). Next, we observe that if $u \in \ker L$ then $L\gamma \cdot u = \gamma \cdot Lu = \gamma \cdot 0 = 0$. Thus $\gamma \cdot u \in \ker L$, and (b) is proved. Finally, let u be in range L; i.e., $u = Lw$ for some $w \in \mathscr{X}$. Then $\gamma \cdot u = \gamma \cdot Lw = L(\gamma \cdot w)$. Thus $\gamma \cdot u \in$ range L, and (c) holds. □

In §1(a) we observe that we can construct the complements (3.10) by defining

$$M = (\ker L)^{\perp}, \qquad N = (\text{range } L)^{\perp}, \tag{3.11}$$

provided L is an elliptic differential operator in the latter case. Suppose the inner product $\langle \cdot, \cdot \rangle$ on $\mathscr{Y}$ is preserved by Γ; in symbols

$$\langle \gamma \cdot u, \gamma \cdot v \rangle = \langle u, v \rangle. \tag{3.12}$$

Then we claim that the orthogonal complements (3.11) are invariant. To show this, suppose $u \in (\ker L)^{\perp}$. Let $\gamma \in \Gamma$ and let $v \in \ker L$. By Lemma 3.2, $\ker L$ is invariant, so $\gamma^{-1} \cdot v \in \ker L$. Thus $\langle u, \gamma^{-1} \cdot v \rangle = 0$. However, by (3.12), $\langle \gamma \cdot u, v \rangle = 0$. Since this holds for all $v \in \ker L$, we see that

$\gamma \cdot u \in (\ker L)^{\perp}$; i.e., $(\ker L)^{\perp}$ is invariant, as claimed. Similarly for $(\text{range } L)^{\perp}$.

The assumption (3.12) is not at all restrictive; it is usually satisfied in applications. For example, the natural inner product on $C_{2\pi}$ given by

$$\langle u, v \rangle = \int_0^{2\pi} u(s)v(s)\, ds$$

is preserved by the action (3.4) of S^1 on $C_{2\pi}$;

$$\begin{aligned}\langle \theta \cdot u, \theta \cdot v \rangle &= \int_0^{2\pi} u(s-\theta)v(s-\theta)\, ds = \int_0^{2\pi} u(t)v(t)\, dt \\ &= \langle u, v \rangle,\end{aligned}$$

since one is integrating over a full period of u and v. Even if the inner product on $\mathscr{Y}$ is not preserved by Γ, we may construct an inner product which is preserved, provided Γ is compact. Specifically, given an inner product $\langle \cdot, \cdot \rangle$ on $\mathscr{Y}$, let

$$[u, v] = \frac{1}{\mathscr{H}(\Gamma)} \int_{\Gamma} \langle \gamma \cdot u, \gamma \cdot v \rangle\, d(\gamma), \tag{3.13}$$

where $\mathscr{H}$ is Haar measure. Then $[\cdot, \cdot]$ *is* preserved by Γ. (*Remark:* Formula (3.11) requires that Γ be compact, so that Γ has finite total measure.) In other words, by averaging over the group, if necessary, we may always construct an inner product which is preserved by Γ; then we define the necessary invariant complements by (3.11). (This construction of averaging over the group is dicussed in more detail in Volume II.)

(c) Proof That Symmetry Is Inherited by the Reduced Equation

Let

$$\Phi: \mathscr{X} \times \mathbb{R}^{k+1} \to \mathscr{Y}, \qquad \Phi(0, 0) = 0,$$

be a smooth mapping between Banach spaces. We want to solve the equation

$$\Phi(u, \alpha) = 0 \tag{3.14}$$

for u as a function of α near $(0, 0)$. Let Γ be a compact Lie group which acts on $\mathscr{Y}$, and suppose Φ commutes with Γ; in symbols

$$\Phi(\gamma \cdot u, \alpha) = \gamma \cdot \Phi(u, \alpha). \tag{3.15}$$

(We assume that the parameters $\alpha_0, \ldots, \alpha_k$ are not affected by Γ.) Let $L = (d\Phi)_{0,0}$; we suppose that L is an elliptic differential operator that is Fredholm of index zero.

The following proposition contains the main analysis of this section. In this proposition we focus on the reduced mapping ϕ obtained in the fourth step of the Liapunov–Schmidt reduction. (Cf. (1.9).) This version of the reduced mapping is more convenient for theoretical analysis, because it is more intrinsic (i.e., not dependent on the choice of coordinates in Step 5). After proving Proposition 3.3 we will discuss the fifth step and the symmetry properties of g.

Proposition 3.3. *In the Liapunov–Schmidt reduction of* (3.14), *if M and N in* (1.6) *are invariant subspaces, then the mapping*

$$\phi\colon \ker L \times \mathbb{R}^{k+1} \to N$$

defined by (1.9) *commutes with the action of* Γ; *in symbols*

$$\phi(\gamma \cdot v, \alpha) = \gamma \cdot \phi(v, \alpha). \tag{3.16}$$

Remark. In (3.16) it is imperative that N be an invariant subspace; otherwise $\gamma \cdot \phi(v, \alpha)$ might not belong to N, which would certainly invalidate (3.16).

Proof of proposition 3.3. Let $E\colon \mathscr{Y} \to \operatorname{range} L$ be the projection with kernel N. We claim that E commutes with Γ. For suppose that $u = v + w$ where $v \in \operatorname{range} L$ and $w \in N$. By linearity

$$\gamma \cdot E(u) = \gamma \cdot v = E(\gamma \cdot v) = E(\gamma \cdot v + \gamma \cdot w) = E(\gamma \cdot u),$$

since both range L and N are invariant subspaces. It follows that $I - E$ also commutes with Γ.

Let $W\colon \ker L \times \mathbb{R}^{k+1} \to M$ be the function defined by (1.8). We claim that

$$W(\gamma \cdot v, \alpha) = \gamma \cdot W(v, \alpha) \tag{3.17}$$

for all $\gamma \in \Gamma$. Assuming this claim, (3.16) follows by manipulating (1.9), using in sequence (3.17), the linearity of γ, (3.15), and the fact that γ commutes with $I - E$.

It remains to prove the claim (3.17). For this we use the uniqueness of solutions in the implicit function theorem. (A special case of this argument was already given in deriving (2.16).) Fix $\gamma \in \Gamma$ and define

$$W_\gamma(v, \alpha) = \gamma^{-1} \cdot W(\gamma \cdot v, \alpha).$$

We compute that

$$\begin{aligned} E\Phi(v + W_\gamma(v, \alpha)) &= E\Phi(\gamma^{-1}(\gamma \cdot v + W(\gamma \cdot u, \alpha))) \\ &= \gamma^{-1} \cdot E\Phi(\gamma \cdot v + W(\gamma \cdot v, \alpha)). \end{aligned}$$

This last term vanishes since (1.8) is valid for all v; in particular, for $\gamma \cdot v$. Thus W_γ also solves the implicit equation (1.7a), and, of course, $W_\gamma(0, 0) = 0$.

By the uniqueness of solutions to the implicit function theorem, we conclude that

$$W_\gamma(u, \alpha) = W(u, \alpha). \qquad \square$$

Finally, let us discuss the choice of bases for ker L and (range $L)^\perp$ in Step 5 of the Liapunov–Schmidt reduction. Let $v_1, \ldots, v_n$ be an arbitrary basis for ker L. Since Γ acts linearly on ker L, for each $\gamma \in \Gamma$ there is a $n \times n$ matrix of scalars, $a_{ij}(\gamma)$, such that

$$\gamma \cdot v_i = \sum_{j=1}^{n} a_{ji}(\gamma) v_j. \tag{3.18}$$

In choosing a basis $v_1^*, \ldots, v_n^*$ for (range $L)^\perp$, we want to arrange that, for the same matrices $a_{ij}(\gamma)$,

$$\gamma \cdot v_i^* = \sum_{j=1}^{n} a_{ji}(\gamma) v_j^*. \tag{3.19}$$

We shall speak of a *consistent choice* of bases if (3.18) and (3.19) hold simultaneously. If we make a consistent choice of bases for ker L and (range $L)^\perp$, then the reduced equation $g\colon \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ will satisfy

$$g(A(\gamma)x, \alpha) = A(\gamma)g(x, \alpha), \tag{3.20}$$

where $A(\gamma)$ is the $n \times n$ matrix defined by (3.18) and (3.19).

In the applications we consider, making a consistent choice of bases presents no problem. A typical situation is as follows. The kernel of L is two dimensional. The first basis element v_1 is even under a reflection R, the second, odd. To satisfy (3.19) we must choose v_1^* and v_2^* to have the same parity under R as v_1 and v_2, respectively.

The proof in the general case that bases can be chosen consistently, requires facts from the representation theory of Lie groups. Let us sketch the issues. Any representation of Γ may be decomposed into a direct sum of irreducible representations. Now, the linear mapping L is an isomorphism between M and range L. Since L commutes with Γ, L induces an isomorphism of the representations of Γ on $\mathscr{X}$ and $\mathscr{Y}$ which are left over, namely, the representations of Γ on ker L and (range $L)^\perp$, must be isomorphic. This statement is not hard to prove when $\mathscr{X} = \mathscr{Y}$ and both are finite dimensional, but it requires more care in the infinite-dimensional case. Let us define "isomorphic representations". Let $v_1, \ldots, v_n$ and $v_1^*, \ldots, v_n^*$ be bases for ker L and (range $L)^\perp$, respectively. Then define $n \times n$ matrices $A(\gamma)$, by (3.18), and $B(\gamma)$, by

$$\gamma \cdot v_i^* = \sum_{j=1}^{n} b_{ji}(\gamma) v_j^*.$$

To say that the representations are isomorphic means that there is an invertible matrix S such that for all γ

$$B(\gamma) = S^{-1}A(\gamma)S.$$

Therefore, by a change of basis in (range $L)^{\perp}$, we can always satisfy (3.19).

To conclude, let us consider (3.20) in a special case; specifically, let us suppose that

(i) $\Gamma = \mathbf{Z}_2 = \{I, R\}$, where $R^2 = I$.
(ii) $\dim \ker L = 1$.

In this case the Liapunov–Schmidt reduction leads to a scalar function $g\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ which must satisfy

$$g(A(R)x, \alpha) = A(R)g(x, \alpha). \tag{3.21}$$

Since $R^2 = I$, we have $A(R) = \pm 1$. If $A(R) = +1$, then (3.21) is no restriction whatsoever—cf. Remark 3.1 that a symmetry must be *broken* to make any difference. On the other hand, if $A(R) = -1$, then (3.21) shows that g is $\mathbf{Z}_2$-symmetric, as defined in Chapter VI.

§4. The Liapunov–Schmidt Reduction of Scalar Reaction–Diffusion Equations

(a) Description of the Problem

Reaction–diffusion equations are a common source of problems exhibiting bifurcation. The simplest such problem is

$$u_t = Du_{\xi\xi} - f(u) \tag{4.1a}$$

for $0 < \xi < l$, $t > 0$ subject to some boundary conditions at $\xi = 0$ and l, say for definiteness Dirichlet conditions

$$u(0, t) = u(l, t) = 0. \tag{4.1b}$$

In this section we suppose u is a scalar function. In applications u might represent the concentration of some chemical or a population density. If this concentration were spatially uniform (this situation is called *well stirred* in the chemical engineering literature), then u would evolve according to the ODE $du/dt = -f(u)$. However, we are interested in situations where spatial inhomogenity prevails, in which case diffusion also influences the evolution of u. The right-hand side of (4.1a) includes both these effects. For simplicity we consider diffusion in just one space dimension.

We shall suppose that

$$f(0) = 0 \quad \text{and} \quad f'(0) < 0. \tag{4.2}$$

Thus $u = 0$ is an isolated equilibrium point of the ODE which, moreover, is unstable. With our choice of boundary conditions (4.1b), $u = 0$ is also a solution of the PDE. In the present subsection, we will establish the following two points concerning the stability of the zero solution of the PDE.

(i) If diffusion is sufficiently small (resp. large), the zero solution of the PDE is unstable (resp. stable).
(ii) Bifurcation of steady-state solutions of the PDE with nontrivial spatial structure is associated with this changeover of stability

In more picturesque language, point (i) may be rephrased as saying that diffusion can stabilize an unstable solution of the ODE. For a single scalar equation this is about all that can happen. (Cf. §5 for *systems* of equations.) This is the reason we assume the solution of the ODE is unstable—if we started with a stable solution of the ODE, diffusion in the PDE would only make it "more stable."

Physically the most natural way to vary the effects of diffusion is to keep D constant but vary the length l of the interval on which the PDE is posed. In doing this it is convenient to introduce a scaled variable $\eta = \xi/l$ so that all problems are posed on the same domain, independent of l. This scaling yields the equation

$$u_t = \frac{D}{l^2} u_{\eta\eta} - f(u).$$

Thus we may write the equilibrium equation associated to (4.1) as

$$\begin{aligned} &\text{(a)} \quad -u_{\eta\eta} + \lambda f(u) = 0, \\ &\text{(b)} \qquad u(0) = u(1) = 0, \end{aligned} \tag{4.3}$$

where $\lambda = l^2/D$ is our bifurcation parameter. (Apart from boundary conditions, equation (2.1) for the buckling beam is a special case of (4.3).)

(b) Stability of the Trivial Solution

Use the left-hand side of (4.3a) to define a mapping

$$\Phi: \mathscr{X} \times \mathbb{R} \to C^0(0, 1),$$

where $\mathscr{X} = \{u \in C^2(0, 1): u(0) = u(1) = 0\}$. (We introduced the minus sign in (4.3a) so that positive eigenvalues of $d\Phi$ correspond to stability.) The linearization of (4.3) at the trivial solution $u = 0$, for a given value of λ, is

$$Lu = -u'' + \lambda f'(0)u.$$

This differs from $-(d/d\eta)^2$ merely by $\lambda f'(0)$ times the identity. Thus the spectrum of L consists of the eigenvalues of $-(d/d\eta)^2$ shifted by $\lambda f'(0)$, namely

$$n^2\pi^2 + \lambda f'(0); \qquad n = 1, 2, \ldots.$$

Note that all eigenvalues of L are positive provided

$$\lambda < \frac{\pi^2}{|f'(0)|}; \tag{4.4}$$

thus $u = 0$ is a stable solution of (4.3) for such λ's. The unstable solution of The ODE has been stabilized through diffusion. The PDE has been tied to its equilibrium value at the boundary, and provided diffusion is large enough relative to the length of the interval, this stabilizes u throughout the interval. This establishes point (i) above.

(c) The Liapunov–Schmidt Reduction

Consider (4.3) when λ is close to $\lambda_0 = \pi^2/|f'(0)|$. At λ_0, $(d\Phi)_{0,\lambda_0}$ is singular; it has a one-dimensional kernel spanned by $u_0(\eta) = \sin \pi\eta$. We analyze solutions of (4.3) through the Liapunov–Schmidt reduction, as discussed in §1. In Step 1 of the reduction we choose orthogonal complements; i.e.,

$$M = [\mathbb{R}\{\sin \pi\eta\}]^\perp,$$
$$N = [\text{range } L]^\perp = \ker L^* = \mathbb{R}\{\sin \pi\eta\},$$

the last equality because L is self-adjoint. In Step 5, we choose

$$v_1 = v_1^* = \sin \pi\eta.$$

This reduction leads to a single equation in one variable $g(x, \lambda) = 0$ whose solutions locally are in one-to-one correspondence with solutions of (4.3). (*Remark*: It is not possible to determine an explicit formula for $g(x, \lambda)$. Rather, we use (1.14) to compute enough of the derivatives of g at the bifurcation point so as locally to determine the form of the bifurcation.)

At $x = 0$, $\lambda = \lambda_0$, we have that

$$g = g_x = g_\lambda = 0.$$

In Exercise 4.1 the reader is asked to show that

$$g_{xx} = \lambda_0 f''(0)\int_0^1 v_1^3\, d\eta, \qquad g_{\lambda x} = \lambda_0 f'(0)\int_0^1 v_1^2\, d\eta < 0. \tag{4.5}$$

Let us suppose that $f''(0) \neq 0$. Then g_{xx} and $g_{\lambda x}$ are both nonzero. Moreover, $g_{\lambda\lambda} = 0$. Thus by Proposition II,9.4 the trivial solution of (4.5) undergoes a transcritical bifurcation at $\lambda = \lambda_0$. By an appropriate change of coordinates $g(x, \lambda)$ may be transformed to the normal form

$$\pm x^2 - (\lambda - \lambda_0)x. \tag{4.6}$$

For $\lambda \neq \lambda_0$, (4.7) has two zeros; namely, the trivial solution $x = 0$ and a nontrivial solution $x = \pm(\lambda - \lambda_0)$.

To understand the significance of this latter zero, we recall from the Liapunov–Schmidt reduction the correspondence between solutions of $g(x, \lambda) = 0$ and solutions of the full equations, $\Phi(u, \lambda) = 0$. Specifically

$$g(x, \lambda) = 0 \text{ iff } \Phi(xv_1 + W(xv_1, \lambda), \lambda) = 0,$$

where W is defined implicitly by (1.8). In other words, to a solution (x, λ) of $g(x, \lambda) = 0$ we associate the solution $u = xv_1 + W(xv_1, \lambda)$ of the full problem.

We see from this correspondence that the nontrivial zeros of $g(x, \lambda)$ are associated to solutions of (4.3) of the form

$$u(\eta) = xv_1(\eta) + W(x, \lambda)(\eta). \tag{4.7}$$

Moreover, $W_x(0, \lambda_0) = W_\lambda(0, \lambda_0) = 0$, the first equality by (I,3.15) and the second because the fact that $u = 0$ is a trivial solution implies that $W(0, \lambda) \equiv 0$. Thus

$$u = xv_1 + \mathcal{O}(x^2). \tag{4.8}$$

In other words, the nontrivial solutions of (4.3) have the spatial structure of v_1 near the bifurcation point.

It may happen, however, that $f''(0) = 0$—for example, this occurs if $f(u)$ is an odd function. In this case $g_{xx} = 0$ and

$$g_{xxx} = \lambda_0 f'''(0) \int_0^1 v_1^4 \, d\eta. \tag{4.9}$$

(Exercise 4.1.) Now the bifurcation is supercritical or subcritical according as $f'''(0)$ is positive or negative, with normal form

$$\pm x^3 - \lambda x.$$

(See Proposition II,9.2.) As above, the nontrivial zero of g corresponds to nontrivial solutions of (4.3), and the remarks about the spatial structure of these solutions continue to apply.

Continuing this sequence, if

$$f''(0) = f'''(0) = \cdots = f^{(k-1)}(0) = 0, \qquad f^{(n)}(0) \neq 0,$$

then we get a bifurcation problem with the canonical form

$$\pm x^n - \lambda n = 0.$$

We do not pursue this issue here. Likewise, we do not discuss possible imperfections here.

Exercise

4.1. Use formulas (1.14b, c, e) to verify (4.5) and (4.9).

§5. The Brusselator

(a) Description of the Problem

Rather more interesting phenomena can occur for *systems* of reaction–diffusion equations than for a single equation. We illustrate some of these phenomena with a specific model, often called the Brusselator. The Brusselator was designed as the simplest model consistent with chemical kinetics that exhibits oscillatory behavior like the Belusov–Zhabotinsky reaction. (Prigogine and Lefever [1974].) (Its somewhat whimsical name is meant to suggest "the oscillator created in Brussels.") This model is described by the two partial differential equations

$$\frac{\partial X}{\partial t} = D_1 \frac{\partial^2 X}{\partial \xi^2} + X^2 Y - (B + 1)X + A,$$

$$\frac{\partial Y}{\partial t} = D_2 \frac{\partial^2 Y}{\partial \xi^2} - X^2 Y + BX. \tag{5.1}$$

In these equations X, Y, A, and B all represent chemical concentrations; X and Y are unknown, while A and B are assumed fixed, independent of ξ and t. As is customary, we shall treat B as the bifurcation parameter. D_1 and D_2 are diffusion constants. We consider (5.1) on the interval $0 \leq \xi \leq l$ for $t \geq 0$, subject to the boundary conditions

$$\begin{aligned} X(0, t) &= X(l, t) = A, \\ Y(0, t) &= Y(l, t) = B/A. \end{aligned} \tag{5.2}$$

Note that $X = A$, $Y = B/A$ is an equilibrium solution of the ODE, the unique equilibrium point in fact. With the boundary condition (5.2), $X = A$, $Y = B/A$ is also a solution of the PDE, which we call the trivial solution. (*Remark*: To avoid possible confusion, let us define PDE and ODE in this context. By the PDE we mean (5.1); by the ODE we mean the equation which results from discarding the terms in (5.1) with ξ-derivatives (i.e., the diffusion terms). The ODE describes the evolution of a spatially homogeneous system.)

The Brusselator exhibits a wide variety of bifurcation phenomena, and it is therefore useful as a pedagogical example. In our analysis we shall establish the following five points.

(i) Diffusion can destabilize a stable equilibrium solution of the ODE. More specifically we shall show the following. For both the ODE and PDE, the solution $X = A$, $Y = B/A$ is stable when B is sufficiently small. For the ODE this solution is stable if and only if

$$B < 1 + A^2; \tag{5.3}$$

for the PDE this solution is stable if and only if $B < B_*$, where B_* depends on the various parameters in the problem. It may well happen that

$$B_* < 1 + A^2. \tag{5.4}$$

If so, then in the range $B_* < B < 1 + A^2$ diffusion destabilizes a stable rest point of the ODE.

(ii) As in the scalar case, steady-state solutions of the PDE with nontrivial spatial structures bifurcate from the trivial solution at $B = B_*$. Unlike the previous case, however, these solutions may have spatial structure based on *any* eigenfunction of $(\partial/\partial\xi)^2$, not just the first. Specifically, provided l is fairly large, the spatial structure of the bifurcating solutions has the form $\sin \mu\xi$, where

$$\mu \approx \pi\{A^2/D_1 D_2\}^{1/4}. \tag{5.5}$$

This is an absolutely fascinating phenomenon. These solutions have their own length scale, determined by (5.5), and it is almost completely independent of l, the length of the interval. In other words, a periodic structure with a length scale determined from parameters in the equation spontaneously develops from an undifferentiated interval as the bifurcation parameter is increased. Many people feel that the emergence of periodic structures in growing organisms, such as hair, teeth, feathers, gills, etc., involves a mechanism of the kind studied here.

(iii) We study the bifurcating solutions with the Liapunov–Schmidt technique. This reduces the problem to a single scalar equation $g = 0$. However, there are several auxiliary parameters in this problem, and for certain values of these parameters additional low-order derivatives of g may vanish at the singularity. In this way we may obtain higher-order singularities by varying the parameters.

(iv) The equation (5.1) and boundary conditions (5.2) commute with the reflection

$$\xi \to l - \xi. \tag{5.6}$$

Thus our problem provides a nice illustration of a reflectional symmetry of the form (3.3). Let us elaborate. At the bifurcation point $\ker L$ is spanned by

$$\sin(\pi\kappa\xi/l), \tag{5.7}$$

where κ is an integer such that $\pi\kappa/l$ is approximately equal to (5.5). If κ is odd, the eigenfunction (5.7) is invariant under (5.6); in the notation of (3.20) we have $A(R) = +1$. If κ is even, $A(R) = -1$. It follows from the discussion in §3(c) that when κ is odd, symmetry plays no role in our problem; when κ is even, the reduced equation is $\mathbf{Z}_2$-symmetric. In particular, when κ is even, the minimally degenerate bifurcation is a pitchfork, even though the nonlinearity in the equation has quadratic terms which by formula (1.14) would appear to contribute the second derivative g_{xx}.

(v) We also consider the effect of imperfections on the model. This shows by example how to include imperfections in calculations with the Liapunov–Schmidt reduction.

We divide the analysis of this model into two parts as follows. In the remainder of §5 we study the stability of the trivial solution; i.e., we address points (i) and (ii) above. In §6 we perform the Liapunov–Schmidt reduction, addressing points (iii), (iv), and (v) in the process.

(b) Stability of the Trivial Solution

We are interested in bifurcation from the trivial solution. Thus we introduce incremental variables $u = X - A$, $v = Y - B/A$ into (5.1). This yields

$$\Phi = \frac{\partial}{\partial t}\begin{pmatrix} u \\ v \end{pmatrix} + L\begin{pmatrix} u \\ v \end{pmatrix} + N(u, v) = 0, \tag{5.8}$$

where the linear part is given by

$$L\begin{pmatrix} u \\ v \end{pmatrix} = -\begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}\frac{\partial^2}{\partial \xi^2}\begin{pmatrix} u \\ v \end{pmatrix} - \begin{pmatrix} B - 1 & A^2 \\ -B & -A^2 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} \tag{5.9}$$

and the nonlinear part by

$$N(u, v) = -\left(\frac{B}{A}u^2 + 2Auv + u^2 v\right)\begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{5.10}$$

First let us investigate the stability of the ODE. Its linearization about the rest point is the 2×2 matrix appearing in the second term in (5.9); we denote this linearization by L_2. We have

$$\det L_2 = A^2, \qquad \operatorname{tr} L_2 = 1 + A^2 - B. \tag{5.11}$$

Since $\det L_2 \neq 0$, both eigenvalues of L_2 are nonzero. We claim that the real parts of these eigenvalues are of the same sign. To see this, first note that L_2 has real entries, so that either the eigenvalues are both real or are a pair of complex conjugates. In the latter case the eigenvalues have the *same* real part, so the claim is true. In the former case, also, the eigenvalues must have the same sign, since their product, $\det L_2$, is positive. The sum of the eigenvalues, of course, equals $\operatorname{tr} L_2$, so we may determine this common sign by inspection of (5.11). In fact, we find that $(A, B/A)$ is a stable rest point of the ODE if and only if (5.3) is satisfied. (The change in stability that the ODE undergoes at $B = 1 + A^2$ is accompanied by what is called a Hopf bifurcation. See Chapter VIII for a discussion of this phenomenon—we do not consider it further here.)

We now turn to the stability of the PDE, which we investigate through its linearization, given by (5.9). We claim that all eigenfunctions of L have the form

$$\sin(\pi k\xi/l)\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \tag{5.12}$$

where k is a positive integer. To see this, first observe that the two parameters, c_1 and c_2, in (5.12) yield a two-dimensional subspace of functions which is invariant for L; moreover, these functions vanish at the boundary. Therefore, for each k there exist two, linearly independent eigenfunctions of L of the form (5.12). But by Fourier analysis,

$$\{\sin(\pi k\xi/l: k \text{ a positive integer}\}$$

is a complete set for scalar functions; arguing componentwise, it follows that (5.12) provides a complete set for vector functions. This proves the claim.

Restriction of L to the two-dimensional subspace (5.12) gives the matrix

$$\begin{pmatrix} \dfrac{\pi^2 k^2 D_1}{l^2} - B + 1 & -A^2 \\ B & \dfrac{\pi^2 k^2 D_2}{l^2} + A^2 \end{pmatrix}. \tag{5.13}$$

We look at the eigenvalues of (5.13) through the trace and determinant of this matrix. We find

$$\text{trace} = \frac{\pi^2 k^2}{l^2}(D_1 + D_2) + A^2 + 1 - B,$$

$$\text{determinant} = \frac{\pi^4 k^4}{l^4} D_1 D_2 + \frac{\pi^2 k^2}{l^2}[D_1 A^2 + D_2(1 - B)] + A^2.$$

Note that the trace changes sign when

$$B = 1 + A^2 + \frac{\pi^2 k^2}{l^2}(D_1 + D_2); \tag{5.14}$$

i.e., at a larger value of B than where tr L_2 in (5.11) changes sign. This fact is in keeping with the generally stabilizing effects of diffusion. However, here the determinant may vanish, unlike in the previous case. Indeed, it vanishes when $B = B_k$ where

$$B_k = 1 + \frac{D_1}{D_2}A^2 + D_1\left(\frac{\pi k}{l}\right)^2 + \frac{A^2}{D_2}\left(\frac{l}{\pi k}\right)^2. \tag{5.15}$$

If $B > B_k$, the two eigenvalues of L on the subspace (5.12) are both real and *of opposite sign*; in particular, one of the eigenvalues is negative, so instability obtains in this mode.

The trivial solution of the PDE loses stability if for any k one of these eigenvalues acquires a negative real part. It follows from the above analysis that the smallest value of B at which this occurs is

$$B = \min\{B_*, 1 + A^2 + \frac{\pi^2}{l^2}(D_1 + D_2)\},$$

where

$$B_* = \min_k B_k. \tag{5.16}$$

We are trying to show that it is possible for B_* to be less than $1 + A^2$. It is clear from an inspection of the first two terms in (5.15) that we must choose $D_1 < D_2$ to achieve this, and it appears that if $D_1 \ll D_2$ the third and fourth terms will be manageable. To make this quantitative we estimate the minimum in (5.16) by calculus. In this direction, consider (5.15) for a moment for all positive real values of k, not just integer values. This function is graphed in Figure 5.1. Its minimum occurs at

$$\bar{k} = \frac{l}{\pi}(A^2/D_1D_2)^{1/4} \tag{5.17}$$

and the minimum value is

$$\bar{B} = 1 + \frac{D_1}{D_2}A^2 + 2\sqrt{\frac{D_1}{D_2}}A. \tag{5.18}$$

Observe that $\bar{B} < 1 + A^2$ if and only if

$$D_1 < D_2 \text{ and } A > 2\sqrt{\frac{D_1}{D_2}}\left(1 - \frac{D_1}{D_2}\right)^{-1}. \tag{5.19}$$

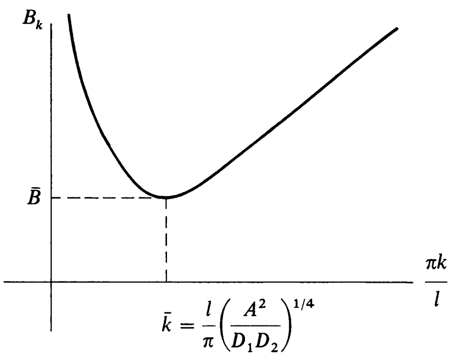


Figure 5.1. Instability curve for the Brusselator as a function of wave number.

When is (5.4), $B_* < 1 + A^2$, satisfied? We shall indicate three circumstances. In all three cases we suppose that (5.19) holds so that

$$\bar{B} < 1 + A^2. \tag{5.20}$$

As the first circumstance, suppose that the minimum in Figure 5.1 occurs at an integer value of k; i.e., suppose that $\bar{k}$ given by (5.17) is an integer. Then $B_* = \bar{B}$, and it follows from (5.20) that $B_* < 1 + A^2$. As the second circumstance, we claim that $B_* < 1 + A^2$ if l is sufficiently large. Note that $B_* \geq \bar{B}$, since $\bar{B}$ is the minimum of the function in Figure 5.1 over all real values of the argument while B_* is the minimum over only integer multiples of π/l. However, if l is large, the integer multiples of π/l are closely spaced. Thus by taking l sufficiently large, we may make $B_* - \bar{B}$ as small as desired; in particular, by (5.20), we may satisfy $B_* < 1 + A^2$. As the third circumstance, we claim that if $D_1 < D_2$ then for sufficiently large A we have $B_* < 1 + A^2$. At first this may seem obvious; however, B_* depends on A. The point is that B_* grows as $(D_1/D_2)A^2$ as $A \to \infty$, so that if $D_1 < D_2$, we may satisfy $B_* < 1 + A^2$ by choosing A large.

The above calculation also contains a derivation of (5.5). As B is increased, instability first occurs in a mode having spatial dependence $\sin(\pi k\xi/l)$, where k is approximately given by (5.17). Formula (5.5) follows from this observation.

§6. The Liapunov–Schmidt Reduction of the Bifurcation

Throughout §§5 and 6, we are studying the bifurcation of steady-state solutions of (5.8) with spatial structure from the trivial solution $u = v = 0$. In §5 we showed that the linearized operator L is invertible for $B < B_*$, where B_* is defined by (5.16), but L is singular for $B = B_*$. When $B = B_*$, $\ker L$ has dimension one, provided the minimum in (5.15) is achieved at a single integer; this is true generically, and we assume it in our analysis below. In the present section we analyze the bifurcation using the Liapunov–Schmidt reduction. We have organized this material into six subsections.

(a) Results and their interpretation in the generic case without symmetry.
(b) Calculations in the generic case without symmetry.
(c) Imperfections in the generic case without symmetry.
(d) Higher-order singularities by varying parameters.
(e) The occurrence of symmetry.
(f) Limitations of the analysis.

(a) Results and their Interpretation in the Generic Case Without Symmetry

Let κ be the integer (assumed unique) which realizes the minimum in (5.16). In subsections (a)–(d) we shall suppose κ is *odd* so that the reflectional symmetry (5.6) only enters trivially into the problem.

The Liapunov–Schmidt technique reduces the study of steady-state solutions of (5.8) to the study of the zeros of a single scalar equation

$$g(x, \lambda) = 0. \tag{6.1}$$

Here $\lambda = B - B_*$ and x parametrizes ker L. In subsection (b) we shall set up this reduction and compute a few low-order derivatives of g. Here we report the result of the calculation and interpret it.

Our problem has the trivial solution $u = v = 0$, from which it follows that $g(0, \lambda) \equiv 0$. Thus at the singularity at the origin we have

$$g = g_x = g_\lambda = g_{\lambda\lambda} = 0. \tag{6.2a}$$

In the next item we show that

$$g_{\lambda x}(0, 0) < 0 \tag{6.2b}$$

and

$$\operatorname{sgn} g_{xx}(0, 0) = \operatorname{sgn}(A^2 - D_2/D_1). \tag{6.2c}$$

In particular, $g_{xx}(0, 0)$ vanishes iff

$$A^2 = D_2/D_1. \tag{6.3}$$

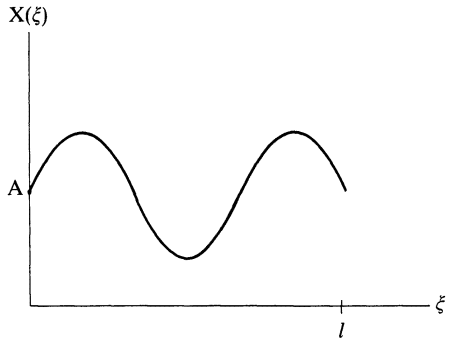
Assuming (6.3) does *not* hold, then g is equivalent to the canonical form

$$\delta x^2 - \lambda x \tag{6.4}$$

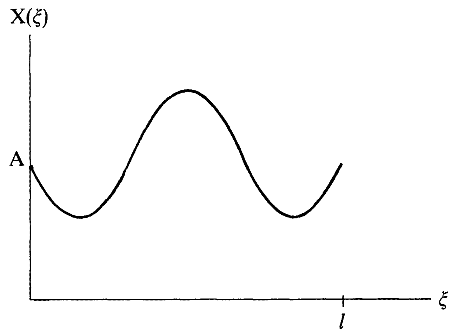
where $\delta = \pm 1$ equals $\operatorname{sgn}(A^2 - D_2/D_1)$. This normal form describes transcritical bifurcation; in particular (6.4) vanishes when $x = \delta\lambda$. Recall from the Liapunov–Schmidt reduction that to each solution (x, λ) of the reduced equations there corresponds a solution of the full equations (5.8) of the form

$$\begin{pmatrix} u \\ v \end{pmatrix} = x \sin(\pi\kappa\xi/l)\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + O(x^2).$$

Thus for each nonzero λ, there is an equilibrium solution of (5.8) with spatial structure described by $\sin(\pi\kappa\xi/l)$. The sign of δ in (6.4) determines the sign of x along the nontrivial solution branch; the sign of x in turn determines the the phase of the associated solution of (5.8), as sketched in Figure 6.1 for



(i) $A^2 > D_2/D_1$



(ii) $A^2 < D_2/D_1$

Figure 6.1. Graphs of bifurcating solutions when $\kappa = 3$, $B > B_*$.

$\kappa = 3$, $\lambda > 0$. (Note that in the figure we have used the original variables in (5.1).)

Remark. In the title of this item, we used the phrase "generic case without symmetry." By "generic case" we mean:

(a) The minimizing integer κ in (5.14) is unique.
(b) $A^2 \neq D_2/D_1$.

By "without symmetry" we invoke the restriction to odd κ.

(b) Calculations in the Generic Case Without Symmetry

Let us scale the interval $(0, l)$ to $(0, \pi)$. This scales the diffusion coefficients by a factor $(\pi/l)^2$, but we do not incorporate this change explicitly in our notation.

The integer κ which minimizes (5.16) satisfies

$$\kappa \approx (A^2/D_1D_2)^{1/4} \tag{6.5}$$

if l is moderately large. Below we derive approximate expressions for the derivatives of $g(x, \lambda)$ assuming (6.5) is exact. (These expressions are exact if the right-hand side of (6.5) is an integer.)

We write the equilibrium equation associated to (5.8) abstractly as

$$\Phi(w, \lambda) = 0, \tag{6.6}$$

where $w = (u, v)$, $\lambda = B - B_*$, and $\Phi(w, \lambda) = Lw + N(w)$. Of course, $(d\Phi)_{0,0} = L$. Let us reduce (6.6) near (0, 0) using the Liapunov–Schmidt technique. In Step 1 of the reduction we take orthogonal complements in (1.6); i.e.,

$$M = (\ker L)^\perp, \qquad N = (\text{range } L)^\perp. \tag{6.7}$$

The one-dimensional kernel of L is spanned by w_1, where

$$w_1(\xi) = \sin \kappa\xi \begin{pmatrix} D_2\kappa^2 \\ -1 - D_1\kappa^2 \end{pmatrix}. \tag{6.8}$$

By the Fredholm alternative, in (6.7) we have

$$N = (\text{range } L)^\perp = \ker L^* = \mathbb{R}\{w_1^*\},$$

where

$$w_1^*(\xi) = \sin \kappa\xi \begin{pmatrix} 1 + \dfrac{\kappa^2 D_2}{A^2} \\ 1 \end{pmatrix}. \tag{6.9}$$

In Step 5 of the reduction we choose the functions (6.8) and (6.9) as bases for ker L and (range $L)^\perp$, respectively.

Remark. In Theorem I,4.1, we showed that when reducing a finite-dimensional system, stability of a bifurcating solution could be determined from the sign of g_x, provided $\langle w_1^*, w_1 \rangle > 0$. Although we do not prove it here, this result can be extended to reaction–diffusion equations. Let us show that for the choices (6.8), (6.9), we have $\langle w_1^*, w_1 \rangle > 0$. Note that

$$\langle w_1^*, w_1 \rangle = \left\{ (D_2 - D_1)\kappa^2 + \frac{D_2^2\kappa^4}{A^2} - 1 \right\} \int_0^\pi \sin^2 \kappa\xi \, d\xi. \tag{6.10}$$

We are interested in cases where (5.4) is satisfied, which requires that

$$D_1 < D_2. \tag{6.11}$$

Now (6.11) implies that the first term in (6.10) is positive, and (6.11) in conjunction with (6.5) implies that the second term is also positive, as desired.

We calculate the derivatives (6.2b, c) using (1.14b, e):

$$\begin{aligned}&\text{(a)}\quad g_{xx}(0,0) = \langle w_1^*, d^2\Phi(w_1, w_1)\rangle,\\ &\text{(b)}\quad g_{\lambda x}(0,0) = \left\langle w_1^*, d\left(\frac{\partial \Phi}{\partial \lambda}\right)\cdot w_1\right\rangle.\end{aligned} \tag{6.12}$$

Here we have used the fact that $\partial\Phi/\partial\lambda(0, 0) = 0$ to discard the second term in (1.14e). Now

$$\begin{aligned}&\text{(a)}\quad d^2\Phi(w_1, w_1) = -2\left(\frac{B}{A}u_0^2 + 2Au_0v_0\right)\begin{pmatrix}1\\-1\end{pmatrix},\\ &\text{(b)}\quad d\left(\frac{\partial\Phi}{\partial\lambda}\right)\cdot w_1 = -u_0\begin{pmatrix}1\\-1\end{pmatrix},\end{aligned} \tag{6.13}$$

where (u_0, v_0) are the components of the eigenfunction (6.8). On substitution of (6.8) into (6.13a) we find

$$d^2\Phi(w_1, w_1) = -2\sin^2\kappa\xi\left\{\frac{B}{A}(D_2\kappa^2)^2 + 2A(D_2\kappa^2)(-1 - D_1\kappa^2)\right\}\begin{pmatrix}1\\-1\end{pmatrix}.$$

Next we use (6.5) to rewrite (5.18) as

$$B = 1 + \frac{D_1}{D_2}A^2 + 2D_1\kappa^2;$$

we substitute this expression for B and simplfy to obtain

$$d^2\Phi(w_1, w_1) = -2\sin^2\kappa\xi\frac{D_2\kappa^2}{A}\{(D_2 - D_1A^2)\kappa^2 + 2(D_1D_2\kappa^4 - A^2)\}\begin{pmatrix}1\\-1\end{pmatrix}.$$

Observe that the second term vanishes when the value (6.5) is used for κ; the first term can be reduced to

$$d^2\Phi(w_1, w_1) = -2\sqrt{D_1D_2}\sin^2\kappa\xi\left(\frac{D_2}{D_1} - A^2\right)\begin{pmatrix}1\\-1\end{pmatrix}. \tag{6.14}$$

Finally, we substitute into (6.12a) to obtain

$$g_{xx}(0,0) = +2\frac{D_2}{A}\left(A^2 - \frac{D_2}{D_1}\right)\int_0^\pi \sin^3\kappa\xi\, d\xi. \tag{6.15}$$

This integral equals $4/3\kappa$, assuming κ is odd, but its exact value is not important for (6.2c)—only the fact that it is positive. A similar calculation, left as an exercise, yields

$$g_{x\lambda}(0,0) = -\frac{D_2}{D_1}\int_0^\pi \sin^2\kappa\xi\, d\xi,$$

from which (6.2b) follows.

(c) Imperfections in the Generic Case Without Symmetry

There is a wide variety of imperfections one might consider in this problem. For example, we might prescribe boundary data in (5.2) close to the equilibrium values but not quite equal; this would obliterate the trivial solution. Or we might prescribe Robin type boundary conditions, corresponding to a nonzero resistance to influx of X and Y. Yet another perturbation is to replace the parameter A in (5.1) by the function

$$A(\xi, \varepsilon) = A_0 \frac{\cosh\sqrt{\varepsilon}\left(\xi - \frac{\pi}{2}\right)}{\cosh\sqrt{\varepsilon}\frac{\pi}{2}}, \varepsilon > 0 \tag{6.16}$$

which is motivated by the following considerations. In the derivation of (5.1), A measures a chemical concentration which is fixed by the experimenter. In practice, concentrations can only be fixed at the boundary of the domain; in the interior concentrations must be determined by solving a boundary problem

$$\frac{\partial^2 A}{\partial \xi^2} - \varepsilon A = 0 \quad \text{on} \quad (0, \pi),$$

$$A(0) = A(\pi) = A_0,$$

which has (6.16) as solution. Here ε measures the rate at which A is depleted relative to its diffusivity. This perturbation also eliminates the trivial solution. Similarly, we might allow for depletion of B in the interior of the interval.

Although there are many possible perturbations of the bifurcation problem, the universal unfolding of (6.4) only contains one parameter. Hence these different perturbations all have the same qualitative effect (provided an appropriate derivative is nonzero.) We choose the imperfection (6.16) for analysis. This illustrates how to handle imperfections within the Liapunov–Schmidt reduction. It should be noted, however, that we must have $\varepsilon \geq 0$ for this perturbation to have physical significance.

We repeat the Liapunov–Schmidt reduction on the one-parameter family of bifurcation problems obtained when A in (5.1) is replaced by (6.16). This gives an unfolding $G(x, \lambda, \varepsilon) = 0$ of the original reduced equation $g(x, \lambda) = 0$, where $G: \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. According to Table IV,3.2, in order to show that G is a universal unfolding of g it suffices to show that $G_\varepsilon \neq 0$ at the origin. We have from (1.14d)

$$G_\varepsilon(0, 0, 0) = \left\langle w_1^*, \frac{\partial \Phi}{\partial \varepsilon} \right\rangle. \tag{6.17}$$

We differentiate (5.1) to obtain

$$\frac{\partial \Phi}{\partial \varepsilon} = \frac{\partial A}{\partial \varepsilon}(\xi, 0)\begin{pmatrix}1\\0\end{pmatrix}.$$

Expanding (6.16) in powers of ε gives

$$A(\xi, \varepsilon) = A_0\left[1 + \left(\frac{\pi}{2}\right)^2 \frac{\varepsilon}{2} + \cdots\right]^{-1}\left[1 + \left(\xi - \frac{\pi}{2}\right)^2 \frac{\varepsilon}{2} + \cdots\right],$$

so that

$$\begin{aligned}\frac{\partial A}{\partial \varepsilon}(\xi, 0) &= \frac{A_0}{2}\left\{\left(\xi - \frac{\pi}{2}\right)^2 - \left(\frac{\pi}{2}\right)^2\right\}\\ &= -\frac{A_0}{2}\xi(\pi - \xi).\end{aligned}$$

On substitution in (6.17) we find

$$G_\varepsilon(0, 0, 0) = -\frac{A_0}{2}\left(1 + \frac{D_2\kappa^2}{A^2}\right)\int_0^\pi \xi(\pi - \xi)\sin \kappa\xi \, d\xi.$$

The integral here is positive so that $G_\varepsilon(0, 0, 0) < 0$. This gives rise to the diagrams sketched in Figure 6.2. Note that $\varepsilon > 0$ in both diagrams. The reason there are two cases in Figure 6.2 is because the sign of $A^2 - D_2/D_1$ determines the unspecified sign in (6.4), and this sign interacts with the perturbation to determine the effect of the perturbation. Thus we have the amusing situation that both perturbed diagrams may be obtained from a perturbation of a specified sign, albeit with different parameter values.

(d) Higher-Order Singularities by Varying Parameters

We now consider briefly what happens if (6.3) is satisfied. Then $g_{xx}(0, 0) = 0$, so we must compute higher derivatives. It turns out that $g_{xxx}(0, 0)$ is positive. The resulting pitchfork bifurcation is illustrated in

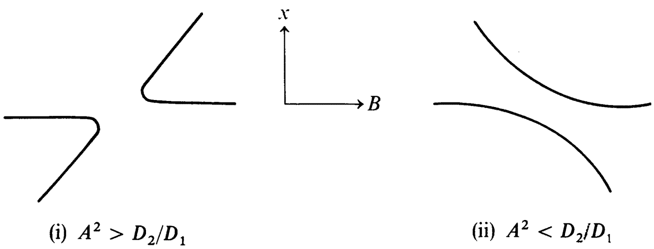


Figure 6.2. Perturbed bifurcation diagrams with $\varepsilon > 0$.

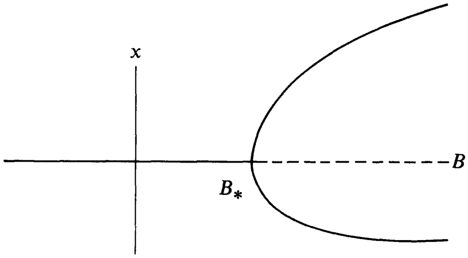
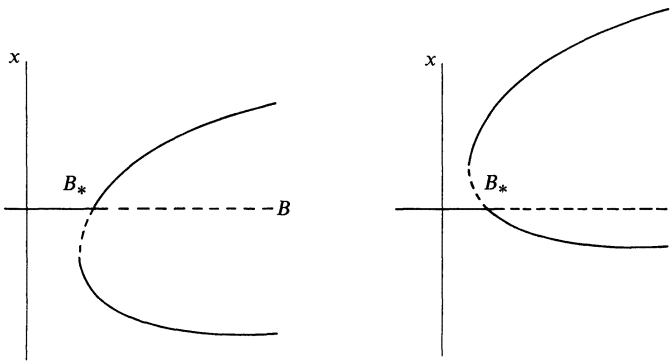


Figure 6.3. Asymmetric pitchfork in the Brusselator.

Figure 6.3. We have drawn the pitchfork somewhat asymmetrically to emphasize that this pitchfork does not arise from symmetry, but by varying parameters.

Both branches of the pitchfork are, of course, stable. If we perturb one of the parameters in (6.3), for example, A, we get an unfolding of the pitchfork as sketched in Figure 6.4. Let us interpret these diagrams. For values of the parameters which satisfy (6.3) approximately, *but not exactly*, we know from above that there is a transcritical bifurcation at $B = B_*$. This behavior is apparent in Figure 6.4, but there is additional information—the unstable branch of bifurcating solutions which exists for $B < B_*$ quickly turns around and becomes a stable branch. Experience suggests this behavior often persists for values of the parameters far from those satisfying (6.3), which is one of the reasons a local theory has proven so useful. In a specific example we must resort to numerical calculation to see whether this behavior does in fact persist.

In cases such as this when a degeneracy in g occurs through varying a parameter, we have to calculate a higher-order derivative in g than we would expect at first—here $g_{xxx}(0, 0)$. Often this calculation can be quite difficult; as discussed in Subsection (e), terms involving L^{-1} must be



(i) $A^2 > D_2/D_1$ (ii) $A^2 < D_2/D_1$

Figure 6.4. Perturbations of asymmetric pitchfork preserving the trivial solution.

evaluated. In the present case, however, it turns out that the troublesome term vanishes because (6.3) holds, and $g_{xxx}(0, 0)$ can be evaluated rather easily. See Exercise 6.1.

(e) The Occurrence of Symmetry

We now turn to the case where the integer κ minimizing (5.14) is *even*. As we remarked above, the symmetry of our problem under the reflection $\xi \to \pi - \xi$ plays an important role here because for even κ, $\sin \kappa\xi$ has odd parity with respect to this reflection. Thus we expect a pitchfork bifurcation at $B = B_*$ in this case.

Suppose we ignored this symmetry and attempted to calculate $g_{xx}(0, 0)$ as above. Formula (6.15) is equally valid here, with the important difference that for κ even the integral vanishes, since $\sin^3 \kappa\xi$ is an odd function under reflection through $\pi/2$. This provides a more traditional argument that $g_{xx}(0, 0)$ must vanish.

Thus we are led to compute $g_{xxx}(0, 0)$. This is a rather more difficult calculation. (Unlike the situation of subsection (d), we do not have (6.3) to help us.) We shall indicate the difficulties but not carry through the calculations. There is an important point to be illustrated here: the Liapunov–Schmidt reduction is fairly tractible if the lowest-order terms which can be nonzero are in fact nonzero, but the difficulty escalates rapidly if one must go beyond this first stage.

From (1.14c) we have

$$g_{xxx}(0, 0) = \langle w_1^*, d^3\Phi(w_1, w_1, w_1) - 3d^2\Phi(w_1, L^{-1}Ed^2\Phi(w_1, w_1))\rangle.$$

The first term is similar to our earlier calculations; it only requires the evaluation of certain integrals. The second term, specifically the L^{-1}, is the source of the difficulties. (Compare with Exercise 6.1.) Let us elaborate. If $(U, V) = L^{-1}Ed^2\Phi(w_1, w_1)$, then (U, V) is obtained by solving an equation

$$L\begin{pmatrix} U \\ V \end{pmatrix} = Ed^2\Phi(w_1, w_1). \tag{6.18}$$

This is a two-point boundary problem for (U, V). For reference we write (6.18) out explicitly *without the projection E* as follows.

$$\text{(a)} \quad -D_1\frac{d^2U}{d\xi^2} - (B - 1)U - A^2V = 2A\left(A^2 - \frac{D_2}{D_1}\right)\sin^2 \kappa\xi,$$

$$\text{(b)} \quad -D_2\frac{d^2V}{d\xi^2} + BU + A^2V = -2A\left(A^2 - \frac{D_2}{D_1}\right)\sin^2 \kappa\xi, \tag{6.19}$$

with boundary conditions

$$\text{(c)} \quad U(0) = U(\pi) = V(0) = V(\pi) = 0.$$

The projection E is needed because, in general, (6.19) is not solvable. The operator L is not invertible; thus (6.19) is solvable only if the right-hand side is in range L; the projection E instructs us to subtract off an appropriate multiple of w_1^* so that the right-hand side *does* belong to range L. Also the solution of (6.19) is not unique; since we may add any multiple of w_1 to a solution. The definition of the generalized inverse L^{-1} tells us to choose the unique solution which is orthogonal to w_1.

We may solve (6.19) in either of two ways. The first method is an ODE method: (6.19) is a system of ODE's with constant coefficients and the right-hand side is a linear combination of exponentials. Thus, we can find an explicit particular solution of (6.19) and then subtract off an appropriate solution of the homogeneous equation to satisfy the boundary conditions. The other method is to expand the right-hand side in an infinite series of eigenfunctions of L and invert in that way. The latter method was followed by Auchmuty and Nicolis [1975]. It turns out that the finfinite series which this method yields is explicitly summable. We refer to their paper for details. Their formula for $g_{xxx}(0, 0)$ is rather longer than seems appropriate to reproduce here, as our interest in the Brusselator is primarily pedagogical.

It is noteworthy that $g_{xxx}(0, 0)$ can have either sign or be zero, depending on the parameters A, D_1, D_2. Thus for certain values of the parameters the Brusselator undergoes a bifurcation governed by the canonical form

$$\pm x^5 - \lambda x = 0, \tag{6.20}$$

provided that the fifth-order derivative is nonzero. To our knowledge no one has calculated this derivative. The canonical form (6.20) has codimension one within the class of functions which respect the symmetry; see Chapter VI, §2 for a discussion.

We close with a brief mention of imperfections for the case of κ even. Above we listed quite a few possible imperfections. However, all of them respect the $\mathbf{Z}_2$-symmetry. Therefore, by the stability result of Chapter VI, these perturbations have no effect on the qualitative structure of the bifurcation, at least for small perturbations. This is a rather surprising result, in that it seems, at first, the perturbations would destroy the trivial solution and distort the bifurcation diagrams significantly. If we wish to find parameters which provide a universal unfolding of the bifurcation, it is essential to break the $\mathbf{Z}_2$-symmetry, for example, by imposing different boundary conditions on the two ends. We do not pursue the matter further here.

(f) Limitations of the Analysis: Multiple Eigenvalues

There is a rich structure to the complete set of solutions to the Brusselator, as the many papers in the literature on this subject will attest. We have not attempted to give a complete description of even the steady-state solutions to the Brusselator and, indeed, we have hardly scratched the surface. Our purpose was merely, however, to give a flavor for the types of equilibrium solutions which may be found by perturbations from the trivial, spatially homogeneous state at the first bifurcation. Even for this limited goal, our analysis is incomplete, and in this subsection we discuss these limitations.

Specifically, there are two principal points around which our analysis is incomplete, and these are related.

(A) We have shown above that the first bifurcation from the trivial solution in (5.8) occurs when $B = B_*$, but as a glance at Figure 5.1 will show, the subsequent bifurcation points are not far behind. This is especially true if l is large so that the multiples of π/l are closely spaced in Figure 5.1. Alternatively, we may derive this conclusion analytically from (5.14), the formula for the critical values of B at which bifurcation occurs.

It follows that when l is large, it may be technically correct to analyze the first eigenvalue as a simple eigenvalue, but to do so is to ignore much of the complicated structure of the solution set. Specifically, the fact that other bifurcations occur soon after the initial one means that the analysis based on the assumption of a simple eigenvalue is only valid in a rather small neighborhood. Moreover, the situation depicted in Figure 5.1 is typical of many of the more interesting applications of bifurcation theory; namely, that there are many modes available for bifurcation, the bifurcation point depending on the wave number as in Figure 5.1. (Cf. Case Study 3.) In our opinion, to develop techniques for dealing with this class of difficulties is one of the most challenging and interesting open problems of the subject.

(B) Even when l is relatively small, the assumption that the first eigenvalue is a simple eigenvalue is true generically, but not always. Indeed if

$$A^2 = D_1 D_2 \left(\frac{\pi\kappa}{l}\right)^2 \left[\left(\frac{\pi\kappa}{l}\right) + 1\right]^2 \tag{6.21}$$

for some integer κ, then

$$B_* = B_\kappa = B_{\kappa+1};$$

i.e., the first eigenvalue is itself double and the analysis above is not valid.

Moreover, suppose that (6.21) is only approximately valid. Then again, although it is correct to assume that the first eigenvalue is simple, the subsequent analysis is valid only on a rather small neighborhood.

The point of view described in this volume suggests that a better way to analyze problems, where (6.21) is approximately valid, is to assume that (6.21) holds exactly, analyze the resulting double eigenvalue (a kind of organizing center) and perturb. (Cf. Keener [1976], Schaeffer and Golubitsky [1981].)

We end our discussion by noting that these double eigenvalue problems always have a nontrivial symmetry. Observe that double eigenvalues result from competition between modes of wave number κ and $\kappa + 1$. Since one of these wave numbers is even and one is odd, the reflection (5.6) acts as minus the identity on one eigenfunction and as the identity on the other. It follows from the analysis of §3(c) that the Liapunov–Schmidt reduction for the double eigenvalue problem leads to a reduced bifurcation mapping $g: \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ which has a trivial solution and which commutes with the reflectional symmetry

$$(x_1, x_2) \to (-x_1, x_2).$$

Such bifurcation problems will be studied in Volume II.

Exercise

6.1. Assume (6.3) is valid; namely, $A^2 = D_2/D_1$. For the operator Φ defined in (5.8) show that

$$g_{xxx}(0, 0) = \frac{9\pi}{4}\frac{(D_2\kappa^2)^3(1 + D_1\kappa^2)}{A^2}.$$

Hint: Recall from (1.14c) that

$$g_{xxx} = \langle w_1^*, d^3\Phi(w_1, w_1, w_1) - 3d^2\Phi(w_1, w_2)\rangle,$$

where w_1^* and w_1 are defined in (6.9) and (6.8) and $w_2 = L^{-1}E(d^2\Phi)(w_1, w_1)$. Use (6.3) and (6.14) to show that $w_2 = 0$; thus simplifying substantially the calculation.

Bibliographical Comments

Several references on the Liapunov–Schmidt technique were mentioned in the bibliographical comments for Chapter I. Our treatment of the Liapunov–Schmidt reduction with symmetry is based on Sattinger [1979]; we have included this material in considerable generality because we will need it in Volume II.

Our motivation to apply singularity theory methods to the elastica stemmed from the heuristic article by Zeeman [1976]. We discussed the Brusselator partly to illustrate reaction–diffusion equations and partly to follow the fashions—the literature on this problem has become enormous. Turing [1952] was the first to observe that solutions of reaction–diffusion equations with their own length scale could bifurcate from a spatially homogeneous solution.

APPENDIX 3

Smooth Mappings Between Banach Spaces

In this appendix we consider differentiable mappings between Banach spaces. We must deal with such operators when reducing nonlinear differential equations by the Liapunov–Schmidt technique. We follow the notation introduced in Chapter I, §3(e) for such calculations in the finite-dimensional case. Indeed, for the most part, the generalization to infinite dimensions only requires checking that (at least one version of) the definitions in the finite-dimensional case remain meaningful in infinite dimensions. However, our presentation here is self-contained, although terse. (See Chow and Hale [1982] for a more complete discussion of differentiable mappings between Banach spaces.)

In §1(b), we defined the differential of a mapping $\Phi: \mathscr{X} \to \mathscr{Y}$ between Banach spaces using difference quotients. Another definition is

$$(d\Phi)_u \cdot v = \frac{d}{dt} \Phi(u + tv)\bigg|_{t=0}.$$

The latter definition is more useful for generalization to higher-order derivatives.

The kth-order derivative of such a mapping at the point $u \in \mathscr{X}$, when it exists, is defined as the multi-argument mapping

$$d^k: \underbrace{\mathscr{X} \times \cdots \times \mathscr{X}}_{k \text{ times}} \to \mathscr{Y} \tag{A3.1}$$

given by the following formula,

$$(d^k\Phi)_u(v_1, \ldots, v_k) = \frac{\partial}{\partial t_1} \cdots \frac{\partial}{\partial t_k} \Phi\left(u + \sum_{i=1}^{k} t_i v_i\right)\bigg|_{t=0}. \tag{A3.2}$$

In finite dimensions (A3.2) has a simple representation in terms of the higher-order partial derivatives of Φ (Cf. Chapter I, §3(e).); but in infinite dimensions, the invariant notation of (A3.2) is preferable because it avoids the irrelevant convergence questions which the use of components entails.

In the following discussion of (A3.2) we will assume that for every $u, v_i \in \mathscr{X}$ the mapping $\mathbb{R}^k \to \mathscr{Y}$ given by $t \to \Phi(u + \sum_1^k t_i v_i)$ is of class C^k, so that the derivative in (A3.2) is meaningful. Note that $d^k\Phi$ is a symmetric function of its arguments; i.e., $d^k\Phi(v_1, \ldots, v_k)$ is unchanged by permutation of the v_i's. In cases of interest $d^k\Phi$ is bounded and multilinear; these terms are defined as follows. We shall say that $d^k\Phi$ is *multilinear* if

$$d^k\Phi(av_1 + bv'_1, v_2, \ldots, v_k) = ad^k\Phi(v_1, v_2, \ldots, v_k) + bd^k\Phi(v'_1, v_2, \ldots, v_k), \tag{A3.3}$$

where $a, b \in \mathbb{R}$. Since $d^k\Phi$ is a symmetric function of its arguments, analogous formulas hold for linear combinations of v's in any argument of $d^k\Phi$. We shall call $d^k\Phi$ bounded if there exists a constant C such that

$$\|d^k\Phi(v_1, \ldots, v_k)\| \le C\|v_1\| \cdots \|v_k\|. \tag{A3.4}$$

Of course, $\|\cdot\|$ on the right-hand side of (A3.4) refers to the $\mathscr{X}$-norm; on the left-hand side, to the $\mathscr{Y}$-norm.

In this notation Taylor's formula for approximation by polynomials assumes the form

$$\left\|\Phi(u + v) - \sum_{j=0}^{k} \frac{1}{j!}(d^k\Phi)_u(v, \ldots, v)\right\| = o(\|v\|^k) \tag{A3.5}$$

as $v \to 0$, where by convention $(d^0\Phi)_u = \Phi(u)$. We make (A3.5) the basis of the following definition.

Definition A3.1. A mapping $\Phi: \mathscr{X} \to \mathscr{Y}$ is of class C^s if:

(i) For every $k \le s$, (A3.2) defines a bounded, symmetric, k-linear map $(d^k\Phi)_u$ which depends continuously on u (in the norm topologies).
(ii) For every $k \le s$ the estimate (A3.5) holds.

Definition A3.1 reduces to the definition of C^1 in Appendix 1, Chapter 1 when $s = 1$.

Consider, for example, a mapping $\Phi: C(0, \pi) \to C(0, \pi)$ of the form $\Phi(u) = \phi(u)$ where ϕ is a C^∞-function of one argument. (The nonlinear term in (1.1) has this form, with ϕ the sine function.) We ask the reader to show that ϕ is of class C^∞ and that

$$(d^k\Phi)_u(v_1, \ldots, v_k) = \phi^{(k)}(u)v_1, \ldots, v_k, \tag{A3.6}$$

where $\phi^{(k)}$ is the kth derivative of ϕ.

The mappings which arise in physical applications typically are differential operators. It is most important to distinguish between the differentiability of such an operator and the differentiability of functions in its domain.

Consider, for example, the linear operator $(d/d\xi)^2\colon C^2(0, \pi) \to C(0, \pi)$. In this case $d^k\Phi$ vanishes for $k \geq 2$, and the conditions of Definition A3.1 are readily verified. Thus $(d/d\xi)^2$ is infinitely differentiable, even though it operates between spaces of finitely differentiable functions. Indeed, any bounded linear transformation defines an infinitely differentiable mapping, and $(d/d\xi)^2$ is bounded because of our choice of $\mathscr{X}$ and $\mathscr{Y}$.

The following version of the chain rule will be needed below. Let $U(t)$ and $W_i(t)$, $i = 1, \ldots, k$, be smooth functions from $\mathbb{R} \to \mathscr{X}$. Then

$$\frac{d}{dt}(d^k\Phi)_U(W_1, \ldots, W_k) = (d^{k+1}\Phi)_U\left(\frac{\partial U}{\partial t}, W_1, \ldots, W_k\right) + \sum_{j=1}^{k} (d^k\Phi)_U\left(W_1, \ldots, \frac{\partial W_j}{\partial t}, \ldots, W_k\right). \tag{A3.7}$$

Here we assume that Φ does not depend explicitly on t; otherwise we must add a term with $d^k(\partial\Phi/\partial t)$ on the right-hand side of (A3.7). We leave it to the reader to derive (A3.7).

APPENDIX 4
Some Properties of Linear Elliptic Differential Operators

In this appendix we discuss some properties of elliptic boundary problems that are relevant for the applications in this book. We expect this material will be most useful as a reference for the reader with at least a vague exposure to elliptic theory. (Schechter [1977], Chapters 8–10 give a more expansive treatment.) Our approach is very myopic—we discuss only what is needed for our applications.

This appendix is divided into three parts. In the first part we develop the theme "elliptic boundary problems define Fredholm operators," primarily by example. Indeed, the examples we mention give rise to Fredholm operators of index zero. In the second part we briefly discuss the adjoint of linear elliptic operators; in the third part, elliptic regularity.

(a) Elliptic Boundary Problems as Fredholm Operators of Index Zero

Let L be a second-order, linear partial differential operator on $\mathbb{R}^N$, say

$$Lu = \sum_{i,j=1}^{N} a_{ij}(\xi)\frac{\partial^2 u}{\partial \xi_i\,\partial \xi_j} + \sum_{j=1}^{N} b_j(\xi)\frac{\partial u}{\partial \xi_j} + c(\xi)u, \tag{A4.1}$$

where $\xi = (\xi_1, \ldots, \xi_N) \in \mathbb{R}^N$. We suppose that $a_{ij}(\xi)$ is a symmetric matrix; i.e., $a_{ij}(\xi) = a_{ji}(\xi)$. (This involves no loss of generality, since a skew-symmetric matrix would sum to zero in (A4.1).) We shall call L *elliptic* if for every $\xi \in \mathbb{R}^N$ the matrix $a_{ij}(\xi)$ is positive definite. For example, if $N = 1$

this condition simply means that $a_{11}(\xi) > 0$. If L has the form

$$Lu = \Delta u + \sum_{j=1}^{N} b_j(\xi)\frac{\partial u}{\partial \xi_j} + c(\xi)u, \tag{A4.2}$$

where

$$\Delta u = \sum_{i=1}^{N} \frac{\partial^2 u}{\partial \xi_i^2}$$

is the Laplacian, then L is elliptic. We temporarily restrict our attention to operators of the form (A4.2). We assume that $b_i(\xi)$, $c(\xi)$ are smooth functions of ξ.

In discussing boundary problems for L we follow the modern practice of making L a *bounded* operator by an appropriate choice of domain and range. Let us consider one specific example in some detail; viz., the Dirichlet problem. Let $\Omega \subset \mathbb{R}^N$ be a bounded domain with smooth boundary. We want to solve

$$Lu = f \quad \text{in } \Omega \tag{A4.3a}$$

subject to the boundary condition

$$u = 0 \quad \text{on } \partial\Omega. \tag{A4.3b}$$

We formulate this problem in operator terms as follows. For $0 < s < 1$, let $C^s(\Omega)$ denote the space of functions on the closure of Ω that are Hölder continuous of exponent s, and let $C^{k+s}(\Omega)$ denote the space of functions u such that all partial derivatives of u of order k or less belong to $C^s(\Omega)$. These spaces are Banach spaces with respect to the norms

$$\|u : C^s\| = \sup_x |u(x)| + \sup_{\substack{x,y \\ (x \neq y)}} \frac{|u(x) - u(y)|}{|x - y|^s},$$

$$\|u : C^{k+s}\| = \sum_{|\alpha| \leq k} \|D^\alpha u : C^s\|.$$

Choose some number s in $(0, 1)$. We define L as an operator $L: \mathscr{X} \to \mathscr{Y}$, where

$$\begin{aligned} &\text{(a)} \quad \mathscr{X} = \{u \in C^{2+s}(\Omega) : u = 0 \text{ on } \partial\Omega\}, \\ &\text{(b)} \quad \mathscr{Y} = C^s(\Omega). \end{aligned} \tag{A4.4}$$

Then L is a bounded operator between these spaces. If $u \in \mathscr{X}$, $f \in \mathscr{Y}$, and $Lu = f$, then u is a solution to the boundary problem (A4.3).

The following proposition expresses the theme of subsection (a) in a special case.

Proposition A4.1. *The operator* (A4.2) *between the spaces* (A4.4) *is Fredholm of index zero.*

This result is proved in Berger [1977].

Remark. Suppose $N = 1$; i.e., suppose (A4.3) is a two-point boundary problem for an ODE. Then we may simplify the above example by taking $s = 0$. More precisely, for an *ordinary* differential operator, L is Fredholm of index zero between the spaces

$$\mathscr{X} = \{u \in C^2(a, b) : u(a) = u(b) = 0\},$$
$$\mathscr{Y} = C^0(a, b).$$

However, when $N > 1$, we must take $s > 0$ in order to ensure that range L be closed. (Another alternative when $N > 1$ is to work in Sobolev spaces rather than Hölder spaces. We do not pursue this here.)

Proposition A4.1 is only a very special manifestation of the principle "elliptic boundary problems define Fredholm operators." Much more general elliptic boundary problems than (A4.3) define Fredholm operators of index zero. In this book we shall need to generalize (A4.3) in the following three directions:

(i) more general boundary conditions;
(ii) systems of elliptic equations; and
(iii) operators of order κ, $\kappa \neq 2$.

All of the cases we consider lead to Fredholm operators of index zero. Let us elaborate on these cases in sequence.

(i) *More General Boundary Conditions*
Suppose we replace (A4.3b) by Neumann boundary condition; i.e.,

$$\frac{\partial u}{\partial N} = 0 \quad \text{on } \partial\Omega,$$

where $\partial/\partial N$ indicates the normal derivatives. In operator terms, let

$$\mathscr{X} = \left\{u \in C^{2+s}(\Omega) : \frac{\partial u}{\partial N} = 0\right\},$$

$\mathscr{Y} = C^s(\Omega)$. Then $L : \mathscr{X} \to \mathscr{Y}$ is Fredholm of index zero.

Similarly, we can formulate a mixed boundary problem; i.e., Dirichlet boundary conditions on one portion of $\partial\Omega$, Neumann boundary conditions on the rest. This also leads to a Fredholm operator of index zero. (In Hölder spaces the points on $\partial\Omega$ where the boundary conditions change from one type to the other cause some problems. These difficulties can be avoided entirely by working in Sobolev spaces rather than Hölder spaces. We do not discuss this further here.)

(ii) *Systems of Elliptic Equations*
The Brusselator (§§5, 6) leads to a *system* of elliptic differential equations; specifically, to equations of the form

$$Lu = Au + \sum_{j=1}^{N} B_j(\xi)\frac{\partial u}{\partial \xi_j} + C(\xi)u, \tag{A4.5}$$

where $u = (u_1, \ldots, u_k)$ is a *vector* of unknown functions and A, $B_j(\xi)$, $C(\xi)$ are $k \times k$ matrices. Suppose we define

$$\mathscr{X} = \{(u_1, \ldots, u_k) \in C^{2+s}(\Omega, \mathbb{R}^k) : u_i = 0 \text{ on } \partial\Omega\},$$

and

$$\mathscr{Y} = C^s(\Omega, \mathbb{R}^k),$$

where $C^s(\Omega, \mathbb{R}^k)$ indicates a space of vector-valued functions. Then $L: \mathscr{X} \to \mathscr{Y}$ is Fredholm of index zero.

(iii) *Operators of Order* κ, $\kappa \neq 2$
The above definition of elliptic may be generalized to operators of arbitrary order. In Case Study 3 we will encounter the fourth-order elliptic operator in $\mathbb{R}^2$

$$L = \Delta^2 + \lambda\left(\frac{\partial}{\partial \xi_1}\right)^2, \tag{A4.6}$$

where Δ^2 is the biharmonic operator. Of course a fourth-order operator needs two boundary conditions. For example, we might define

$$\mathscr{X} = \{u \in C^{4+s}(\Omega) : u = u_N = 0 \text{ on } \partial\Omega\},$$

$\mathscr{Y} = C^s(\Omega)$. Then $L: \mathscr{X} \to \mathscr{Y}$ is Fredholm, with index zero. Many other choices of boundary conditions also lead to a Fredholm operator of index zero.

It turns out that a partial differential operator with real coefficients can be elliptic only if it is of even order. For *ordinary* differential operators, however, there is no such restriction. Indeed in Chapter VIII a first-order system of ordinary differential equations with periodic boundary conditions arises. Specifically, we have

$$\begin{aligned} \mathscr{X} &= \{u \in C^1((0, b), \mathbb{R}^k) : u(0) = u(b)\}, \\ \mathscr{Y} &= \{u \in C^0((0, b), \mathbb{R}^k) : u(0) = u(b)\}, \\ Lu &= u' + Au, \end{aligned} \tag{A4.7}$$

where A is a $k \times k$ constant matrix. Then $L: \mathscr{X} \to \mathscr{Y}$ is Fredholm of index zero.

(b) The Adjoint Operator

Let $L: \mathscr{X} \to \mathscr{Y}$ be a linear, elliptic differential operator where, as above, the proper number of boundary conditions is incorporated into the definition of $\mathscr{X}$. We define the adjoint as a map $L^*: \mathscr{X} \to \mathscr{Y}$ that satisfies

$$\langle w, Lu \rangle = \langle L^*w, u \rangle \tag{A4.8}$$

for all $u, w \in \mathscr{X}$. Here brackets refer to the inner product (1.2). For example, the adjoint of (A4.2) is

$$L^*w = \Delta w - \sum_{j=1}^{n} \frac{\partial}{\partial \xi_j} [b_j(\xi)w] + c(\xi)w. \tag{A4.9}$$

This formula holds for *all* the boundary conditions for L considered above; of course, the precise domain of L^* in (A4.9) depends on the choice of boundary conditions. Formula (A4.9) may be derived by integration by parts.

Similarly, for (A4.5) we have

$$L^*w = A^t \Delta w = \sum_{j=1}^{n} \frac{\partial}{\partial \xi_j} [B_j^t(\xi)w] + C^t(\xi)w, \tag{A4.10}$$

where A^t indicates the matrix transpose. For (A4.6), L^* equals L; in other words, L is *self-adjoint*. For (A4.7)

$$L^*w = -w + A^t w.$$

(c) Elliptic Regularity

Let $L: \mathscr{X} \to \mathscr{Y}$ be a linear elliptic differential operator where, as above, suitable boundary conditions are incorporated into the definition of $\mathscr{X}$. We use the phrase *elliptic regularity* to refer to the following property of elliptic equations:

Suppose $u \in \mathscr{X}$, $f \in \mathscr{Y}$, and $Lu = f$. If $f \in C^\infty(\Omega)$, then so is u. In particular, if $Lu = 0$, then $u \in C^\infty(\Omega)$.

Elliptic regularity was important in our discussion of orthogonal complements in §1(a), especially the Fredholm alternative

$$(\text{range } L)^\perp = \ker L^*.$$

Of course

$$\ker L^* = \{u \in \mathscr{X} : L^*u = 0\}.$$

Since L^* is elliptic, *all the elements of* $\ker L^*$ *are actually* C^∞*-functions*. Because of this fact, in §1(a), it does not matter whether $(\text{range } L)^\perp$ is computed in $\mathscr{Y}$ or in $\mathscr{Y}^*$ and it does not matter whether $\ker L^*$ is regarded as a subspace of $\mathscr{Y}$ or of $\mathscr{Y}^*$—both interpretations give the same answer.

CHAPTER VIII

The Hopf Bifurcation

§0. Introduction

The term *Hopf bifurcation* refers to a phenomenon in which a steady state of an evolution equation evolves into a periodic orbit as a bifurcation parameter is varied. The Hopf bifurcation theorem (Theorem 3.2) provides sufficient conditions for determining when this behavior occurs. In this chapter, we study Hopf bifurcation for systems of ODE using singularity theory methods. The principal advantage of these methods is that they adapt well to degenerate Hopf bifurcations; i.e., cases where one or more of the hypotheses of the traditional theory fail. The power of these methods is illustrated by Case Study 2, where we present the analysis by Labouriau [1983] of degenerate Hopf bifurcation in the clamped Hodgkin–Huxley equations.

In §1 of this chapter we introduce the phenomena of Hopf bifurcation by examples. In §2 we show how periodic orbits may be characterized as the zeros of a certain mapping, and we apply the Liapunov–Schmidt reduction to this mapping to obtain a simple equation whose solutions enumerate periodic orbits. This approach is due to Cesari and Hale (See Hale [1969], Chow and Hale [1982]). At first it seems surprising that an essentially dynamic phenomenon (viz., periodic orbits) may be analyzed by steady-state techniques (viz., determining the zeros of a mapping). In §§3 and 4 we present the standard Hopf theory—existence and uniqueness of periodic orbits in §3 and stability of periodic orbits in §4. In §5 we study degenerate cases.

In each section we have tried to put the statements and discussion of all results at the beginning of the section and the proofs at the end. It should therefore be possible for the casual reader to absorb the main points of the discussion without being buried by the myriad of details needed to obtain rigorous proofs.

Our exposition follows Golubitsky and Langford [1981], in that we stress the following symmetry in the equations: We are working with spaces of periodic functions on which the circle group S^1 acts through the change of phase action (VII,3.4). Of course, this symmetry is propagated by the Liapunov–Schmidt reduction. Indeed, our reduction is a two-stage process—the first stage leads to a mapping $\phi: \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^2$ which commutes with rotations in the plane; the second stage leads to a scalar function $g: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Moreover, g is a $\mathbf{Z}_2$-symmetric bifurcation problem as analyzed in Chapter VI. Indeed, we study degenerate Hopf bifurcation by means of the unfolding theory for $\mathbf{Z}_2$-symmetric bifurcation problems in Chapter VI.

There have been several books in recent years on the Hopf bifurcation. We mention three: Marsden and McCracken [1976], Hassard *et al.* [1981], and Carr [1981]. Also, the paper of Crandall and Rabinowitz [1978] is an excellent reference for the traditional theory; we drew from this paper in writing this chapter.

§1. Simple Examples of Hopf Bifurcation

In this section we introduce the phenomena of Hopf bifurcation by describing several examples. Consider an autonomous system of ODE

$$\frac{du}{dt} + F(u, \lambda) = 0, \tag{1.1}$$

where $F: \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ is C^∞ and λ is the bifurcation parameter. Suppose that

$$F(0, \lambda) \equiv 0;$$

so $u = 0$ is a steady-state solution to (1.1) for all λ.

Hopf showed that a one-parameter family of periodic solutions to (1.1) emanating from $(u, \lambda) = (0, 0)$ could be found if two hypotheses on F were satisfied. Let $A(\lambda) = (dF)_{0,\lambda}$ be the $n \times n$ Jacobian matrix of F along the steady state solutions. The first Hopf assumption is:

$$\begin{aligned} &A(0) \quad \text{has simple eigenvalues } \pm i; \text{ and} \\ &A(0) \quad \text{has no other eigenvalues lying on the imaginary axis.} \end{aligned} \tag{1.2}$$

Remarks (i) Note that if we rescale the time t in (1.1) by setting $t = \gamma s$ for γ fixed and positive, (1.1) changes to

$$\frac{du}{ds} + \gamma F(u, \lambda) = 0.$$

Under this scaling the linearization $A(\lambda)$ is multiplied by γ. As a result we may interpret (1.2) as stating that $A(0)$ has a pair of nonzero, purely imaginary eigenvalues which have been rescaled to equal $\pm i$.

(ii) There is no difficulty in proving that periodic orbits for (1.1) exist if $A(0)$ has other eigenvalues on the imaginary axis, provided none of these is an integer multiple of $\pm i$. However, (1.2) is vital for the analysis of stability. For simplicity, we make the assumption (1.2) throughout.

We claim that $A(\lambda)$ has simple eigenvalues of the form $\sigma(\lambda) \pm i\omega(\lambda)$, where $\sigma(0) = 0$, $\omega(0) = 1$, and σ and ω are smooth. This follows from the fact that $A(\lambda)$ has real entries which depend smoothly on λ and that the eigenvalues $\pm i$ of $A(0)$ are simple. The second Hopf assumption is:

$$\sigma'(0) \neq 0; \tag{1.3}$$

that is, the imaginary eigenvalues of $A(\lambda)$ cross the imaginary axis with nonzero speed as λ crosses zero.

Hopf's first theorem states that there is a one-parameter family of periodic solutions to (1.1) if assumptions (1.2) and (1.3) hold. An elementary and instructive example of this is the simplest linear example in the plane defined by

$$F(u, \lambda) = -\begin{pmatrix} \lambda & -1 \\ 1 & \lambda \end{pmatrix} u. \tag{1.4}$$

We can compute the phase portraits for the system (1.4) as λ varies by solving the equations explicitly. With initial condition $u(0) = (a, 0)$, the solution to (1.1) is given by $u(t) = ae^{\lambda t}(\cos t, \sin t)$. The phase portrait for this system is given in Figure 1.1. For $\lambda < 0$ the steady state $u = 0$ is stable (i.e., orbits spiral into the origin), while for $\lambda > 0$ the steady state $u = 0$ is unstable (i.e., orbits spiral away from the origin). However, for $\lambda = 0$, the steady state $u = 0$ is neutrally stable, and *each* orbit is 2π-periodic. This is the one-parameter family of periodic orbits guaranteed by Hopf. We may parametrize these orbits by their amplitudes.

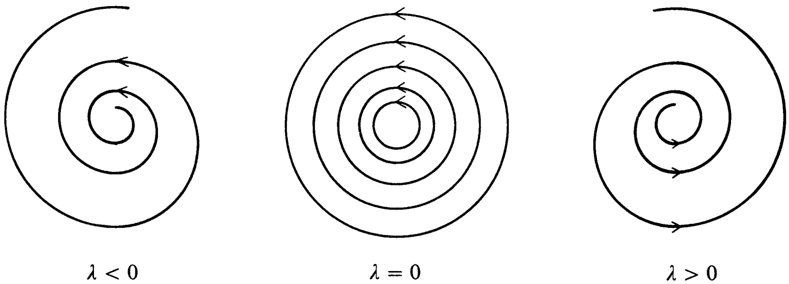


Figure 1.1. Phase portraits for the linear system (1.4).

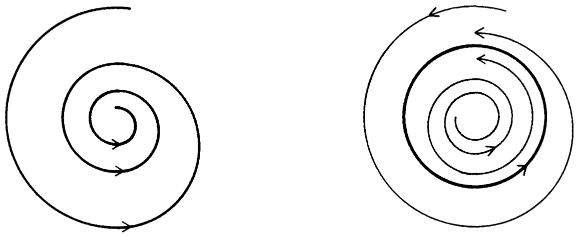


Figure 1.2. Phase portraits for the nonlinear system (1.5).

In a sense, Hopf's result above states that this family of periodic solutions persists even when higher-order terms (in u and λ) are added to F. However, this one-parameter family of periodic solutions need not remain in the plane $\lambda = 0$. In fact, the generic situation is that when higher-order terms are added to F, for each fixed λ there is at most one periodic orbit remaining near the origin.

For example, consider the system defined by

$$F(u, \lambda) = -\begin{pmatrix} \lambda & -1 \\ 1 & \lambda \end{pmatrix} u + |u|^2 u. \tag{1.5}$$

The phase portraits for this system are given in Figure 1.2. The new phenomenon in this example is that for each $\lambda > 0$ there is exactly one periodic solution of (1.5). Moreover, this periodic solution is *stable* in the sense that all nearby orbits approach this periodic solution. (Such a periodic solution is called a *stable limit cycle*.) In other words, there has been an *exchange of stability* from the steady state $u = 0$ when $\lambda < 0$ to the newly created periodic solution when $\lambda > 0$. The second Hopf theorem states that this behavior occurs typically. To better understand this phenomenon, note that the cubic terms in (1.5) push u towards the interior of circles $|u| = \text{const.}$; for $|u|$ large, these dominate, thus forcing orbits towards the origin. On the other hand, when $|u|$ is small the linear terms in (1.5) dominate, and if $\lambda > 0$ the linear terms force orbits away from the origin as in Figure 1.1. The existence of a periodic solution results from the competition of these forces.

The "bifurcation diagrams" for these two examples are presented in Figure 1.3. We have graphed there the amplitudes of the periodic and

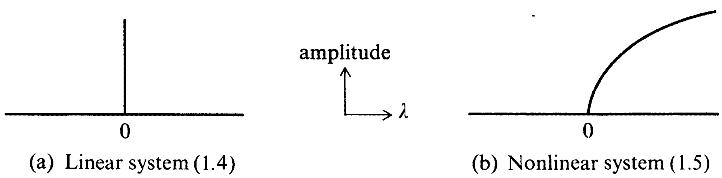


Figure 1.3. Schematic diagrams of steady state and periodic solutions in examples (1.4) and (1.5).

steady-state solutions versus the bifurcation parameter λ. The second diagram in Figure 1.3 should remind the reader of the symmetric pitchfork. We will make this association precise below by means of the Liapunov–Schmidt reduction.

§2. Finding Periodic Solutions by a Liapunov–Schmidt Reduction

In analyzing (1.1) it will be convenient to allow the equation to depend on auxiliary parameters from the start. Let $F: \mathbb{R}^n \times \mathbb{R}^{k+1} \to \mathbb{R}^n$, and consider the equation

$$\frac{du}{dt} + F(u, \alpha) = 0, \tag{2.1}$$

where $\alpha = (\alpha_0, \ldots, \alpha_k)$ combines the bifurcation parameter $\lambda = \alpha_0$ with k auxiliary parameters. In all of Chapter VIII, we suppose that

$$F(0, \alpha) \equiv 0, \tag{2.2}$$

and that $A(\alpha)$ satisfies (1.2), where $A(\alpha) = (dF)_{0,\alpha}$.

In this section, first we construct from (2.1) an operator Φ which has the property that solutions to $\Phi = 0$ correspond to periodic solutions of (2.1) with period approximately 2π, then we apply the Liapunov–Schmidt reduction to Φ, and finally we process the reduced equation to derive the following theorem, which is the main result of this section.

Theorem 2.1. *Assume that the system* (2.1) *satisfies the simple eigenvalue hypothesis* (1.2). *Then there exists a smooth germ* $g(x, \alpha)$ *of the form*

$$g(x, \alpha) = r(x^2, \alpha)x, \qquad r(0, 0) = 0$$

such that locally solutions to $g(x, \alpha) = 0$ *with* $x \geq 0$ *are in one-to-one correspondence with orbits of small amplitude periodic solutions to the system* (2.1) *with period near* 2π.

Remarks. (i) In proving Theorem 2.1, we shall show that the period of the periodic solutions obtained from this result varies smoothly with x^2.

(ii) The bifurcation diagrams in Figure 1.3 are just pictures of $g = 0$, $x \geq 0$ for the two examples. Theorem 2.1 justifies the drawing of those figures.

This section divides into three parts, in which we do the following:

(a) Define the operator Φ.
(b) Describe the Liapunov–Schmidt reduction for Φ and use its properties to prove Theorem 2.1.
(c) Derive the properties of the Liapunov–Schmidt reduction.

(a) The Definition of the Operator Φ

We want to view the system (2.1) as an operator Φ on the space of periodic functions and to reduce the problem of finding periodic solutions to (2.1) to finding solutions to $\Phi = 0$. However, there is a technical problem with this approach. We need to consider functions with various periods, and the set of all periodic functions is not a linear space—the sum of two periodic functions with different periods is not, in general, periodic. However, it is possible to circumvent this difficulty by introducing an extra parameter τ corresponding to a rescaled time. Specifically, let

$$s = (1 + \tau)t.$$

In terms of s, (2.1) may be rewritten

$$(1 + \tau)\frac{du}{ds} + F(u, \alpha) = 0. \tag{2.3}$$

We shall look for 2π-periodic solutions of (2.3). In the process we will be able to determine τ. For a given value of τ, a 2π-periodic solution of (2.3) corresponds to a periodic solution of (2.1) with period $2\pi/(1 + \tau)$. The small amplitude periodic solutions of (2.1) have periods close to 2π; thus we shall find that $\tau \approx 0$.

Let $C_{2\pi}$ be the Banach space of continuous, 2π-periodic functions from $\mathbb{R}$ into $\mathbb{R}^n$ with the norm

$$\|u\| = \max_s |u(s)|;$$

let $C^1_{2\pi}$ be the Banach space of such mappings that are continuously differentiable, with the norm

$$\|u\|_1 = \|u\| + \left\|\frac{du}{ds}\right\|.$$

We define

$$\Phi\colon C^1_{2\pi} \times \mathbb{R}^{k+1} \times \mathbb{R} \to C_{2\pi} \tag{2.4}$$

by (2.3); i.e.,

$$\Phi(u, \alpha, \tau) = (1 + \tau)\frac{du}{ds} + F(u, \alpha). \tag{2.5}$$

Then the equation $\Phi(u, \alpha, \tau) = 0$ characterizes the 2π-periodic solutions of (2.3). Note that for all α, τ,

$$\Phi(0, \alpha, \tau) \equiv 0. \tag{2.6}$$

The following point is of fundamental importance to our analysis. The circle group S^1 acts on $C_{2\pi}$ through the change of phase action (VII,3.4). We repeat this formula: for $\theta \in S^1$ and $u \in C_{2\pi}$ let

$$(\theta \cdot u)(s) = u(s - \theta). \tag{2.7}$$

The operator Φ commutes with this group action; in symbols

$$\Phi(\theta \cdot u, \alpha, \tau) = \theta \cdot \Phi(u, \alpha, \tau). \tag{2.8}$$

The first term in (2.5) commutes with this action because translation and differentiation commute. The second term commutes because the differential equation is autonomous; i.e., F does not depend explicitly on s.

(b) Properties of the Liapunov–Schmidt Reduction

Having characterized periodic solutions of (2.1) as solutions of the equation

$$\Phi(u, \alpha, \tau) = 0, \tag{2.9}$$

we now solve (2.9) using the Liapunov–Schmidt reduction. The linearization of Φ about $(u, \alpha, \tau) = (0, 0, 0)$ is given by

$$Lu = \frac{du}{ds} + A_0 u,$$

where A_0 is the $n \times n$ matrix $(dF)_{0,0}$. (In (1.2) we used the notation $A(0)$ for A_0.) The operator $L: C^1_{2\pi} \to C_{2\pi}$ is Fredholm of index zero. (Cf. Appendix 4.) The main facts needed for setting up the Liapunov–Schmidt reduction are summarized in the following proposition.

Proposition 2.2. *Assume that* (2.1) *satisfies the simple eigenvalue hypothesis* (1.2). *Then*

(a) dim ker $L = 2$.
(b) *There is a basis* v_1, v_2 *for* ker L *with the following property: If we identify* ker L *with* $\mathbb{R}^2$ *via the mapping*

$$(x, y) \to xv_1 + yv_2, \tag{2.10}$$

then the action of S^1 *on* ker L *is given by*

$$\theta \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{2.11}$$

In words, θ *acts on* $\mathbb{R}^2$ *by rotation counterclockwise through the angle* θ.
(c) *There is an invariant splitting of* $C_{2\pi}$ *given by*

$$C_{2\pi} = \text{range } L \oplus \textit{ker } L. \tag{2.12a}$$

This splitting induces a splitting of $C^1_{2\pi}$

$$C^1_{2\pi} = \ker L \oplus M, \tag{2.12b}$$

where $M = (\text{range } L) \cap C^1_{2\pi}$.

We prove this proposition in subsection (c) below. Let us now use this information to set up the Liapunov–Schmidt reduction. For our discussion we refer to the five steps outlined in Chapter VII, §1. In Step 1 of the reduction we choose M and N according to the splittings (2.12); i.e.,

$$M = (\text{range } L) \cap C^1_{2\pi}, \qquad N = \ker L. \tag{2.13}$$

Steps 2, 3, and 4 do not require specific information from the particular application intended, and we do not carry out Step 5 here. Rather, we stop after Step 4 with the coordinate free form of the reduced mapping ϕ given by (VII,1.9). For our choice (2.13) of complements.

$$\phi\colon \ker L \times \mathbb{R}^{k+1} \times \mathbb{R} \to \ker L;$$

i.e., the same space $\ker L$ occurs in both the domain and range of ϕ. (Indeed, this is one reason for not using orthogonal complements in (2.13).) Moreover, M and N are invariant complements; thus we conclude from Proposition VII,3.3 that ϕ commutes with the action of S^1 on $\ker L$; in symbols,

$$\phi(\theta \cdot v, \alpha, \tau) = \theta \cdot \phi(v, \alpha, \tau), \tag{2.14}$$

where the action of S^1 is given by (2.11). (Since $\ker L$ occurs in both the domain and range of ϕ, the situation here is simpler than that of Chapter VII, §3—specifically, the issue of making a consistent choice of bases for $\ker L$ and $(\text{range } L)^\perp$ does not arise.)

Proposition 2.3. *In the coordinates on* $\ker L$ *defined by* (2.10), *the reduced mapping* ϕ *has the form*

$$\phi(x, y, \alpha, \tau) = p(x^2 + y^2, \alpha, \tau)\begin{pmatrix} x \\ y \end{pmatrix} + q(x^2 + y^2, \alpha, \tau)\begin{pmatrix} -y \\ x \end{pmatrix}, \tag{2.15}$$

where p and q are smooth germs satisfying

$$\begin{aligned} &\text{(a)} \quad p(0,0,0) = 0, \qquad \text{(b)} \quad q(0,0,0) = 0, \\ &\text{(c)} \quad p_\tau(0,0,\tau) = 0, \qquad \text{(d)} \quad q_\tau(0,0,\tau) = -1. \end{aligned} \tag{2.16}$$

We prove Proposition 2.3 in subsection (c) below. We conclude this subsection by deriving Theorem 2.1 from Propositions 2.2 and 2.3.

Proof of Theorem 2.1. Observe from (2.15), the explicit form for ϕ, that $\phi = 0$ if and only if one of the following relations holds:

$$\begin{aligned} &\text{(a)} \quad x = y = 0, \\ &\text{(b)} \quad p = q = 0. \end{aligned} \tag{2.17}$$

Solutions to (2.17a) correspond to the trivial, steady state solution $u = 0$, while solutions to (2.17b) correspond to 2π-periodic solutions of the system (2.3); the latter are nonconstant if $z = x^2 + y^2 > 0$. It is convenient to eliminate the redundancy in solutions to (2.17b) associated to the S^1 action. Thus, we assume that $y = 0$ and $x \geq 0$, as any vector may be put into this form by a suitable rotation of the plane. Equations (2.17) then have the form

$$\begin{aligned} &\text{(a)} \quad x = 0, \\ &\text{(b)} \quad p(x^2, \lambda, \tau) = q(x^2, \lambda, \tau) = 0. \end{aligned} \tag{2.18}$$

Now we claim that near the origin the equation

$$q(x^2, \alpha, \tau) = 0 \tag{2.19}$$

may be solved for $\tau = \tau(x^2, \alpha)$. Indeed, (2.16b, d) provide the input needed to derive the claim from the implicit function theorem. Let us define

$$r(z, \alpha) = p(z, \alpha, \tau(z, \alpha)), \qquad g(x, \alpha) = r(x^2, \alpha)x. \tag{2.20}$$

Then the equation

$$\phi(x, y, \tau, \alpha) = 0 \tag{2.21}$$

has solutions with $x^2 + y^2 > 0$ only if $\tau = \tau(x^2 + y^2, \alpha)$; moreover, all solutions of (2.21) may be obtained from a solution of

$$g(x, \alpha) = 0,$$

with $x \geq 0$, by an appropriate rotation. On the other hand, solutions of (2.21) locally are in one-to-one correspondence with periodic solutions of (2.1). □

(c) Proofs of Propositions 2.2 and 2.3

Proof of proposition 2.2. We consider points (a), (b), (c) in sequence.

(a) Consider the linear system of ODE's with constant coefficients

$$Lu = 0, \qquad u \in C^1_{2\pi},$$

where $L = d/ds + A_0$. The general solution of $Lu = 0$ is a sum of exponentials times eigenvectors of A_0. Only the eigenvalues $\pm i$ lead to 2π-periodic solutions. Let $c \in \mathbb{C}^n$ satisfy

$$A_0 c = -ic, \qquad \bar{c}^t \cdot c = 2, \tag{2.22}$$

where the row vector $\bar{c}^t$ is formed from the column vector c by taking the transpose and replacing every entry by its complex conjugate. Then

$$v_1(s) = \operatorname{Re}(e^{is}c), \qquad v_2(s) = \operatorname{Im}(e^{is}c) \tag{2.23}$$

form a basis for $\ker L$. In particular, $\dim \ker L = 2$.

(b) We consider the basis (2.23) for $\ker L$. A simple calculation using (2.23) shows that

$$\theta \cdot v_1(s) = v_1(s - \theta) = \cos(\theta)v_1(s) + \sin(\theta)v_2(s),$$
$$\theta \cdot v_2(s) = v_2(s - \theta) = -\sin(\theta)v_1(s) + \cos(\theta)v_2(s).$$

Formula (2.11) follows on recalling the identification (2.10).

(c) We choose a computational proof of this point, so as to minimize the use of functional analysis. First, let us construct a basis for $\ker L^*$, where

$$L^*w = -\frac{dw}{ds} + A_0^t w \tag{2.24}$$

is the adjoint operator with respect to the inner product

$$\langle u, v \rangle = \frac{1}{2\pi} \int_0^{2\pi} \overline{v(s)^t u(s)}\, ds. \tag{2.25}$$

(Cf. Appendix 4.) For the moment we consider only real-valued functions, so the complex conjugation in (2.25) has no content; below it will be convenient to consider complex-valued functions.) Now A_0 and A_0^t have the same eigenvalues; in particular, the hypothesis (1.2) holds for A_0^t. Let d be a nonzero vector in $\mathbb{C}^n$ satisfying

$$A_0^t d = \text{id}, \tag{2.26}$$

and let

$$v_1^*(s) = \text{Re}(e^{is}d), \qquad v_2^*(s) = \text{Im}(e^{is}d). \tag{2.27}$$

Then v_1^*, v_2^* are a basis for $\ker L^*$. It is convenient to normalize d in (2.26) according to the following lemma.

Lemma 2.4. *The eigenvector d may be chosen such that*

$$\text{(a)} \quad \bar{d}^t c = 2, \qquad \text{(b)} \quad d^t c = 0. \tag{2.28}$$

Remark. This lemma expresses the biorthogonality of the right and left eigenvectors of the matrix A_0. (Cf. Noble [1969], §10.7.)

PROOF OF LEMMA 2.4. Let a be any eigenvector of A_0^t; say $A_0^t a = \mu a$. Then

$$-ia^t c = a^t(A_0 c) = (A_0^t a)^t c = \mu a^t c. \tag{2.29}$$

Thus $a^t c = 0$ if $\mu \neq -i$. In particular, (2.28b) follows. We claim that $\bar{d}^t c \neq 0$. Suppose otherwise. Then c is orthogonal to *all* the eigenvectors of A_0^t—by hypothesis, c is orthogonal to $\bar{d}$, which is the eigenvector of A_0^t associated to the eigenvalue $-i$; by (2.29), c is orthogonal to all the others. This implies that $c = 0$, a contradiction. Thus $\bar{d}^t c \neq 0$, as claimed, and we may scale d so that (2.28a) holds. (*Remark:* This argument requires more attention if A_0 has multiple eigenvalues. However, by (1.2), $\pm i$ are *not* multiple eigenvalues, and this is the essential point. If the remaining eigenvalues are not simple we can show that c is orthogonal to their generalized eigenspaces.) □

With this normalization for d, we have the following formulas for various inner products of v_j and v_j^*: for $j, k = 1, 2$

$$\begin{aligned} &\text{(a)} \quad \langle v_j^*, v_k^* \rangle = \tfrac{1}{2}\bar{d}^t d\delta_{jk}, \\ &\text{(b)} \quad \langle v_j^*, v_k \rangle = \delta_{jk}, \\ &\text{(c)} \quad \langle v_j, v_k \rangle = \delta_{jk}. \end{aligned} \tag{2.30}$$

We leave it for the reader to verify these formulas by substituting (2.23) and (2.27) into (2.25).

Now let us verify the splitting (2.12a). Since L is Fredholm with index zero,

$$\operatorname{codim} \operatorname{range} L = \dim \ker L = 2.$$

In other words, $\ker L$ has the right dimension to be a complementary subspace to range L. To verify (2.12a) it suffices to show that

$$(\operatorname{range} L) \cap (\ker L) = \{0\}. \tag{2.31}$$

It follows from the Fredholm alternative (VII,1.4) that

$$\operatorname{range} L = (\ker L^*)^{\perp}. \tag{2.32}$$

Suppose v belongs to the intersection (2.31). Since $v \in \ker L$, we may write $v = xv_1 + yv_2$. Since $v \in \operatorname{range} L$, we deduce from (2.32) that $\langle v, v_j^* \rangle = 0$; on recalling (2.30b) we conclude that $x = y = 0$. This verifies (2.31) and hence (2.12a).

Regarding (2.12b), we see from (2.32) that

$$M = \{u \in C_{2\pi}^1 : \langle u, v_1^* \rangle = \langle u, v_2^* \rangle = 0\}.$$

The decomposition (2.12b) now follows from an argument similar to the one justifying (2.12a). □

We now turn to the proof of Proposition 2.3. The fact that the reduced equations have the form (2.15) is a general consequence of symmetry. Let us isolate what is involved in the next lemma.

Lemma 2.5. *Let $\phi\colon \mathbb{R}^2 \to \mathbb{R}^2$ be a smooth mapping that commutes with* (2.11). *Then there exist smooth functions $p(z)$, $q(z)$ of one real variable such that*

$$\phi(x, y) = p(x^2 + y^2)\begin{pmatrix} x \\ y \end{pmatrix} + q(x^2 + y^2)\begin{pmatrix} -y \\ x \end{pmatrix}. \tag{2.33}$$

PROOF. Let us write ϕ in coordinates as

$$\phi(x, y) = (\phi_1(x, y), \phi_2(x, y)).$$

Since ϕ commutes with S^1, ϕ commutes with rotation through the angle π. Now $\pi \cdot (x, y) = (-x, -y)$. Thus

$$\phi(-x, -y) = (-\phi_1(x, y), -\phi_2(x, y)).$$

In particular, if $x = s$ and $y = 0$, then

$$\phi_j(-s, 0) = -\phi_j(s, 0), \qquad j = 1, 2.$$

In other words, each component $\phi_j(s, 0)$ is an odd function of s. By Corollary VI,2.2

$$\phi_1(s, 0) = p(s^2)s; \qquad \phi_2(s, 0) = q(s^2)s$$

for some smooth functions p and q.

More generally, given a point (x, y), choose an angle θ such that $\theta \cdot (s, 0) = (x, y)$ where $s^2 = x^2 + y^2$. It follows that

$$\begin{aligned}\phi(x, y) &= \phi(\theta \cdot (s, 0)) = \theta \cdot \phi(s, 0) \\ &= \theta \cdot (p(s^2)s, q(s^2)s) \\ &= p(x^2 + y^2)[\theta \cdot (s, 0)] + q(x^2 + y^2)[\theta \cdot (0, s)],\end{aligned}$$

the last equality following from the linearity of the action (2.11). Now $\theta \cdot (s, 0) = (x, y)$ and we claim that $\theta \cdot (0, s) = (-y, x)$. It is clear that $\theta \cdot (0, s)$ is perpendicular to $\theta \cdot (s, 0)$, since $(0, s)$ and $(s, 0)$ are perpendicular and rotation preserves angles. Moreover $\theta \cdot (0, s)$ has length $(x^2 + y^2)^{1/2}$. Thus to verify the claim it suffices to check the sign, and this is easy. □

PROOF OF PROPOSITION 2.3. We have already discussed (2.15) in the above lemma. Formula (2.15) differs slightly from Lemma 2.5, in that there are auxiliary parameters in (2.15). However, there is no difficulty in extending the lemma to allow ϕ to depend on parameters.

We verify (2.16) using the formulas (VII,1.14) for the derivatives of the reduced function. In order to apply these formulas we now carry out Step 5 of the Liapunov–Schmidt reduction. Specifically, for $j = 1, 2$, let

$$\phi_j(x, y, \alpha, \tau) = \langle v_j^*, \phi(xv_1 + yv_2, \alpha, \tau)\rangle. \tag{2.34}$$

Then

$$\begin{aligned}&\text{(a)} \quad \phi_1(x, 0, \alpha, \tau) = p(x^2, \alpha, \tau)x, \\ &\text{(b)} \quad \phi_2(x, 0, \alpha, \tau) = q(x^2, \alpha, \tau)x.\end{aligned} \tag{2.35}$$

Thus

$$p(0, 0, 0) = \frac{\partial \phi_1}{\partial x}(0, 0, 0, 0),$$

$$q(0, 0, 0) = \frac{\partial \phi_2}{\partial x}(0, 0, 0, 0),$$

which vanish by (VII,1.14a). This proves (2.16a, b). Also from (2.35)

$$p_\tau(0, 0, \tau) = \frac{\partial^2 \phi_1}{\partial x\, \partial \tau}(0, 0, 0, \tau),$$

$$q_\tau(0, 0, \tau) = \frac{\partial^2 \phi_2}{\partial x\, \partial \tau}(0, 0, 0, \tau).$$

By (VII,1.14e)

$$\frac{\partial^2 \phi_j}{\partial x\, \partial \tau} = \langle v_j^*, d(\Phi_\tau) \cdot v_1 - d^2\Phi(v_1, L^{-1}E\Phi_\tau)\rangle. \tag{2.36}$$

But from (2.5)

$$\Phi_\tau(u, \alpha, \tau) = \frac{du}{ds}.$$

Thus $\Phi_\tau(0, \alpha, \tau) = 0$, so the second term in (2.36) vanishes. For the first we have

$$d(\Phi_\tau) \cdot v_1 = \frac{dv_1}{ds}.$$

We claim that

$$\frac{dv_1}{ds} = -v_2. \tag{2.37}$$

We substitute (2.37) into (2.36) and recall (2.30b) to prove (2.16c, d).

It remains to prove (2.37). This is a special case of the following formulas:

$$\text{(a)}\quad \frac{dv_1}{ds} = -v_2, \qquad \text{(b)}\quad \frac{dv_2}{ds} = v_1,$$

$$\text{(c)}\quad \frac{dv_1^*}{ds} = -v_2^*, \qquad \text{(d)}\quad \frac{dv_2^*}{ds} = v_1^*. \tag{2.38}$$

These may be obtained by differentiating (2.22) and (2.27); we leave this for the reader. □

§3. Existence and Uniqueness of Solutions

In this section we apply the reduction of §2 to discuss periodic solutions of a system of ordinary differential equations,

$$\frac{du}{dt} + F(u, \alpha) = 0. \tag{3.1}$$

Here, as in (2.1), $\alpha = (\alpha_0, \ldots, \alpha_k)$, where $\alpha_0 = \lambda$ is the bifurcation parameter and $\alpha_1, \ldots, \alpha_k$ are auxiliary parameters. Our goal is to prove the first two Hopf theorems. The first theorem provides sufficient conditions for a family of periodic orbits to (3.1) to exist; the second provides sufficient conditions for the orbits to be parametrized by λ, the bifurcation parameter.

This section is divided into three parts. In subsection (a) we formulate our results and apply them to the example (1.5) in §1 above. The last two parts contain calculations which may be omitted without loss of continuity. The calculations of subsection (b) support subsection (a). The calculations of subsection (c) will not be needed until §4, but we include them here, as the same techniques are used.

(a) Statement and Discussion of the Result

In Theorem 2.1 we reduced the study of periodic orbits of (2.1) to solving a single scalar equation

$$g(x, \alpha) = 0, \tag{3.2}$$

provided the simple eigenvalue hypothesis (1.2) is satisfied. Now $g(x, \alpha)$ has the form $r(x^2, \alpha)x$ for some function r; the nontrivial solutions of (3.2) may be obtained by solving

$$r(x^2, \alpha) = 0. \tag{3.3}$$

The information in the Hopf theorems is readily obtained if the function r in (3.3) is available; the difficult point is to derive the information directly from the differential equation (3.1). Suppose for the moment that r is known. Let us suppose further that

$$r_\lambda(0, 0) \neq 0, \tag{3.4}$$

where $\lambda = \alpha_0$ is the bifurcation parameter. Then by the implicit function theorem we may solve (3.3) for λ as a function of x^2 and $\alpha' = (\alpha_1, \ldots, \alpha_k)$; in symbols

$$\lambda = \mu(x^2, \alpha'). \tag{3.5}$$

In other words, if (3.4) holds, then (3.1) has a $(k + 1)$-parameter family of periodic solutions which bifurcate from the trivial solution. (If there are no auxiliary parameters in (3.1) (i.e., if $k = 0$), then we find a one-parameter family of periodic solutions.) In Proposition 3.3 below we express $r_\lambda(0, 0)$ in terms of F in (3.1) as follows. Assuming (1.2) holds, the matrices $A(\alpha) = (dF)_{0,\alpha}$ have simple eigenvalues close to $\pm i$ that vary smoothly with α; we let

$$\Sigma(\alpha) = \sigma(\alpha) - i\omega(\alpha) \tag{3.6}$$

be the eigenvalue of $A(\alpha)$ satisfying $\sigma(0) = 0$, $\omega(0) = 1$. In Proposition 3.3 we show that

$$r_\lambda(0, 0) = \sigma_\lambda(0). \tag{3.7}$$

Thus (3.4) holds if and only if (1.3) holds. We shall refer to either condition as the *eigenvalue crossing condition.*

We now formulate the first Hopf theorem.

Theorem 3.1. *Let the system of* ODE's (3.1) *satisfy*:

(H1) *the simple eigenvalue condition* (1.2); *and*
(H2) *the eigenvalue crossing condition* (1.3) (*i.e.*, $\sigma_\lambda(0) \neq 0$).

Then there is a $(k + 1)$*-parameter family of periodic orbits of* (3.1) *bifurcating from the steady-state solution* $u = 0$ *at* $\alpha = 0$.

Theorem 3.1 follows from the above discussion, apart from the verification of (3.7) in Proposition 3.3 below.

The above discussion yields no information about how these periodic solutions depend on the bifurcation parameter λ. To address this issue we consider the power series expansion of the function $\mu(x^2, \alpha')$ in (3.5):

$$\mu(x^2, \alpha') = \mu_0(\alpha') + \mu_2(\alpha')x^2 + \mu_4(\alpha')x^4 + \cdots,$$

where $\mu_0(0) = 0$. If we assume that

$$\mu_2(0) \neq 0, \tag{3.8}$$

then for each λ, (3.3) has precisely one or no solutions x close to the origin with $x \geq 0$ according as

$$\mu_2(0)[\lambda - \mu_0(\alpha')]$$

is positive or negative, respectively. (In other words, the bifurcation is supercritical or subcritical according as μ_2 is positive or negative, respectively.) Now it follows by implicit differentiation of the identity $r(z, (\mu(z, \alpha'), \alpha')) \equiv 0$ that

$$\mu_2(0) = -r_z(0, 0)/r_\lambda(0, 0). \tag{3.9}$$

Thus conditions (3.4) and (3.9) are equivalent to

$$r_z(0, 0) \neq 0, \qquad r_\lambda(0, 0) \neq 0.$$

When $k = 0$, these are precisely the conditions needed to prove that $g(x, \lambda)$ is $\mathbf{Z}_2$-equivalent to the pitchfork bifurcation. Moreover, the pitchfork is $\mathbf{Z}_2$-persistent, so auxiliary parameters have no qualitative effect. Let us formalize this discussion as the second Hopf theorem.

Theorem 3.2. *Let* (3.1) *satisfy hypotheses* (H1) *and* (H2) *of Theorem* 3.1 *and*

(H3) $$r_z(0, 0) \neq 0.$$

Then for each $\alpha' = (\alpha_1, \ldots, \alpha_k)$ *the reduced bifurcation equation* g *is* Z_2 *equivalent to the pitchfork* $\varepsilon x^3 + \delta\lambda x$, *where* $\varepsilon = \operatorname{sgn} r_z(0, 0)$ *and* $\delta = \operatorname{sgn} \sigma_\lambda(0)$.

Proposition 3.3 below gives a formula for $r_z(0, 0)$ in terms of F in (3.1). (More properly, the conjunction of Theorem 3.2 and this formula for $r_z(0, 0)$ should be called the second Hopf theorem.)

The calculations relating derivatives of r to the function F in (3.1) are summarized in the following proposition. In the proposition we use the notation (3.6) for the relevant eigenvalues of $A(\alpha)$ and we continue the convention $z = x^2$. Also, c and d are the eigenvectors of A_0 and A_0^t defined by (2.22) and by (2.26), (2.28), respectively.

Proposition 3.3.

$$\begin{aligned} &\text{(a)} \quad r_\lambda(0, 0) = \sigma_\lambda(0) = \tfrac{1}{2} \operatorname{Re} \bar{d}^t A_\lambda(0) c, \\ &\text{(b)} \quad r_z(0, 0) = \tfrac{1}{4} \operatorname{Re}\{\bar{d}^t \cdot [d^2F(c, b_0) + d^2F(\bar{c}, b_2) + \tfrac{1}{4} d^3F(c, c, \bar{c})]\}, \end{aligned} \tag{3.10}$$

where $b_0, b_2 \in \mathbb{C}^n$ *are defined by*

$$\begin{aligned} &\text{(a)} \quad A_0 b_0 = -\tfrac{1}{2} d^2F(c, \bar{c}), \\ &\text{(b)} \quad (A_0 + 2iI) b_2 = -\tfrac{1}{4} d^2F(c, c). \end{aligned} \tag{3.11}$$

We prove this proposition in subsection (b) below.

In general, fairly difficult calculations are required to apply Theorems 3.1 and 3.2. However, in certain specific cases the calculations simplify. We mention two.

Remark 3.4. (a) If (3.1) is a 2×2 system then

$$\sigma(\alpha) = \tfrac{1}{2} \operatorname{trace} A(\alpha). \tag{3.12}$$

Formula (3.12) follows trivially from the fact that the trace of a matrix is the sum of its eigenvalues.

(b) If $(d^2F)_{0,0} \equiv 0$ then the computation of $r_z(0, 0)$ simplifies substantially. Indeed, from (3.12c)

$$r_z(0, 0) = \tfrac{1}{16} \operatorname{Re} \bar{d}^t \cdot d^3F(c, c, \bar{c}). \tag{3.13}$$

This simplication occurs, for example, if $F(u, \lambda)$ is odd in u.

Let us apply the theorems to the example (1.5) of §1. This example exhibits both of the simplifications listed in Remark 3.4. Recall that in (1.5)

$$F(u, \lambda) = -\begin{pmatrix} \lambda & -1 \\ 1 & \lambda \end{pmatrix} u + |u|^2 u. \tag{3.14}$$

We will show that for this system $r_\lambda(0, 0) = -1 < 0$ and $r_z(0, 0) = 1 > 0$. It then follows that the reduced bifurcation is $\mathbf{Z}_2$-equivalent to $x^3 - \lambda x$, as pictured in Figure 1.3(b). In particular, this system has a unique periodic solution for each $\lambda > 0$ and no periodic solutions when $\lambda < 0$. This conclusion supports the description of the phase portrait given in Figure 1.2.

Let us compute $r_\lambda(0, 0)$ and $r_z(0, 0)$. From (3.14) we see that $A(\lambda) = -\left(\begin{smallmatrix} \lambda & -1 \\ 1 & \lambda \end{smallmatrix}\right)$. Hence by (3.12), $\sigma(\lambda) = -\lambda$. We conclude from (3.10a) that $r_\lambda(0, 0) = -1$. Concerning $r_z(0, 0)$, we compute from (3.14) the multiderivative needed to evaluate (3.13) as follows: If $u, v, w \in \mathbb{R}^2$, then

$$(d^3F)_{0,0}(u, v, w) = 2\begin{pmatrix} 3u_1v_1w_1 + (u_2v_2w_1 + u_2v_1w_2 + u_1v_2w_2) \\ 3u_2v_2w_2 + (u_1v_1w_2 + u_1v_2w_1 + u_2v_1w_1) \end{pmatrix}. \tag{3.15}$$

Here u_1, u_2 are the components of $u \in \mathbb{R}^2$, etc. Using (2.22), (2.26), and (2.28) we find that $c = d = (1, -i)$. Substitution into (3.15) yields $d^3F(c, c, \bar{c}) = 8(1, -i)$. Finally, we conclude from (3.13) that $r_z(0, 0) = 1$.

Remark. Theorems 3.1 and 3.2 do not address stability of the periodic orbits. This is the subject of Hopf's third theorem, which we discuss in §4.

(b) Proof of Proposition 3.3

The proof of this proposition is a calculation similar to the proof of Proposition 2.3 above. Both proofs are based on an analysis of the Liapunov–Schmidt reduction of §2. We assume the reader is familiar with the reduction.

We recall from (2.20) that

$$r(z, \alpha) = p(z, \alpha, \tau(z, \alpha)).$$

Differentiating by the chain rule we find

$$\begin{aligned} &\text{(a)} \quad r_\lambda(0, 0) = p_\lambda(0, 0, 0) + p_\tau(0, 0, 0)\tau_\lambda(0, 0), \\ &\text{(b)} \quad r_z(0, 0) = p_z(0, 0, 0) + p_\tau(0, 0, 0)\tau_z(0, 0). \end{aligned} \tag{3.16}$$

By (2.16c), $p_\tau(0, 0, 0) = 0$, so we may drop the second terms in (3.16). To evaluate the first terms we recall from (2.35a) that

$$p(x^2, \alpha, \tau)x = \phi_1(x, 0, \alpha, \tau), \tag{3.17}$$

where ϕ_1 is the first coordinate of the reduced mapping ϕ: $\ker L \times \mathbb{R}^{k+1} \times \mathbb{R} \to \ker L$; in symbols

$$\phi_1(x, y, \alpha, \tau) = \langle v_1^*, \phi(xv_1 + yv_2, \alpha, \tau)\rangle.$$

(Cf. (2.34).) Differentiating (3.17) and combining with (3.16) we have

$$\begin{aligned} &\text{(a)} \quad r_\lambda(0, 0) = \frac{\partial^2 \phi_1}{\partial\lambda\, \partial x}(0, 0, 0, 0), \\ &\text{(b)} \quad r_z(0, 0) = \frac{\partial^3 \phi_1}{\partial x^3}(0, 0, 0, 0). \end{aligned} \tag{3.18}$$

We shall evaluate the derivatives of ϕ in (3.18) with (VII,1.14).

First consider (3.18a). We have from (VII,1.14e) that

$$\frac{\partial^2 \phi_1}{\partial\lambda\, \partial x} = \langle v_1^*, d(\Phi_\lambda)\cdot v_1 - d^2\Phi(v_1, L^{-1}E\Phi_\lambda)\rangle. \tag{3.19}$$

The operator Φ is defined by (2.5); differentiating (2.5) we see that

$$\Phi_\lambda(u, \alpha, \tau) = \frac{\partial F}{\partial\lambda}(u, \alpha).$$

In particular

$$\Phi_\lambda(0, \alpha, \tau) = 0, \qquad d(\Phi_\lambda)\cdot v_1 = A_\lambda(0)\cdot v_1.$$

Thus (3.19) becomes

$$\frac{\partial^2 \phi_1}{\partial\lambda\, \partial x} = \langle v_1^*, A_\lambda(0)\cdot v_1\rangle. \tag{3.20}$$

Now from (2.23) and (2.27)

$$\begin{aligned} &\text{(a)} \quad v_1 = \operatorname{Re}(e^{is}c) = \tfrac{1}{2}(e^{is}c + e^{-is}\bar{c}), \\ &\text{(b)} \quad v_1^* = \operatorname{Re}(e^{is}d) = \tfrac{1}{2}(e^{is}d + e^{-is}\bar{d}). \end{aligned} \tag{3.21}$$

We substitute these into (3.20) and find

$$r_\lambda(0, 0) = \frac{\partial^2 \phi_1}{\partial\lambda\, \partial x}(0, 0, 0, 0) = \tfrac{1}{2}\operatorname{Re} \bar{d}^t A_\lambda(0)c. \tag{3.22}$$

This verifies the first half of formula (3.10a) in Proposition 3.3.

Remark 3.5. Let us isolate a point in the derivation of (3.22) that will be needed below. When we substitute (3.21) into (3.20) and expand using the bilinearity of the inner product, we get *four* terms. However, two of them vanish, because

$$\langle e^{ims}, e^{ins}\rangle = 0 \quad \text{if} \quad m \neq n. \tag{3.23}$$

To verify the other half of (3.10a) we evaluate the inner product on the right-hand side of (3.20). Let $\Sigma(\alpha) = \sigma(\alpha) - i\omega(\alpha)$ be the eigenvalue (3.6) of $A(\alpha)$, and let $c(\alpha)$ be the associated eigenvector such that $c(0) = c$; thus

$$A(\alpha)c(\alpha) = \Sigma(\alpha)c(\alpha). \tag{3.24}$$

We multiply (3.24) by $\bar{d}^t$, differentiate with respect to λ, and evaluate at $\alpha = 0$; this yields

$$\overline{d^t} A_\lambda(0)c + \bar{d}^t A_0 c_\lambda = \Sigma_\lambda(0)\bar{d}^t c + \Sigma(0)\bar{d}^t c.$$

We claim that second terms on the right and left cancel; indeed, $\Sigma(0) = -i$, while

$$\bar{d}^t A_0 c_\lambda = (\overline{A_0 d})^t c_\lambda = (i\bar{d})^t c_\lambda = -i\bar{d}^t c_\lambda.$$

Canceling these terms we have

$$\bar{d}^t A_\lambda(0)c = \Sigma_\lambda(0)\bar{d}^t c. \tag{3.25}$$

On recalling that $\bar{d}^t c = 2$ and taking real parts, we complete the proof of (3.10a).

We now turn to (3.18b). We have from (VII,1.14c) that

$$\frac{\partial^3 \phi_1}{\partial x^3} = \langle x_1^*, d^3\Phi(v_1, v_1, v_1) + 3d^2\Phi(v_1, W_{11})\rangle, \tag{3.26}$$

where

$$W_{11} = -L^{-1}Ed^2\Phi(v_1, v_1). \tag{3.27}$$

Regarding the first term in (3.26), we substitute formulas (3.21) for v_1 and v_1^* into (3.26) and use the multilinearity to expand into 16 terms, many of which vanish as in Remark 3.5; we find

$$\langle v_1^*, d^3\Phi(v_1, v_1, v_1)\rangle = \tfrac{3}{8}\,\mathrm{Re}\,\bar{d}^t d^3F(c, c, \bar{c}). \tag{3.28}$$

To evaluate the second term in (3.26) we must solve the differential equation

$$\frac{d}{ds} W_{11} + A_0 W_1 = -Ed^2F(v_1, v_1). \tag{3.29}$$

We recall (3.21a) and use the bilearity of d^2F to show

$$d^2F(v_1, v_1) = \tfrac{1}{4}\{e^{2is}d^2F(c, c) + 2d^2F(c, \bar{c}) + e^{-2is}d^2F(\bar{c}, \bar{c})\}. \tag{3.30}$$

Note that $d^2F(v_1, v_1) \in$ range L, since by (3.23)

$$\langle v_j^*, d^2F(v_1, v_1)\rangle = 0, \qquad j = 1, 2.$$

Thus the projection E in (3.29) acts as the identity on $d^2F(v_1, v_1)$. Now (3.29) is an inhomogeneous ODE with constant coefficients. Thus there exists a particular solution of (3.29) of the form

$$b_0 + e^{2is}b_2 + e^{-2is}\bar{b}_2; \tag{3.31}$$

indeed, (3.31) satisfies (3.29) if and only if b_0 and b_2 are given by (3.11). Moreover, (3.31) is orthogonal to v_j^*, so it belongs to range L. Thus $W_{11} = L^{-1}Ed^2\Phi(v_1, v_1)$ is given by (3.31). We substitute (3.31) into (3.26) and simplify as in Remark 3.5 to obtain

$$\langle v_1^*, d^2\Phi(v_1, W_{11})\rangle = \tfrac{1}{2}\,\mathrm{Re}\{\bar{d}^t\cdot[d^2F(c, b_0) + d^2F(\bar{c}, b_2)]\}. \tag{3.32}$$

We derive (3.10b) by combining (3.18b), (3.26), (3.28), and (3.32). □

(c) Another Calculation

In §4 it will be useful to have a formula for $r_{\lambda\lambda}(0, 0)$. This derivative may be calculated by the same techniques as above; it seems natural to include the calculation here.

Proposition 3.6.

$$r_{\lambda\lambda}(0, 0) = \sigma_{\lambda\lambda}(0) = \tfrac{1}{2}\,\mathrm{Re}\,\bar{d}^t[A_{\lambda\lambda}(0)c + 2A_\lambda(0)b_1], \tag{3.33}$$

where $b_1 \in C^n$ *is defined by*

$$(A_0 + iI)b_1 = -[A_\lambda(0) - \Sigma_\lambda(0)]c, \qquad \bar{d}^t b_1 = 0. \tag{3.34}$$

Proof. First we differentiate the formula

$$r(z, \alpha) = p(z, \alpha, \tau(z, \alpha))$$

twice with respect to λ. Recalling that $p(0, 0, \tau) \equiv 0$, we find

$$r_{\lambda\lambda}(0, 0) = p_{\lambda\lambda}(0, 0, 0) + 2p_{\lambda\tau}(0, 0, 0)\tau_\lambda(0, 0). \tag{3.35}$$

We see from differentiation of (3.17) that

$$r_{\lambda\lambda}(0, 0) = \frac{\partial^3\phi_1}{\partial\lambda^2\,\partial x}(0, 0, 0, 0) + 2\tau_\lambda(0, 0)\frac{\partial^3\phi_1}{\partial\tau\,\partial\lambda\,\partial x}(0, 0, 0, 0). \tag{3.36}$$

There is a slight problem here, in that the derivatives in (3.36) are not included on the list (VII,1.14). However, using the fact that $\Phi(0, \alpha, \tau) \equiv 0$ to simplify the calculations, we find that for any two parameters α_l and α_m

$$\frac{\partial^3\phi_1}{\partial\alpha_l\,\partial\alpha_m\,\partial x} = \langle v_1^*, d(\Phi_{\alpha_l\alpha_m})\cdot v_1 + d(\Phi_{\alpha_l})W_{x\alpha_m} + d(\Phi_{\alpha_m})W_{x\alpha_l}\rangle, \tag{3.37}$$

where

$$W_{x\alpha_j} = -L^{-1}Ed(\Phi_{\alpha_j})v_1. \tag{3.38}$$

(We ask the reader to verify this in Exercise 3.1.)

Let us use (3.37) to evaluate the two derivatives of ϕ_1 in (3.36). First, we show that the $\tau\lambda x$-derivative vanishes. By direct calculation

$$\Phi_{\tau\lambda} = 0,$$

so the first term in (3.37) vanishes in this case. Next,

$$W_{x\tau} = L^{-1}E\frac{dv_1}{ds}.$$

From (2.38a), $dv_1/ds = -v_2$. But $Ev_2 = 0$. Thus (3.37) collapses to a single term,

$$\frac{\partial^3\phi}{\partial\tau\,\partial\lambda\,\partial x} = \langle v_1^*, d(\Phi_\tau)W_{x\lambda}\rangle.$$

However $d(\Phi_\tau)\cdot v = dv/ds$, and

$$\left\langle v_1^*, \frac{d}{ds}W_{x\lambda}\right\rangle = -\left\langle \frac{dv_1^*}{ds}, W_{x\lambda}\right\rangle = \langle v_2^*, W_{x\lambda}\rangle = 0.$$

The last equality follows because the derivative $W_{x\lambda}$ belongs to range L and is therefore orthogonal to v_i^*. Thus the $\tau\lambda x$-derivative in (3.36) vanishes.

For the $\lambda\lambda x$-derivative in (3.36), we compute

$$\frac{\partial^3\phi_1}{\partial\lambda^2\,\partial x} = \langle x_1^*, A_{\lambda\lambda}(0)v_1 + 2A_\lambda(0)W_{x\lambda}\rangle, \tag{3.39}$$

where

$$\frac{d}{ds}W_{x\lambda} + A_0W_{x\lambda} = -EA_\lambda(0)v_1. \tag{3.40}$$

We claim that the solution of the ODE (3.40) is

$$W_{x\lambda} = \mathrm{Re}(e^{is}b_1), \tag{3.41}$$

where $b_1 \in \mathbb{C}^n$ is defined by (3.34). On combining (3.41), (3.39), and (3.36) we obtain

$$r_{\lambda\lambda}(0, 0) = \tfrac{1}{2}\,\mathrm{Re}\,\overline{d^t}[A_{\lambda\lambda}(0)c + 2A_\lambda(0)b_1]. \tag{3.42}$$

This proves half of (3.33), modulo the claim (3.41).

To prove (3.41), we write the right-hand side of (3.40) in complex notation as

$$-EA_\lambda(0)v_1 = -\tfrac{1}{2}EA_\lambda(0)\{e^{is}c + e^{-is}\bar{c}\}. \tag{3.43}$$

We substitute in sequence these two terms into (3.40). Substituting the first yields

$$\frac{d}{ds}w + A_0w = -\tfrac{1}{2}E(A_\lambda(0)e^{is}c). \tag{3.44}$$

The projection E is needed to guarantee that (3.44) has a solution, since the operator $d/ds + A_0$ annihilates one function with the spatial dependence e^{is}; viz., $e^{is}c$. E projects onto the range of $d/ds + A_0$ by subtracting off an appropriate multiple of $e^{is}c$; in symbols,

$$E(A_\lambda(0)e^{is}c) = A_\lambda(0)e^{is}c - ke^{is}c$$

for some constants k. Since range $L = \mathbb{R}\{v_1^*, v_2^*\}^\perp$, the constant k is determined by

$$\bar{d}^t(A_\lambda(0) - kI)c = 0.$$

We evaluate k with the following trick. We observe from (3.25) that

$$\bar{d}^t[A_\lambda(0) - \Sigma_\lambda(0)]c = 0;$$

thus $k = \Sigma_\lambda(0) = \sigma_\lambda(0) - i\omega_\lambda(0)$. It now follows that if b_1 is defined by (3.44), then $e^{is}b_1$ solves (3.44). Moreover, the condition $\bar{d}^t b_1 = 0$ guarantees that $e^{is}b_1 \in$ range L. Taking real parts, we obtain the claim (3.40).

Finally, we complete the proof of (3.33) by operating on the eigenvalue relationship (3.24). Specifically, we claim that

$$\begin{aligned} &\text{(a)} \quad \Sigma_{\lambda\lambda}(0)\bar{d}^t c = \bar{d}^t A_{\lambda\lambda}(0)c + 2\bar{d}^t[A_\lambda(0) - \Sigma_\lambda(0)]c_\lambda(0), \\ &\text{(b)} \quad c_\lambda(0) = b_1, \end{aligned} \tag{3.45}$$

where b_1 is defined by (3.34). Equation (3.45a) follows by differentiating (3.24) twice, multiplying by $\bar{d}^t$, and regrouping terms. To derive (3.45b) we differentiate (3.24) once and regroup terms to find that

$$(A_0 - \Sigma(0))c_\lambda(0) = -(A_\lambda(0) - \Sigma_\lambda(0))c. \tag{3.46}$$

Since $\Sigma(0) = -i$, (3.46) is the same equation as (3.34). Thus b_1 is a solution of (3.46). Of course, for any constant k, $b_1 + kc$ is also a solution of (3.46). This lack of uniqueness comes from the fact that any multiple of $c(\alpha)$, say $k(\alpha)c(\alpha)$, would also satisfy (3.24). However, if we require that

$$c(\lambda) = c + \tilde{c}(\lambda) \quad \text{where} \quad \tilde{c}(\lambda) \in \text{range}(A_0 + iI),$$

then (3.45b) holds.

We substitute (3.45b) into (3.45a), take real parts, and use the fact that $\bar{d}^t c = 2$; this yields

$$\sigma_{\lambda\lambda}(0) = \tfrac{1}{2}\,\text{Re}\,\bar{d}^t[A_{\lambda\lambda}(0) + 2[A_\lambda(0) - \Sigma_\lambda(0)]b_1]. \tag{3.47}$$

But $\Sigma_\lambda(0)$ is a scalar and $\bar{d}^t b_1 = 0$. Thus (3.47) provides the missing part of (3.33). □

EXERCISE

3.1. Making use of the fact that $\Phi(0, \alpha, \tau) \equiv 0$ verify the formula for $\partial^3\phi_1/\partial\alpha_l\,\partial\alpha_m\,\partial x$ given in (3.37) and (3.38).

§4. Exchange of Stability

There are three major themes in Hopf bifurcation: existence, uniqueness, and stability of small amplitude periodic solutions. In §3 we discussed existence and uniqueness; here we discuss stability. The main theorem,

Theorem 4.1, states that the stability of the bifurcating periodic solutions can be determined from the reduced function $g(x, \alpha)$ which is defined in Theorem 2.1 above. The result is similar to Theorem I,4.1 concerning the stability of equilibrium solutions of an ODE—similar both as to content and the techniques used in the proof.

We divide this section into four parts. In subsection (a), we formulate Theorem 4.1, our main result, and discuss the principle of exchange of stability. In subsection (b), we summarize the essentials of Floquet theory, which we use to prove Theorem 4.1. We set the context for proving Theorem 4.1 in subsection (c), and we carry out the proof in subsection (d); the former section contains the new ideas not present in Chapter I.

(a) Formulation of the Main Result

Let $u_0(t)$ be a periodic solution of a $n \times n$ autonomous system of ODE,

$$\frac{du}{dt} + F(u) = 0. \tag{4.1}$$

Roughly speaking, u_0 is asymptotically stable if any solution of (4.1) with initial data close to (the orbit of) u_0 actually tends to (the orbit of) u_0 as $t \to \infty$. More formally, define $\mathcal{O}_{u_0} \subset \mathbb{R}^n$, the orbit of u_0, by

$$\mathcal{O}_{u_0} = \{u_0(t) : t \in \mathbb{R}\}.$$

We say that u_0 is *asymptotically stable* if there is an $\varepsilon > 0$ and $\delta > 0$ such that for any solution $u(t)$ of (4.1),

$$\operatorname{dist}(u(0), \mathcal{O}_{u_0}) < \varepsilon \Rightarrow \operatorname{dist}(u(t), \mathcal{O}_{u_0}) < \delta, t > 0$$

and

$$\lim_{t \to \infty} \operatorname{dist}(u(t), \mathcal{O}_{u_0}) = 0.$$

Otherwise we will call u_0 *unstable*.

Consider a $(k + 1)$-parameter family of ODE of the form (4.1),

$$\frac{du}{dt} + F(u, \alpha) = 0. \tag{4.2}$$

As above, $\alpha = (\alpha_0, \ldots, \alpha_k)$, where $\alpha_0 = \lambda$ is the bifurcation parameter. Suppose that for all α, $u = 0$ is an equilibrium solution of (4.2). In this section we strengthen the simple eigenvalue hypothesis (1.2) as follows: If $\zeta_1, \ldots, \zeta_n$ are the eigenvalues of $A(0) = (dF)_{0,0}$, then we suppose that

$$\zeta_1 = i, \qquad \zeta_2 = -i, \qquad \operatorname{Re} \zeta_j > 0, \qquad j = 3, \ldots, n. \tag{4.3}$$

By Theorem 2.1, the small amplitude periodic solutions of (4.2) with period close to 2π are parametrized by solutions of the scalar equation

$$g(x, \alpha) = 0, \tag{4.4}$$

where $g: \mathbb{R} \times \mathbb{R}^{k+1} \to \mathbb{R}$ has the form $g(x, \alpha) = r(x^2, \alpha)x$. In particular, nontrivial solutions of (4.4) satisfy the equation

$$r(x^2, \alpha) = 0. \tag{4.5}$$

The following theorem is the main result of §4.

Theorem 4.1. *Let the linearization of* (4.2) *at* $u = 0$, $\alpha = 0$ *satisfy* (4.3). *Then the periodic solution of* (4.2) *corresponding to a solution* (x, α) *of* $g(x, \alpha) = 0$ *is asymptotically stable if* $g_x(x, \alpha) > 0$ *and unstable if* $g_x(x, \alpha) < 0$.

Let us illustrate the theorem in the nondegenerate case described in Theorem 3.2; i.e., when

$$\begin{aligned} &\text{(a)} \quad r_z(0, 0) \neq 0 \qquad (\mu_2 \neq 0), \\ &\text{(b)} \quad r_\lambda(0, 0) \neq 0 \qquad (\sigma'(0) \neq 0). \end{aligned} \tag{4.6}$$

For simplicity we suppose $k = 0$, so that there are no auxiliary parameters and we write $\alpha = \lambda$. Also for definiteness we shall assume that the equilibrium solution

$$u = 0 \quad \text{is asymptotically stable if} \quad \lambda < 0. \tag{4.7}$$

First let us show that (4.7) implies that

$$r_\lambda(0, 0) < 0. \tag{4.8}$$

By Theorem I,4.1, $u = 0$ is unstable if one of the eigenvalues of $A(\lambda) = (dF)_{0,\lambda}$ has a negative real part. We see from (4.3) that for small λ, only the eigenvalues $\Sigma(\lambda) = \sigma(\lambda) \pm i\omega(\lambda)$ defined by (3.6) could possibly have a negative real part. Since $u = 0$ is asymptotically stable for $\lambda < 0$, no eigenvalue has a negative real part; thus we have

$$\sigma_\lambda(0) \leq 0. \tag{4.9}$$

By Proposition 3.3,

$$\sigma_\lambda(0) = r_\lambda(0, 0) \neq 0. \tag{4.10}$$

Thus (4.7) follows from (4.9) and (4.10).

Consider the periodic solutions of (4.2) associated to nontrivial solutions of (4.4); i.e., to solutions of (4.5). It follows from Theorem 3.2 that these solutions exist for $\lambda > 0$ if $r_z(0, 0) > 0$ (supercritical) and for $\lambda < 0$ if $r_z(0, 0) < 0$ (subcritical). We apply Theorem 4.1 to ascertain the stability of these solutions. Now

$$g_x(x, \lambda) = 2x^2 r_z(x^2, \lambda) + r(x^2, \lambda).$$

Figure 4.1. Exchange of stability in Hopf bifurcation.

By (4.5) the second term here vanishes. In a sufficiently small neighborhood of the origin, $r_z(x^2, \lambda)$ does not change sign. Thus by Theorem 4.1 these solutions are asymptotically stable if supercritical, unstable if subcritical. These two cases are illustrated in Figure 4.1. The phrase "exchange of stability" is often used to describe the fact that nontrivial solutions on one side of the bifurcation point have the same stability as the trivial solution on the other side of the bifurcation point.

Remarks (4.2). (i) The discussion above completes the justification of the phase portrait for (1.5) shown in Figure 1.2.

(ii) The proof that supercritical solutions are stable and subcritical solutions unstable can be extended to the case where $r_z(0, 0) = 0$ but some higher-order z-derivative of r is nonzero. (We still need $r_\lambda(0, 0) < 0$.) See Exercise 4.1.

(b) Floquet Theory

Floquet theory provides a sufficient condition for a periodic solution of an ODE to be asymptotically stable. This condition involves the eigenvalues of a certain matrix obtained from the solution. It is similar in spirit to the condition of linear stability defined in Chapter I, §4, which provides a sufficient condition for an equilibrium solution of an ODE to be asymptotically stable. In this subsection we present the essential results of Floquet theory, in preparation for the proof of Theorem 4.1.

In our discussion of Floquet theory we temporarily supress all parameters in the equation. Let $u(t)$ be a 2π-periodic solution of (4.1). Consider the linear system

$$\frac{dz}{dt}(t) + B(t)z(t) = 0, \tag{4.11}$$

where $B(t) = (dF)_{u(t)}$ is the linearization of F along the periodic solution u. Choose n linearly independent solutions to (4.11), $z_j(t)$, such that $z_j(0) = e_j$, where e_j is the unit vector in the jth direction. Let

$$M_u = (z_1(2\pi) | \cdots | z_n(2\pi)) \tag{4.12}$$

be the $n \times n$ matrix whose columns are $z_j(2\pi)$, $j = 1, \ldots, n$. It follows by linearity that if $z(t)$ is a solution to (4.11) with initial condition $z(0)$, then

$$z(2\pi) = M_u z(0). \tag{4.13}$$

Definition 4.3. The eigenvalues $\gamma_1, \ldots, \gamma_n$ of the matrix M_u defined by (4.12) are called the *Floquet multipliers* of the periodic solution $u(t)$.

Let us show that there always is one Floquet multiplier equal to $+1$. (By convention we will take $\gamma_1 = 1$.) Note from (4.13) that $+1$ is an eigenvalue of M_u if and only if there is a 2π-periodic solution of (4.11). We claim that du/dt is such a 2π-periodic solution of (4.11). To see this, differentiate the relation

$$\frac{du}{dt}(t) + F(u(t)) = 0$$

with respect to t; this yields

$$\frac{d^2u}{dt^2} + B(t)\frac{du}{dt} = 0.$$

Thus (4.11) always admits at least one 2π-periodic solution, so one Floquet multiplier equals $+1$.

The basic result concerning Floquet multipliers is the following.

Proposition 4.4. *Let $\gamma_1 = 1, \gamma_2, \ldots, \gamma_n$ be the Floquet multipliers associated to a 2π-periodic solution of* (4.1).

(a) *If $|\gamma_2|, \ldots, |\gamma_n| < 1$, then $u(t)$ is asymptotically stable.*
(b) *If $|\gamma_j| > 1$ for some j, then $u(t)$ is unstable.*

Remark. If $|\gamma_2|, \ldots, |\gamma_n| \leq 1$ with $|\gamma_j| = 1$ for some j, we will call $u(t)$ *neutrally stable*. In this case it cannot be determined from Floquet theory whether $u(t)$ is asymptotically stable or unstable.

A proof of Proposition 4.4 is given in Chapter 14 of Coddington and Levinson [1955]. Let us discuss the idea behind the proof, as this is relatively simple to explain. Consider a solution $v(t)$ to (4.1) with initial condition $v(0)$ near $\mathcal{O}_u$, the orbit of u. For simplicity, we assume that $v(0)$ is close to $u(0)$. Let the solution $v(t)$ flow for time 2π to $v(2\pi)$. The main point to prove is that $v(2\pi)$ is approximately equal to $u(2\pi) + z(2\pi)$, where z is the solution to the linear system (4.11) with initial condition $z(0) = v(0) - u(0)$. Moreover,

$$u(2\pi) - v(2\pi) \approx z(2\pi) = M_u z(0) = M_u[v(0) - u(0)].$$

Based on this approximation, it can be shown that the solution $v(t)$ decays to the periodic solution $u(t)$ if M_u is a contraction in the directions transverse to $\mathcal{O}_u$. Of course, M_u is such a contraction if $|\gamma_2|, \ldots, |\gamma_n| < 1$. Conversely, if $|\gamma_j| > 1$ for some j, then there is a direction in which M_u is an expanding; the same approximation shows that there is a solution which escapes from any small neighborhood of $\mathcal{O}_u$.

(c) Preliminaries to the Proof of Theorem 4.1

In proving Theorem 4.1, we will want to determine the Floquet multipliers of a periodic solution of (4.2) that is obtained from the Liapunov–Schmidt reduction as in §2. To see what is involved let us consider a small amplitude periodic solution of (4.2), assuming $\alpha \approx 0$. Since $|u(t)|$ is small for all t, we have in (4.11)

$$B(t) = (dF)_{u(t),\alpha} \approx A_0, \tag{4.14}$$

where $A_0 = (dF)_{0,0}$. Let $\zeta_1, \ldots, \zeta_n$ be the eigenvalues of A_0. We now argue heuristically that in this situation the following approximate relationship exists between these eigenvalues of A_0 and the Floquet multipliers of $u(t)$:

$$\gamma_j \approx e^{-2\pi\zeta_j}, \qquad j = 1, \ldots, n. \tag{4.15}$$

To show this, we use (4.14) to approximate (4.11) by the linear system with constant coefficients,

$$\frac{dz}{dt}(t) + A_0 z(t) = 0. \tag{4.16}$$

Let $z_j(t)$ be linearly independent solutions of (4.16) such that $z_j(0) = e_j$. Form a matrix M_0 with $z_j(2\pi)$ as its jth column, as in (4.12). Then for any solution $z(t)$ of (4.16),

$$z(2\pi) = M_0 z(0). \tag{4.17}$$

Now for each eigenvalue ζ_j of A_0, (4.16) has a solution of the form $z(t) = e^{-\zeta_j t}v$, where v is the associated eigenvector. Substituting this $z(t)$ into (4.17) we find

$$M_0 z(0) = z(2\pi) = e^{-2\pi\zeta_j} z(0).$$

Thus for each j, $e^{-2\pi\zeta_j}$ is an eigenvalue of M_0. However, by (4.14), $M_0 \approx M_u$, so the eigenvalues of M_0 will be close to the eigenvalues of M_u. This justifies (4.15), at least heuristically.

Let us examine (4.15) for eigenvalues which satisfy (4.3). From (4.3) we have

$$e^{-2\pi\zeta_1} = e^{-2\pi\zeta_2} = 1, \qquad |e^{-2\pi\zeta_j}| < 1 \quad \text{for} \quad j = 3, \ldots, n.$$

Thus according to (4.15)

$$\gamma_1 = 1, \qquad \gamma_2 \approx 1, \qquad |\gamma_j| < 1 \quad \text{for} \quad j = 3, \ldots, n.$$

(We have inserted explicitly the fact that $\gamma_1 = 1$, although this does not follow from (4.15).) In other words, a small amplitude periodic solution $u(t)$ of (4.2) is asymptotically stable if $|\gamma_2| < 1$, unstable if $|\gamma_2| > 1$.

In the next subsection we shall give a proof of Theorem 4.1 based on these ideas. In the rest of this subsection we develop the mathematics needed to make these ideas rigorous. The next proposition begins this program. In this proposition we exhibit a relationship between the Floquet multipliers of $u(t)$ and the eigenvalues of a linear operator related to $u(t)$; in contrast to (4.15), this relationship is *exact*. Specifically, we recall from §2 the operator $\Phi: C^1_{2\pi} \times \mathbb{R}^{k+1} \times \mathbb{R} \to C_{2\pi}$ from which periodic solutions to (4.2) are constructed; viz.

$$\Phi(u, \alpha, \tau) = (1 + \tau)\frac{du}{ds} + F(u, \alpha). \tag{4.18}$$

Let $u(s)$ be a (2π-periodic) solution of

$$\Phi(u, \alpha, \tau) = 0. \tag{4.19}$$

The proposition relates the spectrum of $d\Phi$ at $u(s)$ to the Floquet multipliers of $u(s)$.

Proposition 4.5. *Let* $u(s)$ *be a* 2π*-periodic solution of* (4.19), *and let* $\gamma_1 = 1, \gamma_2, \ldots, \gamma_n$ *be the Floquet multipliers of* u. *Then the spectrum of* $(d\Phi)_{u,\alpha,\tau}$ *is*

$$\bigcup_{l \in \mathbb{Z}} \bigcup_{j=1}^{n} \left\{(1 + \tau)\left[-\frac{1}{2\pi}\log \gamma_j + li\right]\right\}. \tag{4.20}$$

In words, the spectrum consists of all possible determinations of $\log \gamma_j$ *multiplied by* $-(1 + \tau)/2\pi$.

Proof. A number $\zeta \in \mathbb{C}$ is an eigenvalue of $(d\Phi)_{u,\alpha,\tau}$ if there exists a nonzero function $v(s)$ such that

$$(d\Phi)_{u,\alpha,\tau} \cdot v = \zeta v, \qquad v(2\pi) = v(0). \tag{4.21}$$

On the other hand, we may rephrase the definition of Floquet multipliers as follows: $\gamma \in \mathbb{C}$ is a Floquet multiplier of $u(s)$ if there is a nonzero function $v(s)$ such that

$$(d\Phi)_{u,\alpha,\tau} \cdot v = 0, \qquad v(2\pi) = \gamma v(0). \tag{4.22}$$

For the reader's convenience, we write out the formula for $d\Phi$:

$$(d\Phi)_{u,\alpha,\tau} \cdot v = (1 + \tau)\frac{dv}{ds} + (dF)_{u,\alpha} \cdot v. \tag{4.23}$$

The proposition follows from the fact that solutions of (4.21) and (4.22) may be related, as follows. If $v(s)$ is a solution of (4.21), then

$$\tilde{v}(s) = e^{-(\zeta/(1+\tau))s}v(s)$$

is a solution of (4.22) with Floquet multiplier

$$\gamma = e^{-2\pi\zeta/(1+\tau)}. \tag{4.24}$$

Conversely, if $v(s)$ satisfies (4.22), then for any determination of $\log \gamma$

$$\tilde{v}(s) = e^{-(\log \gamma/2\pi)s} v(s)$$

satisfies (4.21) with eigenvalue

$$\zeta = -\frac{1+\tau}{2\pi} \log \gamma.$$

If the Floquet multipliers $\gamma_1, \ldots, \gamma_n$ are all distinct, then (4.20) follows immediately from the above correspondence between solutions of (4.21) and of (4.22). If some Floquet multipliers are repeated, then some additional analysis is required to show that multiplicities match in (4.20). We leave this task for the reader. □

Since $d\Phi$ is a linear operator on an infinite-dimensional space, we expect $d\Phi$ to have infinitely many eigenvalues. Formula (4.20) shows that this is the case. Nonetheless, the spectrum of $d\Phi$ is still rather uncomplicated. In particular, the spectrum is contained in the union of n lines in the complex plane,

$$\bigcup_{j=1}^{n} \left\{ \zeta \in \mathbb{C} : \operatorname{Re} \zeta = -\frac{1+\tau}{2\pi} \log|\gamma_j| \right\}. \tag{4.25}$$

Proposition 4.5 includes the hypothesis that $\Phi(u, \alpha, \tau) = 0$, but this is not necessary to show that the spectrum of $d\Phi$ is contained in the union of n such lines. Indeed, consider an arbitrary linear $n \times n$ ordinary differential operator $\mathscr{L}$ with 2π-periodic coefficients, say

$$\mathscr{L}u = \frac{du}{ds} + B(s)u. \tag{4.26}$$

We claim that the spectrum of $\mathscr{L}$ is contained in the union of n lines; in symbols

$$\sigma(\mathscr{L}) \subset \bigcup_{j=1}^{n} \{\zeta \in \mathbb{C} : \operatorname{Re} \zeta = \mu_j\}. \tag{4.27}$$

If $\mathscr{L}$ has no multiple eigenvalues, we may prove (4.27) by exhibiting a complete set of eigenfunctions of $\mathscr{L}$ of the form

$$e^{ils} z_j(s); \qquad l \in \mathbb{Z}, \quad j = 1, \ldots, n. \tag{4.28}$$

If $\mathscr{L}$ has repeated eigenvalues, we must modify (4.28) accordingly. In either case, we leave the details of the justification of (4.27) to the reader.

Let us argue that the lines (4.27) depend continuously on the matrix $B(s)$ in (4.26); i.e., the numbers $\{\mu_j\}$ depend continuously on $B(s)$. The essential point here is the following. If

$$\mathscr{L}_i = \frac{d}{ds} + B_i, \qquad i = 1, 2,$$

then $\mathscr{L}_1 - \mathscr{L}_2 = B_1 - B_2$, which is a *bounded* operator. In other words, $\mathscr{L}_2$ is a bounded perturbation of $\mathscr{L}_1$. Thus we may relate the spectrum of $\mathscr{L}_2$ to that of $\mathscr{L}_1$ with relatively elementary methods. (Cf. Chapter 4, §3 of Kato [1976].) Let us formalize this continuity property in the case where it will be of use in proving Theorem 4.1; viz., the spectrum of $d\Phi$.

Proposition 4.6. *For any* $(u, \alpha, \tau) \in C^1_{2\pi} \times \mathbb{R}^{k+1} \times \mathbb{R}$, *the spectrum of* $(d\Phi)_{u,\alpha,\tau}$ *is contained in the union of n lines,* (4.27). *Moreover, these lines depend continuously on* (u, α, τ).

Remark 4.7. Although $\{\mu_j\}$ depend *continuously* on u, α, τ, this dependence on u, α, τ is definitely *not differentiable*. The problem is associated with multiple eigenvalues. Indeed, the numbers μ_j are better behaved than the individual eigenvalues—near a multiple eigenvalue, branch point phenomena may make it impossible to define the eigenvalues as continuous functions of u, α, τ. However, this difficulty can be avoided for the μ_j; because they are real quantities, one may, for example, order them

$$\mu_1 \leq \mu_2 \leq \cdots \leq \mu_n,$$

thereby avoiding difficulties concerning their being multiple-valued.

In connection with Proposition 4.6, let us record the following fact for reference below. If hypothesis (4.3) is satisfied, then for $u = 0$, $\alpha = 0$, $\tau = 0$, the spectrum of $(d\Phi)_{0,0,0}$ is contained in the set (4.27), where

$$\mu_1 = \mu_2 = 0; \qquad \mu_j = \operatorname{Re} \zeta_j > 0, \quad j = 3, \ldots, n. \tag{4.29}$$

Above we suggested that in proving Theorem 4.1 we could test for the stability of a small amplitude periodic solution of (4.2) by inspection of the single Floquet multiplier γ_2. We now establish this rigorously. Let $u(s)$ be a solution of (4.19) with $\|u\|$, α, and τ all small. Now the spectrum of $(d\Phi)_{u,\alpha,\tau}$ is contained in

$$\bigcup_{j=1}^{n} \left\{\zeta : \operatorname{Re} \zeta = -\frac{1+\tau}{2\pi} \log|\gamma_j|\right\}. \tag{4.30}$$

Of course $\gamma_1 = 1$. According to Proposition 4.6, by making $\|u\|$, α, and τ sufficiently small, we can arrange that the lines (4.30) are close to the lines in (4.29); in particular, we can arrange that

$$\log|\gamma_j| < 0 \quad \text{for} \quad j = 3, \ldots, n.$$

Thus $|\gamma_j| < 1$ for $j = 3, \ldots, n$. Therefore, $u(s)$ is asymptotically stable if $|\gamma_2| < 1$, unstable if $|\gamma_2| > 1$.

In the next subsection we use these ideas to prove Theorem 4.1. Specifically, we will show for the periodic solution of (4.2) associated to a solution (x, α) of the reduced equation $g(x, \alpha) = 0$ that

$$|\gamma_2| < 1 \quad \text{if} \quad g_x(x, \alpha) > 0, \qquad |\gamma_2| > 1 \quad \text{if} \quad g_x(x, \alpha) < 0. \tag{4.31}$$

Let us consider a difficulty which arises in this proof. It is natural to attempt to prove (4.31) by mimicing the proof of the corresponding result of Chapter I, Theorem 4.1; viz., to define $\log|\gamma_2|$ as a smooth function of x and α to show that the quotient

$$\frac{\log|\gamma_2|}{g_x(x, \alpha)}$$

defines a C^∞, nonvanishing function near the origin. However, this program runs into trouble right at the start because of the fact that $(d\Phi)_{0,0,0}$ has a double eigenvalue at zero. As discussed in Remark 4.7, this makes the eigenvalues of $d\Phi$ nondifferentiable functions. The resolution of this difficulty comes from the fact that at solutions of $\Phi(u, \alpha, \tau) = 0$, zero is always an eigenvalue of $d\Phi$, with du/ds as the associated eigenfunction. Therefore, we consider the eigenvalue problem for $d\Phi$ on the quotient space

$$C_{2\pi}/\mathbb{R}\left\{\frac{du}{ds}\right\};$$

this makes the double eigenvalue at zero simple. In this way we are able to define a smooth function $\mu(x, \alpha)$ which equals $\log|\gamma_2|$ when $g(x, \alpha) = 0$. In the rest of this subsection we discuss the construction of μ more fully; in subsection (d) we prove Theorem 4.1 by applying the techniques of Chapter I to the quotient μ/g_x.

It is helpful for this discussion to recall the notation used in §2 to solve the equation

$$\Phi(u, \alpha, \tau) = 0 \tag{4.32}$$

with the Liapunov–Schmidt reduction. In Proposition 2.2(c), we decomposed the domain and range of Φ,

$$\begin{aligned} &\text{(a)} \quad C_{2\pi} = \text{range } L \oplus \ker L, \\ &\text{(b)} \quad C^1_{2\pi} = \ker L \oplus M, \end{aligned} \tag{4.33}$$

where $M = (\text{range } L) \cap C^1_{2\pi}$. We invoked the implicit function theorem to determine a mapping $W\colon \ker L \times \mathbb{R}^{k+1} \times \mathbb{R} \to M$ such that

$$E\Phi(v + W(v, \alpha, \tau), \alpha, \tau) \equiv 0, \tag{4.34}$$

where E is projection onto range L. Let v_1, v_2 and v_1^*, v_2^* be the bases for $\ker L$ and $(\text{range } L)^\perp$ given by (2.23) and (2.27), respectively. Then we defined $\tau(x^2, \alpha)$ by solving the equation

$$\langle v_2^*, \Phi(xv_1 + W(xv_1, \alpha, \tau), \alpha, \tau)\rangle = 0, \tag{4.35}$$

and we let

$$g(x, \alpha) = \langle v_1^*, \Phi(xv_1 + W(xv_1, \alpha, \tau(x^2, \alpha)), \alpha, \tau(x^2, \alpha))\rangle. \tag{4.36}$$

This construction led to the following conclusion: If $g(x, \alpha) = 0$, then

$$u = xv_1 + W(xv_1, \alpha, \tau(x^2, \alpha)) \tag{4.37}$$

satisfies (4.32); conversely, up to changes of phase, every small amplitude periodic solution of (4.32) arises through this construction. Let us define a mapping $\Omega\colon \mathbb{R} \times \mathbb{R}^{k+1} \to C^1_{2\pi}$ by the right-hand side of (4.37); viz.,

$$\Omega(x, \alpha) = xv_1 + W(xv_1, \alpha, \tau(x^2, \alpha)). \tag{4.38}$$

Below we shall indicate derivatives with respect to s by a dot; thus $\dot{\Omega} = d\Omega/ds$. By (2.38a)

$$\dot{\Omega} = -xv_2 + \dot{W}. \tag{4.39}$$

Moreover $\dot{W} \in M$, since $M = (\ker L^*)^\perp$ and by (2.38c, d),

$$\left\langle v_j^*, \frac{dW}{ds} \right\rangle = -\left\langle \frac{dv_j^*}{ds}, W \right\rangle = 0.$$

In words, we define $\mu(x, \alpha)$ as the continuously varying eigenvalue of $d\Phi$ at the point $(\Omega(x, \alpha), \alpha, \tau(x^2, \alpha))$ on the quotient space $C_{2\pi}/\mathbb{R}\{\dot{\Omega}\}$, such that $\mu(0, 0) = 0$. Let us write an equation for μ:

$$(d\Phi)_{\Omega(x,\alpha),\alpha,\tau(x^2,\alpha)} \cdot (v_1 + w) + \eta\dot{\Omega} = \mu(v_1 + w). \tag{4.40}$$

Here $w \in M$ and $\eta \in \mathbb{R}$. Equation (4.40) states that $v_1 + w$ is an eigenfunction of $d\Phi$ with eigenvalue μ, modulo an error that is proportional to $\dot{\Omega}$. Note that the projection of the eigenfunction onto $\ker L$ is v_1. By comparison, we see from (4.39) that the projection of the error term $\dot{\Omega}$ onto $\ker L$ is (a multiple of) v_2.

Unfortunately, the error term $\dot{\Omega}$ in (4.40) vanishes in the limit $x \to 0$. Specifically we claim that

$$\dot{\Omega}(0, \alpha) = 0, \qquad \lim_{x\to 0} \frac{\dot{\Omega}(x, 0)}{x} = v_2.$$

The basis for this claim is the fact that

$$W(0, \alpha) = W_x(0, 0) = 0. \tag{4.41}$$

(Cf. (I,3.15).) Indeed, the claim follows on substituting (4.41) into (4.39). We must replace $\dot{\Omega}$ by $\dot{\Omega}/x$ in (4.40) to avoid difficulties as $x \to 0$; i.e., μ is defined by

$$(d\Phi)_{\Omega(x,\alpha),\alpha,\tau(x^2,\alpha)} \cdot (v_1 + w) + \eta(\dot{\Omega}/x) = \mu(v_1 + w). \tag{4.42}$$

The following two lemmas contain the properties of the function $\mu(x, \alpha)$ needed for the proof of Theorem 4.1.

Lemma 4.8. *Equation* (4.42) *defines a* C^∞*-function* $\mu(x, \alpha)$.

Lemma 4.9. *If* $g(x, \alpha) = 0$, *then*

$$\mu(x, \alpha) = -\frac{1 + \tau(x^2, \alpha)}{2\pi} \log|\gamma_2|,$$

where γ_2 *is the second Floquet multiplier of the periodic solution* $u = \Omega(x, \alpha)$ *of* (4.32).

Proof of Lemma 4.8. Equation (4.42) contains two scalar unknowns, μ and η, and one vector unknown, w. We shall write (4.42) as an equation

$$\Psi(\mu, \eta, w; x, \alpha) = 0 \tag{4.43}$$

and solve (4.43) for μ, η, and w by the implicit function theorem. Specifically, let

$$\Psi: \mathbb{R} \times \mathbb{R} \times M \times \mathbb{R} \times \mathbb{R}^{k+1} \to C_{2\pi}, \tag{4.44}$$

be defined by

$$\Psi(\mu, \eta, w; x, \alpha) = (d\Phi)_{\Omega(x,\alpha),\alpha,\tau(x^2,\alpha)} \cdot (v_1 + w) + \eta\dot{\Omega}(x, \alpha)/x - \mu(v_1 + w).$$

Let us compute the derivatives of Ψ at the origin. For the two scalar unknowns we have

$$\frac{\partial\Psi}{\partial\mu} = -v_1, \qquad \frac{\partial\Psi}{\partial\eta} = v_2;$$

for w, since $(d\Phi)_{0,0,0} = L$, we have

$$\text{range}(d_w\Psi) = \text{range}(L \mid M) = \text{range } L.$$

Now by (4.33a),

$$C_{2\pi} = \mathbb{R}\{v_1\} \oplus \mathbb{R}\{v_2\} \oplus \text{range } L.$$

Thus the differential $d\Psi$ at the origin maps the first three factors in (4.44) isomorphically onto $C_{2\pi}$. Therefore, by the implicit function theorem, we may solve (4.43) near the origin for μ, η, and w as smooth functions of x and α. □

PROOF OF LEMMA 4.9. First we show that $\mu(x, \alpha)$ is an eigenvalue of $d\Phi$ at the point $(\Omega(x, \alpha), \alpha, \tau(x^2, \alpha))$ on the whole space $C_{2\pi}$. This is in addition to the eigenvalue zero associated to the eigenvector $\dot{\Omega}(x, \alpha)$. Let $V = v_1 + w$ be the eigenvector of $d\Phi$ on $C_{2\pi}/\mathbb{R}\{\dot{\Omega}\}$, as defined by (4.40). Thus we have two equations

$$\begin{aligned} &\text{(a)} \quad d\Phi \cdot \dot{\Omega} = 0, \\ &\text{(b)} \quad d\Phi \cdot V = \mu V - \eta \dot{\Omega}. \end{aligned} \tag{4.45}$$

If $\mu(x, \alpha) \neq 0$, then

$$d\Phi \cdot \left(V - \frac{\eta}{\mu}\dot{\Omega}\right) = \mu\left(V - \frac{\eta}{\mu}\dot{\Omega}\right);$$

i.e., μ is an eigenvalue of $d\Phi$ associated to the eigenfunction $V - (\eta/\mu)\dot{\Omega}$. If $\mu = 0$, then

$$(d\Phi)^2 V = 0;$$

i.e., zero is a double eigenvalue of $d\Phi$, with V as a generalized eigenfunction.

Now the spectrum of $d\Phi$ is contained in the union of n lines, each associated to $\log|\gamma_j|$ for some j, as in (4.30); $\mu(x, \alpha)$ must lie on one of these n lines. For $j = 3, \ldots, n$, $\log|\gamma_j|$ is bounded away from zero, so $\mu(x, \alpha)$ cannot lie on one of these. Also $\log|\gamma_1| = 0$, and this line is already accounted for by the eigenfunction $\dot{\Omega}$. Thus μ must lie on the line associated to $\log|\gamma_2|$, and the lemma follows. □

(d) Proof of Theorem 4.1

From this point on, our proof of Theorem 4.1 follows the proof of Theorem 4.1 in Chapter I. Specifically, let $g(x, \alpha)$ be defined by (4.4). As discussed above, it suffices to prove that $\mu(x, \alpha)$ and $g_x(x, \alpha)$ have the same sign in some neighborhood of the origin. We do this using Proposition I,4.2 to show that the quotient μ/g_x defines a positive, C^∞-function.

Before starting this, let us record several formulas which occur in the Liapunov–Schmidt reduction; these formulas will be needed for the proof. First, for any function $u \in C_{2\pi}$,

$$u = 0 \quad \text{iff} \quad \langle v_1^*, u\rangle = 0, \qquad \langle v_2^*, u\rangle = 0 \quad \text{and} \quad Eu = 0. \tag{4.46}$$

Next, if $\Omega(x, \alpha)$ is defined by (4.38), then

$$\begin{aligned} &\text{(a)} \quad E\Phi(\Omega(x, \alpha), \alpha, \tau(x^2, \alpha)) \equiv 0, \\ &\text{(b)} \quad \langle v_2^*, \Phi(\Omega(x, \alpha), \alpha, \tau(x^2, \alpha))\rangle \equiv 0, \\ &\text{(c)} \quad g(x, \alpha) = \langle v_1^*, \Phi(\Omega(x, \alpha), \alpha, \tau(x^2, \alpha))\rangle. \end{aligned} \tag{4.47}$$

These equations repeat (4.34), (4.35), and (4.36) in our present notation. Finally, we compute from (4.18) that

$$\Phi_\tau(u, \alpha, \tau) = \dot{u}. \tag{4.48}$$

We now show that

$$g_x(x, \alpha) = 0 \quad \text{implies} \quad \mu(x, \alpha) = 0;$$

i.e., we verify condition (a) in Proposition I,4.2. Suppose that $g_x(x, \alpha) = 0$. From an inspection of (4.40), we see that to show that $\mu(x, \alpha) = 0$ we must prove

$$(d\Phi) \cdot V + \eta\dot{\Omega} = 0 \tag{4.49}$$

for some $\eta \in \mathbb{R}$ and some function V of the form $V = v_1 + w$, where $w \in M$. We will verify (4.49) with

$$V = \Omega_x(x, \alpha), \qquad \eta = 2x\tau_z(x^2, \zeta), \tag{4.50}$$

where $z = x^2$. Specifically, differentiate (4.47c) with respect to x and use (4.48) to show that

$$g_x = \langle v_1^*, d\Phi \cdot \Omega_x + 2x\tau_z\dot{\Omega}\rangle;$$

since $g_x = 0$, we conclude that the left-hand side of (4.49) is orthogonal to v_1^* for the choices (4.50). Similarly, we show that the left-hand side of (4.49) is orthogonal to v_2^* by differentiating (4.47*b*). Finally, we deduce that E applied to the left-hand side of (4.49) vanishes by differentiating (4.47a). Applying (4.46), we see that the left-hand side of (4.49) itself vanishes; in other words, $\mu(x, \alpha) = 0$.

We turn to condition (b) in Proposition I,4.2. As in Chapter I, we can only verify this condition by inserting another parameter into the differential equation. Thus we define

$$\tilde{F}(u, \alpha, \beta) = F(u, \alpha) + \beta u. \tag{4.51}$$

Similarly, let

$$\tilde{\Phi}(u, \alpha, \beta, \tau) = (1 + \tau)\frac{du}{ds} + \tilde{F}(u, \alpha, \beta). \tag{4.52}$$

In general, we write $\tilde{\mu}(x, \alpha, \beta)$, $\tilde{g}(x, \alpha, \beta)$, etc. for the various functions with this extra parameter inserted. We now show that

$$\text{(a)} \quad \mu_\beta(0, 0, 0) > 0, \qquad \text{(b)} \quad g_{x\beta}(0, 0, 0) > 0; \tag{4.53}$$

i.e., we verify condition (b) of Proposition I,4.2. This will complete the proof of Theorem 4.1.

We prove (4.53a) by deducing from (4.42) that

$$\tilde{\mu}(0, 0, \beta) \equiv \beta. \tag{4.54}$$

First observe that

$$\Omega(0, 0, \beta) = 0, \tag{4.55}$$

as this satisfies (4.47a). We see from (4.52) that

$$(d\Phi)_{0,0,\beta,\tau} \cdot V = (d\Phi)_{0,0,0,0} \cdot V + \tau\dot{V} + \beta V. \tag{4.56}$$

Setting $V = v_1$, $\tau = \tilde{\tau}(0, 0, \beta)$, and recalling that $v_1 \in \ker L$, we find that

$$(d\Phi)_{0,0,\beta,\tilde{\tau}(0,0,\beta)} \cdot v_1 - \tilde{\tau}(0, 0,\beta)\dot{v}_1 = \beta v_1. \tag{4.57}$$

We deduce (4.54) by matching terms in (4.57) with (4.42).

Next consider (4.53b). Let

$$\tilde{\sigma}(\alpha, \beta) \pm i\tilde{\omega}(\alpha, \beta)$$

be the eigenvalue of $A(\alpha, \beta) = (dF)_{0,\alpha,\beta}$ that equals $\pm i$ at the origin. It follows from Proposition 3.3(a) that

$$g_{x\beta}(0, 0, 0) = \tilde{\sigma}_\beta(0, 0) \tag{4.58}$$

(Although Proposition 3.3 is only stated for the derivative with respect to λ, the proof shows that the same formula holds for a derivative with respect to any parameter.) It is readily seen from (4.51) that

$$\tilde{\sigma}(0, \beta) \equiv \beta.$$

Thus (4.53b) follows from (4.58). □

Exercise

4.1. Let the reduced bifurcation equation

$$g(x, \lambda) = r(x^2, \lambda)x$$

satisfy the eigenvalue crossing condition $r_\lambda(0, 0) \neq 0$. Assume g has finite $\mathbf{Z}_2$-codimension and show that exchange of stability holds. *Hint:* Show that finite codimension implies that there exists k such that

$$r = \frac{\partial}{\partial z} r = \cdots = \left(\frac{\partial}{\partial z}\right)^{k-1} r = 0$$

at (0, 0) and $(\partial/\partial z)^k r(0, 0) \neq 0$. Use Proposition IV,2.14 to conclude that g is $\mathbf{Z}_2$-equivalent to $h = \varepsilon x^{2k+1} + \delta\lambda x$. Now apply Theorem 4.1 to h.

§5. Degenerate Hopf Bifurcation

In our study of Hopf bifurcation above, we have introduced various hypotheses on the equation; viz.,

(H1) The simple eigenvalue condition (1.2).
(H2) The eigenvalue crossing condition (1.3).
(H3) The condition $r_z(0, 0) \neq 0$ (cf. Theorem 3.2).
(H4) The stability condition (4.3).

(Note that (H4) implies (H1).) By *degenerate Hopf bifurcation* we mean a Hopf bifurcation where one or more of the above assumptions fails. In this section we study cases where (H2) and/or (H3) fail. (The breakdown of (H1) has been much studied in the dynamical systems literature; see for example §7.5 of Guckenheimer and Holmes [1983]. We do not consider this case here.)

We divide this section into two parts. In subsection (a) we discuss the simplest kinds of degenerate Hopf bifurcation on a theoretical level. Our analysis is based on the $\mathbf{Z}_2$-symmetric techniques developed in Chapter VI. In subsection (b) we study a simple model in which a degenerate Hopf bifurcation occurs. (In Case Study 2 we apply our methods to a much more interesting set of equations, the clamped Hodgkin–Huxley equations.)

(a) Theoretical Discussion

Consider a $(k + 1)$-parameter family of $n \times n$ systems of ODE

$$\frac{du}{dt} + F(u, \lambda, \alpha) = 0, \qquad F(0, \lambda, \alpha) \equiv 0. \tag{5.1}$$

Here we return to writing the bifurcation parameter λ explicitly; thus $\alpha = (\alpha_1, \ldots, \alpha_k)$. Suppose $A_0 = (dF)_{0,0,0}$ satisfies hypothesis (H4), and hence, also (H1). By Theorem 2.1, the small amplitude periodic solutions of (5.1) with period near 2π are in one-to-one correspondence with solutions of a scalar equation

$$g(x, \lambda, \alpha) = 0, \qquad x \geq 0, \tag{5.2}$$

where g has the form

$$g(x, \lambda, \alpha) = r(x^2, \lambda, \alpha)x. \tag{5.3}$$

According to (5.3), g is a family of $\mathbf{Z}_2$-symmetric bifurcation problems as defined in Chapter VI. We study degenerate Hopf bifurcation by means of the techniques of Chapter VI. Let us elaborate. Suppose that for $\alpha = 0$, (5.2) exhibits a degenerate singularity of some kind; specifically, suppose that for $\alpha = 0$ (5.2) is $\mathbf{Z}_2$-equivalent to a normal form $h(x, \lambda)$. We regard $g(x, \lambda, \alpha)$ as an unfolding of $g(x, \lambda, 0)$. Let $H(x, \lambda, \beta)$, where $\beta = (\beta_1, \ldots, \beta_l)$ be a universal unfolding of h (in the $\mathbf{Z}_2$-symmetric category). Then there is a mapping A of parameter spaces such that

$$g(\cdot, \cdot, \alpha) \sim H(\cdot, \cdot, A(\alpha)), \tag{5.4}$$

where $\sim$ indicates $\mathbf{Z}_2$-equivalence. Consequently, the small amplitude periodic solutions of (5.1) are enumerated by bifurcation diagrams in the universal unfolding of h. (Cf. Chapter VII, §1(c).)

We extend the above discussion to include stability, as follows. Recall from Theorem 4.1 that the stability of a small amplitude periodic solution of (5.1) is determined by the sign of $g_x(x, \lambda, \alpha)$. Now in (5.4), we claim that at solutions of (5.2), g_x and H_x have the same sign—indeed, this is precisely the content of Lemma 5.1. Thus in the situation described above, the stability of small amplitude periodic solutions of (5.1) may be determined from the universal unfolding H.

Lemma 5.1. *Suppose g and h are $\mathbf{Z}_2$-equivalent bifurcation problems*; i.e., *suppose*

$$g(x, \lambda) = S(x^2, \lambda)h(X(x, \lambda), \Lambda(\lambda)), \tag{5.5}$$

where $S(0, 0) > 0$, $\Lambda'(0, 0) > 0$, $X(x, \lambda) = a(x^2, \lambda)x$, *and* $a(0, 0) > 0$. *If* $g(x_0, \lambda_0) = 0$ *for some* (x_0, λ_0), *then*

$$\operatorname{sgn} g_x(x_0, \lambda_0) = \operatorname{sgn} h_x(X(x_0, \lambda_0), \Lambda(\lambda_0)).$$

PROOF. Let $(x_1, \lambda_1) = (X(x_0, \lambda_0), \Lambda(\lambda_0))$. Since $g(x_0, \lambda_0) = 0$ it follows that $h(x_1, \lambda_1) = 0$. Now we differentiate (5.5) with respect to x and evaluate at (x_0, λ_0), obtaining

$$g_x(x_0, \lambda_0) = S(x_0^2, \lambda_0)h_x(x_1, \lambda_1)X_x(x_0, \lambda_0).$$

The lemma follows on observing $S(0, 0) > 0$ and $X_x(0, 0) > 0$. □

With the above discussion in mind, let us recall the simplest degenerate $\mathbf{Z}_2$-symmetric bifurcation problems from Chapter VI; i.e., those of $\mathbf{Z}_2$-codimension one. We reduce the number of cases to consider by supposing that the steady-state solution is stable subcritically. This implies that in (5.3), $r(0, \lambda) > 0$ for $\lambda < 0$. Given this restriction, there are four Z_2-bifurcation problems with $\mathbf{Z}_2$-codimension one. Their normal forms are:

$$\begin{aligned} &\text{(a)} \quad x^3 + \lambda^2 x, \\ &\text{(b)} \quad -x^3 + \lambda^2 x, \\ &\text{(c)} \quad x^5 - \lambda x, \\ &\text{(d)} \quad -x^5 - \lambda x. \end{aligned} \tag{5.6}$$

The corresponding bifurcation diagrams are given in Figure 5.1. Unstable solutions are indicated by dashed lines.

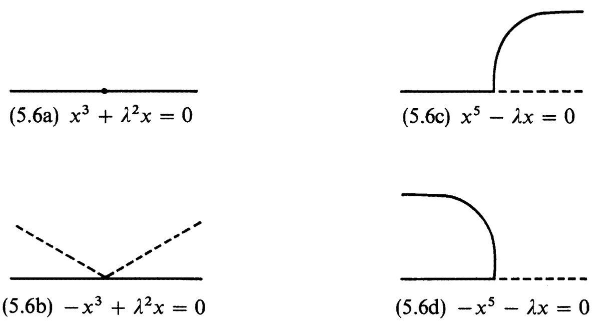


Figure 5.1. $\mathbf{Z}_2$-bifurcation problems with codimension one.

Normal forms (5.6a, b) represent the simplest degenerate Hopf bifurcations where hypothesis (H2) fails; in other words, the eigenvalues do not cross the imaginary axis with nonzero speed (i.e., $\sigma_\lambda(0) = 0$). We see from (5.6a) in Figure 5.1 that there need not exist a family of periodic solutions in degenerate Hopf bifurcations. This example shows the necessity of hypothesis (H2) in Theorem 2.1. Normal form (5.6b) gives an example where supercritical periodic solutions are unstable. This example shows the importance of hypothesis (H2) for exchange of stability (cf. §4(a)). Normal forms (5.6c, d) represent the simplest degenerate Hopf bifurcations when hypothesis (H3) fails. These examples exhibit interesting behavior only when they are perturbed, as in their universal unfoldings.

Universal unfoldings for the normal forms (5.6) are as follows:

$$\begin{aligned} &\text{(a)} \quad x^3 + \lambda^2 x + \alpha x, \\ &\text{(b)} \quad -x^3 + \lambda^2 x + \alpha x, \\ &\text{(c)} \quad x^5 - \lambda x + \alpha x^3, \\ &\text{(d)} \quad -x^5 - \lambda x + \alpha x^3. \end{aligned} \tag{5.7}$$

The perturbed bifurcation diagrams are shown in Figure 5.2.

In any of the diagrams of Figure 5.2, for $\alpha \neq 0$ only nondegenerate Hopf bifurcations occur from the steady-state solution. However, the unfolding of a degenerate Hopf bifurcation predicts some interesting global behavior. To elaborate on this, let's focus on two examples, (5.7a, c).

As shown in Figure 5.2, when $\alpha < 0$, (5.7a) exhibits two nondegenerate Hopf bifurcations, the first supercritical and the second subcritical. The interesting feature in this diagram is that the two branches of periodic solutions bifurcating from the trivial solution in fact coincide. It is difficult to derive this behavior with classical techniques. (The example of subsection (b) has this behavior.)

As shown in Figure 5.2, for (5.7c) a branch of periodic solutions starts off subcritically at a nondegenerate Hopf bifurcation and regains stability at a limit point bifurcation. For a range of λ values (between λ_0 and λ_1 in Figure 5.2)

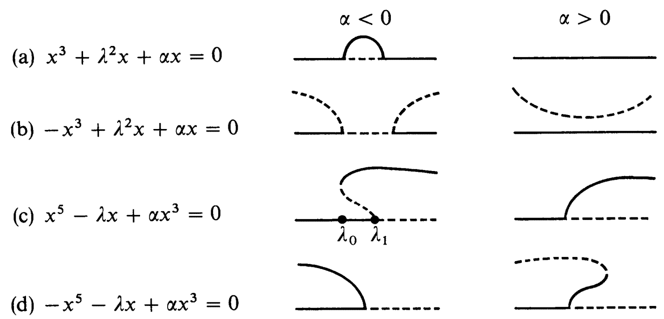


Figure 5.2. Perturbations of $\mathbf{Z}_2$-codimension one bifurcation problems.

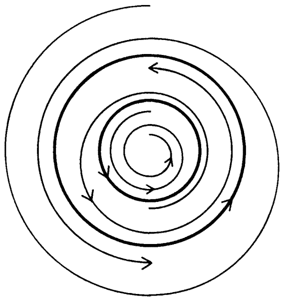


Figure 5.3. Possible phase portrait associated with $x^5 - \lambda x + \alpha x^3 = 0, \alpha < 0$.

there are two periodic solutions, one unstable and one stable. In Figure 5.3 we show a possible phase portrait (of a 2×2 system) for λ fixed between λ_0 and λ_1. In this case, the bifurcation behavior for quasistatic variation of the parameter λ is interesting. As λ increases the steady state solution loses stability at λ_1, and there is a jump to a finite amplitude periodic solution. If λ is then decreased the periodic solution persists to λ_0, where there is a jump to the steady-state solution. Note that at $\lambda = \lambda_0$ the two periodic solutions merge, and they disappear for $\lambda < \lambda_0$.

We solved the recognition problem for these singularities in Chapter VI, §3. We recall the answers here. Let $g(x, \lambda) = r(x^2, \lambda)x$ and let $z = x^2$.

(a) Suppose $r = r_\lambda = 0$ and $\varepsilon = \operatorname{sgn} r_z \neq 0$, $\delta = \operatorname{sgn}(r_{\lambda\lambda}) \neq 0$. Then g is $\mathbf{Z}_2$-equivalent to $\varepsilon x^3 + \delta\lambda^2 x$.

(b) Suppose $r = r_z = 0$ and $\varepsilon = \operatorname{sgn} r_{zz} \neq 0$, $\delta = \operatorname{sgn}(r_\lambda) \neq 0$. Then g is $\mathbf{Z}_2$-equivalent to $\varepsilon x^5 + \delta\lambda x$.

Thus, in order to find the normal forms (5.6a, b) we must compute $r_{\lambda\lambda}$. Proposition 3.6 gives a formula for this coefficient. Our example of degenerate Hopf bifurcation in subsection (b) will lead to the normal form $x^3 + \lambda^2 x$, (5.6a).

In order to find the normal forms (5.6c, d) we must compute r_{zz}. The computation of this coefficient is rather lengthy. The calculation of r_{zz} was first given in Hassard and Wan [1978] using center manifold theory. An alternative calculation using the Liapunov–Schmidt reduction is given in Golubitsky and Langford [1981]. The formula for r_{zz} resulting from this calculation is sufficiently complicated so as to make computing r_{zz} by hand virtually impossible, except in the simplest cases.

From a theoretical point of view, the most interesting degenerate bifurcation is the simplest singularity where *both* hypotheses (H2) and (H3) fail; that is, where $r_z = r_\lambda = 0$. Here we find a singularity with $\mathbf{Z}_2$-codimension three, one modal parameter, and topological $\mathbf{Z}_2$-codimension two. This singularity was analyzed in detail in Chapter VI, §5. Moreover, this singularity serves as an organizing center for the analysis by Labouriau [1983] of the clamped Hodgkin–Huxley equations. (Cf. Case Study 2.)

(b) A Model System from the Study of Glycolysis

The following system of ODE's has been suggested as a simple model to explain certain oscillations observed in experimental studies of glycolysis. (See Tyson and Kauffman [1975] or Ashkenazi and Othmer [1978].) We analyze this system, looking for degenerate Hopf bifurcations. This system models a product activated reaction. In scaled concentration variables X and Y it reads

$$\begin{aligned} \frac{dX}{ds} - \lambda + \kappa X + XY^2 &= 0, \\ \frac{dY}{ds} - \kappa X + Y - XY^2 &= 0, \end{aligned} \tag{5.8}$$

where $\lambda > 0$ is the bifurcation parameter representing the feed rate and κ is a low-reaction rate $(0 < \kappa < 1)$. This equation has the unique steady state

$$X_0 = \frac{\lambda}{\lambda^2 + \kappa}, \qquad Y_0 = \lambda.$$

Defining new variables $u_1 = X - X_0$, $u_2 = Y - Y_0$ and rescaling time by $t = (\lambda^2 + \kappa)^{1/2} s$ yield the system

$$\frac{du}{dt} + A(\lambda, \kappa)u + f(u)\begin{pmatrix} 1 \\ -1 \end{pmatrix} = 0, \tag{5.9}$$

where

$$A(\lambda, \kappa) = \frac{1}{\sqrt{\lambda^2 + \kappa}} \begin{pmatrix} \lambda^2 + \kappa & \dfrac{2\lambda^2}{\lambda^2 + \kappa} \\ -(\lambda^2 + \kappa) & \dfrac{\kappa - \lambda^2}{\lambda^2 + \kappa} \end{pmatrix}, \tag{5.10}$$

and

$$f(u) = \frac{1}{\sqrt{\lambda^2 + \kappa}} \left(2\lambda u_1 u_2 + \frac{\lambda}{\lambda^2 + \kappa} u_2^2 + u_1 u_2^2 \right). \tag{5.11}$$

We calculate

$$\begin{aligned} &\det A(\lambda, \kappa) = 1, \\ &\operatorname{tr} A(\lambda, \kappa) = (\lambda^2 + \kappa)^{-3/2}[\lambda^4 + (2\kappa + 1)\lambda^2 + \kappa^2 + \kappa]. \end{aligned} \tag{5.12}$$

Therefore, the linearization $A(\lambda, \kappa)$ has eigenvalues $\pm i$ when $\operatorname{tr} A(\lambda, \kappa) = 0$; i.e., when

$$\lambda^4 + (2\kappa - 1)\lambda^2 + \kappa(\kappa + 1) = 0. \tag{5.13}$$

When $\kappa < \frac{1}{8}$, equation (5.13) has two positive real roots, denoted by λ_+ and λ_-, which satisfy

$$\lambda_\pm^2 = \tfrac{1}{2}[1 - 2\kappa \pm \sqrt{1 - 8\kappa}]. \tag{5.14}$$

These correspond to Hopf bifurcations of (5.9). When $\kappa > \frac{1}{8}$, (5.13) has no real solutions. We assume for now that $\kappa < \frac{1}{8}$.

Note that (5.9) is a 2×2 system. According to Remark 3.4(a), the speed at which its eigenvalues cross the imaginary axis is given by

$$\sigma_\lambda(\lambda_\pm, \kappa) = \tfrac{1}{2}\frac{\partial}{\partial\lambda}\operatorname{tr} A(\lambda, \kappa)|_{\lambda=\lambda_\pm} = 2\lambda_\pm(\lambda_\pm^2 + \kappa)^{-3/2}[2\lambda_\pm^2 + 2\kappa - 1].$$

We compute from (5.14) that

$$\begin{aligned} &\text{(a)} \quad \operatorname{sgn} \sigma_\lambda(\lambda_+, \kappa) = \operatorname{sgn}(\sqrt{1 - 8\kappa}) > 0, \\ &\text{(b)} \quad \operatorname{sgn} \sigma_\lambda(\lambda_-, \kappa) = \operatorname{sgn}(-\sqrt{1 - 8\kappa}) < 0. \end{aligned} \tag{5.15}$$

Thus the steady-state solution goes from stable to unstable via a Hopf bifurcation at λ_-, and then it regains stability via a second Hopf bifurcation at λ_+. Calculations by Ashkenazi and Othmer [1978] show that $r_z > 0$ for both of these Hopf bifurcations.

By the second Hopf theorem (Theorem 3.2), the periodic solutions which bifurcate at λ_- are supercritical; according to Theorem 4.1, these solutions are stable. Similarly, the periodic solutions which bifurcate at λ_+ are subcritical and also stable. This leads to the (incomplete) bifurcation diagram Figure 5.4. This figure immediately suggests the question: Do the two branches of periodic solutions in Figure 5.4 meet one another? Besides the figure, there is another piece of related evidence. As $\kappa \to \frac{1}{8}$ from below, $\sigma_\lambda(\lambda_\pm, \kappa) \to 0$. This affects the argument above in which we used the nondegenerate Hopf theory to obtain Figure 5.4. Specifically, the nondegenerate theory predicts that the bifurcating periodic solutions exist in a domain which *shrinks to a point as* $\kappa \to \frac{1}{8}$.

We shall prove that the two branches in Figure 5.4 do in fact meet one another, at least for κ close to $\frac{1}{8}$. We do this as follows. Let

$$g(x, \lambda, \kappa) = 0 \tag{5.16}$$

be the reduced bifurcation equation obtained from (5.9). In Lemma 5.2 below we will prove that for $\kappa = \frac{1}{8}$, (5.16) has a singularity which is $\mathbf{Z}_2$-equivalent to the normal form (5.6a). Now (5.16) is a one-parameter unfolding of $g(x, \lambda, \frac{1}{8})$. As such it may be factored through the universal unfolding (5.7a). This shows that for κ close to $\frac{1}{8}$ we may obtain bifurcation diagrams for (5.16) from Figure 5.2(a). In particular, comparing Figures 5.4

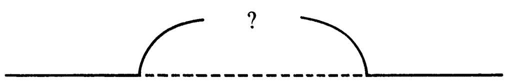


Figure 5.4. Periodic solution structure for (5.9) with $\kappa < \frac{1}{8}$.

and 5.2(a), we see that the two branches in Figure 5.4 must meet one another.

Speaking philosophically, we are able to derive this conclusion with singularity theory methods because these methods yield results that are valid on an open neighborhood in x, λ, κ-space of the bifurcation point. By contrast, the straightforward application of the nondegenerate Hopf theory yields results that are only valid on disjoint neighborhoods in x, λ-space of $(0, \lambda_-)$ and $(0, \lambda_+)$; these neighborhoods shrink to a point as $\kappa \to \frac{1}{8}$.

Remark. The universal unfolding (5.7a) has two bifurcation points if $\alpha < 0$ and none if $\alpha > 0$. Equation (5.16) has two if $\kappa < \frac{1}{8}$ and none if $\kappa > \frac{1}{8}$. Thus when (5.16) is factored through the universal unfolding (5.7a), $\kappa < \frac{1}{8}$ corresponds to $\alpha < 0$, and $\kappa > \frac{1}{8}$ to $\alpha > 0$. It is not difficult to show that $d\alpha/d\kappa$ is nonzero, and hence positive. This is related to the recognition problem for universal unfoldings in the $\mathbf{Z}_2$-symmetric context—specifically, (5.16) provides a universal unfolding of $g(x, \lambda, \frac{1}{8})$. However, we did not address this problem in Chapter VI, and we do not pursue it here.

Lemma 5.2. *If* $\kappa = \kappa_c = \frac{1}{8}$, *then* (5.16) *has a singularity at* $x = 0$, $\lambda = \lambda_c = \sqrt{\frac{3}{8}}$ *which is* $\mathbf{Z}_2$*-equivalent to* (5.6*a*).

Proof. The recognition problem for (5.6a) is solved in (5.8a). To prove the lemma, we must show that for $\kappa = \kappa_c$

$$r_\lambda = 0, \qquad r_{\lambda\lambda} > 0, \qquad r_z > 0 \quad \text{at} \quad (x, \lambda, \kappa) = (0, \lambda_c, \kappa_c),$$

where $r(x, \lambda, \kappa)$ is defined by $g(x, \lambda, \kappa) = r(x^2, \lambda, \kappa)x$. We already know that $r_\lambda = 0$; we will show that

$$\text{(a)} \quad r_{\lambda\lambda} = \frac{3}{\sqrt{8}}, \qquad \text{(b)} \quad r_z = \frac{3}{16}\sqrt{2}. \tag{5.17}$$

Calculation of $r_{\lambda\lambda}(0, \lambda_c, \kappa_c)$: We recall from Proposition 3.6 that

$$r_{\lambda\lambda}(0, \lambda_c, \kappa_c) = \sigma_{\lambda\lambda}(\lambda_c, \kappa_c).$$

Since $A(\lambda, \kappa)$ is a 2×2 matrix,

$$\sigma(\lambda, \kappa) = \tfrac{1}{2} \operatorname{tr} A(\lambda, \kappa).$$

We differentiate (5.12) and substitute $\lambda_c = \sqrt{\frac{3}{8}}$, $\kappa_c = \frac{1}{8}$ to show $r_{\lambda\lambda}(0, \lambda_c, \kappa_c) = 3/\sqrt{8}$.

Calculation of $r_z(0, \lambda_c, \kappa_c)$: Following Proposition 3.3, we calculate r_z in four steps:

(i) we compute c and d;
(ii) we find the multi-derivatives d^2F and d^3F;
(iii) we solve (3.11); and
(iv) we combine the various pieces.

(i) We have from (5.12)

$$A_0(\lambda_c, \kappa_c) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 3 \\ -1 & -1 \end{pmatrix}.$$

Using the normalizations (2.22) and (2.28) we find

$$c = \frac{1}{\sqrt{2}}(1 - \sqrt{2}i, -1), \qquad d = -i(1, 1 - \sqrt{2}i).$$

(ii) Evaluating (5.11) at λ_c, κ_c yields

$$f(u) = \sqrt{3}(u_1 u_2 + u_2^2) + \sqrt{2}u_1 u_2^2.$$

Hence

$$d^2F(\xi, \eta) = \sqrt{3}(\xi_1\eta_2 + \xi_2\eta_1 + 2\xi_2\eta_2)\begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

$$d^3F(\xi, \eta, \zeta) = 2\sqrt{2}(\xi_1\eta_2\zeta_2 + \xi_2\eta_1\zeta_2 + \xi_2\eta_2\zeta_1)\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

(iii) We now compute that

$$d^2F(c, \bar{c}) = 0, \qquad d^2F(c, c) = \sqrt{6}i\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

$$d^3F(c, c, \bar{c}) = (3 - \sqrt{2}i)\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

It follows from the equations (3.11) that

$$b_0 = 0, \qquad b_2 = \frac{1}{\sqrt{6}}\left(-1 + \frac{1}{\sqrt{2}}i, 1\right).$$

(iv) We evaluate the terms in (3.10)

$$d^2F(c, b_0) = 0, \qquad d^2F(\bar{c}, b_2) = -\frac{3\sqrt{2}}{4}i(1, -1).$$

Now $\bar{d}^t \cdot (1, -1) = \sqrt{2}$. Substituting into (3.10) we obtain

$$r_z(0, \lambda_c) = \frac{3}{16}\sqrt{2}. \qquad \square$$

Bibliographical Comments

The literature on Hopf bifurcation is extensive; in the introduction to this chapter we mentioned several recent texts.

Broadly speaking, the occurrence Hopf bifurcation may be established either by a Liapunov–Schmidt reduction or by a center manifold reduction.

The latter approach generalizes to the study of more degenerate local phase portraits. This approach is developed in Arnold [1983], Guckenheimer and Holmes [1983], Carr [1981], and Chow and Hale [1982]. As we indicated in §5, the former approach generalizes to the study of degenerate Hopf bifurcation and multiple periodic orbits. Studies of degenerate Hopf bifurcation using the Liapunov–Schmidt approach may be found in Chafee [1968, 1978], Flocherzi [1979] and Vanderbauwhede [1982], among others. Combining $\mathbf{Z}_2$-equivariant singularity theory with the Liapunov–Schmidt method first occurs in Golubitsky and Langford [1981]. See also Labouriau [1983].

Hopf bifurcation may be generalized to PDE using infinite dimensional techniques. Cf. Marsden and McCracken [1976], Kielhofer [1979], and Henry [1981].

Finally, a further generalization concerns bifurcation of periodic orbits of a system of ODE or PDE which possesses additional symmetry. This topic will be covered in detail in Volume II. Cf. Ruelle [1973], Sattinger [1983], and Golubitsky and Stewart [1984].

CASE STUDY 2

The Clamped Hodgkin–Huxley Equations

The clamped Hodgkin–Huxley equations are a 4×4 system of nonlinear ODE that model electrical activity in the giant axon of a squid under certain controlled experimental conditions. These equations have periodic solutions which bifurcate from an equilibrium solution as a parameter is varied. The equations also contain several auxiliary parameters, and the dependence of the periodic solutions on the various parameters is rather involved. In this Case Study we discuss these bifurcation phenomena with an eye towards illustrating how singularity theory can contribute to understanding such problems. In contrast to the other two Case Studies, additional work is required to complete the analysis of the problem considered in this Case Study. To rephrase this more positively, the present Case Study may be more stimulating than the other two in that it suggests new questions for research, including some important theoretical issues.

This Case Study is divided into four subsections. In subsection (a) we discuss the derivation of the clamped Hodgkin–Huxley equations and in subsection (b) we present the basic information concerning Hopf bifurcations in these equations. We apply singularity theory methods in the last two subsections—to the basic bifurcation phenomena in subsection (c) and to a degnerate Hopf bifurcation in subsection (d).

(a) Origins of the Equations

In this initial subsection we briefly discuss the physical origins of the clamped Hodgkin–Huxley equations (equations (C2.7) below). Although the rest of the Case Study does not depend explicitly on this discussion, we suspect that the subsequent analysis will be more meaningful if the reader is familiar with

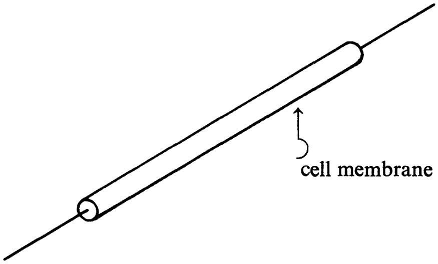


Figure C2.1. Schematic of the giant axon.

the origins of the equation. We refer to Rinzel [1978], Cronin [1981], or Labouriau [1983] for more detailed information on the derivation of the equations.

As we stated above, the clamped Hodgkin–Huxley equations model electrical activity in the giant axon of a squid under certain controlled experimental conditions. We have sketched such an axon in Figure C2.1. The axon is typically a couple of centimeters in length and 0.5 millimeters in diameter. The electrical properties of the axon are primarily determined by the cell membrane, which is a thin membrane (10^{-6} cm) separating the interior of the cell from the exterior. The reason the membrane is so important is the following. The interior and exterior of the cell have relatively low electrical resistance because they contain many charged ions; by contrast the membrane has a very high resistance. Thus the electrical potential is approximately constant throughout the interior of the cell and throughout the exterior of the cell. However, a significant potential difference may develop across the membrane. (*Remark*: In reality the resistance of the interior of the cell is not completely negligible—although the conductivity per unit volume of the material there is large, the long, thin geometry of this region leads to a significant resistance over the length of the axon. However, the *clamped* Hodgkin–Huxley equations describe a situation where the resistance inside the cell is *made* negligible by inserting a thin metal wire along the axis of the axon. Without this wire the potential difference across the membrane varies significantly along the length of the axon. This variation in the potential is related to the main function of the axon; i.e., the propagation of an electrical signal along its length. Specifically, a PDE is needed to describe the situation where the potential varies, and the traveling wave solutions of this PDE model signal propagation. In this Case Study, however, we consider only the clamped case; i.e., we assume that the electrical potential inside the axon is constant, so that the potential difference across the membrane does not vary along the length of the axon.)

Considered as a circuit element, the cell membrane acts both as a capacitor and as a resistor. It acts as a capacitor in that charged ions may accumulate on its surface and cause a potential difference, exactly as in a parallel plate capacitor. It acts as a resistor in that a small number of ions may pass through

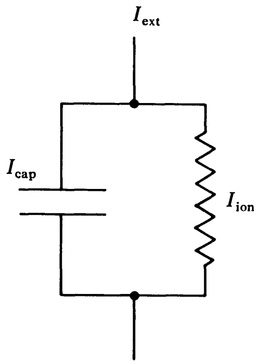


Figure C2.2. Schematic circuit diagram for the cell membrane.

the membrane, the number depending on the potential difference. These two functions are indicated schematically in Figure C2.2. As indicated by the figure, the total external current may be decomposed into a capacitative current and an ionic current; in symbols,

$$I_{ext} = I_{cap} + I_{ion}. \tag{C2.1}$$

In its function as a capacitor, the membrane acts as a linear circuit element; i.e., the capacitative current may be represented

$$I_{cap} = C\frac{dV}{dt}, \tag{C2.2}$$

where V is the potential difference across the membrane and C is the (constant) capacitance of the membrane. (C is the total capacitance, not the capacitance per unit area; this choice is more natural for the clamped case that we are dealing with here.) By contrast, in its function as a resistor, the behavior of the membrane is exceedingly nonlinear.

Let us discuss the function of the membrane as a resistor more fully. This behavior is extremely complex; indeed, characterizing this behavior was perhaps the most important contribution of Hodgkin–Huxley [1952], and they received the Nobel prize for this work. Primarily the ionic current is due to sodium and potassium ions passing through the membrane. Hodgkin and Huxley decomposed the ionic current into three terms

$$I_{ion} = I_{Na} + I_K + I_L, \tag{C2.3}$$

where I_L (L for leakage) measures the (small) current due to all other species. Each current on the right in (C2.3) depends linearly on the potential difference V across the membrane; in symbols

$$\begin{aligned} &\text{(a)} \quad I_{Na} = g_{Na}(V - V_{Na}), \\ &\text{(b)} \quad I_K \ = g_K(V - V_K), \\ &\text{(c)} \quad I_L \ = \bar{g}_L(V - V_L), \end{aligned} \tag{C2.4}$$

where V_{Na}, V_K, and V_L define the voltage levels at which no current will flow in the indicated species and g_{Na}, g_K, $\bar{g}_L$ are the conductances of the indicated species. In (C2.4) the quantities V_{Na}, V_K, V_L, and $\bar{g}_L$ are constants; by contrast, g_{Na} and g_K are time dependent, within the limits

$$\begin{aligned} &\text{(a)} \quad 0 \le g_{Na} \le \bar{g}_{Na}, \\ &\text{(b)} \quad 0 \le g_K \le \bar{g}_K, \end{aligned} \tag{C2.5}$$

where $\bar{g}_{Na}$ and $\bar{g}_K$ are constants. Hodgkin and Huxley described the evolution in time of g_{Na} and g_K as follows. They introduced three auxiliary variables M, N, and H such that

$$\begin{aligned} &\text{(a)} \quad g_{Na} = \bar{g}_{Na} M^3 H, \\ &\text{(b)} \quad g_K = \bar{g}_K N^4. \end{aligned} \tag{C2.6}$$

The variables M, N, and H are dimensionless and lie in the interval [0, 1], so that (C2.5) follows trivially from (C2.6). The evolution in time of each variable M, N, and H is governed by an empirical differential equation (of first order), and the evolution of g_{Na} and g_K may in turn be obtained from that of M, N, and H through (C2.6). (*Remark*: There is an intuitive model which may make the introduction of M, N, and H seem less arbitrary. In this model one supposes that the sodium and potassium ions pass through certain holes or ports in the membrane, which may be open or closed. The maximum conductivities $\bar{g}_{Na}$ and $\bar{g}_K$ in (C2.5) occur if all these hypothetical ports are open. By virtue of (C2.6), the variables M, N, and H characterize the fraction of hypothetical ports that are open. It should be noted, however, that there is little direct experimental evidence supporting this model.)

Let us now synthesize the above information by presenting the clamped Hodkin–Huxley equations. There are four unknowns, the potential difference V across the membrane and the three variables M, N, and H; the temperature T (measured in degrees Celsius) appears as an auxiliary parameter via the function $\Phi(T) = 3^{(T-6.3)/10}$. The equations are fourfold:

$$\begin{aligned} &\text{(a)} \quad \frac{dV}{dt} = \frac{1}{C}\{\bar{g}_{Na} M^3 H(V - V_{Na}) + \bar{g}_K N^4 (V - V_K) + \bar{g}_L (V - V_L) - I_{ext}\} \\ &\text{(b)} \quad \frac{dM}{dt} = -\frac{\Phi(T)}{\tau_M(V)}(M - M_\infty(V)), \\ &\text{(c)} \quad \frac{dN}{dt} = -\frac{\Phi(T)}{\tau_N(V)}(N - N_\infty(V)), \\ &\text{(d)} \quad \frac{dH}{dt} = -\frac{\Phi(T)}{\tau_H(V)}(H - H_\infty(V)). \end{aligned} \tag{C2.7}$$

The first equation here expresses the equality (C2.1), where we have used the representations (C2.2) for I_{cap} and (C2.3) for I_{ion}, combining with (C2.4) and

(C2.6) in the latter case. The last three equations are the empirical relations deduced by Hodgkin–Huxley. These three equations all have the form

$$\frac{dY}{dt} = -\frac{1}{\tau}(Y - Y_\infty). \tag{C2.8}$$

If τ and Y_∞ are constants, solutions of (C2.8) decay exponentially to the equilibrium value Y_∞ at the characteristic rate τ^{-1}. In the Hodgkin–Huxley system, however, equations (C2.7b, c, d) are coupled to (C2.7a) in that the equilibrium values and characteristic rates depend on the potential difference V. (The rate constants also depend on temperature through the scaling factor $\Phi(T)$.) For our purposes it does not matter exactly how the equilibrium values and the characteristic rates depend on V; rather than list this information here, we refer the reader to Rinzel [1978] or Labouriau [1983]. Similarly, with one exception, we do not give the specific values of the empirical constants C, $\bar{g}_{\text{Na}}$, $\bar{g}_{\text{K}}$, $\bar{g}_{\text{L}}$, V_{Na}, V_{K}, and V_{L}. The one exception concerns $\bar{g}_{\text{Na}}$; for future reference we mention now that

$$\frac{\bar{g}_{\text{Na}}}{C} = 1.2 \times 10^5 \text{ sec}^{-1}. \tag{C2.9}$$

(Dividing by C in (C2.9) makes the units more convenient.)

We also remark that the voltage levels V_{Na}, V_{K}, and V_{L} are determined by the ion concentrations inside and outside the cell. In analyzing equations (C2.7), one usually takes the values for these voltage levels that occur *in vivo*. However an experimenter can readily vary the ion concentrations outside the cell; thus all values of the parameters V_{Na}, V_{K}, and V_{L} have physical validity. (Indeed, Hodgkin and Huxley separated the different contributions to I_{ion} in (C2.3) precisely by varying the voltage levels in this manner.)

(b) Hopf Bifurcations in the Clamped Hodgkin–Huxley Equations

In the clamped Hodgkin–Huxley equations (C2.7), the only variables that are directly observable in the laboratory are the potential difference V and the current I_{ext}. Experimentally it is possible to fix either V or I_{ext} to a prescribed level and to measure the other. In the original experiments from which (C2.7) was derived, Hodgkin and Huxley fixed V and measured I_{ext}. Below we shall consider fixing I_{ext} and measuring V. In other words, we shall regard (C2.7) as defining a bifurcation problem with I_{ext} as the bifurcations parameter. Observe that this bifurcation problem depends on several auxiliary parameters; most notable among these is the temperature T, since for the time being we assume that the other parameters take the values which occur *in vivo*.

In this Case Study, our interest in the clamped Hodgkin–Huxley equations derives from the fact that for a large range of values of I_{ext}, these equations possess stable periodic solutions (provided the temperature is not too high). These solutions were first discovered numerically; more recently bifurcation theory was used to establish their existence analytically. (See Rinzel [1978] for references to earlier papers and Hassard [1978], Rinzel and Miller [1980], and Labouriau [1983] for more recent work.) This mathematical fact, the existence of periodic solutions, has an approximate experimental counterpart in that when I_{ext} is fixed, the axon emits several pulses in sequence before settling into a new equilibrium. However, in the laboratory only finitely many pulses occur, perhaps as few as four; by contrast (C2.7) predicts an infinite sequence of pulses. This discrepancy of course represents an inadequacy of (C2.7) as a model; specifically, (C2.7) ignores the fact that the axon saturates or "adapts" under continuous stimulation. Although there are more accurate models which incorporate adaption, we shall only work with (C2.7)—our goal is to illustrate the use of singularity theory in exploring the properties of a mathematical model, not to understand the physiology.

The periodic solutions of the clamped Hodgkin–Huxley equations bifurcate from a trivial solution as I_{ext} is varied; in other words, (C2.7) exhibits Hopf bifurcation. The salient points concerning this bifurcation are the following:

(i) For every value of I_{ext} there is a unique, steady-state solution of (C2.7). (Cf. Hassard [1978].)
(ii) Except for very high temperatures ($T > 28°C$), there are two distinct Hopf bifurcations as I_{ext} is varied, both subcritical. (Cf. Hassard [1978] and Labourian [1983].)
(iii) The solutions emerging from the two bifurcations are connected to one another globally as sketched in Figure C2.3. (Several bifurcation diagrams corresponding to different temperatures are shown in the figure. A dotted line indicates an unstable solution. For a given temperature, the trivial solution is unstable between the two bifurcation points and stable elsewhere.) (Cf. Rinzel and Miller [1980].)

Figure C2.3 suggests that the loops of periodic solutions of (C2.7) shrinks to zero as the temperature is increased, and this is in fact the case. In her thesis, Labouriau [1983] investigated this behavior quantitatively. As one might expect, the loop of periodic solutions disappears when the two bifurcation points meet one another and move off into the complex plane. However, it turns out that before this happens, the first bifurcation (i.e., the one on the left in Figure C2.3) changes from subcritical to supercritical. Specifically Labouriau's results are the following:

(i) For $T < T_2 = 28.859°C$ there are two Hopf bifurcations from the trivial solution and for $T > T_2$, none.
(ii) For $T < T_1 = 28.853°C$ both bifurcations are subcritical, while for $T_1 < T < T_2$ the first bifurcation is supercritical, the second, subcritical.

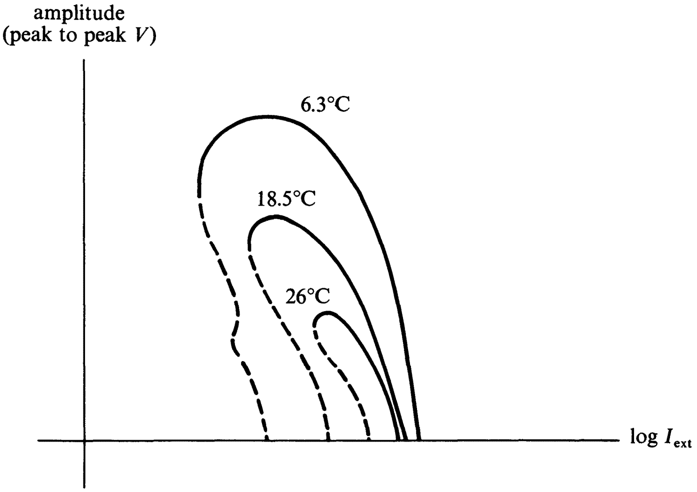


Figure C2.3. Periodic solutions of the clamped Hodgkin–Huxley equations. (Cf. Rinzel and Miller [1980].)

Bifurcation diagrams for the three cases $T < T_1$, $T_1 < T < T_2$, and $T > T_2$ are shown in Figure C2.4.

Remark C2.1. Rinzel and Miller [1980] found that there is an extra "knee" in the bifurcation diagram of Figure C2.3 when $T = 6.3$°C. In other words, there is a small range of volume of I_{ext} for which the clamped Hodgkin–Huxley equations have *four* periodic solutions. However, only the periodic solution with the largest amplitude is asymptotically stable.

(c) Discussion of the High-Temperature Behavior Using Singularity Theory

Our purpose in this Case Study is to illustrate how singularity theory can help in synthesizing a coherent understanding of the behavior of a mathematical model. To this end, in this subsection we discuss the disappearance of periodic solutions of the clamped Hodgkin–Huxley equations at high temperatures, taking the point of view of singularity theory. In subsection (b), the temperature T was the only auxiliary parameter. Thus from a generic point of view, one would expect only singularities of codimension one to be

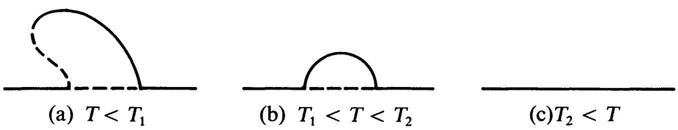


Figure C2.4. Bifurcations at high temperature.

involved in this disappearance of periodic solutions. In this subsection we discuss what insight may be gained by looking at singularities of codimension one. (In subsection (d) we consider varying other parameters besides T and look at singularities of higher codimension which might arise.)

As we stated above, Labouriau [1983] found that when the temperature is increased, one of the two Hopf bifurcations of (C2.7) changes from subcritical to supercritical before the two bifurcation points meet one another and disappear. (Cf. Figure C2.4.) Let us now argue that from the singularity theory point of view this behavior was inevitable. To this end, we recall the identification established in Chapter VIII between periodic solutions of a system of ODE near a bifurcation point and solutions of a single scalar equation that has $\mathbf{Z}_2$-symmetry. Let $g(x, \lambda) = 0$ be the reduced bifurcation problem obtained in this way from (C2.7) when $T = T_2$; i.e., at the temperature where the two bifurcation points meet one another. Generically we expect g to be a singularity of codimension one or less, since T is the only auxiliary parameter in the problem. Moreover g cannot be of codimension zero, since g is not persistent; specifically, an arbitrary small change in temperature splits the bifurcation point into two bifurcation points or eliminates all bifurcation points according as the change is negative or positive, respectively. Thus g must be of codimension one. We see from Table VI,5.1 that there are two classes of $\mathbf{Z}_2$-symmetric bifurcation problems of codimension one, namely,

$$\begin{aligned} &\text{(a)} \quad \varepsilon x^3 + \delta\lambda^2 x, \\ &\text{(b)} \quad \varepsilon x^5 + \delta\lambda x, \end{aligned} \tag{C2.10}$$

where ε and δ are ± 1. The singularity g must be equivalent to one of these normal forms. Indeed, varying the temperature in (C2.7) leads to a one-parameter unfolding $G(x, \lambda, T)$ of $g(x, \lambda)$, where $G(x, \lambda, T_2) = g(x, \lambda)$; by the universal unfolding theorem, the unfolding G can be factored through the universal unfolding of the relevant normal form in (C2.10). To summarize the argument so far, up to equivalence, the bifurcation diagrams of (C2.7) for T near T_2 must be one of those illustrated in Figures VI,3.1, 3.2, and 3.3, the figures in which universal unfoldings of (C2.10) are graphed.

The crucial point in our argument is the following: In Figures VI,3.1, 3.2, and 3.3, *whenever a bifurcation diagram contains two bifurcation points, one bifurcation is subcritical and the other is supercritical.* Therefore, for the two Hopf bifurcation points of (C2.7) to meet one another at a codimension-one singularity as the temperature is increased, necessarily one of the bifurcations must first become supercritical.

Only one of the normal forms in (C2.10) is consistent with the information given in subsection (b) above; viz., $x^3 + \lambda^2 x$. We see from Figures VI,3.1, 3.2, and 3.3 that (C2.10*a*) is the normal form associated with two bifurcation points meeting one another as a parameter is varied. If $\varepsilon\delta = +1$ in (C2.10), the two bifurcating solution branches connect to one another to form a loop (cf. Figure VI,3.3); if $\varepsilon\delta = -1$, the branches do not form such a loop (cf.

Figure VI, 3.2). It is clear from Figure C2.4(b) that $\varepsilon\delta = +1$ is the right choice for the problem at hand. Finally, we observe that since the trivial solution is stable below the bifurcation point, we must take $\delta = +1$. This completes the argument that $x^3 + \lambda^2 x$ is the appropriate normal form when the two bifurcation points meet one another.

Conversely, these ideas may be reversed to obtain an analytical proof that, at least for T near T_2, the solution branches emerging from the two bifurcation points join up to form a loop as in Figure C2.4(b). The argument would run as follows. As above, let $g(x, \lambda)$ be the reduced bifurcation problem obtained from (C2.7) via Liapunov–Schmidt reduction when $T = T_2$, and let $G(x, \lambda, T)$ be the one-parameter unfolding that results from varying the temperature. Suppose we show that at the bifurcation point

$$g_x = 0, \qquad g_{\lambda x} = 0, \qquad g_{xxx} > 0, \quad \text{and} \quad g_{\lambda\lambda x} > 0;$$

then it would follow from Theorem VI,5.1 that g is equivalent to $x^3 + \lambda^2 x$. (Cf. also Table VI,5.3 for the solution of the recognition problem.) Moreover G may be factored through the universal unfolding of $x^3 + \lambda^2 x$; indeed if $G_{xT} \neq 0$, then G provides a universal unfolding of g, and we may deduce that for T near T_2 the bifurcation diagrams of G have a loop connecting the two bifurcation points, simply from an inspection of Figure VI,3.3. (*Remark*: Labouriau [1983] performed calculations of this type for a related, more complicated problem; see subsection (d) below.)

The above argument is rigorously applicable only for T near T_2. However, as is so often the case, the structure of the bifurcation diagram which is established near the degenerate bifurcation point persists far away from this point; to rephrase this in the present context, the loop of periodic solutions connecting the two bifurcation points continues to exist at moderate and low temperatures. (Its structure does change slightly at low temperatures—cf. Remark C2.1.)

Incidentally, (C2.10b) is the normal form of the reduced equations for (C2.7) when $T = T_1$; i.e., at the temperature where the subcritical bifurcation changes to supercritical. By an argument similar to the above, it can be shown that $+x^5 - \lambda x$ is the only choice of signs consistent with the data of subsection (b).

(d) Singularities of Higher Codimension

Above we saw that singularities of codimension one occur in the clamped Hodgkin–Huxley equations at the temperatures $T_1 = 28.853°\text{C}$ and $T_2 = 28.859°\text{C}$. These two temperatures are exceedingly close to one another—the relative separation is only 0.02%. The fact that they are so close suggests that it may be possible to make them merge and form a more degenerate singularity by a small variation in an additional parameter. This degenerate singularity would provide an organizing center for the problem. By examining the universal unfolding of this degenerate singularity, one could deduce that at slightly modified values of the parameters the clamped Hodgkin–Huxley equations exhibit more complicated behavior than what has been seen up to now.

Labouriau [1983] carried out this program for analyzing the Hodgkin–Huxley equations, including the necessary numerical calculations. We summarize her work in part (ii) of this subsection, after presenting some conjectures of our own in part (i). In part (ii) we also discuss certain unresolved differences in our points of view which suggest further research. Regarding these differences, let us state clearly that Labouriau's work came first. This means not only that it was her idea first to apply singularity theory to this problem, but also that we had the benefit of her work in forming our conjectures.

(i) *Conjectures on the Application of Singularity Theory*

When $T = T_1$ the clamped Hodgkin–Huxley equations exhibit a singularity equivalent to the normal form (C2.10b); when $T = T_2$, (C2.10a). The codimension-one singularities (C2.10) arise from the failure of one of the nondegeneracy conditions in the Hopf theorem. Specifically, the eigenvalue crossing condition fails in (C2.10a) and the coefficient of the cubic terms vanishes in (C2.10b). The simplest singularity in which both these conditions fail simultaneously is

$$\varepsilon x^5 + 2m\lambda x^3 + \delta\lambda^2 x, \tag{C2.11}$$

where ε and δ equal ± 1 and m, a modal parameter, can assume any real value. Apart from the exceptional values $m = 0, \pm\sqrt{\varepsilon\delta}, \pm\infty$, this singularity has ($C^\infty$) codimension three and topological codimension two.

It is natural to conjecture that if the codimension-one singularities of the clamped Hodgkin–Huxley equations can be made to merge through varying a parameter, then (C2.11), for some choice of ε, δ, and m, is the normal form which describes the resulting degenerate singularity. Let us argue that in (C2.11) only $\varepsilon = \delta = +1$ is consistent with the information in subsection (b). For the trivial solution to be stable below the bifurcation point we must have $\delta = +1$. If (C2.11) is perturbed by restoring all physical parameters to their values *in vivo*, we want the bifurcating solutions to form a closed loop, as in Figure C2.3; on inspecting Figures VI,7.3 and 7.4, we see that such a loop is formed only if $\varepsilon = +1$.

By contrast, the available data do not select an exact value for the modal parameter m. Let us discuss the possibilities. We anticipate that m should be positive since there are small perturbations of (C2.11) with two subcritical bifurcations only if $m > 0$. (Cf. Figure VI,7.4.) However, even assuming $m > 0$, the available data do not choose decisively between the two ranges of topological triviality, $0 < m < 1$ and $m > 1$. There is a significant difference in the behavior associated with the two ranges which at first seem to offer a basis for choosing between them. Let us elaborate. Consider the (presumably small) perturbation of (C2.11) associated to restoring the parameters in the clamped Hodgkin–Huxley equations to their original values. This perturbation should lead to diagrams as shown in Figure C2.3. Refering to Figure VI,7.4, we see that Case 5 for $m > 1$ and Case 1′ for $0 < m < 1$ both have this behavior, with the difference that in Case 5 for $m > 1$ there is a branch of solutions completely disconnected from the trivial solution or branches which bifurcate from it. The fact that no such solutions have been reported in the literature would seem to argue for the other case. However, when the clamped Hodgkin–Huxley equations have been solved numerically, there does not seem to have been a serious effort to search for isolated branches of solutions. Thus in our opinion this difference does not offer a reliable guide for which range the parameter m in (C2.11) should lie in.

In the absence of a clear-cut choice between $0 < m < 1$ and $m > 1$, we present a plausibility argument that $0 < m < 1$ is the more likely occurrence. The starting point for this argument is the extreme proximity of T_1 and T_2. It seems that the clamped Hodgkin–Huxley equations (with parameters set at their values *in vivo*) just miss having an accidental degeneracy; i.e., a singularity of codimension higher than the number of parameters being varied. Our argument will be that the degree of accidental degeneracy is less if $0 < m < 1$ than if $m > 1$.

First we establish some notation. We have supposed that one of the parameters in the clamped Hodgkin–Huxley equations has been varied from its value *in vivo* until the two codimension-one singularities merge. Let us call this parameter a. We write a_* for its value *in vivo*, a_0 for the value at which the degenerate singularity occurs. (Similarly let T_0 be the temperature at which the degenerate singularity occurs.) We shall assume throughout this argument that $a_0 - a_*$ is small. This is simply a reformulation of our observation that the separation between the two singularities of codimension one is small, so that a small change in an auxiliary parameter should suffice to cause them to merge.

For T near T_0 and a near a_0, let $G(x, \lambda, T, a)$ be the $\mathbf{Z}_2$-symmetric bifurcation problem obtain by the Liapunov–Schmidt reduction of the clamped Hodgkin–Huxley equations, the parameter a having been changed from its value *in vivo* as indicated. Then $G(x, \lambda, T_0, a_0)$ is equivalent to (C2.11), and $G(x, \lambda, T, a)$ is a two-parameter unfolding of $G(x, \lambda, T_0, a_0)$. According to the universal unfolding theorem, $G(x, \lambda, T, a)$ may be factored through the universal unfolding of (C2.11); viz;

$$x^5 + 2m\lambda x^3 + \lambda^2 x + \alpha x + \beta x^3. \tag{C2.12}$$

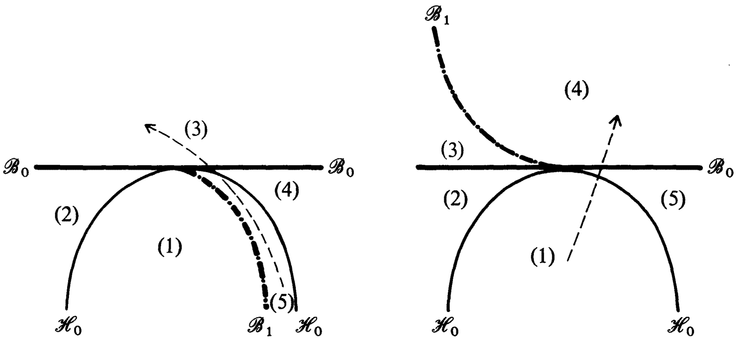
(We have set $\varepsilon = \delta = +1$.) The theorem guarantees only that this factorization is possible in some neighborhood of T_0, a_0; however, in keeping with the idea that $a_0 - a_*$ is small, *we assume* that the domain of validity of this representation is large enough to allow a to be restored to its value *in vivo*. It follows from this assumption that periodic solutions of the clamped Hodgkin–Huxley equations with *in vivo* parameters may be associated with solutions of the equation $G(x, \lambda, T, a_*) = 0$. Moreover for each T, $G(x, \lambda, T, a_*)$ is equivalent to (C2.12) for some choice of α, β, and m; specifically let us define $\alpha(T)$, $\beta(T)$, $m(T)$ so that

$$G(x, \lambda, T, a_*) \sim x^5 + 2m(T)\lambda x^3 + \lambda^2 x + \alpha(T)x + \beta(T)x^3. \quad \text{(C2.13)}$$

Since the unfolding (C2.12) is topologically trivial in the ranges $0 < \mathrm{m} < 1$ and $m > 1$, we shall ignore the coefficient $m(T)$ in our discussion below.

The functions $\alpha(T)$, $\beta(T)$ define a curve in the α,β-plane. Let us see how this curve is placed relative to the transition variety of the universal unfolding (C2.12). In Figure C2.5 we have reproduced slices at constant m of the transition varieties in case $0 < m < 1$ and $m > 1$; these come from Figure VI,7.4. Of course, the curve $\alpha(T)$, $\beta(T)$ is subject to the restriction that as T varies it must produce the sequence of bifurcation diagrams of the original equations; i.e., those shown in Figure C2.4. In both Figures C2.5(a) and (b) we have drawn a curve which has this property; in fact, only curves crossing the indicated regions in the same sequence will have this property.

To conclude our argument, we claim that if $m > 1$ the curve $\alpha(T)$, $\beta(T)$ must be positioned very carefully in order to have the properties required of it, while if $0 < m < 1$ there is much more freedom in the placement of this curve. Let us elaborate. Since we have assumed that $a_0 - a_*$ is small, we expect the curves in Figure C2.5 to lie rather close to the origin. Consider the possibility $m > 1$. Since T_1 and T_2 are so close to one another, the curve in Figure C2.5(a) must cross region 4 very quickly, although it dwells in region 3 and 5 rather longer. It is clear from the figure that the position, slope, and curvature of the curve must all lie within narrow limits to achieve this. By contrast, if $0 < m < 1$ it is a very minimal restriction on a curve to quickly cross region 5 of Figure C2.5(b) and to dwell in regions 1 and 4. In this sense, the degree of coincidence required by consistency with the data is lower if $0 < m < 1$. For this reason we consider $0 < m < 1$ the more likely possibility.



(a) $m > 1$ (b) $0 < m < 1$

Figure C2.5. Transition varieties for (C2.12).

(ii) *The Work of Labouriau*

Labouriau [1983] chose $\bar{g}_{Na}$, the sodium conductivity, as the parameter to vary in the clamped Hodgkin–Huxley equations. Her calculations show that a 9% decrease of this parameter leads to a singularity equivalent to (C2.11) with $\varepsilon = \delta = +1$ and $m \approx 12.69$. We shall first summarize her work and then discuss the fact that her results are at variance with our conjecture above.

To begin, let us recall from Table VI,5.3 the solution to the recognition problem for the normal form (C2.11). For a $\mathbf{Z}_2$-symmetric germ $g(x, \lambda)$ to be equivalent to (C2.11), g must satisfy defining conditions

$$g_x = g_{\lambda x} = g_{xxx} = 0 \tag{C2.14}$$

and certain nondegeneracy conditions. In stating the latter, we use Corollary VI,2.2 to express g in the form $g(x, \lambda) = r(x^2, \lambda)x$ for some germ $r(u, \lambda)$; as in Chapter VI we write $u = x^2$. The nondegeneracy conditions for equivalence with (C2.11) are

$$r_{uu} \neq 0, \qquad r_{\lambda\lambda} \neq 0, \qquad r_{\lambda u}^2 - r_{uu}r_{\lambda\lambda} \neq 0. \tag{C2.15}$$

Assuming (C2.15), the parameters in (C2.11) are given by

$$\varepsilon = \operatorname{sgn} r_{uu}, \qquad \delta = \operatorname{sgn} r_{\lambda\lambda}, \qquad m = \frac{r_{\lambda u}}{\sqrt{|r_{uu}r_{\lambda\lambda}|}}. \tag{C2.16}$$

Equations (C2.14) determine the values of the parameters in the clamped Hodgkin–Huxley equations for which the degenerate singularity occurs. Labouriau solves equations (C2.14) as follows. Let $G(x, I_{ext}, T, a)$ be the reduced function obtained from the Liapunov–Schmidt reduction of the clamped Hodgkin–Huxley equations when $\bar{g}_{Na}$ is replaced by a free parameter a. First, for each value of a, Labouriau solves (numerically) the two equations

$$G_x = \frac{\partial}{\partial I_{ext}} G_x = 0 \tag{C2.17}$$

for I_{ext} and T. (*Warning*: Equations (C2.17) must be solved with $x = 0$; x is *not* to be determined from the equations. To see this, recall from the Liapunov–Schmidt reduction of Chapter VIII, §2 that x parametrizes the amplitude of a possible periodic solution. We are studying bifurcation from a steady-state solution, which means that $x = 0$.) Then for the values of I_{ext} and T determined from (C2.17), Labouriau evaluates G_{xxx}, the third function which must vanish in (C2.14). The result is plotted in Figure C2.6. Note that G_{xxx} vanishes when $a/C \approx 1.1 \times 10^5$ sec^{-1}. (Regarding the comparison with *in vivo* parameters, we see from (C2.9) that $a_*/C \approx 1.2 \times 10^5$ sec^{-1}.) Finally, for the distinguished values of I_{ext}, T, and a selected by this procedure, Labouriau computes the higher-order derivatives of G, thereby verifying the nondegeneracy conditions (C2.15). In this way she proves that the clamped Hodgkin–Huxley equations exhibit a singularity equivalent to (C2.11) with $\varepsilon = \delta = +1$ and $m \approx 12.69$ when $\bar{g}_{Na}$ is reduced by approximately 9%.

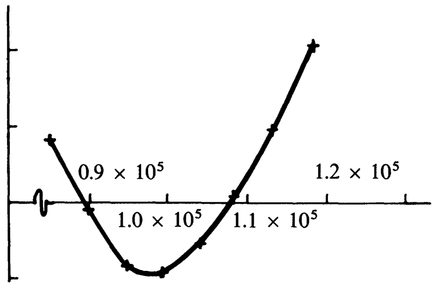


Figure C2.6. G_{xxx} as a function of a. (Cf. Labouriau [1983], p. 90.)

No doubt the reader has noticed that Labouriau's results differ from what we conjectured above. Specifically, she found a modal parameter in the range $m > 1$, not $0 < m < 1$ as we conjectured. Her results are based on actual computations while our conjectures are based on a loose plausibility argument. The reader may well wonder why we chose to present unsubstantiated conjectures. The reason is twofold: Partially we suspect that the story on this application of singularity theory is not yet complete, and partially the unresolved issues point to questions for further research.

The fact that Labouriau found a degenerate singularity where $m > 1$ with a parameter change of 9% does not rule out the possibility that there is another degenerate singularity, much closer to the physical values that could only be found by varying a different parameter. Indeed, consider varying all three conductivities $\bar{g}_{\mathrm{Na}}$, $\bar{g}_{\mathrm{K}}$, and $\bar{g}_{\mathrm{L}}$ and all three voltage levels V_{Na}, V_{K}, and V_{L} in (C.2) and searching for degenerate singularities in this six-dimensional parameter space. (For simplicity let us not consider changes in $\Phi(T)$, $\tau_M(V)$, $M_\infty(V)$, etc.) Since (C2.11) has topological codimension two, generically we would expect there to be a four-dimensional subvariety in $\mathbb{R}^6$ of parameter values for which the equations exhibit a singularity equivalent to (C2.11). How can one find the singularity which has the greatest influence on the physical problem with *in vivo* parameters? Might there be even more degenerate singularities which play a significant role? We hope that the general theory will be able to provide insight concerning these questions, but more work is needed.

Incidentally, we note from Figure C2.6 that Labouriau actually found *two* choices of parameter values which yield a singularity equivalent to (C2.11)—the second such singularity occurs when $a/C \approx 0.9 \times 10^5\ \mathrm{sec}^{-1}$, a change of 25% from the *in vivo* parameter values. At this singularity $\varepsilon = \delta = +1$ but $m \approx -6.93$. Presumably this singularity has little influence on the physical domain, since it is so far removed.

Let us discuss the primary degenerate singularity that Labouriau found, where $m = 12.69$. We recall from part (i) that when a degenerate singularity with $m > 1$ is perturbed, the resulting bifurcation diagrams contain an isolated branch of solutions. In other words, it follows from the work of

Labouriau that there are periodic solutions of the clamped Hodgkin–Huxley equations with $\bar{g}_{Na}$ reduced about 9% which are not connected to the trivial solution in any way. Further investigation is needed to ascertain whether this branch of solutions persists as $\bar{g}_{Na}$ is restored to its value *in vivo*.

In this connection it is unfortunate that Labouriau varied $\bar{g}_{Na}$ rather than V_{Na}. As we noted in subsection (a), V_{Na} may be adjusted experimentally by changing the concentration of sodium ions outside the axon; there is no such way to alter $\bar{g}_{Na}$ experimentally. Had she chosen V_{Na}, the behavior of the equations near the degenerate singularity would be of direct experimental significance. With $\bar{g}_{Na}$, the link with experiment is less direct.

CHAPTER IX

Two Degrees of Freedom Without Symmetry

§0. Introduction

The main focus of the present volume has been bifurcation problems in one state variable. In this chapter and the next we anticipate Volume II by discussing certain limited aspects of the singularity theory of bifurcation problems with several state variables. For the most part, such problems have rather high codimensions, at least in the absence of symmetry. For example, 8 is the lowest possible codimension for a bifurcation problem in three or more state variables. In two state variables, 3 is the minimum codimension. In this chapter we study bifurcation problems in two state variables with codimension 3.

The number 3 is significant, since this was the cutoff codimension in our classification theorem in Chapter IV. Thus the work of the present chapter is necessary to complete the classification theorem. Recall that in Chapter IV, §1 we motivated the classification theorem by identifying codimension as a rough measure of the likelihood of finding a singularity in applications. However, the number of state variables did not affect this likelihood. Thus it is important to list *all* bifurcation problems of codimension 3 or less, not just those with one state variable.

This chapter is divided into three sections. In §1 we give the basic definitions for bifurcation problems in n state variables and estimate their codimensions. In §§2 and 3 we study bifurcation problems in two state variables of codimension three. (Following Thompson and Hunt [1979], we call these *hilltop bifurcation*.) In §2 we solve the recognition problem, and in §3 we enumerate perturbed bifurcation diagrams.

As we stated above, in the absence of symmetry, bifurcation problems with many state variables have high codimension. This means that such problems do not occur in applications very often. However, codimensions are changed drastically by symmetry; in particular, bifurcation problems with many state variables occur frequently in applications when there is a symmetry group acting. We shall explore one instance of this in the next chapter. In Volume II, the consequences of symmetry will be a major focus.

§1. Bifurcation with *n* State Variables

In this section, first we define the notion of bifurcation problems in n state variables, then we define equivalence for such problems, next we estimate the codimension of such problems, and finally we discuss the relation of equivalence and linearized stability in n dimensions.

Definition 1.1. Let $g\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ be the germ of a smooth map. Then we call $g(z, \lambda)$ a *bifurcation problem in n state variables* if

$$g(0, 0) = 0, \qquad (dg)_{(0,0)} = 0, \tag{1.1}$$

where dg is the $n \times n$ Jacobian matrix obtained by differentiation of g in the z directions.

Remarks. (i) We require that $(dg)_{0,0} = 0$ in (1.1). Should the rank $(dg)_{0,0}$ equal $k > 0$, then we could apply the Liapunov–Schmidt reduction to obtain a reduced bifurcation problem in $n - k$ state variables.

(ii) Usually bifurcation problems in n state variables result from a Liapunov–Schmidt reduction of some larger problem where the kernel of the linearized operator has dimension n.

(iii) Definition 1.1 implies that a bifurcation problem g in one state variable satisfies $g = g_x = 0$; that is, g is required to have a singularity at the origin.

Definition 1.2. Let $g, h\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ be bifurcation problems with n state variables with $n > 1$. Then g and h are *equivalent* if

$$g(z, \lambda) = S(z, \lambda)h(Z(z, \lambda), \Lambda(\lambda)), \tag{1.2}$$

where S is a $n \times n$ invertible matrix depending smoothly on z and λ and the mapping $\Phi(z, \lambda) = (Z(z, \lambda), \Lambda(\lambda))$ is an invertible change of coordinates preserving the orientation in λ. In particular, $Z(0, 0) = 0$, $\Lambda(0) = 0$, $\det(dZ)_{(0,0)} \neq 0$, and $\Lambda'(0) > 0$.

Remarks. (i) Since S in (1.2) is invertible we see that

$$\Phi(\{(z, \lambda)\colon g(z, \lambda) = 0\}) = \{(z, \lambda)\colon h(z, \lambda) = 0\}.$$

Thus equivalences preserve bifurcation diagrams. It also preserves the orientation of the parameter λ.

(ii) Note that in Definition 1.2 we do not assume that Z or S is orientation preserving. As a result, this definition of equivalence for one state variable bifurcation problems does not agree with the notion of equivalence given earlier. The reason for this change involves linearized stability, which we will discuss below.

Let us now discuss codimension for bifurcation problems in n state variables. The following proposition shows that the codimension of such problems is very high indeed if n is at all large. (*Remark:* The proposition assumes that no symmetry is present.)

Proposition 1.3. *If g is a bifurcation problem with n state variables, then*

$$\operatorname{codim} g \geq n^2 - 1. \tag{1.3}$$

We shall not set up the machinery needed to give a formal definition of codimension. (We will do this in Volume II.) Hence we cannot give a precise proof of the proposition. Rather, we will give an argument for (1.3) based on analogy with Corollary III,2.6. In that corollary we showed that in one state variable the codimension of g is equal to the number of defining conditions for g less two. The correction term two came from the fact that g was defined on two variables x and λ. A corresponding result is true for bifurcation problems with n state variables; viz., the codimension of g equals the number of defining conditions for g less $n + 1$. Here the correction term $n + 1$ comes from the fact that g depends on $n + 1$ variables, $z_1, \ldots, z_n$, and λ. This is the basis for our heuristic proof of Proposition 1.3.

A bifurcation problem g in n state variables must satisfy the $n + n^2$ conditions given in (1.1); i.e., there are $n^2 + n$ defining conditions. Subtracting the correction term $n + 1$ we arrive at our lower bound of $n^2 - 1$ for the codimension of g.

It follows from (1.3) that bifurcation problems in n state variables where $n \geq 3$ have codimension at least eight. According to the thesis of Chapter IV, §1, we should expect such singularities only in a bifurcation problem with at least eight auxiliary parameters (after nondimensionalization). In this sense bifurcation problems with many state variables are unlikely in applications. (Furthermore, these problems are so complex as to defeat the purpose of singularity theory methods; i.e., to provide a mathematical structure which facilitates the understanding of bifurcation phenomena.)

However, in two state variables there are bifurcation problems with codimension three. Moreover, three is the number we (somewhat arbitrarily) chose as our limit for the classification theorem of Chapter IV. In the remaining two sections of this chapter, we study those bifurcation problems in two state variables which have codimension three.

We end this section with a brief discussion of linearized stability, especially the incompatibility of equivalence with linearized stability.

Let h be a bifurcation problem with n state variables, and suppose $h(z_0, \lambda_0) = 0$. We say that (z_0, λ_0) is *linearly stable* if every eigenvalue of $(dh)_{z_0, \lambda_0}$ has a positive real part. (Here we are thinking of (z_0, λ_0) as a steady-state solution to the system of ODE's $\dot{z} + h(z, \lambda) = 0$.) We call this solution *linearly unstable* if some eigenvalue has a negative real part.

We ask the question: Are the signs of the real parts of the eigenvalues of dh invariants of equivalence? The answer is clearly no; let $S = -I$. When $n = 1$ we restricted the notion of equivalence by requiring $S(0, 0) > 0$ and $X_x(0, 0) > 0$. Consequently, in this case stability or instability *is* an invariant of equivalence. (Cf. Theorem I,4.1.) In this regard, one might be tempted to require in Definition 1.1 that $\det S$ and $\det dZ$ be positive. However, this restriction is not sufficient to obtain the invariance of stability assignments under equivalence. For example, if n is even, then the determinant of $-I$ is $+1$; thus, as in the above example, we may change all the signs of the eigenvalues of dh. For this reason when $n \geq 2$, we have not imposed any positivity conditions on S or X in Definition 1.2.

We end our discussion of stability assignments with an observation: Although the question of invariance of stability assignments under (an appropriate form of) equivalence is a complicated issue, symmetry often makes the problem easier. We shall explore one instance of this in the next Chapter X, §3 and many more in Volume II.

§2. Hilltop Bifurcation

We now complete the classification theorem for bifurcation problems in codimension three or less. (Cf. Theorem IV, 2.1.)

Theorem 2.1. *Let* $g: \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ *be a bifurcation problem with* codim $g \leq 3$. (*As in Definition* 1.2, *we assume* $g(0, 0)$ *and* $(dg)_{0,0}$ *vanish.*) *Then either* $n = 1$ *or* $n = 2$. *If* $n = 1$, *then* g *is equivalent to one of the eleven normal forms listed in Table* IV,2.1. *If* $n = 2$ *then* g *is equivalent to one of the normal forms*

$$(12) \quad (x^2 - y^2 + \lambda, 2xy), \quad \text{or}$$

$$(13) \quad (x^2 + \varepsilon\lambda, y^2 + \delta\lambda),$$

where $\varepsilon, \delta = \pm 1$.

Remarks. (i) The numbers (12) and (13) are intended to continue the list of Table IV,2.1.

(ii) The two cases in normal form (13) in which $\varepsilon\delta = -1$ are equivalent. Otherwise, all normal forms in (12) and (13) are inequivalent.

Although we have many of the pieces at our disposal, we shall not give a complete proof of Theorem 2.1. Specifically we postpone for Volume II:

(i) The proof of Proposition 1.3 which shows that $n \leq 2$;
(ii) The characterization of higher-order terms for the normal forms (12) and (13).

Let us discuss the second point. Let $\mathcal{M}$ be the maximal ideal in $\mathscr{E}_{x,y,\lambda}$; i.e., $\mathcal{M}$ is the set of germs of real valued functions on $\mathbb{R}^2 \times \mathbb{R}$ which vanish at the origin. In Volume II, we will show that all terms which lie in

$$\mathcal{M}^3 + \mathcal{M}\langle\lambda\rangle \tag{2.1}$$

are higher-order terms for the normal forms (12) and (13). This means that the normal form plus a perturbation lying in (2.1) is equivalent to the original normal form. (*Remark*: Note that the terms in (12) and (13) belong to $\mathcal{M}^2 + \langle\lambda\rangle$.)

In this section we concentrate on one important aspect of the proof of Theorem 2.1; viz., the analysis of intermediate-order terms. Specifically, we shall introduce a notion of nondegenerate bifurcation problems in two state variables. A bifurcation problem is nondegenerate provided it satisfies several inequalities as in (H1) and (H2) below. (*Remark:* These inequalities are invariants of equivalence.) Subject to the limitations of the preceding paragraph, we will prove that any nondegenerate bifurcation problem in two state variables is equivalent to (12) or (13). (Cf. Theorems 2.2 and 2.3.) This fact relates to Theorem 2.1 as follows. If g is a *degenerate* bifurcation problem, then one of the inequalities must fail; in other words, an inequality becomes an equality. This equality represents an extra defining condition on g, in addition to (1.1). Thus codim $g \geq 4$. Alternatively put, nondegenerate bifurcation problems are precisely those bifurcation problems associated with the defining conditions (1.1); such problems have codimension three.

We now introduce notational conventions that we will use for the remainder of this section. Let g be a bifurcation problem with two state variables. It follows from (1.1) that $g(0, 0) = 0$ and $(dg)_{0,0} = 0$. Thus we may write

$$g(x, y, \lambda) = f(x, y) + \lambda r + \cdots,$$

where $\cdots$ indicates terms lying in (2.1) and

$$\begin{aligned}
\text{(a)}\quad f(x, y) &= (p(x, y), q(x, y)) \\
&= (p_1x^2 + p_2xy + p_3y^2, q_1x^2 + q_2xy + q_3y^2),
\end{aligned} \tag{2.2}$$

$$\text{(b)}\quad r = (r_1, r_2) = g_\lambda(0, 0).$$

Often we shall abbreviate (x, y) to z and write

$$h(z, \lambda) = f(z) + \lambda r \tag{2.3}$$

for the intermediate-order terms in g.

To begin our discussion, we show that if two germs g and $\tilde{g}$ are equivalent, then the corresponding intermediate-order terms h and $\tilde{h}$ are equivalent via a *linear* equivalence. Suppose that

$$\tilde{g}(z, \lambda) = S(z, \lambda)g(Z(z, \lambda), \Lambda(\lambda)). \tag{2.4}$$

Matching terms in (2.4) modulo the ideal (2.1), we see that

$$\tilde{h}(z, \lambda) = Bh(Az, \sigma\lambda), \tag{2.5}$$

where $B = S(0, 0)$, $A = (dZ)_{(0,0)}$, and $\sigma = \Lambda'(0)$.

Our first nondegeneracy condition is

(H1) $$g_\lambda(0, 0) \neq 0.$$

Note that (H1) is an invariant of equivalence.

Our second nondegeneracy condition is less obvious. Let

$$Q(z) = \det(df)_z.$$

Observe that $Q(z)$ is a homogeneous, quadratic polynomial in x and y, say

$$Q(z) = ax^2 + bxy + cy^2.$$

Thus the zero set of $\{z: Q(z) = 0\}$ consists of either two crossed lines, one line with multiplicity two, or just the origin. More briefly, in these cases we shall say that Q has two distinct real roots, two equal roots, or complex roots, respectively. Now the structure of the zero set of $Q(z)$ is determined by the sign of the discriminant

$$D = b^2 - 4ac.$$

Specifically,

$$\begin{aligned} &\text{(a)} \quad D > 0 \Leftrightarrow Q \quad \text{has distinct real roots,} \\ &\text{(b)} \quad D = 0 \Leftrightarrow Q \quad \text{has two equal roots,} \\ &\text{(c)} \quad D < 0 \Leftrightarrow Q \quad \text{has complex roots.} \end{aligned} \tag{2.6}$$

Let us compute D explicitly for the case at hand. We find

$$Q(z) = 2(p_1q_2 - p_2q_1)x^2 + 4(p_1q_3 - p_3q_1)xy + 2(p_2q_3 - p_3q_2)y^2, \tag{2.7}$$

so that

$$D = 16\{(p_1q_3 - p_3q_1)^2 - (p_1q_2 - p_2q_1)(p_2q_3 - p_3q_2)\}. \tag{2.8}$$

Our second nondegeneracy condition is

(H2) $$D \neq 0.$$

Let us show that (H2) is an invariant of equivalence. Indeed, the sign of D is such an invariant. We give a geometric argument. If g and $\tilde{g}$ are equivalent, then we apply the chain rule to (2.5) to deduce that

$$\tilde{Q}(z) = cQ(Az), \tag{2.9}$$

where $c = \det S(0, 0) \det A$. In words, (2.9) states that $\tilde{Q}$ is a nonzero multiple of Q evaluated in the new coordinates Az. Then the zero sets of Q and $\tilde{Q}$ have the same structure. By (2.6), D and $\tilde{D}$ have the same sign.

Condition (H2) divides nondegenerate bifurcation problems into two cases, D positive and D negative. We now state the appropriate normal form if D is negative. (We defer the proof until the end of the section.)

Theorem 2.2. *Let g be a bifurcation problem in two state variables satisfying*

(H1) $$g_\lambda(0, 0) \neq 0$$

and

(H2)$^-$ $$D < 0.$$

Then g is equivalent to

$$(x^2 - y^2 + \lambda, 2xy). \tag{2.10}$$

If $D > 0$, we need an additional nondegeneracy condition in order to derive normal forms. Since $D > 0$, there are precisely two lines along which $Q(z) = \det(df)_z$ vanishes. Choose nonzero vectors z_1 and z_2, one on each line, and let $w_i = f(z_i)$, $i = 1, 2$. Our last nondegeneracy condition is that $g_\lambda(0, 0)$ not be parallel to either w_1 or w_2. In terms of the cross product in two dimensions we may reformulate this condition as

(H3) $$g_\lambda(0, 0) \times w_i \neq 0, \qquad i = 1, 2.$$

Let us interpret the vectors w_i in condition (H3) geometrically. We claim that the image of the mapping $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ is the wedge spanned by the two vectors w_1 and w_2, as illustrated in Figure 2.1. The most elegant proof of

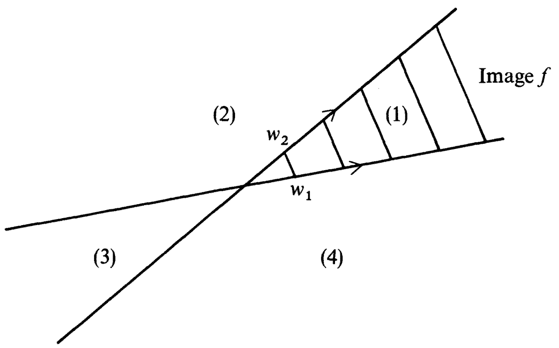


Figure 2.1. The image of f.

this claim is to change coordinates in the domain of f so that z_1 and z_2 become e_1 and e_2 (i.e., unit vectors in the coordinate directions), and to change coordinates in the range of f so that w_1 and w_2 also become e_1 and e_2. In these coordinates f takes the explicit form

$$f(z) = (x^2, y^2) \tag{2.11}$$

which occurs in (13). Indeed, this is precisely the construction we will use in obtaining the normal form (13). We refer to the proof of Theorem 2.3 for the details of this construction. For the mapping (2.11), the claim may be verified by a trivial, explicit calculation.

Next we argue that (H3) is an invariant of equivalence. It suffices to consider linear equivalences as in (2.5). Coordinate transformations on the domain do not change either w_i or $g_\lambda(0, 0)$—for g_λ, this is clear; for w_i, this follows from the above geometric interpretation. As regards coordinate transformations on the range, we observe that for two vectors $u, v \in \mathbb{R}^2$,

$$(Bu) \times (Bv) = (\det B)(u \times v). \tag{2.12}$$

Thus (H3) is not changed by coordinate transformations on the range.

We use condition (H3) to divide nondegenerate bifurcation problems with $D > 0$ into three cases. Referring to Figure 2.1, we identify four quadrants in the plane associated with the image of f. In words, $\operatorname{Im} f$ is the first quadrant, and the quadrants are numbered counterclockwise. The three cases are as follows:

$$\begin{aligned} &\text{(i)} \quad g_\lambda(0, 0) \in \text{quadrant } 1,\\ &\text{(ii)} \quad g_\lambda(0, 0) \in \text{quadrant } 3,\\ &\text{(iii)} \quad g_\lambda(0, 0) \in \text{quadrant } 2 \text{ or } 4. \end{aligned} \tag{2.13}$$

These three cases are invariant under equivalence. (*Remark*: It might seem more natural to divide Case (iii) into two subcases, $g_\lambda(0, 0)$ in quadrant 2 or 4. However, these two subcases are not invariants of equivalence. Specifically, multiplication of g by a matrix B with $\det B < 0$ interchanges them.)

We now state the normal form theorem for nondegenerate bifurcation problems with $D > 0$.

Theorem 2.3. *Let g be a bifurcation problem in two state variables satisfying*

(H2)$^+$ $\qquad D > 0,$ *and*

(H3) $\qquad g_\lambda(0, 0) \times w_i \neq 0, \qquad i = 1, 2.$

Then g is equivalent to

$$\begin{aligned} &\text{(i)} \quad (x^2 + \lambda, y^2 + \lambda),\\ &\text{(ii)} \quad (x^2 - \lambda, y^2 - \lambda), \quad \textit{or}\\ &\text{(iii)} \quad (x^2 + \lambda, y^2 - \lambda), \end{aligned} \tag{2.14}$$

according as Case (i), (ii), *or* (iii) *occurs in* (2.13).

(*Remark*: Hypothesis (H3) implies (H1).)

In proving Theorems 2.2 and 2.3, we shall assume that $g(z, \lambda)$ has the form

$$g(z, \lambda) = f(z) + \lambda r,$$

with f and r as in (2.2). This is based on our (unproved) assertion that terms in (2.1) are higher-order terms for normal forms (12) and (13) and may be transformed away by an appropriate choice of equivalence.

PROOF OF THEOREM 2.3. Choose z_1, z_2, w_1, w_2 as above. Define 2×2 matrices A and B such that

$$A\begin{pmatrix}1\\0\end{pmatrix} = z_1, \qquad A\begin{pmatrix}0\\1\end{pmatrix} = z_2,$$
$$Bw_1 = \begin{pmatrix}1\\0\end{pmatrix}, \qquad Bw_2 = \begin{pmatrix}0\\1\end{pmatrix}.$$

Then

$$\tilde{g}(z, \lambda) = Bg(Az, \lambda)$$

has the form

$$\tilde{g}(z, \lambda) = (x^2 + a\lambda, y^2 + b\lambda). \tag{2.15}$$

With this normalization, quadrants in Figure 2.1 coincide with the usual quadrants in the xy-plane. In (2.15), if (a, b) belongs to the first, third, or second quadrants, we perform the equivalence

$$\tilde{B}\tilde{g}(\tilde{A}, \lambda), \tag{2.16}$$

where

$$\tilde{A} = \begin{pmatrix}|a|^{1/2} & 0\\ 0 & |b|^{1/2}\end{pmatrix}, \tilde{B} = \begin{pmatrix}|a|^{-1} & 0\\ 0 & |b|^{-1}\end{pmatrix};$$

this reduces (2.15) to (2.14i), (2.14ii), or (2.14iii), respectively. If (a, b) belongs to the fourth quadrant, we perform a preliminary equivalence (2.16) with

$$\tilde{A} = \tilde{B} = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}.$$

This interchanges a and b in (2.15), so that (a, b) belongs to the second quadrant. We then proceed as before. □

PROOF OF THEOREM 2.2. To begin, we prove a preliminary result; viz., if $q(z)$ is a homogeneous quadratic polynomial in x and y which vanishes only at the origin, then there is a matrix A such that $q(Az) = \pm(x^2 + y^2)$. First recall that there is a rotation matrix

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

such that $q(Az) = ax^2 + by^2$. By rescaling x and y we may put q into the form $\pm x^2 \pm y^2$. Since q vanishes only at the origin, only the cases $\pm(x^2 + y^2)$ can occur.

Now form the mapping $f(z) = (p(z), q(z))$ from the quadratic terms of the mapping g in Theorem 2.2. We claim that q, the second coordinate of f, vanishes along two distinct lines in the plane. Suppose otherwise; then either

(i) q vanishes only at the origin, or
(ii) q vanishes (to order 2) along some line.

In the first case, we may assume that $q = x^2 + y^2$; this combines the above result with a possible multiplication of g by -1. Moreover, there is a rotation A such that $p(Az) = p_1x^2 + p_3y^2$. This rotation does not affect q, since q is rotationally invariant. We now compute from (2.8) that $D = 16(p_1 - p_3)^2 \geq 0$, which contradicts hypothesis $(\text{H2})^-$ of Theorem 2.2. This rules out possibility (i) above. Similarly, regarding possibility (ii), after a preliminary change of coordinates, we may assume that $q = x^2$. Substituting in (2.8) we find $D = 16p_3^2 \geq 0$, which is again a contradiction.

Therefore $\{q = 0\}$ consists of two distinct lines through the origin. Perform a linear transformation A so that $\{q = 0\}$ consists of the lines $\{x = 0\}$ and $\{y = 0\}$. Thus

$$p(x, y) = p_1x^2 + p_2xy + p_3y^2, \qquad q(x, y) = q_2xy.$$

We compute from (2.8) that $D = 16p_1p_3q_2^2$. Hence p_1 and p_3 have opposite signs. We rescale x and y and we multiply p and q by constants to obtain $p_1 = 1$, $p_3 = -1$, and $q_2 = 2$; this corresponds to an equivalence transformation $Bg(Az, \lambda)$ with A and B diagonal. Finally, we multiply g by an upper triangular matrix

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$$

to eliminate the cross term p_2xy. In this way we arrange that

$$f(z) = (x^2 - y^2, 2xy).$$

We normalize the term $\lambda r = \lambda g_\lambda(0, 0)$ as follows. Observe that for any rotation matrix A,

$$Af(A^{-1/2}z) = f(z).$$

Choose a rotation A such that $Ar = a(1, 0)$. We rescale λ to obtain the desired normal form (12). □

§3. Persistent Bifurcation Diagrams for Hilltop Bifurcation

In this section we study the perturbed bifurcation diagrams associated to the normal forms

$$\begin{aligned} &\text{(a)} \quad (x^2 - y^2 + \lambda, 2xy), \\ &\text{(b)} \quad (x^2 - \lambda, y^2 - \lambda), \\ &\text{(c)} \quad (x^2 + \lambda, y^2 - \lambda). \end{aligned} \tag{3.1}$$

Formula (3.1a) is the normal form (12) of Theorem 2.1. Formulas (3.1b, c) list two of the three inequivalent cases in normal form (13); pictures for the remaining case $(x^2 + \lambda, y^2 + \lambda)$ can be obtained from our pictures for (3.1b) below by mapping λ to $-\lambda$. For reference, the unperturbed bifurcation diagrams associated to all three normal forms (3.1) are shown in Figure 3.1. (*Remarks*: The bifurcation diagrams for (3.1) are curves in $\mathbb{R}^3$. In Figure 3.1 we have also drawn the x, λ and y, λ coordinate planes as an aid in visualizing the bifurcation diagrams. Portions of the bifurcation diagram are shown as dotted to indicate that they are "hidden" by the coordinate planes. In particular, *dotted curves do not indicate an unstable branch*; no stability assignments are indicated in Figure 3.1.)

We obtain perturbed bifurcation diagrams for (3.1) with the same methods that we developed in Chapter III for bifurcation problems in one state variable. Specifically,

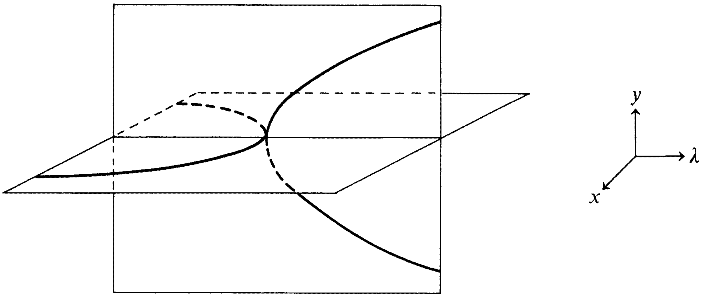
(i) we find a universal unfolding for the normal form;
(ii) we partition the parameter space ($\mathbb{R}^3$ in this case, since (3.1) has codimension three) into several regions with the transition variety Σ;
(iii) we obtain an enumeration of the perturbed bifurcation diagrams from the regions of $\mathbb{R}^3 \sim \Sigma$.

In Volume II, we will present the theory which justifies this method of analysis. Here we concentrate on the calculations for perturbations of (3.1).

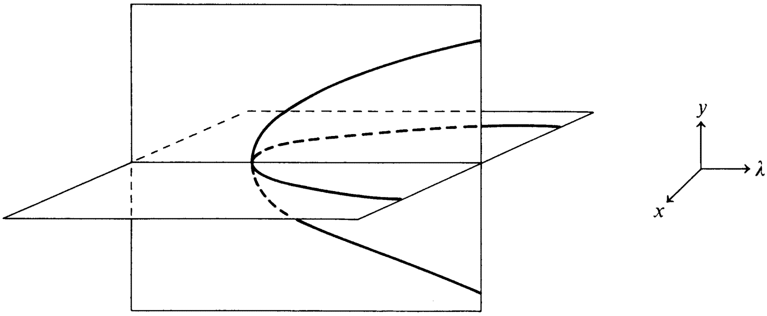
Although this section is fairly short, we have divided it into four subsections, as follows:

(a) Universal unfoldings.
(b) Theoretical characterization of the transition variety.
(c) Formulas and pictures for the transition variety.
(d) The perturbed bifurcation diagrams.

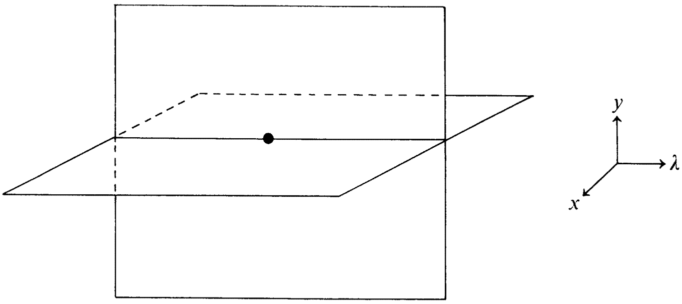
In subsections (a), (c), and (d) we carry out steps (i), (ii), and (iii) above. In subsection (b) we discuss how to obtain equations for Σ. Although with several state variables the same three phenomena (viz., bifurcation, hysteresis, and double limit points) contribute to Σ as with one state variable, the equations characterizing each variety appear slightly different. We hope that the explanation of subsection (b) will enable the reader to reproduce our subsequent calculations.



(i) $\lambda = y^2 - x^2$,
$xy = 0$



(ii) $\lambda = x^2$,
$x^2 = y^2$



(iii) $\lambda = y^2$,
$x^2 + y^2 = 0$

Figure 3.1. Unperturbed hilltop bifurcation, as in (3.1).

(a) Universal Unfoldings

The following are universal unfoldings of the normal form (3.1); note that they depend on three auxiliary parameters, α, β, and γ.

$$\begin{aligned}
&\text{(a)}\quad (x^2 - y^2 + \lambda + 2\alpha x - 2\beta y, 2xy + \gamma),\\
&\text{(b)}\quad (x^2 - \lambda + 2\alpha y - \gamma, y^2 - \lambda + 2\beta x + \gamma),\\
&\text{(c)}\quad (x^2 + \lambda + 2\alpha y + \gamma, y^2 - \lambda + 2\beta x + \gamma).
\end{aligned}\tag{3.2}$$

We do not prove this in the present volume. Indeed we have not even defined universal unfoldings for problems with n state variables. However, intuitively the situation is clear—the universal unfoldings (3.2) enumerate all small perturbations of (3.1), up to equivalence (as defined in Definition 1.2).

For reference, we present the following proposition which solves the recognition problem for universal unfoldings of hilltop bifurcation. We defer its proof for Volume II.

Theorem 3.1. *Let* $G(z, \lambda, \alpha, \beta, \gamma) = (P(z, \lambda, \alpha, \beta, \gamma), Q(z, \lambda, \alpha, \beta, \gamma))$ *be a three-parameter unfolding of a bifurcation problem* $g(z, \lambda)$ *in two state variables where* g *satisfies the nondegeneracy conditions* H1, H2, *and* H3 (*if* $D > 0$) *of* §2. *Then* G *is a universal unfolding of* g *if and only if*

$$\det\begin{pmatrix}
0 & P_{xx} & 2P_{xy} & 0 & Q_{xx} & 2Q_{xy}\\
0 & 2P_{xy} & P_{yy} & 0 & 2Q_{xy} & Q_{yy}\\
P_\lambda & P_{x\lambda} & P_{y\lambda} & Q_\lambda & Q_{x\lambda} & Q_{y\lambda}\\
P_\alpha & P_{\alpha x} & P_{\alpha y} & Q_\alpha & Q_{\alpha x} & Q_{\alpha y}\\
P_\beta & P_{\beta x} & P_{\beta y} & Q_\beta & Q_{\beta x} & Q_{\beta y}\\
P_\gamma & P_{\gamma x} & P_{\gamma y} & Q_\gamma & Q_{\gamma x} & Q_{\gamma y}
\end{pmatrix} \neq 0,$$

when evaluated at $(z, \lambda, \alpha, \beta, \gamma) = (0, 0, 0, 0, 0)$.

(b) Theoretical Characterization of the Transition Variety

Let $G: \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}^n$ be a k-parameter unfolding of a bifurcation problem g with n state variables. The main result concerning the transition variety $\Sigma \subset \mathbb{R}^k$ is the following: For any two choices of parameters, α_1 and α_2, in a given connected component of $\mathbb{R}^k \sim \Sigma$, $G(\cdot, \cdot, \alpha_1)$ and $G(\cdot, \cdot, \alpha_2)$ are equivalent. As in one state variable, Σ is a union of three subvarieties,

$$\Sigma = \mathscr{B} \cup \mathscr{H} \cup \mathscr{D},$$

where $\mathscr{B}$, $\mathscr{H}$, and $\mathscr{D}$ are associated with bifurcation, hysteresis, and double limit points, respectively. In this subsection we derive equations for $\mathscr{B}$, $\mathscr{H}$, and $\mathscr{D}$. For simplicity, we consider only the case of two state variables, although the differences with the general case are minimal.

Let $G(z, \lambda, \alpha)$ be a k-parameter unfolding of a bifurcation problem in two state variables; thus $z = (x, y) \in \mathbb{R}^2$, $\lambda \in \mathbb{R}$, and $\alpha \in \mathbb{R}^k$. We write G in components, $G = (P, Q)$. Let DG be the 2×3 Jacobian matrix

$$DG = \begin{pmatrix} P_x & P_y & P_\lambda \\ Q_x & Q_y & Q_\lambda \end{pmatrix}.$$

This is in contrast to dG, which denotes the 2×2 Jacobian matrix of derivatives of G with respect to x and y only.

The *bifurcation variety* $\mathscr{B}$ consists of those $\alpha \in \mathbb{R}^k$ for which the curve

$$\{(z, \lambda) \in \mathbb{R}^2 \times \mathbb{R} \colon G(z, \lambda, \alpha) = 0\} \tag{3.3}$$

contains a singular point. It follows from the implicit function theorem that (z, λ) can be a singular point of (3.3) only if $(DG)_{z,\lambda,\alpha}$ fails to be surjective; otherwise, we could solve $G = 0$ for two of the variables as functions of the third. Therefore, we define

$$\begin{aligned} \mathscr{B} = \{\alpha \in \mathbb{R}^k \colon \exists (z, \lambda) \text{ such that } & G(z, \lambda, \alpha) = 0 \\ & \text{and rank } (DG)_{z,\lambda,\alpha} \le 1\}. \end{aligned} \tag{3.4}$$

The *hysteresis variety* $\mathscr{H}$ consists of those $\alpha \in \mathbb{R}^k$ for which the curve (3.3) makes at least quadratic contact with the vertical planes $\{\lambda = \text{const}\}$. Let us define quadratic contact. Suppose that we parametrize a portion of the curve (3.3), say

$$\phi(t) = (x(t), y(t), \lambda(t)); \tag{3.5}$$

we say that ϕ makes *quadratic contact* with the plane $\{\lambda = \text{const}\}$ at $t = 0$ if the third component in (3.5) satisfies

$$\text{(a)} \quad \lambda'(0) = 0, \qquad \text{(b)} \quad \lambda''(0) = 0. \tag{3.6}$$

(To avoid trivialities, we insist that $\phi'(0) \neq 0$.) We derive necessary conditions for quadratic contact as follows. Differentiating the identity $G(\phi(t), \alpha) \equiv 0$ with respect to t and evaluating at $t = 0$, we find

$$\begin{aligned} &\text{(a)} \quad (DG_{\phi(0),\alpha} \cdot \phi'(0) = 0, \\ &\text{(b)} \quad (D^2G)_{\phi(0),\alpha}(\phi'(0), \phi'(0)) + (DG)_{\phi(0),\alpha} \cdot \phi''(0) = 0. \end{aligned} \tag{3.7}$$

Equation (3.7a) is shorthand for a sum involving three terms; we group the terms as

$$(dG)_{\phi(0),\alpha} \cdot v_0 + (G_\lambda)_{\phi(0),\alpha} \cdot \lambda'(0) = 0,$$

where $v_0 = (x'(0), y'(0))$. If (3.6a) holds, the second term here vanishes. In other words, if (3.6a) holds, there is a nonzero vector v_0 such that

$$(dG)_{\phi(0),\alpha} \cdot v_0 = 0.$$

(In particular, $\det(dG)_{\phi(0),\alpha} = 0$.) Similarly, if (3.6b) holds, we deduce from (3.7b) that there is a vector w_0 such that

$$(dG)_{\phi(0),\alpha} \cdot w_0 = (d^2G)_{\phi(0),\alpha}(v_0, v_0);$$

viz., $w_0 = -(x''(0), y''(0))$. Thus we define

$$\begin{aligned}\mathscr{H} = \{\alpha \in \mathbb{R}^k\colon \exists(z, \lambda, v) \text{ such that } & G(z, \lambda, \alpha) = 0, v \neq 0,\\ &(dG)_{z,\lambda,\alpha} \cdot v = 0, \text{ and}\\ &(d^2G)_{z,\lambda,\alpha}(v, v) \in \text{range } (dG)_{z,\lambda,\alpha}\}.\end{aligned} \tag{3.8}$$

We extract an equation for the *double limit point variety* $\mathscr{D}$ from the above analysis as follows. Limit points satisfy (3.6a). Above we showed that if (3.6a) holds, then $\det(dG)_{\phi(0),\alpha} = 0$. Thus we define

$$\begin{aligned}\mathscr{D} = \{\alpha \in \mathbb{R}^k\colon \exists(z_1, z_2, \lambda) \text{ such that } & z_1 \neq z_2, G(z_i, \lambda, \alpha) = 0, \text{ and}\\ &\det(dG)_{z_i,\lambda,\alpha} = 0, i = 1, 2\}.\end{aligned} \tag{3.9}$$

(c) Formulas and Pictures for the Transition Variety

In Table 3.1 we list the equations of the three varieties $\mathscr{B}$, $\mathscr{H}$, and $\mathscr{D}$ for the three universal unfoldings (3.2). We leave it to the reader to check these formulas. The computations for $\mathscr{B}$ are easy, while those for $\mathscr{H}$ and $\mathscr{D}$ are somewhat lengthy.

Even though we have listed explicit formulas for Σ, it is still a complicated task to enumerate the connected components of $\mathbb{R}^3 \sim \Sigma$. We simplify this enumeration by making a singular change of coordinates in the unfolding parameters α, β, γ. This transformation changes the order of contact of the varieties $\mathscr{B}$, $\mathscr{H}$ and $\mathscr{D}$, but it leaves the number of components of $\mathbb{R}^3 \sim \Sigma$ and their relative positions unchanged. To illustrate this, let us consider (3.2a). We define ρ, θ, and δ by the equations

$$\alpha = \rho \cos\theta, \qquad \beta = \rho \sin\theta \quad \text{and} \quad \gamma = \rho^2\delta^3. \tag{3.10}$$

Thus (ρ, θ, δ) are a singular modification of cylindrical coordinates in $\alpha\beta\gamma$-space. In these coordinates $\mathscr{B}$ and $\mathscr{H}$ have the equations

$$\begin{aligned}&\mathscr{B}\colon \{\delta = 0\},\\ &\mathscr{H}\colon \{4\delta^3 - 3\delta = -\sin 2\theta\}.\end{aligned} \tag{3.11}$$

Table 3.1. Equations for the Transition Variety in Hilltop Bifurcation.

	$\mathscr{B}$	$\mathscr{H}$	$\mathscr{D}$
(a)	$\gamma = 0$	$(2\gamma + \alpha\beta)^3 = \frac{27}{8}(\alpha^2 + \beta^2)^2\gamma$	$\varnothing$
(b)	$\gamma = \frac{1}{2}(\alpha^2 - \beta^2)$	$\gamma = -\frac{3}{2}\alpha^{2/3}\beta^{2/3}(\alpha^{2/3} - \beta^{2/3})$	$\{\alpha = 0, \gamma \leq 0\} \cup \{\beta = 0, \gamma \geq 0\}$
(c)	$\gamma = \frac{1}{2}(\alpha^2 + \beta^2)$	$\gamma = -\frac{3}{2}\alpha^{2/3}\beta^{2/3}(\alpha^{2/3} + \beta^{2/3})$	$\{\alpha = 0, \gamma \leq 0\} \cup \{\beta = 0, \gamma \leq 0\}$

In other words, ρ *drops out of the equations*. Thus we may describe $\mathscr{B}$ and $\mathscr{H}$ by two-dimensional graphs in θ, δ-space. We have graphed (3.11) in Figure 3.2(a). We view this figure as the intersection of Σ with the cylinder $\{\rho = 1\}$ in (α, β, γ)-space. The full three-dimensional graph of Σ may be obtained from Figure 3.2(a) by considering all values of ρ and scaling according to (3.10). In this way, we conclude from Figure 3.2(a) that $\mathbb{R}^3 \sim \Sigma$ contains ten connected components.

Similarly, we can find the connected components of the complement of Σ for the examples (3.2b) and (3.2c) with the singular scalings:

$$\alpha = \rho^3 \cos^3 \theta, \qquad \beta = \rho^3 \sin^3 \theta, \qquad \gamma = \rho^6 \delta.$$

In these coordinates, $\mathscr{B}$, $\mathscr{H}$, and $\mathscr{D}$ have equations as follows:

for (3.2b)

$$\mathscr{B}: \{\delta = \cos(2\theta)(4 - \sin^2(2\theta))/8\},$$

$$\mathscr{H}: \{\delta = -3\cos(2\theta)\sin^2(2\theta)/8\},$$

$$\mathscr{D}: \left\{\theta = \frac{\pi}{2}, \frac{3\pi}{2}; \delta \leq 0\right\} \cup \left\{\theta = 0, \pi; \delta \geq 0\right\},$$

and for (3.2c)

$$\mathscr{B}: \{\delta = (4 - 3\sin^2(2\theta))/8\},$$

$$\mathscr{H}: \{\delta = -3\sin^2(2\theta)/8\},$$

$$\mathscr{D}: \left\{\theta = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}; \delta \leq 0\right\}.$$

We have drawn these graphs in Figure 3.2(b), (c); in the figures we have also enumerated the components of $\mathbb{R}^3 \sim \Sigma$.

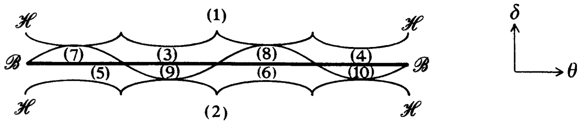
Finally, for the normal form (3.2c) we have sketched the full three-dimensional graph of Σ in $\alpha\beta\gamma$-space in Figure 3.3.

(d) The Perturbed Bifurcation Diagrams

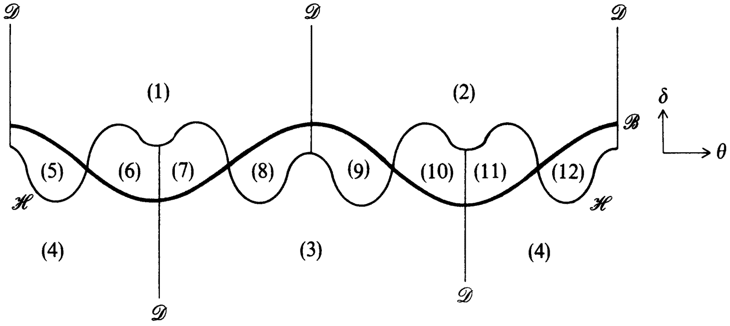
Now we discuss the perturbed bifurcation diagrams associated to the various regions in Figure 3.2. We consider (3.2a) in some detail, and we merely present the results for (3.2b, c). Our analysis relies heavily on the fact that for certain values of the parameters α, β, γ, a pitchfork bifurcation occurs in the unfoldings (3.2). (Cf. the analysis of the winged cusp in Chapter III, §8.)

Consider the bifurcation problem obtained by setting $\beta = \gamma = 0$ in the unfolding (3.2a); viz.,

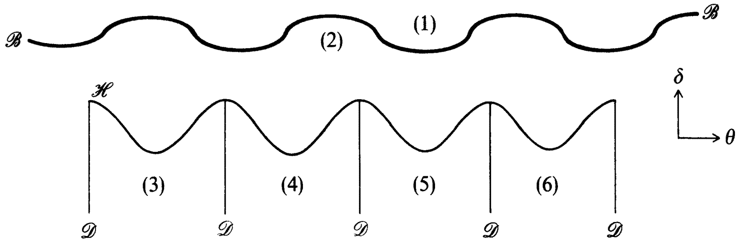
$$\begin{aligned} &\text{(a)} \quad x^2 - y^2 + \lambda + 2\alpha x = 0, \\ &\text{(b)} \quad 2xy = 0, \end{aligned} \tag{3.12}$$



(a) Transition variety for $(x^2 - y^2 + \lambda, 2xy)$.



(b) Transition variety for $(x^2 - \lambda, y^2 - \lambda)$.



(c) Transition variety for $(x^2 + \lambda, y^2 - \lambda)$.

Figure 3.2. Transition varieties for hilltop bifurcation.

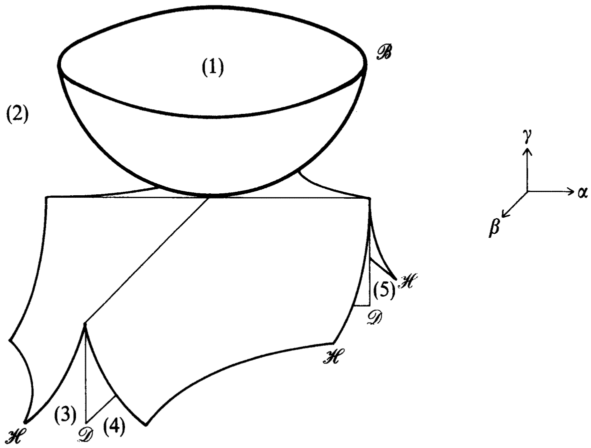


Figure 3.3. Three-dimensional sketch of the transitions variety for $(x^2 + \lambda, y^2 - \lambda)$. See Fig. 3.2(c).

where $\alpha \neq 0$. The solution set of (3.12) consists of two parabolas,

$$\begin{aligned} &\text{(a)} \quad x = 0, \qquad -y^2 + \lambda = 0, \\ &\text{(b)} \quad y = 0, \qquad x^2 + 2\alpha x + \lambda = 0. \end{aligned} \tag{3.13}$$

These parabolas intersect at $x = y = \lambda = 0$, as illustrated in Figure 3.4. At first glance, Figure 3.4 may seem rather similar to the unperturbed bifurcation diagram of Figure 3.1(a); the difference is that in Figure 3.1(a) the intersection occurs at the vertex of both parabolas, while in Figure 3.4 it occurs off center of the parabola (3.13b). The intersection point is actually a pitchfork bifurcation. This means that at the intersection point, the Jacobian of (3.12) has rank 1 and the Liapunov–Schmidt reduction of (3.12) leads to a bifurcation problem in one variable that is equivalent to the pitchfork. We ask the reader to verify this in Exercise 3.1. (*Remark:* Note that both (3.1a) and the perturbed problem (3.12) commute with the reflection $(x, y) \to (x, -y)$. It is no accident that the pitchfork survives the perturbation.)

Let us locate the perturbation (3.12) in Figure 3.2(a). We see from (3.10) that for the perturbation (3.12), $\delta = 0$ and $\theta = 0$ or π according as $\alpha > 0$ or $\alpha < 0$, respectively. Now we know from the one variable theory that at a pitchfork bifurcation point, the bifurcation variety $\mathscr{B}$ and the hysteresis variety $\mathscr{H}$ intersect one another and have cubic contact. In Figure 3.2(a), $\mathscr{B}$ and $\mathscr{H}$ do indeed intersect one another at $\delta = 0$ and $\theta = 0, \pi$, but the intersection appears to be transverse. However, this is misleading—because of the scaling (3.10), in $\alpha\beta\gamma$-space, $\mathscr{B}$ and $\mathscr{H}$ have cubic contact.

For definiteness let us take $\alpha < 0$ in (3.12). This corresponds to $\delta = 0$, $\theta = \pi$ in Figure 3.2(a). Four separate regions of $\mathbb{R}^3 \sim \Sigma$ abut the point $\delta = 0$, $\theta = \pi$ in Figure 3.2(a); viz., 3, 6, 8, and 9. This is to be expected from the one variable theory, as the parameters β and γ provide a universal unfolding of (3.12). (See Exercise 3.1.) The bifurcation diagrams associated to these regions may be obtained from Figure 3.4 by splitting apart the

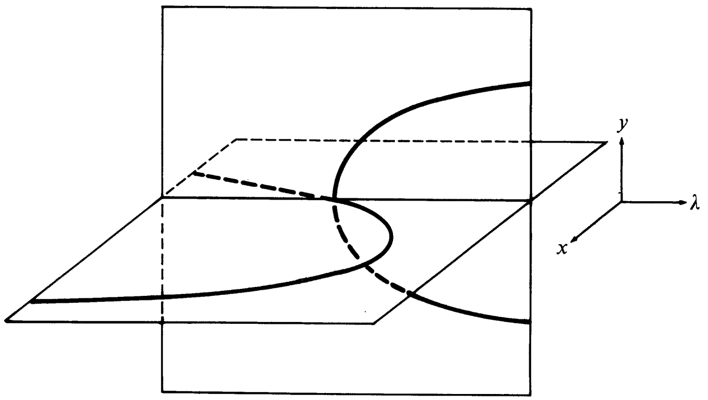
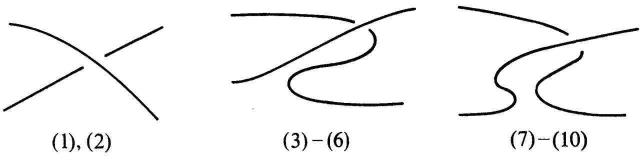


Figure 3.4 Bifurcation diagram of (3.12).

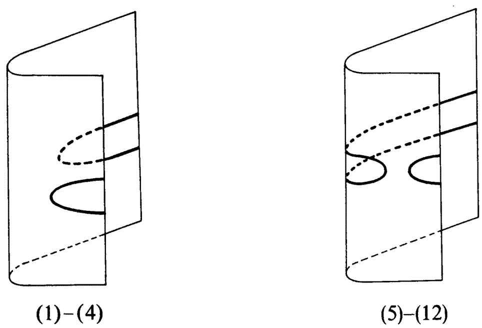
pitchfork in various ways. The resulting bifurcation diagrams have either two or four limit points, as illustrated in Figure 3.5(a). We associate the diagrams containing four limit points with regions 7 and 8, because after the scaling (3.10) these are the "thin" regions in the universal unfolding of the pitchfork provided by β and γ.

There are four pitchfork points in Figure 3.2(a), at $\theta = 0$, $\pi/2$, π, and $3\pi/2$. Two of these are associated with (3.12), and two with a similar perturbation where $\alpha = \gamma = 0$. By perturbing these four pitchfork points we may account for the bifurcation diagrams of regions 3 through 10 in Figure 3.2(a). The remaining two regions, 1 and 2, are associated with bifurcation diagrams that have no limit points, as shown in Figure 3.5(a).

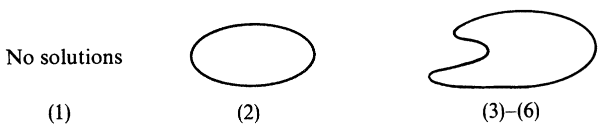
The analysis of (3.2b) is similar. There are four points in Figure 3.2(b) where $\mathscr{B}$ and $\mathscr{H}$ intersect, and these yield pitchfork bifurcations. Every region in Figure 3.2(b) abuts a pitchfork point. Thus every perturbed diagram can be obtained from one of the pitchforks. The possible diagrams are shown in Figure 3.5(b); they contain either two or four limit points.



(a) Persistent perturbations of $(x^2 - y^2 + \lambda, 2xy)$.



(b) Persistent perturbations of $(x^2 - \lambda, y^2 - \lambda)$.



(c) Persistent perturbations of $(x^2 + \lambda, y^2 - \lambda)$.

Figure 3.5. Persistent perturbations in hilltop bifurcation. Numbers refer to regions of Fig. 3.2.

The last example, (3.2c), is a two-dimensional version of an isola center. Again, the number of limit points determines the perturbed diagrams; zero, two, or four are possible. These are shown in Figure 3.5(c).

Note that double limit points occur for (3.2b) and (3.2c). Indeed, there are points where $\mathscr{D}$ intersects $\mathscr{H}$. The singularity which occurs at such intersections is the quartic fold $x^4 - \lambda$. We ask the reader to prove this in Exercise 3.2.

Exercises

3.1. (a) Show that when $\alpha \neq 0$ the bifurcation problem (3.12) has a pitchfork bifurcation $\varepsilon(y^3 - \lambda y)$ near $(x_0, y_0, \lambda_0) = (0, 0, 0)$ where $\varepsilon = \text{sgn}(\alpha)$. *Hint*: Use the implicit function theorem to solve (3.12a) for $x = x(y, \lambda)$, $x(0, 0) = 0$. Next show that $x_\lambda(0, 0) = -1/2\alpha$, $x_y(0, 0) = 0$ and $x_{yy}(0, 0) = 1/\alpha$.

(b) Verify that (3.2a) is a universal unfolding of (3.12a) using Proposition III,4.4.

3.2. (a) Show that (3.2b, c) have quartic folds $(\varepsilon x^4 + \delta\lambda)$ at parameter values in $\mathscr{H} \cap \mathscr{D}$. (See Table 3.1. $\mathscr{H} \cap \mathscr{D}$ consists of the two lines $\alpha = \gamma = 0$, $\beta \neq 0$ and $\beta = \gamma = 0$, $\alpha \neq 0$.)

(b) Show that (3.2b, c) is a universal unfolding of the quartic folds using Table IV,3.2(7).

Bibliographical Comments

The primary sources for this chapter are Golubitsky and Schaeffer [1979a] and Thompson and Hunt [1973].

CHAPTER X

Two Degrees of Freedom with $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-Symmetry

§0. Introduction

In this chapter we explore the role of symmetry in the study of bifurcation problems. Specifically, we analyze a family of bifurcation problems in two state variables that commute with the group $\mathbf{Z}_2 \oplus \mathbf{Z}_2$. Although even in this special case our discussion is necessarily incomplete, it indicates the directions of the general theory.

There are several specific benefits to be derived from understanding this material. Most pragmatically, we will use the results of this chapter in Case Study 3. Pedagogically, the examples of this chapter provide a wonderful illustration of the importance of moduli in the general theory—the simplest bifurcation problems in the $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-symmetric context have codimension three, and *two of the three unfolding parameters are moduli*. Finally, this chapter may help to overcome a somewhat pessimistic impression left from Chapter IX about the intractibility of bifurcation problems in several state variables. When symmetry is present, group theory offers powerful techniques for organizing complicated information, and these techniques have not yet been pushed even close to their limits. On the contrary, it seems quite plausible to us that further research in this area may lead to striking new results.

There are four sections in this chapter. In §1 we describe the general setting for $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-symmetry. In §2, we present the basic singularity theory results for the least degenerate $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-symmetric bifurcation problems. In §3, we discuss the invariance of stability under $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-equivalence. (We saw in Chapter IX, §1 that in general equivalence transformations in many state variables do not respect the stability of solutions; somewhat surprisingly, symmetry alters this situation.) Finally, in §4 we draw the bifurcation diagrams.

§1. Bifurcation Problems with $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-Symmetry

In this section we discuss three points, as follows:

(a) The action of the group $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ on $\mathbb{R}^2$.
(b) Restrictions on bifurcation problems g commuting with this action of $\mathbf{Z}_2 \oplus \mathbf{Z}_2$.
(c) Solution types of the equation of $g = 0$.

There is an intimate connection between points (a) and (c). Specifically, we shall see that solutions of the equation $g = 0$ are naturally classified into several different types which correspond exactly to the group action.

(a) The Action of $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ on $\mathbb{R}^2$.

The group $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ has four elements (ε, δ) where $\varepsilon = \pm 1$ and $\delta = \pm 1$. The group element (ε, δ) acts on the point $(x, y) \in \mathbb{R}^2$ by

$$(\varepsilon, \delta) \cdot (x, y) = (\varepsilon x, \delta y). \tag{1.1}$$

We may, of course, think of the action of (ε, δ) on $\mathbb{R}^2$ as a linear mapping; the matrix associated to the action of (ε, δ) is the diagonal matrix

$$\begin{pmatrix} \varepsilon & 0 \\ 0 & \delta \end{pmatrix}. \tag{1.2}$$

The behavior of the action of $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ on $\mathbb{R}^2$ is different at different points in $\mathbb{R}^2$. Moreover, these differences are the root cause of much of the structure we find in bifurcation problems with symmetry. We describe these differences in two ways: through orbits and through isotropy subgroups.

The *orbit* of a point (x, y) under the action of $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ is the set of points

$$\{(\varepsilon, \delta) \cdot (x, y) : (\varepsilon, \delta) \in \mathbf{Z}_2 \oplus \mathbf{Z}_2\}.$$

It is easy to see that there are four orbit types:

(a) The origin, $(0, 0)$,
(b) Points on the x-axis, $(\pm x, 0)$ with $x \neq 0$,
(c) Points on the y-axis, $(0, \pm y)$ with $y \neq 0$,
(d) Points off the axes, $(\pm x, \pm y)$ with $x \neq 0$, $y \neq 0$. (1.3)

From (1.3) we see that orbits have either 1, 2, or 4 points; the origin is distinguished, being the unique one point orbit.

Isotropy subgroups provide another way of describing the differences between different points in $\mathbb{R}^2$ under this action. The isotropy subgroup of a point (x, y) is the set of symmetries preserving that point. In symbols, the *isotropy subgroup* of the point (x, y) is

$$\{(\varepsilon, \delta) \in \mathbf{Z}_2 \oplus \mathbf{Z}_2 : (\varepsilon, \delta) \cdot (x, y) = (x, y)\}.$$

Again, it is easy to see that there are four isotropy subgroups:

$$\begin{aligned}
&\text{(a)} \quad \mathbf{Z}_2 \oplus \mathbf{Z}_2 \text{ corresponding to the origin,}\\
&\text{(b)} \quad \mathbf{Z}_2 = \{(1, \delta)\} \text{ corresponding to } (x, 0) \text{ with } x \neq 0,\\
&\text{(c)} \quad \mathbf{Z}_2 = \{(\varepsilon, 1)\} \text{ corresponding to } (0, y) \text{ with } y \neq 0,\\
&\text{(d)} \quad \mathbf{1} = \{(1, 1)\} \text{ corresponding to } (x, y) \text{ with } x \neq 0, y \neq 0.
\end{aligned} \tag{1.4}$$

Note that orbit types in (1.3a, b, c, d) have isotropy subgroups (1.4a, b, c, d), respectively.

The ideas of orbit type and isotropy subgroup extend to a general group action. As we will see in Volume II, these ideas play an important role in determining the structure of bifurcation problems with symmetry. In this chapter, however, we will restrict the discussion to bifurcation problems with $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-symmetry.

(b) The Form of $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-Symmetric Bifurcation Problems

Let $g: \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ be a bifurcation problem with two state variables; that is, let g be C^∞ and satisfy

$$g(0, 0) = 0, \qquad (dg)_{0,0} = 0. \tag{1.5}$$

(Cf. Definition IX,1.1.) We say that the bifurcation problem g *commutes with the group* $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ if

$$g((\varepsilon, \delta) \cdot (x, y), \lambda) = (\varepsilon, \delta) \cdot g(x, y, \lambda). \tag{1.6}$$

(Cf. Chapter VII, §3(a).) In the next lemma we describe explicitly the form that (1.6) imposes on g. (Cf. Lemma VI,2.1 and Corollary VI,2.2 for the analogous properties when g has only one state variable and commutes with the group $\mathbf{Z}_2$.)

Lemma 1.1. *Let $g: \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ be a bifurcation problem in two state variables commuting with the action of $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ in* (1.1). *Then there exist smooth functions $p(u, v, \lambda)$, $q(u, v, \lambda)$ such that*

$$g(x, y, \lambda) = (p(x^2, y^2, \lambda)x, q(x^2, y^2, \lambda)y), \tag{1.7a}$$

where

$$p(0, 0, 0) = 0, \qquad q(0, 0, 0) = 0. \tag{1.7b}$$

PROOF. We write g in coordinates

$$g(x, y, \lambda) = (a(x, y, \lambda), b(x, y, \lambda)).$$

The commutativity relation (1.6) implies that

$$\begin{aligned} a(\varepsilon x, \delta y, \lambda) &= \varepsilon a(x, y, \lambda), \\ b(\varepsilon x, \delta y, \lambda) &= \delta b(x, y, \lambda). \end{aligned} \tag{1.8}$$

When $\varepsilon = -1$ and $\delta = +1$, equation (1.8) shows that a is odd in x and b is even in x. When $\varepsilon = +1$ and $\delta = -1$, equation (1.8) shows that a is even in y and b is odd in y. It follows from Taylor's theorem that we may factor these functions,

$$\begin{aligned} a(x, y, \lambda) &= \tilde{a}(x, y, \lambda)x, \\ b(x, y, \lambda) &= \tilde{b}(x, y, \lambda)y, \end{aligned} \tag{1.9}$$

where $\tilde{a}$ and $\tilde{b}$ are even in both x and y. Applying Lemma VI,2.1, first to x and then to y, we conclude that g has the desired form (1.7a).

Regarding (1.7b), we recall from (1.5) that the linear terms in g vanish. The only linear terms compatible with the symmetry are

$$(p(0, 0, 0)x, q(0, 0, 0)y);$$

thus $p(0, 0, 0) = q(0, 0, 0) = 0$.

(c) Solution Types for g

Consider solving the equation $g = 0$ when g has the form (1.7a). There are four solution types of such an equation, which occur according as the first or second factor in $p(x^2, y^2, \lambda)x$ vanishes and the first or second factor in $q(x^2, y^2, \lambda)y$ vanishes. Specifically, we have solution types

$$\begin{aligned} &\text{(a)} \quad x = y = 0, \\ &\text{(b)} \quad p(x^2, 0, \lambda) = 0, y = 0, x \neq 0, \\ &\text{(c)} \quad x = 0, q(0, y^2, \lambda) = 0, y \neq 0, \\ &\text{(d)} \quad p(x^2, y^2, \lambda) = 0, q(x^2, y^2, \lambda) = 0, x \neq 0, y \neq 0. \end{aligned} \tag{1.10}$$

Moreover, these solution types correspond exactly to the orbit types listed in (1.3) of the action of $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ on $\mathbb{R}^2$.

We use the following terminology for these four types of solutions:

(a) trivial solution,
(b) x-mode solutions, (1.11)
(c) y-mode solutions,
(d) mixed mode solutions.

We see from (1.3) that each solution type has its own characteristic multiplicity. Specifically, x-mode and y-mode solutions always come in pairs, $(\pm x, 0)$ and $(0, \pm y)$; and mixed mode solutions always come four at a time, $(\pm x, \pm y)$.

It is instructive to compare the present two state variables, $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-context with the one state variable, $\mathbf{Z}_2$-context of Chapter VI. In the latter context, there are two solution types; the trivial solution $(x = 0)$ and nontrivial solutions $(x \neq 0)$. Moreover, nontrivial solutions always come in pairs, $\pm x$. Thus the present context exhibits similar, but richer, structure. Group theoretic methods are the natural tool for exploring this structure systematically.

The buckling problems described in Chapter I, §1; Chapter VI, §1; and Chapter VII, §1 all lead to one state variable bifurcation problems with $\mathbf{Z}_2$-symmetry. In these models it is simple to relate the group theory concepts to the physical situation. For example, the trivial solution corresponds to the unbuckled state; nontrivial solutions occur in pairs, because for each "buckled-up" state, there is a symmetric "buckled-down" state. Similarly, in Case Study 3 we study a bifurcation problem with $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-symmetry; the group theory concepts will be clarified by seeing them in this physical situation. In particular, the terminology in (1.11) derives from this situation.

§2. Singularity Theory Results

We divide this section into three subsections, as follows:

(a) Equivalence in the $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-symmetric context.
(b) The recognition problem for the simplest bifurcation problems with $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-symmetry.
(c) Universal unfoldings for these simplest singularities.

Thus, subsections (a) and (b) discuss how to generalize the ideas of Chapter II to the present context; subsection (c), how to generalize the first half of Chapter III. (In §4, we will consider the generalization of the second half of Chapter III; i.e., the perturbed bifurcation diagrams.)

The singularities we describe here have codimension three and modality two. This emphasizes the need to study moduli when symmetry groups are present.

(a) $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-Equivalence

Let $g, h\colon \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ be bifurcation problems with two state variables commuting with the action of $\mathbf{Z}_2 \oplus \mathbf{Z}_2$. We say that g and h are $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-*equivalent* if g and h are equivalent (in the sense of Definition IX,1.2) and, in addition, the equivalence preserves the symmetry. Let us elaborate. Recall that g and h are equivalent if there exists a 2×2 invertible matrix $S(x, y, \lambda)$ depending smoothly on x, y, and λ and a diffeomorphism $\Phi(x, y, \lambda) = (Z(x, y, \lambda), \Lambda(\lambda))$ satisfying

$$g(x, y, \lambda) = S(x, y, \lambda)h(Z(x, y, \lambda), \Lambda(\lambda)), \tag{2.1}$$

such that

$$\Phi(0, 0, 0) = (0, 0, 0) \quad \text{and} \quad \Lambda'(0) > 0. \tag{2.2}$$

We say that the equivalence S, Φ preserves the symmetry if

$$\begin{aligned} &\text{(a)} \quad Z(\varepsilon x, \delta y, \lambda) = (\varepsilon, \delta) \cdot Z(x, y, \lambda), \\ &\text{(b)} \quad S(\varepsilon x, \delta y, \lambda)\begin{pmatrix} \varepsilon & 0 \\ 0 & \delta \end{pmatrix} = \begin{pmatrix} \varepsilon & 0 \\ 0 & \delta \end{pmatrix} S(x, y, \lambda). \end{aligned} \tag{2.3}$$

The restriction (2.3a) states that Z commutes with the group $\mathbf{Z}_2 \oplus \mathbf{Z}_2$. The restriction (2.3b) is precisely the condition on S needed to guarantee that our equivalence transformations have the following property: If h is a bifurcation problem commuting with the group $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ and if g is $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-equivalent to h, then g commutes with $\mathbf{Z}_2 \oplus \mathbf{Z}_2$.

Condition (2.3) restricts the form of Z and S in the following ways. Since Z commutes with $\mathbf{Z}_2 \oplus \mathbf{Z}_2$, (2.3a), we may apply Lemma 1.1 to show that

$$Z(x, y, \lambda) = (a(x^2, y^2, \lambda)x, b(x^2, y^2, \lambda)y). \tag{2.4}$$

Thus

$$(dZ)_{0,0,0} = \begin{pmatrix} a(0, 0, 0) & 0 \\ 0 & b(0, 0, 0) \end{pmatrix};$$

in words, $(dZ)_{0,0,0}$ is diagonal. Turning now to S, we write out the entries of S as

$$\begin{pmatrix} S_1(x, y, \lambda) & S_2(x, y, \lambda) \\ S_3(x, y, \lambda) & S_4(x, y, \lambda) \end{pmatrix}.$$

A short calculation using (2.3b) shows that S_1 and S_4 are even in both x and y, and that S_2 and S_3 are odd in both x and y. Thus, Lemma VI,2.1 coupled with Taylor's theorem implies that

$$S(x, y, \lambda) = \begin{pmatrix} c_1(x^2, y^2, \lambda) & c_2(x^2, y^2, \lambda)xy \\ c_3(x^2, y^2, \lambda)xy & c_4(x^2, y^2, \lambda) \end{pmatrix}. \tag{2.5}$$

In particular

$$S(0, 0, 0) = \begin{pmatrix} c_1(0, 0, 0) & 0 \\ 0 & c_4(0, 0, 0) \end{pmatrix}; \tag{2.6}$$

Thus $S(0, 0, 0)$ is also diagonal.

We shall require that $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-equivalences satisfy

$$a(0, 0, 0) > 0, \qquad b(0, 0, 0) > 0, \qquad c_1(0, 0, 0) > 0, \qquad c_4(0, 0, 0) > 0. \tag{2.7}$$

Restriction (2.7) stems from a desire to have $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-equivalences preserve linearized stability. This issue will be discussed in detail in §3.

To summarize we have:

Definition 2.1. Two bifurcation problems g and h, both commuting with the group $\mathbf{Z}_2 \oplus \mathbf{Z}_2$, are $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-*equivalent* if there exists S and $\Phi = (Z, \Lambda)$ as above satisfying (2.1), (2.2), (2.4), (2.5), and (2.7).

(b) The Recognition Problem for the Simplest Examples

Let g be a bifurcation problem with two state variables commuting with the group $\mathbf{Z}_2 \oplus \mathbf{Z}_2$. Thus g has the form (1.7). Let us split off the lowest-order terms in (1.7), say

$$g(x, y, \lambda) = k(x, y, \lambda) + \text{hot}, \tag{2.8a}$$

where

$$k(x, y, \lambda) = (Ax^3 + Bxy^2 + \alpha\lambda x, Cx^2y + Dy^3 + \beta\lambda y). \tag{2.8b}$$

The higher-order terms in (2.8a) include all monomials $x^r y^s \lambda^t$ satisfying at least one of the following conditions:

$$\begin{aligned} &\text{(a)} \quad r + s \geq 5, \\ &\text{(b)} \quad t = 1, r + s \geq 3, \\ &\text{(c)} \quad t \geq 2. \end{aligned}$$

The simplest bifurcation problems of the form (2.8) are those which satisfy the following list of nondegeneracy conditions. It is these singularities that we study for the remainder of this chapter.

Definition 2.2. The bifurcation problem g in (2.8) is *nondegenerate* if all the following conditions are satisfied:

$$\begin{aligned} &\text{(a)} \quad A \neq 0, \qquad D \neq 0, \\ &\text{(b)} \quad \alpha \neq 0, \qquad \beta \neq 0, \\ &\text{(c)} \quad A\beta - C\alpha \neq 0, \qquad B\beta - D\alpha \neq 0, \\ &\text{(d)} \quad AD - BC \neq 0. \end{aligned} \tag{2.9}$$

In the following proposition we solve the recognition problem for nondegenerate bifurcation problems commuting with $\mathbf{Z}_2 \oplus \mathbf{Z}_2$.

Proposition 2.3. *Let $g\colon \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ be a bifurcation problem in two state variables commuting with the group $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ and satisfying the nondegeneracy conditions* (2.9). *Then g is* ($\mathbf{Z}_2 \oplus \mathbf{Z}_2$)*-equivalent to*

$$h(x, y, \lambda) = (\varepsilon_1 x^3 + mxy^2 + \varepsilon_2 \lambda x, nx^2 y + \varepsilon_3 y^3 + \varepsilon_4 \lambda y), \tag{2.10}$$

where

$$\begin{aligned} &\text{(a)}\quad \varepsilon_1 = \operatorname{sgn}(A), && \varepsilon_3 = \operatorname{sgn}(D),\\ &\text{(b)}\quad \varepsilon_2 = \operatorname{sgn}(\alpha), && \varepsilon_4 = \operatorname{sgn}(\beta),\\ &\text{(c)}\quad m = \left|\frac{\beta}{D\alpha}\right| B, && n = \left|\frac{\alpha}{A\beta}\right| C. \end{aligned} \tag{2.11}$$

Moreover

$$m \neq \varepsilon_2\varepsilon_3\varepsilon_4, \qquad n \neq \varepsilon_1\varepsilon_2\varepsilon_4, \quad mn \neq \varepsilon_1\varepsilon_3. \tag{2.12}$$

Remarks. (i) The normal form h in (2.10) depends on two parameters m and n satisfying the nondegeneracy conditions (2.12). These are the two modal parameters promised above.

(ii) The proof of this proposition divides into two parts. In the first part, one uses linear ($\mathbf{Z}_2 \oplus \mathbf{Z}_2$)-equivalences (i.e., scalings) to transform k to the normal form h. In the second part, one shows that the higher-order terms may be transformed away by a nonlinear ($\mathbf{Z}_2 \oplus \mathbf{Z}_2$)-equivalence. The second part will be presented in Volume II; here, we give only the scaling argument.

PROOF. The most general linear ($\mathbf{Z}_2 \oplus \mathbf{Z}_2$)-equivalence is given by

$$Z(x, y, \lambda) = (ax, by),$$

$$\Lambda(\lambda) = \sigma\lambda,$$

$$S(x, y, \lambda) = \begin{pmatrix} c & 0 \\ 0 & d \end{pmatrix},$$

where a, b, σ, c, and d are positive constants. Letting this equivalence act on $k(x, y, \lambda)$, which is given by (2.8b), we find

$$\begin{pmatrix} c & 0 \\ 0 & d \end{pmatrix} k(ax, by, \sigma\lambda) = (ca^3Ax^3 + cab^2Bxy^2 + ca\sigma\alpha\lambda x,$$

$$da^2bCx^2y + db^3Dy^3 + db\sigma\lambda y). \tag{2.13}$$

To obtain the normal form (2.10), we need

$$\begin{aligned} &\text{(a)}\quad ca^3|A| = 1,\\ &\text{(b)}\quad ca\sigma|\alpha| = 1,\\ &\text{(c)}\quad db^3|D| = 1,\\ &\text{(d)}\quad db\sigma|\beta| = 1. \end{aligned}$$

We solve these equations by setting

$$\begin{aligned}
&\text{(a)}\quad c = \frac{1}{a^3|A|},\\
&\text{(b)}\quad d = \frac{1}{b^3|D|},\\
&\text{(c)}\quad \sigma = a^2\left|\frac{A}{\alpha}\right|,\\
&\text{(d)}\quad \frac{a}{b} = \sqrt{\left|\frac{D\alpha}{A\beta}\right|}.
\end{aligned} \tag{2.14}$$

Substitution of (2.14) into the right-hand side of (2.13) yields the normal form (2.10) with m and n given by (2.11c). The restrictions (2.12) follow from the nondegeneracy conditions (2.9c, d). □

(c) Universal Unfoldings for Nondegenerate ($\mathbf{Z}_2 \oplus \mathbf{Z}_2$)-Bifurcation Problems

Universal unfoldings in the ($\mathbf{Z}_2 \oplus \mathbf{Z}_2$)-symmetric context are defined in the natural way using ($\mathbf{Z}_2 \oplus \mathbf{Z}_2$)-equivalence. (Cf. Definitions III,2.1 and VI,2.5.) We defer giving precise definitions until Volume II. Also in Volume II we shall derive the following universal unfolding for the normal form (2.10).

Theorem 2.4. *Let $h(x, y, \lambda)$ be the normal form* (2.10) *satisfying the nondegeneracy conditions* (2.12). *Then*

$$H(x, y, \lambda, \tilde{m}, \tilde{n}, \sigma) = (\varepsilon_1 x^3 + \tilde{m}xy^2 + \varepsilon_2 \lambda x, \tilde{n}x^2 y + \varepsilon_3 y^3 + \varepsilon_4(\lambda - \sigma)y) \tag{2.15}$$

is a ($\mathbf{Z}_2 \oplus \mathbf{Z}_2$)-universal unfolding of h. Here $(\tilde{m}, \tilde{n}, \sigma)$ varies on a neighborhood of $(m, n, 0)$.

One consequence of Theorem 2.4 is that the nondegenerate normal forms h have ($\mathbf{Z}_2 \oplus \mathbf{Z}_2$)-codimension three and modality two. Moreover, the universal unfolding (2.15) is topologically trivial provided m and n are nonzero and (2.12) holds. Thus according to our definition in Chapter V, §6, the ($\mathbf{Z}_2 \oplus \mathbf{Z}_2$)-topological codimension of h is one. In other words, up to topological equivalence, small ($\mathbf{Z}_2 \oplus \mathbf{Z}_2$)-symmetric perturbations of h are characterized by the one-parameter σ in (2.15).

In Case Study 3 it will be necessary to have more complete information about this parameter σ as given in the following proposition. Let $G(x, y, \lambda, \alpha)$

be a one-parameter unfolding of a germ g which is $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-equivalent to (2.10). Then G may be factored through the universal unfolding (2.15); in symbols,

$$G(\cdot, \cdot, \cdot, \alpha) \sim H(\cdot, \cdot, \cdot, \tilde{m}(\alpha), \tilde{n}(\alpha), \sigma(\alpha)). \tag{2.16}$$

In the Proposition we specify the sign of $d\sigma/d\alpha(0)$. (*Remark:* The magnitude of $d\sigma/d\alpha(0)$ is not an invariant of equivalence.)

Proposition 2.5. *If* $G(x, y, \lambda, \alpha) = (p(x^2, y^2, \lambda, \alpha)x, q(x^2, y^2, \lambda, \alpha)y)$, *then in* (2.16).

$$\operatorname{sgn} \frac{d\sigma}{d\alpha}(0) = \operatorname{sgn}\left\{\frac{p_\alpha(0)}{p_\lambda(0)} - \frac{q_\alpha(0)}{q_\lambda(0)}\right\}. \tag{2.17}$$

We shall use (2.17) in Case Study 3 to relate the sign of σ to that of α. We leave the proof of (2.17) for the reader. (See Schaeffer and Golubitsky [1979], §9.)

EXERCISE

2.1. Let $g(x, y, \lambda)$ be a bifurcation problem with $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-symmetry, and let $h(x, \lambda)$ be $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-equivalent to g.
(a) Show that h has $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-symmetry.
(b) Show that if g is nondegenerate, so is h.

§3. Linearized Stability and $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-Symmetry

Let $g: \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ be a bifurcation problem commuting with the group $\mathbf{Z}_2 \oplus \mathbf{Z}_2$. We shall call a solution (x, y, λ) of the equation $g(x, y, \lambda) = 0$ *linearly stable* if both the eigenvalues of dg at (x, y, λ) have a positive real part; *unstable* if at least one of them has a negative real part. In this section we show that in most cases linear stability or instability is an invariant of $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-equivalence. This is in marked contrast to the situation when there is no symmetry—as we saw in Chapter IX, §1, without symmetry linear stability is definitely *not* an invariant of equivalence. Of course, the reader may protest that in the present chapter we have imposed several restrictions in our definition of equivalence that were not a part of our definition in Chapter IX. The point is that when symmetry is present, often there is a natural set of conditions that may be imposed on equivalences which make linear stability an invariant of equivalence; while when symmetry is absent, there is no such set of conditions. (*Remark:* The relation between symmetry and stability is an exciting topic for current research.)

We begin the analysis with a preliminary discussion of how to exploit symmetry in computing the eigenvalues of dg. Since g commutes with

$(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-symmetry, g has the form (1.7) which we recall here:

$$g(x, y, \lambda) = (p(u, v, \lambda)x, q(u, v, \lambda)y), \tag{3.1}$$

where $u = x^2$, $v = y^2$, and $p(0, 0, 0) = q(0, 0, 0) = 0$. The Jacobian matrix is then

$$dg = \begin{pmatrix} p + 2up_u & 2p_v xy \\ 2q_u xy & q + 2vq_v \end{pmatrix}. \tag{3.2}$$

We ask the reader to verify the following consequences of (3.2).

Lemma 3.1. *Let* (x, y, λ) *be a solution to* $g = 0$.

(a) *If* (x, y, λ) *is a trivial or a pure mode solution* (*i.e.*, $x = 0$ *or* $y = 0$) *then* dg *is diagonal and its eigenvalues* (*which are real*) *have the signs listed as follows, for the three cases*:

(i) *Trivial solution*: sgn $p(0, 0, \lambda)$, sgn $q(0, 0, \lambda)$.
(ii) *x-mode solution*: sgn $p_u(x, 0, \lambda)$, sgn $q(x, 0, \lambda)$.
(iii) *y-mode solution*: sgn $p(0, y, \lambda)$, sgn $q_v(0, y, \lambda)$.

(b) *If* (x, y, λ) *is a mixed mode solution* (*i.e.*, $x \neq 0$ *and* $y \neq 0$) *then*

$$\begin{aligned} &\text{(a)} \quad \operatorname{sgn}\det(dg) = \operatorname{sgn}(p_u q_v - p_v q_u) \quad \textit{at} \quad (x, y, \lambda), \\ &\text{(b)} \quad \operatorname{sgn}\operatorname{tr}(dg) = \operatorname{sgn}(up_u + vq_v) \quad \textit{at} \quad (x, y, \lambda). \end{aligned} \tag{3.3}$$

Remarks. (i) Here we have shown by direct computation that at a pure mode solution to $g = 0$ the matrix dg is diagonal. We could have shown this abstractly by using the chain rule and the fact that at a pure mode solution there is a nontrivial element in the isotropy subgroup.

(ii) Recall that for a 2×2 matrix the signs of the real parts of the eigenvalues may be determined from the signs of the trace and determinant. In particular, if sgn det $dg < 0$, then the solution must be unstable, as the eigenvalues are real and of opposite signs. If sgn det $dg > 0$ then the solution is stable if sgn tr $dg > 0$ and unstable if sgn tr $dg < 0$. However, even with the formula (3.3), the computation of the signs of the eigenvalues of dg often requires some care.

We now turn to the main question: Is linearized stability an invariant of equivalence? We begin by defining what this means. Let g be a bifurcation problem in two state variables commuting with the group $\mathbf{Z}_2 \oplus \mathbf{Z}_2$, and let (z, λ) be a solution of the equation $g = 0$, where z is a shorthand for (x, y). We say that the stability of (z, λ) is *invariant under equivalence* if for every h that is $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-equivalent to g, the real parts of the eigenvalues of dh at the corresponding solution of $h = 0$ have the same signs as for dg. If the equivalence is given explicitly by

$$g = Sh(Z, \Lambda), \tag{3.4}$$

then the eigenvalues of dh are to be computed at $(Z(z, \lambda), \Lambda(\lambda))$.

The following proposition lists the cases in which stability of a solution to $g = 0$ is invariant under equivalences.

Proposition 3.2. *Let g be a bifurcation problem in two state variables commuting with the group $\mathbf{Z}_2 \oplus \mathbf{Z}_2$, and let (z, λ) be a solution to $g = 0$. Then the stability of (z, λ) is invariant under equivalence if any of the following hold:*

(a) *(z, λ) is a trivial or pure mode solution;*
(b) *(z, λ) is a mixed mode solution and* $\det(dg)_{z,\lambda} < 0$;
(c) *(z, λ) is a mixed mode solution,* $\det(dg)_{z,\lambda} > 0$ *and $p_u \cdot q_v > 0$ at the origin.*

Remarks. (i) In case (c), the proposition only applies to solutions which are close to the origin.

(ii) For the normal form (2.10), we have $p_u q_v = \varepsilon_1 \varepsilon_3$; in particular, if $\varepsilon_1 = \varepsilon_3$, then $p_u q_v > 0$. It follows from the proposition that if $\varepsilon_1 = \varepsilon_3$, then the stability of *any* solution of the equation is an invariant of equivalence.

We shall prove Proposition 3.2 using the following lemma.

Lemma 3.3. *Let g be a bifurcation problem in two state variables commuting with $\mathbf{Z}_2 \oplus \mathbf{Z}_2$, and let (z, λ) be a solution to $g = 0$. Then the stability of (z, λ) is invariant under equivalence if for every matrix function S satisfying* (2.3b) *and* (2.7), *the signs of the real parts of the eigenvalues of $S \cdot dg$ at (z, λ) are the same as those of dg at (z, λ).*

Remark. Lemma 3.3, appropriately restated, is valid for any number of state variables and any symmetry group.

We prove Lemma 3.3 below, after proving Proposition 3.2.

Proof of Proposition 3.2. We consider the three cases, in turn, using Lemma 3.3.

(a) From (2.5) and (2.7) we see that along a trivial or pure mode solution S is a diagonal matrix with positive entries on the diagonal. Since dg is also diagonal (cf. Lemma 3.1(a)) the signs of the eigenvalues of $S \cdot dg$ and dg are the same.

(b) Since $\det S > 0$, we see that

$$\operatorname{sgn}(\det dg) = sgn(\det(S \cdot dg)), \tag{3.5}$$

Both determinants are negative, and therefore both dg and $S \cdot dg$ have real eigenvalues of opposite signs.

(c) Equation (3.5) is still valid, in this case, only now both determinants are positive. Thus we must compare $\operatorname{tr}(S \cdot dg)$ with $\operatorname{tr} dg$, and in general, we cannot hope that the two traces have the same sign. However, under the hypothesis $p_u q_v > 0$, we can prove this near the origin.

According to (1.10), at a mixed mode solution we have $p = q = 0$; thus we may compute from (3.2) that

$$\operatorname{tr}(dg) = 2(up_u + vq_v). \tag{3.6}$$

(Cf. (3.3b).) Similarly, if S has the form (2.5), then

$$\operatorname{tr}(S \cdot dg) = 2\{u(c_1 p_u + c_2 vq_u) + v(c_4 q_v + c_3 up_v)\}, \tag{3.7}$$

where $c_i = c_i(u, v, \lambda)$. Now the term $c_2 vq_u$ in (3.7) vanishes at the origin and c_1 is positive, so near the origin

$$\operatorname{sgn}(c_1 p_u + c_2 vq_u) = \operatorname{sgn}(c_1 p_u) = \operatorname{sgn}(p_u);$$

similarly $\operatorname{sgn}(c_4 q_v + c_3 up_v) = \operatorname{sgn} q_v$. Thus corresponding terms in (3.6) and (3.7) have the same sign. Moreover, if $p_u q_v > 0$, the terms in (3.6) and (3.7) add, rather than cancel; thus if $p_u q_v > 0$,

$$\operatorname{sgn} \operatorname{tr}(dg) = \operatorname{sgn} \operatorname{tr}(S \cdot dg)$$

near the origin. □

We end this section with:

Proof of Lemma 3.3. The general $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-equivalence may be obtained by composing the three elementary forms of equivalence:

$$\begin{aligned} &\text{(a)} \quad g(z, \lambda) = h(z, \Lambda(\lambda)), \\ &\text{(b)} \quad g(z, \lambda) = h(Z(z, \lambda), \lambda), \\ &\text{(c)} \quad g(z, \lambda) = S(z, \lambda)h(z, \lambda). \end{aligned} \tag{3.8}$$

Trivially an equivalence of type (3.8a) does not change the stability of a solution—it does not even change the Jacobian dg. Moreover, in formulating the lemma we have hypothesized that equivalences of type (3.8c) leave the signs of the real parts of the eigenvalues of dg unchanged. Thus we need only consider type (3.8b).

We rewrite (3.8b) as

$$g(z, \lambda) = (dZ)_{z,\lambda} k(z, \lambda), \tag{3.9}$$

where

$$k(z, \lambda) = (dZ)_{z,\lambda}^{-1} h(Z(z, \lambda), \lambda). \tag{3.10}$$

We claim that $(dZ)_{z,\lambda}$ satisfies the same symmetry constraints (2.3b) and (2.7) as $S(z, \lambda)$; to see this, differentiate the identity (2.3a) using the chain rule. Thus we conclude from (3.9) and the invariance of stability under equivalences of the form (3.8c) that the signs of the real parts of the eigenvalues of dg are the same as those of dk. We finish the proof by showing that the eigenvalues of $(dk)_{z,\lambda}$ are the same as those of $(dh)_{Z(z,\lambda),\lambda}$. Indeed, differentiating (3.10) and using the fact that $h(Z(z_0, \lambda_0), \lambda_0) = 0$, we obtain

$$(dk)_{z_0,\lambda_0} = (dZ)^{-1}_{z_0,\lambda_0}(dh)_{Z(z_0,\lambda_0),\lambda_0}(dZ)_{z_0,\lambda_0};$$

thus dk and dh are similar matrices and have the same eigenvalues. □

EXERCISE

3.1. Prove Lemma 3.1.

§4. The Bifurcation Diagrams for Nondegenerate Problems

Our goal in this section is to draw bifurcation diagrams, both unperturbed and perturbed, for the normal form (2.10). For easy reference we recall the normal form here:

$$h(x, y, \lambda) = (p(x^2, y^2, \lambda)x, q(x^2, y^2, \lambda)y), \tag{4.1}$$

where

$$\begin{aligned} &\text{(a)} \quad p(x^2, y^2, \lambda) = \varepsilon_1 x^2 + my^2 + \varepsilon_2 \lambda, \\ &\text{(b)} \quad q(x^2, y^2, \lambda) = nx^2 + \varepsilon_3 y^2 + \varepsilon_4 \lambda. \end{aligned} \tag{4.2}$$

The modal parameters m and n satisfy the nondegeneracy condition

$$mn \neq \varepsilon_1\varepsilon_3, \qquad m \neq \varepsilon_2\varepsilon_3\varepsilon_4, \quad n \neq \varepsilon_1\varepsilon_2\varepsilon_4. \tag{4.3}$$

Unfortunately, there are sixteen possible choices of sign $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ in (4.2). To reduce the effort to manageable proportions, we draw the diagrams only for the two cases

$$\text{(A)} \quad \varepsilon_1 = \varepsilon_3 = 1, \qquad \varepsilon_2 = \varepsilon_4 = -1,$$

and

$$\text{(B)} \quad \varepsilon_1 = 1, \qquad \varepsilon_2 = \varepsilon_3 = \varepsilon_4 = -1.$$

The reasons for choosing these two are as follows.

Generally, in applications the trivial solution is stable subcritically; that is, for $\lambda < 0$. It follows from Lemma 3.1(a)(i) that the trivial solution is stable subcritically precisely when

$$\varepsilon_2 = \varepsilon_4 = -1. \tag{4.4}$$

Note that (4.4) holds for both cases (A) and (B) above.

Frequently, in applications the pure mode solutions bifurcate supercritically. This happens when

$$\varepsilon_1 = \varepsilon_3 = +1. \tag{4.5a}$$

Of course (4.4) and (4.5a) define case (A). In particular, the bifurcation problem that we study in Case Study 3 satisfies (4.4) and (4.5a).

For more theoretical reasons we also consider the case

$$\varepsilon_1 = +1, \qquad \varepsilon_3 = -1. \tag{4.5b}$$

We study (4.5b) to illustrate two points. First we will show that for certain perturbations in this case, a Hopf bifurcation must occur along a mixed mode branch. Second, we will clarify why invariance of stability does not hold, in general, along mixed mode branches.

We shall present the explanatory calculations for case (A) in some detail. The calculations for case (B) are similar, and we leave much for the reader to verify.

Case (A). $\varepsilon_1 = \varepsilon_3 = +1, \varepsilon_2 = \varepsilon_4 = -1.$

The universal unfolding of h is

$$H(x, y, \lambda) = (x^3 + mxy^2 - \lambda x, nx^2y + y^3 - (\lambda - \sigma)y). \tag{4.6}$$

The nondegeneracy conditions (4.3) are, in this case,

$$mn \neq 1, \qquad m \neq 1, \qquad n \neq 1. \tag{4.7}$$

These nondegeneracy conditions divide the modal parameter plane into seven regions as indicated by the solid lines in Figure 4.1.

The first fact to note towards our goal of drawing bifurcation diagrams is that the unperturbed (i.e., $\sigma = 0$) bifurcation diagrams for (4.6) are topologically equivalent for all values of m and n within each region. In addition, the universal unfolding H of (4.6) is topologically trivial in each region except in region 4. Region 4 divides into four sub-regions, indicated by the dashed lines, in which the universal unfoldings are topologically trivial. We shall draw the diagrams for regions 1, 2, 3, 4a,b,c, and 5.

Remarks. (i) Interchanging x and y has the effect of interchanging m and n and reversing the sign of σ in (4.6). (Cf. (2.17).) Thus diagrams in regions 2′, 3′ and 4b′ can be obtained rather easily from those in regions 2, 3 and 4b, respectively. Note that the interchanging of x and y is an equivalence, but it is not a $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-equivalence since it does not commute with $\mathbf{Z}_2 \oplus \mathbf{Z}_2$.

(ii) We will not deal with the distinguished values (4.7) of the modal parameters where h has infinite $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-codimension. (As in Chapters V and VI, appropriate higher-order terms reduce the codimension to three.) Likewise, we will not deal with infinite values for the modal parameter. (As

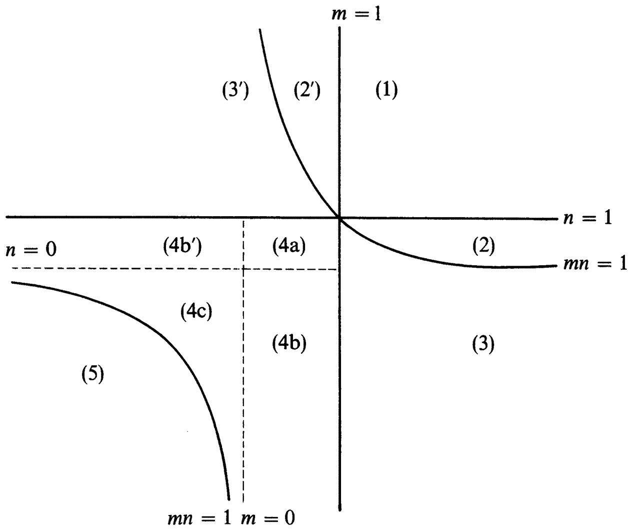


Figure 4.1. Degeneracies in the modal plane for (4, 6).

in Chapters V and VI, our scaling makes this case appear more singular than it really is.)

When drawing our bifurcation diagrams we take advantage of the ($\mathbf{Z}_2 \oplus \mathbf{Z}_2$)-symmetry to simplify their presentation. Specifically, we saw in §1 that pure mode solutions come in pairs and mixed mode solutions come four at a time. Drawing this multiplicity of solutions tends to be more confusing than helpful. Even though the bifurcation diagrams in $\mathbb{R}^3$ are quite pretty we refrain from attempting to draw them. Rather we draw schematic bifurcation diagrams, in which the lines refer to orbits of solutions as described in (1.3). We have already used such a convention in the $\mathbf{Z}_2$-symmetric context in Chapter VI.

The unperturbed (i.e., $\sigma = 0$) bifurcation diagrams for the various regions are given in Figure 4.2, and the perturbed bifurcation diagrams ($\sigma < 0$ and $\sigma > 0$) are given in Figure 4.3. In these figures we have indicated the stability assignments along each branch by $++$, $+-$, $-+$, or $--$. For example, $++$ indicates that both eigenvalues of dg have positive real parts; such a solution branch is stable. Similarly, $+-$ or $-+$ indicates that the two eigenvalues of dg have opposite signs, and $--$ indicates that both eigenvalues have negative real parts; in all these latter cases the solution is unstable. We have also drawn the stable solutions in a heavy black line for easy reference. Observe that condition $p_u q_v > 0$ in (c) of Proposition 3.2 is satisfied for H in this case. Thus the stability assignments along *all* the branches are invariants of ($\mathbf{Z}_2 \oplus \mathbf{Z}_2$)-equivalence.

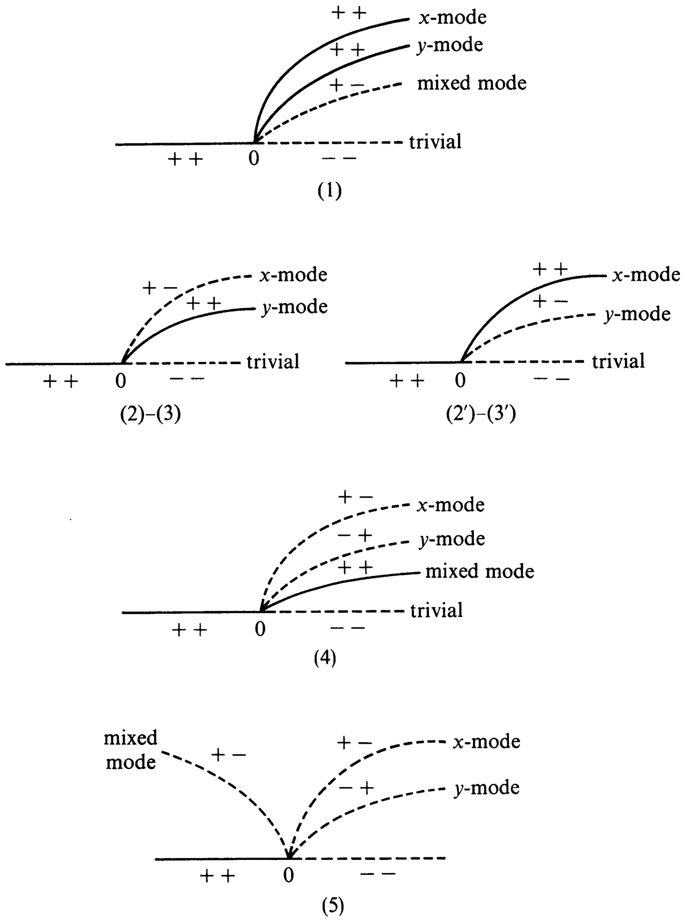


Figure 4.2. Unperturbed schematic bifurcation diagrams. Numbers in parentheses refer to Figure 4.1

We now discuss briefly the information needed to obtain these drawings. Using (1.10) we solve $H = 0$ explicitly as in (4.8).

$$
\begin{aligned}
&\text{(a)} \quad x = y = 0, && \text{trivial solutions,} \\
&\text{(b)} \quad \lambda = x^2; \quad y = 0, && x\text{-mode solutions,} \\
&\text{(c)} \quad x = 0; \quad \lambda - \sigma = y^2, && y\text{-mode solutions,} \\
&\text{(d)} \quad (1 - n)x^2 + (m - 1)y^2 = \sigma, && \text{mixed mode solutions.} \\
&\qquad \lambda = x^2 + my^2.
\end{aligned}
\tag{4.8}
$$

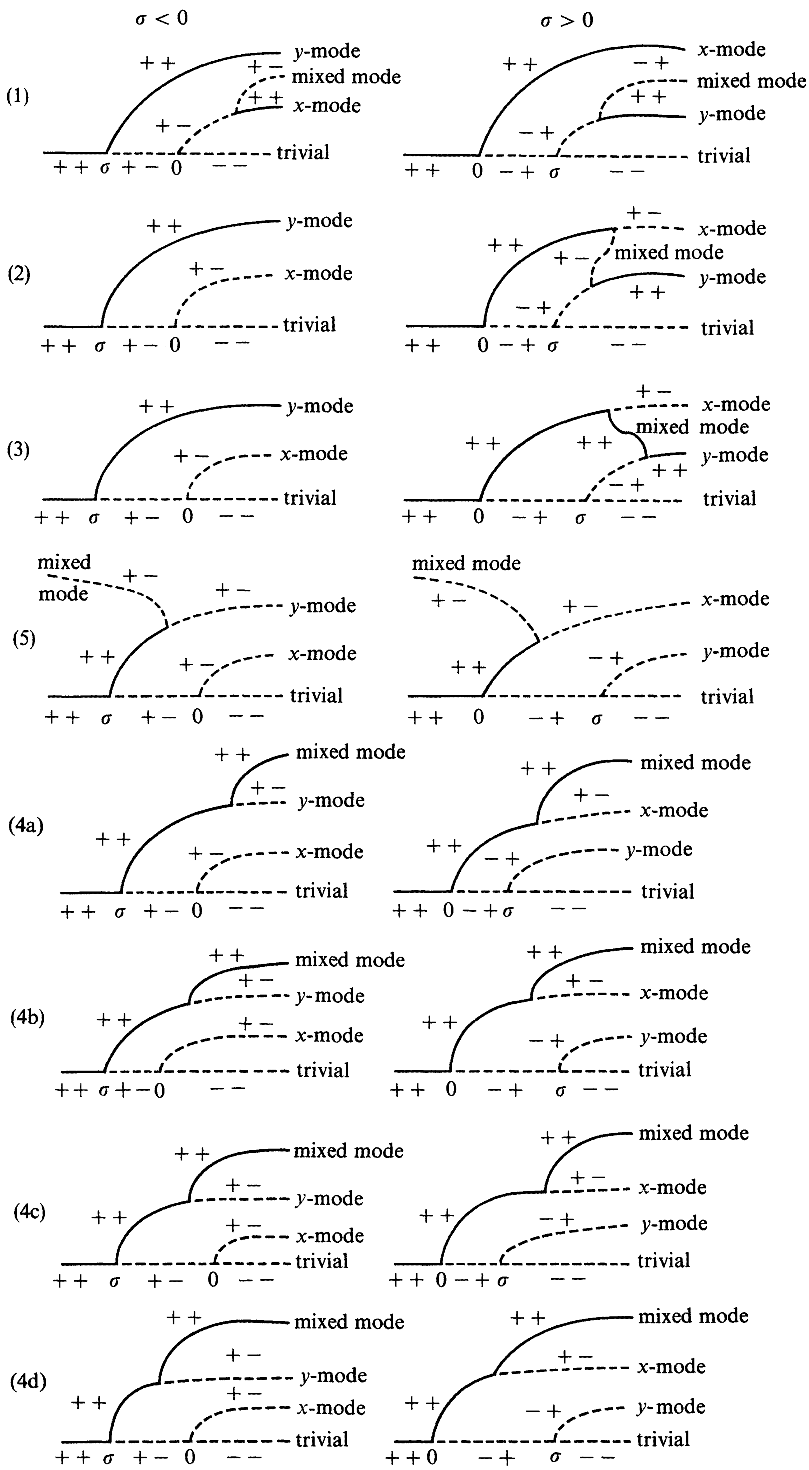


Figure 4.3. Persistent perturbations of (4.6). Numbers refer to regions in Fig. 4.1.

Next we use Lemma 3.1 to compute the signs of the eigenvalues of dH along $H = 0$. This information is recorded in (4.9).

$$\begin{array}{lll} \text{(a)} & \operatorname{sgn}(-\lambda), \operatorname{sgn}(\sigma - \lambda), & \text{trivial solutions,} \\ \text{(b)} & +, \operatorname{sgn}((n-1)\lambda + \sigma), & x\text{-mode solutions,} \\ \text{(c)} & \operatorname{sgn}((m-1)\lambda - m\sigma), +, & y\text{-mode solutions,} \\ \text{(d)} & \operatorname{sgn}(\det dH) = \operatorname{sgn}(1 - mn), & \text{mixed mode solutions.} \\ & \operatorname{sgn} \operatorname{tr}(dH) = +. & \end{array} \tag{4.9}$$

From the information in (4.8) and (4.9) it is an easy task to draw the unperturbed bifurcation diagrams ($\sigma = 0$). We note only that mixed mode solutions occur in the unperturbed bifurcation diagrams precisely when

$$\operatorname{sgn}(m - 1) = \operatorname{sgn}(n - 1). \tag{4.10}$$

(Cf. (4.8d).)

The bifurcation diagrams are more complicated when $\sigma \neq 0$. Here secondary bifurcation occurs in a persistent way; this is a consequence of the $(Z_2 \oplus Z_2)$-symmetry. If we considered perturbations which break the symmetry, then the persistent perturbations would not have any bifurcation points in the bifurcation diagrams. (Cf. Theorem III,6.1 and Chapter IX, §3).

Let us discuss these diagrams. The most striking feature is the differences in the perturbed bifurcation diagrams for $\sigma > 0$ in the three regions 1, 2, and 3. The diagram for region 1 shows a primary bifurcation to a stable x-mode solution which remains stable for large λ and a subsequent bifurcation from the trivial solution to a y-mode solution which gains stability through a secondary bifurcation. In a controlled experiment where λ is the only parameter varied and λ is varied quasistatically, there is little likelihood that the y-mode solution will be observed. Contrasted with this, in the perturbations of regions 2 and 3, the x-mode solution looses stability at a secondary bifurcation and a transition to the y-mode takes place. In region 3 this transition takes place smoothly along a stable mixed mode solution branch; while in region 2 this transition occurs with a jump to the stable y-mode solution. These considerations will play a fundamental role in our discussion of mode jumping in a rectangular plate described in Case Study 3.

Let us record some facts which should help the reader in verifying the above statements. Secondary bifurcation of mixed mode solutions occur somewhere along an x-mode solution branch if $\operatorname{sgn}(\sigma) = \operatorname{sgn}(1 - n)$. These secondary bifurcations occur somewhere along a y-mode branch $\operatorname{sgn}(\sigma) = \operatorname{sgn}(m - 1)$.

These bifurcations occur at λ values λ_x and λ_y, respectively, where

$$\begin{aligned} &\text{(a)} \quad \lambda_x = \frac{\sigma}{1-n}, \\ &\text{(b)} \quad \lambda_y = \frac{m\sigma}{m-1}, \\ &\text{(c)} \quad \operatorname{sgn}(\lambda_x - \lambda_y) = \operatorname{sgn}\left[\sigma \frac{mn-1}{(m-1)(1-n)}\right]. \end{aligned} \tag{4.11}$$

We now discuss another item of interest concerning Figure 4.3. In region 4 when $\sigma > 0$, the second bifurcation along the trivial branch is to a y-mode solution branch and occurs at $\lambda = \sigma$. According to (4.11a), the secondary bifurcation to the mixed mode solution from the x-mode branch occurs at $\lambda_x = \sigma/(1-n)$. In this region $n < 1$, and

$$\begin{aligned} &\text{(a)} \quad \lambda_x > \sigma \quad \text{if} \quad n > 0, \\ &\text{(b)} \quad \lambda_x < \sigma \quad \text{if} \quad n < 0. \end{aligned} \tag{4.12}$$

Since the order in which bifurcations occur is an invariant under equivalence, topological triviality fails along $n = 0$. A similar situation occurs along $m = 0$ in region 4.

One might suspect that a similar division of region 5 occurs. However, in region 5 both m and n are negative; hence the secondary bifurcation always occurs at a λ value less than that of the second primary bifurcation.

Let us show that the stability assignments in Figure 4.3 are in fact largely *determined by the bifurcation diagram itself.* When $\sigma \neq 0$ bifurcations occur only along trivial and pure mode solutions, and at these bifurcation points, rank $dH = 1$; that is, only one eigenvalue of dH is zero. Near such points we could perform a Liapunov–Schmidt reduction to obtain a reduced bifurcation equation in one state variable which describes the branching. All of these bifurcations are pitchfork bifurcations; this is to be expected, since along pure mode solutions or the trivial solution there is at least one nontrivial reflectional symmetry in the isotropy group and the bifurcation breaks this symmetry. At a pitchfork bifurcation we have the principle of exchange of stability. However for $|\lambda| \gg 0$, the stability assignments in Figure 4.3 must be the same as in the unperturbed diagrams of Figure 4.2. Thus by working in "from infinity," using exchange of stability at bifurcation points, the stabilities may be verified.

Case (B). $\varepsilon_1 = +1, \varepsilon_2 = \varepsilon_3 = \varepsilon_4 = -1.$

The nondegeneracy conditions in this case are:

$$mn \neq -1, \qquad m \neq -1, \qquad n \neq 1. \tag{4.13}$$

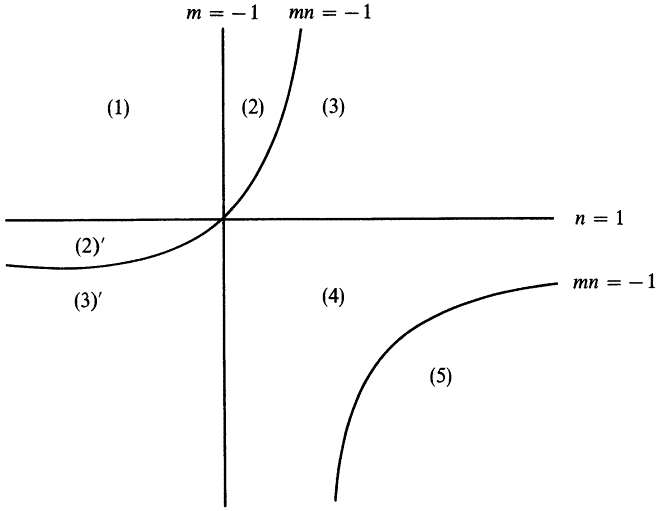


Figure 4.4. Degeneracies in the modal plane for (4, 2) $\varepsilon_2 = \varepsilon_4 = -1, \varepsilon, \varepsilon_3 = -1$.

The seven regions of modal parameter space are now located as shown in Figure 4.4. We shall consider only region 4 and restrict to positive σ. The perturbed bifurcation diagram for m and n in region 4 and $\sigma > 0$ is shown in Figure 4.5.

Let us focus on the mixed mode solution in this figure. As in case (A) above, the stability assignments in the figure can be checked using exchange of stability; in particular, the mixed mode solution must be stable for values of λ near λ_x and unstable for λ near λ_y. However, the eigenvalues of dH vary continuously along the mixed mode branch, and they are always nonzero. The only way that the stability assignment $++$ can change to $--$ is for the eigenvalues to cross the imaginary axis somewhere away from 0. These considerations suggest that a Hopf bifurcation occurs along the

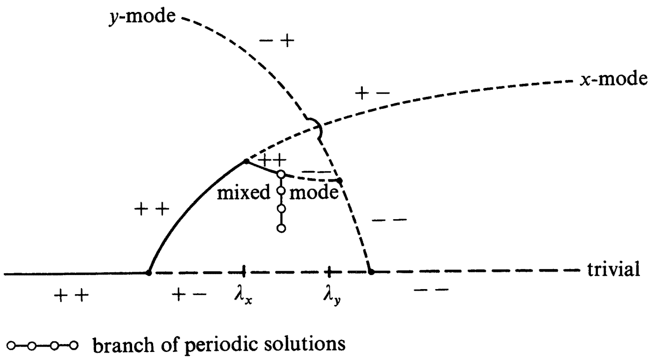


Figure 4.5 The existence of Hopf bifurcation.

mixed mode solution branch. Indeed this is basically true, but there are certain complications we must explore before making a definite statement.

In deriving the normal form H, we have performed a $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-equivalence to transform away all terms of higher order than three in x and y. However, equivalences of the form we use may change the dynamics of a system of ODE's. Let us elaborate. Consider a system of ODE

$$\dot{z} + g(z, \lambda) = 0. \tag{4.14}$$

Only equivalences of the form $g \to Sh(Z, \lambda)$ where

$$S = (dZ)^{-1}_{z,\lambda} \tag{4.15}$$

will preserve the dynamics of (4.14). Of course, the set of coordinate transformations which satisfy (4.15) is a proper subset of all equivalances. In particular, there exist mappings $g: \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ which are $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-equivalent to the normal form h but which contain fifth- (and higher-) order terms that cannot be transformed away by an equivalence satisfying (4.15). These fifth-order terms play an important role in the Hopf bifurcation discussed above.

The full analysis of Hopf bifurcation along the mixed mode solution branch in Figure 4.5 is beyond the scope of this volume, but we summarize the main issues. It turns out that there exists a nondegeneracy condition on the fifth-order term in g which guarantees that the eigenvalues of dg along the mixed mode branch cross the imaginary axis exactly once and with nonzero speed. If this condition is satisfied, then by Theorem VIII,3.1 a Hopf bifurcation occurs along the mixed mode branch. Moreover, there is a second nondegeneracy condition associated with the condition $\mu_2 \neq 0$ of Theorems VIII,4.1 and VIII,3.2. Depending on the sign of this term, the Hopf bifurcation may be supercritical and therefore stable.

We end our discussion by noting that the exact λ value where the Hopf bifurcation occurs is *not* an invariant of $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-equivalence, even for equivalences such that $\Lambda(\lambda) = \lambda$. This shows why the stability of mixed mode solutions is, in general, not an invariant of $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-equivalence.

Bibliographical Comments

The material in this chapter is taken from §4 of Golubitsky and Schaeffer [1979b]. However, let us mention that we were led to this presentation through trying to understand the important paper of Bauer *et al.* [1975] from the singularity theory point of view. A Hopf bifurcation similar to that of Figure 4.5 was studied by Iooss and Langford [1980].

CASE STUDY 3

Mode Jumping in the Buckling of a Rectangular Plate

(a) Synopsis

The buckling of a long rectangular plate involves rather more complex behavior than the buckling of a rod, which we studied in Chapter VII, §2. Consider the situation sketched in Figure C3.1 of a plate subject to a compressive load λ, uniformly distributed over the end faces. When λ is sufficiently large the plate begins to buckle into a pattern involving several small, localized buckles, rather than a single large arch as for a rod. (The reason for this different behavior is that we assume the plate is supported along all four sides, and this makes the displacement vanish at the boundary.) More precisely, let us scale the coordinates so that the undeformed plate is parametrized by

$$\Omega = \{(\xi, \eta) : 0 < \xi < l\pi, 0 < \eta < \pi\},$$

where l is the aspect ratio of the plate; we assume $l > 1$. We will show below that just after buckling the lateral displacement function $w(\xi, \eta)$ is proportional to

$$\sin \frac{k\xi}{l} \sin \eta, \tag{C3.1}$$

where k is a positive integer. We shall call k the *wave number*. (To be precise, (C3.1) only holds for what are called simply supported boundary conditions; this will be discussed in detail below.) We will also show that the buckles try to be square, i.e., that the value of k in (C3.1) after the bifurcation is nearly equal to l.

As λ is increased beyond the buckling load, nonlinear effects begin to influence the form of the displacement (C3.1), but (C3.1) still continues to

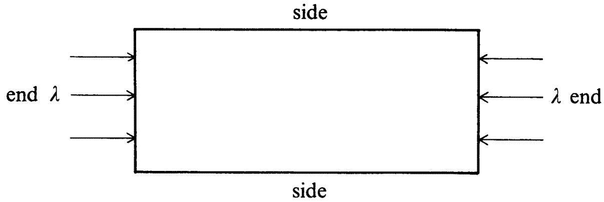


Figure C3.1. Plate subject to compressive load.

describe the displacement approximately—*up to a point*. In certain experiments, when the load is increased sufficiently far beyond the buckling load the plate jumps to a new configuration approximately described by (C3.1) but *with k increased by unity*. This phenomenon, called *mode jumping*, is the focus of this Case Study.

For the reader's convenience, let us give a few specifics concerning Stein's [1959] experiment. His plate had an aspect ratio $l = 5.38$. The initial buckle pattern had five buckles (i.e., was described by (C3.1) with $k = 5$) and persisted up to approximately 1.7 times the buckling load. At this point the plate jumped "suddenly and violently" to the mode with six buckles. Further increases in λ led to jumps to seven buckles, eight buckles, and finally to eventual collapse of the plate. We shall only discuss the first mode jump because, among other reasons, the plate was judged to enter the plastic regime between the jump from $k = 6$ to $k = 7$.

In this Case Study we analyze mode jumping as follows. First, we will show that for certain distinguished values of the aspect ratio l the mathematical idealization of this experiment leads to a problem involving bifurcation from a double eigenvalue with $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-symmetry as studied in Chapter X. Consider one such distinguished aspect ratio, say $l = l^*$; thus for $l = l^*$ the reduced equations are $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-equivalent to the normal form (X,2.10). This normal form will be the organizing center in our analysis of the problem. As our second step we will perform the Liapunov–Schmidt reduction of the problem when $l = l^*$, thereby determining the modal parameters m and n in the canonical form (X,2.10). Third, we will relate $l - l^*$ to the unfolding parameter σ in (X,4.6). With this information we may read off the bifurcation diagram of the physical problem when $l \approx l^*$ from the appropriate part of Figure X,4.3. In particular, this analysis predicts mode jumping if and only if region 2 with $\sigma > 0$ of Figure X,4.1 is selected by this process. Our conclusion will be that mode jumping occurs for the boundary conditions considered most realistic for the experiment (clamped) and does not occur for the commonly analyzed boundary conditions (simply supported).

It should be pointed out that we do not include imperfections in this analysis. We expect that imperfections will be unimportant, even though the full, symmetry breaking unfolding of (X,2.10) is quite complicated and a complete rigorous analysis seems remote. Let us elaborate. Consider Figure

X,4.3, region 2 with $\sigma > 0$; more precisely imagine increasing λ quasi-statically from zero in the bifurcation diagram of this figure. Mode jumping occurs when the x-mode becomes unstable; this happens through a pitchfork bifurcation as sketched in Figure C3.2(a). A typical perturbation of Figure C3.2(a) is shown in Figure C3.2(b). Both diagrams share the essential feature that the solution follows one smooth branch until that branch becomes unstable. The only qualitatively significant effect of the imperfection is to introduce a preferred direction for the jump.

There is a helpful intuitive explanation of how the double eigenvalue arises, related to the fact that the buckles try to be approximately square. The buckles given by (C3.1) have width π and length $\pi(l/k)$. Since l need not be an integer, the buckles cannot in general be exactly square. Roughly speaking, the first mode to become unstable is the one for which l/k is closest to unity. Thus as l increases, the preferred wave number k increases along with it. However, there are certain values of l between two adjacent integers, say k and $k + 1$, such that the modes (C3.1) with wave number k and $k + 1$ become unstable at exactly the same load; hence, these distinguished values of l lead to a double eigenvalue. (*Warning:* The above intuition is correct in spirit but not in detail. In particular, the double eigenvalue does not occur for l equal exactly to a half integer.)

Let us give a preliminary discussion of how the symmetry group $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ arises in this problem, as the situation here is typical of many applications. The mathematical description of the plate problem that we study below is a system of PDE which commutes with the following two reflections:

$$R_1 w = -w \tag{C3.2}$$

(i.e., reflection through the plane of the undeformed plate) and

$$R_2 w = w \circ \Pi, \tag{C3.3}$$

where

$$\Pi(\xi, \eta) = (l\pi - \xi, \eta)$$

(i.e., reflection about the minor axis of Ω). Note that R_1 and R_2 commute with one another and generate a four element group $\{I, R_1, R_2, R_1R_2\}$ that is abstractly isomorphic to $\mathbf{Z}_2 \oplus \mathbf{Z}_2$. To make the correspondence with

(a) symmetric (b) perturbed

Figure C3.2. Typical symmetry breaking of pitchfork bifurcations in (X,2.10).

Chapter X complete, let us consider the Liapunov–Schmidt reduction of the PDE for a value of l such that the first bifurcation is from a double eigenvalue, say with associated eigenvectors

$$w_1(\xi, \eta) = \sin\frac{k\xi}{l}\sin\eta \qquad w_2(\xi, \eta) = \sin\frac{(k+1)\xi}{l}\sin\eta.$$

This reduction leads to a reduced bifurcation equation

$$g\colon \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2;$$

here $\mathbb{R}^2 \cong \{xw_1 + yw_2\}$ parametrizes the kernel of the linearized operator. According to Chapter VII, §3(c), g also commutes with the reflections R_i. Now

$$R_1 w_i = -w_i, \qquad i = 1, 2,$$

and

$$R_2 w_1 = (-1)^{k+1} w_1, \qquad R_2 w_2 = (-1)^k w_2.$$

We obtain the representation (X,1.1) for these reflections by identifying

$$I \leftrightarrow (1, 1), \qquad R_1 \leftrightarrow (-1, -1),$$
$$R_2 \leftrightarrow ((-1)^{k+1}, (-1)^k), \qquad R_1 R_2 \leftrightarrow ((-1)^k, (-1)^{k+1}).$$

Thus the reduced problem g commutes with the action of $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ studied in Chapter X. (*Remark:* The full set of PDE also commutes with reflection about the *major* axis of Ω; i.e., $\tilde{\Pi}(\xi, \eta) = (\xi, \pi - \eta)$. However, this symmetry has no effect on the problem, since it leaves the eigenfunctions w_1 and w_2 invariant. (Cf. Remark VII,3.1 that a symmetry must be *broken* to be important.))

The remainder of the Case Study is divided into three parts as follows. In part (b) we give a mathematical formulation of the plate problem. This includes a discussion of simply supported and clamped boundary conditions. The bifurcation analysis of simply supported boundary conditions is given in part (c); clamped boundary conditions, in part (d).

(b) Formulation of a Specific Mathematical Model

The von Kármán equations for a plate involve two unknown functions, the lateral displacement $w(\xi,\eta)$ and the Airy stress function $\phi(\xi, \eta)$. The equations are posed in the domain Ω parametrizing the undeformed plate; they read as follows.

$$\Delta^2 w = [\phi, w] - \lambda w_{\xi\xi}, \tag{C3.4a}$$

$$\Delta^2 \phi = -\tfrac{1}{2}[w, w]. \tag{C3.4b}$$

Here λ is the external load, Δ^2 is the biharmonic operator in ξ and η, and the bracket operation is defined by

$$[u, v] = u_{\xi\xi}v_{\eta\eta} - 2u_{\xi\eta}v_{\xi\eta} + u_{\eta\eta}v_{\xi\xi}. \tag{C3.4c}$$

Let us say a few words about interpretation. Continuum mechanics models the effect of internal forces in a continuous medium by a stress tensor σ_{ij}, which is characterized by the following property. Suppose that one isolates (conceptually) a region $\mathcal{O} \subset \Omega$ of the plate; then the ith component of the force exerted on $\mathcal{O}$ by the material outside $\mathcal{O}$ is

$$F_i = \sum_{j=1}^{2} \int_{\partial\mathcal{O}} \sigma_{ij} N_j \, dS, \qquad i = 1, 2,$$

where N_j is the unit exterior normal. For equilibrium the net force on every subregion must vanish, which by the divergence theorem implies that

$$\sum_{j=1}^{2} \frac{\partial \sigma_{ij}}{\partial x_j} = 0; \qquad i = 1, 2.$$

Recall that for a divergence free vector field V in two dimensions (on a simply connected domain) there exists a stream function ψ such that $V = (\psi_\eta, -\psi_\xi)$. Using this fact twice and the fact that σ_{ij} is a symmetric tensor we may deduce that there is a function ϕ such that

$$\sigma_{11} = \phi_{\eta\eta}, \qquad \sigma_{12} = \sigma_{21} = -\phi_{\xi\eta}, \qquad \sigma_{22} = \phi_{\xi\xi};$$

this ϕ is the Airy stress function. Now (C3.4b) says that if certain second-order derivatives of w are nonzero, this creates stresses in the plate. Likewise, (C3.4a) describes the equilibrium displacement: $\Delta^2 w$, a combination of fourth-order derivatives, is nonzero only in response to stresses, either internal stresses transmitted through ϕ or the external stress λ.

Of course (C3.4) must be supplemented by boundary conditions. For ϕ we will require that

$$\frac{\partial \phi}{\partial N} = \frac{\partial}{\partial N} \Delta\phi = 0 \quad \text{on } \partial\Omega; \tag{C3.5}$$

where $\partial/\partial N$ indicates the normal derivative. These conditions are derived in Schaeffer and Golubitsky [1979]. Other boundary conditions on ϕ are derived in Holder and Schaeffer [1984].

For w, the most convenient boundary conditions from a mathematical point of view are what are called simply supported boundary conditions, namely

$$w = 0, \qquad \frac{\partial^2 w}{\partial N^2} = 0. \tag{C3.6}$$

In loose terms, (C3.6) describes a hinged boundary: the edge of the plate is free to rotate, although it cannot translate up or down. If $\partial^2 w/\partial N^2 \neq 0$ (i.e., if the end of the plate is curved), the plate will rotate to eliminate the torque caused by nonzero curvature. (*Remark:* Physically, it is virtually impossible to maintain a boundary condition like this on all four sides after the plate begins to buckle. Indeed, Stein had to exercise considerable ingenuity in devising an experiment where (C3.6) was appropriate on *two* of the four sides.) We claim that simply supported boundary conditions may be written

$$w = \Delta w = 0 \quad \text{on } \partial\Omega. \tag{C3.7}$$

The term Δw equals $\partial^2 w/\partial N^2 + \partial^2 w/\partial T^2$, and by the first boundary condition we have $\partial^2 w/\partial T^2 = 0$ along a straight side; thus (C3.7) is equivalent to (C3.6), as claimed.

The other canonical choice of boundary conditions for w is

$$w = 0, \qquad \frac{\partial w}{\partial N} = 0, \tag{C3.8}$$

normally called clamped boundary conditions. (C3.8) is easier to interpret—the plate cannot rotate or translate. Whatever stresses this sets up at the edge of the plate are absorbed by the boundary. Stein [1959] considered clamped boundary conditions the better approximation along the end faces in his experiment. We quote: "The plate was subject . . . to 'flat end' loading which results in almost complete clamping of the loaded edges."

Thus the most accurate mathematical description of Stein's experiment imposes clamped boundary conditions on the ends, simply supported on the sides. However, in the next subsection we analyze the problem with simply supported boundary conditions on all four sides; this is because we believe the fastest entrance into the subject is first to understand the theory in the simply supported case. After that anlysis we return to the case of clamped end faces.

Note that zero is an eigenvalue of Δ^2 on Ω with boundary conditions (C3.5), with any constant function as eigenfunction. Thus it is not immediately obvious that (C3.4b) is solvable at all. However, it follows from Lemma C3.1 below with $u = 1$, $v = w$ that for any smooth function w which vanishes at $\partial\Omega$

$$\int_\Omega [w, w]\, dA = 0.$$

Therefore (C3.4b) is in fact solvable and the solution is unique up to an additive constant which has no physical significance. We denote this solution by $\phi = -\frac{1}{2}\Delta^{-2}[w, w]$ and substitute it into (C3.4a) to obtain an integro-differential equation

$$\Delta^2 w + \lambda w_{\xi\xi} = C(w), \tag{C3.9}$$

where the cubic term $C(w)$ is given by

$$C(w) = -\tfrac{1}{2}[\Delta^{-2}[w, w], w]. \tag{C3.10}$$

Lemma C3.1. *If u, v, w are smooth functions on Ω and if v and w vanish on $\partial\Omega$ then*

$$\int_\Omega u[v, w]\,dA = \int_\Omega [u, v]w\,dA.$$

PROOF. Ignoring the possible boundary terms in an integration by parts we find that

$$\begin{aligned}\int_\Omega u[v, w]\,dA &= \int_\Omega \frac{\partial^2}{\partial\eta^2}(uv_{\xi\xi}) - 2\frac{\partial^2}{\partial\xi\,\partial\eta}(uv_{\xi\eta}) + \frac{\partial^2}{\partial\xi^2}(uv_{\eta\eta})w\,dA \\ &= \int_\Omega [u, v]w\,dA,\end{aligned}$$

the latter equality because terms cancel. It remains to show the boundary terms vanish. In the first integration by parts the boundary term is $\int_{\partial\omega} u\langle V, N\rangle\,dS$, where the two component vector V is given by

$$V = (v_{\eta\eta}w_\xi - v_{\xi\eta}w_\eta, -v_{\xi\eta}w_\xi + v_{\xi\xi}w_\eta).$$

On a portion of the boundary where $N = (0, \pm 1)$, both terms in the second component of V vanish, since w_ξ and $v_{\xi\xi}$ are tangential derivatives and both v and w vanish along $\partial\Omega$; thus $\langle V, N\rangle = 0$. Similarly where $N = (\pm 1, 0)$. In the second integration by parts the boundary integrand contains a factor of w and therefore vanishes. □

(c) Simply Supported Boundary Conditions

In this part of the Case Study we perform a bifurcation analysis on (C3.9) subject to boundary conditions (C3.7). The subsection divides naturally into two parts:

(i) An analysis of the linearized problem, including a derivation of the aspect ratios for which a double eigenvalue occurs.
(ii) Calculation of the Liapunov–Schmidt reduction. The subsection leads to the prediction that mode jumping does not occur with boundary conditions (C3.7).

Observe that $w = 0$ is a solution of (C3.9) for any value of λ. Since $C(w)$ is homogeneous of degree three, the linearization of (C3.9) at $w = 0$ is

$$\Delta^2 w + \lambda w_{\xi\xi} = 0. \tag{C3.11}$$

Now for any integers k and m

$$\left\{\sin\frac{k\xi}{l}\sin m\eta\right\} \tag{C3.12}$$

is an eigenfunction of (C3.11) which satisfies the boundary conditions (C3.7). Moreover, the functions (C3.12) are precisely the eigenfunctions of the Laplace operator on Ω with Dirichlet boundary conditions, and therefore these functions form a complete orthogonal set in $L^2(\Omega)$. Thus, to determine for what values of λ (C3.11) has a nonzero solution, it suffices to substitute (C3.12) into (C3.11) and solve for λ. This yields

$$\lambda = \left(\frac{k}{l} + m^2 \frac{l}{k}\right)^2. \tag{C3.13}$$

Equation (C3.13) gives the load at which the undeformed solution of (C3.9) becomes unstable with respect to the mode (C3.12). Which mode becomes unstable first? It is clear that (C3.13) is an increasing function of m. Therefore the first mode to become unstable has $m = 1$. Let

$$\lambda_k = \left(\frac{k}{l} + \frac{l}{k}\right)^2,$$

the right-hand side of (C3.13) when $m = 1$. Figure C3.3 shows a plot of λ_k versus k, regarding k as a continuous variable. Note that this function achieves its minimum at $k = l$. Of course, only integer values of k in Figure C3.3 represent possible bifurcation points. Thus, the first bifurcation from $w = 0$ in (C3.9) occurs at

$$\lambda = \min\{\lambda_k : k = 1, 2, \ldots\}. \tag{C3.14}$$

If the minimization in (C3.14) is unique, the generic situation, then the first eigenvalue will be simple. However, mode jumping involves a competition between two modes. To treat two separate modes with local methods we adjust the aspect ratio to get bifurcation from a double eigenvalue, thereby obtaining an organizing center for this problem. For

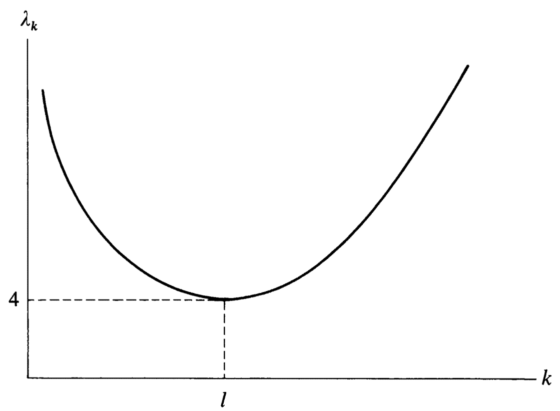


Figure C3.3. Graph of instability curve $\lambda_k = (k/l + l/k)^2$.

what values of l does the first bifurcation point involve two distinct modes, say k and $k + 1$? In other words, when is

$$\lambda_k = \lambda_{k+1}? \tag{C3.15}$$

Substituting (C3.13) into (C3.15) gives

$$l = \sqrt{k(k+1)} \tag{C3.16}$$

as the condition for the first eigenvalue to be double. Note that if l is given by (C3.16) then for large k we have

$$l = k + \tfrac{1}{2} + O(k^{-1}) \tag{C3.17}$$

and the first bifurcation point occurs at

$$\min_k \lambda_k = \sqrt{\frac{k}{k+1}} + \sqrt{\frac{k+1}{k}} = 2 + O(k^{-1}). \tag{C3.18}$$

We now turn to the Liapunov–Schmidt reduction of (C3.9), which we write abstractly as

$$\Phi(w, \lambda) = 0. \tag{C3.19}$$

Let l be given by (C3.16) so that the first bifurcation of (C3.18) from $w = 0$ is at a double eigenvalue, say $l = l^*$. Then the Liapunov–Schmidt reduction of (C3.19) near $w = 0$, $\lambda = \lambda^*$ leads to a pair of equations

$$g_i(x, y, \lambda) = 0, \qquad i = 1, 2. \tag{C3.20}$$

We claim that the first equation in (C3.20) is odd in x and even in y; the second, even in x and odd in y. (In short, we are claiming that g commutes with $\mathbf{Z}_2 \oplus \mathbf{Z}_2$. We proved this above, but we reprove it here, giving more complete references.) To prove the claim, first observe that (C3.9) is equivariant with respect to the two reflections R_1 and R_2 defined by (C3.2) and (C3.3); i.e.

$$\Phi(R_i w, \lambda) = R_i \Phi(w, \lambda). \tag{C3.21}$$

We showed in Chapter VII, §7(a) that symmetry in the full problem may be carried over to the reduced equation. To elaborate on this point, let us regard g in coordinate free way as a map

$$g\colon \ker L \times \mathbb{R} \to (\operatorname{range} L)^{\perp},$$

where $L = (d\Phi)_{0,\lambda^*}$. Now the kernel of L is spanned by the two functions

$$v_1(\xi, \eta) = \sin\left(\frac{k\xi}{l}\right) \sin \eta, \qquad v_2(\xi, \eta) = \sin\left(\frac{(k+1)}{l}\, \xi\right) \sin \eta. \tag{C3.22}$$

Observe that

$$\begin{aligned} &\text{(a)}\quad R_1 v_1 = -v_1, && R_1 v_2 = -v_2,\\ &\text{(b)}\quad R_2 v_1 = (-1)^{k+1} v_1, && R_2 v_2 = (-1)^k v_2. \end{aligned} \tag{C3.23}$$

Note that one of the signs in (C3.23b) is minus, one plus. Thus the four element group $\{I, R_1, R_2, R_1R_2\}$ is isomorphic to $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ and its action on the two-dimensional space ker L is isomorphic to (X,1.1). Moreover, the linearized operator $d\Phi$ is self-adjoint, so that (range $L)^\perp$ = ker L, and the group action on the range of g is identical. It was shown in Chapter X, §1(b) that under these conditions the reduced equations have the form (X,1.7a), which is precisely the claim above.

Chapter X, §2 gives the basic singularity theory results for bifurcation problems with $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-symmetry. In particular, Proposition X,2.3 gives the normal form for the reduced equations (C3.20) and Theorem X,2.4 gives the universal unfolding. Note that these theorems cover a variety of possible cases, only one of which occurs here, namely,

$$\varepsilon_1 = \varepsilon_3 = 1, \qquad \varepsilon_2 = \varepsilon_3 = -1. \tag{C3.24}$$

(See Chapter X, §4 for an intuitive discussion of why this case is most relevant for applications.) The normal form (X,4.6), which is obtained from Theorem X,2.4 when (C3.24) is assumed, provides a convenient formula for reference.

To apply the results of Chapter X we must compute various derivatives of g. We do not present all the details of the calculation here, even though this problem is one of those tractible problems where all the calculations can be done entirely by hand. Rather, our goal is to describe the calculation sufficiently well, including the tricks, so that the interested reader can carry out the calculations himself. We refer the reader to Schaeffer and Golubitsky [1979] for a more complete treatment.

Because of symmetry, the third-order derivatives of g with respect to x and y will be the first nonvanishing derivatives. In general, evaluation of third-order derivatives according to (VII,1.14c) requires inverting the linearized operator L. However, equation (C3.9) is odd in w (this is the symmetry (C3.2)), and this implies that $(d^2\Phi)_{0,\lambda} = 0$; therefore, this complication does not arise here. (Cf. Chapter VII, §2(d).)

Let us discuss the calculation of derivatives in a typical case. We have from (VII,1.14c)

$$\frac{\partial^3 g_1}{\partial x\, \partial y^2} = \langle v_1, d^3\Phi(v_1, v_2, v_2)\rangle, \tag{C3.25}$$

where v_1, v_2 are the eigenfunctions (C3.21). Now from (C3.9) $d^3\Phi = -d^3C$; indeed, since C is homogeneous cubic, $d^k\Phi = 0$ for $k \geq 4$, but this is no help. From formula (A3.2) we have

$$d^3C(v_1, v_2, v_3) = [\Delta^{-2}[v_1, v_2], v_3] + [\Delta^{-2}[v_2, v_3], v_1] + [\Delta^{-2}[v_3, v_1], v_2]. \tag{C3.26}$$

We substitute this into (C3.25) and find

$$\frac{\partial^3 g_1}{\partial x\, \partial y^2} = 2\langle v_1, [\Delta^{-2}[v_1, v_2], v_2]\rangle + \langle v_1, [\Delta^{-2}[v_2, v_2], v_1]\rangle.$$

By Lemma C3.1 we may integrate by parts in both terms to obtain

$$\frac{\partial^3 g_1}{\partial x\, \partial y^2} = 2\langle [v_1, v_2], \Delta^{-2}[v_1, v_2]\rangle + \langle [v_1, v_1], \Delta^{-2}[v_2, v_2]\rangle, \quad \text{(C3.27)}$$

Now v_1 and v_2 are given by (C3.22), and these formulas may be substituted into (C3.4c) to get formulas for $[v_i, v_j]$. It turns out that $[v_i, v_j]$ is a finite linear combination (two terms if $i = j$, four terms if not) of products

$$\cos \frac{\kappa_1 \xi}{l} \cos \kappa_2 \xi, \quad \text{(C3.28)}$$

where κ_1 and κ_2 are nonnegative integers. Most fortunately, (C3.28) is an eigenfunction of Δ^2 with boundary conditions (C3.5). Thus $\Delta^{-2}[w_i, w_j]$ is again a finite linear combination of eigenfunctions (C3.28); i.e., a finite Fourier cosine series. By Fourier analysis one may evaluate the inner products in (C3.27) as a finite sum of products of corresponding Fourier coefficients.

In §7 of Schaeffer and Golubitsky [1979] all the third-order derivatives $\partial^3 g_i/\partial x_j\, \partial x_k\, \partial x_l$ and the mixed derivatives $\partial^2 g_i/\partial \lambda\, \partial x_j$ are evaluated in this way. The values are then scaled as in Chapter X, §2 to determine the modal parameters m and n in the normal form (X,2.10). The result is

$$m = \frac{(k+1)^2}{k^2 + (k+1)^2}\left\{6 + \frac{1}{[2(2k+1)^2 + 1]^2}\right\},$$

$$n = \frac{k^2}{k^2 + (k+1)^2}\left\{6 + \frac{1}{[2(2k+1)^2 + 1]^2}\right\}. \quad \text{(C3.29)}$$

The modal parameters (C3.29) certainly satisfy

$$m > 1, \qquad n > 1 \quad \text{(C3.30)}$$

for all values of k; indeed $m > 3$ and n tends to 3 as k tends to infinity.

The above calculations assumed l was given by (C3.16). For values of l near $\sqrt{k(k+1)}$ we may analyze the problem as a one-parameter unfolding of the ideal case where (C3.16) is satisfied. Such an unfolding may be factored through the universal unfolding (X,4.6). The bifurcation diagrams of the perturbed problem may be read off from Figure X,4.3. Because of (C3.30) the relevant diagrams are those of region (1), for which there is no mode jumping. Rather the first mode to bifurcate remains stable as λ increases.

It was not necessary to relate l to the unfolding parameter σ of (X,4.6) to carry out the preceding analysis. This can nonetheless be done, and in fact we shall have to do so in studying clamped boundary conditions below.

(d) Clamped Boundary Conditions

Finally, we turn to the boundary conditions which seem to provide the most accurate description of Stein's [1959] experiment. We now consider (C3.9) subject to the following boundary conditions.

$$\begin{aligned} &\text{(a)} \quad w = \frac{\partial w}{\partial N} = 0 \quad \text{on ends } (\xi = 0 \text{ or } l\pi), \\ &\text{(b)} \quad w = \Delta w = 0 \qquad \text{on sides } (\eta = 0 \text{ or } \pi). \end{aligned} \tag{C3.31}$$

(We retain the boundary conditions (C3.5) for ϕ; hence $C(w)$ is unchanged.) The calculations for this case have much in common with the simply supported case considered above, and we couch our discussion in terms of what is different.

In general, it is impossible to find explicitly the values of λ for which (C3.11) has a nonzero solution with these boundary conditions. However, we are interested in those special values of l for which the first eigenvalue is double, and remarkably, this is an easier problem than trying to find the general eigenfunction of (C3.11). We may still separate variables in (C3.11) by looking for solutions of the form

$$f(\xi) \sin m\eta. \tag{C3.32}$$

As above we take $m = 1$ since we are looking for the lowest eigenvalue. Substitution of (C3.32) with $m = 1$ into (C3.11) yields the two point boundary problem

$$\begin{aligned} &\text{(a)} \qquad f^{(iv)} + (\lambda - 2)f'' + f = 0, \\ &\text{(b)} \quad f(0) = f'(0) = f(l\pi) = f'(l\pi) = 0. \end{aligned} \tag{C3.33}$$

For any λ, (C3.33a) has four linearly independent solutions, which have exponential form $e^{p\xi}$ for some $p \in \mathbb{C}$. If $\lambda \leq 4$ the exponentials are real (i.e., nonoscillatory), and it is impossible to satisfy the boundary conditions. Thus we assume that $\lambda > 4$. For such λ the general solution of (C3.33a) has the form

$$C_1 \cos(a\xi) + C_2 \sin(a\xi) + C_3 \cos(b\xi) + C_4 \sin(b\xi), \tag{C3.34}$$

where

$$a = \sqrt{L - M}, \qquad b = \sqrt{L + M}, \qquad L = \frac{\lambda}{2} - 1, \qquad M = \sqrt{\frac{\lambda^2}{4} - \lambda}. \tag{C3.35}$$

The two boundary conditions at $\xi = 0$ eliminate two of the four constants in (C3.34). Thus f may be written as a linear combination of the two functions

$$\phi(\xi) = b \sin(a\xi) - a \sin(b\xi), \qquad \psi(\xi) = \cos(a\xi) - \cos(b\xi).$$

We want to choose l so that the eigenvalue is double, so *both* these must be eigenfunctions. Thus we must have

$$\phi(l\pi) = \psi(l\pi) = \psi'(l\pi) = 0. \tag{C3.36}$$

(*Remark:* Since $\phi' = ab\psi$, the fourth condition $\phi'(l\pi) = 0$ is redundant.) It follows from (C3.36) that

$$a = \frac{k}{l}, \qquad b = \frac{k + 2n}{l} \tag{C3.37}$$

for some positive integers. From (C3.35) we obtain

$$a^2 + b^2 = \lambda - 2, \qquad b^2 - a^2 = \sqrt{\lambda^2 - 4\lambda}.$$

Elimination of λ from these equations yields

$$ab = 1.$$

Thus (C3.37) implies that $l = \sqrt{k(k + 2n)}$. Since we are interested in the lowest eigenvalue we take $n = 1$:

$$l = \sqrt{k(k + 2)}. \tag{C3.38}$$

Note the close resemblence of (C3.38) with (C3.16), the condition for a double eigenvalue with simply supported boundary conditions.

Thus if l satisfies (C3.38), the first bifurcation is from a double eigenvalue. The associated eigenfunctions are

$$\begin{aligned} &\text{(a)}\quad v_1(\xi, \eta) = \left\{\frac{k+2}{k}\sin\frac{k\xi}{l} - \sin\frac{(k+2)\xi}{l}\right\}\sin\eta, \\ &\text{(b)}\quad v_2(\xi, \eta) = \left\{\cos\frac{k\xi}{l} - \cos\frac{(k+2)\xi}{l}\right\}\sin\eta. \end{aligned} \tag{C3.39}$$

The remainder of the Liapunov–Schmidt reduction is exactly parallel to that for the previous case. In particular, the action of the two reflection operators R_i on the eigenfunctions (C3.39) is identical to the action (C3.23). Thus the reduced equations here have the same $(\mathbf{Z}_2 \oplus \mathbf{Z}_2)$-symmetry.

The eigenfunctions (C3.39a) are more complex than the eigenfunctions (C3.22), and this has one sticky consequence for the Liapunov–Schmidt reduction. In the previous case $[v_i, v_j]$ was a finite linear combination of eigenfunctions of Δ^{-2}, and this greatly simplified the evaluation of (C3.27). In the present case $[v_i, v_j]$ is again such a finite linear combination of eigenfunctions if $i = j$, but not if $i \neq j$. Although this difficulty requires a slight change in approach, it is still possible to do all calculations by hand. See §8 of Schaeffer and Golubitsky [1979].

The modal parameters in the canonical form (X,2.10) of the reduced equations are calculated for all values of k in Schaeffer and Golubitsky [1979]. Here we consider only $k = 5$ for the following reason. In the experiment Stein [1959] observed mode jumping from a mode with five buckles to one with six. If we take $k = 5$, the two eigenfunctions v_1 and v_2 in (C3.39) describe buckled configurations with five and six buckles, respectively. This may not be immediately apparent, but it may be seen by graphing these functions. (There is further supporting evidence for this statement in that v_1 is even under the reflection (C3.3), v_2 odd.) When $k = 5$ the value of the aspect ratio given by (C3.38) is

$$l^* = 5.92, \tag{C3.40}$$

and the calculations of Schaeffer and Golubitsky [1979] show that the modal parameters in (X,2.10) have the values

$$m = 1.0875, \qquad n = 0.9715. \tag{C3.41}$$

Note that $mn = 1.0565 > 1$; i.e., these values of m and n lie in region 2 of Figure X,4.1. The associated bifurcation diagram may be read off from Figure X,4.2. In particular, the only bifurcating solutions are pure modes; one is stable, the other unstable. The values (C3.41) mean that the six buckle mode associated with v_2 is stable, the five buckle mode unstable.

The above analysis applies to the aspect ratio (C3.40), while in Stein's experiment

$$l = 5.38. \tag{C3.42}$$

We consider the one-parameter family of bifurcation problems obtained by letting l vary as an unfolding of the idealized case (C3.40). Such an unfolding can be factored through the universal unfolding (X,4.6), at least for l near l^*; we assume that (C3.42) is close enough to l^* for this factorization to be possible. We want to read off the bifurcation diagram of the perturbed problem (C3.42) from Figure X,4.3. In order to do this, it is essential to relate $l - l^*$ to the unfolding parameter σ in (X,4.6). This is done in a rigorous way in §9 of Schaeffer and Golubitsky [1979]; here we content ourselves with the following heuristic analysis.

The role of σ in (X,4.6) is to split the double eigenvalues of the ideal problem into two simple eigenvalues; one mode or the other is favored, according to the sign of σ. It may be seen from Figure X,4.3 that mode jumping occurs when the mode which is unstable in the ideal problem is the first to bifurcate in the perturbed problem. Now for the plate problem, recall that the buckles try to be roughly square. Indeed, the double eigenvalue for $l = l^*$ arises because neither the five buckle mode nor the six buckle mode fits very well into the alloted length but both fit equally poorly. If $l < l^*$ the five buckle mode will experience a better fit and will be the first to bifurcate. But the five buckle mode is unstable in the ideal problem. Therefore, this analysis predicts mode jumping for the aspect ratio (C3.42), in agreement with the experimental result.

It is possible, based on the above ideas, to make a quantative prediction of the load at which the mode jump occurs. However, this analysis seems to yield a rather poor estimate of this load. There are several possible explanations of this discrepancy. One is the fact that the modal parameters (C3.41) lie rather close to the boundary of region 2 of Figure X,4.1 and different qualitative behavior obtains for parameter values across the boundary. Another is the fact that there are other modes which bifurcate shortly after the two modes studied here, and there may be some complicated interactions. It would be worthwhile to pursue these ideas further, but we cannot do so here.

Bibliography

Adams, J. F. [1969]. *Lectures on Lie Groups*. Benjamin, New York.

Ahlfors, L. [1979]. *An Introduction to the Theory of Analytic Functions of One Complex Variable*. McGraw-Hill, New York.

Arnold, V. I. [1976]. Local normal forms of functions. *Invent. Math.*, **35**, 87–109.

Arnold, V. I. [1981]. *Singularity Theory*. London Mathematical Society Lecture Notes Series, 53, Cambridge University Press, Cambridge.

Arnold, V. I. [1983]. *Geometrical Methods in the Theory of Ordinary Differential Equations*. Grundlehren, 250. Springer-Verlag, New York.

Ashkenazi, M. and Othmer, H. [1978]. Spatial patterns in coupled oscillators. *J. Math. Biol.*, **5**, 305–350.

Auchmuty, J. F. G. and Nicolis, G. [1975]. Bifurcation analysis of non-linear reaction-diffusion equations, I. *Bull. Math. Biol.*, **37**, 323–365.

Balakotaiah, V. and Luss, D. [1981]. Analysis of the multiplicity patterns of a CSTR. *Chem. Eng. Commun.*, **13**, 111–132.

Balakotaiah, V. and Luss, D. [1982a]. Exact steady-state multiplicity criteria for two consecutive or parallel reactions in lumped-parameter systems. *Chem. Eng. Sci.*, **37**, No. 3, 433–445.

Balakotaiah, V. and Luss, D. [1982b]. Structure of the steady-state solutions of lumped-parameter chemically reacting systems. *Chem. Eng. Sci.*, **37**, No. 11, 1611–1623.

Balakotaiah, V. and Luss, D. [1983]. Multiplicity features of reacting systems. Dependence of the steady-states of a CSTR on the residence time. *Chem. Eng. Sci.*, **38**, No. 10, 1709–1721.

Balakotaiah, V. and Luss, D. [1984]. Global analysis of the multiplicity features of multi-reaction lumped-parameter systems. *Chem. Eng. Sci.*, **39**, No. 5, 865–881.

Berger, M. [1977]. *Nonlinearity and Functional Analysis*. Academic Press, New York.

Brocker, [1975]. *Differential Germs and Catastrophes*. (Transl. by L. Lander.) London Mathematical Society Lecture Notes Series, 17. Cambridge University Press, Cambridge.

Buchner, M., Marsden, J., and Schecter, S. [1983]. Applications of the blowing-up construction and algebraic geometry to bifurcation problems, *J. Diff. Eqns.*, **48**, 404–433.

Carr, J. [1981]. *Applications of Centre Manifold Theory*. Applied Mathematical Sciences, 35. Springer–Verlag, New York.

Chafee, N. [1968]. The bifurcation of one or more closed orbits from an equilibrium point of an autonomous differential system. *J. Diff. Eqns.*, **4**, 661–679.

Chafee, N. [1978]. Generalized Hopf bifurcation and perturbation in a full neighborhood of a given vector field. *Ind. Univ. Math. J.*, **27**, 173–194.

Chow, S. N., and Hale, J. [1982], *Methods of Bifurcation Theory*. Grundlehren, 251. Springer-Verlag, New York.

Coddington, E., and Levinson, N. [1955]. *Theory of Ordinary Differential Equations*. McGraw-Hill, New York.

Crandall, M. and Rabinowitz, P. [1971]. Bifurcation from simple eigenvalues. *J. Funct. Anal.*, **8**, 321–340.

Crandall, M. and Rabinowitz, P. [1973]. Bifurcation, perturbation of simple eigenvalues, and linearized stability. *Arch. Rational Mech. Anal.*, **52**, 161–180.

Crandall, M. and Rabinowitz, P. [1978]. The Hopf bifurcation theorem in infinite dimensions. *Arch. Rational Mech. Anal.*, **19**, 1083–1102.

Cronin, J. [1981]. *Mathematics of Cell Electrophysiology*. Marcel Dekker, New York.

Damon, J. [1980]. Private communication.

Dancer, E. N. [1980]. On the existence of bifurcating solutions in the presence of symmetries. *Proc. Royal Soc. Edin.* A, **85**, 321–336.

Dangelmayr, G. and Stewart, I. [1984]. Classification and unfoldings of sequential bifurcations. *SIAM J. Math. Anal.*, **15**, 423–445.

Flockerzi, D. [1979]. Existence of small periodic solutions of ordinary differential equations in $\mathbb{R}^2$. *Arch. Math.*, **33**, 263–278.

Gibson, C. G. [1979]. *Singular Points of Smooth Mappings*. Research Notes in Mathematics, 25. Pitman, London.

Golubitsky, M. [1978]. An introduction to catastrophe theory and its applications. *SIAM Rev.*, **20**, 352–387.

Golubitsky, M. and Guillemin, V. [1973]. *Stable Mappings and Their Singularities*. Graduate Texts in Mathematics, 14. Springer-Verlag, New York.

Golubitsky, M. and Keyfitz, B. L. [1980]. A qualitative study of the steady state solutions for a continuous flow stirred tank chemical reactor. *Siam J. Math. Anal.*, **11**, 316–339.

Golubitsky, M., Keyfitz, B. L., and Schaeffer, D. [1981]. A singularity theory analysis of a thermal-chainbranching model for the explosion peninsula. *Commun. Pure Appl. Math.*, **34**, 433–463.

Golubitsky, M. and Langford, W. F. [1981]. Classification and unfoldings of degenerate Hopf bifurcations. *J. Diff. Eqns.*, **41**, 375–415.

Golubitsky, M. and Schaeffer, D. [1979a]. A theory for imperfect bifurcation theory via singularity theory. *Commun. Pure Appl. Math.*, **32**, 21–98.

Golubitsky, M. and Schaeffer, D. [1979b]. Imperfect bifurcation in the presence of symmetry. *Commun. Math. Phys.*, **67**, 205–232.

Golubitsky, M. and Stewart, I. N. [1984]. Hopf bifurcation in the presence of symmetry. *Arch. Rational Mech. Anal.* To appear.

Gray, B. F. and Yang, C. [1965]. On the unification of the thermal and chain theories of the explosion limits. *J. Phys. Chem.*, **69**, 2747–2750.

Guckenheimer, J. and Holmes, P. [1983]. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Applied Mathematical Sciences, 42. Springer-Verlag, New York.

Hale, J. K., Magalhaes, L. T., and Oliva, W. M. [1984]. *An Introduction to Infinite Dimensional Dynamical Systems—Geometric Theory*. Applied Mathematical Sciences, 47. Springer-Verlag, New York.

Hassard, B. [1978]. Bifurcation of periodic solutions of the Hodgkin–Huxley model for the squid giant axon. *J. Theor. Biol.*, **71**, 401–420.

Hassard, B., Kazarinoff, N., and Wan, Y.-H. [1981]. *Theory and Applications of Hopf Bifurcation.* London Mathematical Society Lecture Notes Series, 41. Cambridge University Press, Cambridge.

Hassard, B. and Wan, Y.-H. [1978]. Bifurcation formulae derived from the center manifold theory. *J. Math. Anal. Appl.*, **63**, 297–312.

Henry, D. [1981]. *Geometric Aspects of Semilinear Parabolic Equations.* Lecture Notes in Mathematics, 840. Springer-Verlag, New York.

Hirsch, M. and Smale, S. [1974]. *Differential Equations, Dynamical Systems, and Linear Algebra.* Academic Press, New York.

Hodgkin, A. and Huxley, A. [1952]. A qualitative description of membrane current and its application to conduction and excitation in nerves. *J. Physiol. (London)*, **117**, 500–544.

Holder, E. J. and Schaeffer, D. [1984]. Boundary conditions and mode jumping in the von Karman equations. *SIAM J. Math. Anal.*, **15**, 446–458.

Iooss, G. and Joseph, D. D. [1981]. *Elementary Stability and Bifurcation Theory.* Springer-Verlag, New York.

Iooss, G. and Langford, W. [1980], Interaction of Hopf and pitchfork bifurcation, in *Bifurcation Problems and Their Numerical Solution* (H. Mittelmann and H. Weber, eds.). ISNM, 54, Birkhauser Verlag, Basel.

Kato, T. [1976]. *Perturbation Theory for Linear Operators.* Grundlehren, 132. Springer-Verlag, New York.

Keener, J. [1976]. Secondary bifurcation in non-linear diffusion reaction equations. *Stud. Appl. Math.*, **55**, 187–211.

Keyfitz, B. L. [1984]. Classification of one state variable singularities with distinguished parameter up to codimension seven. University of Houston. Preprint.

Kielhofer, H. [1976]. On the Lyapunov stability of stationary solutions of semilinear parabolic differential equations. *J. Diff. Eqns.*, **22**, 193–208.

Kielhofer, H. [1979]. Hopf bifurcation at multiple eigenvalues. *Arch. Rational Mech. Anal.*, **69**, 53–83.

Koiter, W. T. [1945]. Over de stabiliteit van het elastische evenwicht. Thesis, University of Delft.

Labouriau, I. [1983]. Applications of singularity theory to neurobiology. Thesis, University of Warwick. (*SIAM J. Math. Anal.* To appear.)

Marsden, J. and McCracken, M. [1976]. *The Hopf Bifurcation and Its Applications.* Applied Mathematical Sciences, 19. Springer-Verlag, New York.

Martinet, J. [1974]. Dèploiements versels des applications diffèrentiables et classification des applications stables, in *Singularités d'Applications Différentiables, Plans-sur-Bex 1975* (O. Burlet and F. Ronga, eds.). Lecture Notes in Mathematics, 535. Springer-Verlag, Berlin, pp. 1–44.

Martinet, J. [1982]. *Singularities of Smooth Functions and Maps.* (Transl. by C. Simon.) London Mathematical Society Lecture Notes Series, 58. Cambridge University Press, Cambridge.

Mather, J. N. [1968]. Stability of C^∞ mappings, III. Finitely determined map germs. *Publ. Math. I.H.E.S.*, **35**, 127–156.

Mather, J. N. [1969a]. Stability of C^∞ mappings, II. Infinitesimal stability implies stability. *Ann. Math.*, **89**, 254–291.

Mather, J. N. [1969b]. Right equivalence. Lecture Notes, University of Warwick.

Mather, J. N. [1971]. Stability of C^∞ mappings, VI. The nice dimensions, in *Proceedings of Liverpool Singularities Symposium*, Vol. I (C. T. C. Wall, ed.). Lecture Notes Mathematics, 192. Springer-Verlag, Berlin, pp. 207–253.

Matkowsky, B. J. and Reiss, E. L. [1977]. Singular perturbations of bifurcations. *SIAM J. Appl. Math.*, **33**, 230–255.

Michel, L. [1972]. Nonlinear group action. Smooth actions of compact Lie groups on manifolds, in *Statistical Mechanics and Field Theory* (R. N. Sen and C. Weil, eds.). Israel University Press, Jerusalem, pp. 133–150.

Milnor, J. [1968]. *Singular Points of Complex Hypersurfaces.* Annals of Mathematical Studies. Princeton University Press, Princeton, NJ.

Mumford, D. [1965]. *Geometric Invariant Theory.* Springer-Verlag, Berlin.

Noble, B. [1969]. *Applied Linear Algebra.* Prentice-Hall, Englewood Cliffs, NJ.

Percell, P. B. and Brown, P. N. [1984]. Finite determination of bifurcation problems. *SIAM J. Math. Anal.* To appear.

Poston, T. and Stewart, I. [1978]. *Catastrophe Theory and Its Applications.* Pitman, San Francisco.

Prigogine, I. and Lefever, R. [1974]. Stability and self-organization in open systems, in *Membranes, Dissapative Structures, and Evolution* (G. Nicolis and R. Lefever, eds.). Wiley, New York.

Reiss, E. [1969]. Column buckling—an elementary example of bifurcation, in *Bifurcation Theory and Nonlinear Eigenvale Problems* (J. Keller and S. Antman, eds.). W. A. Benjamin, New York, pp. 1–16.

Rinzel, J. [1978]. Integration and propagation of neuroelectric signals, in *Studies in Mathematical Biology*, Vol. 15 (S. Levin, ed.). Mathematical Association of America, Washington, DC, pp. 1–66.

Rinzel, J. and Miller, R. [1980]. Numerical calculation of stable and unstable periodic solutions to the Hodgkin–Huxley equations. *Math. Biosci.*, **49**, 27–59.

Roorda, J. [1965]. The stability of structures with small imperfections. *J. Eng. Mech. Div., Amer. Soc. Civ. Eng.*, **91**, 87.

Roorda, J. [1968]. On the bucking of symmetric structural systems with first- and second-order imperfections. *Intern. J. Struct.*, **4**, 1137–1148.

Ruelle, D. [1973]. Bifurcations in the presence of a symmetry group. *Arch. Rational Mech. Anal.*, **51**, 136–152.

Sattinger, D. H. [1973]. *Topics in Stability and Bifurcation Theory.* Lecture Notes in Mathematics, 309. Springer-Verlag, Berlin.

Sattinger, D. H. [1978]. Group representation theory, bifurcation theory and pattern formation. *J. Funct. Anal.*, **28**, 58–101.

Sattinger, D. H. [1979]. *Group Theoretic Methods in Bifurcation Theory.* Lecture Notes in Mathematics, 762. Springer-Verlag, Berlin.

Sattinger, D. H. [1983]. *Branching in the Presence of Symmetry.* CBMS-NSF Conference Notes, 40. SIAM, Philadelphia.

Schaeffer, D. and Golubitsky, M. [1979]. Boundary conditions and mode jumping in the bucking of a rectangular plate. *Commun. Math. Phys.*, **69**, 209–236.

Schaeffer, D. and Golubitsky, M. [1981]. Bifurcation analysis near a double eigenvalue of a model chemical reaction. *Arch. Rational Mech. Anal.*, **75**, 315–347.

Schechter, M. [1977]. *Modern Methods in Partial Differential Equations.* McGraw-Hill, New York.

Seeley, R. [1964]. Extension of C^∞ functions defined in a half space. *Proc. Amer. Math. Soc.*, **15**, 625–626.

Shearer, M. [1981]. Coincident bifurcation of equilibrium and periodic solutions of evolution equations. *J. Math. Anal. Appl.*, **84**, 113–132.

Stein, M. [1959]. The phenomenon of change of buckle patterns in elastic structures. NASA Technical Report R-39.

Stewart, I. N. [1981]. Bifurcation and hysteresis varieties for the thermal-chain-branching model with a negative modal parameter. *Math. Proc. Camb. Phil. Soc.*, **90**, 127–139.

Stewart, I. N. and Woodcock, A. E. R. [1984]. Bifurcation and hysteresis varieties for the thermal-chainbranching model with a positive modal parameter. *Math. Proc. Camb. Phil. Soc.* To appear.

Taylor, A. E. and Mann, W. R. [1983]. *Advanced Calculus.* Wiley, New York.

Thompson, J. M. T. and Hunt, G. W. [1973]. *General Theory of Elastic Stability.* Wiley, New York.

Thompson, J. M. T. and Hunt, G. W. [1977]. A bifurcation theory for the instabilities of optimization and design. *Synthese*, **36**, 315–351.

Turing, A. [1952]. The chemical basis of morphogenesis. *Phil. Trans. Roy. Soc.* B, **237**, 37–72.

Tyson, J. and Kauffman, S. [1975]. Control of mitosis by a continuous biochemical oscillation. *J. Math. Biol.*, **1**, 289.

Uppal, A., Ray, W. H., and Poore, A. [1974]. On the dynamic behavior of continuous stirred tank reactors. *Chem. Eng. Sci.*, **29**, 967–985.

Uppal, A., Ray, W. H., and Poore, A. [1976]. The classification of the dynamic behavior of continuous stirred tank reactors—influence of reactor residence time. *Chem. Eng. Sci.*, **31**, 205–214.

Vanderbauwhede, A. [1982]. *Local Bifurcation and Symmetry.* Research Notes in Mathematics, 75. Pitman, Boston.

Wall, C. T. C. [1983]. Classification of unimodal isolated singularities of complete intersections, in *Singularities* (P. Orlik, ed.). Proceedings Symposium of Pure Mathematics, 40, part 2. American Mathematical Society, Providence, RI, pp. 625–640.

Zariski, O. and Samuel, P. [1960]. *Commutative Algebra*, Vol. II. Van Nostrand, Princeton, NJ.

Zeeman, E. C. [1976]. Euler buckling, in *Structural Stability, the Theory of Catastropes, and Applications to the Sciences* (P. Hilton, ed.). Lecture Notes in Mathematics, 525. Springer-Verlag, New York, pp. 373–395.

Zeldovich, Y. V. and Zisin, U. A. [1941]. On the theory of thermal stress. Flow in an exothermic stirred reactor, II. Study of heat loss in a flow reactor. *J. Tech. Phys.*, **11**, No. 6, 501–508 (Russian).

Symbol Index

Index

Applied Mathematical Sciences

(continued from page ii)

(continued on next page)

Applied Mathematical Sciences

(continued from previous page)

118. *Godlewski/Raviart:* Numerical Approximation of Hyperbolic Systems of Conservation Laws.
119. *Wu:* Theory and Applications of Partial Functional Differential Equations.
120. *Kirsch:* An Introduction to the Mathematical Theory of Inverse Problems.
121. *Brokate/Sprekels:* Hysteresis and Phase Transitions.
122. *Gliklikh:* Global Analysis in Mathematical Physics: Geometric and Stochastic Methods.
123. *Le/Schmitt:* Global Bifurcation in Variational Inequalities: Applications to Obstacle and Unilateral Problems.
124. *Polak:* Optimization: Algorithms and Consistent Approximations.
125. *Arnold/Khesin:* Topological Methods in Hydrodynamics.
126. *Hoppensteadt/Izhikevich:* Weakly Connected Neural Networks.
127. *Isakov:* Inverse Problems for Partial Differential Equations.
128. *Li/Wiggins:* Invariant Manifolds and Fibrations for Perturbed Nonlinear Schrödinger Equations.
129. *Müller:* Analysis of Spherical Symmetries in Euclidean Spaces.
130. *Feintuch:* Robust Control Theory in Hilbert Space.
131. *Ericksen:* Introduction to the Thermodynamics of Solids, Revised ed.
132. *Ihlenburg:* Finite Element Analysis of Acoustic Scattering.
133. *Vorovich:* Nonlinear Theory of Shallow Shells.
134. *Vein/Dale:* Determinants and Their Applications in Mathematical Physics.
135. *Drew/Passman:* Theory of Multicomponent Fluids.
136. *Cioranescu/Saint Jean Paulin:* Homogenization of Reticulated Structures.
137. *Gurtin:* Configurational Forces as Basic Concepts of Continuum Physics.
138. *Haller:* Chaos Near Resonance.
139. *Sulem/Sulem:* The Nonlinear Schrödinger Equation: Self-Focusing and Wave Collapse.
140. *Cherkaev:* Variational Methods for Structural Optimization.
141. *Naber:* Topology, Geometry, and Gauge Fields: Interactions.
142. *Schmid/Henningson:* Stability and Transition in Shear Flows.