

Tian Ma · Shouhong Wang

Phase Transition Dynamics

Second Edition



Springer

Phase Transition Dynamics

Tian Ma • Shouhong Wang

Phase Transition Dynamics

Second Edition



Springer

Tian Ma
Department of Mathematics
Sichuan University
Sichuan, China

Shouhong Wang
Department of Mathematics
Indiana University
Bloomington, IN, USA

ISBN 978-3-030-29259-1 ISBN 978-3-030-29260-7 (eBook)
<https://doi.org/10.1007/978-3-030-29260-7>

Mathematics Subject Classification: 76xxx, 82xxx, 35Qxx, 37Lxx

1st edition: © Springer New York Heidelberg Dordrecht London 2014
© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

For Li and Ping

Preface to the Second Edition

Since the publication of the **first edition** of this book, progresses have been made on statistical physics, quantum physics, and topological phase transitions (TPTs). It is clear now that all phase transitions in nature that we have encountered can be categorized into the following two types:

- (a) Dynamical phase transitions
- (b) Topological phase transitions

The fundamental law governing a physical system is given in the form of mathematical equations, often partial differential equations (PDEs). The solutions of these equations determine the states of the system. A dynamic phase transition refers to transitions of the underlying physical system from one state to another, as the control parameter crosses certain critical threshold. The notion of dynamic phase transitions is applicable to all dissipative systems, including nonlinear dissipative systems in statistical physics, quantum physics, fluid dynamics, atmospheric and oceanic sciences, biological and chemical systems, etc.

TPTs are entirely different from dynamic phase transitions. Intuitively speaking, a TPT refers to the change of the topological structure in the physical space as certain system control parameter crosses a critical threshold. The PDE model of a physical system may undergo a dynamic phase transition at a certain critical threshold, and yet the topological phase transition may occur at the same or entirely different critical thresholds. In other words, dynamic transition studies transition and evolution of physical states in the infinite-dimensional phase space of the PDE model, the TPT studies topological structure change in the physical space, and yet they are highly intertwined.

The **first edition** of this book focuses mainly on the theory and applications of dynamic phase transitions. In this **second edition**, three new chapters, Chaps. 7–9, together with revised Sects. 1.1, 2.7, 3.1, and 3.2, are added to provide a comprehensive **first-principle approach** to the following new topics:

1. The nature and essence of physical laws (Sect. 1.1)
2. Dynamic theory of thermodynamical phase transitions (Sect. 3.1)
3. Dynamic theory of fluctuations and critical exponents (Sect. 3.2)

4. Fundamental principles of statistical physics (Chap. 7)
5. Quantum mechanism of condensates and high Tc superconductivity (Chap. 8)
6. Topological phase transitions (Chap. 9)

Hereafter, we present a brief description of the new sections and chapters.

Section 1.1

FIRST, we recall the two guiding principles of theoretical physics introduced in Ma and Wang (2015a, 2017a), inspired by the vision and work of Albert Einstein and Paul Dirac. The first guiding principle says that the entire theoretical physics is built upon a few fundamental first principles and the laws of nature are simple and beautiful with clear physical pictures. The second guiding principle states that for each physical system, the underlying symmetry of the system dictates uniquely the functional F also called Lagrangian action, which then determines the mathematical equations, representing fundamental law of the system. The functional F represents proper energy of the system. For example, the Einstein principle of general relativity (PGR), a symmetry principle, dictates the Einstein–Hilbert functional and consequently determines the fundamental law of gravity.

SECOND, the physical law is then derived from the energy functional F . For the four fundamental interactions, theoretical physics faces great challenges and mysteries, including (1) the dark matter and dark energy phenomena, (2) the nature of Higgs fields for the weak interactions, (3) the quark confinement for the strong interactions, (4) the asymptotic freedom, and (5) the unification of the four interactions. These challenges lead to our recent discovery of three fundamental principles: the principle of interaction dynamics (PID), the principle of representation invariance (PRI) (Ma and Wang, 2014b, 2015b, 2012), and the principle of symmetry-breaking (PSB) (Ma and Wang, 2014a).

Basically, PID takes the variation of the functional F under the energy-momentum conservation constraints. For gravity, PID is the direct consequence of the presence of dark energy and dark matter. For the weak interaction, PID is the requirement of the presence of the Higgs field. For the strong interaction, we demonstrated that PID is the consequence of the quark confinement and asymptotic freedom phenomena.

PRI requires that the gauge theory be independent of the choices of the representation generators. PRI is purely a logic requirement for the gauge theory and leads to profound consequences. One important consequence is as follows: for an $SU(N)$ gauge theory with $N \geq 2$, there are $N^2 - 1$ gauge potentials. PRI combines these potentials into a total gauge potential, giving rise to a precise force formula of the interaction and the precise meaning of the coupling constant as the charge of the interaction.

Another consequence of PRI is that the unification for the four interactions can only be achieved through coupling rather than through a large symmetry group. In fact, in deriving the law for a system consisting of several subsystems, the subsystems must be coupled in the energy level by adding up the energies (functionals) F_i of the subsystems: $F = \sum_i F_i$. Then, the combined F dictates the physical law for the

coupled system, where part of the symmetries of the subsystems must be broken (PSB). In other words, the essence of physics and unification is coupling.

THIRD, we discovered in Ma and Wang (2017a) a general dynamic law for all physical motion systems, based on a new variational principle with constraint-infinities.

Sections 3.1 and 3.2

FIRST, Chap. 3 studies equilibrium phase transitions in statistical physics. In revised Sect. 3.1, we introduce the basic dynamical model of thermodynamical systems, also called the standard model of statistical physics. This model is based on the new potential-descending principle (PDP) discovered by Ma and Wang (2017a), which will be systematically introduced in Sect. 7.1. With the standard model, the main issue boils down to finding a better and more accurate account of the thermodynamic potentials, which are dictated by the symmetries of the underlying physical systems. Currently, there are two approaches for the derivation of these potentials and Hamiltonian energy functionals. The first is the classical approach based on the Landau's mean field theoretical approach. The second is based only on the underlying symmetries and first principles and is presented in Sects. 7.6, 8.3, and 8.4 in this book.

The general theoretical framework and fluctuation theory in Sects. 3.1 and 3.2 are based on the second approach, the first-principle approach; while applications in the remaining sections of Chap. 3 is using the Landau's mean field theory approach as in the first edition. Of course, the same results can also be obtained using the second approach with the functionals in Sects. 7.6, 8.3, and 8.4.

SECOND, with the standard model, the dynamical transition theory can then be applied, and we obtain in Sect. 3.1 three basic theorems, first established by Ma and Wang (2017b), which provide a full theoretical characterization of thermodynamic phase transitions. The first theorem, Theorem 3.1.3, states that as soon as the linear instability occurs, the system always undergoes a dynamical transition to one of the three types: continuous, catastrophic, and random. This theorem offers the detailed information for the *phase diagram* and the *critical threshold* λ_c in the control parameter $\lambda = (\lambda_1, \dots, \lambda_N) \in \mathbb{R}^N$.

The second theorem, Theorem 3.1.4, provides the corresponding relationship between the Ehrenfest classification and the dynamical classification. The Ehrenfest classification identifies experimental quantities to determine the transition types, while the dynamical classification provides a full theoretical characterization of the transition. This theorem establishes a natural bridge between theory and experiments. With the corresponding relationship given in this theorem, we obtain a precise and easy theoretical approach to determine the transition order, the critical exponents, and the transition diagram. Also, this theorem shows that there are only first-, second-, and third-order thermodynamic transitions. The third-order transition can hardly be determined by thermodynamic parameters experimentally. The

dynamical transition theory, however, offers an easy and clear approach to identify the third-order transition.

The last theorem, Theorem 3.1.5, states that both catastrophic and random transitions lead to saddle-node bifurcations, and both latent heat, superheated and supercooled, states always accompany the saddle-node bifurcation associated with the first-order transitions.

THIRD, there have been extensive studies of critical exponents for phase transitions of thermodynamic systems; see among many others (Stanley, 1971; Sengers and Shanks, 2009). However, there are still two basic open questions regarding critical exponents. The first is whether the critical exponents are universal. It is believed, though not proven, that they are indeed universal. The second question is why there is a discrepancy between theoretical and experiment values of critical exponents.

Section 3.2, based on Liu et al. (2017a), shows that the critical exponents are indeed universal and the discrepancy is due to fluctuations. More precisely, the aim of this section is

1. To establish a dynamical law of fluctuations based on the standard model of thermodynamics
2. To derive the theoretical critical exponents using the standard model
3. To deduce the critical exponents based on the standard model with fluctuations, leading to correct critical exponents in agreement with experimental results
4. To show that the standard model of statistical physics derived from first principles (Ma and Wang, 2017a; Liu et al., 2019), as described in Sect. 3.1, is in agreement with experiments

Chapter 7

This chapter presents the recently developed first-principle approach to statistical physics and quantum physics in the spirit of the guiding principles of physics, Principles 1.1.1 and 1.1.2, as mentioned earlier.

One aim of this chapter is to introduce the potential-descending principle (PDP) discovered recently by Ma and Wang (2017a).

PDP consists of two parts. First, for a given thermodynamic system, the order parameters (state functions) $u = (u_1, \dots, u_N)$, the control parameters λ , and the thermodynamic potential (or potential in short) F are well-defined quantities, fully describing the system. Second, for a nonequilibrium state $u(t; u_0)$ with initial state $u(0, u_0) = u_0$, the potential $F(u(t; u_0); \lambda)$ is strictly decreasing as time evolves, and the order parameters $u(t; u_0)$, as time evolves to infinity, tends to an equilibrium of the system, which is a minimal point of the potential F .

We show that PDP, together with the classical principle of equal a priori probabilities (PEP), leads to the first and second laws of thermodynamics, irreversibility, and the three classical statistics: the Maxwell–Boltzmann, the Fermi–Dirac, and the Bose–Einstein statistics. Consequently, PDP and PEP are the first principles of statistical physics.

Also, PDP leads to the standard model of statistical physics given in the form of (7.1.1) and (7.1.2), as described earlier for the revised Sect. 3.1. Again, the basic issue boils down to deriving precise form of thermodynamic potentials F , which are fully dictated by underlying symmetries of the physical system as presented in Sect. 7.6.

We remark that in all classical derivation of the potential-decreasing property from the second law, the potential-decreasing property is itself used as a hidden assumption. Also, the three statistics are derived classically using the entropy being at the maximum, which is not a principle of nature. Finally, we remark that PDP offers a clear description of the irreversibility of thermodynamical systems as the potential decreases. Notice that entropy S is a state function, which is the solution of basic thermodynamic equations. Hence, the thermodynamic potential is a higher-level physical quantity than entropy and, consequently, is the correct physical quantity, rather than the entropy, for describing irreversibility for all thermodynamic systems.

The second aim of this chapter is to present a statistical theory of heat first discovered in Ma and Wang (2017d). One motivation of the theory is the lack of physical carriers of heat in classical thermodynamics. Another motivation is the recent discovered photon cloud structure of electrons (Ma and Wang, 2015a). This leads to a natural conjugate relation between electrons and photons, reminiscent of the conjugate relation between temperature and entropy. The new theory of heat provides a natural connection between these two conjugate relations: at the equilibrium of absorption and radiation, the average energy level of the system maintains unchanged, and represents the temperature of the system; at the same time, the number (density) of photons in the sea of photons represents the entropy (entropy density) of the system. The theory contains four parts: (1) the photon number formula of entropy, (2) the energy level formula of temperature, (3) the temperature theorem, and (4) the thermal energy formula. In particular, the photon number entropy formula is equivalent to the Boltzmann entropy formula and, however, possesses new physical meaning that the physical carrier of heat is the photons.

The third aim is to introduce a new field theoretical interpretation of quantum mechanical wave functions, first discovered in Ma and Wang (2019). We postulate that the wave function $\psi = |\psi|e^{i\varphi}$ is the common wave function for all particles in the same class determined by the external potential $V(x)$, $|\psi(x)|^2$ represents the distribution density of the particles, and $\frac{\hbar}{m}\nabla\varphi$ is the velocity field of the particles. This is an entirely different interpretation from the classical Bohr interpretation. Also, this new interpretation of wave functions does not alter the basic theories of quantum mechanics, and instead offers new understanding of quantum mechanics, and plays a fundamental role for the quantum theory of condensed matter physics and quantum physics.

Again, this interpretation of wave functions is a **field** theoretical interpretation. In fact, fields are used as the fundamental level of physical quantities to describe the nature.

Chapter 8

The study in this chapter is strongly motivated by the lack of physical mechanism of condensates and high-temperature superconductivity. The phenomenon of superconductivity was first discovered by Kamerlingh Onnes in 1911, in metallic mercury below 4 K. Since then, the main goals of study have been to attempt to observe superconductivity at increasing temperatures and to derive its physical mechanism.

The classical BCS theory (Bardeen et al., 1957) is a microscopic theory based on the Cooper electron-pair mechanism, which was used to explain superconducting currents, the Meissner effect, the isotope effect, and to calculate specific heats and penetration depths. However, the BCS theory is applicable only for $T < 35$ K.

In 1986, Bednorz and Müller (1986) discovered superconductivity in a lanthanum-based cuprate perovskite material, which had a transition temperature of 35 K (Nobel Prize in Physics, 1987). Bednorz and Müller's work opened the door for high T_c superconductors. The current highest temperature of superconductors reaches $T_c = 130$ K and $T_c = 160$ K under high pressure.

The mechanism presented in this chapter is based on Ma and Wang (2019). There are three main ingredients for the quantum mechanism of high T_c superconductivity. The first ingredient is the field theoretical interpretation of quantum mechanical wave functions as discussed earlier. With this interpretation, we show that the key for condensation of bosonic particles is that their interaction is sufficiently weak to ensure that a large collection of boson particles are in a state governed by the same condensation wave function field ψ under the same bounding potential V . For superconductivity, the formation of superconductivity comes down to conditions (1) for the formation of electron-pairs and (2) for the electron-pairs to share a common wave function.

The second ingredient is the layered PID interaction potential of electrons (Ma and Wang, 2015a), modifying the classical Coulomb law. As partially described before, PID and PRI are the first principle for deriving the correct interaction potentials and force formulas. The layered PID interaction potential of electrons, together with the average energy-level formula of temperature (Ma and Wang, 2017d), and the two conditions in the above mechanism of superconductivity are explicitly derived.

The third ingredient is the derivation of both microscopic and macroscopic formulas for the critical temperature. The field and topological phase transition equations for condensates are also derived.

Chapter 9

As mentioned earlier, all phase transitions in nature that we have encountered can be classified into two types: (a) dynamical phase transitions and (b) topological phase transitions (TPTs). TPTs are entirely different from dynamic phase transitions. Intuitively speaking, a TPT refers to the change of the topological structure in the physical space as certain system control parameter crosses a critical threshold. Basically, the state function u of a physical system is a solution of the partial

differential equation (PDE) model, and we are searching for the change of the topological structure of the solution u in the “physical space.” Consequently, one needs to see both the evolution of the PDE model and the evolution of its solution structure in the physical space. This nature of the problem makes the study of TPTs extremely challenging.

This chapter aims to develop a systematic theory of TPTs and explores a few typical examples, including

- (a) Quantum phase transitions (QPTs)
- (b) Galactic spiral structures
- (c) Electromagnetic eruptions on solar surface
- (d) Boundary-layer separation of fluid flows
- (e) Interior separation of fluid flows

As mentioned in the preface to the first edition, the book is intended for graduate students and researchers in mathematics, physics, and other related science fields. Portions of the book have been used as the textbook for graduate courses in nonlinear dynamics, statistical physics, and quantum physics by the authors in Sichuan University and Indiana University.

Acknowledgments

We would like to thank three anonymous referees for their insightful suggestions. We are grateful to Ms. Donna Chernyk at Springer for her professional effort and support.

We would like to express all of our gratitude to our wives, Li and Ping, and children, Jiao, Wayne, and Melinda, for their unflinching love, support, and appreciation.

Sichuan, China
Bloomington, IN, USA

Tian Ma
Shouhong Wang

Preface to the First Edition

Most problems in the natural sciences are described by either dissipative or conservative systems. Phase transition dynamics for both types of systems is of central importance in the nonlinear sciences. In this book, the term “phase transition” is to be understood in a broad sense, including both classical phase transitions and general transitions as found in nature. This book is an introduction to a comprehensive and unified dynamic transition theory for dissipative systems and to applications of the theory to a wide range of problems in the nonlinear sciences.

The main objectives of this book are to derive a general principle of dynamic transitions for dissipative systems, to establish a systematic dynamic transition theory, and to explore the physical implications of applications of the theory to a wide range of problems in the nonlinear sciences.

As a general principle, dynamic transitions of all dissipative systems are classified into three categories: continuous, catastrophic, and random. In comparison with the classical classification scheme for equilibrium phase transitions, whereby phase transitions are identified by the lowest-order derivative of the free energy that is discontinuous at the transition, the dynamic classification scheme is suitable for both equilibrium and nonequilibrium phase transitions. Once the type of a dynamic transition is determined for a given equilibrium system, the order of a transition in the classical sense immediately becomes transparent.

The philosophical basis of dynamic transition theory is to search for a system’s full set of transition states, giving a complete characterization of stability and transition. The set of transition states is represented by a local attractor near or away from the basic state. Following this approach, dynamic transition theory is developed to identify the transition states and to classify them both dynamically and physically.

The theory is strongly motivated by phase transition problems in the nonlinear sciences. Namely, the mathematical theory is developed with close links to physics, and in return, the theory is applied to physical problems, leading to physical predictions and new insights into both theoretical and experimental studies for the underlying physical problems.

The study of equilibrium phase transitions presented in this book involves a combination of modeling, mathematical analysis, and physical predictions. We adopt the

general idea of the Ginzburg–Landau phenomenological approach, and introduce a unified time-dependent Ginzburg–Landau model. Mathematically, this model is a quasigradient flow. Physically, the model is proposed using the Le Châtelier principle. A few typical and important equilibrium phase transition problems are addressed in this book, including the physical vapor transport (PVT) system, ferromagnetism, binary systems (Cahn–Hilliard equation), and the superconductivity and superfluidity of liquid helium (helium-3, helium-4 and their mixture).

The main focus for nonequilibrium transitions is on typical problems in classical and geophysical fluid dynamics, in climate dynamics, in chemical reactions, and in biology, including in particular Rayleigh–Bénard convection, the Taylor problem, Taylor–Couette–Poiseuille flow, and the rotating convection problem in classical fluid dynamics; the El Niño–Southern Oscillation (ENSO) and wind-driven and thermohaline circulation in geophysical fluid dynamics and climate dynamics; the Belousov–Zhabotinsky chemical reactions; and chemotaxis and population models in biology.

Our intent has been to address the interests and backgrounds not only of applied mathematicians but also of sophisticated students and researchers working on nonlinear problems including those in physics, meteorology, oceanography, biology, chemistry, and the social sciences. The audience for this book includes second-year or more-advanced graduate students and researchers in mathematics and physics as well as in other related fields. Chapter 1 is designed so that most readers who have completed it will be able to jump directly to the related applications in Chaps. 3–6, after going quickly through the main ideas in this chapter as well as in the general introduction.

The mathematical background for physicists and other researchers in other fields is a good graduate-level mathematical physics course. For example, any student who is capable of studying graduate-level quantum mechanics or statistical physics should be able to follow Chap. 1 and the applications in Chaps. 3–6. Most graduate students in applied mathematics nowadays have to learn some basic physics related to the applied problems they are working on, and they should have no additional difficulty in reading the related topics of their interest in this book.

Acknowledgments

We have greatly benefited from illuminating discussions with Jerry Bona, Mickael Chekroun, Henk Dijkstra, Ciprian Foias, Susan Friedlander, Michael Ghil, Chun-Hsiung Hsia, Paul Newton, Hans Kaper, Louis Nirenberg, Benoit Perthame, Jie Shen, Roger Temam, Xiaoming Wang, and Kevin Zumbrun. We owe a special debt of gratitude to all of them. We are grateful to Honghu Liu and Taylan Sengul for their critical reading, several times, of an earlier draft of the book and for their many constructive comments. We extend our wholehearted thanks to the anonymous referees; their careful reading and thoughtful remarks have led to a more effective book.

Also, we are grateful to Wen Masters and Reza Malek-Madani, of the Office of Naval Research, for their constant support and encouragement. We express our sincerest thanks to Achi Dosanjh, of Springer-Verlag, for her great effort and support on this book project. The research presented in this book was supported in part by grants from the ONR, the NSF, and the Chinese NSF.

The authors owe a debt beyond our ability to express it to Li and Ping for their kindness, love, patience, and dedications to our families. This book could never have been written without their devoted support from both Li and Ping.

Sichuan, China
Bloomington, IN, USA

Tian Ma
Shouhong Wang

Introduction

Transitions are to be found throughout the natural world. The laws of nature are usually represented by differential equations, which can be regarded as dynamical systems—both finite- and infinite-dimensional. There are two types of dynamical systems: dissipative and conservative. The term “dissipative structure” was coined by Ilya Prigogine. A conservative system is often described by a Hamiltonian structure, and a dissipative structure is closely associated with the concept of absorbing sets and global attractors (see among many others Lorenz 1963a; Prigogine and Lefever 1968; Foiaş and Temam 1979; Ladyzenskaja 1991; Temam 1997).

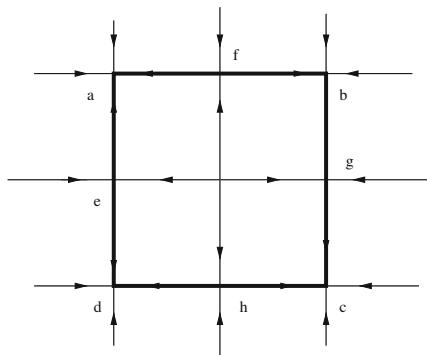
In thermodynamics, a phase transition is understood as the transformation of a thermodynamic system from one phase or state of matter to another. However, many problems in nature involve the study of transitions of one state to new states and the stability/robustness of the new states, and so in this book, the term “phase transition” will be understood in a broad sense, to include both the classical phase transitions in physics and the general transitions found in nature. The study of dynamic transitions also involves the dynamic behavior of transitions and transition states.

This main objectives of this book are threefold. The first is to derive, in Theorem 2.1.3, the following general principle on dynamic transitions for dissipative systems:

Principle 1 *Dynamic transitions of all dissipative systems can be classified into three categories: continuous, catastrophic, and random.*

The second objective is to develop a systematic dynamic transition theory to study the types and structure of dynamic transitions for dissipative systems. The third is to apply dynamic transition theory in order to explore the physical implications of dynamic transitions in a wide range of scientific problems. The applications include most of the typical types of dissipative systems, including (1) gradient-type systems for equilibrium phase transitions in statistical physics, (2) Navier–Stokes-type equations for incompressible fluid flows, (3) Boussinesq-type equations for geophysical fluid dynamics and climate dynamics, and (4) systems of reaction-diffusion equations modeling chemical reactions, chemotaxis, and population dynamics in biology.

Fig. 1 The states after the transition is given by a bifurcated attractor containing four nodes (the points a, b, c, and d), four saddles (the points e, f, g, h), and orbits connecting these eight points



The states of a dissipative system can be a complicated object in the phase space and may include steady states, periodic states, and transients. Hence, the starting point of our study is the following key philosophy:

Philosophy 2 *The key philosophy of dynamic transition theory is to search for the full set of transition states, giving a complete characterization of stability and transition. The set of transition states is a local attractor, representing the physical reality after the transition.*

It is natural to understand this philosophy from the physical point of view. For example, according to von Neumann (1960), the motion of the atmosphere can be divided into three categories depending on the time scale of the prediction. They are motions corresponding, respectively, to the short-, medium-, and long-term behavior of the atmosphere. Climate, which corresponds to the long-term behavior, should be represented by an attractor containing not only climate equilibria but also the transients.

The general principle for dynamic transitions for dissipative systems given by Principle 1, along with Philosophy 2, provides a global viewpoint for the study of dynamic transitions. To illustrate the basic motivation and ideas behind this point of view, we consider the following simple example.

Example 3 For $x = (x_1, x_2)^t \in \mathbb{R}^2$, the system

$$\dot{x}_1 = \lambda x_1 - x_1^3, \quad \dot{x}_2 = \lambda x_2 - x_2^3 \quad (1)$$

undergoes a dynamic transition from a basic stable state $x = 0$ to a local attractor $\Sigma_\lambda = S^1$, as λ crosses $\lambda_0 = 0$. This local transition attractor is as shown in Fig. 1. It contains exactly four nodes (the points a, b, c, and d), four saddles (the points e, f, g, h), and orbits connecting these eight points. The connecting orbits represent transient states.

From the physical transition point of view, as λ crosses 0, the new transition states are represented by the whole local attractor Σ_λ , rather than by any of the steady states or any of the connecting orbits.

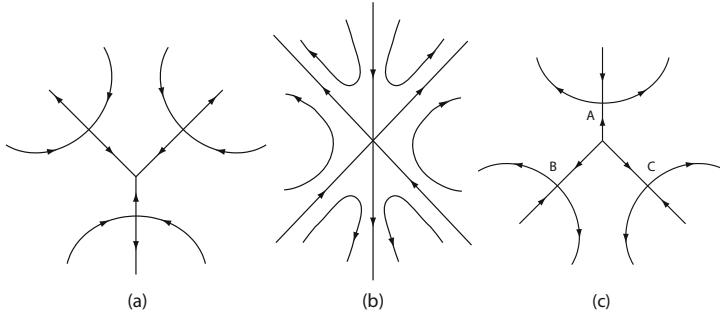


Fig. 2 Transition structure from a singular point with index $\text{ind}(F, 0) = -2$: (a) $\lambda < 0$; (b) $\lambda = 0$; (c) $\lambda > 0$

To demonstrate this point of view further, we consider another example:

Example 4 For $x = (x_1, x_2)^t \in \mathbb{R}^2$, consider the system

$$\begin{aligned}\dot{x}_1 &= \lambda x_1 + x_1^2 + x_1 x_2 - 10x_1^3, \\ \dot{x}_2 &= \lambda x_2 - 2x_1 x_2 - x_2^2 - 10x_2^3.\end{aligned}\tag{2}$$

In this example, the lowest-order nonlinear terms are quadratic, given by $F(x) = (x_1^2 + x_1 x_2, -2x_1 x_2 - x_2^2)^t$. As we shall prove in Chap. 2, the index of the vector field $F(\cdot)$ at $x = 0$ is -2 , leading to a dynamic transition as shown in Fig. 2. For $\lambda > 0$ and near 0, there are seven steady-state solutions: the origin $O = (0, 0)^t$ and six nonzero points A, B, C, d, e , and f such that the three fixed points A, B , and C are close to the origin O and converge to O as $\lambda \rightarrow 0$ and the other three fixed points are away from O and do not converge to the origin O as $\lambda \rightarrow 0$. The three line segments OA, OB , and OC divide a neighborhood of O into three regions, and for any initial value in one of the regions, the solution converges to one of the three fixed points d, e , and f , away from O . Hence, as λ crosses 0, the transition is catastrophic, also called in this book a jump (Type-II) transition. The three points A, B , and C near the origin are the three bifurcation points, which do not represent the transition state in any way.

We note that the six nonzero fixed points are connected by heteroclinic orbits, and the region enclosed by these orbits is the global attractor of the system. \square

The above two simple examples clearly demonstrate the necessity of the key philosophy, which will become more transparent in the wide range of applications of the theory throughout the book.

Dissipative systems are governed by differential equations—both ordinary and partial—which can be written in the following unified abstract form:

$$\frac{du}{dt} = L_\lambda u + G(u, \lambda), \quad u(0) = u_0,\tag{3}$$

where $u : [0, \infty) \rightarrow X$ is the unknown function, $\lambda \in \mathbb{R}^1$ is the system control parameter, and X is a Banach space.

We shall always consider u the deviation of the unknown function from some equilibrium state \bar{u} . Hence, $L_\lambda : X \rightarrow X$ is a linear operator, and $G : X \times \mathbb{R}^1 \rightarrow X$ is a nonlinear operator. For example, for the classical incompressible Navier–Stokes equations, u represents the velocity function, and for the time-dependent Ginzburg–Landau equations of superconductivity, we have $u = (\psi, A)$, where ψ is a complex-valued wave function and A is the magnetic potential.

To address the dynamic transition of a given dissipative system, the first step is to study the linear eigenvalue problem for system (3), which is closely related to the principle of exchange of stability (PES), leading to precise information on linear unstable modes. The precise statement of the PES is as follows. Let $\{\beta_j(\lambda) \in \mathbb{C} \mid j \in \mathbb{N}\}$ be the eigenvalues (counting multiplicity) of L_λ , and assume that

$$\operatorname{Re} \beta_i(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_0, \\ = 0 & \text{if } \lambda = \lambda_0, \\ > 0 & \text{if } \lambda > \lambda_0, \end{cases} \quad \forall 1 \leq i \leq m, \quad (4)$$

$$\operatorname{Re} \beta_j(\lambda_0) < 0 \quad \forall j \geq m + 1. \quad (5)$$

Much of the linear theory on stability and transitions involves establishing the PES. There is a vast literature devoted to linear theory, including, among many others, Chandrasekhar (1981), Drazin and Reid (1981) for classical fluid dynamics and Pedlosky (1987) for geophysical fluid dynamics.

With the linear theory and the PES at our disposal, we can immediately show that the system always undergoes a dynamical transition at the critical threshold and that Principle 1 holds (see Theorem 2.1.3 for details).

The details of the transition behavior are then dictated by the nonlinear interactions of the system. The dynamic transition theory presented in Chap. 2 provides a systematic approach to the study of nonlinear interactions, leading to detailed information on the types of transitions and the structure of transition states and their physical implications.

The general principle of dynamic transitions for dissipative systems, Principle 1, classifies all dynamic transitions of a dissipative system into three categories, continuous, catastrophic, and random, which are also called Type-I, Type-II, and Type-III in this book. To say that a transition is continuous amounts to saying that the control parameter crosses the critical threshold and the transition states stay in a close neighborhood of the basic state. The transition at $\lambda = 0$ in Example 3 is a continuous transition.

In fact, continuous transitions are essentially characterized by the attractor bifurcation theorem, Theorem 2.2.2, first proved in Ma and Wang (2005c,b). The attractor bifurcation theorem states that when the PES holds and the basic state is asymptotically stable at the critical parameter value λ_0 , the system undergoes a continuous dynamic transition, which is described by a bifurcated attractor.

A key assumption in the attractor bifurcation theorem is the asymptotic stability of the basic solution at the critical parameter value λ_0 . One important idea of the proof is a refined upper semicontinuity of attractors, which also plays an important role in analyzing the other two types of transitions.

There are many physical systems that can undergo a continuous transition. For example, consider the classical Bénard convection. As the Rayleigh number crosses the critical Rayleigh number, the system undergoes a continuous transition to an attractor, homeomorphic to an $(m - 1)$ -dimensional sphere S^{m-1} , that consists of steady states and transients. Here, m is the number of unstable modes of the linearized eigenvalue problem at the critical Rayleigh number, dictated by the spatial geometry, which also defines the pattern formation mechanism of the problem. As we shall see in Sect. 4.1, steady states occupy only a zero-measure subset of this sphere S^{m-1} , and the transients occupy most the sphere.

When there is no longer asymptotic stability of the basic state at the critical parameter, the system undergoes either catastrophic or random transitions, as dictated by the nonlinear interactions. The dynamic transition theory presented in this book gives a systematic approach to distinguishing these transitions.

Intuitively speaking, catastrophic transition corresponds to the case in which the system undergoes a more drastic change as the control parameter crosses the critical threshold. The transition given in Example 4 and the transitions in many examples given in this book are typical catastrophic transitions.

A random transition corresponds to the case in which a neighborhood (fluctuations) of the basic state can be divided into two regions such that fluctuations in one of them lead to continuous transitions and those in the other lead to catastrophic transitions.

Given the above observations, it is then crucial to determine the asymptotic behavior of a system at the critical threshold. For this purpose and for determining the structure of the local attractor representing the transition states, the most natural approach is to project the underlying system to the space generated by the most unstable modes, preserving the dynamic transition properties. This is achieved with center manifold reduction.

An important part of dynamic transition theory deals with the analysis of reduced systems with close links to physics and with the philosophy of searching for a complete set of transition states. First, the transition type is completely dictated by the flow structure of the reduced system at the critical threshold, leading to precise information on the phase transition diagram. Also, for a continuous (Type-I) transition, the reduced system provides a complete description of the set of transition solutions. It is worth mentioning that the far-field solutions in the catastrophic and random cases will have to be derived from the original partial differential equations, although the transition types and phase diagrams are completely determined by the reduced system.

In a nutshell, this book establishes a general principle for dynamic transitions for dissipative systems. Namely, all transitions of a dissipative system can be dynamically classified into one of the three types: continuous, catastrophic, and random. A systematic dynamic transition theory is then developed to identify the types and

structure of transitions and is further applied to a wide range of problems in physics, fluid dynamics, climate dynamics, chemistry, and biology.

It is believed that the methods and ideas presented in this book can be readily applied to many scientific problems in related fields as well, as evidenced by the wide range of applications presented in this book.

The main outline of this book is as follows. The philosophy and general principles of dynamic transition theory with applications to a few typical problems are addressed in Chap. 1. Mathematical aspects of dynamic transition theory are addressed in Chap. 2.

Chapter 3 consists of applications of the general theory to a few typical equilibrium phase transition problems, including the physical vapor transport (PVT) system, ferromagnetism, binary systems, superconductivity, and superfluidity. As mentioned earlier, the study of these equilibrium phase transition problems involves a combination of modeling, mathematical analysis, and physical predictions.

Chapter 4 focuses on transition problems in classical fluid dynamics. The transition of the classical Bénard convection problem is studied in Sect. 4.1, leading to a continuous (Type-I) transition at the first critical Rayleigh number. For the Taylor problem, the Taylor–Couette–Poiseuille (TCP) flow problem, and the rotating convection problem, the associated linear operators are not symmetric, leading to the existence of all three types of transitions as well as the existence of periodic solutions. This chapter gives a detailed account of these transitions and their physical implications. In the TCP flow case, for example, we show that the Taylor vortices do not appear immediately after the first transition, and that appear only after the system has undergone a transition in its structure in the physical space as we increase the Taylor number to another critical number. This result is derived using a combination of the dynamic transition theory presented in this book and the geometric theory of incompressible flows developed recently in Ma and Wang (2005d).

Chapter 5 addresses dynamic transitions in geophysical fluid dynamics and climate dynamics. One important aspect of the study is the introduction of turbulent frictional terms in the model, leading to a transition to circulations over the tropics, consistent with the Walker circulation. Another important component of this chapter is the introduction of a new mechanism of El Niño–Southern Oscillation (ENSO) that is consistent with observations. Dynamic transitions corresponding to typical meridional circulations and the thermohaline circulations are examined as well.

Application-oriented readers may go directly to the study of the scientific problems starting from Chaps. 3 after reading through Chap. 1.

Contents

1	Introduction to Dynamic Transitions	1
1.1	First Principles and Dynamic Models	1
1.1.1	Guiding Principles of Theoretical Physics	1
1.1.2	Theory of Four Fundamental Interactions	3
1.1.3	Dynamic Law of Physical Motion	4
1.1.4	Rayleigh–Bénard Convection	10
1.1.5	Mathematical Formulation of Physical Problems	12
1.2	Introduction to Dynamic Transition Theory	14
1.2.1	Motivations and Key Philosophy	14
1.2.2	Principle of Exchange of Stability	14
1.2.3	Equation of Critical Parameters	15
1.2.4	Classifications of Dynamic Transitions	16
1.2.5	Structure and Characterization of Dynamic Transitions	18
1.2.6	General Features of Dynamic Transitions	19
1.3	Examples of Typical Phase Transition Problems	21
1.3.1	Rayleigh–Bénard Convection	21
1.3.2	El Niño Southern Oscillation (ENSO)	22
1.3.3	Dynamic Transition vs Transition in the Physical Space	24
1.3.4	Andrews Critical Point and Third-Order Gas-Liquid Transition	26
1.3.5	Binary Systems	28
2	Dynamic Transition Theory	31
2.1	General Dynamic Transition Theory	31
2.1.1	Classification of Dynamic Transitions	31
2.1.2	Characterization of Transition Types	34
2.1.3	Local Topological Structure of Transitions	36
2.2	Continuous Transition	40
2.2.1	Finite Dimensional Systems	40
2.2.2	S^1 -Attractor Bifurcation	41
2.2.3	S^m -Attractor Bifurcation	48

2.2.4	Structural Stability of Dynamic Transitions	48
2.2.5	Infinite Dimensional Systems	50
2.3	Transition from Simple Eigenvalues	52
2.3.1	Real Simple Eigenvalues	52
2.3.2	Transitions from Complex Simple Eigenvalues	55
2.3.3	Computation of b	61
2.4	Transition from Eigenvalues with Multiplicity Two	63
2.4.1	Index Formula for Second-Order Nondegenerate Singularities	63
2.4.2	Bifurcation at Second-Order Singular Points	66
2.4.3	Case where $\text{Ind}(F, 0) = -2$	71
2.4.4	Case Where $\text{Ind}(F, 0) = 2$	76
2.4.5	Case Where $\text{Ind}(F, 0) = 0$	83
2.4.6	Indices of k -th Order Nondegenerate Singularities	85
2.4.7	Structure of k -th Order Nondegenerate Singularities	96
2.4.8	Transition from k -th Order Nondegenerate Singularities	99
2.4.9	Bifurcation to Periodic Orbits	102
2.4.10	Application to Parabolic Systems	105
2.5	Singular Separation	107
2.5.1	General Principle	107
2.5.2	Saddle-Node Bifurcation	109
2.5.3	Singular Separation of Periodic Orbits	111
2.6	Perturbed Systems	113
2.6.1	General Eigenvalues	113
2.6.2	Simple Eigenvalues	114
2.6.3	Complex Eigenvalues	119
2.7	Notes	126
3	Equilibrium Phase Transitions in Statistical Physics	129
3.1	Dynamical Theory of Thermodynamic Phase Transitions	132
3.1.1	Standard Model of Statistical Physics	132
3.1.2	Ehrenfest Classification	135
3.1.3	Three Basic Theorems of Thermodynamic Phase Transitions	136
3.2	Dynamic Theory of Fluctuations and Critical Exponents	140
3.2.1	Statistical Theory of Fluctuation	141
3.2.2	Dynamical Theory of Fluctuations	144
3.2.3	Fluctuation Effect on Critical Parameters	149
3.2.4	Theoretical Critical Exponents of Standard Models	152
3.2.5	Critical Exponents with Fluctuation Influences	158
3.3	Third-Order Gas-Liquid Phase Transition	164
3.3.1	Introduction	164
3.3.2	Time-Dependent Models for PVT Systems	165
3.3.3	Phase Transition Dynamics for PVT Systems	167
3.3.4	Physical Conclusions	169

3.4	Ferromagnetism	173
3.4.1	Classical Theory of Ferromagnetism	173
3.4.2	Dynamic Transitions in Ferromagnetism	177
3.4.3	Physical Implications	180
3.4.4	Asymmetry of Fluctuations	182
3.5	Phase Separation in Binary Systems	183
3.5.1	Modeling	184
3.5.2	Phase Transition in General Domains	185
3.5.3	Phase Transition in Rectangular Domains	190
3.5.4	Spatial Geometry, Transitions, and Pattern Formation	201
3.5.5	Phase Diagrams and Physical Conclusions	204
3.6	Superconductivity	211
3.6.1	Ginzburg-Landau Model	211
3.6.2	TGDL as a Gradient-Type System	216
3.6.3	Phase Transition Theorems	219
3.6.4	Model Coupled with Entropy	226
3.6.5	Physical Conclusions	227
3.7	Liquid Helium-4	237
3.7.1	Dynamic Model for Liquid Helium-4	238
3.7.2	Dynamic Phase Transition for Liquid ^4He	242
3.8	Superfluidity of Helium-3	248
3.8.1	Dynamic Model for Liquid ^3He with Zero Applied Field ..	250
3.8.2	Critical Parameter Curves and PT -Phase Diagram	253
3.8.3	Classification of Superfluid Transitions	258
3.8.4	Liquid ^3He with Nonzero Applied Field	265
3.8.5	Physical Remarks	268
3.9	Mixture of He-3 and He-4	269
3.9.1	Model for Liquid Mixture of ^3He - ^4He	270
3.9.2	Critical Parameter Curves	271
3.9.3	Transition Theorems	273
3.9.4	Physical Conclusions	278
4	Fluid Dynamics	279
4.1	Rayleigh–Bénard Convection	280
4.1.1	Bénard Problem	280
4.1.2	Boussinesq Equations	281
4.1.3	Dynamic Transition Theorems	283
4.1.4	Topological Structure and Pattern Formation	288
4.1.5	Asymptotic Structure of Solutions for the Bénard problem ..	297
4.1.6	Structure of Bifurcated Attractors	299
4.1.7	Physical Remarks	307
4.2	Taylor–Couette Flow	307
4.2.1	Taylor Problem	308
4.2.2	Governing Equations	309
4.2.3	Narrow-gap Case with Axisymmetric Perturbations	312

4.2.4	Asymptotic Structure of Solutions and the Taylor vortices	317
4.2.5	Taylor Problem with z -Periodic Boundary Condition	323
4.2.6	Other Boundary Conditions	339
4.2.7	Three-Dimensional Perturbation for Narrow-Gap Case	344
4.2.8	Physical Remarks	346
4.3	Boundary-Layer and Interior Separations in the Taylor–Couette–Poiseuille Flow	347
4.3.1	Model for the Taylor–Couette–Poiseuille Problem	348
4.3.2	Phase Transition of the TCP Problem	352
4.3.3	Boundary-Layer Separation from the Couette–Poiseuille Flow	356
4.3.4	Interior Separation from the Couette–Poiseuille Flow	362
4.3.5	Nature of Boundary-Layer and Interior Separations	367
4.4	Rotating Convection Problem	371
4.4.1	Rotating Boussinesq Equations	372
4.4.2	Eigenvalue Problem	373
4.4.3	Principle of Exchange of Stabilities	378
4.4.4	Transition from First Real Eigenvalues	385
4.4.5	Transition from First Complex Eigenvalues	391
4.4.6	Physical Remarks	396
4.5	Convection Scale Theory	399
5	Geophysical Fluid Dynamics and Climate Dynamics	405
5.1	Modeling and General Characteristics of Geophysical Flows	406
5.2	El Niño–Southern Oscillation (ENSO)	408
5.2.1	Walker Circulation and ENSO	409
5.2.2	Equatorial Circulation Equations	410
5.2.3	Walker Circulation Under Idealized Conditions	412
5.2.4	Walker Circulation Under Natural Conditions	419
5.2.5	ENSO: Metastable Oscillation Theory	425
5.3	Thermohaline Ocean Circulation	428
5.3.1	Boussinesq Equations	430
5.3.2	Linear Analysis	432
5.3.3	Nonlinear Dynamic Transitions	441
5.3.4	Convection Scales and Dynamic Transition	448
5.4	Large Scale Meridional Atmospheric Circulation	458
5.4.1	Polar, Ferrel, and Hadley Cells	458
5.4.2	β -plane Assumption	459
5.4.3	Meridional Circulation Under Idealized Condition	460
5.4.4	Physical Implications	464
6	Dynamical Transitions in Chemistry and Biology	467
6.1	Modeling	468
6.1.1	Dynamical Equations of Chemical Reactions	468
6.1.2	Population Models of Biological Species	470

6.2	Belousov–Zhabotinsky Chemical Reactions: Oregonator	472
6.2.1	Field–Körös–Noyes Equations	472
6.2.2	Transition Under the Dirichlet Boundary Condition	475
6.2.3	Transitions Under the Neumann Boundary Condition	478
6.2.4	Phase Transition in the Realistic Oregonator	490
6.3	Belousov–Zhabotinsky Reactions: Brusselator	495
6.3.1	Prigogine–Lefever Model	496
6.3.2	Linearized Problem	497
6.3.3	Transition From Real Eigenvalues	499
6.3.4	Transition from Complex Eigenvalues	501
6.4	Bacterial Chemotaxis	502
6.4.1	Keller–Segel Models	503
6.4.2	Dynamic Transitions for Rich Stimulant System	509
6.4.3	Transition of Three-Component Systems	519
6.4.4	Biological Conclusions	526
7	Fundamental Principles of Statistical and Quantum Physics	531
7.1	Potential-Descending Principle	532
7.1.1	Potential-Descending Principle	532
7.1.2	Standard Model of Thermodynamic Systems	533
7.1.3	Irreversibility	535
7.1.4	Problems in Boltzmann Equation	536
7.2	First and Second Laws of Thermodynamics	541
7.2.1	Derivation of the First and Second Laws from PDP	541
7.2.2	Thermodynamic Potentials	542
7.3	Foundations of Quantum Physics	547
7.3.1	Principles of Quantum Mechanics	547
7.3.2	New Interpretation of Quantum Wave Functions	550
7.3.3	Introduction to Elementary Particles	553
7.3.4	Photon Cloud Model of Electrons	555
7.3.5	Energy Levels of Particles	558
7.4	Statistics of Equilibrium Systems	562
7.4.1	Maxwell–Boltzmann Statistics for Classical Systems	562
7.4.2	Quantum Systems	564
7.5	Statistical Theory of Heat	566
7.5.1	Energy Level Formula of Temperature	566
7.5.2	Physical Meaning of the Temperature Formula	571
7.5.3	Theory of Entropy	572
7.5.4	Essence of Heat	575
7.6	Thermodynamic Potentials	579
7.6.1	$SO(n)$ Symmetry and Basic Forms of Thermodynamic Potentials	580
7.6.2	PVT Systems	586
7.6.3	N-Component Systems	590
7.6.4	Magnetic and Dielectric Systems	592

8 Quantum Mechanism of Condensates and High Tc Superconductivity	597
8.1 Quantum Mechanism of Condensation	597
8.1.1 Quantum Mechanism of Condensates	597
8.1.2 Field Equations of Condensates	599
8.1.3 Topological Structure Equations of Condensates	600
8.2 Theory for High-Temperature Superconductivity	602
8.2.1 Physical Mechanism of Superconductivity	602
8.2.2 PID Interaction Potential of Electrons	604
8.2.3 Condition for Formation of Electron-Pairs	606
8.2.4 Binding Energy of Superconducting Electron-Pairs	610
8.2.5 Formulas for the Critical Temperature T_c	612
8.3 Thermodynamic Potentials for Condensates	616
8.3.1 Quantum Rules of Condensates	616
8.3.2 Ginzburg-Landau Free Energy of Superconductivity	618
8.3.3 Thermodynamic Potential for Liquid ^4He	620
8.3.4 Liquid ^3He Superfluidity	622
8.3.5 Gibbs Free Energy for Gaseous Condensates (BEC)	625
8.4 Quantum Systems of Condensates	629
8.4.1 Spin Operators of Spinors	629
8.4.2 Hamiltonian Energy of Superconductors	631
8.4.3 Superfluid Systems	633
8.4.4 Gaseous BEC Systems	635
8.5 Field Equations for Condensates	636
8.5.1 Field Equations of Superconductivity	636
8.5.2 Field Equations of Superfluidity and BEC	637
9 Topological Phase Transitions	639
9.1 Topological Theory of 2D Incompressible Flows	639
9.1.1 Basic Concepts	639
9.1.2 Structural Stability Theorems	640
9.1.3 Structural Bifurcations on Boundary	642
9.1.4 Interior Separation	643
9.2 Quantum Phase Transitions	644
9.2.1 Definition and Examples	644
9.2.2 Topological Index of Quantum Condensates	649
9.2.3 Microscopic Mechanism of the Meissner Effect	652
9.2.4 Abrikosov Vortex–Meissner Phase Transition	654
9.2.5 QPT Theorem for Scalar BEC Systems	657
9.3 Formation of Galactic Spiral Structure	659
9.3.1 Formation Mechanism of Spiral Galaxies	659
9.3.2 Spiral Pattern Formation	661
9.3.3 Momentum Fluid Model and Gravitational Radiation	662
9.3.4 Model for Galactic Dynamics	666
9.3.5 Theory for Galactic Spiral Structure	670

9.4	Solar Surface Eruptions and Sunspots	674
9.4.1	Sun's Surface Fluid Dynamics	674
9.4.2	Theory on Formation of Sunspots and Solar Eruptions	679
9.4.3	Plasma Fuel	687
9.5	Dynamic Theory of Boundary-Layer Separations	688
9.5.1	Phenomena and Problems on Boundary-Layer Separations	688
9.5.2	Basic Boundary-Layer Theory	691
9.5.3	Predictable Conditions and Critical Thresholds	695
9.5.4	Boundary-Layer Separation of Ocean Circulations	701
9.6	Interior Separations and Cyclone Formation Theory	706
9.6.1	Horizontal Heat-Driven Fluid Dynamical Equations	706
9.6.2	Interior Separation Equations	708
9.6.3	U-Flow Theory	710
9.6.4	Tornado and Hurricane Formations	715
Appendix A	719	
A.1	Formulas for Center Manifold Functions	719
A.2	Dynamics of Gradient-Type Systems	725
A.3	Constraint Variation	732
A.3.1	General Orthogonal Decomposition Theorem	732
A.3.2	Variations with Constrained-Infinitesimals	734
References	737	
Index	753	



Chapter 1

Introduction to Dynamic Transitions

The study of phase transitions is an active field with a long history. The interested reader is referred to Prigogine and Lefever (1968), Glansdorff and Prigogine (1971), Nicolis and Prigogine (1977), Onuki (2002), Reichl (1998), Stanley (1971), Chandrasekhar (1981), and Drazin and Reid (1981) for some classical treatments and various historical accounts.

This book aims to provide a comprehensive, unified, and balanced account of both dynamic and topological phase transition theories and their applications to statistical systems, quantum systems, classical and geophysical fluid dynamics, and climate dynamics. The dynamic phase transition theory establishes a dynamic transition principle, Principle 1, following the philosophy of searching for a complete set transition states. We present in this chapter a brief introduction to this dynamic transition theory together with an introduction to modeling and to fundamental issues of dynamic phase transitions motivated by problems in the nonlinear sciences. The application-oriented reader may go directly to the study of the scientific problems starting from Chap. 3 after quickly reading this chapter.

Also, new chapters in this **Second Edition** include Chap. 7: Fundamental Principles of Statistical and Quantum Physics, Chap. 8: Quantum Mechanism of Condensates and High Tc Superconductivity, and Chap. 9: Topological Phase Transitions.

1.1 First Principles and Dynamic Models

1.1.1 Guiding Principles of Theoretical Physics

Physics studies fundamental interactions, motion, and formation of matter in our Universe. The heart of fundamental physics is to seek experimentally verifiable, fundamental laws and principles of Nature. The language to express these laws and principles is mathematics. Namely, physical concepts and theories are transformed

into mathematical models, and the predictions derived from these models can be verified experimentally and conform to the reality.

Modeling is a crucial step to understand a physical phenomena. A good model should be derived based on fundamental first principles, and often presented in the form of differential equations. The first principles are often connected with symmetries, which also dictate specific forms of the mathematical models, represented by differential equations.

In a nutshell, we have the following two guiding first principles of physics (Ma and Wang, 2015a).

Principle 1.1.1 (Guiding Principle of Physics) *The essence of physics includes the following:*

1. *the entire theoretical physics is built upon a few fundamental first principles; and*
2. *the laws of Nature are simple and beautiful with clear physical pictures.*

Historically, Albert Einstein and Paul Dirac were the main proponents of this guiding principle, which has been adopted by the physicists.

Principle 1.1.2 (Guiding Principle of Physics) *All physical systems obey laws and principles of Nature, and possess the following three characteristics:*

1. *For each system, there is a group of functions $u = (u_1, \dots, u_N)$ describing its states, and the laws and principles obeyed by the system can be expressed as:*

$$\text{physical laws} = \text{mathematical equations},$$

with the state functions u as solutions of the mathematical equations;

2. *For each system, there is a functional $F(u)$, which dictates mathematical equations;*
3. *All physical systems obey certain symmetries, which essentially determine the mathematical forms of the functionals F .*

This guiding principle shows that to derive the physical law of a physical system, we need to identify the basic symmetry of the system. For example, one of the most remarkable revolution in sciences is that the Einstein principle of general relativity (PGR), a symmetry principle, dictates the Einstein–Hilbert functional, and consequently dictates the fundamental law of gravity.

Symmetry plays a fundamental role in understanding Nature. In mathematical terms, each symmetry, associated with a particular physical law, consists of three main ingredients (Ma and Wang, 2015a):

- (1) the underlying space,
- (2) the symmetry group, and
- (3) the tensors, describing the objects which possess the symmetry.

One important point to make is that different physical systems enjoy different symmetries. For example, gravitational interaction enjoys the symmetry, stated as the principle of general relativity, which, amazingly, dictates the Einstein–Hilbert

functional for the law of gravity. Also, the other three fundamental interactions obey the gauge and the Lorentz symmetries, which dictate also the corresponding functionals F , also called actions. The Newtonian motion systems obey the Galileo symmetry, and all motion systems satisfy the $SO(n)$ symmetry.

In searching for laws of Nature, one inevitably encounters a system consisting of a number of subsystems, each of which enjoys its own symmetry principle with its own symmetry group. To derive the basic law of the system, we have demonstrated in Ma and Wang (2015a) the following principle of symmetry-breaking (PSB), which is of fundamental importance for deriving physical laws for both fundamental interactions and motion dynamics. We recapitulate this principle as follows.

Principle 1.1.3 (PSB) *Physical systems in different levels obey different laws, which are dictated by corresponding symmetries. For a system coupling different levels of physical laws, part of these symmetries must be broken:*

1. *There are several main symmetries—the $SO(n)$ invariance, the Galileo invariance, the Lorentz invariance, the $SU(N)$ gauge invariance, and the general relativistic invariance—which are mutually independent and dictate in part the physical laws in different levels of Nature; and*
2. *for a system coupling several different physical laws, part of these symmetries must be broken.*

Physically, the functional F is a certain form of energy for the underlying physical system. For a system consisting of a number of subsystems, the above PSB amounts to saying that the subsystems are coupled in the energy level by adding up the energies (functionals) F_i of the subsystems:

$$F = \sum_i F_i, \quad (1.1.1)$$

as the energy is an intensive variable. Then the combined F dictates the physical law for the coupled system. In other words, the essence of physics and unification is coupling:

$$\text{the essence of physics is coupling.} \quad (1.1.2)$$

1.1.2 Theory of Four Fundamental Interactions

There are four fundamental interactions: the gravity, the electromagnetism, the strong, and the weak interactions. Based on the above guiding principles, the fundamental theory of the four interactions must contain the following three basic constituents:

- the symmetries,
- the Lagrangian actions, also called functionals, and
- the field equations.

The relation between these three components is

$$\text{symmetries} \implies \text{unique actions (functionals)} \implies \text{field equations}$$

1. *Symmetries* The symmetry for gravity is the invariance under general coordinate transformations, which is precisely described by the Einstein principle of general relativity (PGR), and the symmetry for the electromagnetism, the weak and the strong interactions is the gauge symmetry, originally proposed by Herman Weyl.
2. *Uniqueness of Actions* The symmetries determine uniquely the actions (functionals): the PGR uniquely determines the Einstein–Hilbert functional, and the gauge symmetry uniquely dictates the Yang–Mills action.
Of course, the uniqueness is derived under the principle that the law of nature must be simple; simplicity implies stability and beauty.
3. *Field Equations by PID* The principle of interaction dynamics (PID) takes variation of the actions subject to generalized energy-momentum conservation constraints. It is the direct consequence of the presence of dark energy and dark matter, is the requirement of the presence of the Higgs field for the weak interaction, and is the consequence of the quark confinement phenomena for the strong interaction. Hence

$$\begin{aligned} \text{PID is the principle for deriving the field equations} \\ \text{for the fundamental interactions.} \end{aligned} \tag{1.1.3}$$

In summary, the fundamental theory of the four interactions is now complete. The symmetry for gravity is different from the gauge symmetry for the electromagnetism, the weak and the strong interactions, leading to different actions and field equations. In essence, the electromagnetic, the weak and the strong interactions are unified by the gauge field theory. The symmetry for gravity was discovered by Einstein and the action by Einstein and Hilbert. For the electromagnetism, the weak and the strong interactions, the gauge symmetry was discovered by Herman Weyl, and the general $SU(N)$ action was introduced by Yang and Mills (1954). We refer the interested readers to Ma and Wang (2015a) for details.

1.1.3 Dynamic Law of Physical Motion

Theoretical physics studies (1) the laws for the four fundamental interactions (gravity, electromagnetism, the weak, and the strong interactions), and (2) the laws for motion dynamics of physical systems.

According to their scales, physical systems are classified into

1. classical mechanical systems (planetary scale),
2. quantum systems (micro level),
3. fluid mechanics systems,
4. astrophysical systems, and

5. statistical systems (relating microscopic properties of individual particles to the macroscopic properties).

The basic laws for describing classical mechanical systems, quantum systems, and astrophysical systems are given in terms of Newtonian laws, the principle of Lagrangian dynamics, and the principle of Hamiltonian dynamics.

By a careful examination of basic motion laws of different physical systems and by the mathematical study, we discover a general dynamic law (Ma and Wang, 2017a), suitable for all physical motion systems in (1)–(5) mentioned above.

To state this general dynamic law, we start with the definition of constraint variation. Let H, H_1 be two Hilbert spaces, and

$$L : H_1 \rightarrow H, \quad L^* : H \rightarrow H_1^* \quad (1.1.4)$$

be a pair of dual linear bounded operators. Consider a functional defined on H :

$$F : H \rightarrow \mathbb{R}^1. \quad (1.1.5)$$

Let

$$\mathcal{N}^* = \{v \in H \mid L^*v = 0\} \neq \{0\}. \quad (1.1.6)$$

Definition 1.1.4 Let (1.1.6) hold.

1. For any $u \in H$, the derivative operator of F at u with L^* -constraint, denoted by $\delta_{L^*}F(u)$, is defined as follows

$$\langle \delta_{L^*}F(u), v \rangle_H = \frac{d}{dt} \Big|_{t=0} F(u + tv) \quad \forall v \in \mathcal{N}^*. \quad (1.1.7)$$

2. The derivative operator of F at u with L -constraint, denoted by $\delta_L F(u)$, is defined as follows

$$\langle \delta_L F(u), \varphi \rangle_{H_1} = \frac{d}{dt} \Big|_{t=0} F(u + tL\varphi) \quad \forall \varphi \in H_1. \quad (1.1.8)$$

Then using the general orthogonal decomposition theorem, Theorem A.3.1, we show that the above variations with constraint infinitesimals take the following form:

$$\delta_{L^*}F(u) = \delta F(u) + Lp, \quad (1.1.9)$$

$$\delta_L F(u) = L^* \delta F(u), \quad (1.1.10)$$

for some function p , which plays a similar role as the pressure in incompressible fluid flows. Here $\delta F(u)$ is the usual derivative operator. We are now ready to state the dynamical law of physical motion.

Dynamical Law 1.1.5 There are two types of physical motion systems: dissipative systems and conservation systems. For each physical motion system, there are a set of state functions $u = (u_1, \dots, u_N)$, a functional $F(u)$, and an operator \mathcal{L} , which is either $\mathcal{L} = L$ or $\mathcal{L} = L^*$ for some differential operator L , such that the following statements hold true:

1. $-\delta_{\mathcal{L}} F(u)$ is the driving force of the system;
2. F is $SO(n)$ ($n = 2, 3$) invariant;
3. for an isolated system, the dynamic equation can be expressed in the form

$$\frac{du}{dt} = -A\delta_{\mathcal{L}} F(u), \quad (1.1.11)$$

where A is the coefficient matrix, A is symmetric and positive definite if and only if the system is a dissipative system, and A is anti-symmetry if and only if the system is a conservation system;

4. for a system coupling different subsystems, the motion equations of (1.1.11) become

$$\frac{du}{dt} = -A\delta_{\mathcal{L}} F(u) + B(u), \quad (1.1.12)$$

where $B(u)$ represents the symmetry-breaking.

A few remarks are in-order.

First, in (1.1.11) and (1.1.12), the derivative d/dt is physical, namely

$$\begin{aligned} \frac{d}{dt} &= \frac{\partial}{\partial t} + ig\mathbf{A}_0 && \text{for gauge fields ,} \\ \frac{d}{dt} &= \frac{\partial}{\partial t} + (\mathbf{u} \cdot \nabla) && \text{for fluid fields.} \end{aligned}$$

Second, the operator \mathcal{L} ($= L$ or L^*) in physical systems is an operator. When $\mathcal{L} = id$ is the identity operator, $\delta_{\mathcal{L}}$ is the usual derivative operator. Typical form of the operator L is the gradient operator with L^* being the corresponding divergence-operator.

Third, $-\delta_{\mathcal{L}} F(u)$ represents the generalized force, and (1.1.11) is the generalized Newtonian Law. For example, consider the classical Newtonian Second Law:

$$\frac{d^2x}{dt^2} = -\nabla\varphi, \quad (1.1.13)$$

where φ is an interaction potential. Equation (1.1.13) is equivalent to

$$\frac{dx}{dt} = y, \quad \frac{dy}{dt} = -\nabla\varphi. \quad (1.1.14)$$

Now let

$$\Phi = \frac{y^2}{2} + \varphi, \quad u = (x, y)^T.$$

Let

$$J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$

be the symplectic matrix. Then we have

$$J\nabla\Phi = (-y, \nabla\varphi)^T,$$

and the Newtonian Second Law can be rewritten in the form of (1.1.11) as follows:

$$\frac{du}{dt} = -J\nabla\Phi. \quad (1.1.15)$$

Here there is no constraint; namely, the operator $\mathcal{L} = L$ is the identity operator.

Fourth, the $B(u)$ term in (1.1.12) represents the symmetry-breaking. For example, consider the damped wave equation

$$u_{tt} = \Delta u - ku_t + f(u).$$

Let $v = u_t$ and $\psi = (u, v)^T$, then the wave equation can be written as

$$\begin{aligned} \frac{d}{dt}\psi &= -J\delta F(\psi) + B\psi, \\ F(\psi) &= \int \left[\frac{1}{2}v^2 + |\nabla u|^2 + g(u) \right], \\ B\psi &= \begin{pmatrix} 0 \\ -kv \end{pmatrix}, \end{aligned}$$

where $g(u)$ is the primitive function of f , i.e., $g' = f$. Here the functional F inherits the symmetry of the wave equation *without damping term ku_t* . But the damping term cannot be included in the functional F , and breaks the symmetry of the non-damped wave equation. As mentioned earlier, symmetry-breaking is the essence for modeling a physical system coupling subsystems obeying different symmetries.

Hereafter we examine classical individual dynamical laws of motion.

(1) Newtonian Law

A group of bodies with masses (m_1, \dots, m_N) motion in a potential field $V(x)$, the motion equations are

$$\frac{dp_i}{dt} = -m_i \delta_i V(x) \quad (1 \leq i \leq N), \quad (1.1.16)$$

where $\delta_i = \partial/\partial x_i$. In (1.1.15), we have shown that (1.1.16) has the form of (1.1.11).

(2) Fluid Dynamics

We consider here two examples—the compressible Navier–Stokes equations and the Boussinesq equations. We start with the compressible Navier–Stokes equations:

$$\begin{aligned}\frac{\partial u}{\partial t} + (u \cdot \nabla)u &= \frac{1}{\rho} [\mu \Delta u - \nabla p + f], \\ \frac{\partial \rho}{\partial t} &= -\operatorname{div}(\rho u) \\ u|_{\partial\Omega} &= 0.\end{aligned}\tag{1.1.17}$$

Let the constraint operator $L = -\nabla$ be the gradient operator, with dual operator $L^* = \operatorname{div}$. Also, let

$$\Phi(u, \rho) = \int_{\Omega} \left[\frac{\mu}{2} |\nabla u|^2 - fu + \frac{1}{2} \rho^2 \operatorname{div} u \right] dx.$$

Then the equations of (1.1.17) are written as

$$\begin{aligned}\frac{du}{dt} &= -\frac{1}{\rho} \frac{\delta L^*}{\delta u} \Phi(u, \rho), \\ \frac{d\rho}{dt} &= -\frac{\delta}{\delta \rho} \Phi(u, \rho),\end{aligned}\tag{1.1.18}$$

which is in the form of (1.1.11) with coefficient matrix $A = \operatorname{diag}(1/\rho, 1)$. Also, the pressure is given by

$$p = p_0 - \lambda \operatorname{div} u - \frac{1}{2} \rho^2.$$

Here

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x_i} \frac{dx_i}{dt} = \frac{\partial u}{\partial t} + (u \cdot \nabla)u, \quad \frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + u \cdot \nabla \rho.$$

The Boussinesq equations in classical fluid dynamics are written as

$$\begin{aligned}\frac{\partial u}{\partial t} + (u \cdot \nabla)u - \rho_0^{-1} [\nu \Delta u + \nabla p] &= -g k \rho(T), \\ \frac{\partial T}{\partial t} + (u \cdot \nabla)T - \kappa \Delta T &= 0, \\ \operatorname{div} u &= 0,\end{aligned}\tag{1.1.19}$$

where ν, κ, g are constants, $u = (u_1, u_2, u_3)$ is the velocity field, p is the pressure function, T is the temperature function, \bar{T}_0 is a constant representing the lower surface temperature at $x_3 = 0$, and $k = (0, 0, 1)$ is the unit vector in the x_3 -direction.

As before, we take the linear constraint operator $L = -\nabla$ with dual $L^* = \operatorname{div}$. Let

$$\Phi(u, T) = \int_{\Omega} \left[\frac{\mu}{2} |\nabla u|^2 + \frac{\kappa}{2} |\nabla T|^2 \right] dx.$$

Then the Boussinesq equations of (1.1.19) are written as

$$\begin{aligned}\frac{du}{dt} &= -\frac{1}{\rho_0} \frac{\delta L^*}{\delta u} \Phi(u, T) + B(u, T), \\ \frac{dT}{dt} &= -\frac{\delta}{\delta T} \Phi(u, T),\end{aligned}\tag{1.1.20}$$

which is in the form of (1.1.11) with coefficient matrix $A = \text{diag}(1/\rho_0, 1)$. The term $B(u, T)$ is given by

$$B(u, T) = g\rho_0 k\rho(T),$$

which breaks the symmetry of the functional $\Phi(u, T)$, caused by the coupling between momentum equations and the temperature equation. Also, the incompressibility condition $\text{div } u = 0$ is built into the construction of the basic function space, which we omit the details.

(3) Hamiltonian Dynamics

The principle of Hamiltonian dynamics (PHD) is a universal principle, which governs all energy conservative systems, including the astronomic mechanics, wave motion equations, quantum mechanics, and the Maxwell equations.

PHD amounts to saying that for a conservative system, there are two sets of conjugate functions $u = (u_1, \dots, u_N)$, $v = (v_1, \dots, v_N)$ and a Hamiltonian energy $H = H(u, v)$, such that u and v satisfy the equations; see among others (Ma and Wang, 2015a):

$$\begin{aligned}\frac{\partial u}{\partial t} &= k \frac{\delta}{\delta v} H(u, v), \\ \frac{\partial v}{\partial t} &= -k \frac{\delta}{\delta u} H(u, v),\end{aligned}$$

which can be expressed in form of (1.1.11) as

$$\frac{dv}{dt} = -A\delta H(u, v), \quad A = k \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}.\tag{1.1.21}$$

(4) Lagrangian Dynamics

The principle of Lagrangian dynamics (PLD) is also a universal principle in physics, which governs the physical motion systems such as Newtonian mechanics, quantum physical system, elastic waves, classical electrodynamics, etc.; see among others (Landau and Lifshitz, 1975; Ma and Wang, 2015a).

PLD amounts to saying that for an isolated motion system, there are state functions $u = (u_1, \dots, u_N)$ and a functional $F(u)$ called the Lagrangian action, such that u satisfies

$$\delta F(u) = 0,\tag{1.1.22}$$

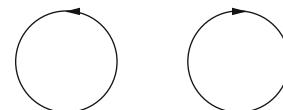
which can be equivalently referred to the Hamiltonian system (1.1.21) with conjugates u and $v = u_t$.

1.1.4 Rayleigh–Bénard Convection

The study of transitions in classical and geophysical fluid dynamics and in climate dynamics has huge economical and societal impacts. Classical fluid dynamics has a long history, and the models for typical fluid flows have been developed based on first principles. The atmosphere and ocean around the earth are rotating geophysical fluids, which are also two important components of the climate system. The phenomena of the atmosphere and ocean are extremely rich in their organization and complexity, with scales ranging from meters to thousands of kilometers. The modeling of important geophysical fluid flows is still an important part of the current and future studies.

To demonstrate the basic ideas on the modeling and the mathematical settings, we consider the classical Rayleigh–Bénard convection problem. Convection is the well-known phenomena of fluid motion induced by buoyancy when a fluid is heated from below. It is familiar as the driving force in atmospheric and oceanic phenomena. The Rayleigh–Bénard convection problem was originated in the famous experiments conducted by Bénard in 1900. Bénard investigated a liquid in a horizontal dish with its diameter much larger than its height, keeping the upper surface free touch with air and heating from below. He noticed that as the lower surface temperature T is less than some critical temperature T_c the liquid is in a static state, and as T is greater than T_c the liquid suddenly breaks into a rather regular cellular pattern of hexagonal convection cells. The cells have the convection roll structure in their vertical section as illustrated in Fig. 1.1.

Fig. 1.1 The vertical section of thermal convection



The modeling of the Bénard problem includes the following main ingredients.

FIRST, the energy conservation law is needed to take the effect of the heat into consideration. The basic physical law is the first law of thermodynamics, which is an expression of the more universal physical law of the conservation of energy:

$$\frac{\partial T}{\partial t} + (u \cdot \nabla)T - \kappa \Delta T = 0. \quad (1.1.23)$$

Here as we discussed before, the term $-\kappa\Delta T$ is due to the energy diffusion, and κ is the diffusion coefficient.

SECOND, an equation of state is needed to provide a relation between state variables: the pressure p , the temperature T , and the density ρ . Boyle's Law was perhaps the first expression of an equation of state. For water as well as for many other fluids, an empirical relationship derived using the Taylor expansion near a basic state is often used. In particular, we shall use the following linear approximation of the equation of state in our discussions for the Rayleigh–Bénard convection:

$$\rho = \rho_0(1 - \alpha_T(T - \bar{T}_0)), \quad (1.1.24)$$

where \bar{T}_0 is the reference temperature and α_T is the expansion coefficient.

THIRD, the external forcing here is given by $-\rho g k$, where g is the gravitational constant, and $k = (0, 0, 1)$ is the vertical unit vector. We obtain the following system of partial differential equations governing the motion and states of the flow:

$$\begin{aligned} \rho \left[\frac{\partial u}{\partial t} + (u \cdot \nabla)u \right] &= \mu \Delta u - \nabla p - \rho g k, \\ \frac{\partial T}{\partial t} + (u \cdot \nabla)T - \kappa \Delta T &= 0, \\ \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho u) &= 0, \\ \rho &= \rho_0(1 - \alpha_T(T - \bar{T}_0)). \end{aligned}$$

As mentioned before, modeling of a given physical problem consists of the mathematical representations of the physical laws, and approximations are made based on the specific characteristics of the underlying physics problem. In this case, we have already made a simplification on the equation of state. Another simplification is often made using the classical Boussinesq assumption, which amounts to saying that density differences are sufficiently small to be neglected, except where they appear in terms multiplied by g , the acceleration due to gravity, and in the equation of state. With this assumption, the above system of equations is reduced to the following equations; see among others (Rayleigh, 1916; Drazin and Reid, 1981; Chandrasekhar, 1981):

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla)u - \nu \Delta u + \rho_0^{-1} \nabla p &= -gk[1 - \alpha(T - \bar{T}_0)], \\ \frac{\partial T}{\partial t} + (u \cdot \nabla)T - \kappa \Delta T &= 0, \\ \operatorname{div}u &= 0, \end{aligned} \quad (1.1.25)$$

where $\nu = \mu/\rho_0$. This system of equations is often called the *Boussinesq equations*.

The spatial domain is usually given by $\Omega = D \times (0, h)$, where $D \subset \mathbb{R}^2$ is a bounded domain in \mathbb{R}^2 and h is the vertical height of the container. The system is often supplemented with the following boundary conditions:

$$\begin{aligned} u_3 = 0, \quad & \frac{\partial(u_1, u_2)}{\partial x_3} = 0, T = \bar{T}_0 \quad \text{at } x_3 = 0, \\ u_3 = 0, \quad & \frac{\partial(u_1, u_2)}{\partial x_3} = 0, T = \bar{T}_1 \quad \text{at } x_3 = h, \\ u_n = 0, \quad & \frac{\partial u_\tau}{\partial n} = 0, \quad \frac{\partial T}{\partial n} = 0 \quad \text{on } \partial D \times (0, h), \end{aligned} \quad (1.1.26)$$

where n and τ are the unit normal and tangent vector on $\partial D \times [0, h]$ respectively, and $u_n = u \cdot n, u_\tau = u \cdot \tau$.

1.1.5 Mathematical Formulation of Physical Problems

As discussed in the previous sections, many problems in sciences are modeled by partial differential equations. With these models in our disposal, one basic level question is the well-posedness of the model. The well-posedness question consists of the existence, the uniqueness, regularity of solutions, and the dependence of the solutions on physical parameters. There has been a long history of studies in this direction in the last century. With the well-posedness, we can put the model into the perspective of a dynamical system, both finite and infinite dimensional, as follows:

$$\frac{du}{dt} = L_\lambda u + G(u, \lambda), \quad u(0) = \varphi, \quad (1.1.27)$$

where $u : [0, \infty) \rightarrow X$ is the unknown function, and X is the phase space equipped with norm $\|\cdot\|$. Often times, the system depends on a control parameter $\lambda \in \mathbb{R}^N$ ($N \geq 1$). In general, $L_\lambda : X_1 \rightarrow X$ is a linear operator, and $G(\cdot, \lambda) : X_1 \rightarrow X$ is a nonlinear operator of higher order of u :

$$G(u, \lambda) = o(\|u\|) \quad (\text{higher order of } u), \quad (1.1.28)$$

which amounts to saying that $u = 0$ is always a steady-state solution of (1.1.27), and represents the basic solution of the underlying physical system. For a system with nonzero basic state, the function u here always represents the *deviation* from the basic solution, as demonstrated in the example below.

To demonstrate the basic ideas, consider the Boussinesq equations (1.1.25) with the boundary conditions (1.1.26). We proceed as follows.

Basic State and Its Deviation Equations The problem possesses a basic solution given by

$$\bar{u} = 0, \quad \bar{T} = \bar{T}_0 - \frac{\bar{T}_0 - \bar{T}_1}{h} x_3, \quad \bar{p} = p_0 - \rho_0 g x_3 - \frac{\alpha_T (\bar{T}_0 - \bar{T}_1)}{2h} x_3^2.$$

Consider the deviation (u', T', p') from the basic solution $(0, \bar{T}, \bar{p})$ defined by:

$$u = u', \quad T = \bar{T} + T', \quad p = \bar{p} + p'.$$

Then the resulting equations of the deviations are given by (drop primes for simplicity):

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla)u - \nu \Delta u + \rho_0^{-1} \nabla p &= g k \alpha_T T, \\ \frac{\partial T}{\partial t} + (u \cdot \nabla)T - \frac{\bar{T}_0 - \bar{T}_1}{h} u_3 - \kappa \Delta T &= 0, \\ \operatorname{div} u &= 0. \end{aligned} \tag{1.1.29}$$

Non-dimensionalization To study the dynamics and its physical implications, we need to non-dimensionalize the problem. For this purpose, we set

$$\begin{aligned} (x, t) &= (hx', h^2 t' / \kappa), & (u, T, p) &= (\kappa u' / h, \beta h(T' / \sqrt{R}), \rho_0 \kappa^2 p' / h^2), \\ \Pr &= \nu / \kappa, & R &= \frac{g \alpha \beta}{\kappa \nu} h^4. \end{aligned}$$

Once again, omitting the primes, the equations (1.1.29) can be rewritten as

$$\begin{aligned} \frac{1}{\Pr} \left[\frac{\partial u}{\partial t} + (u \cdot \nabla)u + \nabla p \right] - \Delta u - \sqrt{R} T k &= 0, \\ \frac{\partial T}{\partial t} + (u \cdot \nabla)T - \sqrt{R} u_3 - \Delta T &= 0, \\ \operatorname{div} u &= 0. \end{aligned} \tag{1.1.30}$$

The spatial domain now becomes, still denoted by Ω , $\Omega = D \times (0, 1)$, and the boundary conditions are

$$\begin{aligned} u_3 &= 0, \quad \frac{\partial(u_1, u_2)}{\partial x_3} = 0, T = 0 \quad \text{at } x_3 = 0, 1, \\ u_n &= 0, \quad \frac{\partial u_\tau}{\partial n} = 0, \quad \frac{\partial T}{\partial n} = 0 \quad \text{on } \partial D \times (0, 1). \end{aligned} \tag{1.1.31}$$

Mathematical Set-Up For the problem (1.1.30) with (1.1.31) we define the function spaces as

$$\begin{aligned} H &= \{(u, T) \in L^2(\Omega)^4 \mid \operatorname{div} u = 0, u_n = 0 \text{ on } \partial\Omega\}, \\ H_1 &= \{(u, T) \in H^2(\Omega)^4 \mid \operatorname{div} u = 0, (u, T) \text{ satisfies (1.1.31)}\}. \end{aligned} \tag{1.1.32}$$

Also, let $L_\lambda = -A + B_\lambda : H_1 \rightarrow H$, and $G : H_1 \rightarrow H$ be defined by

$$\begin{aligned} A\phi &= (-\Pr P(\Delta u), -\Delta T), \\ B_\lambda \phi &= \lambda(\Pr P[Tk], u_3), \\ G(\phi) &= (-P[(u \cdot \nabla)u], -(u \cdot \nabla)T), \end{aligned} \tag{1.1.33}$$

for any $\phi = (u, T) \in H_1$. Here $\lambda = \sqrt{R}$, and

$$P : L^2(\Omega)^3 \rightarrow \{u \in L^2(\Omega)^3 \mid \operatorname{div} u = 0, \quad u \cdot n|_{\partial\Omega} = 0\}$$

is the Leray projection. Then it is easy to see that the Boussinesq equations (1.1.30) with (1.1.31) can be put into the operator form (1.1.27).

1.2 Introduction to Dynamic Transition Theory

1.2.1 Motivations and Key Philosophy

As mentioned in the Introduction, for many problems in sciences, we need to understand the transitions from one state to another, and the stability/robustness of the new states. For this purpose, a dynamic transition theory for nonlinear partial differential equations is developed recently by the authors, and is applied to both equilibrium and nonequilibrium phase transitions in nonlinear sciences. This theory is based on the Key Philosophy 2 that we search for the full set of transition states representing the physical reality after the transition. The set of transition states is often represented by a local attractor.

With this philosophy and under the close link to the physics, the dynamic transition theory is developed, and in return the theory is applied to physical problems yielding new insights and predictions to the underlying physical phenomena. An important part of the study presented in this book is to identify and to classify the local attractors after the transition.

1.2.2 Principle of Exchange of Stability

It is clear that linear theory plays a crucial role in studying nonlinear dynamic transitions. Classically, one starts with the following normal mode solution of the problem (1.1.27):

$$u = e^{\beta t} \phi, \tag{1.2.1}$$

where ϕ is time-independent. Inserting this normal form solution into (1.1.27) and dropping the nonlinear term, we find that

$$L_\lambda \phi = \beta \phi. \tag{1.2.2}$$

It is then easy to derive the following general statements in many nonlinear problems of the form (1.1.27):

- (1) If all eigenvalues of the eigenvalue problem (1.2.2) are on the left-hand side of the imaginary axis in the complex plane, the basic solution $u = 0$ is linearly stable;
- (2) If an eigenvalue of (1.2.2) is on the right-hand side of the imaginary axis, the basic solution is linearly unstable; and

- (3) If some eigenvalues of (1.2.2) cross the imaginary axis from the left to the right as the control parameter λ crosses a critical value λ_0 , the stable basic solution losses its stability and a dynamic transition likely occurs.

In the dynamic transition theory and its applications presented in this book, we always take a special path in the control parameter space $\lambda \in \mathbb{R}^N$. Namely, when we discuss dynamic transitions, it is always understood that the path we take can often be represented as

$$\lambda(s) = (\lambda_1(s), \dots, \lambda_N(s)) \quad \text{for } s \in \mathbb{R}. \quad (1.2.3)$$

However, for brevity, we misuse the notations and simply write $\lambda = (\lambda_1, \dots, \lambda_N)$.

Physically, the eigenvalue crossing condition in item (3) here is called the *principle of exchange of stability (PES)*; see among others (Chandrasekhar, 1981; Drazin and Reid, 1981).

Definition 1.2.1 Let $\{\beta_j(\lambda) \in \mathbb{C} \mid j \in \mathbb{N}\}$ be the eigenvalues (counting multiplicity) of L_λ .¹ These eigenvalues satisfy the principle of exchange stability at λ_0 with multiplicity m , if

$$\operatorname{Re} \beta_i(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_0, \\ = 0 & \text{if } \lambda = \lambda_0, \\ > 0 & \text{if } \lambda > \lambda_0, \end{cases} \quad \forall 1 \leq i \leq m, \quad (1.2.4)$$

$$\operatorname{Re} \beta_j(\lambda_0) < 0 \quad \forall j \geq m + 1. \quad (1.2.5)$$

It is worth noting that if (1.1.27) represents a dissipative system, the linear operator $L_\lambda : X_1 \rightarrow X$ is often a sectorial operator and is a completely continuous field.² In fact, in many applications, the linear operator is often an elliptic operator such as a Laplacian plus low-order differential terms, which is always a sectorial operator. In this case, there are countable number of eigenvalues for (1.2.2), and the above PES makes sense.

Finally, the dynamic transition theory in the next chapter is to identify and to classify the dynamical transitions for the nonlinear problem (1.1.27). Applications to various problems in science and engineering are then given in the subsequent chapters.

1.2.3 Equation of Critical Parameters

The PES and Statement (3) above suggest that the critical value $\lambda_0 = (\lambda_1^0, \dots, \lambda_N^0)$ of the control parameter $\lambda = (\lambda_1, \dots, \lambda_N)$ should satisfy the following equation:

$$\operatorname{Re} \beta_1(\lambda_1, \dots, \lambda_N) = 0, \quad (1.2.6)$$

¹For brevity, we use here $\lambda \in \mathbb{R}$.

²See the beginning of Chap. 2 for the precise definition, and Ma and Wang (2005b) for more detailed discussions.

which is called the *equation of critical parameters* in physical literatures. Here β_1 is the first eigenvalue.

Equation (1.2.6) defines a hypersurface in the control parameter space \mathbb{R}^N , which divides \mathbb{R}^N into several disjoint regions. The graph of (1.2.6) in \mathbb{R}^N is called phase diagram for (1.1.27). As the control parameter crosses the hypersurfaces from one region to another, the type of transition is derived by the nonlinear dynamic transition theory.

1.2.4 Classifications of Dynamic Transitions

The concept of phase transition originates from the statistical physics and thermodynamics. In physics, “phase” is usually used as a state of matter. In this book, we use phase to stand for a (stable) state in the systems of nonlinear sciences including physics, chemistry, biology, ecology, economics, fluid dynamics and geophysical fluid dynamics, etc. Hence, the content of phase transition has been endowed with more general significance. In fact, the phase transition dynamics introduced in this book treats a wide variety of topics involving the universal critical phenomena of state changes in Nature in a unified mathematical viewpoint and manner.

Based on statistical physics, the phase transitions of nonlinear dissipative systems can be divided into two types: the equilibrium phase transition and the nonequilibrium phase transition. Namely, if the system (1.1.27) possesses gradient-type structure, then the phase transitions are called equilibrium phase transition; otherwise they are called the nonequilibrium phase transitions.

From the physical point of view, the equilibrium phase transition means that the thermodynamic systems undergo a transition at a critical point from an equilibrium state to another, maintaining minimal free energy. We shall see later that the equilibrium phase transitions appear extensively in Statistical Physics, Solid-state Physics, Materials Science, Thermodynamics, and Elastic mechanics. The nonequilibrium phase transitions appear mainly in chemical reaction dynamics, biology, ecology, economics, fluid dynamics, and geophysical fluid dynamics. In addition, these two types of phase transitions have remarkable differences in their dynamical properties. In particular, an equilibrium phase transition does not lead to chaos, while a nonequilibrium phase transition may.

Equilibrium phase transitions are usually classified by the Ehrenfest classification scheme, which groups phase transitions based on the degree of non-analyticity involved; see among many others (Reichl, 1998; Stanley, 1971) for detailed discussions.

One important ingredient of dynamic transition theory presented in this book is to establish a following dynamic classification scheme for dynamic transitions; see also Definition 2.1.4. The dynamic classification scheme is applicable to both the equilibrium and nonequilibrium phase transitions, which was first presented in Ma and Wang (2007b, 2009b,c, 2008a).

Definition 1.2.2 (Dynamic Classification of Phase Transition) Let $\lambda_0 \in \mathbb{R}$ be a critical point of (1.1.27), and (1.1.27) always undergoes a transition from a basic state Σ_λ^1 to Σ_λ^2 . The transitions can be classified into three types: Type-I, Type-II, and Type-III, as λ crosses λ_0 , depending on their dynamic properties.

Mathematically, these three types of transitions are also respectively called continuous, jump, and mixed transitions, as shown in Figs. 1.2, 1.3, and 1.4; we refer interested readers to Chap. 2 for the precise definition, and to Chaps. 3–5 for concrete examples.

Fig. 1.2 Schematic of Type-I transition: The transition states are represented by a local attractor Σ_λ , which attracts a neighborhood of the basic solution

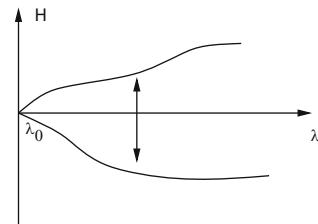


Fig. 1.3 Schematic of Type-II transition: The transition states are represented by some local attractors **away** from the basic state at the critical λ_0 . In particular, the system undergoes a drastic change

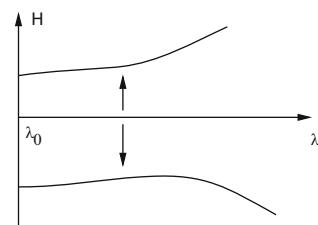
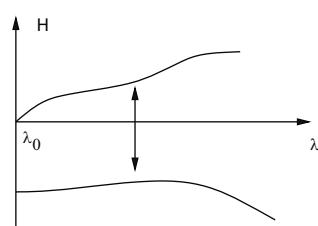


Fig. 1.4 Schematic of Type-III transition: The transition states are represented by two local attractors, with one as in a Type-II transition, and the other as in a Type-I transition



Among other features, the dynamic classification scheme provides a classification scheme for both equilibrium and nonequilibrium phase transitions, includes the dynamic behavior of the system, and offers clear information on the transition types in the classical sense for equilibrium transition problems. In equilibrium phase transition problems studied in Chap. 3, Type-I transitions correspond to second or higher-order transitions in terms of the Ehrenfest classification scheme, and Type-II and III transitions correspond to first-order transitions, with which the latent heat, metastable states, and hysteresis occur. These phenomena will be explored in detail in Chap. 3, leading to some physical predictions. Hence the dynamic classification scheme provides not only the type of transitions, but also the dynamic properties of the system.

1.2.5 Structure and Characterization of Dynamic Transitions

With the key Philosophy 2, the linear theory, and the dynamic classification scheme presented earlier, the main objective now becomes the identification and classification of dynamic transitions and the transition states. Concrete methods and ideas are developed in detail in Chap. 2, and are applied to a wide range of science problems in Chaps. 3–6. Here we resent a few general remarks.

FIRST, the dynamic classification scheme is universal for dissipative dynamic systems, and is applicable for both equilibrium and nonequilibrium phase transitions. A Type-I transition is *essentially* determined by the attractor bifurcation theory (Ma and Wang, 2005b), although as we shall see in the next chapter that there exist Type-I transitions with no bifurcated attractors. Given a dissipative dynamic system, Type-II and Type-III transitions always give rise to a singular separation such as the saddle-node bifurcation, and the existence of metastable states.

SECOND, the dynamic transition theory characterizes dynamic transitions both near and far away from the basic equilibrium state. This global nature of the theory is often derived by combining the detailed local characterizations as given in Chap. 2, with the global properties of the problem such as the existence of global attractor. For example, a Type-II transition leads to the transition states *far* from the basic solution, and a Type-III transition gives rise to two types of transition states: one near and one far away from the basic solution. The singular separation is another global feature of the Type-II and Type-III transitions.

THIRD, one crucial component of the dynamic transition theory is the center manifold reduction; see Appendix A.1 for details. Under the PES assumption (1.2.4) and (1.2.5), by reducing (1.1.27) to the center manifold, we know that the type of transitions for (1.1.27) at $(0, \lambda_0)$ is completely dictated by its reduced equation near $\lambda = \lambda_0$, which can be expressed as:

$$\frac{dx}{dt} = J_\lambda x + PG(x + \Phi(x, \lambda), \lambda) \quad \text{for } x \in \mathbb{R}^m, \quad (1.2.7)$$

where J_λ is the $m \times m$ order Jordan matrix corresponding to the eigenvalues given by (2.1.4), $\Phi(x, \lambda)$ is the center manifold function of (2.1.1) near λ_0 , $P : X \rightarrow E_\lambda$ is the canonical projection, and

$$E_\lambda = \bigcup_{1 \leq i \leq m} \bigcup_{k \in \mathbb{N}} \{u \in X_1 \mid (L_\lambda - \beta_i(\lambda))^k u = 0\}$$

is the eigenspace of L_λ corresponding to the first m eigenvalues.

In many applications, it is crucial to calculate the center manifold function Φ . For this purpose, a systematic approach is given in Appendix A.1. This approach provides useful formulas, which are used in the applications presented in remaining chapters of this book. For example, if $G(u, \lambda)$ has the Taylor expansion

$$G(u, \lambda) = G_k(u, \lambda) + o(\|u\|^k) \quad \text{for some } k \geq 2, \quad (1.2.8)$$

where $G_k(u, \lambda)$ is a k -multilinear operator, and J_λ is diagonal near $\lambda = \lambda_0$, then by (A.1.14), we have

$$\begin{aligned}\Phi(x, \lambda) &= \sum_{j=m+1}^{\infty} \Phi_j(x, \lambda) e_j, \\ \Phi_j(x, \lambda) &= \frac{-1}{\beta_j(\lambda)(e_j, e_j^*)} (G_k(x, \lambda), e_j^*) + o(k),\end{aligned}\tag{1.2.9}$$

where $x = \sum_{i=1}^m x_i e_i$, and $\beta(\lambda) = (\beta_1(\lambda), \dots, \beta_m(\lambda))$ is as in (1.2.4). Here e_j and e_j^* ($1 \leq j \leq m$) are the eigenvectors of L_λ and L_λ^* respectively corresponding to the eigenvalues $\beta_j(\lambda)$ as in (1.2.4). Here (\cdot, \cdot) stands for the dual product between the function space X and its dual X^* ; see also (1.1.27).

1.2.6 General Features of Dynamic Transitions

In a nutshell, here are some general features of dynamic transitions for dissipative systems.

First, we derive in this book a general principle, Principle 1, that dynamic transitions of all dissipative systems can be classified into three categories: continuous, catastrophic, and random.

Second, when a thermodynamic system undergoes a phase transition at some critical temperature and pressure, it will always reach a new equilibrium state where its thermodynamic potential reaches its minimum. From the viewpoint of minimal potential energy, equilibrium phase transitions are the changes of minimal value point of the energy functional at some critical point λ_0 of the system control parameter λ .

Third, in 1960s, B. B. Belosov and A. M. Zhabotinsky found a class of chemical reactions, called the Belosov-Zhabotinsky reactions, which lie in non-thermodynamic states. In their study for the Belosov-Zhabotinsky reactions, Prigogine and his colleague found that the phase transitions for thermodynamic equilibrium states are universal for nonlinear dissipative systems. Prigogine's self-organization theory suggests that a stable state far from equilibrium needs to be maintained by energy input from the exterior.

Fourth, mathematically speaking, both equilibrium and nonequilibrium phase transitions can be put into the perspective of a dynamical system with dissipative structure, which we often call a dissipative system. The dissipation here is associated with dissipation of the generalized energy. The global dynamical behavior of a dissipative system can be described by the existence of an absorbing set and the global attractor.

The dynamic transition theory developed in this book is aimed to provide a systematic approach for studying the phase transitions and critical phenomena, and some general dynamic features are summarized as follows:

Physical States and Universality of Dynamic Transitions A dissipative system always lives in a stable state, which is often represented by a local attractor of the system, and dynamic transitions of a dissipative system is a universal phenomena in nature.

Generalized Energy In each stable state, there is some energy dissipation. Therefore a dissipative system needs energy supply in the generalized sense.

Each nonlinear dissipative system in nature has only finite number of stable states, denoted by $\Sigma_1^\lambda, \dots, \Sigma_n^\lambda$. The stability of each state Σ_i^λ is sustained in an energy range

$$\Sigma_i^\lambda : \lambda_i^1 < \lambda < \lambda_i^2 \quad \text{for } 1 \leq i \leq n,$$

where λ represents the generalized supplied energy. Namely, only when the supplied generalized energy λ from the external is maintained in the range $\lambda_i^1 < \lambda < \lambda_i^2$, the system can sustain the state Σ_i^λ ; otherwise the system will undergo a transition to another state.

The lowest energy state Σ_1^λ is a basic state of the system, and its energy range is

$$\Sigma_1^\lambda : \lambda_1^1 \leq \lambda < \lambda_1^2.$$

Namely, Σ_1^λ needs the minimum energy to sustain.

Metastable States It is possible that several stable states coexist at a certain energy level. We call them metastable states, and dynamic transitions among them can occur due to fluctuations and uncertainty.

Chaotic State A chaotic state of a system may be defined as an attractor $\mathcal{A} \subset U$, which attracts its open neighborhood U , and contains no minimal attractors of singular elements (singular points and periodic orbits). It is clear then from this definition that a gradient type flow does not have a chaotic state.

If a system has a chaotic state, then the chaotic state is Σ_n^λ :

$$\Sigma_n^\lambda : \lambda_n^1 < \lambda < \infty,$$

which needs the maximal energy to sustain. Moreover, once a chaotic state Σ_n^λ appears, there will be no new ordered states occurring for any $\lambda > \lambda_n^1$, unless the generalized energy drops its level below λ_n^1 .

In general, if (1.1.27) undergoes successive transitions at $\lambda_0 < \dots < \lambda_k$ ($k \geq 1$), such that the global attractor \mathcal{A}_λ of (1.1.27) satisfies that

$$\mathcal{A}_\lambda = \{0\} \quad \text{for } \lambda \leq \lambda_0,$$

and \mathcal{A}_λ has no minimal attractors of singular elements on $\lambda > \lambda_k$, then with the successive transitions, the system undergoes a transition from ordered states to a chaotic state.

1.3 Examples of Typical Phase Transition Problems

Hereafter we shall present a few typical phase transition problems addressed in this book, demonstrating the need and advantage of the dynamical transition theory. The main focus is on physical implications and predictions derived from the dynamic transition theory, and the readers are referred to later chapters for further details.

1.3.1 Rayleigh–Bénard Convection

As mentioned earlier, the Rayleigh–Bénard convection together with the Taylor problem has become a paradigm of nonequilibrium phase transitions with its instability, bifurcations, pattern formation, and chaotic behaviors in fluid dynamics as well as in physical systems in general.

Consider the nondimensional equation (1.1.30) governing the motion and states of the Bénard convection problem. By solving the linearized eigenvalue problem, as given, e.g., by (4.1.12) in Chap. 4, the critical parameter equation (1.2.6) takes the following form

$$R = R_c, \quad (1.3.1)$$

where R_c is the first critical Rayleigh number.

With the dynamical transition theory, we can show (Ma and Wang, 2004b) that as the Rayleigh number R crosses the critical parameter R_c , the system is always under a continuous (Type-I) transition to an $(m - 1)$ -dimensional sphere $\mathcal{A}_R = S^{m-1}$, where m is the dimension of the eigenspace for the linearized problem, corresponding to the critical Rayleigh number R_c .

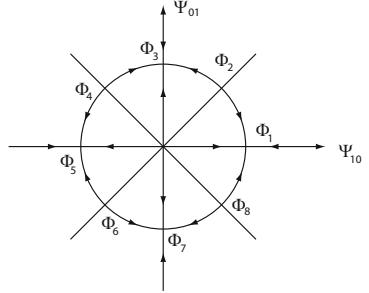
The structure of the set $\mathcal{A}_R = S^{m-1}$ of transition states and the related pattern formation mechanism depend on the physical boundary conditions and the spatial geometry (of the domain); see Ma and Wang (2007a) and Sect. 4.1 for more details.

First, in the case where the (nondimensional) spatial domain is given by $\Omega = (0, L)^2 \times (0, 1)$ with free slip lateral boundary condition and with L satisfying (4.1.62), the transition attractor $\mathcal{A}_R = S^1$ (i.e., $m = 2$) consists of exactly eight singular points Φ_i ($1 \leq i \leq 8$) and eight heteroclinic orbits connecting the singular points, as shown in Fig. 1.5, where $\Phi_1, \Phi_3, \Phi_5, \Phi_7$ are minimal attractors, and $\Phi_2, \Phi_4, \Phi_6, \Phi_8$ are saddle points. See Theorem 4.1.14. We note here that the connecting orbits are the transients, which represent most of the states in the transition attractor.

Second, in the case where the problem is equipped with periodic boundary conditions in the horizontal directions with periods L_1 and L_2 , the transition attractor \mathcal{A}_R is given by an attractor \mathcal{A}_R homeomorphic to a sphere, i.e.,

$$\mathcal{A}_R = \begin{cases} S^5 & \text{if } L_2 = \sqrt{k^2 - 1} L_1, \quad k = 2, 3, \dots, \\ S^3 & \text{otherwise.} \end{cases}$$

Fig. 1.5 The singular points $\Phi_1, \Phi_3, \Phi_5, \Phi_7$ are the minimal attractors, and the other singular points are the saddles



Due to the translational invariants in the horizontal directions, the high dimensional attractor \mathcal{A}_R consists of finite number of two-dimensional tori (containing only steady states) and the transients containing connecting the tori. The collection of the tori of steady states is a zero measure set in \mathcal{A}_R . Namely, as in the previous case, the most states in \mathcal{A}_R are the transients.

It is worth mentioning that the Key Philosophy 2 and Principle 1 play crucial role in deriving the above results, which cannot be directly derived by any existing methods. In fact, without the Key Philosophy 2 and Principle 1, one may not have any motivation to envision these results. Also, these results are global in nature.

Finally, the dynamic transition theory opens a new avenue to pattern formation, as the patterns and structure of transition states are dictated by solutions in \mathcal{A}_R , which are again determined by eigenvectors of the corresponding linearized problem (with high-order approximations). With the theory, for example, we can easily examine the formation and stability of different structures such as roles, rectangles, and hexagons. However, due to the limitation of the scope of this book, we touch only occasionally pattern formation issue in this book, and we intend to leave further exploration of pattern from the dynamic transition point of view in the future.

1.3.2 *El Niño Southern Oscillation (ENSO)*

ENSO is a global coupled ocean-atmosphere phenomenon, associated with floods, droughts, and other disturbances around the world. ENSO is the most prominent known source of inter-annual variability in weather and climate (about 3–8 years). In spite of its importance and a long history of studies, the understanding of its nature and mechanism is still lacking, and a careful fundamental level examination of the problem is crucial.

ENSO is closely related to the Walker circulation over the tropics, as shown in Figs. 1.6 and 1.7. When the convective activity weakens or reverses, an El Niño phase takes place, causing the ocean surface to be warmer than average, reducing or terminating the upwelling of cold water. On the other hand, a particularly strong Walker circulation causes a La Niña event, resulting in cooler sea-surface temperature (SST) due to stronger upwelling.

Fig. 1.6 Global atmospheric Walker circulation over the tropics

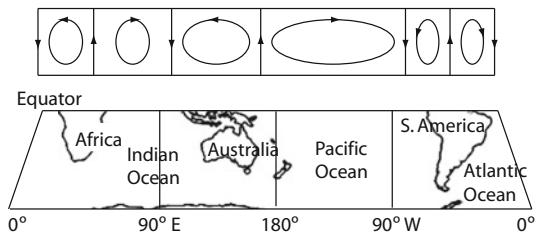


Fig. 1.7 Walker circulation over the tropic pacific ocean

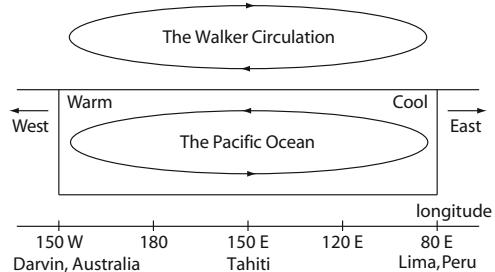
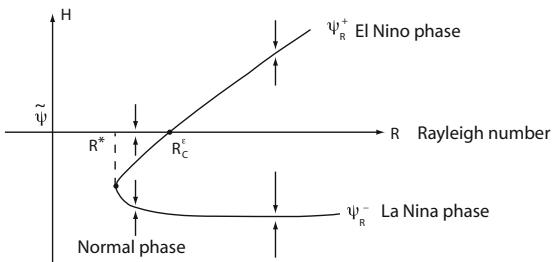


Fig. 1.8 The transition diagram for the equatorial atmospheric circulation, where R^* is the saddle-node bifurcation point, and R_c^ϵ is the first critical Rayleigh number. Here the Rayleigh number R is proportional to SST



An interesting current debate is whether ENSO is best modeled as a stochastic or chaotic system—linear and noise-forced, or nonlinear oscillatory and unstable system (Philander and Fedorov, 2003)?

With the dynamic transition theory and the geometric theory of incompressible flows (Ma and Wang, 2005d), we derive a new mechanism of ENSO, as a self-organizing and self-excitation system, with two highly coupled processes:

- the oscillation between the two metastable warm (El Niño phase) and cold (La Niña phase) events, and
- the spatiotemporal oscillation of the sea surface temperature (SST) field.

The symbiotic interplay between these two processes gives rise to the climate variability associated with the ENSO, leads to both the random and deterministic features of the ENSO, and defines a new natural *feedback* mechanism, which drives the sporadic oscillation of the ENSO. This new mechanism is schematically illustrated in Fig. 1.8, and see also Sect. 5.2 for more details.

We emphasize in particular that the transition behavior of ENSO phenomena depicted here offers a natural explanation of the randomness associated with mixed (Type-III) transitions, hence the terminology of random transition. The randomness is due to fluctuations of initial data for deterministic systems.

Another good example is the pipe flow. Under some natural conditions, the basic flow is always linearly and nonlinearly stable for all Reynolds number (Eckhardt et al., 2007). However, for large Reynolds numbers, both experimental and numerical simulations lead to chaotic behaviors. The crucial point of the problem is that the stable basic solution possesses a very narrow basin of attraction, and small fluctuations lead immediately to chaotic motion outside of the basin of attraction of the basic flow.

1.3.3 Dynamic Transition vs Transition in the Physical Space

The nature of a flow's boundary-layer separation from the boundary and interior separation plays a fundamental role in many physical problems. Typical examples of boundary layer separation include the vortices forming from the bank of a river, and the formation of vortices around an airfoil. The formation of tornadoes is a typical example of interior separation. Basically, in the boundary layer, the shear flow can detach/separate from the boundary, generating a slow backflow and leading to more complicated turbulent behavior; see Fig. 1.9. The interior separation is often connected with spin-offs.

Precise characterization of the boundary layer and interior separations of fluid flows is a long-standing problem in fluid mechanics, going back to the pioneering work of Prandtl (1904). Progress has been made recently using a systematic geometric theory of incompressible flows (Ma and Wang, 2005d).

The main philosophy of studying the flow structures in the physical space is to classify the topological structure and its transitions of the *instantaneous* velocity field (i.e., streamlines in the Eulerian coordinates), treating the time variable as a parameter as watching a movie. Based on this philosophy, the study includes researches in two directions:

- (1) the development of a global geometric theory of divergence-free vector fields on general two-dimensional compact manifolds with or without boundaries, and
- (2) the connections between the solutions of the Navier–Stokes (or Euler) equations and the dynamics of the velocity fields in the physical space.

As part of this program, a rigorous characterization of both boundary layer and interior separations are derived by the authors, in collaboration in part with Michael Ghil; see, e.g., Ghil et al. (2001, 2005), Ma and Wang (2001, 2005d).

To demonstrate some of the ideas for this characterization and its applications, we consider the Taylor–Couette flow, a viscous fluid between two coaxial rotating cylinders. For the classical Taylor problem, the basic (Couette) flow becomes unstable as soon as the rotation speed of the inner cylinder exceeds a critical value. This

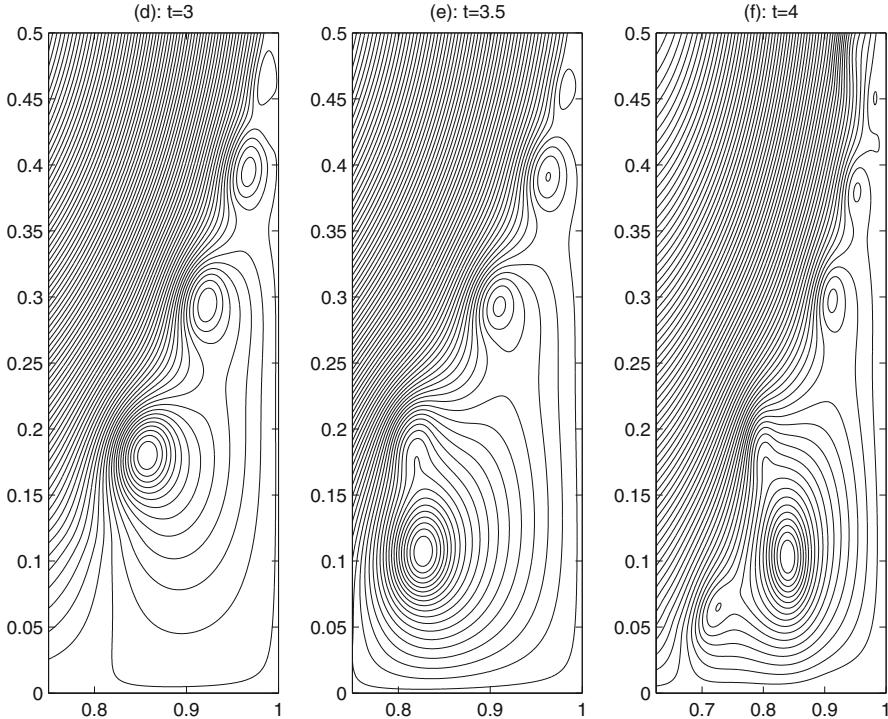


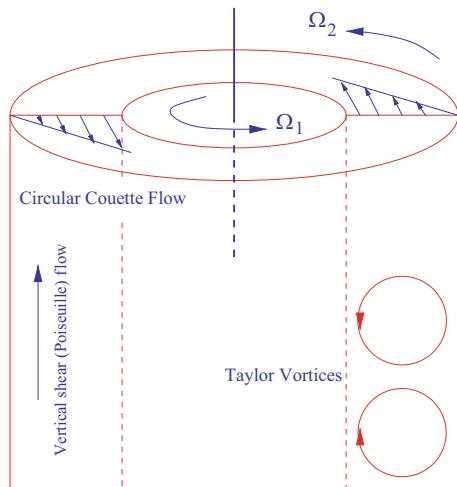
Fig. 1.9 Boundary layer separation of shear flow

instability gives rise to a stationary axisymmetric counter-rotating vortices that fill the whole annular region. The associated flow is referred to as Taylor–Couette (TC) flow.

When a through flow driven by a pressure gradient along the rotating axis is added, the base flow consists of a superposition of circular Couette flow and annular Poiseuille flow, called Couette–Poiseuille (CP) flow; see Fig. 1.10. The axial through-flow suppresses the basic stationary instability, and as the axial pressure gradient increases while the rotation speed of the inner cylinder is held fixed, the first bifurcation gives rise to a traveling train of axisymmetric Taylor vortices, commonly referred to as propagating Taylor vortices (PTV). Henceforth, the term Taylor–Couette–Poiseuille (TCP) flow is used to refer to all hydrodynamic phenomena pertaining to the open system described above; see among others (Raghu and Georgiadis, 2004) and the references therein.

With the dynamic transition theory, combined with characterization of boundary-layer and interior separations (Ma and Wang, 2005d), we show that there are two transition processes associated with the formation of the propagating Taylor vortices; see Ma and Wang (2009a):

Fig. 1.10 Taylor–Couette–Poiseuille flow



- The first transition is a dynamic transition at the first critical Taylor number, leading to a CP-like flow. In this case, there is no formation of PTV.
- The second transition occurs at another critical Taylor number T_1 and is the transition of flow structure in the physical space. This transition leads to the formation of PTV, as shown in Fig. 1.11.

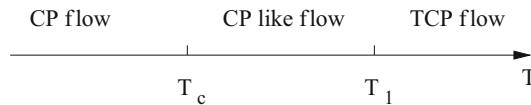


Fig. 1.11 Dynamic and structural transitions leading to the formation of the Taylor vortices

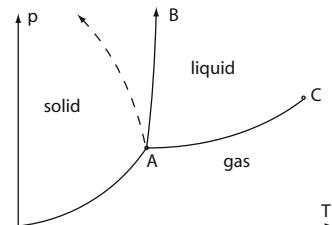
In a nutshell, contrary to what is commonly believed, we show that the PTV do not appear after the first dynamical transition, and they appear only when the Taylor number is further increased to cross another critical value so that a structural bifurcation in the physical spaces occurs.

1.3.4 Andrews Critical Point and Third-Order Gas-Liquid Transition

The physical-vapor transport (PVT) system is one of the most basic models to study in equilibrium phase transitions. As we know, a PVT system is a system composed of one type of molecules, and the interaction between molecules is governed by the Van der Waals law. The molecules generally have a repulsive core and a short-range attraction region outside the core. Such systems have a number of phases: gas, liquid, and solid, and a solid can appear in a few phases. The most typical example of a

PVT system is water. In general, the phase transitions of *PVT* systems mainly refer to the gas-liquid, gas-solid, and liquid-solid phase transitions. These transitions are all first order in the Ehrenfest sense (i.e., discontinuous) and are accompanied by a latent heat and a change in density.

Fig. 1.12 Coexistence curves of a typical *PVT* system: *A* is the triple point, *C* is the critical point, and the dashed curve is a melting line with negative slope



A *PT*-phase diagram of a typical *PVT* system is schematically illustrated in Fig. 1.12, where point *A* is the triple point at which the gas, liquid, and solid phases can coexist. Point *C* is the Andrews critical point at which the gas-liquid coexistence curve terminates (Reichl, 1998; Stanley, 1971; Kleman and Laverntovich, 2007; Fisher, 1964). The fact that the gas-liquid coexistence curve has a critical point means that we can go continuously from a gaseous state to a liquid state without ever meeting an observable phase transition, if we choose the right path.

A natural question to ask is why the Andrews critical point exists and what is the order of transition going beyond this critical point. This is a longstanding problem, which is currently solved using the dynamic transition theory in Ma and Wang (2008b, 2011e):

First, a dynamic model for the phase transition in a *PVT* system is established based on (a) the unified dynamic models for equilibrium phase transitions established in Sect. 3.1, (b) the classical phase diagram given in Fig. 1.12, and (c) the van der Waals equations.

Second, with this dynamic model, we are able to derive a theory on the Andrews critical point *C*:

1. the Andrews point *C* is derived using a dynamic approach different from the classical derivation,
2. the gas-liquid coexistence curve can be extended beyond the Andrews critical point, and
3. the transition is *first order* before the critical point, *second order* at the critical point, and *third order* beyond the Andrews critical point.

This clearly explains why it is hard to observe the gas-liquid phase transition beyond the Andrews critical point. Furthermore, the analysis leads naturally to the introduction of a general asymmetry principle of fluctuations and the preferred transition mechanism for a thermodynamic system. The theoretical results derived are in agreement with the experimental results obtained in Nishikawa and Morita (1998). Also, the derived second-order transition at the critical point is consistent with the result obtained in Fisher (1964).

In addition, a careful examination of the dynamic transitions of the PVT system suggests that there are two possible phase transition behaviors near a critical point, and theoretically each of them has the same probability to take place. However only one of them can appear in reality, and this suggests the asymmetry of fluctuations given in Conjecture 3.4.3. Namely, the symmetry of fluctuations for general thermodynamic systems may not be universally true. In other words, in some systems with multi-equilibrium states, the fluctuations near a critical point occur only in one basin of attraction of some equilibrium states, which are the ones that can be physically observed.

1.3.5 Binary Systems

Materials, consisting of two components A and B , such as binary alloys, binary solutions, and polymers, are called binary systems. Sufficient cooling of a binary system may lead to phase separations. Namely, at the critical temperature, the concentrations of both components A and B with homogeneous distribution undergo changes, leading to heterogeneous spatial distributions (Kleman and Laverntovich, 2007; Reichl, 1998; Pismen, 2006; Novick-Cohen and Segel, 1984; Cahn and Hillard, 1957).

A commonly used model for a binary system is the classical Cahn–Hilliard equation, which can also be derived using the unified model presented early in this chapter; see Chap. 3 for more details. Let u_A and u_B ($= 1 - u_A$) be the concentrations of components A and B respectively. The Helmholtz free energy is given by

$$F(u) = F_0 + \int_{\Omega} \left[\frac{\mu}{2} |\nabla u_B|^2 + f(u_B) \right] dx, \quad (1.3.2)$$

where the energy density function f can be explicitly given using the Bragg–Williams or the Hildebrand theory (Reichl, 1998); see (3.5.2):

$$\begin{aligned} f = & \mu_A(1 - u_B) + \mu_B u_B + RT(1 - u_B) \ln(1 - u_B) \\ & + RT u_B \ln u_B + au_B(1 - u_B), \end{aligned} \quad (1.3.3)$$

where μ_A, μ_B are the chemical potentials of A and B respectively, R the molar gas constant, and $a > 0$ the measure of repulsion action between A and B .

In a homogeneous state, u_B is a constant in space and is equal to its mean value, i.e., $u_B = \bar{u}_B$. Let

$$u = u_B - \bar{u}_B, \quad u_0 = \bar{u}_B.$$

The classical Cahn–Hilliard equation is then derived using the Landau mean field theory:

$$\frac{\partial u}{\partial t} = -k\Delta^2 u + \Delta[b_1 u + b_2 u^2 + b_3 u^3], \quad (1.3.4)$$

where the coefficients k , b_1 , b_2 , and b_3 can be given explicitly in terms of the temperature T , the pressure p , and the homogeneous mol fraction u_0 of component B , using the explicit expression of f ; see (3.5.4).

To illustrate the main ideas, for simplicity, we consider here only the case where the binary solution is contained in a cube $(0, L_d)^3$; see Chap. 3 and Ma and Wang (2009b,c) for more general cases.

Critical Parameter Equation First the critical parameter curve equation is derived by setting the first eigenvalue of the linearized problem to zero, $\beta_1 = 0$, leading to

$$T_c = \frac{u_0(1-u_0)}{RD} \left(a - \frac{k\pi^2}{2L_d^2} \right), \quad (1.3.5)$$

where T_c is the critical temperature.

Order of Phase Transitions and Phase Diagrams By the nonlinear dynamic transition theory, then we derive that the order of transition is completely determined by the sign of the following number K_d (the subscript d refers to dimensional form):

$$K_d = \frac{26L_d^2 b_2^2}{27\pi^2 k} - b_3, \quad (1.3.6)$$

which depends on the temperature T , the pressure p , and the homogeneous mol fraction u_0 of component B . Namely, as $K_d > 0$, the transition is of the first order with latent heat, corresponding to catastrophic transition in the dynamic classification scheme, and as $K_d < 0$, the transition is of the second order.

To examine the order of separations, we start with two important formulas for the critical parameters. First, by (3.5.4), solving $K_d = 0$ gives a precise formula for the critical length scale L_d in terms of the critical temperature T_c and the mol fraction u_0 :

$$L_d = \frac{3\sqrt{3k}\pi\sqrt{u_0(1-u_0)(1-3u_0+3u_0^3)}}{2\sqrt{26DRT_c}|u_0 - \frac{1}{2}|}. \quad (1.3.7)$$

We are ready now to examine the order of separation in terms of the length scale L_d and mol fraction u_0 . First, using (1.3.7), we derive the Lu_0 -phase diagram given in Fig. 1.13. Then using Fig. 1.13 and (1.3.5), we derive the Tu_0 -phase diagram given in Fig. 1.14.

Fig. 1.13 Lu_0 -phase diagram derived using (3.5.60) or (3.5.61): region II is the first-order transition region with latent heat, and region I is the second-order transition region

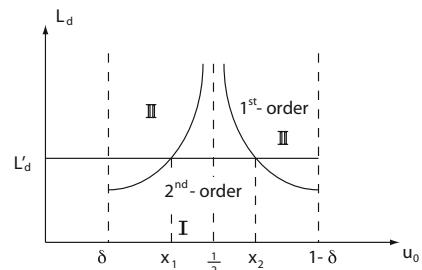
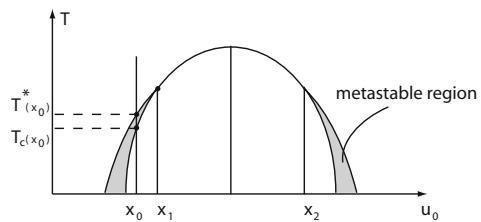


Fig. 1.14 Tu_0 -phase diagram for a fixed length scale L' , derived using Fig. 1.13 and (1.3.5): the shadowed region stands for the metastable region, and $T^*(x_0)$ is a critical temperature corresponding to a saddle-node bifurcation as shown in Chap. 3, and $T_c(x_0)$ is the critical temperature



From these two phase diagrams, we obtain immediately the following physical conclusions:

1. For a fixed length scale $L = L'$, there are numbers $x_1 < \frac{1}{2} < x_2$ such that the transition is of second order if the molar fraction $x_1 < u_0 < x_2$, and the transition is of first order if $u_0 > x_2$ or $u_0 < x_1$. See Fig. 1.13.
2. The phase diagram Fig. 1.14 is for this fixed length scale L' . The points x_1 and x_2 are the two molar concentrations where there is no metastable region and no hysteresis phenomena for $x_1 < u_0 < x_2$. In other words,

$$T^*(u_0) = T_c(u_0) \quad \text{for } x_1 < u_0 < x_2,$$

$T^*(x_0)$ and $T_c(x_0)$ are critical temperatures.

It is worth mentioning that the results derived using the Cahn–Hilliard equation (1.3.4) is also valid for the model using the full Bragg–Williams potential (1.3.3).

Pattern Formation and Structure of Transition Attractor As indicated earlier, in the case where $K_d > 0$, the transition is catastrophic. For the case where $K_d < 0$, the transition is of the second order, and the local attractor representing the complete transition states $\Sigma_T = S^2$ consists of 26 steady states and the transients connecting these 26 steady states, as shown in Fig. 1.15. This clearly demonstrates again the point mentioned before for the Bénard convection, the dynamic transition theory opens a new avenue toward to pattern formation, as the patterns and structure of transition states are dictated by solutions in Σ_T .

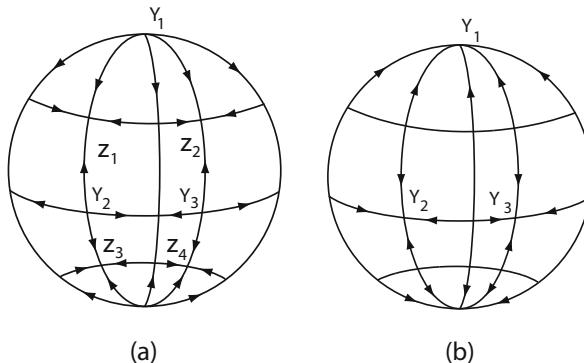


Fig. 1.15 Structure of $\Sigma_T = S^2$ with either (a) $\pm Z_k$ ($1 \leq k \leq 4$) are attractors, or (b) $\pm Y_k$ ($1 \leq k \leq 3$) are attractors, determined by a simple relationship between b_2 and b_3 ; see Theorem 3.5.5



Chapter 2

Dynamic Transition Theory

This chapter introduces the dynamic transition theory for nonlinear dissipative systems developed recently by the authors. The main focus is to derive a general principle, Principle 1, on dynamic transitions for dissipative systems and to study the types and structure of dynamic transitions. Our study is based on the key philosophy, Philosophy 2, which states that we should search for a complete set of transition states.

2.1 General Dynamic Transition Theory

2.1.1 Classification of Dynamic Transitions

Let X and X_1 be two Banach spaces, and $X_1 \subset X$ be a compact and dense inclusion. In this chapter, we always consider the following nonlinear evolution equations

$$\frac{du}{dt} = L_\lambda u + G(u, \lambda), \quad u(0) = \varphi, \quad (2.1.1)$$

where $u : [0, \infty) \rightarrow X$ is the unknown function. Here we assume that the system control parameter $\lambda \in \mathbb{R}^1$. As explained in (1.2.3), in the case where $\lambda \in \mathbb{R}^N$, we always take a one-dimensional path for λ , and we can always treat λ as in the case for $N = 1$.

Assume that¹ $L_\lambda : X_1 \rightarrow X$ is a parameterized linear completely continuous field depending continuously on λ , which satisfies

¹ We follow here the notation used in Ma and Wang (2005b). In particular, a linear operator $L : X_1 \rightarrow X$ is called a completely continuous field if $L = -A + B : X_1 \rightarrow X$, $A : X_1 \rightarrow X$ is a linear homeomorphism, and $B : X_1 \rightarrow X$ is a linear compact operator. Also, we refer the interested readers to classical books, e.g., Kato (1995), for the basic knowledge of linear operators, and to Henry (1981) and Pazy (1983) for semigroups of linear operators and sectorial operators.

- $L_\lambda = -A + B_\lambda$ is a sectorial operator,
 $A : X_1 \rightarrow X$ is a linear homeomorphism, and
 $B_\lambda : X_1 \rightarrow X$ is a linear compact operator.

In this case, we can define the fractional order spaces X_σ for $\sigma \in \mathbb{R}^1$ using interpolation, and we refer interested readers to Henry (1981), Pazy (1983), and Ma and Wang (2005b) for their precise definitions. Also we assume that $G(\cdot, \lambda) : X_\alpha \rightarrow X$ is a C^r ($r \geq 1$) bounded mapping for some $0 \leq \alpha < 1$, depending continuously on $\lambda \in \mathbb{R}^1$, and

$$G(u, \lambda) = o(\|u\|_{X_\alpha}) \quad \forall \lambda \in \mathbb{R}^1. \quad (2.1.3)$$

Hereafter we always assume the conditions (2.1.2) and (2.1.3), which imply that the system (2.1.1) has a dissipative structure.

Definition 2.1.1 We say that the system (2.1.1) undergoes a (dynamic) transition from $(u, \lambda) = (0, \lambda_0)$ at $\lambda = \lambda_0$ if the following two conditions hold true:

1. if $\lambda < \lambda_0$, $u = 0$ is locally asymptotically stable for (2.1.1), and
2. if $\lambda > \lambda_0$, there exists a neighborhood $U \subset X$ of $u = 0$ independent of λ , such that for any $\varphi \in U \setminus \Gamma_\lambda$ the solution $u_\lambda(t, \varphi)$ of (2.1.1) satisfies that

$$\limsup_{t \rightarrow \infty} \|u_\lambda(t, \varphi)\| \geq \delta(\lambda) > 0, \quad \lim_{\lambda \rightarrow \lambda_0} \delta(\lambda) \geq 0,$$

where Γ_λ is the stable manifold of $u = 0$, with $\text{codim } \Gamma_\lambda \geq 1$ in X for $\lambda > \lambda_0$.

The above definition amounts to saying that the basic state $u = 0$ is stable for $\lambda < \lambda_0$, and lose completely its stability for $\lambda > \lambda_0$. Obviously, the attractor bifurcation of (2.1.1) is a dynamic transition. However, bifurcation and transition are two different, but related, concepts. Definition 2.1.1 defines the transition of (2.1.1) from a stable equilibrium point to other, not necessarily equilibrium, states. In general, we can define transitions from one attractor to another as follows.

Definition 2.1.2 Let $\Sigma_\lambda \subset X$ be an invariant set of (2.1.1). We say that (2.1.1) undergoes a transition from $(\Sigma_{\lambda_0}, \lambda_0)$ at $\lambda = \lambda_0$ if the following conditions are satisfied:

- when $\lambda < \lambda_0$, Σ_λ is a local minimal attractor, and
- when $\lambda > \lambda_0$, there exists a neighborhood $U \subset X$ of Σ_λ independent of λ such that for any $\varphi \in U \setminus (\Gamma_\lambda \cup \Sigma_\lambda)$, the solution $u(t, \varphi)$ of (2.1.1) satisfies that

$$\limsup_{t \rightarrow \infty} \text{dist}(u(t, \varphi), \Sigma_\lambda) \geq \delta(\lambda) > 0, \quad \lim_{\lambda \rightarrow \lambda_0} \delta(\lambda) = \delta \geq 0,$$

where Γ_λ is the stable manifolds of Σ_λ with $\text{codim } \Gamma_\lambda \geq 1$.

The main focus of this book is on dynamic transitions from equilibrium states. We start with the principle of exchange of stabilities (PES). Let the (generalized) eigenvalues (counting multiplicity) of L_λ be given by $\{\beta_j(\lambda) \in \mathbb{C} \mid j = 1, 2, \dots\}$, and let

$$\operatorname{Re} \beta_i(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_0, \\ = 0 & \text{if } \lambda = \lambda_0, \\ > 0 & \text{if } \lambda > \lambda_0 \end{cases} \quad \forall 1 \leq i \leq m, \quad (2.1.4)$$

$$\operatorname{Re} \beta_j(\lambda_0) < 0 \quad \forall j \geq m + 1. \quad (2.1.5)$$

A few technical tools for verifying the PES for applications are introduced in Ma and Wang (2008d), by calculating the derivatives of the first eigenvalues, which cross the imaginary axis at the critical parameter.

The following theorem is a basic principle of transitions from equilibrium states, and provides sufficient conditions and a basic classification for transitions of nonlinear dissipative systems. The proof of this theorem is given at the end of this subsection.

Theorem 2.1.3 *Let the conditions (2.1.4) and (2.1.5) hold true. Then, the system (2.1.1) always undergoes a dynamic transition from $(u, \lambda) = (0, \lambda_0)$, and there is a neighborhood $U \subset X$ of $u = 0$ such that the transition is one of the following three types:*

- (1) **CONTINUOUS TRANSITION:** *there exists an open and dense set $\tilde{U}_\lambda \subset U$ such that for any $\varphi \in \tilde{U}_\lambda$, the solution $u_\lambda(t, \varphi)$ of (2.1.1) satisfies*

$$\lim_{\lambda \rightarrow \lambda_0} \limsup_{t \rightarrow \infty} \|u_\lambda(t, \varphi)\| = 0.$$

- (2) **JUMP TRANSITION:** *for any $\lambda_0 < \lambda < \lambda_0 + \varepsilon$ with some $\varepsilon > 0$, there is an open and dense set $U_\lambda \subset U$ such that for any $\varphi \in U_\lambda$,*

$$\limsup_{t \rightarrow \infty} \|u_\lambda(t, \varphi)\| \geq \delta > 0,$$

where $\delta > 0$ is independent of λ .

- (3) **MIXED TRANSITION:** *for any $\lambda_0 < \lambda < \lambda_0 + \varepsilon$ with some $\varepsilon > 0$, U can be decomposed into two open (not necessarily connected) sets U_1^λ and U_2^λ : $\overline{U} = \overline{U}_1^\lambda \cup \overline{U}_2^\lambda$, $U_1^\lambda \cap U_2^\lambda = \emptyset$, such that*

$$\lim_{\lambda \rightarrow \lambda_0} \limsup_{t \rightarrow \infty} \|u(t, \varphi)\| = 0 \quad \forall \varphi \in U_1^\lambda,$$

$$\limsup_{t \rightarrow \infty} \|u(t, \varphi)\| \geq \delta > 0 \quad \forall \varphi \in U_2^\lambda,$$

where U_1^λ and U_2^λ are called metastable domains.

With this theorem in our disposal, it is natural to classify all dynamic transitions into the three categories as described in this theorem. We remark that the concept of phase transition originates from the statistical physics and thermodynamics. For brevity, we use “phase” to refer to a stable state for a nonlinear system from many sciences fields including physics, chemistry, biology, ecology, economics, fluid dynamics, geophysical fluid dynamics, etc. In fact, the dynamic transition the-

ory presented in this book provides a unified mathematical approach/methodology applicable to a wide variety of topics.

Definition 2.1.4 (Classification of Dynamic Transitions) *The dynamic transitions for (2.1.1) at $\lambda = \lambda_0$ is classified using their dynamic properties: continuous, jump, and mixed as given in Theorem 2.1.3, which are called Type-I, Type-II, and Type-III respectively.*

If the system (2.1.1) possesses the gradient-type structure, then the dynamic transitions are associated with the equilibrium phase transitions in statistical physics; otherwise they relate to the nonequilibrium phase transitions.

For equilibrium phase transition problems in statistical physics, classically, there are several ways to classify phase transitions. The most notable classification scheme of phase transitions is the classical Ehrenfest classification scheme, where phase transitions are labeled by the lowest derivative of the free energy that is discontinuous at the transition.

Proof of Theorem 2.1.3. Using the center manifold reduction outlined in Chap. 1, we only have to consider the reduced equation (1.2.7). By the PES assumption (2.1.4) and (2.1.5), we know that for $\lambda > \lambda_0$, $x = 0$ is a repeller. Namely, there exists a neighborhood O_λ of $x = 0$ such that for any $x_0 \in O_\lambda$,

$$\liminf_{t \rightarrow \infty} \|x(t, x_0)\| > 0.$$

As $\lambda \rightarrow \lambda_0^+$, there exists a neighborhood O of $x = 0$ such that $O = \overline{O}_1 \cup \overline{O}_2$ and

$$O_1 = \{x_0 \in O \mid \lim_{\lambda \rightarrow \lambda_0} \limsup_{t \rightarrow \infty} \|x(t, x_0)\| = 0\},$$

$$O_2 = \{x_0 \in O \mid \lim_{\lambda \rightarrow \lambda_0} \limsup_{t \rightarrow \infty} \|x(t, x_0)\| > \delta > 0\},$$

where $\delta > 0$ is obtained using the unstable manifold theorem and finite dimensionality of the reduced problem. Then there can only be the following three cases:

- (a) O_2 is nowhere dense and O_1 is open and dense in O ,
- (b) O_1 is nowhere dense and O_2 is open and dense in O , and
- (c) Both O_1 and O_2 contains interior points.

We note that by the continuity of orbits, the case where both O_1 and O_2 are dense in O cannot happen. Again by the continuity of orbits, we see that cases (a)–(c) correspond to the three types of transitions respectively for (1.2.7). \square

2.1.2 Characterization of Transition Types

An important aspect of the transition theory is to determine which of the three types of transitions given by Theorem 2.1.3 occurs in a specific problem. The following theorem is useful to distinguish the transition types of (2.1.1) at $(u, \lambda) = (0, \lambda_0)$, which

can be deduced directly from Theorem 2.1.3, the attractor bifurcation theorem, Theorem 2.2.2, and Theorem 2.1.15 in Ma and Wang (2005d) on the stability of extended orbits.

Theorem 2.1.5 *Let the conditions (2.1.4) and (2.1.5) hold true, then the transition types of (2.1.1) at $\lambda = \lambda_0$ is completely dictated by its reduced equation (1.2.7). In particular, let $U \subset \mathbb{R}^m$ be a neighborhood of $x = 0$. Then we have the following assertions:*

- (1) *If the transition of (2.1.1) at $(0, \lambda_0)$ is continuous, then there is an open and dense set $\tilde{U} \subset U$ such that for any $x_0 \in \tilde{U}$ the solution $x(t, x_0)$ of (1.2.7) at $\lambda = \lambda_0$ with $x(0, x_0) = x_0$ satisfies that*

$$\lim_{t \rightarrow \infty} x(t, x_0) = 0.$$

- (2) *If there exists an open and dense set $\tilde{U} \subset U$ such that for any $x_0 \in \tilde{U}$ the solution $x(t, x_0)$ of (1.2.7) at $\lambda = \lambda_0$ satisfies*

$$\limsup_{t \rightarrow \infty} |x(t, x_0)| \neq 0,$$

then the transition is a jump transition.

- (3) *If the transition is mixed, then there exist two open sets $\tilde{U}_1, \tilde{U}_2 \subset U$ such that*

$$\begin{aligned} \lim_{t \rightarrow \infty} x(t, x_0) &= 0 & \forall x_0 \in \tilde{U}_1, \\ \limsup_{t \rightarrow \infty} |x(t, x_0)| &\neq 0 & \forall x_0 \in \tilde{U}_2. \end{aligned}$$

In general, the conditions in Assertions (1)–(3) of Theorem 2.1.5 are not sufficient. They, however, do give sufficient conditions when (2.1.1) has a variational structure. To see this, let (2.1.1) be a gradient-type equation. Under the conditions (2.1.4) and (2.1.5), by Lemmas A.2.7 and A.2.8, in a neighborhood $U \subset X$ of $u = 0$, the center manifold M^c in U at $\lambda = \lambda_0$ consists of three subsets

$$M^c = W^u + W^s + D,$$

where W^s is the stable set, W^u is the unstable set, and D is the hyperbolic set of (1.2.7).

Theorem 2.1.6 *Let (2.1.1) be a gradient-type equation, and the conditions (2.1.4) and (2.1.5) hold true. If $u = 0$ is an isolated singular point of (2.1.1) at $\lambda = \lambda_0$, then we have the following assertions:*

- (1) *The transition of (2.1.1) at $(u, \lambda) = (0, \lambda_0)$ is continuous if and only if $u = 0$ is locally asymptotically stable at $\lambda = \lambda_0$, i.e., the center manifold is stable: $M^c = W^s$.*
- (2) *The transition of (2.1.1) at $(u, \lambda) = (0, \lambda_0)$ is jump if and only if there is an open and dense set \tilde{U} in a neighborhood $U \subset H$ of $u = 0$ such that for any $\varphi \in \tilde{U}$, the solution $u(t, \varphi)$ of (2.1.1) at $\lambda = \lambda_0$ satisfies*

$$\limsup_{t \rightarrow \infty} \|u(t, \varphi)\| \neq 0.$$

Proof. The sufficiency of Assertion (1) follows from Theorem 2.2.2. By Lemmas A.2.7 and A.2.8, the hyperbolic set $D \subset M^c$ is open, and if $D = \emptyset$, then the center manifold M^c satisfies

$$M^c = W^s \quad \text{or} \quad M^c = W^u. \quad (2.1.6)$$

Thus, the necessity of Assertion (1) follows from Assertion (1) of Theorem 2.1.5 and (2.1.6). Assertion (2) follows from Assertion (2) of Theorem 2.1.5. The proof is complete. \square

2.1.3 Local Topological Structure of Transitions

The study of both local and global topological structure of transitions is an important topic in the phase transition dynamics in nonlinear sciences. In this subsection we briefly discuss the local topological structure of transitions of (2.1.1) near $(u, \lambda) = (0, \lambda_0)$. Equivalently, we only have to address the topological structure of transitions of the finite dimensional system (1.2.7) near $(x, \lambda) = (0, \lambda_0)$.

Let $O \subset \mathbb{R}^m$ be a neighborhood of $x = 0$. By Theorem 2.1.5, for the transition of (2.1.1) at $(u, \lambda) = (0, \lambda_0)$, there are two subsets in O , the stable set M^s and the complementary set $O_2 = O \setminus M^s$. These sets are related to transitions as follows:

$$\begin{aligned} \overline{M^s} = \overline{O}, \quad \overset{\circ}{O}_2 = \emptyset &\quad \Leftarrow \quad \text{Continuous transition,} \\ \overset{\circ}{M^s} = \emptyset, \quad \overline{O}_2 = \overline{O} &\quad \Rightarrow \quad \text{Jump transition,} \\ \overline{O} = \overline{M^s} + \overline{O}_2, \quad \overset{\circ}{M^s} \neq \emptyset, \quad \overset{\circ}{O}_2 \neq \emptyset &\quad \Leftarrow \quad \text{Mixed transition,} \end{aligned}$$

where $\overset{\circ}{U}$ stands for the set of interior points of a set U .

We now examine continuous transitions based on the properties of the stable set.

CASE WHERE THE STABLE SET $M^s = O$: In this case, the transition is an attractor bifurcation, and the topological structure of transitions in O is characterized by the attractor, which is an $(m - 1)$ -dimensional homological sphere.

CASE WHERE $\overline{M^s} = \overline{O}$ AND $M^s \neq O$: In this case, $\overset{\circ}{M^s}$ may not be connected, and both the following two cases can happen:

- (a) there is an attractor in $\overset{\circ}{M^s}$ bifurcated from $(0, \lambda_0)$; and
- (b) there is no attractor in $\overset{\circ}{M^s}$ bifurcated from $(0, \lambda_0)$.

Example 2.1. Then this example shows that the case (b) does happen:

$$\begin{aligned} \frac{dx_1}{dt} &= 4\lambda x_1 + \lambda x_2 + x_1^2 - 4x_1x_2 - x_2^2, \\ \frac{dx_2}{dt} &= -6\lambda x_1 + 6x_1x_2. \end{aligned} \quad (2.1.7)$$

The linear and nonlinear operators in (2.1.7) are given by

$$A_\lambda = \begin{pmatrix} 4\lambda & \lambda \\ -6\lambda & 0 \end{pmatrix}, \quad G(x) = \begin{pmatrix} x_1^2 - 4x_1x_2 - x_2^2 \\ 6x_1x_2 \end{pmatrix}.$$

The matrix A_λ has two eigenvalues $\beta_1(\lambda) = \lambda(2 + i\sqrt{2})$, $\beta_2(\lambda) = \lambda(2 - i\sqrt{2})$, which satisfy (2.1.4) with $\lambda_0 = 0$. We can calculate that

$$\text{ind}(G, 0) = 2. \quad (2.1.8)$$

To further study the transitions of equation (2.1.7), we need to introduce a lemma, called the Poincaré formula, which provides some connections between the local topological structure of singular points of a two-dimensional vector field and the indices of the singular points. This formula is also useful in later discussions.

Lemma 2.1.7 *Let $v \in C^r(\Omega, \mathbb{R}^2)$ ($r \geq 0$) be a 2D vector field, and $v(0) = 0$. Then the following Poincaré index formula holds true:*

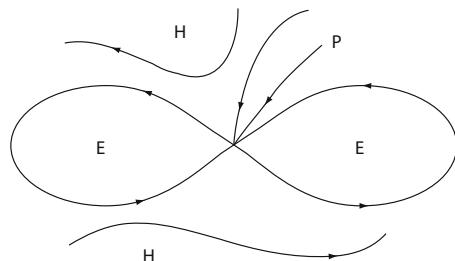
$$\text{ind}(v, 0) = 1 + \frac{1}{2}(e - h), \quad (2.1.9)$$

where e is the number of elliptic regions, h is the number of hyperbolic regions, and the elliptic, hyperbolic, and parabolic regions E, H , and P in a neighborhood $U \subset \mathbb{R}^2$ of $x = 0$ are defined by (see Fig. 2.1):

$$\begin{aligned} E &= \{x \in U \mid \lim_{t \rightarrow \infty} S(t)x = \lim_{t \rightarrow -\infty} S(-t)x = 0\}, \\ H &= \{x \in U \mid S(t)x, S(-t)x \notin U \text{ as } t > t_0 \text{ for some } t_0 > 0\}, \\ P &= \{x \in U \mid \text{either } \lim_{t \rightarrow \infty} S(t)x = 0, S(-t)x \notin U(t > t_0), \\ &\quad \text{or } \lim_{t \rightarrow -\infty} S(-t)x = 0, S(t)x \notin U, \\ &\quad \text{or } S(t)x, S(-t)x \in U, \forall t \geq 0\}, \end{aligned}$$

where $S(t)$ is the semigroup generated by v .

Fig. 2.1 Schematic of elliptic, parabolic, and hyperbolic regions



By the Poincaré index formula (2.1.9), we deduce from (2.1.8) that the vector field $G(x)$ has at least two elliptic regions near $x = 0$. In fact, $G(x)$ has just two elliptic regions. To see this, it is clear that the following equation

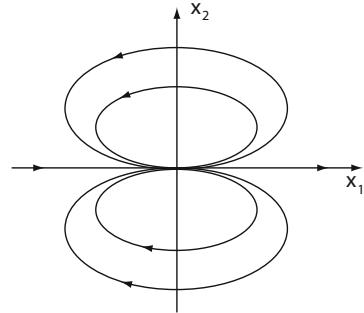
$$k = \frac{x_2}{x_1} = \frac{G_2(x)}{G_1(x)} = \frac{6x_1x_2}{x_1^2 - 4x_1x_2 - x_2^2} = \frac{6k}{1 - 4k - k^2}$$

has only one real solution $k = 0$, which implies that $G(x)$ has only two straight line orbits on the x_1 -axis which reach to $x = 0$. Therefore, the topological structure of $G(x)$ is as shown in Fig. 2.2, and for each point $x_0 \in \mathbb{R}^2$ and $x_0 \neq (x_1, 0)$ with $x_1 > 0$, we have

$$\lim_{t \rightarrow \infty} S(t)x_0 = 0,$$

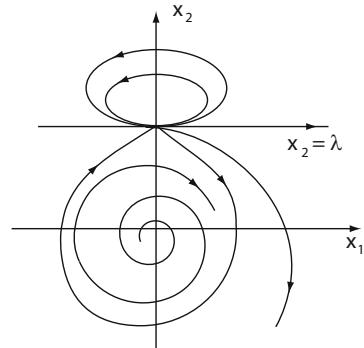
where $S(t)$ is the semigroup generated by $G(x)$.

Fig. 2.2 Topological structure of G



As $\lambda > \lambda_0 = 0$, the system (2.1.7) has a continuous transition, and bifurcates from $(0, \lambda_0)$ to a unique singular point $x_\lambda = (0, \lambda)$. The topological structure of this transition is as shown in Fig. 2.3. Obviously, there is no attractor bifurcated from $(0, \lambda_0)$ on $\lambda > \lambda_0$. \square

Fig. 2.3 Topological structure after the transition



By Theorem 2.1.5 we know that if the interior of the stable set is empty, i.e., $\mathring{M}^s = \emptyset$, then $\overline{O}_2 = \overline{O}$, and the transition of (2.1.1) at $(0, \lambda_0)$ is jump. However, the converse statement is not true. The following example shows that there exist systems

which have jump transition, but their stable sets M_s at the critical state at $\lambda = \lambda_0$ are open and dense in a neighborhood $O \subset \mathbb{R}^m$ of $x = 0$.

Example 2.2. Consider the system given by

$$\begin{aligned}\frac{dx_1}{dt} &= 2\lambda x_1 + 3\lambda x_2 + x_1^2 - x_2^2, \\ \frac{dx_2}{dt} &= -\lambda x_1 + x_1 x_2.\end{aligned}\tag{2.1.10}$$

The linear operator in (2.1.10) has the eigenvalues

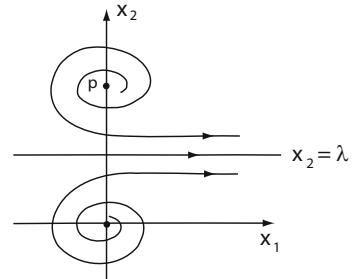
$$\beta_1(\lambda) = \lambda(1 + i\sqrt{2}), \quad \beta_2(\lambda) = \lambda(1 - i\sqrt{2}).$$

It is easy to check that the nonlinear operator $G(x) = (x_1^2 - x_2^2, x_1 x_2)^t$ is topologically equivalent to the structure as shown in Fig. 2.2, i.e., the stable set M^s of G is open and dense in \mathbb{R}^2 . However, as $\lambda > \lambda_0 = 0$, the system (2.1.10) bifurcates from $(0, \lambda_0)$ to a unique singular point $x_\lambda = (x_1^\lambda, x_2^\lambda) = (0, 3\lambda)$, and the eigenvalues of the Jacobian matrix at x_λ of the vector field in (2.1.10) is

$$\tilde{\beta}_1(\lambda) = \lambda(1 + i\sqrt{5}), \quad \tilde{\beta}_2(\lambda) = \lambda(1 - i\sqrt{5}).$$

It shows that the transition of (2.1.10) at $(0, \lambda_0) = (0, 0)$ on $\lambda > \lambda_0$ is jumping, which is schematically illustrated in Fig. 2.4. \square

Fig. 2.4 Jump transition for Example 2.2, with $p = (0, 3\lambda)$



An important topic related to jump transitions is the saddle node bifurcation, found in phase transition problems in nonlinear sciences, which will be addressed in later sections.

The structural problem of mixed transitions deals mainly with the stable and unstable domains and the structure of the bifurcated attractor in stable domains. Mathematically, a related interesting problem is to study the topological structure of bifurcated invariant sets for (2.1.1).

2.2 Continuous Transition

2.2.1 Finite Dimensional Systems

In this subsection, we are devoted to study the attractor bifurcation for the following finite dimensional systems

$$\frac{dx}{dt} = A_\lambda x + G(x, \lambda), \quad \lambda \in \mathbb{R}^1, \quad x \in \mathbb{R}^n \quad (n \geq 1), \quad (2.2.1)$$

where A_λ is an $n \times n$ matrix, whose entries are continuous functions of λ , and $G : \mathbb{R}^n \times \mathbb{R}^1 \rightarrow \mathbb{R}^n$ is C^r ($r \geq 1$) on $x \in \mathbb{R}^n$, and continuous on $\lambda \in \mathbb{R}^1$:

$$A_\lambda = \begin{pmatrix} a_{11}(\lambda) & \cdots & a_{1n}(\lambda) \\ \vdots & \ddots & \vdots \\ a_{n1}(\lambda) & \cdots & a_{nn}(\lambda) \end{pmatrix}, \quad G(x, \lambda) = o(|x|) \quad \forall \lambda \in \mathbb{R}^1. \quad (2.2.2)$$

Let all eigenvalues (counting multiplicities) of A_λ be given by $\beta_1(\lambda), \dots, \beta_n(\lambda)$. It is known that the eigenvalues $\beta_i(\lambda)$ ($1 \leq i \leq n$) depend continuously on λ .

In the following, we give some definitions of dynamic bifurcation for (2.2.1), which are also suitable for infinite dimensional systems; see also Ma and Wang (2005b,c).

Definition 2.2.1

- We say that the system (2.2.1) bifurcates from $(x, \lambda) = (0, \lambda_0)$ to an invariant set Σ_λ , if there exists a sequence of invariant sets $\{\Sigma_{\lambda_n}\}$ such that

$$0 \notin \Sigma_{\lambda_n}, \quad \lim_{n \rightarrow \infty} \lambda_n = \lambda_0, \quad \lim_{n \rightarrow \infty} d(\Sigma_{\lambda_n}, 0) = \lim_{n \rightarrow \infty} \max_{x \in \Sigma_{\lambda_n}} |x| = 0.$$

- If the invariant sets Σ_λ are attractors of (2.2.1), then the bifurcation is called attractor bifurcation.
- If Σ_λ are attractors of (2.2.1) which have the same homology as an m -dimensional sphere, then we say that the system has an S^m -attractor bifurcation at $(0, \lambda_0)$, and Σ_λ is called an m -dimensional homological sphere.

For the eigenvalues $\beta_i(\lambda)$ of A_λ , we assume that

$$\operatorname{Re}\beta_i(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_0, \\ = 0 & \text{if } \lambda = \lambda_0, \\ > 0 & \text{if } \lambda > \lambda_0 \end{cases} \quad \forall 1 \leq i \leq m, \quad (2.2.3)$$

$$\operatorname{Re}\beta_j(\lambda_0) < 0 \quad \forall m+1 \leq j \leq n. \quad (2.2.4)$$

Let E_0 be the eigenspace of A_λ at λ_0 associated with the first m eigenvalues; then $\dim E_0 = m$.

The main results in this section are the following S^m -attractor bifurcation theorems for the finite dimensional system (2.2.1); see Ma and Wang (2005b).

Theorem 2.2.2 Assume that (2.2.2)–(2.2.4) hold, and $x = 0$ is locally asymptotically stable for (2.2.1) at $\lambda = \lambda_0$. Then we have the following assertions.

- (1) The system (2.2.1) bifurcates from $(0, \lambda_0)$ to an attractor Σ_λ for $\lambda > \lambda_0$ with $m - 1 \leq \dim \Sigma_\lambda \leq m$, which is connected when $m \geq 2$.
- (2) There exists a neighborhood $U \subset \mathbb{R}^n$ of $x = 0$ such that Σ_λ attracts U/Γ , where Γ is the stable manifold of $x = 0$ with codimension $\text{codim } \Gamma = m$ in \mathbb{R}^n .
- (3) The bifurcated attractor Σ_λ is an $(m - 1)$ -dimensional homological sphere, more precisely, $\Sigma_\lambda = \bigcap_{k=1}^{\infty} M_k$, where $M_{k+1} \subset M_k$ are the m -dimensional annulus. In particular, if Σ_λ is a finite simplicial complex, then Σ_λ is a deformation retract of m -annulus.
- (4) For any $x_\lambda \in \Sigma_\lambda$, x_λ can be expressed as

$$x_\lambda = z_\lambda + o(|z_\lambda|), \quad z_\lambda \in E_0.$$

- (5) If the number of singular points of (2.2.1) in Σ_λ is finite, then we have the following index formula:

$$\sum_{x \in \Sigma_\lambda} \text{ind}(-(A_\lambda + G), x) = \begin{cases} 2 & \text{if } m = \text{odd}, \\ 0 & \text{if } m = \text{even}. \end{cases}$$

Theorem 2.2.3 Assume that (2.2.2)–(2.2.4) hold true and $x = 0$ is globally asymptotically stable for (2.2.1) at $\lambda = \lambda_0$. Then for any bounded open set $U \subset \mathbb{R}^n$ with $0 \in U$, there is an $\varepsilon > 0$ such that if $\lambda_0 < \lambda < \lambda_0 + \varepsilon$, the attractor Σ_λ of (2.2.1) bifurcated from $(0, \lambda_0)$ attracts $U \setminus \Gamma$, where $\Gamma \subset \mathbb{R}^n$ is the stable manifold of $x = 0$ with $\dim \Gamma = n - m$. In particular, if (2.2.1) has global attractors near $\lambda = \lambda_0$, then Σ_λ attracts $\mathbb{R}^n \setminus \Gamma$.

2.2.2 S^1 -Attractor Bifurcation

We now prove that the bifurcated attractor Σ_λ of (2.2.1) from an eigenvalue with multiplicity two is homeomorphic to S^1 .

Let $v \in C^r(\Omega, \mathbb{R}^{m+1})$ ($r \geq 1$) be an n -dimensional vector field given by

$$v_\lambda(x) = \lambda x - G(x, \lambda) \quad \text{for } x \in \Omega \subset \mathbb{R}^{m+1}, \quad (2.2.5)$$

where $G(x, \lambda) = G_k(x, \lambda) + o(|x|^k)$, and G_k is a k -multilinear field, which satisfies

$$c|x|^{k+1} \leq (G_k(x, \lambda), x) \quad \forall \lambda \in \mathbb{R}^1, \quad (2.2.6)$$

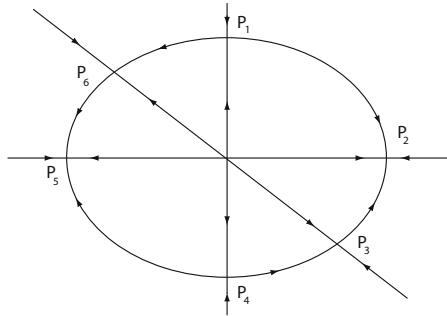
for some constant $c > 0$, and $k = 2p + 1$ with $p \geq 1$.

Theorem 2.2.4 Let $v_\lambda \in C^r(\Omega, \mathbb{R}^2)$ satisfy (2.2.5) and (2.2.6). Then v_λ bifurcates from $(x, \lambda) = (0, 0)$ on $\lambda > 0$ to an attractor Σ_λ , which is homeomorphic to a circle S^1 . Moreover, one and only one of the following is true:

- (1) Σ_λ is a periodic orbit.
- (2) Σ_λ contains some arc segments $\Gamma_i \subset \Sigma_\lambda$ ($1 \leq i \leq l$) such that Γ_i consist of singular points, or
- (3) Σ_λ contains at most $2(k+1) = 4(p+1)$ singular points. Furthermore Σ_λ has an even number, say $2N$, saddle points, and $2N$ number of stable node points (possibly degenerate), and n ($\leq 4(p+1)-4N$) singular points with index zero; see Fig. 2.5 for $N=1$ and $n=2$.

Proof. We proceed in the following five steps.

Fig. 2.5 $\Sigma_\lambda = S^1$ has $4N + n$ ($N = 1, n = 2$) singular points, where P_1, P_4 are saddles, P_2, P_5 are nodes, P_3, P_6 are singular points with index zero



STEP 1. Obviously, (2.2.6) implies that $x = 0$ is asymptotically stable for (2.2.5) at the critical state $\lambda = 0$. Hence, by Theorem 2.2.2, the vector field v_λ bifurcates from $(x, \lambda) = (0, 0)$ to an attractor Σ_λ on $\lambda > 0$, which has the homology type of a circle S^1 .

STEP 2. Let Σ_λ have no singular points. Then, Σ_λ must contain at least a periodic orbit. We need to show that Σ_λ contains only one periodic orbit.

In the polar coordinate system $(x_1, x_2) = (r \cos \theta, r \sin \theta)$, the vector field v_λ becomes

$$\frac{dr}{d\theta} = r \frac{\cos \theta v_1 + \sin \theta v_2}{\cos \theta v_2 - \sin \theta v_1}. \quad (2.2.7)$$

We see that

$$\begin{aligned} \cos \theta v_1 &= \lambda r \cos^2 \theta - \cos \theta g_1(r \cos \theta, r \sin \theta, \lambda), \\ \sin \theta v_2 &= \lambda r \sin^2 \theta - \sin \theta g_2(r \cos \theta, r \sin \theta, \lambda), \\ \cos \theta v_2 &= \lambda r \cos \theta \sin \theta - \cos \theta g_2(r \cos \theta, r \sin \theta, \lambda), \\ \sin \theta v_1 &= \lambda r \sin \theta \cos \theta - \sin \theta g_1(r \cos \theta, r \sin \theta, \lambda), \end{aligned}$$

where $G(x, \lambda) = (g_1(x, \lambda), g_2(x, \lambda))$. Let $g_i(x, \lambda) = g_{ki}(x, \lambda) + o(|x|^k)$ for $i = 1, 2$. By assumption, $g_{ki}(x, \lambda)$ ($i = 1, 2$) are $k = 2m + 1$ multilinear functions. Hence (2.2.7) is rewritten as

$$\frac{dr}{d\theta} = \frac{\lambda - r^{2m}(\cos \theta g_{k1} + \sin \theta g_{k2}) + o(r^{2m})}{r^{2m-1}(\sin \theta g_{k1} - \cos \theta g_{k2}) + o(r^{2m-1})}. \quad (2.2.8)$$

Based on (2.2.6), we have

$$0 < c \leq \cos \theta g_{k1}(\cos \theta, \sin \theta, \lambda) + \sin \theta g_{k2}(\cos \theta, \sin \theta, \lambda). \quad (2.2.9)$$

On the other hand, by assumption, Σ_λ contains a periodic orbit for any $\lambda > 0$ sufficiently small. Hence there exists $C > 0$ such that

$$0 < C \leq \sin \theta g_{k1}(\cos \theta, \sin \theta, \lambda) - \cos \theta g_{k2}(\cos \theta, \sin \theta, \lambda), \quad (2.2.10)$$

for any $0 \leq \theta \leq 2\pi$ and $\lambda > 0$ small. The condition (2.2.10) amounts to saying that the orbits of v_λ are circular around $x = 0$.

Let $r(\theta, r_0)$ be the solution of (2.2.8) with initial value $r(0, r_0) = r_0$. Then we have the Taylor expansion

$$r^{2m}(\theta, r_0) = r_0^{2m} + R(\theta) \cdot o(r_0^{2m}), \quad R(0) = 0. \quad (2.2.11)$$

It follows from (2.2.8) and (2.2.11) that

$$\begin{aligned} \frac{1}{2m} [r^{2m}(2\pi, r_0) - r^{2m}(0, r_0)] &= \int_0^{2\pi} \frac{\lambda - r^{2m}\alpha(\theta) + o(r^{2m})}{\beta(\theta) + o(1)} d\theta \\ &= a\lambda - br_0^{2m} + o(r_0^{2m}), \end{aligned} \quad (2.2.12)$$

where

$$\begin{aligned} a &= \int_0^{2\pi} \frac{1}{\beta(\theta) + o(1)} d\theta, & b &= \int_0^{2\pi} \frac{\alpha(\theta)}{\beta(\theta) + o(1)} d\theta, \\ \alpha(\theta) &= \cos \theta g_{k1} + \sin \theta g_{k2}, & \beta(\theta) &= \sin \theta g_{k1} - \cos \theta g_{k2}. \end{aligned}$$

From (2.2.12) we see that the periodic solutions of v_λ near $x = 0$ correspond to positive solutions of

$$a\lambda - br_0^{2m} + o(r_0^{2m}) = 0. \quad (2.2.13)$$

By (2.2.9) and (2.2.10), $a > 0, b > 0$. Therefore, (2.2.13) has a unique positive solution near $r = 0$:

$$r_0 = \left(\frac{a\lambda}{b} \right)^{1/2m} + o(\lambda^{1/2m}),$$

for any $\lambda > 0$ sufficiently small. Thus, Σ_λ consists of a unique periodic orbit.

STEP 3. In the case where Σ_λ contains a singular point, we assert that Σ_λ contains either a finite number of singular points or a few arc segments $\Gamma_i \cap \Sigma_\lambda$ ($1 \leq i \leq l$) consisting of singular points, and if it contains finite number of singular points, then there are at most $2(k+1)$ of them near $x = 0$.

In fact, if

$$\frac{g_1(x, \lambda)}{g_2(x, \lambda)} = \frac{x_1}{x_2} + f(x, \lambda),$$

where $f(x, \lambda)$ are cut-off function as follows

$$f(r, \theta) \begin{cases} = 0 & \forall \theta \in [\theta_i, \theta_{i+1}], \\ \neq 0 & \theta \notin [\theta_i, \theta_{i+1}] \end{cases} \quad \forall 1 \leq i \leq l,$$

where (r, θ) is the polar coordinate system, then Σ_λ contains l ($l \geq 1$) arc segments Γ_i which consist of singular points. In particular if $f(x, \lambda) = 0$, namely

$$\frac{g_1(x, \lambda)}{g_2(x, \lambda)} = \frac{x_1}{x_2}, \quad \forall x \in \mathbb{R}^2, \quad (2.2.14)$$

then Σ_λ has a cycle of singular points. Otherwise, by (2.2.6), the number of singular points of v_λ is finite. The maximal number of singular points for v_λ is determined by the following equation

$$\lambda x - G_k(x, \lambda) = 0. \quad (2.2.15)$$

Since G_k is a k -multilinear vector field, the singular points of (2.2.15) must be on the straight lines $x_2 = zx_1$, where z satisfies

$$z = \frac{g_{k2}(x_1, x_2, \lambda)}{g_{k1}(x_1, x_2, \lambda)} = \frac{g_{k2}(1, z, \lambda)}{g_{k1}(1, z, \lambda)}. \quad (2.2.16)$$

The number of solutions of (2.2.16) is at most $k + 1$. Since $k = \text{odd}$, the number of solutions of (2.2.15) is at most $2(k + 1)$.

STEP 4. Let Σ_λ contain a cycle S^1 of singular points, then we shall see that $\Sigma_\lambda = S^1$. Under the polar coordinate system, we have

$$v_r(\theta, r) = (v_\lambda(x), x) = \lambda r^2 - r^{k+1} \alpha(\theta) + o(r^{k+1}),$$

where $\alpha(\theta)$ is as in (2.2.12). By (2.2.6),

$$0 < c \leq \alpha(\theta) \leq c_1, \quad \forall 0 \leq \theta \leq 2\pi.$$

It is clear that for each θ ($0 \leq \theta \leq 2\pi$), v_r has a unique zero point

$$r_\lambda(\theta) = (\lambda/\alpha(\theta))^{1/k-1} + o(|\lambda|^{1/k-1}),$$

near $r = 0$ for $\lambda > 0$. Hence, the set

$$\tilde{\Omega}_\lambda = \{(\theta, r_\lambda(\theta)) \mid v_r(\theta, r_\lambda(\theta)) = 0, 0 \leq \theta \leq 2\pi\}$$

is homeomorphic to a cycle S^1 , and all singular points of v_λ near $x = 0$ are in $\tilde{\Omega}_\lambda$. It implies that $\tilde{\Omega}_\lambda \subset \Sigma_\lambda$. Let

$$g_{kl} = \sum_{i+j=k} a_{ij}^l x_1^i x_2^j, \quad l = 1, 2. \quad (2.2.17)$$

By (2.2.14), we know that $x_2 g_{k1} = x_1 g_{k2}$. Then we infer from (2.2.6) that

$$0 < \alpha_{k0}^1 = \alpha_{k-11}^2. \quad (2.2.18)$$

At the singular point $(\tilde{x}_1, 0) \in \tilde{\Omega}_\lambda$ of v_λ , for any $\lambda > 0$ sufficiently small, we have

$$\begin{aligned}\operatorname{div} v_\lambda(\tilde{x}_1, 0) &= 2\lambda - k\alpha_{k0}^1 \tilde{x}_1^{k-1} - \alpha_{k-11}^2 \tilde{x}_1^{k-1} + o(|\tilde{x}_1|^{k-1}) \\ &= (\text{by } \alpha_{k0}^1 \tilde{x}_1^{k-1} = \lambda \text{ and (2.2.18)}) \\ &= -(k-1)\lambda + o(|\lambda|) \\ &< 0.\end{aligned}$$

In the same fashion, for any point $\tilde{x} \in \tilde{\Omega}_\lambda$, we can take an orthogonal system transformation such that \tilde{x} is on the x_1 -axis, then we can prove that

$$\operatorname{div} v_\lambda(x) < 0, \quad \forall x \in \tilde{\Omega}_\lambda,$$

which implies that $\Sigma_\lambda = \tilde{\Omega}_\lambda = S^1$.

STEP 5. Σ_λ contains finite number of singular points. We prove that $\Sigma_\lambda = S^1$.

By the Brouwer degree theory, it follows from (2.2.6) that

$$\deg(v_\lambda, \Omega, 0) = 1 \quad \forall |\lambda| > 0 \text{ sufficiently small},$$

in some neighborhood $\Omega \subset \mathbb{R}^2$ of $x = 0$. It is known that

$$\operatorname{ind}(v_\lambda, 0) = 1, \quad \forall |\lambda| \neq 0.$$

Hence, we have

$$\sum_{z_i \in \Sigma_\lambda} \operatorname{ind}(v_\lambda, z_i) = 0. \quad (2.2.19)$$

Let $z \in \Sigma_\lambda$ be a singular point of v_λ . Without loss of generality, we take the orthogonal coordinate system such that $z = (x_1, 0)$. Then by (2.2.6) and (2.2.17), the Jacobian matrix of v_λ at z is given by

$$Dv_\lambda(z) = \begin{pmatrix} -(k-1)\lambda + o(\lambda) & * \\ 0 & (1 - \alpha_{k-11}^2 / \alpha_{k0}^1)\lambda \end{pmatrix},$$

where $\alpha_{k0}^1 > 0$. Obviously, $Dv_\lambda(z)$ has an eigenvalue $\beta = -(k-1)\lambda + o(\lambda) < 0$. Hence, for any singular point $z \in \Sigma_\lambda$ of v_λ , the index of v_λ at z can only be either 1, -1, or 0. It is easy to see that if the index is 1, then z is a stable node point.

Let the index of v_λ at z be -1:

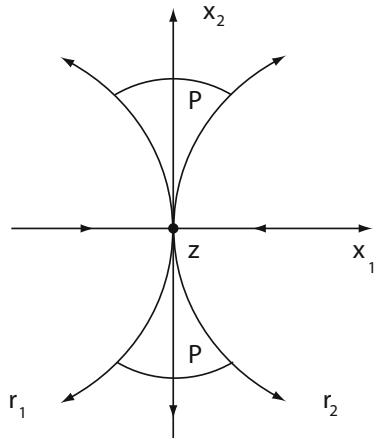
$$\operatorname{ind}(v_\lambda, z) = -1. \quad (2.2.20)$$

When $\alpha_{k-11}^2 \neq \alpha_{k0}^1$, z is nondegenerate. Therefore, v_λ has a unique unstable manifold at z . When $\alpha_{k-11}^2 = \alpha_{k0}^1$,

$$\operatorname{div} v_\lambda(z) = -(k-1)\lambda + o(\lambda) < 0. \quad (2.2.21)$$

If the unstable manifold of v_λ at z is not unique, then the local structure of v_λ at z is topologically equivalent to that as shown in Fig. 2.6.

Fig. 2.6 r_1 and r_2 are unstable manifolds of v_λ at z , the flow in P is outward from z



On the other hand, (2.2.21) means that there is a neighborhood $O \subset \mathbb{R}^2$ of $z = 0$, such that

$$\operatorname{div} v_\lambda(x) < 0, \quad \forall x \in O,$$

which implies that for any open set $\tilde{O} \subset O$, the area of \tilde{O} satisfies

$$|\tilde{O}| > |\tilde{O}_t|, \quad \forall 0 < t < t_0, \quad (2.2.22)$$

where $t_0 > 0$ depends on \tilde{O} , $\tilde{O}_t = S(t)\tilde{O}$, and $S(t)$ is the operator semigroup generated by v_λ .

However, it is clear that for any open set $\tilde{O} \subset O$ in a domain P as shown in Fig. 2.6, the property (2.2.22) is not true. Hence, the unstable manifold of v_λ at z must be unique.

We can prove in the same fashion that if the index of v_λ at z is zero, the unstable manifold of v_λ at z is also unique.

By the Poincare-Bendixson theorem, all unstable manifolds of the singular points of v_λ with indices -1 and 0 are connected to the singular points with indices 1 and 0 as shown in Fig. 2.5. Thus by the uniqueness of unstable manifolds for each singular point with either index -1 or index 0 , the set of all singular points and unstable manifolds is a cycle S^1 , i.e., $\Sigma_\lambda = S^1$ because Σ_λ is homological to S^1 .

STEP 6. Finally, if

$$\Sigma_\lambda = \bigcup_{i=1}^r \Gamma_i + \Sigma,$$

where Γ_i are arc segments consisting of singular points and Σ contains only finite number of singular points, we can apply the methods as used in Steps 4 and 5 to prove that $\Sigma_\lambda = S^1$.

Thus the theorem is proved. \square

More generally, consider a two-dimensional vector field given by

$$v(x, \lambda) = \begin{pmatrix} \beta_1(\lambda) & 0 \\ 0 & \beta_2(\lambda) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + G(x, \lambda), \quad G(x, \lambda) = o(|x|), \quad (2.2.23)$$

where

$$\beta_i(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_0, \\ = 0 & \text{if } \lambda = \lambda_0, \\ > 0 & \text{if } \lambda > \lambda_0 \end{cases} \quad \text{for } i = 1, 2. \quad (2.2.24)$$

In the polar coordinate system (r, θ) , the projection v_r of v on r -direction is expressed as

$$v_r(r, \theta, \lambda) = (v(x, \lambda), \hat{x}) = r(\beta_1(\lambda) \cos^2 \theta + \beta_2(\lambda) \sin^2 \theta) - g(r, \theta, \lambda),$$

where $\hat{x} = (\cos \theta, \sin \theta)$. Assume that for some $\delta > 0$,

$$(G(x, \lambda), x) = g(x, \lambda) > 0 \quad \forall 0 < |x| < \delta. \quad (2.2.25)$$

If for any $\lambda - \lambda_0 > 0$ sufficiently small and $0 \leq \theta \leq 2\pi$ there exists a unique $r_\theta > 0$ satisfies that

$$\begin{aligned} r_\theta(\beta_1 \cos^2 \theta + \beta_2 \sin^2 \theta) - g(r_\theta, \theta, \lambda) &= 0, \quad 0 \leq \theta \leq 2\pi, \\ \frac{\partial g(r_\theta, \theta, \lambda)}{\partial r} &> \beta_1(\lambda) \cos^2 \theta + \beta_2(\lambda) \sin^2 \theta, \end{aligned} \quad (2.2.26)$$

then the vector field $v(x, \lambda)$ given by (2.2.23) bifurcates from $(x, \lambda) = (0, \lambda_0)$ on $\lambda > \lambda_0$ to a circle attractor $\Sigma_\lambda = S^1$.

In fact, by Theorem 2.2.2, it follows from (2.2.24) and (2.2.25) that $v(x, \lambda)$ bifurcates from $(0, \lambda_0)$ to an attractor Σ_λ which is homological to a circle S^1 , i.e., Σ_λ may be one of the cases as shown in Fig. 2.7a–c.

Intuitively, the conditions (2.2.24)–(2.2.26) exclude the cases as shown in Fig. 2.7b, c.

To see this, we know that the function $r_\theta = r(\theta, \lambda)$ satisfying (2.2.26) represents a circle S^1 , i.e., the set

$$\Sigma_\lambda = \{(r(\theta, \lambda), \theta) \in \mathbb{R}^2 \mid 0 \leq \theta \leq 2\pi\}$$

is homeomorphic to S^1 , and all singular points of v in Σ_λ are also in Σ_λ , which implies, by (2.2.26), that for any singular point $x_0 \in \Sigma_\lambda$ of v , the Jacobian matrix $Dv(x_0)$ has at least an eigenvalue λ such that $\operatorname{Re}\lambda < 0$. It follows that the unstable manifold of a singular point of v in Σ_λ is unique. Thus, $\Sigma_\lambda = S^1$.

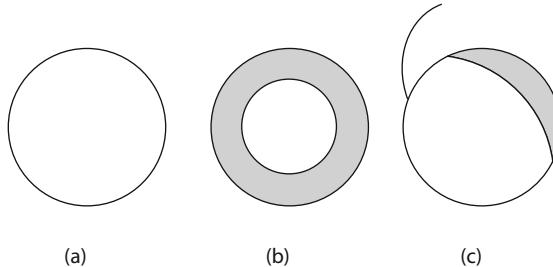


Fig. 2.7 (a) A circle S^1 , (b) an annulus, and (c) a set having the homotopical type of S^1

2.2.3 S^m -Attractor Bifurcation

We now consider the attractor bifurcation of m -dimensional sphere S^m for the vector fields $v \in C^r(\Omega, \mathbb{R}^{m+1})$ given by (2.2.5). Let

$$v(x) = Ax + G(x) \quad \forall x \in \Omega \subset \mathbb{R}^{m+1},$$

where A is an $(m+1) \times (m+1)$ matrix, and $G(x) = o(|x|)$. The following is the S^m attractor bifurcation theorem; see Ma and Wang (2007b):

Theorem 2.2.5 *Let $v_\lambda \in C^r(\Omega, \mathbb{R}^{m+1})$ satisfy (2.2.5) and (2.2.6). Then v bifurcates from $(x, \lambda) = (0, 0)$ to an attractor Σ_λ on $\lambda > 0$, which is homeomorphic to an m -dimensional sphere S^m .*

2.2.4 Structural Stability of Dynamic Transitions

In this subsection, we introduce the stability of dynamic bifurcation for the parameterized vector field $v(x, \lambda)$, which is not only the stability of bifurcated branches, but also the local structural stability of bifurcated solutions. We shall see later that the stability of dynamic bifurcation is important in the transition theory of the perturbed systems, which is related to the phase transition stability in science and engineering.

Let $\Omega \subset \mathbb{R}^n$ be an open set, $x = 0 \in \Omega$ and $\lambda \in I = (a, b) \subset \mathbb{R}^1$. We denote the space of parameterized vector fields by

$$C_0^{k,1}(\Omega \times I, \mathbb{R}^n) = \{v : \Omega \times I \rightarrow \mathbb{R}^n \mid v(0, \lambda) = 0, \quad \forall \lambda \in I\}$$

endowed with norm

$$\|v\|_{C^{k,1}} = \sup_{(x, \lambda) \in \Omega \times I} \left[\sum_{p=0}^k |D_x^p v| + |D_\lambda v| + |D_{\lambda x}^2 v| \right].$$

Obviously, if $v \in C_0^{k,1}(\Omega \times I, \mathbb{R}^n)$ ($k \geq 1$), then v is k -th differentiable at $x \in \Omega$ and differentiable at $\lambda \in I$, and v can be expressed as

$$v(\cdot, \lambda) = A_\lambda + G(\cdot, \lambda), \quad (2.2.27)$$

where A_λ and $G(\cdot, \lambda)$ are the same as in (2.2.1) and (2.2.2).

It is known that the simple real and complex eigenvalues of A_λ , are differentiable on λ (Kato, 1995), which can be expressed by

$$\beta_i(\lambda) = \alpha_i + \sigma_i \lambda + o(|\lambda|) \quad \forall 1 \leq i \leq n.$$

In the following, we introduce the definitions of stability of dynamic bifurcation.

Definition 2.2.6 Let $v_1, v_2 \in C_0^{k,1}(\Omega \times I, \mathbb{R}^n)$ and $\lambda_i \in I$ ($i = 1, 2$) be a bifurcation point of invariant set $\Gamma_i(\rho)$ of $v_i(x, \rho)$. We say that both bifurcation points λ_1 and λ_2 have the same structure if v_1 and v_2 are locally topologically equivalent at $\Gamma_1(\rho_1)$ and $\Gamma_2(\rho_2)$ with $\rho_1 - \lambda_1 = \rho_2 - \lambda_2$, i.e., there are neighborhoods $U_i \subset \mathbb{R}^n$ of $\Gamma_i(\rho)$ and a homeomorphism of $\varphi : U_1 \rightarrow U_2$ such that φ takes the orbits of v_1 in U_1 to orbits of v_2 in U_2 preserving orientation.

Definition 2.2.7 Let $v \in C_0^{k,1}(\Omega \times I, \mathbb{R}^n)$, $\lambda_0 \in I$ be a bifurcation point of $v(x, \lambda)$. We say that λ_0 is stable in $C_0^{k,1}(\Omega \times I, \mathbb{R}^n)$ if for any $\varepsilon > 0$ sufficiently small there is a neighborhood $O \subset C_0^{k,1}(\Omega \times I, \mathbb{R}^n)$ of v such that for any $u \in O$, u has a bifurcation point λ_1 with $|\lambda_1 - \lambda_0| < \varepsilon$, and λ_1 and λ_0 have the same structure. If all bifurcation points of v are stable, we say that v is stable for bifurcation in $C_0^{k,1}(\Omega \times I, \mathbb{R}^n)$.

Let $v \in C_0^{k,1}(\Omega \times I, \mathbb{R}^n)$ ($k \geq 1$). The point $\lambda_0 \in I$ is called an eigen-parameter with multiplicity m if the eigenvalues $\beta_i(\lambda)$ of $A_\lambda = Dv(0, \lambda)$ satisfy

$$\begin{aligned} \operatorname{Re}\beta_i(\lambda_0) &= 0 && \text{if } 1 \leq i \leq m, \\ \operatorname{Re}\beta_j(\lambda_0) &\neq 0 && \text{if } m+1 \leq j \leq n. \end{aligned} \quad (2.2.28)$$

$\lambda_0 \in I$ is simple if $m = 1$ for $\beta_1(\lambda_0) = 0$ and $m = 2$ for $\beta_1(\lambda_0) = \bar{\beta}_2(\lambda_0) = i\beta(\beta \neq 0)$. If $\lambda_0 \in I$ is simple, and the eigenvalues in (2.2.28) satisfy

$$\frac{d}{d\lambda} \operatorname{Re}\beta_i(\lambda_0) \neq 0 \quad \forall 1 \leq i \leq m, \quad (2.2.29)$$

then λ_0 is called a regular eigen-parameter.

The following two theorems provide sufficient and necessary conditions for structural stability of dynamic bifurcation of parameterized vector fields in $C_0^{3,1}(\Omega \times I, \mathbb{R}^n)$; they are proved in Ma and Wang (2005b).

Theorem 2.2.8 Let $v \in C_0^{3,1}(\Omega \times I, \mathbb{R}^n)$ and $\lambda_0 \in I$ be a simple eigen-parameter of v . Then there exists a number $b(\lambda_0)$, called the bifurcation number of v at λ_0 , which continuously depends on $\|v\|_{C^{3,1}}$, such that the following assertions hold true.

- (1) λ_0 is a stable bifurcation point of v if and only if λ_0 is regular and the bifurcation number $b(\lambda_0) \neq 0$.
- (2) If λ_0 has multiplicity $m = 1$ and $b(\lambda_0) \neq 0$, then $v(x, \lambda)$ bifurcates to a unique branch of singular points $x(\lambda)(\lambda \neq \lambda_0)$ on each side of $\lambda = \lambda_0$, which are hyperbolic, and $x(\lambda)$ has Morse index $k + 1$ for $\beta_1(\lambda) < 0$ and Morse index k for $\beta_1(\lambda) > 0$ where $k = \text{number of the eigenvalues with } \text{Re}\beta_j(\lambda_0) > 0$.
- (3) If λ_0 has multiplicity $m = 2$, then $v(x, \lambda)$ bifurcates to a unique branch of periodic orbits Γ_λ for $\text{Re}\beta_1(\lambda) < 0$ as $b(\lambda_0) > 0$ and for $\text{Re}\beta_1(\lambda) > 0$ as $b(\lambda_0) < 0$, which are hyperbolic.

Theorem 2.2.9 A parameterized vector field $v \in C_0^{3,1}(\Omega \times I, \mathbb{R}^n)$ is structurally stable for bifurcation if and only if

- (1) all eigen-parameters of v are simple and regular,
- (2) each bifurcation number of v is nonzero.

Moreover, the set of all vector fields with stable bifurcation is open and dense in $C_0^{3,1}(\Omega \times I, \mathbb{R}^n)$.

Remark 2.2.10 We note that in $C_0^{1,1}(\Omega \times I, \mathbb{R}^n)$ there are no vector fields which are stable for bifurcation except the vector fields which have no eigen-parameters in I . In $C_0^{2,1}(\Omega \times I, \mathbb{R}^n)$ the set of all vector fields with stable bifurcation are those whose eigen-parameter are regular with multiplicity $m = 1$.

2.2.5 Infinite Dimensional Systems

For infinite dimensional systems, consider (2.1.1). Let the eigenvalues of L_λ be given by $\{\beta_j(\lambda) \mid j = 1, 2, \dots\} \subset \mathbb{C}$. Assume the conditions (2.1.4) and (2.1.5) hold true. Let E_0 be the eigenspace of L_λ at $\lambda = \lambda_0$; then $\dim E_0 = m$.

The following theorem generalizes Theorems 2.2.2 and 2.2.3 to the infinite dimensional system (2.1.1), provides sufficient conditions for continuous transitions, and gives local transition structure.

Theorem 2.2.11 Under the conditions (2.1.4) and (2.1.5), if $u = 0$ is a locally asymptotically stable equilibrium point of (2.1.1), then we have the following assertions:

1. Equation (2.1.1) bifurcates from $(u, \lambda) = (0, \lambda_0)$ on $\lambda > \lambda_0$ to an attractor \mathcal{A}_λ with $m - 1 \leq \dim \mathcal{A}_\lambda \leq m$, i.e., (2.1.1) has a continuous transition at $\lambda = \lambda_0$, and \mathcal{A}_λ is connected if $m \geq 2$.
2. There exists a neighborhood $U \subset X$, such that \mathcal{A}_λ attracts $U \setminus \Gamma$, where Γ is the stable manifold of $u = 0$ with $\text{codim } \Gamma = m$.
3. \mathcal{A}_λ is an $(m - 1)$ -dimensional homological sphere, in particular, if \mathcal{A}_λ is finite simplicial complex, then \mathcal{A}_λ has the homotopy type of the $(m - 1)$ -dimensional sphere S^{m-1} .

4. For any $u_\lambda \in \mathcal{A}_\lambda$, u_λ can be expressed as

$$u_\lambda = v_\lambda + o(\|v_\lambda\|), \quad v_\lambda \in E_0.$$

5. If the number of the singular points of (2.1.1) in \mathcal{A}_λ is finite, then we have the index formula

$$\sum_{u_i \in \Sigma_\lambda} \text{ind}(-(L_\lambda + G), u_i) = \begin{cases} 2 & \text{if } m = \text{odd}, \\ 0 & \text{if } m = \text{even}. \end{cases}$$

6. If $u = 0$ is globally asymptotically stable for (2.1.1) at $\lambda = \lambda_0$, then for any bounded open set $U \subset H$, there is an $\varepsilon > 0$ such that as $\lambda_0 < \lambda < \lambda_0 + \varepsilon$, the attractor \mathcal{A}_λ attracts $U \setminus \Gamma$ in H . In particular, if (2.1.1) has a global attractor for all λ near λ_0 , then \mathcal{A}_λ attracts $H \setminus \Gamma$.

Proof. By the reduction of (2.1.1) on the center manifold, the bifurcation equation of (2.1.1) is given by (1.2.7). By (2.1.3) and the center manifold function is of high order, i.e., $\Phi(x, \lambda) = o(\|x\|)$, we have

$$PG(x, \Phi(x, \lambda), \lambda) = o(\|x\|).$$

Since $u = 0$ is asymptotically stable for (2.1.1), $x = 0$ is local asymptotically stable singular point of (1.2.7) at $\lambda = \lambda_0$. Hence, Assertions (1)–(5) follow from Theorem 2.2.2, and Assertion (6) can be deduced by Theorem 5.4 in Ma and Wang (2005b). The proof is complete. \square

When the multiplicity in (2.1.4) and (2.1.5) is $m = 1$, then we have the pitchfork attractor bifurcation theorem for (2.1.1) as follows.

Theorem 2.2.12 Under the assumptions of Theorem 2.2.11, if $m = 1$ in (2.1.4) and $G(u, \lambda)$ is analytic at $u = 0$, then there exists an open set $U \subset X$ with $0 \in U$ such that (2.1.1) bifurcates from $(u, \lambda) = (0, \lambda_0)$ on $\lambda > \lambda_0$ to exactly two singular points $u_1(\lambda)$ and $u_2(\lambda) \in U$, and U can be decomposed into two open sets U_1^λ and U_2^λ satisfying the following properties:

- (1) $\overline{U} = \overline{U}_1^\lambda + \overline{U}_2^\lambda$, $U_1^\lambda \cap U_2^\lambda = \emptyset$, $\Gamma = \partial U_1^\lambda \cap \partial U_2^\lambda$ is the stable manifold of $u = 0$, and $u_i(\lambda) \in U_i^\lambda$ ($i = 1, 2$).
- (2) Let $u(t, \varphi)$ be the solution of (2.1.1), then

$$\lim_{n \rightarrow \infty} \|u(t, \varphi) - u_i(\lambda)\| = 0 \quad \forall \varphi \in U_i^\lambda \quad (i = 1, 2).$$

Proof. Assertion (1) and (2) can be deduced from Theorem 5.5 in Ma and Wang (2005b), and the conclusion that (2.1.1) bifurcates to exactly two singular points can be directly derived from the Krasnoselski theorem; see Theorem 1.11 in Ma and Wang (2005b). The proof is complete. \square

2.3 Transition from Simple Eigenvalues

2.3.1 Real Simple Eigenvalues

In this section, we consider the transition of (2.1.1) from a simple critical eigenvalue. Let the eigenvalues $\beta_j(\lambda)$ of L_λ satisfy (2.1.4) and (2.1.5) with $m = 1$. In this case, $\beta_1(\lambda)$ is a real eigenvalue. Let $e_1(\lambda)$ and $e_1^*(\lambda)$ be the eigenvectors of L_λ and L_λ^* respectively corresponding to $\beta_1(\lambda)$ with

$$L_{\lambda_0} e_1 = 0, \quad L_{\lambda_0}^* e_1^* = 0, \quad (e_1, e_1^*) = 1.$$

Let $\Phi(x, \lambda)$ be the center manifold function of (2.1.1) near $\lambda = \lambda_0$, and assume that

$$(G(xe_1 + \Phi(x, \lambda_0), \lambda_0), e_1^*) = \alpha x^k + o(|x|^k), \quad (2.3.1)$$

where $k \geq 2$ an integer and $\alpha \neq 0$ a real number.

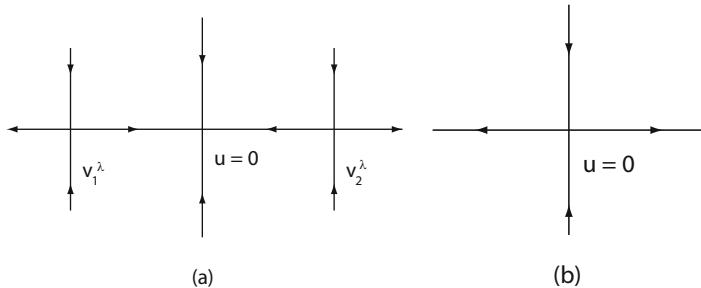


Fig. 2.8 Topological structure of jump transition of (2.1.1) when $k=\text{odd}$ and $\alpha > 0$: (a) $\lambda < \lambda_0$; (b) $\lambda \geq \lambda_0$. Here the horizontal line represents the center manifold

Theorem 2.3.1 Let (2.1.4) and (2.1.5) with $m = 1$, and (2.3.1) hold true. If $k=\text{odd}$ and $\alpha \neq 0$ in (2.3.1), then the following assertions hold true:

- (1) If $\alpha > 0$, then (2.1.1) has a jump transition from $(0, \lambda_0)$, and bifurcates on $\lambda < \lambda_0$ to exactly two saddle points v_1^λ and v_2^λ with the Morse index one, as shown in Fig. 2.8.
- (2) If $\alpha < 0$, then (2.1.1) has a continuous transition from $(0, \lambda_0)$, which is an attractor bifurcation as described in Theorem 2.2.12, and as shown in Fig. 2.9.
- (3) The bifurcated solutions v_1^λ and v_2^λ can be expressed as

$$v_{1,2}^\lambda = \pm |\beta_1(\lambda)/\alpha|^{1/(k-1)} e_1(\lambda) + o(|\beta_1|^{1/(k-1)}).$$

Theorem 2.3.2 Let (2.1.4) and (2.1.5) with $m = 1$, and (2.3.1) hold true. If $k=\text{even}$ and $\alpha \neq 0$, then we have the following assertions:

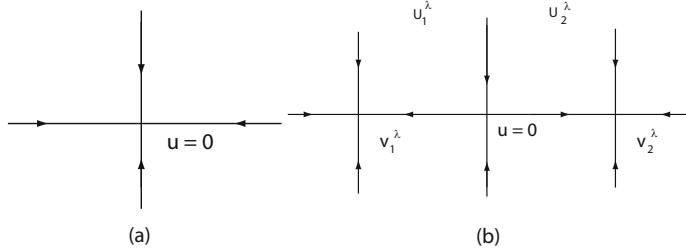


Fig. 2.9 Topological structure of continuous transition of (2.1.1) when $k=$ odd and $\alpha < 0$: (a) $\lambda \leq \lambda_0$; (b) $\lambda > \lambda_0$

1. Equation (2.1.1) has a mixed transition from $(0, \lambda_0)$. More precisely, there exists a neighborhood $U \subset X$ of $u = 0$ such that U is separated into two disjoint open sets U_1^λ and U_2^λ by the stable manifold Γ_λ of $u = 0$ satisfying that the local transition structure is as shown in Fig. 2.10, and

- a. $U = U_1^\lambda + U_2^\lambda + \Gamma_\lambda$,
 - b. the transition in U_1^λ is jump, and
 - c. the transition in U_2^λ is continuous.
2. Equation (2.1.1) bifurcates in U_2^λ to a unique singular point v^λ on $\lambda > \lambda_0$, which is an attractor such that for any $\varphi \in U_2^\lambda$,
- $$\lim_{t \rightarrow \infty} \|u(t, \varphi) - v^\lambda\| = 0.$$
3. Equation (2.1.1) bifurcates on $\lambda < \lambda_0$ to a unique saddle point v^λ with Morse index one.
4. The bifurcated singular point v^λ can be expressed as

$$v^\lambda = -(\beta_1(\lambda)/\alpha)^{1/(k-1)} e_1 + o(|\beta_1|^{1/(k-1)}).$$

Remark 2.3.3 Theorem 2.3.2 shows that for the mixed transition from real simple eigenvalues, there is a bifurcated attractor in the stable domain, which is a singular point. However, we shall see later that in general this property is not true. In other words, a mixed transition may have no bifurcated attractor in its stable domain.

Remark 2.3.4 The local topological structure of the transitions of (2.1.1) from simple eigenvalues is schematically shown in the center manifold in Figs. 2.8 and 2.9.

Proof of Theorems 2.3.1 and 2.3.2. Under the assumption, the reduction equation of (2.1.1) on the center manifold is given by

$$\frac{dx}{dt} = \beta_1(\lambda)x + (G(xe_1(\lambda) + \Phi(x, \lambda)), e_1^*(\lambda)). \quad (2.3.2)$$

By (2.3.1), the equation (2.3.2) is rewritten as

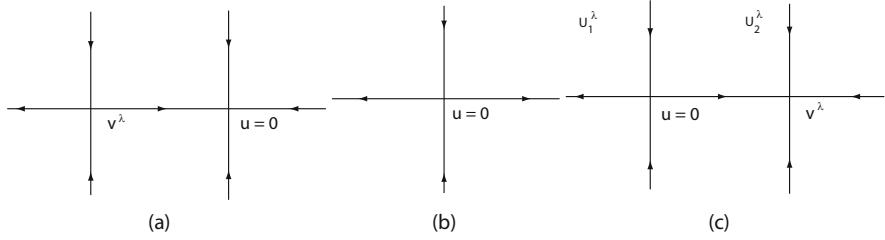


Fig. 2.10 Topological structure of mixed transition of (2.1.1) when $k=\text{even}$ and $\alpha \neq 0$: (a) $\lambda < \lambda_0$; (b) $\lambda = \lambda_0$; (c) $\lambda > \lambda_0$, U_1^λ is the unstable domain, and U_2^λ the stable domain

$$\frac{dx}{dt} = \beta_1(\lambda)x + \alpha_\lambda x^k + o(|x|^k) \quad \text{for } x \in \mathbb{R}^1, \quad (2.3.3)$$

where $\lim_{\lambda \rightarrow \lambda_0} \alpha_\lambda = \alpha \neq 0$.

It is known that the transition of (2.1.1) and its local topological structure are determined completely by (2.3.3). Obviously, when $k=\text{odd}$ and $\alpha > 0$, for $\lambda > \lambda_0$ ($\beta_1(\lambda) > 0$), the solutions of (2.3.3) satisfy that

$$\limsup_{t \rightarrow \infty} |x(t, x_0)| > \delta > 0 \quad \forall 0 < |x_0| < \varepsilon, \quad (2.3.4)$$

where $\delta > 0$ is a constant, and $x(0, x_0) = x_0$. From (2.3.4) we know that the transition of (2.3.3) at $\lambda = \lambda_0$ is jump, which implies that the transition of (2.1.1) from $(0, \lambda_0)$ is jump.

It is clear that (2.3.3) has exactly two bifurcated solutions on $\lambda < \lambda_0$ ($\beta_1(\lambda) < 0$) as follows

$$x_\pm = \pm |\beta_1(\lambda)/\alpha_\lambda|^{1/(k-1)} + o(|\beta_1(\lambda)|^{1/(k-1)}).$$

Therefore

$$v_{1,2}^\lambda = x_\pm e_1 + \Phi(x_\pm, \lambda) \quad (2.3.5)$$

is the bifurcated singular points of (2.1.1) on $\lambda < \lambda_0$. By Theorem 3.10 in Ma and Wang (2005b), the Morse index of v_i^λ ($i = 1, 2$) is one. Thus Assertion (1) of Theorem 2.3.1 is proved, and the expression (2.3.5) is as in Assertion (3).

Other conclusions and Theorem 2.3.2 can be derived from (2.3.3) in the same fashion, and we omit the details. The proof is complete. \square

We remark that Theorems 2.3.1 and 2.3.2 are very convenient for applications. However, a complete theory for the transition from real simple eigenvalues is given by the following theorem together with Theorems 2.3.1 and 2.3.2.

Let $U \subset X$ be a neighborhood of $u = 0$. An open set $D \subset U$ is called a connected unstable domain if for any $\varphi \in D$, the solution $u_\lambda(t, \varphi)$ of (2.1.1) satisfies

$$\limsup_{t \rightarrow \infty} \|u_\lambda(t, \varphi)\| \geq \delta > 0 \quad \forall \lambda > \lambda_0,$$

and ∂D is an invariant set of (2.1.1) with

$$\lim_{\lambda \rightarrow \lambda_0} \lim_{t \rightarrow \infty} \|u_\lambda(t, \varphi)\| = 0 \quad \forall \varphi \in \partial D.$$

Theorem 2.3.5 Let (2.1.4) and (2.1.5) with $m = 1$ hold true. The following assertions hold true for (2.1.1):

- (1) The transition is continuous if and only if the bifurcation of singular points is on the side of $\lambda > \lambda_0$, and this transition is an attractor bifurcation.
- (2) The transition is mixed if and only if there exists a bifurcation of singular points on each side of $\lambda = \lambda_0$, and there are exactly one connected unstable domain and one connected stable domain, and (2.1.1) bifurcates on $\lambda > \lambda_0$ to an attractor which attracts the stable domain.
- (3) This transition is jump if and only if there is no bifurcation of singular points on the side of $\lambda > \lambda_0$, i.e., the bifurcation only can occur on $\lambda = \lambda_0$ or $\lambda < \lambda_0$, and there are exactly two connected unstable domains.

Theorem 2.3.5 can be directly derived from the reduction equation (2.3.2) of (2.1.1) on the center manifold, and we omit the details here.

2.3.2 Transitions from Complex Simple Eigenvalues

Dynamic Transitions We now study the transition from a pair of complex eigenvalues. Assume that the eigenvalues of L_λ satisfy

$$\begin{cases} \operatorname{Re}\beta_1(\lambda) = \operatorname{Re}\beta_2(\lambda) \\ \operatorname{Im}\beta_1(\lambda_0) = -\operatorname{Im}\beta_2(\lambda_0) \neq 0, \end{cases} \begin{cases} < 0 & \text{if } \lambda < \lambda_0, \\ = 0 & \text{if } \lambda = \lambda_0, \\ > 0 & \text{if } \lambda > \lambda_0, \end{cases} \quad (2.3.6)$$

$$\operatorname{Re}\beta_j(\lambda_0) < 0 \quad \forall j \geq 3. \quad (2.3.7)$$

It is known that with the conditions (2.3.6) and (2.3.7), (2.1.1) undergoes a Hopf bifurcation from $(0, \lambda_0)$. The following theorem amounts to saying that the transition of (2.1.1) from $(0, \lambda_0)$ has only two types: continuous and jump transitions.

Theorem 2.3.6 Under the conditions (2.3.6) and (2.3.7), the transition of (2.1.1) from $(0, \lambda_0)$ is either continuous or jump, and the following assertions hold true:

- (1) The transition is continuous if and only if (2.1.1) bifurcates to a periodic orbit on $\lambda > \lambda_0$, and this transition is an attractor bifurcation.
- (2) The transition is jump if and only if (2.1.1) has no bifurcation to periodic orbits on $\lambda > \lambda_0$, and there is only one connected metastable domain.

Proof. Without loss of generality, let $\operatorname{Im}\beta_1(\lambda) = -\operatorname{Im}\beta_2(\lambda) = 1$. Thus, the reduction equation of (2.1.1) can be rewritten as

$$\frac{dx}{dt} = \begin{pmatrix} \alpha(\lambda) & -1 \\ 1 & \alpha(\lambda) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} g_1(x, \lambda) \\ g_2(x, \lambda) \end{pmatrix}, \quad (2.3.8)$$

where $x = (x_1, x_2)^t \in \mathbb{R}^2$, $g_i(x, \lambda) = o(|x|)$ ($i = 1, 2$), and $\alpha(\lambda) = \operatorname{Re}\beta_1(\lambda)$ satisfies that

$$\alpha(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_0, \\ = 0 & \text{if } \lambda = \lambda_0, \\ > 0 & \text{if } \lambda > \lambda_0. \end{cases} \quad (2.3.9)$$

At the critical value $\lambda = \lambda_0$, (2.3.8) is in the form

$$\frac{dx}{dt} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} g_1(x, \lambda_0) \\ g_2(x, \lambda_0) \end{pmatrix}. \quad (2.3.10)$$

By the dynamical system theory, it is known that the singular point $x = 0$ can only be one of the three kinds as singular points of (2.3.10): (a) a stable focus, (b) an unstable focus, and (c) a center with infinite number of periodic orbits in a neighborhood $U \subset \mathbb{R}^2$ of $x = 0$.

For case (a), $x = 0$ is locally asymptotically stable. Hence, by Theorem 2.2.2, the transition is an attractor bifurcation in $\lambda > \lambda_0$.

For cases (b) and (c), the solution $x(t, x_0)$ of (2.3.10) satisfies that for any $0 < |x_0| < \varepsilon$, there is a $\delta_0 > 0$ such that

$$\limsup_{t \rightarrow \infty} |x(t, x_0)| \geq \delta_0 > 0, \quad (2.3.11)$$

where $x(0, x_0) = x_0$, and $\varepsilon > 0$ is a constant. By Assertion (2) of Theorem 2.1.5, we infer from (2.3.11) that the transition is jump, and (2.1.1) has no bifurcated periodic orbits on $\lambda > \lambda_0$.

Conversely if the transition is an attractor bifurcation, then by Assertion (1) of Theorem 2.1.5, $x = 0$ is a stable focus of (2.3.10). \square

Transition Types We give a criterion to determine the transition types of (2.1.1) from complex simple eigenvalues.

Let $e_1(\lambda)$, $e_2(\lambda)$ and $e_1^*(\lambda)$ and $e_2^*(\lambda)$ be the eigenvectors of L_λ and L_λ^* respectively corresponding to the complex eigenvalues $\beta_2(\lambda) = \bar{\beta}_1(\lambda) = \alpha(\lambda) + i\sigma(\lambda)$, where α satisfies (2.3.9), $\sigma_0 = \sigma(\lambda_0) \neq 0$, and

$$\begin{cases} L_\lambda e_1(\lambda) = \alpha(\lambda)e_1(\lambda) + \sigma(\lambda)e_2(\lambda), \\ L_\lambda e_2(\lambda) = -\sigma(\lambda)e_1(\lambda) + \alpha(\lambda)e_2(\lambda), \end{cases} \quad (2.3.12)$$

$$\begin{cases} L_\lambda^* e_1^*(\lambda) = \alpha(\lambda)e_1^* - \sigma(\lambda)e_2^*, \\ L_\lambda^* e_2^*(\lambda) = \sigma(\lambda)e_1^* + \alpha(\lambda)e_2^*. \end{cases} \quad (2.3.13)$$

By the spectral theorem, Theorem 3.4 in Ma and Wang (2005b), we can take

$$(e_i(\lambda), e_j^*(\lambda)) = \delta_{ij} \quad \forall 1 \leq i, j \leq 2.$$

Let $\Phi(x, \lambda)$ be the center manifold function of (2.1.1) near $\lambda = \lambda_0$, $x = x_1 e_1 + x_2 e_2$, and $e_i = e_i(\lambda_0)$ ($i = 1, 2$). Assume that for $i = 1, 2$,

$$(G(x + \Phi(x, \lambda_0), \lambda_0), e_i^*) = \sum_{2 \leq p+q \leq 3} a_{pq}^i x_1^p x_2^q + o(|x|^3). \quad (2.3.14)$$

For (2.3.14) we introduce a number, which is the bifurcation number in Theorem 2.2.8, given by

$$\begin{aligned} b = & \frac{3\pi}{4}(a_{30}^1 + a_{03}^2) + \frac{\pi}{4}(a_{12}^1 + a_{21}^2) + \frac{\pi}{2\sigma_0}(a_{02}^1 a_{02}^2 - a_{20}^1 a_{20}^2) \\ & + \frac{\pi}{4\sigma_0}(a_{11}^1 a_{20}^1 + a_{11}^1 a_{02}^1 - a_{11}^2 a_{20}^2 - a_{11}^2 a_{02}^2). \end{aligned} \quad (2.3.15)$$

The following theorem is essentially the generalization of a theorem (Andronov et al., 1973) in the finite-dimensional case to the infinite dimensional case here.

Theorem 2.3.7 *Let the conditions (2.3.6) and (2.3.7) hold true.*

1. *If $b < 0$, then the transition of (2.1.1) is continuous, and the bifurcated periodic orbit is an attractor. Furthermore, the bifurcated periodic solution can be expressed as*

$$\begin{aligned} u &= x_1 e_1 + x_2 e_2 + o(|\alpha(\lambda)|), \\ x_1(t) &= \left(\frac{\alpha(\lambda)}{|b|} \right)^{1/2} \cos \sigma t + o(|\alpha(\lambda)|), \\ x_2(t) &= \left(\frac{\alpha(\lambda)}{|b|} \right)^{1/2} \sin \sigma t + o(|\alpha(\lambda)|), \end{aligned}$$

where $\alpha(\lambda)$ and $\sigma = \sigma(\lambda)$ are as in (2.3.12).

2. *If $b > 0$, then the transition is jump, and (2.1.1) bifurcates on $\lambda < \lambda_0$ to a unique unstable periodic orbit.*

Proof. The reduced equation of (2.1.1) on the center manifold is given by

$$\begin{aligned} \frac{dx_1}{dt} &= \alpha(\lambda)x_1 - \sigma(\lambda)x_2 + (G(x + \Phi(x, \lambda), \lambda), e_1^*(\lambda)), \\ \frac{dx_2}{dt} &= \sigma(\lambda)x_1 + \alpha(\lambda)x_2 + (G(x + \Phi(x, \lambda), \lambda), e_2^*(\lambda)), \end{aligned} \quad (2.3.16)$$

where $\alpha(\lambda)$ satisfies (2.3.9). Therefore, by (2.3.14) the equations (2.3.16) at $\lambda = \lambda_0$ become

$$\begin{aligned} \frac{dx_1}{dt} &= -\sigma x_2 + g_{21}(x) + g_{31}(x) + o(|x|^3), \\ \frac{dx_2}{dt} &= \sigma x_1 + g_{22}(x) + g_{32}(x) + o(|x|^3), \end{aligned} \quad (2.3.17)$$

where $\sigma = \sigma(\lambda_0) \neq 0$, and

$$g_{ki}(x) = \sum_{p+q=k} a_{pq}^i x_1^p x_2^q \quad \text{for } k = 2, 3, \quad i = 1, 2.$$

Taking the polar coordinate system $(x_1, x_2) = (r \cos \theta, r \sin \theta)$, we obtain from (2.3.17) that

$$\begin{aligned} \frac{dr}{d\theta} &= \frac{r \cos \theta \frac{dx_1}{dt} + r \sin \theta \frac{dx_2}{dt}}{\cos \theta \frac{dx_2}{dt} - \sin \theta \frac{dx_1}{dt}} \\ &= \frac{r^2 [\cos \theta(g_{21} + rg_{31}) + \sin \theta(g_{22} + rg_{32})] + o(r^3)}{\sigma + r [\cos \theta(g_{22} + rg_{32}) - \sin \theta(g_{21} + rg_{31})] + o(r^2)}, \end{aligned} \quad (2.3.18)$$

where $g_{ki} = \sum_{p+q=k} a_{pq}^i \cos^p \theta \sin^q \theta$ for $k = 2, 3$ and $i = 1, 2$.

Near $r = 0$, the equation (2.3.18) can be rewritten as

$$\frac{1}{r^2} \frac{dr}{d\theta} = \frac{1}{\sigma} (u_1 + r(u_2 + \frac{1}{\sigma} u_1 v_1)) + o(r), \quad (2.3.19)$$

where

$$\begin{aligned} u_1(\theta) &= \cos \theta \sum_{p+q=2} a_{pq}^1 \cos^p \theta \sin^q \theta + \sin \theta \sum_{p+q=2} a_{pq}^2 \cos^p \theta \sin^q \theta, \\ v_1(\theta) &= \sin \theta \sum_{p+q=2} a_{pq}^1 \cos^p \theta \sin^q \theta - \cos \theta \sum_{p+q=2} a_{pq}^2 \cos^p \theta \sin^q \theta, \\ u_2(\theta) &= \cos \theta \sum_{p+q=3} a_{pq}^1 \cos^p \theta \sin^q \theta + \sin \theta \sum_{p+q=3} a_{pq}^2 \cos^p \theta \sin^q \theta. \end{aligned}$$

Let $r(\theta, a)$ be the solution of (2.3.19) with the initial value $r(0, a) = a$. Integrating (2.3.19) on θ , we have

$$\frac{r(\theta, a) - a}{ar(\theta, a)} = \frac{1}{\sigma} \int_0^\theta [u_1 + r(u_2 + \frac{1}{\sigma} u_1 v_1) + o(r)] d\theta,$$

which implies that $r(\theta, a)$ has the Taylor expansion as follows

$$r(\theta, a) = a + \sum_{k=2} f_k(\theta) a^k = a + o(a^2). \quad (2.3.20)$$

It is easy to see that

$$\int_0^{\pm 2\pi} u_1(\theta) d\theta = 0. \quad (2.3.21)$$

We know that the orbits of (2.3.17) around $x = 0$ rotate clockwise if $\sigma < 0$, and counterclockwise if $\sigma > 0$. Hence, we deduce from (2.3.19)–(2.3.21) that if $\sigma > 0$,

$$\frac{r(2\pi, a) - r(0, a)}{r(2\pi, a)} = \frac{1}{\sigma} a^2 \int_0^{2\pi} (u_2 + \frac{1}{\sigma} u_1 v_1) d\theta + o(a^2),$$

and if $\sigma < 0$,

$$\begin{aligned} \frac{r(-2\pi, a) - r(0, a)}{r(-2\pi, a)} &= \frac{1}{\sigma} a^2 \int_0^{-2\pi} (u_2 + \frac{1}{\sigma} u_1 v_1) d\theta + o(a^2) \\ &= \frac{1}{|\sigma|} a^2 \int_0^{2\pi} (u_2 + \frac{1}{\sigma} u_1 v_1) d\theta + o(a^2). \end{aligned}$$

Namely, for $\sigma \neq 0$ we obtain

$$\frac{r(\pm 2\pi, a) - r(0, a)}{r(\pm 2\pi, a)} = \frac{1}{|\sigma|} b a^2 + o(a^2), \quad (2.3.22)$$

where

$$b = \int_0^{2\pi} (u_2(\theta) + \frac{1}{\sigma} u_1(\theta) v_1(\theta)) d\theta. \quad (2.3.23)$$

From (2.3.22) we can see that when $b < 0$, $x = 0$ is a stable focus of (2.3.17), which implies that the transition of (2.3.16) at $\lambda = \lambda_0$ is continuous, and when $b > 0$, $x = 0$ is an unstable focus and the transition is jump. By direct calculation we can derive that the number b in (2.3.23) equals to that given in (2.3.15).

Furthermore, we see that when $b < 0$, the bifurcated periodic solution of (2.3.16) is given by

$$(x_1(t), x_2(t)) = \left(a e^{\alpha(\lambda)} \cos \sigma t + o(|\alpha(\lambda)|), a e^{\alpha(\lambda)} \sin \sigma t + o(|\alpha(\lambda)|) \right),$$

where a is as (2.3.22). It is known that $ba^2 = -\alpha(\lambda)$.

The proof of the theorem is complete. \square

We now consider the special case where the nonlinear term $G(u, \lambda)$ of (2.1.1) has the Taylor expansion as follows

$$G(u, \lambda) = G_k(u, \lambda) + G_{k+1}(u, \lambda) + o(\|u\|_{X_1}^{k+1}), \quad (2.3.24)$$

where $k \geq 3$ is an integer, and $G_r(u, \lambda)$ is the r -multilinear operator for $r = k, k+1$. Let e_1, e_2 and e_1^*, e_2^* be as in (2.3.14). Then for $i = 1, 2$, we have

$$(G_r(x_1 e_1 + x_2 e_2, \lambda_0), e_i^*) = \sum_{p+q=r} a_{pq}^i x_1^p x_2^q.$$

For the coefficients a_{pq}^i , we define a new number by

$$b_0 = \begin{cases} \sum_{l=0}^{(k-1)/2} \alpha_{kl} (a_{(k-2l)2l}^1 + a_{2l(k-2l)}^2) & \text{for } k = \text{odd}, \\ \sum_{l=0}^{k/2} \alpha_{(k+1)l} (a_{(k+1-2l)2l}^1 + a_{2l(k+1-2l)}^2) & \text{for } k = \text{even}. \end{cases} \quad (2.3.25)$$

where for $r = k, k + 1$,

$$\alpha_{rl} = \int_0^{2\pi} \sin^{r+1-2l} \theta \cos^{2l} \theta d\theta.$$

Theorem 2.3.8 Under the conditions (2.3.6), (2.3.7), and (2.3.24), the following assertions hold true:

1. If $b_0 > 0$, the system (2.1.1) has a jump transition at $\lambda = \lambda_0$, and bifurcates on $\lambda < \lambda_0$ to a unique periodic orbit, which is a repeller.
2. If $b_0 < 0$, (2.1.1) has a continuous transition and bifurcates on $\lambda > \lambda_0$ to a unique periodic orbit which is an attractor.

Proof. Under the condition (2.3.24), the reduction equation (2.3.16) at λ_0 is in the following form:

$$\begin{aligned} \frac{dx_1}{dt} &= -\sigma x_2 + g_k^1(x) + g_{k+1}^1(x) + o(|x|^{k+1}), \\ \frac{\partial x_2}{\partial t} &= \sigma x_1 + g_k^2(x) + g_{k+1}^2(x) + o(|x|^{k+1}), \end{aligned} \quad (2.3.26)$$

where

$$g_r^i(x) = (G_r(x_1 e_1 + x_2 e_2, \lambda_0), e_i^*) = \sum_{p+q=r} a_{pq}^i x_1^p x_2^q.$$

Thus, in the polar coordinate system, (2.3.26) can be written as

$$\begin{aligned} \frac{1}{r^k} \frac{dr}{d\theta} &= \frac{\cos \theta g_k^1 + \sin \theta g_k^2 + r(\cos \theta g_{k+1}^1 + \sin \theta g_{k+1}^2) + o(r)}{\sigma - r^{k-1}(\sin \theta g_k^1 - \cos \theta g_k^2) + o(r^{k-1})} \\ &= \begin{cases} \frac{1}{\sigma} u_k(\theta) + o(r) & \text{if } k = \text{odd}, \\ \frac{1}{\sigma} [u_k + r u_{k+1}] + o(r) & \text{if } k = \text{even } (> 2), \end{cases} \end{aligned} \quad (2.3.27)$$

where

$$u_r = \cos \theta \sum_{p+q=r} a_{pq}^1 \cos^p \theta \sin^q \theta + \sin \theta \sum_{p+q=r} a_{pq}^2 \cos^p \theta \sin^q \theta.$$

Note that if either p or q is odd, then

$$\int_0^{2\pi} \cos^p \theta \sin^q \theta d\theta = 0.$$

It follows then from (2.3.27) and (2.3.20) that

$$\frac{r^{k-1}(2\pi, a) - r^{k-1}(0, a)}{r^{k-1}(2\pi, a)} = \begin{cases} \frac{1}{\sigma} ba^{k-1} + o(a^{k-1}) & \text{if } k = \text{odd}, \\ \frac{1}{\sigma} ba^k + o(a^k) & \text{if } k = \text{even}, \end{cases} \quad (2.3.28)$$

where we only consider the case of $\sigma > 0$, and b is given by

$$b = \begin{cases} \int_0^{2\pi} u_k(\theta) d\theta & \text{if } k = \text{odd}, \\ \int_0^{2\pi} u_{k+1}(\theta) d\theta & \text{if } k = \text{even}. \end{cases} \quad (2.3.29)$$

The number in (2.3.29) is the same as in (2.3.25). It is then routine to derive from (2.3.28) the theorem. The proof is complete. \square

2.3.3 Computation of b

Theorems 2.3.7 and 2.3.8 give a number b which dictates the type of transitions from a pair of complex eigenvalues. However, it is important to know how to compute this number in specific problems. In the following we give two examples to show the computing methods.

We first consider a finite dimensional system given by

$$\frac{dx}{dt} = Ax + G(x) + o(|x|^3) \quad \text{for } x \in \mathbb{R}^n, \quad (2.3.30)$$

where A is an $n \times n$ matrix which has eigenvalues

$$\begin{aligned} \lambda_1 &= \bar{\lambda}_2 = -i\alpha & (\alpha \neq 0), \\ \lambda_j &\neq 0 & \forall j = 2, \dots, n. \end{aligned}$$

The eigenvectors e_1, \dots, e_n of A are given by

$$\begin{aligned} Ae_1 &= \alpha e_2, & Ae_2 &= -\alpha e_1, \\ (A - \lambda_j I)^{k_j} e_j &= 0 & \forall j = 3, \dots, n, & k_j \geq 1, \end{aligned}$$

and the eigenvectors e_1^*, \dots, e_n^* of A^* by

$$\begin{aligned} A^* e_1^* &= -\alpha e_2^*, & A^* e_2^* &= \alpha e_1^*, \\ (A^* - \lambda_j I)^{m_j} e_j^* &= 0 & \forall j = 3, \dots, n, & m_j \geq 1, \\ (e_i, e_j^*) &= \delta_{ij}. \end{aligned}$$

The nonlinear term $G(x) = (g_1(x), \dots, g_n(x))$ is a polynomial field

$$g_i(x) = \sum_{2 \leq p_1 + \dots + p_n \leq 3} C_{p_1 \dots p_n}^i x_1^{p_1} \cdots x_n^{p_n}. \quad (2.3.31)$$

The system (2.3.30) can be written in the following form

$$\begin{cases} \frac{dz_1}{dt} = -\alpha z_2 + (B^{-1}G(Bz), a_1 e_1^*) + o(|z|^3), \\ \frac{dz_2}{dt} = \alpha z_1 + (B^{-1}G(Bz), a_2 e_2^*) + o(|z|^3), \end{cases} \quad (2.3.32)$$

$$\frac{d\tilde{z}}{dt} = J_{n-2}\tilde{z} + \begin{pmatrix} (B^{-1}G(Bz), a_3 e_3^*) \\ \vdots \\ (B^{-1}G(Bz), a_n e_n^*) \end{pmatrix} + o(|z|^3), \quad (2.3.33)$$

where $z = (z_1, \dots, z_n)^T$, $\tilde{z} = (z_3, \dots, z_n)$, J_{n-2} is the Jordan matrix associated with the eigenvalues $\lambda_3, \dots, \lambda_n$, and

$$Bz = \begin{pmatrix} e_{11} & e_{21} & \cdots & e_{n1} \\ \vdots & \vdots & & \vdots \\ e_{1n} & e_{2n} & \cdots & e_{nn} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix},$$

with $e_i = (e_{i1}, \dots, e_{in})$ ($1 \leq i \leq n$) being the eigenvectors of A , and

$$a_k = \frac{1}{(e_k, e_k^*)} \quad \forall 1 \leq k \leq n.$$

By the formulas of center manifold functions (Theorem A.1.1), the center manifold function $\Phi(z_1, z_2)$ of (2.3.30) is determined by (2.3.33), whose first approximation $\Phi_0(z_1, z_2)$ is given by

$$\tilde{z}_0 = \Phi_0(z_1, z_2) = \Phi_0^1 + \Phi_0^2 + \Phi_0^3,$$

where

$$\begin{aligned} (-J_{n-2})\Phi_0^1 &= \Gamma_1, \\ (J_{n-2}^2 + 4\alpha^2)(-J_{n-2})\Phi_0^2 &= 2\alpha^2\Gamma_1 - 2\alpha^2\Gamma_2, \\ (J_{n-2}^2 + 4\alpha^2)(-J_{n-2})\Phi_0^3 &= \alpha\Gamma_3 + \alpha\Gamma_4, \end{aligned}$$

and

$$\begin{aligned}\Gamma_1 &= \begin{pmatrix} a_3(B^{-1}G_2(z_1e_1 + z_2e_2), e_3^*) \\ \vdots \\ a_n(B^{-1}G_2(z_1e_1 + z_2e_2), e_n^*) \end{pmatrix}, \\ \Gamma_2 &= \begin{pmatrix} a_3(B^{-1}G_2(z_1e_2 - z_2e_1), e_3^*) \\ \vdots \\ a_n(B^{-1}G_2(z_1e_2 - z_2e_1), e_n^*) \end{pmatrix}, \\ \Gamma_3 &= \begin{pmatrix} a_3(B^{-1}G_2(z_1e_1 + z_2e_2, z_2e_1 - z_1e_2), e_3^*) \\ \vdots \\ a_n(B^{-1}G_2(z_1e_1 + z_2e_2, z_2e_1 - z_1e_2), e_n^*) \end{pmatrix}, \\ \Gamma_4 &= \begin{pmatrix} a_3(B^{-1}G_2(z_2e_1 - z_1e_2, z_1e_1 + z_2e_2), e_3^*) \\ \vdots \\ a_n(B^{-1}G_2(z_2e_1 - z_1e_2, z_1e_1 + z_2e_2), e_n^*) \end{pmatrix},\end{aligned}$$

and G_2 is the bilinear term of G .

Inserting $z = z_1e_1 + z_2e_2 + \Phi_0(z_1, z_2)$ into (2.3.32), we obtain

$$(B^{-1}G(Bz), a_i e_i^*)|_{z=(z_1, z_2, \Phi_0)} = \sum_{2 \leq k+l \leq 3} a_{kl}^i z_1^k z_2^l + o(|z_0|^3).$$

Thus, we derive the coefficients a_{kl}^1 and a_{kl}^2 ($2 \leq k+l \leq 3$) which determine the number b defined in (2.3.15) for the system (2.3.30).

2.4 Transition from Eigenvalues with Multiplicity Two

2.4.1 Index Formula for Second-Order Nondegenerate Singularities

In order to investigate transitions of (2.1.1) from eigenvalues with multiplicity two, it is necessary to discuss the index of the following bilinear vector fields at $x = 0$:

$$F = \begin{pmatrix} a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2 \\ b_{11}x_1^2 + b_{12}x_1x_2 + b_{22}x_2^2 \end{pmatrix}. \quad (2.4.1)$$

We assume that the vector field (2.4.1) is second-order nondegenerate at $x = 0$, which implies that $a_{11}^2 + b_{11}^2 \neq 0$, i.e., $x = 0$ is an isolated singular point of (2.4.1). Without loss of generality, we assume that $a_{11} \neq 0$. Let

$$\Delta = a_{12}^2 - 4a_{11}a_{22}.$$

If $\Delta \geq 0$, we denote

$$\begin{aligned}\alpha_1 &= \frac{-a_{12} + \sqrt{\Delta}}{2a_{11}}, & \alpha_2 &= \frac{-a_{12} - \sqrt{\Delta}}{2a_{11}}, \\ \beta_i &= b_{11}\alpha_i^2 + b_{12}\alpha_i + b_{22} \quad i = 1, 2.\end{aligned}$$

The following index theorem for (2.4.1) will be useful in studying transitions of (2.1.1) from eigenvalues with multiplicity two.

Theorem 2.4.1 *Let the vector field (2.4.1) be second-order nondegenerate at $x = 0$, and $a_{11} \neq 0$. Then*

$$ind(F, 0) = \begin{cases} 0 & \text{if } \Delta < 0 \text{ or } \beta_1\beta_2 > 0, \\ 2 & \text{if } \beta_1 > 0 \text{ and } \beta_2 < 0, \\ -2 & \text{if } \beta_1 < 0 \text{ and } \beta_2 > 0. \end{cases} \quad (2.4.2)$$

Proof. We proceed in several steps as follows.

STEP 1. When $\Delta = a_{12}^2 - 4a_{11}a_{22} < 0$, the following quadratic form is either positively or negatively definite:

$$a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2 > 0 \quad (\text{or } < 0) \quad \forall x \in \mathbb{R}^2, \quad x \neq 0,$$

depending on the sign of a_{11} . Hence the following equations

$$\begin{aligned}a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2 &= -\varepsilon^2 \quad (\text{or } = \varepsilon^2), \\ b_{11}x_1^2 + b_{12}x_1x_2 + b_{22}x_2^2 &= 0,\end{aligned}$$

have no solutions for any $\varepsilon \neq 0$, which implies that

$$ind(F, 0) = 0 \quad \text{if } \Delta < 0.$$

STEP 2. In the case where $\Delta \geq 0$, the vector field F given in (2.4.1) can be rewritten as

$$F = \begin{pmatrix} a_{11}(x_1 - \alpha_1 x_2)(x_1 - \alpha_2 x_2) \\ b_{11}x_1^2 + b_{12}x_1x_2 + b_{22}x_2^2 \end{pmatrix}. \quad (2.4.3)$$

Since F is second-order nondegenerate at $x = 0$, $\beta_1 \cdot \beta_2 \neq 0$. By (2.4.3), $F = (0, \pm\varepsilon^2)$ with $\varepsilon \neq 0$ is equivalent to

$$x_1 = \alpha_i x_2, \quad \beta_i x_2^2 = \pm\varepsilon^2 \quad (i = 1, 2). \quad (2.4.4)$$

If $\beta_1 \cdot \beta_2 > 0$, then one of the systems in (2.4.4), for either $+\varepsilon^2$ or $-\varepsilon^2$, has no solution, which means that the index of F at $x = 0$ is zero.

STEP 3. When $\beta_1 \cdot \beta_2 < 0$, it is easy to see that $\alpha_1 \neq \alpha_2$ and $\Delta > 0$. The vector field $F = (F_1, F_2)$ given in (2.4.3) takes the following form

$$\begin{aligned} F_1 &= a_{11}(x_1 - \alpha_1 x_2)(x_1 - \alpha_2 x_2), \\ F_2 &= \frac{1}{(\alpha_1 - \alpha_2)^2} [\beta_1(x_1 - \alpha_2 x_2)^2 + \beta_2(x_1 - \alpha_1 x_2)^2 \\ &\quad + \gamma(x_1 - \alpha_1 x_2)(x_1 - \alpha_2 x_2)], \end{aligned} \quad (2.4.5)$$

where $\gamma = -(2b_{11}\alpha_1\alpha_2 + b_{12}\alpha_1 + b_{12}\alpha_2 + 2b_{22})$. Let

$$\beta_1 > 0, \quad \beta_2 < 0.$$

Then the solutions $x = (x_1, x_2)$ of (2.4.4) for $+\varepsilon^2$ are given by

$$x_1^\pm = \alpha_1 x_2^\pm, \quad x_2^\pm = \pm \beta_1^{-1/2} \varepsilon. \quad (2.4.6)$$

We take

$$(z_1, z_2) = (x_1 - \alpha_1 x_2, x_1 - \alpha_2 x_2). \quad (2.4.7)$$

Then the Jacobian matrix of F is given by

$$DF = \begin{pmatrix} \frac{\partial F_1}{\partial z_1} & \frac{\partial F_1}{\partial z_2} \\ \frac{\partial F_2}{\partial z_1} & \frac{\partial F_2}{\partial z_2} \end{pmatrix} \begin{pmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} \\ \frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2} \end{pmatrix}.$$

By (2.4.7), it is easy to see that

$$\det \begin{pmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} \\ \frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2} \end{pmatrix} = \det \begin{pmatrix} 1 & -\alpha_1 \\ 1 & -\alpha_2 \end{pmatrix} = \alpha_1 - \alpha_2 = \sqrt{\Delta}/a_{11}. \quad (2.4.8)$$

Hence we find from (2.4.5), (2.4.7), and (2.4.8) that

$$\begin{aligned} \det DF(x) &= (\alpha_1 - \alpha_2) \det \begin{pmatrix} a_{11}z_2 & a_{11}z_1 \\ 2\beta_2 z_1 + \gamma z_2 & 2\beta_1 z_2 + \gamma z_1 \end{pmatrix} \\ &= \frac{a_{11}^2}{\sqrt{\Delta}} \det \begin{pmatrix} z_2 & z_1 \\ 2\beta_2 z_1 + \gamma z_2 & 2\beta_1 z_2 + \gamma z_1 \end{pmatrix}. \end{aligned} \quad (2.4.9)$$

On the other hand, by (2.4.7) we have

$$z_1^\pm = x_1^\pm - \alpha_1 x_2^\pm = 0, \quad z_2^\pm = x_1^\pm - \alpha_2 x_2^\pm = \pm(\alpha_1 - \alpha_2)\beta_1^{-1/2} \varepsilon.$$

Therefore, from (2.4.9) we get

$$\det DF(x^\pm) = \frac{a_{11}^2}{\sqrt{\Delta}} 2\beta_1(z_2^\pm)^2. \quad (2.4.10)$$

By the Brouwer degree theory, we know that

$$\text{ind}(F, 0) = \deg(F, B_r, x_0), \quad x_0 = (0, \varepsilon^2) \in B_r, \quad (2.4.11)$$

where $B_r = \{x \in \mathbb{R}^2 \mid |x| < r\}$ and $r > 0$ sufficiently small. It follows from (2.4.10) and (2.4.11) that if $\beta_1 > 0$,

$$\text{ind}(F, 0) = \text{sign det } DF(x^+) + \text{sign det } DF(x^-) = 2.$$

In the same fashion, we can obtain that

$$\text{ind}(F, 0) = -2 \quad \text{for } \beta_1 < 0, \quad \beta_2 > 0.$$

Thus the formula (2.4.2) is proved. The proof is complete. \square

Remark 2.4.2 If $a_{11} = 0$ and $b_{11} \neq 0$, we let

$$\tilde{\Delta} = b_{12}^2 - 4b_{11}b_{22}.$$

If $\tilde{\Delta} \geq 0$, we define

$$\begin{aligned} \tilde{\alpha}_1 &= \frac{-b_{12} - \sqrt{\tilde{\Delta}}}{2b_{11}}, & \tilde{\alpha}_2 &= \frac{-b_{12} + \sqrt{\tilde{\Delta}}}{2b_{11}}, \\ \tilde{\beta}_i &= a_{11}\tilde{\alpha}_i^2 + a_{12}\tilde{\alpha}_i + a_{22}, \quad i = 1, 2, \end{aligned}$$

then the formula (2.4.2) is written as

$$\text{ind}(F, 0) = \begin{cases} 0 & \text{if } \tilde{\Delta} < 0 \text{ or } \tilde{\beta}_1\tilde{\beta}_2 > 0, \\ 2 & \text{if } \tilde{\beta}_1 < 0 \text{ and } \tilde{\beta}_2 > 0, \\ -2 & \text{if } \tilde{\beta}_1 > 0 \text{ and } \tilde{\beta}_2 < 0. \end{cases}$$

2.4.2 Bifurcation at Second-Order Singular Points

Under the conditions (2.1.4) and (2.1.5), the integers m and r are the algebraic and geometric multiplicities of the eigenvalues $\beta_1(\lambda_0)$ of L_λ at $\lambda = \lambda_0$. Here, we assume that $m = r = 2$, and the operator $L_\lambda + G(\cdot, \lambda)$ is second-order nondegenerate at $(u, \lambda) = (0, \lambda_0)$, i.e.,

$$G(u, \lambda) = G_2(u, \lambda) + o(\|u\|_{X_1}),$$

where $G_2(u, \lambda)$ is a bilinear operator. We denote

$$\begin{aligned} a_{11}(\lambda) &= (G_2(e_1(\lambda), e_1(\lambda), \lambda), e_1^*(\lambda)), \\ a_{22}(\lambda) &= (G_2(e_2(\lambda), e_2(\lambda), \lambda), e_1^*(\lambda)), \\ a_{12}(\lambda) &= (G_2(e_1(\lambda), e_2(\lambda), \lambda) + G_2(e_2(\lambda), e_1(\lambda), \lambda), e_1^*(\lambda)), \\ b_{11}(\lambda) &= (G_2(e_1(\lambda), e_1(\lambda), \lambda), e_2^*(\lambda)), \\ b_{22}(\lambda) &= (G_2(e_2(\lambda), e_2(\lambda), \lambda), e_2^*(\lambda)), \\ b_{12}(\lambda) &= (G_2(e_1(\lambda), e_2(\lambda), \lambda) + G_2(e_2(\lambda), e_1(\lambda), \lambda), e_2^*(\lambda)), \end{aligned} \quad (2.4.12)$$

where $e_i(\lambda)$ and $e_i^*(\lambda)$ ($i = 1, 2$) are the eigenvectors of L_λ and L_λ^* near $\lambda = \lambda_0$, respectively:

$$L_\lambda e_i(\lambda) = \beta_i(\lambda) e_i(\lambda), \quad L_\lambda^* e_i^*(\lambda) = \beta_i(\lambda) e_i^*(\lambda) \quad i = 1, 2.$$

Thus we derive a vector field

$$F(x, \lambda) = \begin{pmatrix} a_{11}(\lambda)x_1^2 + a_{12}(\lambda)x_1x_2 + a_{22}(\lambda)x_2^2 \\ b_{11}(\lambda)x_1^2 + b_{12}(\lambda)x_1x_2 + b_{22}(\lambda)x_2^2 \end{pmatrix}. \quad (2.4.13)$$

Consider the steady-state equation of (2.1.1):

$$L_\lambda u + G(u, \lambda) = 0. \quad (2.4.14)$$

The following theorem will be useful later in this section.

Theorem 2.4.3 *Let the conditions (2.1.4), (2.1.5) and $r = m = 2$. If $u = 0$ is a second-order nondegenerate singular point of $L_\lambda + G$ at $\lambda = \lambda_0$, i.e., $x = 0$ is an isolated zero of $F(x, \lambda_0)$, and the two vectors (a_{11}, a_{12}, a_{13}) and (b_{11}, b_{12}, b_{13}) , defined by (2.4.12), are linearly independent at $\lambda = \lambda_0$, then we have the following assertions.*

1. *There are at most three branches of (2.4.14) bifurcated from $(0, \lambda_0)$ on each side of $\lambda = \lambda_0$.*
2. *If the bifurcated branches on $\lambda > \lambda_0$ (resp. on $\lambda < \lambda_0$) are regular, then the number of branches on this side is either 1 or 3.*
3. *On any given side, if the number of branches is 3, then these 3 branches must be regular.*
4. *If the number of branches on one side is 2, then one of them is regular.*

Proof. We proceed in several steps as follows.

STEP 1. As $m = r = 2$ in (2.1.4) and (2.1.5), the bifurcation equation (2.4.14) can be rewritten as follows

$$\begin{aligned} \beta_1(\lambda)x_1 + a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2 + o(|x|^2) &= 0, \\ \beta_2(\lambda)x_2 + b_{11}x_1^2 + b_{12}x_1x_2 + b_{22}x_2^2 + o(|x|^2) &= 0. \end{aligned} \quad (2.4.15)$$

STEP 2. We now show that there is a $\varepsilon > 0$ such that for any C^∞ function $(f_1(x), f_2(x))$ with $|f_1(x)| + |f_2(x)|$ small near $x = 0$, the following equations

$$a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2 + f_1(x) + o(|x|^2) = 0, \quad (2.4.16)$$

$$b_{11}x_1^2 + b_{12}x_1x_2 + b_{22}x_2^2 + f_2(x) + o(|x|^2) = 0, \quad (2.4.17)$$

have at most four solutions in $|x| < \varepsilon$.

Since the second-order terms in (2.4.16) and (2.4.17) are second-order nondegenerate, at least one of the coefficients a_{11} , a_{22} , b_{11} , and b_{22} is not zero. Without loss of generality, we assume that $a_{22} \neq 0$. Then we obtain from (2.4.16) that

$$x_2^\pm = -\frac{a_{12}}{2a_{22}} x_1 \pm \frac{1}{2a_{22}} \sqrt{(a_{12}^2 - 4a_{11}a_{22})x_1^2 - 4a_{22}f_1(x) + o(|x|^2)},$$

which, by the implicit function theorem, implies that

$$x_2^\pm = -\frac{a_{12}}{2a_{22}} x_1 \pm \frac{1}{2a_{22}} \Delta, \quad (2.4.18)$$

where

$$\begin{aligned} \Delta &= \sqrt{(a_{12}^2 - 4a_{11}a_{22})x_1^2 + g(x_1) + o(|x_1|^2)} \\ g(x_1) &\rightarrow 0 \quad \text{if} \quad f_1(x) \rightarrow 0. \end{aligned}$$

Inserting (2.4.18) into (2.4.17), we have

$$\alpha x_1^2 + b_{22}g(x_1) + 4a_{22}^2f_2(x) + o(|x|^2) = \mp\beta x_1\Delta, \quad (2.4.19)$$

where

$$\alpha = 4a_{22}(a_{22}b_{11} - a_{11}b_{22}) + 2a_{12}(a_{12}b_{22} - a_{22}b_{12}), \quad \beta = 2(a_{22}b_{12} - a_{12}b_{22}).$$

We infer then from (2.4.18) and (2.4.19) that

$$\rho x_1^4 + h_1(x_1)x_1^2 + h_2(x_1) + o(|x_1|^4) = 0, \quad (2.4.20)$$

where

$$\begin{aligned} \rho &= \alpha^2 - \beta^2(a_{11}^2 - 4a_{11}a_{22}), \\ h_1(x_1), h_2(x_1) &\rightarrow 0 \quad \text{if} \quad f_1(x), f_2(x) \rightarrow 0. \end{aligned}$$

It is easy to verify that the condition $\rho \neq 0$ is equivalent to the following conditions

$$b_{22}z_\pm^2 + b_{11}z_\pm + b_{11} \neq 0, \quad z_\pm = \frac{-a_{12} \pm \sqrt{a_{12}^2 - 4a_{11}a_{22}}}{2a_{22}}. \quad (2.4.21)$$

By the assumption of this theorem, conditions in (2.4.21) are valid. Therefore, $\rho \neq 0$, which implies that (2.4.20) has at most four real solutions satisfying that $x_1 \rightarrow 0$ as $f_1(x), f_2(x) \rightarrow 0$.

On the other hand, each solution of (2.4.20) corresponds to one of the signs \pm in (2.4.19) and (2.4.18). Hence equations (2.4.16) and (2.4.17) have at most four real solutions near $x = 0$. Namely, Assertion (1) is proved.

STEP 3. PROOF OF ASSERTION (2). Let $F_\lambda : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be defined by

$$F_\lambda(x) = \begin{pmatrix} \beta_1(\lambda)x_1 + a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2 + o(|x|^2) \\ \beta_2(\lambda)x_2 + b_{11}x_1^2 + b_{12}x_1x_2 + b_{22}x_2^2 + o(|x|^2) \end{pmatrix}.$$

By Theorem 3.9 in Ma and Wang (2005b), the singular points $z(\lambda)$ of (2.4.15) near $x = 0$ are regular if and only if the bifurcated branches of (2.4.14) from $(0, \lambda_0)$ are regular. Hence, we have

$$\text{ind}(F_\lambda, z(\lambda)) = \pm 1 \quad (2.4.22)$$

for $|\lambda - \lambda_0| \neq 0$ sufficiently small. In addition, we know that if $\lambda \neq \lambda_0$,

$$\text{ind}(F_\lambda, 0) = \text{sign}\beta_1(\lambda)\beta_2(\lambda) = 1. \quad (2.4.23)$$

It is known that the topological degree of an even mapping is even. Hence we have

$$\sum_{|z_i|<\varepsilon} \text{ind}(F_\lambda, z_i(\lambda)) = \text{even}, \quad (2.4.24)$$

for any $\varepsilon > 0$ small. Hence Assertion (2) follows from (2.4.22)–(2.4.24) and Assertion (1).

STEP 4. PROOF OF ASSERTIONS (3) AND (4). We know from Step 2 that the solutions of (2.4.16) and (2.4.17) are one to one corresponding to those of (2.4.20). Let $x_0 \in \mathbb{R}^1$ be a solution of (2.4.20) near $x = 0$. Then (2.4.20) can be expressed as

$$(x - x_0)^k g(x) = 0, \quad k \geq 1. \quad (2.4.25)$$

We shall prove that a solution $z_0 \in \mathbb{R}^2$ of (2.4.16) and (2.4.17) corresponding to x_0 is nondegenerate if and only if $k = 1$ in (2.4.25).

Obviously, if $k \geq 2$ in (2.4.25), under a perturbation there are more solutions of (2.4.16) and (2.4.17) near z_0 .

Suppose z_0 is degenerate, i.e.

$$\det DF(z_0) = 0, \quad (2.4.26)$$

where $DF(z_0)$ stands for the Jacobian matrix of F , and

$$F = \begin{cases} a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2 + f_1(x) + o(|x|^2), \\ b_{11}x_1^2 + b_{12}x_1x_2 + b_{22}x_2^2 + f_2(x) + o(|x|^2). \end{cases}$$

Under a suitable coordinate system, $DF(z_0)$ must be one of the following three forms

$$DF(z_0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad (2.4.27)$$

$$DF(z_0) = \begin{pmatrix} \alpha & 0 \\ 0 & 0 \end{pmatrix}, \quad (2.4.28)$$

$$DF(z_0) = \begin{pmatrix} 0 & \alpha \\ 0 & 0 \end{pmatrix}, \quad (2.4.29)$$

for some $\alpha \neq 0$.

Since f_1, f_2 are C^∞ sufficiently small functions, and the quadratic forms in F are second-order nondegenerate, for both cases of (2.4.27) and (2.4.28) it is known that with a linear perturbation the equation (2.4.16) and (2.4.17) will yield one more solution near z_0 , which implies $k \geq 2$ in (2.4.25). We proceed for the case (2.4.29).

Near z_0 , F can be expressed as

$$\tilde{F}(y) = \begin{cases} \alpha y_2 + c_{11}y_1^2 + c_{12}y_1y_2 + c_{22}y_2^2 + o(|y|^2), \\ d_{11}y_1^2 + d_{12}y_1y_2 + d_{22}y_2^2 + o(|y|^2), \end{cases} \quad (2.4.30)$$

where $y = x - z_0$.

The zero point $y = 0$ of (2.4.30) stands for the zero point z_0 of F . Now we consider the perturbation equation

$$\alpha y_2 = -(c_{11}y_1^2 + c_{12}y_1y_2 + c_{22}y_2^2) + o(|y|^2), \quad (2.4.31)$$

$$\varepsilon y_1 = -(d_{11}y_1^2 + d_{12}y_1y_2 + d_{22}y_2^2) + o(|y|^2), \quad (2.4.32)$$

for ε small. By the implicit function theorem, we obtain from (2.4.31) a solution $y_2 = f(y_1)$. Then (2.4.32) implies that

$$\varepsilon y_1 = -(d_{11}y_1^2 + d_{12}y_1f(y_1) + d_{22}f^2(y_1)) + o(|y_1|^2). \quad (2.4.33)$$

By the second-order nondegeneracy of $F(y)$, there exist $a \neq 0$ and $m \geq 2$ such that (2.4.33) becomes

$$\varepsilon y_1 = ay_1^m + o(|y_1|^m). \quad (2.4.34)$$

We infer from (2.4.34) that equations (2.4.31) and (2.4.32) have two solutions $(0, 0)$ and $(\bar{y}_1, f(\bar{y}_1))$, where

$$\bar{y}_1 = (\varepsilon/a)^{1/(m-1)} + o\left(|\varepsilon/a|^{1/(m-1)}\right),$$

and ε is taken such that $\varepsilon \cdot a > 0$.

Thus, we prove that if z_0 is a degenerate zero point of (2.4.15) and (2.4.17), then the exponential $k \geq 2$ of the corresponding solution x_0 in (2.4.25).

We are now in a position to prove Assertions (3)–(4). If (2.4.15) has three bifurcated solutions (resp. two bifurcated solutions) from $(0, \lambda_0)$, then the corresponding solutions $x_i \neq 0$, $i = 1, 2, 3$, (resp. $x_j \neq 0$, $j = 1, 2$) of (2.4.20) can be expressed as

$$(x - x_1)(x - x_2)(x - x_3)xg_1(x) = 0, \\ (\text{resp. } (x - x_1)^2(x - x_2)xg_1(x) = 0,$$

which implies that Assertion (3) (resp. Assertion (4)) holds true.

The proof of the theorem is complete. \square

2.4.3 Case where $\text{Ind}(F, 0) = -2$

The vector field F is second-order nondegenerate at $x = 0$ near $\lambda = \lambda_0$. According to Theorem 2.4.3, under conditions (2.1.4) and (2.1.5) with $m = r = 2$, if the two vectors (a_{11}, a_{12}, a_{13}) and (b_{11}, b_{12}, b_{13}) are linearly independent near $\lambda = \lambda_0$, then there are at most three and at least one bifurcated singular points of (2.1.1) on each side of $\lambda = \lambda_0$.

By Theorem 2.4.1, the index of F defined by (2.4.13) at $x = 0$ takes only three values: $0, \pm 2$. We shall show that the transition of (2.1.1) from $(0, \lambda_0)$ is closely related to the index of F at $x = 0$.

The following theorem describes the transition of (2.1.1) from $(0, \lambda_0)$ for the case where $\text{ind}(F, 0) = -2$.

Theorem 2.4.4 *Let the conditions (2.1.4) and (2.1.5) with $m = r = 2$ hold true, $L_\lambda + G(\cdot, \lambda)$ be second-order nondegenerate at $(u, \lambda) = (0, \lambda_0)$, and $\text{ind}(F, 0) = -2$. Then we have the following assertions:*

- (1) *The transition of (2.1.1) from $(0, \lambda_0)$ is a jump transition, and has exactly three connected unstable domains.*
- (2) *Equation (2.1.1) bifurcates from $(0, \lambda_0)$ to exactly three saddle points on each side of $\lambda = \lambda_0$, which have the Morse index one.*
- (3) *The local topological structure of this transition is as shown in Fig. 2.11.*

Proof. We proceed in several steps as follows.

STEP 1. Under the conditions (2.1.4) and (2.1.5) with $m = r = 2$, the reduction equations of (2.1.1) are given by

$$\begin{aligned} \frac{dx_1}{dt} &= \beta_1(\lambda)x_1 + (G(x_1 e_1 + x_2 e_2 + \Phi(x, \lambda)), e_1^*(\lambda)), \\ \frac{dx_2}{dt} &= \beta_2(\lambda)x_2 + (G(x_1 e_1 + x_2 e_2 + \Phi(x, \lambda)), e_2^*(\lambda)), \end{aligned} \quad (2.4.35)$$

where $\Phi(x, \lambda)$ is the center manifold function satisfying $\Phi(x, \lambda) = o(|x|)$ for $x \in \mathbb{R}^2$. Since F is second-order nondegenerate at $x = 0$, (2.4.35) can be written as

$$\frac{dx}{dt} = J_\lambda x + F(x, \lambda) + o(|x|^2), \quad (2.4.36)$$

where $F(x, \lambda)$ is as in (2.4.13), and

$$J_\lambda x = \begin{pmatrix} \beta_1(\lambda) & 0 \\ 0 & \beta_2(\lambda) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \beta_1(\lambda)x_1 \\ \beta_2(\lambda)x_2 \end{pmatrix}.$$

Since F is second-order nondegenerate at $(x, \lambda) = (0, \lambda_0)$, the vector field on the right-hand side of (2.4.36) is a perturbation of $J_\lambda + F$ near $x = 0$. Hence, by the local structural stability of nondegenerate singular points, it suffices to prove this theorem for the following system, which is the first-order approximation of (2.4.36):

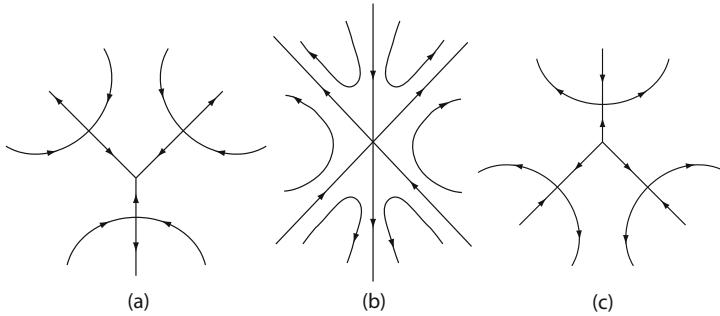


Fig. 2.11 Transition structure from a singular point with index $\text{ind}(F, 0) = -2$: (a) $\lambda < \lambda_0$; (b) $\lambda = \lambda_0$; (c) $\lambda > \lambda_0$

$$\frac{dx}{dt} = J_\lambda x + F(x, \lambda). \quad (2.4.37)$$

STEP 2. PROOF OF ASSERTION (2). By Theorem 2.4.1, if $\text{ind}(F, 0) = -2$, the two vectors (a_{11}, a_{12}, a_{22}) and (b_{11}, b_{12}, b_{22}) are linearly independent. Therefore it follows from Theorem 2.4.3 that (2.4.37) has at most three bifurcated singular points from $(0, \lambda_0)$. We shall prove that (2.4.37) has just three bifurcated singular points on each side of $\lambda = \lambda_0$.

It is known that

$$\text{ind}(J_\lambda + F, 0) = \text{sign}[\beta_1(\lambda)\beta_2(\lambda)] = 1 \quad \text{if } \lambda \neq \lambda_0,$$

$$\sum_{i=1}^k \text{ind}(J_\lambda + F, p_i) + \text{ind}(J_\lambda + F, 0) = \text{ind}(F(\cdot, \lambda_0), 0) = -2,$$

where p_i ($1 \leq i \leq k$) are the bifurcated singular points of (2.4.37) from $(0, \lambda_0)$. Hence if $\lambda \neq \lambda_0$, then

$$\sum_{i=1}^k \text{ind}(J_\lambda + F, p_i) = -3. \quad (2.4.38)$$

If the number $k < 3$ in (2.4.38), then one of the bifurcated singular points, say p_1 , of (2.4.37) satisfies that

$$|\text{ind}(J_\lambda + F, p_1)| \geq 2. \quad (2.4.39)$$

By the Brouwer degree theory, if the Jacobian matrix of $J_\lambda + F$ at p_1 is a nonzero matrix, i.e.,

$$D(J_\lambda + F)(p_1) \neq 0,$$

then we have

$$|\text{ind}(J_\lambda + F, p_1)| \leq 1.$$

Therefore, it follows from (2.4.39) that the Jacobian matrix of $J_\lambda + F$ at p_1 is zero:

$$D(J_\lambda + F)(p_1) = J_\lambda + \left(\frac{\partial F_i(p_1)}{\partial x_j} \right) = 0. \quad (2.4.40)$$

Let $p_1 = (z_1, z_2)$. Then we infer from (2.4.40) that

$$\begin{aligned} \beta_1 + 2a_{11}z_1 + a_{12}z_2 &= 0, \\ a_{12}z_1 + 2a_{22}z_2 &= 0, \\ \beta_2 + 2b_{22}z_2 + b_{12}z_1 &= 0, \\ b_{12}z_2 + 2b_{11}z_1 &= 0, \end{aligned}$$

which, together with $J_\lambda p_1 + F(p_1, \lambda) = 0$, imply that $p_1 = 0$, a contradiction to $p_1 \neq 0$. Thus, we have shown that $k = 3$. From (2.4.38) and Theorem 2.4.3, we deduce that

$$\text{ind}(J_\lambda + F, p_i) = -1 \quad \text{for } i = 1, 2, 3,$$

which implies that p_i ($1 \leq i \leq 3$) are saddle points. By Theorem 3.10 in Ma and Wang (2005b), the Morse index of p_i is one. Thus, Assertion (2) is proved.

STEP 3. PROOF OF ASSERTION (1). By Theorem 2.1.5 and Lemma 2.1.7, it suffices to prove that the vector field $F(x, \lambda_0)$ has only six hyperbolic regions, and has no elliptic and parabolic regions. To this end, we need to introduce a lemma, which is also useful in the later discussion for other cases.

Lemma 2.4.5 *Let $F(x, \lambda_0)$ be second-order nondegenerate at $x = 0$. Then for any λ near λ_0 the vector field $F(x, \lambda)$ has k straight orbit lines with $1 \leq k \leq 3$:*

$$\alpha_i x_1 + \beta_i x_2 = 0, \quad \alpha_i^2 + \beta_i^2 \neq 0, \quad 1 \leq i \leq k, \quad (2.4.41)$$

where $\sigma_i = -\alpha_i/\beta_i$ or $\sigma_i = -\beta_i/\alpha_i$ are the real solutions of the following algebraic equation

$$\begin{aligned} a_{22}\sigma^3 + (a_{12} - b_{22})\sigma^2 + (a_{11} - b_{12})\sigma - b_{11} &= 0, \\ \text{or} \\ b_{11}\sigma^3 + (b_{12} - a_{11})\sigma^2 + (b_{22} - a_{12})\sigma - a_{22} &= 0. \end{aligned} \quad (2.4.42)$$

In particular, if $\text{ind}(F, 0) = -2$, then $k = 3$.

Proof. When $F(x, \lambda)$ are second-order nondegenerate at $x = 0$, then $a_{11}^2 + b_{11}^2 \neq 0$. We assume that $a_{11} \neq 0$. By the homogeneity of F , a straight line $x_2 = \sigma x_1$ is an orbit line of $F(x, \lambda)$ if and only if

$$\sigma = \frac{F_2(x, \lambda)}{F_1(x, \lambda)} = \frac{b_{11}x_1^2 + b_{12}x_1x_2 + b_{22}x_2^2}{a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2} = \frac{b_{11} + b_{12}\sigma + b_{22}\sigma^2}{a_{11} + a_{12}\sigma + a_{22}\sigma^2}.$$

Hence, the straight lines (2.4.41) satisfying (2.4.42) are orbit lines of $F(x, \lambda)$. Obviously, one of the two equations in (2.4.42) must have at least one real solution and at most three real solutions. Thus, we derive that the number of straight orbit lines is k with $1 \leq k \leq 3$, as shown in Fig. 2.11b.

When $\text{ind}(F, 0) = -2$, by Assertion (2), the following system will bifurcate exactly to three singular points at $\lambda = \lambda_0$:

$$\frac{dx}{dt} = (\lambda - \lambda_0)x + F(x, \lambda),$$

and each singular point must be on only one straight orbit line of $F(x, \lambda)$. Hence, if $\text{ind}(F, 0) = -2$, then $k = 3$.

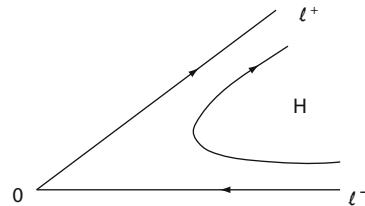
Thus the lemma is proved. \square

Now, we return to prove that $F(x, \lambda_0)$ has only six hyperbolic regions at $x = 0$. By Lemma 2.4.5, $F(x, \lambda_0)$ has exactly three straight orbit lines which intersect at $x = 0$, and the three straight lines divide the plane \mathbb{R}^2 into six invariant regions of $F(x, \lambda_0)$. By the Poincaré index formula (2.1.9), $F(x, \lambda_0)$ possesses at least six hyperbolic regions $H_i \subset \mathbb{R}^2$, and each boundary ∂H_i consists of two rays from $x = 0$:

$$\partial H_i = l_i^+ + l_i^- + \{0\},$$

where l_i^+ is an orbit of $F(x, \lambda_0)$ outward to $x = 0$, and l_i^- is an orbit inward to $x = 0$, as shown in Fig. 2.12. It shows that $F(x, \lambda_0)$ has exactly six hyperbolic regions which

Fig. 2.12 Orbits l^\pm of $F(x, \lambda_0)$



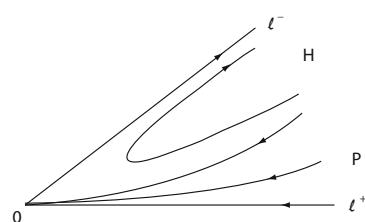
are separated by the three straight orbit lines. Therefore F has no elliptic regions, otherwise the index $\text{ind}(F, 0) \neq -2$.

Finally, we prove F has no parabolic regions P at $x = 0$. Otherwise there must be a parabolic region $P \subset \mathbb{R}^2$, such that for any $x_0 \in P$, the orbit $x(t, x_0)$ of $F(x, \lambda_0)$ satisfies

$$\lim_{t \rightarrow \infty} x(t, x_0) = 0,$$

where $x(0, x_0) = x_0$, and $x(t, x_0)$ is tangent to some straight orbit as shown in Fig. 2.13.

Fig. 2.13 Orbits of x in a sectorial region $S(\theta)$



We shall reach a contradiction. Without loss of generality, we take a straight orbit line of $F(x, \lambda_0)$ as x_1 -axis. Then, in this coordinate system, $F(x, \lambda_0)$ can be expressed as

$$F(x, \lambda_0) = \begin{pmatrix} a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2 \\ x_2(b_{12}x_1 + b_{22}x_2) \end{pmatrix}. \quad (2.4.43)$$

Since $\text{ind}(F, 0) = -2$, $b_{12} \neq 0$; otherwise by Theorem 2.4.1 we have $\text{ind}(F, 0) = 0$. Take a coordinate transformation:

$$x'_1 = x_1 + \frac{b_{22}}{b_{12}}x_2, \quad x'_2 = x_2.$$

Then the vector field (2.4.43) becomes (the primes are dropped for simplicity):

$$F(x, \lambda_0) = \begin{pmatrix} c_{11}x_1^2 + c_{12}x_1x_2 + c_{22}x_2^2 \\ b_{12}x_1x_2 \end{pmatrix}. \quad (2.4.44)$$

Since the coordinate system transformation preserves the topological structure up to orientation, the vector field (2.4.44) is topologically equivalent to F given by (2.4.43).

By Theorem 2.4.1 and $\text{ind}(F, 0) = -2$, we see that

$$b_{12} \cdot \frac{-c_{12} + \sqrt{\Delta}}{2c_{11}} < 0, \quad b_{12} \cdot \frac{-c_{12} - \sqrt{\Delta}}{2c_{11}} > 0,$$

where $\Delta = c_{12}^2 - 4c_{11}c_{22} > 0$. It follows that

$$\frac{b_{12}}{c_{11}} < 0 \quad (c_{11} \neq 0). \quad (2.4.45)$$

By the above conclusion, for $x_0 \in P$, the orbit $x(t, x_0)$ of (2.4.44) near $x = 0$ can be expressed by $x_2 = f(x_1)$ with

$$f(x_1) = o(x_1), \quad x_1 \geq 0. \quad (2.4.46)$$

It follows from (2.4.44) that

$$\frac{df}{dx_1} = \frac{b_{12}x_1x_2}{c_{11}x_1^2 + c_{12}x_1x_2 + c_{22}x_2^2} = \frac{b_{12}}{c_{11}} \frac{f(x_1)}{x_1} \left(1 + \frac{c_{12}}{c_{11}} \frac{f}{x_1} + \frac{c_{22}}{c_{11}} \frac{f^2}{x_1^2} \right)^{-1},$$

which, by (2.4.46), implies that

$$\frac{df}{f} \rightarrow \frac{b_{12}}{c_{11}} \frac{dx_1}{x_1} \quad \text{as } x_1 \rightarrow 0^+.$$

Namely

$$f(x_1) \rightarrow x_1^{-\sigma} \quad \text{as } x_1 \rightarrow 0^+, \quad (2.4.47)$$

where $\sigma = -b_{12}/c_{11} > 0$ (by (2.4.45)).

From (2.4.46) and (2.4.47) we reach a contradiction. Thus we have proved Assertion (1) of Theorem 2.4.4, and consequently we have proved Assertion (3). The proof of Theorem 2.4.4 is complete. \square

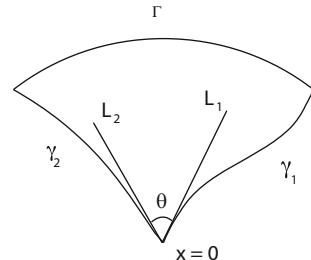
2.4.4 Case Where $\text{Ind}(F, \mathbf{0}) = 2$

For this case, in order to describe transition structure we need to introduce a notation. A set $S(\theta) \subset \mathbb{R}^2$ is called a sectorial region with angle $\theta \in [0, 2\pi]$, if $S(\theta)$ is enclosed by two curves γ_1 and γ_2 starting with $x = 0$ and an arc Γ , and the angle between the tangent lines L_1 and L_2 of γ_1 and γ_2 at $x = 0$ is θ ; see Fig. 2.14. Let $S_r(\theta)$ be the sectorial region with angle θ and radius $r > 0$ given by

$$S_r(\theta) = \{x \in S(\theta) \mid |x| < r\}.$$

The algebraic equations introduced by (2.4.42) is related to transition of (2.1.1). It is clear that the number k of real solutions of (2.4.42) satisfies $1 \leq k \leq 3$, which are associated with the straight orbit lines of vector field $F(x, \lambda)$. The following theorem connects the number k with transition of (2.1.1).

Fig. 2.14 Sectorial region with angle θ



Theorem 2.4.6 Let $\text{ind}(F, \mathbf{0}) = 2$, and the other conditions be the same as Theorem 2.4.4. Then, for the transition of (2.1.1) from $(0, \lambda_0)$ the following assertions hold true:

- (1) When the number $k > 1$, the transition is mixed, and when $k = 1$ the transition is either continuous or mixed depending on the higher-order terms than F .
- (2) If $\beta_1(\lambda) = \beta_2(\lambda)$ near $\lambda = \lambda_0$, the stable transition domain is a sectorial region with angle θ ($\pi < \theta \leq 2\pi$), given by

$$D_\theta = \{u \in U \subset X \mid u = x + v, \quad x = x_1 e_1 + x_2 e_2 \in S(\theta)\},$$

where $U \subset X$ is a neighborhood of $u = 0$.

- (3) If $\beta_1 = \beta_2$, the equation (2.1.1) bifurcates on $\lambda > \lambda_0$ to an attractor \mathcal{A}_λ , which attracts D_θ with $\dim \mathcal{A}_\lambda = 0$ or 1, and contains singular points as its minimal attractors.

- (4) If $k = 1$, the attractor \mathcal{A}_λ consists of a singular point p , and the transition structure is as shown in Fig. 2.15a. If $k = 2$ the attractor \mathcal{A}_λ is either a singular point as shown in Fig. 2.15a, or two singular points p_1, p_2 , and the orbit γ connecting them, as shown in Fig. 2.15b, or three singular points p_0, p_1, p_2 and two orbits γ_1, γ_2 connecting p_0 with p_1, p_2 as shown in Fig. 2.15c. If $k = 3$, \mathcal{A}_λ has the structure as shown in Fig. 2.15c.

Remark 2.4.7 If $\beta_1(\lambda) \neq \beta_2(\lambda)$ near λ_0 and $\lambda \neq \lambda_0$, Assertions (3)–(4) in Theorem 2.4.6 are not valid. In this case, the stable transition domain D may contain no attractor. For example, we consider the following equations

$$\begin{aligned}\frac{dx_1}{dt} &= 4\lambda x_1 + \lambda x_2 + x_1^2 - 4x_1 x_2 - x_2^2, \\ \frac{dx_2}{dt} &= -\lambda x_1 + x_1 x_2.\end{aligned}\quad (2.4.48)$$

It is easy to see that the two eigenvalues of (2.4.48) are

$$\beta_1(\lambda) = (2 + \sqrt{3})\lambda, \quad \beta_2(\lambda) = (2 - \sqrt{3})\lambda.$$

Namely $\beta_1(\lambda) \neq \beta_2(\lambda)$ as $\lambda \neq \lambda_0 = 0$. The nonlinear term

$$G(x_1, x_2) = \begin{pmatrix} x_1^2 - 4x_1 x_2 - x_2^2 \\ x_1 x_2 \end{pmatrix}$$

is second-order nondegenerate, and by Theorem 2.4.1 we can derive that $\text{ind}(G, 0) = 2$. In addition the algebraic equation (2.4.42) is

$$z^3 + 4z^2 = 0,$$

which has two real solutions $z = 0$ and $z = -4$. However, the system (2.4.48) has no attractors, and only one bifurcated singular point $(x_1, x_2) = (0, \lambda)$, the transition structure as shown in Fig. 2.16.

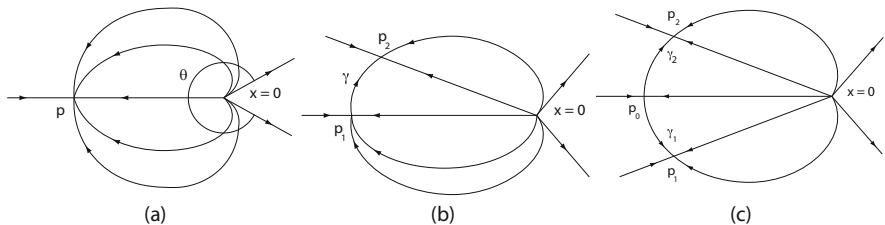
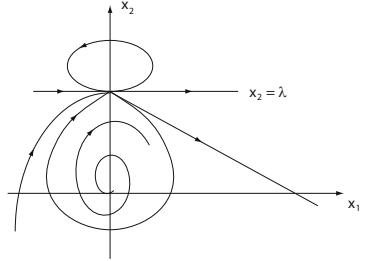


Fig. 2.15 (a) When $k = 1$, the attractor $\mathcal{A}_\lambda = \{p\}$. (b) When $k = 2$, $\mathcal{A}_\lambda = \gamma \cup \{p_1, p_2\}$. (c) When $k = 3$, $\mathcal{A}_\lambda = \gamma_1 \cup \gamma_2 \cup \{p_0, p_1, p_2\}$

Fig. 2.16 Transition for the Example in Remark 2.4.7



Proof of Theorem 2.4.6. As in the proof of Theorem 2.4.4, we only need to consider the system (2.4.37). Theorem 2.4.6 can be achieved in a few lemmas hereafter.

Lemma 2.4.8 *Let $\text{ind}(F, 0) = 2$, and k ($1 \leq k \leq 3$) be the number of real solutions of the algebraic equation (2.4.42). Then we have the following assertions:*

- (1) $F(x, \lambda)$ has no hyperbolic regions at $x = 0$;
- (2) $F(x, \lambda)$ has exactly two elliptic regions E_1 and E_2 ;
- (3) If $k = 1$, then $F(x, \lambda)$ has no parabolic regions. If $k \geq 2$, then $F(x, \lambda)$ has exactly two parabolic regions P_1 and P_2 such that the orbits in P_1 tend to $x = 0$, and in P_2 are away from $x = 0$;
- (4) The elliptic and parabolic regions E and P are sectorial regions $E = S(\theta_1)$, $P = S(\theta_2)$ with $0 < \theta_1, \theta_2 < \pi$, $\theta_1 + \theta_2 = \pi$, and

$$\theta_2 = \cot^{-1} \sigma_1 - \cot^{-1} \sigma_2,$$

where $\sigma_i = x_2/x_1$ ($i = 0, 1, 2$) are the real solutions of (2.4.42) with $\sigma_1 > \sigma_0 \geq \sigma_2$. Moreover, the edges of $S(\theta_1)$ and $S(\theta_2)$ are the straight orbit lines of $F(x, \lambda)$; see Fig. 2.17a–c.

Proof. Based on Lemma 2.4.5, as in the proof of (2.4.44), we can take a coordinate transformation $x' = Ax$ with $\det A > 0$ such that the vector field F can be transformed into the following form, where, for brevity, we omit the primes:

$$F(x, \lambda) = \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} = \begin{pmatrix} a(x_1 - \alpha_1 x_2)(x_1 + \alpha_2 x_2) \\ bx_1 x_2 \end{pmatrix}, \quad (2.4.49)$$

where $a \cdot b \neq 0$. Without loss of generality, we assume that $a, b > 0$. Thus, by Theorem 2.4.1 we have that $\alpha_1, \alpha_2 > 0$.

We can show that (2.4.49) has the topological structure as shown in Fig. 2.17a–c for $a, b > 0$.

To derive the topological structure in Fig. 2.17a–c of (2.4.49), let D_1, D_2, D_3 , and D_4 be the four open quadrants in \mathbb{R}^2 , and the two straight lines $x_1 - \alpha_1 x_2 = 0$ and $x_1 + \alpha_2 x_2 = 0$ also divide the plane \mathbb{R}^2 into four regions

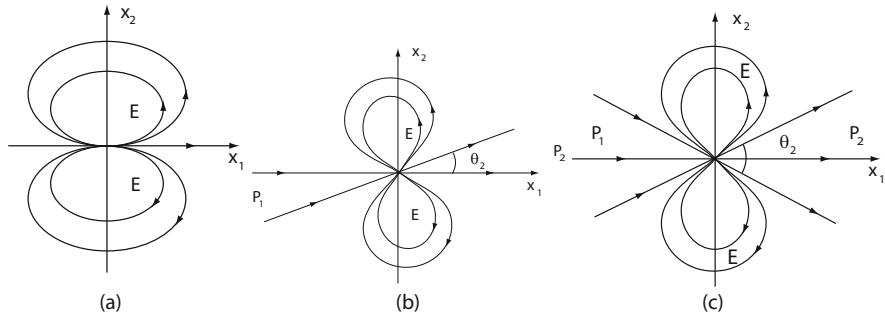


Fig. 2.17 Topological structure of (2.4.49): (a) The number of real solutions of (2.4.42) $k = 1$, (b) $k = 2$, (c) $k = 3$

$$Q_1 = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 - \alpha_1 x_2 > 0, x_1 + \alpha_2 x_2 > 0\},$$

$$Q_2 = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 - \alpha_1 x_2 < 0, x_1 + \alpha_2 x_2 > 0\},$$

$$Q_3 = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 - \alpha_1 x_2 < 0, x_1 + \alpha_2 x_2 < 0\},$$

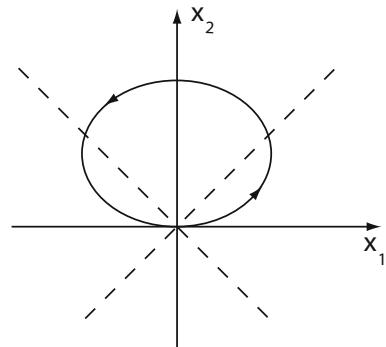
$$Q_4 = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 - \alpha_1 x_2 > 0, x_1 + \alpha_2 x_2 < 0\}.$$

It is easy to see that

$$\begin{aligned} F_1 &> 0 && \text{in } Q_1 \text{ and } Q_3 \\ F_1 &< 0 && \text{in } Q_2 \text{ and } Q_4, \\ F_2 &> 0 && \text{in } D_1 \text{ and } D_3, \\ F_2 &< 0 && \text{in } D_2 \text{ and } D_4. \end{aligned} \tag{2.4.50}$$

The properties (2.4.50) ensure that (2.4.49) has only two elliptic regions E_1 and E_2 , with $E_1 \subset \mathbb{R}_+^2 = D_1 + D_2 = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 > 0\}$ and $E_2 \subset \mathbb{R}_-^2 = D_3 + D_4 = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 < 0\}$; see Fig. 2.18.

Fig. 2.18 The dotted straight lines are $x_1 = -\alpha_2 x_2$ and $x_1 = \alpha_1 x_2$ respectively



Thus, by the Poincaré formula (2.1.9), F has no hyperbolic regions, and Assertions (1) and (2) are proved.

By (2.4.42), it follows from (2.4.49) that the straight orbit lines L_i ($i = 0, 1, 2$) of $F(x, \lambda)$ are given by

$$L_0 : x_2 = 0, \quad L_1 : x_2 = \sigma_1 x_1, \quad L_2 : x_2 = \sigma_2 x_1,$$

and if σ_1, σ_2 are real, then

$$\sigma_1 = \frac{1}{\alpha_1} - \varepsilon_1, \quad \sigma_2 = -\frac{1}{\alpha_2} + \varepsilon_2,$$

for some real numbers $0 < \varepsilon_1 < 1/\alpha_1$ and $0 < \varepsilon_2 < 1/\alpha_2$. Hence we have

$$L_i \subset Q_1 \bigcup Q_3 \quad \text{for } i = 0, 1, 2. \quad (2.4.51)$$

Therefore, Assertions (3) and (4) follow from (2.4.51) and the radial symmetry of $F(x, \lambda)$. The proof of the lemma is complete. \square

Lemma 2.4.9 *If $\beta_1(\lambda) = \beta_2(\lambda)$, and $\text{ind}(F, 0) = 2$, then (2.4.37) bifurcates to an attractor \mathcal{A}_λ on $\lambda > \lambda_0$, which attracts a sectorial region $S_r(\theta)$ with $\pi < \theta \leq 2\pi$. More precisely, $S_r(\theta) \subset (E_1 \cup E_2 \cup P_1) \cap B_r$, where E_1, E_2 are the elliptic regions of F , P_1 is the parabolic region where all orbits of F reach to $x = 0$, $B_r = \{x \in \mathbb{R}^2 \mid |x| < r\}$, $\theta = 2\pi - \theta_0$, and θ_0 the angle of the parabolic region P_2 .*

Proof. We know that under an orthogonal coordinate system transformation, the linear operator

$$J_\lambda = \begin{pmatrix} \beta(\lambda) & 0 \\ 0 & \beta(\lambda) \end{pmatrix}$$

is invariant. Therefore, without loss of generality, we take the vector field F as in the form of (2.4.43). By Theorem 2.4.1, $F(x, \lambda)$ can be written as

$$F = \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} = \begin{pmatrix} a(x_1 - \alpha_1 x_2)(x_1 + \alpha_2 x_2) \\ bx_2(x_1 - \sigma x_2) \end{pmatrix}, \quad (2.4.52)$$

where $a \cdot b \neq 0, \alpha_1 > 0, \alpha_2 > 0$.

We proceed with the case where $a, b > 0$ and $\sigma \geq 0$. The other case can be proved in the same fashion. By Theorem 2.4.1 we know that $\alpha_1 > \sigma > -\alpha_2$, which implies that the lines $x_1 + \alpha_2 x_2 = 0, x_1 - \sigma x_2 = 0, x_1 - \alpha_1 x_2 = 0$, and $x_2 = 0$ are alternatively positioned in \mathbb{R}^2 .

Based on the definition of elliptic and parabolic regions, by Lemma 2.4.8, we obtain that

$$\lim_{t \rightarrow \infty} S_\lambda(t)x = 0 \quad \forall x \in E_1 \bigcup E_2 \bigcup P_1, \quad (2.4.53)$$

where $S_\lambda(t)$ is the operator semigroup generated by $F(x, \lambda)$.

On the other hand, we obtain from (2.4.52) that for any $x \in E_1 \cup E_2 \cup P_1$, there is a $t_0(x) \geq 0$ such that

$$S_\lambda(t)x \in D = \{x \in \mathbb{R}^2 \mid x_1 - \sigma x_2 < 0\} \quad \forall t \geq t_0(x). \quad (2.4.54)$$

It is clear that $P_1 \subset D \subset E_1 \cup E_2 \cup P_1$. Let

$$D(r) = \{x \in D \mid |x| < r\}, \quad D(r_1, r_2) = \{x \in D \mid 0 < r_1 < |x| < r_2\},$$

and $T_\lambda(t)$ be the operator semigroup generated by $J_\lambda + F(\cdot, \lambda)$. It is known that for $\lambda > \lambda_0$ all orbits of $J_\lambda x$ are straight lines emitting outward from $x = 0$. Therefore, by (2.4.54), we deduce that

$$T_\lambda(t)x \in D \quad \forall t > 0, \quad x \in \partial D, \quad x \neq 0. \quad (2.4.55)$$

Now we prove that for any $\lambda - \lambda_0 > 0$ sufficiently small there are $r_1, r_2, r_3 > 0$ with $r_1 < r_2 < r_3$ such that

$$T_\lambda(t)x \in D(r_1, r_2) \quad \forall x \in D(r_3), \quad t > t_x, \quad (2.4.56)$$

for some $t_x \geq 0$.

We know that for $\lambda > \lambda_0$, the singular point $x = 0$ of $J_\lambda + F$ has an unstable manifold M^u with $\dim M^u = 2$. We take $r_1 > 0$ such that the ball $B_{r_1} \subset M^u$. Then, by (2.4.55), we obtain that

$$T_\lambda(t)x \in D(r_1, r_2) \quad \forall x \in D(r_1), \quad t > t_x. \quad (2.4.57)$$

If (2.4.56) is not true, then by (2.4.55) and (2.4.57) there exist $\lambda_n \rightarrow \lambda_0^+, t_n \rightarrow \infty$ and $\{x_n\} \subset D(r_3)$ such that

$$|T_{\lambda_n}(t_n)x_n| \geq r_2 \quad \forall n \geq 1. \quad (2.4.58)$$

Let $x_n \rightarrow x_0 \in D(r_3)$. Then by the stability of extended orbits, Theorem 2.1.15 in Ma and Wang (2005d), and (2.4.58), there is an extended orbit γ of $F(\cdot, \lambda_0)$ with starting point $x_0 \in D(r_3)$, which does not reach to $x = 0$. This is a contradiction to (2.4.53).

It follows from (2.4.55) and (2.4.56) that $D(r_1, r_2)$ is an absorbing set in a neighborhood U of $D(r_1, r_2)$. Hence, by the existence theorem of attractor (Theorem 2.2.11), for $\lambda > \lambda_0$, the set $\mathcal{A}_\lambda = \omega(D(r_1, r_2))$, with $0 \notin \mathcal{A}_\lambda$, is an attractor of (2.4.37), which attracts $D(r_3)$.

Applying Theorem 2.1.15 in Ma and Wang (2005d) again we infer from (2.4.53) that

$$\lim_{\lambda \rightarrow \lambda_0} \max_{x \in \mathcal{A}_\lambda} |x| = 0.$$

Thus, \mathcal{A}_λ is a bifurcated attractor of (2.4.37) from $(0, \lambda_0)$.

We can deduce from (2.4.54) that \mathcal{A}_λ attracts a sectorial region $S_r(\theta) \subset E_1 \cup E_2 \cup P_1$ with $\theta = 2\pi - \theta_0$, where θ_0 is the angle of the parabolic region P_2 . The proof is complete. \square

Lemma 2.4.10 *Under the hypotheses of Lemma 2.4.9, for the attractor \mathcal{A}_λ , we have the following assertions:*

- (1) If the number of real solutions of (2.4.42) $k = 1$, $\mathcal{A}_\lambda = \{p\}$, where p is a singular point of (2.4.37) with the Morse index zero, whose structure is as shown in Fig. 2.15a.
- (2) If $k = 2$, $\mathcal{A}_\lambda = \{p_1, p_2\} \cup \gamma$, where p_1, p_2 are two singular points of (2.4.37), p_1 is degenerate with index zero, p_2 nondegenerate with the Morse index zero, and γ the orbit connecting p_1 and p_2 , as shown in Fig. 2.15b.
- (3) If $k = 3$, $\mathcal{A}_\lambda = \{p_0, p_1, p_2\} \cup \gamma_1 \cup \gamma_2$, where p_0, p_1, p_2 are nondegenerate singular point with p_0 a saddle point, p_1, p_2 two attractors, and γ_1, γ_2 the orbits connecting p_0 with p_1 and p_2 , as shown in Fig. 2.15c.

Proof. Obviously, the attractor \mathcal{A}_λ contains all bifurcated singular points. We first need to prove that \mathcal{A}_λ does not contain extended orbits homeomorphic to S^1 .

By Lemma 2.4.5, all singular points of (2.4.37) must be in the straight orbit lines L of $F(x, \lambda)$, and L are invariant set of (2.4.37) which consist of orbits and singular points.

Using the method as in the proof of Lemma 2.4.9, for any straight orbit line L we can take an orthogonal coordinate system transformation with L as its x_1 -axis. Thus, the vector field $F(x, \lambda)$ takes the form of (2.4.52), and the bifurcated singular point $x_0 = (x_1^0, x_2^0)$ of $L_\lambda + F$ on L is as follows

$$x_0 = (x_1^0, x_2^0) = (-\beta(\lambda)/a, 0),$$

and the Jacobian matrix of $J_\lambda + F$ at x_0 is given by

$$D(J_\lambda + F)(x_0) = \begin{pmatrix} -\beta(\lambda) & * \\ 0 & (1 - \frac{b}{a})\beta(\lambda) \end{pmatrix}. \quad (2.4.59)$$

Hence, on each straight orbit line, there is only one singular point x_0 of $J_\lambda + F$, and there are two orbits r_1 and r_2 in L reaching to x_0 . Moreover, one of r_1 and r_2 connects $x = 0$ to x_0 . It follows that the attractor \mathcal{A}_λ containing all singular points has no closed extended orbits. By the Poincaré-Bendixon theorem we obtain that $\dim \mathcal{A}_\lambda \leq 1$, and the number k of bifurcated singular points is that of the straight orbit lines.

When $k = 1$, by $\text{ind}(F, 0) = 2$ we have

$$\text{ind}(J_\lambda + F, x_0) = 2 - \text{ind}(J_\lambda + F, 0) = 1.$$

Then, it follows from (2.4.59) that x_0 is an attractor. We shall prove that x_0 is nondegenerate. To this end, by (2.4.59) it suffices to show that $b/a > 1$. For (2.4.52) the equation (2.4.42) takes the form

$$z = \frac{bz(1 - \sigma z)}{a(1 - \alpha_1 z)(1 + \alpha_2 z)}, \quad (2.4.60)$$

which has a real solution $z = 0$. Then, (2.4.60) can be written as

$$\alpha_1 \alpha_2 z^2 + \left(\alpha_1 - \frac{b\sigma}{a} - \alpha_2 \right) z + \frac{b}{a} - 1 = 0. \quad (2.4.61)$$

Hence, (2.4.60) has no real solutions other than $z = 0$ if and only if the coefficients of (2.4.61) satisfy

$$\left(\frac{\sigma}{a} b + \alpha_2 - \alpha_1 \right)^2 < 4\alpha_1 \alpha_2 \left(\frac{b}{a} - 1 \right).$$

Since $\alpha_1, \alpha_2 > 0$, it implies that $b/a > 1$. Thus, we deduce that as $k = 1$, the singular point x_0 is nondegenerate, and has the Morse index zero. Therefore, $\mathcal{A}_\lambda = \{x_0\}$ has the structure as shown in Fig. 2.15a.

When $k = 2$, $J_\lambda + F$ has two singular points p_1 and p_2 ,

$$\text{ind}(J_\lambda + F, p_1) = 0, \quad \text{ind}(J_\lambda + F, p_2) = 1,$$

which implies, by Theorem 2.4.3 and (2.4.59), that p_2 is an attractor, and p_1 has exactly two hyperbolic regions. Thus, \mathcal{A}_λ has the topological structure as shown in Fig. 2.15b.

When $k = 3$, $J_\lambda + F$ has three singular points p_0, p_1, p_2 , and by Theorem 2.4.3, they are regular, and

$$\text{ind}(J_\lambda + F, p_0) = -1, \quad \text{ind}(J_\lambda + F, p_1) = \text{ind}(J_\lambda + F, p_2) = 1. \quad (2.4.62)$$

It follows from (2.4.59) and (2.4.62) that p_0 is a saddle point, p_1 and p_2 are attractors. In this case the structure of \mathcal{A}_λ is as shown in Fig. 2.15c. The proof of the lemma is complete. \square

Now, by Lemmas 2.4.8–2.4.10, we can derive Assertions (2)–(4) of Theorem 2.4.6, except the case where $k = 2$ in Assertion (4). In Lemma 2.4.10, as $k = 2$, $J_\lambda + F$ has a singular point p_1 with index zero. Hence, under a perturbation of the higher-order terms than F , near p_1 the system (2.4.36) has either no singular points, or one, or two singular points. The case of no singular points is as shown in Fig. 2.15a, and the case of one singular points as shown in Fig. 2.15b. For the case of two singular points, by (2.4.59), the indices of these singular points are +1 and -1, as shown in Fig. 2.15c. Hence Assertion (4) for $k = 2$ is valid.

Assertion (1) can be deduced by Lemma 2.4.8 and the Poincaré index formula (2.1.9), which illustrate that (2.4.36) bifurcates on $\lambda > \lambda_0$ to at least one singular point p with index one, such that p has at least an elliptic region. Hence, the transition is not a jump transition.

The proof of Theorem 2.4.6 is complete. \square

2.4.5 Case Where $\text{Ind}(F, \mathbf{0}) = 0$

The following theorem gives transition for the case of index zero.

Theorem 2.4.11 Let $\text{ind}(F, 0) = 0$, and the other conditions are the same as in Theorem 2.4.4. Then the transition of (2.1.1) from $(0, \lambda_0)$ is either jump or mixed. If it is a jump transition, then the structure is as shown in Fig. 2.19a. If it is a mixed transition, then the structure is as shown in Fig. 2.19b or c. In particular, for the number k of real solutions of (2.4.42), we have the following assertions:

- (1) if $k = 1$, the transition is jump;
- (2) if $k = 2$, the transition is either jump or mixed; and
- (3) if $k = 3$ and $\beta_1(\lambda) = \beta_2(\lambda)$ near $\lambda = \lambda_0$, then the transition is mixed, and (2.1.1) bifurcates on $\lambda > \lambda_0$ to three singular points p_0, p_1, p_2 , such that p_0 is an attractor which attracts a sectorial region $D_r(\theta)$ with $0 < \theta < \pi$, and p_1, p_2 are saddle points, as shown in Fig. 2.19c.

Proof. In the same fashion as used in the proof of Lemma 2.4.8, one can prove that if $\text{ind}(F, 0) = 0$, then the topological structure of $F(x, \lambda)$ at $x = 0$ near $\lambda = \lambda_0$ is just one of the three types as shown in Fig. 2.20a–c, depending on the number k of real solutions of (2.4.42).

On the other hand, the sum of indices of all singular points near $(x, \lambda) = (0, \lambda_0)$ is zero:

$$\sum_{i=1}^n \text{ind}(J_\lambda + g, z_i) + \text{ind}(J_\lambda + g, 0) = 0,$$

where $g(x, \lambda) = F(x, \lambda) + o(|x|^2)$, which implies that

$$\sum_{i=1}^n \text{ind}(J_\lambda + g, z_i) = -1. \quad (2.4.63)$$

When $J_\lambda + g$ has just one bifurcated singular point z_1 , by (2.4.63), $\text{ind}(J_\lambda + g, z_1) = -1$, which implies that z_1 is a saddle point. The structure of $F(x, \lambda_0)$ at $x = 0$ determines this transition structure of $J_\lambda + g$ as shown in Fig. 2.19a.

When $J_\lambda + g$ has two bifurcated singular points z_1 and z_2 , by Theorem 2.4.3, z_1 is regular and z_2 is degenerate. Hence, by (2.4.63) we can deduce that

$$\text{ind}(J_\lambda + g, z_1) = -1, \quad \text{ind}(J_\lambda + g, z_2) = 0.$$

i.e., z_1 is a saddle point, and z_2 has two hyperbolic regions. Thus, the transition structure is as shown in Fig. 2.19b.

When $J_\lambda + g$ has three bifurcated singular points, by Theorem 2.4.3, they are regular. Therefore, by (2.4.63), we have

$$\text{ind}(J_\lambda + g, z_1) = \text{ind}(J_\lambda + g, z_2) = -1, \quad \text{ind}(J_\lambda + g, z_3) = 1,$$

which show that the transition structure has to be topologically equivalent to that as shown in Fig. 2.19c.

Finally, similar to Theorem 2.4.6, we can obtain Assertions (1)–(3). Thus, the proof is complete. \square

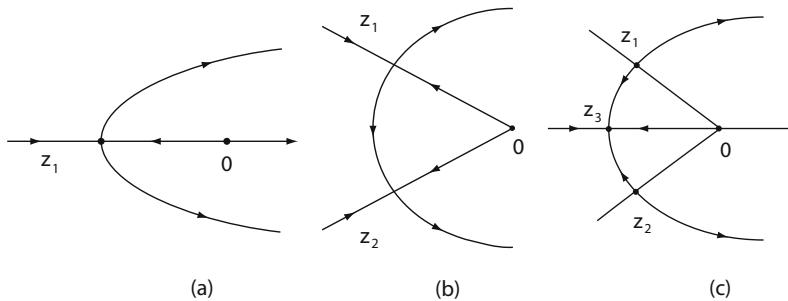
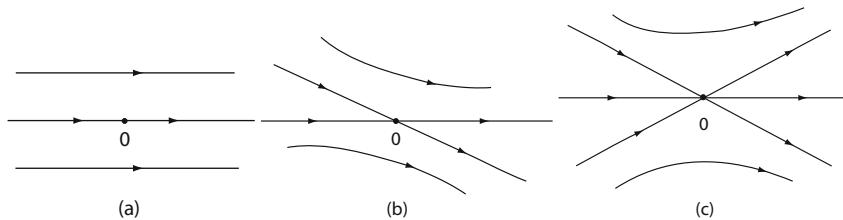


Fig. 2.19 Illustration for the proof of Theorem 2.4.11

Fig. 2.20 Topological structure of $F(x, \lambda)$ at $x = 0$: (a) $k = 1$; (b) $k = 2$; (c) $k = 3$

2.4.6 Indices of k -th Order Nondegenerate Singularities

In order to investigate the transition of (2.1.1) from k -th order nondegenerate singular points, it is necessary to study the index of k -homogeneous vector fields at $x = 0$. Let

$$F = \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} = \begin{pmatrix} \sum_{i+j=k} a_{ij}^1 x_1^i x_2^j \\ \sum_{i+j=k} a_{ij}^2 x_1^i x_2^j \end{pmatrix} \quad (2.4.64)$$

be a k -th order homogeneous field, with $k \geq 2$ an integer.

It is known that the index of F at $x = 0$ satisfies

$$\text{ind}(F, 0) = \begin{cases} \text{even} & \text{if } k = \text{even}, \\ \text{odd} & \text{if } k = \text{odd}. \end{cases}$$

Assume that F is k -th order nondegenerate, i.e., $x = 0$ is an isolated singular point of F . The components $F_i(x)$ ($i = 1, 2$) of F induce two k -th order polynomials as follows

$$F_i(z, 1) = a_{k0}^i z^k + a_{k-11}^i z^{k-1} + \cdots + a_{1,k-1}^i z + a_{0k}^i. \quad (2.4.65)$$

Since F is k -th order nondegenerate, $|a_{k0}^1| + |a_{k0}^2| \neq 0$. Without loss generality, we assume that $a_{k0}^1 \neq 0$; otherwise we can consider the case of $a_{k0}^2 \neq 0$.

The following theorem provides an index formula for the k -th order nondegenerate vector field (2.4.64), which is a generalized version of the second-order index formula (2.4.2).

Theorem 2.4.12 *Let $z_1 \geq \dots \geq z_N$ be N real zero points of $F_1(z, 1)$, i.e., $F_1(z_j, 1) = 0$, for any $1 \leq j \leq N$ (counting multiplicity). Then*

$$\text{ind}(F, 0) = \begin{cases} \sum_{j=1}^N (-1)^{j+1} \text{sign } F_2(z_j, 1) & \text{if } a_{k0}^1 > 0, N \geq 1, \\ \sum_{j=1}^N (-1)^j \text{sign } F_2(z_j, 1) & \text{if } a_{k0}^1 < 0, N \geq 1, \\ 0 & \text{if } N = 0. \end{cases} \quad (2.4.66)$$

Proof. We proceed in several steps as follows.

STEP 1. Since $x = 0$ is an isolated singular point of F , for any real zero point z_0 of $F_1(z, 1), F_2(z_0, 1) \neq 0$; otherwise it is easy to see that all points on the line $x_1 = z_0 x_2$ are singular points of F .

It is clear that when $k = 2$, the formula (2.4.66) is equivalent to (2.4.2).

When $N = 0$, $F_1(x_1, x_2)$ is either positive or negative definite. Hence either

$$(F_1(x_1, x_2), F_2(x_1, x_2)) = (-\varepsilon^2, 0) \quad \text{or} \quad (F_1(x_1, x_2), F_2(x_1, x_2)) = (\varepsilon^2, 0)$$

has no solutions for any $\varepsilon \neq 0$. Hence we have

$$\text{ind}(F, 0) = 0 \quad \text{for } N = 0.$$

STEP 2. Now we need to discuss the case where $N \geq 1$ and $a_{k0}^1 > 0$; the case for $a_{k0}^1 < 0$ can be proved in the similar fashion.

First, we consider the case where the real zero points $z_1 > \dots > z_N$ of $F_1(z, 1)$ are simple, which represent N straight lines:

$$L_i : x_1 = z_i x_2 \quad \text{for } 1 \leq i \leq N.$$

The N straight lines L_i ($1 \leq i \leq N$) divide the region $\mathbb{R}_+^2 = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 > 0\}$ into $N + 1$ sectorial regions D_0, D_1, \dots, D_N with

$$\begin{aligned} \partial D_i &= L_i^+ + L_{i+1}^+ \quad \text{for } i \neq 0, N, \\ \partial D_0 &= L_0^+ + L_1^+, \\ \partial D_N &= L_N^+ + L_0^-, \end{aligned} \quad (2.4.67)$$

where

$$\begin{aligned} L_i^+ &= \{(x_1, x_2) \in L_i \mid x_2 \geq 0\} && \text{for } i \neq 0, \\ L_0^+ &= \{(x_1, x_2) \mid x_1 \geq 0, x_2 = 0\}, \\ L_0^- &= \{(x_1, x_2) \mid x_1 \leq 0, x_2 = 0\}. \end{aligned}$$

We note, by $a_{k0}^1 > 0$, that

$$\begin{aligned} F_1(x_1, x_2) &= 0 && \text{on } L_i : x_1 = z_i x_2, \\ \text{sign } F_1(x_1, x_2) &= (-1)^i && \forall x = (x_1, x_2) \in D_i. \end{aligned} \quad (2.4.68)$$

Let Q_i ($1 \leq i \leq 4$) be the four quadrants in \mathbb{R}^2 . For convenience, we assume that $z_i \neq 0$ ($1 \leq i \leq N$). In fact, $z_i = 0$ corresponds to that L_i is the x_2 -axis, in this case, the discussion is the same. Thus, there is an l such that

$$\begin{aligned} D_i &\subset Q_1 + Q_4 && \forall 0 \leq i \leq l-1, \\ D_j &\subset Q_2 && \forall l+1 \leq j \leq N, \\ D_l \bigcap Q_1 &\neq \emptyset, \\ D_l \bigcap Q_2 &\neq \emptyset. \end{aligned}$$

When $N = 1$, k must be an odd number, and by (2.4.67) and $a_{k0}^1 > 0$, we have

$$F_1(x) \begin{cases} > 0 & \text{if } x \in D_0, \\ < 0 & \text{if } x \in D_1. \end{cases} \quad (2.4.69)$$

By $F(-x) = -F(x)$, it follows from (2.4.69) that $F(x)$ has only a parabolic region at $x = 0$ as shown in Fig. 2.21a and b, the two elliptic and two hyperbolic regions are as shown in Fig. 2.22a (or four, hyperbolic regions are as shown in Fig. 2.22b), and

$$F_2(z_1, 1) > 0 \quad (\text{or } F_2(z_1, 1) < 0). \quad (2.4.70)$$

By the Poincaré formula and (2.4.70) we derive that

$$\text{ind}(F, 0) = 1 + \frac{1}{2}(e - h) = \text{sign } F_2(z_1, 1) = \pm 1.$$

Hence, the formula (2.4.66) is valid for $N = 1$.

STEP 3. Let $N = 2$. We consider the structure of F near L_i^+ ($x_1 = z_i x_2, x_2 \geq 0$). We first discuss the case where $L_i^+ \subset Q_1$ ($i = 1, 2$), i.e., $z_1, z_2 > 0$, and for the case where $L_i^+ \subset Q_2$ ($i = 1, 2$) the discussion is the same.

First, when $F_1(x) > 0$ in D_0 , and $F_2(z_1, 1) > 0, F_2(z_2, 1) < 0$, there is an elliptic region E_1 containing L_1^+ , and the adjacent region in the left-hand side of E_1 , containing L_2^+ , is either an elliptic region or a parabolic region, as shown in Fig. 2.23a orb. In addition, because $N = 2$ and z_1 and z_2 are simple singular points of $F_1(z, 1)$, the number k in (2.4.64) is an even. Hence, if the region E_2 containing L_2^+ is elliptic,

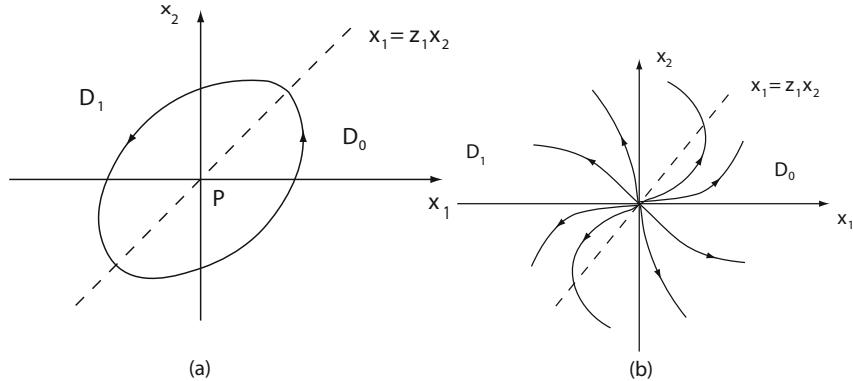


Fig. 2.21 Illustration of Step 2 of the proof of Theorem 2.4.12

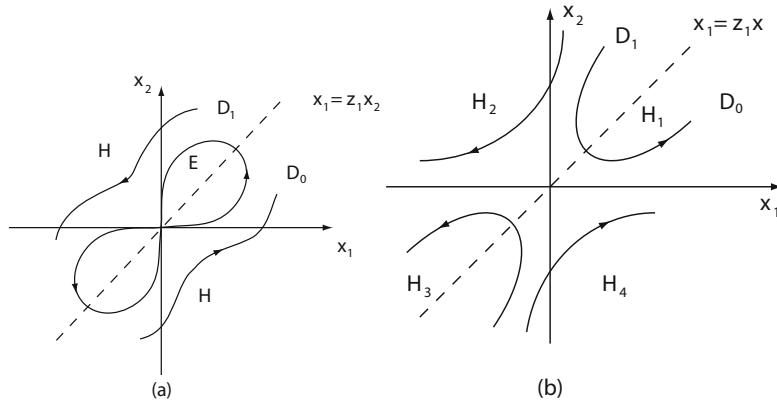


Fig. 2.22 Illustration of Step 2 of the proof of Theorem 2.4.12 (continued)

then the regions in the left-hand side of E_2 and in the right-hand side of E_1 are hyperbolic; see Fig. 2.23a.

Second, when $F_1(x) > 0$ in D_0 , and $F_2(z_1, 1) < 0, F_2(z_2, 1) > 0$, there are exactly three hyperbolic regions H_1, H_2, H_3 of F in $Q_1 + Q_2$, where H_1 contains L_1^+ , H_2 contains L_2^+ , as shown in Fig. 2.24.

Third, when $F_1(x) > 0$ in D_0 , and $F_2(z_1, 1) > 0, F_2(z_2, 1) > 0$, then either there are an elliptic region E_1 and two hyperbolic regions H_2, H_3 of F in $Q_1 + Q_2$ with E_1 containing L_1^+ and H_2 containing L_2^+ , as shown in Fig. 2.25a, or there are a parabolic region P containing L_1^+, L_2^+ and a hyperbolic region H , as shown in Fig. 2.25b.

Fourth, when $F_1(x) > 0$ in D_0 , and $F_2(z_1, 1) < 0, F_2(z_2, 1) < 0$, then either there are two hyperbolic regions H_0, H_1 and an elliptic region E_2 with H_1 containing L_1^+ and E_2 containing L_2^+ , or there are a parabolic region P containing L_1^+, L_2^+ and a hyperbolic region H in $Q_1 + Q_4$, as shown in Fig. 2.26a, b.

Now, we consider the case where $L_1^+ \subset Q_1$ and $L_2^+ \subset Q_2$, i.e., $D_0 \subset Q_1$, $D_2 \subset Q_2$, and $D_1 \cap Q_1 \neq \emptyset$, $D_1 \cap Q_2 \neq \emptyset$.

Fifth, when $F_1(x) > 0$ in D_0 , and $F_2(z_1, 1) > 0, F_2(z_2, 1) < 0$, there is an elliptic region E containing L_1^+ and L_2^+ in $Q_1 + Q_2$, as shown in Fig. 2.27a. As $F_2(z_1, 1) < 0, F_2(z_2, 1) > 0$, there are three hyperbolic regions H_1, H_2, H_3 in $Q_1 + Q_2$, where H_1 contains L_1^+ , H_3 contains L_2^+ , and H_2 is between L_1^+ and L_2^+ ; see Fig. 2.27b.

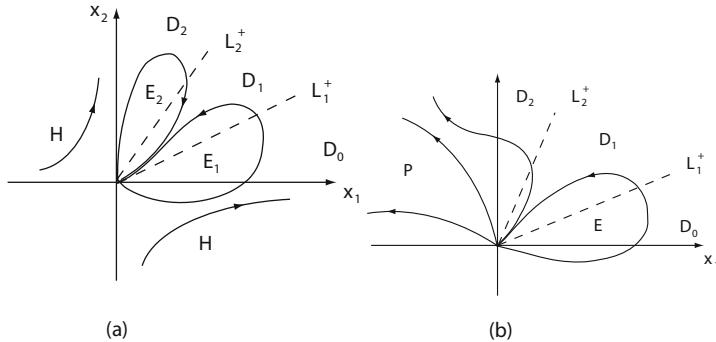
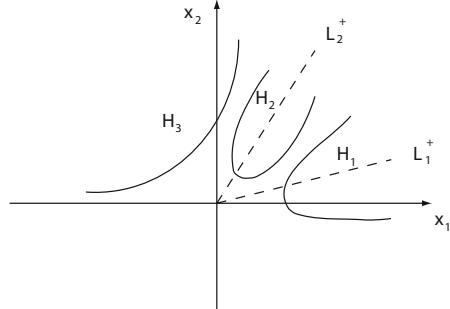


Fig. 2.23 Illustration of Step 3 for the proof of Theorem 2.4.12

Fig. 2.24 Illustration of Step 3 for the proof of Theorem 2.4.12 (continued)



Sixth, when $F_1(x) > 0$ in D_0 , and $F_2(z_1, 1) > 0, F_2(z_2, 1) > 0$, there are a parabolic region P containing L_2^+ and a hyperbolic region H containing L_1^+ as shown in Fig. 2.28a, and as $F_2(z_1, 1) < 0, F_2(z_2, 1) < 0$, the structure of F is as shown in Fig. 2.28b.

From the six properties above, it is easy to see that for $N = 2$ with $z_1 > z_2$, the following formula holds

$$\text{sign}F_2(z_1, 1) - \text{sign}F_2(z_2, 1) = 1 + \frac{1}{2}(e - h), \quad (2.4.71)$$

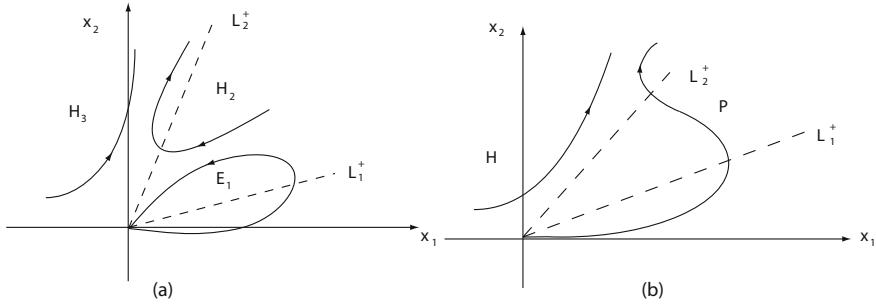


Fig. 2.25 Illustration of Step 3 for the proof of Theorem 2.4.12 (continued)

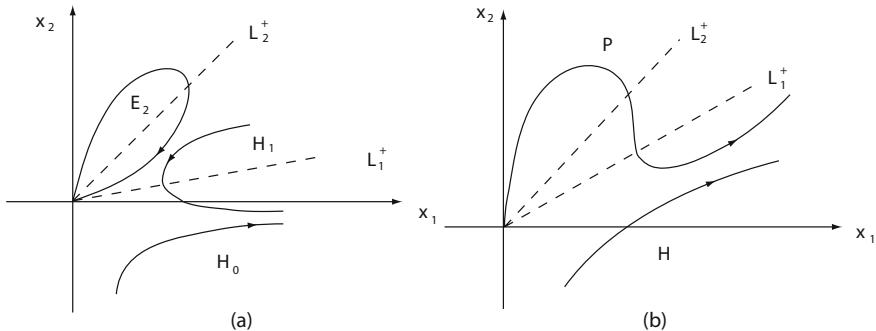


Fig. 2.26 Illustration of Step 3 for the proof of Theorem 2.4.12 (continued)

where \$e\$ and \$h\$ are the numbers of elliptic and hyperbolic regions. Hence the formula (2.4.66) with \$N = 2\$ follows from (2.4.71) and the Poincaré formula (2.1.9).

STEP 4. In the same fashion as above, for \$N \geq 3\$ we can draw from (2.4.68) the following conclusions:

1. Each elliptic region \$E\$ intersects with \$r\$ lines \$L_{i+j}\$ (\$1 \leq j \leq r\$), and the number \$r\$ satisfies

$$r = \begin{cases} 2m + 1 & \text{if } E \subset Q_1 + Q_4 \text{ or } E \subset Q_2 + Q_3, \\ 2m \geq 2 & \text{if } E \cap Q_1 \neq \emptyset \text{ and } E \cap Q_2 \neq \emptyset. \end{cases}$$

In addition, on the lines \$L_{i+j}\$ (\$x_1 = z_{i+j}x_2\$), for \$E \subset Q_1 + Q_4\$ or \$E \subset Q_2 + Q_3\$, we have

$$\text{sign}F_2(z_{i+1}, 1) = \cdots = \text{sign}F_2(z_{i+r}, 1),$$

and for \$E \cap Q_1 \neq \emptyset, E \cap Q_2 \neq \emptyset\$, the numbers of lines \$L_{i+s}^+ \subset E \cap Q_1\$ and \$L_{i+s}^+ \subset E \cap Q_2\$ are odd, i.e., \$l=odd, r-l=odd\$, and

$$\begin{aligned} \text{sign}F_2(z_{i+1}, 1) = \cdots = \text{sign}F_2(z_{i+l}, 1) & \quad \text{for } L_{i+s}^+ \subset Q_1, \forall 1 \leq s \leq l, \\ \text{sign}F_2(z_{i+l+1}, 1) = \cdots = \text{sign}F_2(z_{i+r}, 1) & \quad \text{for } L_{i+s}^+ \subset Q_2, \forall l+1 \leq s \leq r, \\ \text{sign}F_2(z_{i+1}, 1) = -\text{sign}F_2(z_{i+r}, 1) & \quad \text{for } L_{i+1}^+ \subset Q_1, \quad L_{i+r}^+ \subset Q_2. \end{aligned}$$

2. Each hyperbolic region H intersects with r lines L_{i+j} ($1 \leq j \leq r$), with the number r given by

$$r = \begin{cases} 2m + 1 & \text{if } H \subset Q_1 + Q_4 \text{ or } H \subset Q_2 + Q_3, \\ 2m \geq 0 & \text{if } H \cap Q_1 \neq \emptyset \text{ and } H \cap Q_2 \neq \emptyset. \end{cases}$$

On the lines L_{i+j} ($x_1 = z_{i+j}x_2$), for $H \subset Q_1 + Q_4$ or $H \subset Q_2 + Q_3$

$$\text{sign}F_2(z_{i+1}, 1) = \cdots = \text{sign}F_2(z_{i+r}, 1),$$

and for $H \cap Q_1 \neq \emptyset, H \cap Q_2 \neq \emptyset$, the numbers of lines $L_{i+s}^+ \subset E \cap Q_1$ and $L_{i+s}^+ \subset E \cap Q_2$ are even: $l=\text{even}$, $r-l=\text{even}$, and

$$\begin{aligned} \text{sign}F_2(z_{i+1}, 1) &= \cdots = \text{sign}F_2(z_{i+r}, 1) && \text{for } L_{i+s}^+ \subset Q_1, \quad 1 \leq s \leq l, \\ \text{sign}F_2(z_{i+l+1}, 1) &= \cdots = \text{sign}F_2(z_{i+r}, 1) && \text{for } L_{i+s}^+ \subset Q_2, \quad l+1 \leq s \leq r, \\ \text{sign}F_2(z_{i+1}, 1) &= -\text{sign}F_2(z_{i+r}, 1) && \text{for } L_{i+1}^+ \subset Q_1, \quad L_{i+r}^+ \subset Q_2. \end{aligned}$$

3. Each parabolic region P intersects with r lines L_{i+j} ($1 \leq j \leq r$), with

$$r = \begin{cases} 2m \geq 0 & \text{if } P \subset Q_1 + Q_4 \text{ or } P \subset Q_2 + Q_3, \\ 2m + 1 & \text{if } P \cap Q_1 \neq \emptyset \text{ and } P \cap Q_2 \neq \emptyset. \end{cases}$$

Moreover, on the lines L_{i+j} ($x_1 = z_{i+j}x_2$),

- a. if $P \subset Q_1 + Q_4$ or $P \subset Q_2 + Q_3$, then only at even number of successive points $z_{i+s+1}, \dots, z_{i+s+2k}$, the signs of $F_2(z, 1)$ are the same:

$$\text{sign}F_2(z_{i+s+1}, 1) = \cdots = \text{sign}F_2(z_{i+s+2k}, 1), \quad s+2k \leq r,$$

- b. if $P \cap Q_1 \neq \emptyset, P \cap Q_2 \neq \emptyset$, then as $P \cap Q_1$ (or $P \cap Q_2$) contains odd number of lines $L_{i+1}^+, \dots, L_{i+2k+1}^+$ (or $L_{i+2k+1}^+, \dots, L_{i+r}^+$), the signs of $F_2(z, 1)$ are the same at the odd number of successive points $z_{i+2k+1} < \cdots < z_{i+2j+1}$ ($j \leq k$) (or $z_{i+2k+1} > \cdots > z_{i+2j+1}, j \geq k$):

$$\begin{aligned} \text{sign}F_2(z_{i+2j+1}, 1) &= \cdots = \text{sign}F_2(z_{i+2k+1}, 1) && \forall 0 \leq j \leq k, \\ (\text{or } \text{sign}F_2(z_{i+2k+1}, 1) &= \cdots = \text{sign}F_2(z_{i+2j+1}, 1) && \forall k \leq j), \end{aligned}$$

and the signs of $F_2(z, 1)$ are the same at the other even number of successive points of z_{i+j} ($1 \leq j \leq r$).

4. Let E be an elliptic region containing $r \geq 1$ lines L_{i+j} ($1 \leq j \leq r$).

- a. If

$$\begin{aligned} \text{sign}F_2(z_{i+r+1}, 1) &= -\text{sign}F_2(z_{i+r}, 1) \\ (\text{resp. } \text{sign}F_2(z_i, 1) &= -\text{sign}F_2(z_{i+1}, 1)), \end{aligned} \tag{2.4.72}$$

then the adjacent region of E intersecting with L_{i+r+1} (resp. with L_i), which is in the left-hand side (resp. in the right-hand side) of E , is either elliptic or parabolic.

b. If

$$\begin{aligned} \text{sign}F_2(z_{i+r+1}, 1) &= \text{sign}(z_{i+r}, 1) \\ (\text{resp. } \text{sign}F_2(z_i, 1) &= \text{sign}(z_{i+1}, 1)), \end{aligned} \quad (2.4.73)$$

then the adjacent region intersecting with L_{i+r+1} (resp. with L_i) is either hyperbolic or parabolic.

5. Let H be a hyperbolic region intersecting with $r(\geq 1)$ lines L_{i+j} ($1 \leq j \leq r$).

- a. If (2.4.72) holds true, then the adjacent region of H intersecting with L_{i+r+1} (resp. with L_i) is either hyperbolic or parabolic.
- b. If (2.4.73) holds true, then the adjacent region intersecting with L_{i+r+1} (resp. with L_i) is either elliptic or parabolic.

6. Let P be a parabolic region intersecting with $r(\geq 0)$ lines L_{i+j} ($1 \leq j \leq r$). Let $P \subset Q_1$ or $P \subset Q_2$.

- a. As the right adjacent region of P , which intersects with L_i , is elliptic, if

$$\text{sign}F_2(z_i, 1) = -\text{sign}F_2(z_{i+r+1}, 1) \quad (2.4.74)$$

then the left adjacent region of P , which intersects with L_{i+r+1} , must be elliptic, and if

$$\text{sign}F_2(z_i, 1) = \text{sign}F_2(z_{i+r+1}, 1) \quad (2.4.75)$$

then the left adjacent region is hyperbolic.

- b. As the right adjacent region of P is hyperbolic, if (2.4.74) holds true, then the left adjacent region is also hyperbolic, and if (2.4.75) holds true, then the left one is elliptic.

7. Let P be as in (6), and $P \cap Q_1 \neq \emptyset, P \cap Q_2 \neq \emptyset$. As the right adjacent region of P is elliptic, if (2.4.74) holds true, then the left one is hyperbolic, and if (2.4.75) holds, then the left one is elliptic.

For $z_1 > z_2 > \dots > z_N$ ($N \geq 3$), from the Properties 1–7 above, we can see that it suffices to prove (2.4.66) only for the following cases:

- (i) each elliptic region E with $E \subset Q_1$ or $E \subset Q_2$ intersects only with one line L_i ($x_1 = z_i x_2$);
- (ii) the elliptic region E with $E \cap Q_1 \neq \emptyset, E \cap Q_2 \neq \emptyset$ (if exists) intersects only with L_l and L_{l+1} where $L_l \subset Q_1, L_{l+1} \subset Q_2$;
- (iii) each hyperbolic region H with $H \subset Q_1$ or $H \subset Q_2$ intersects only with one line L_i ($1 \leq i \leq N$);
- (iv) the hyperbolic region H with $H \cap Q_1 \neq \emptyset, H \cap Q_2 \neq \emptyset$ (if exists) intersects with no line, which is between L_l and L_{l+1} , where $L_l \subset Q_1$ and $L_{l+1} \subset Q_2$.

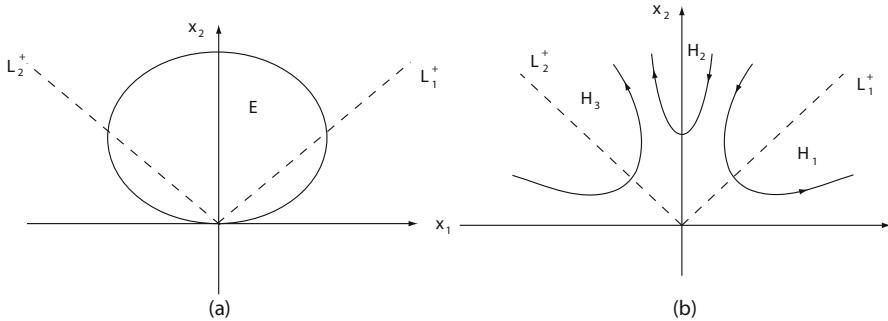


Fig. 2.27 Illustration of Step 3 for the proof of Theorem 2.4.12 (continued)

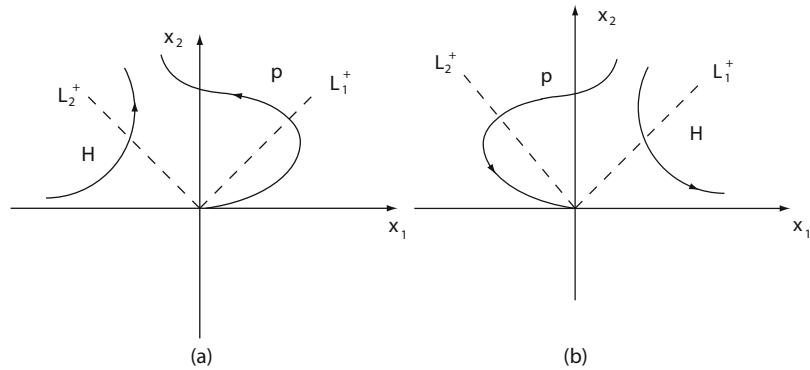


Fig. 2.28 Illustration of Step 3 for the proof of Theorem 2.4.12 (continued)

- (v) no parabolic regions in \$Q_1\$ and \$Q_2\$, and if there is a parabolic region \$P\$ with \$P \cap Q_1 \neq \emptyset, P \cap Q_2 \neq \emptyset\$, then \$P\$ intersects only with either \$L_l\$ or \$L_{l+1}\$, where \$L_l \subset Q_1\$ and \$L_{l+1} \subset Q_2\$.

STEP 5. Based on the properties (i)–(v), we now prove the index formula (2.4.66). Obviously, only one of the elliptic, hyperbolic, and parabolic regions, denoted by \$R\$, satisfies that \$R \cap Q_1 \neq \emptyset\$ and \$R \cap Q_2 \neq \emptyset\$. Hence, we proceed in the following three cases.

CASE 1. There is an elliptic region \$E_l\$ satisfying that \$E_l \cap Q_1 \neq \emptyset\$ and \$E_l \cap Q_2 \neq \emptyset\$. In this case, by property (v) there are no parabolic regions, and all hyperbolic regions are in either \$Q_1\$ or \$Q_2\$.

For \$a_{k0}^1 > 0\$, as the first and third properties above, it is ready to verify that the region containing \$L_1^+\$ is elliptic if and only if \$F_2(z_1, 1) > 0\$. Hence, by the previous properties 4 and 5, we deduce that

$$\begin{aligned} (-1)^{j+1} \operatorname{sign} F_2(z_j, 1) = 1 &\Leftrightarrow \text{the region containing } L_j^+ \text{ is elliptic,} \\ (-1)^{j+1} \operatorname{sign} F_2(z_j, 1) = -1 &\Leftrightarrow \text{the region containing } L_j^+ \text{ is hyperbolic.} \end{aligned} \quad (2.4.76)$$

By properties (i)–(iv), only the elliptic region E_l contains two lines L_l^+ and L_{l+1}^+ , and

$$(-1)^k \operatorname{sign} F_2(z_k, 1) = 1, \quad k = l, l + 1.$$

Hence, the numbers of elliptic and hyperbolic regions are

$$\begin{aligned} e &= 2 \left[-1 + \text{number of } L_j^+ \text{ with } (-1)^{j+1} \operatorname{sign} F_2(z_j, 1) = 1 \right], \\ h &= 2 \times \text{number of } L_j^+ \text{ with } (-1)^{j+1} \operatorname{sign} F_2(z_j, 1) = -1. \end{aligned}$$

Therefore, by the Poincaré formula (2.1.9), we derive that

$$\operatorname{ind}(F, 0) = 1 + \frac{1}{2}(e - h) = \sum_{j=1}^m (-1)^{j+1} \operatorname{sign} F_2(z_j, 1). \quad (2.4.77)$$

CASE 2. There is a hyperbolic region H_l such that $H_l \cap Q_1 \neq \emptyset$ and $H_l \cap Q_2 \neq \emptyset$.

Since there are no parabolic regions in \mathbb{R}^2 , the conclusion (2.4.76) is valid. By Property (iv), this hyperbolic region H_l is the unique one intersecting with no lines of L_j ($1 \leq j \leq m$). Hence, the numbers of hyperbolic and elliptic regions are

$$\begin{aligned} h &= 2 \left[1 + \text{number of } L_j^+ \text{ with } (-1)^{j+1} \operatorname{sign} F_2(z_j, 1) = -1 \right], \\ e &= 2 \times \text{number of } L_j^+ \text{ with } (-1)^{j+1} \operatorname{sign} F_2(z_j, 1) = 1. \end{aligned}$$

Thus we derive (2.4.77).

CASE 3. There is a parabolic region P satisfying that $P \cap Q_1 \neq \emptyset$, and $P \cap Q_2 \neq \emptyset$.

By property (v), the parabolic region P is unique, and P intersects with only one line L_l or L_{l+1} .

We know that the parabolic region $P \subset Q_1 + Q_2$ can be characterized by the orientation as follows

$$\begin{aligned} P^+ &= \{x \in P \mid S(t)x \rightarrow 0, \quad t \rightarrow -\infty\}, \\ P^- &= \{x \in P \mid S(t)x \rightarrow 0, \quad t \rightarrow +\infty\}, \end{aligned}$$

where $S(t)$ is the semigroup generated by $F(x)$. It is not difficult to get that (see Fig. 2.28a, b)

$$\begin{cases} F_1(x) > 0 & \forall x \in D_{l-1} \quad \text{if } P = P^+ \text{ contains } L_l^+, \\ F_1(x) > 0 & \forall x \in D_l \quad \text{if } P = P^+ \text{ contains } L_{l+1}^+, \end{cases} \quad (2.4.78)$$

$$\begin{cases} F_1(x) < 0 & \forall x \in D_{l-1} \quad \text{if } P = P^- \text{ contains } L_l^+, \\ F_1(x) < 0, & \forall x \in D_l \quad \text{if } P = P^- \text{ contains } L_{l+1}^+, \end{cases} \quad (2.4.79)$$

where D_j with $\partial D_j = L_j^+ + L_{j+1}^+$ are the domains as in (2.4.67), and

$$\begin{cases} F_2(z_l, 1) > 0 & \text{if } P = P^+ \text{ contains } L_l^+, \\ F_2(z_{l+1}, 1) > 0 & \text{if } P = P^+ \text{ contains } L_{l+1}^+, \end{cases} \quad (2.4.80)$$

$$\begin{cases} F_2(z_l, 1) < 0 & \text{if } P = P^- \text{ contains } L_l^+, \\ F_2(z_{l+1}, 1) < 0 & \text{if } P = P^- \text{ contains } L_{l+1}^+. \end{cases} \quad (2.4.81)$$

On the other hand, by assumption, $a_{k0}^1 > 0$, i.e., (2.4.68) holds. Thus it follows from (2.4.78) and (2.4.79) that

$$l = \begin{cases} \text{odd} & \text{if } P = P^+ \text{ contains } L_l^+, \\ \text{even} & \text{if } P = P^+ \text{ contains } L_{l+1}^+, \end{cases} \quad (2.4.82)$$

$$l = \begin{cases} \text{even} & \text{if } P = P^- \text{ contains } L_l^+, \\ \text{odd} & \text{if } P = P^- \text{ contains } L_{l+1}^+. \end{cases} \quad (2.4.83)$$

Therefore, from (2.4.80) to (2.4.83), we obtain that

$$(-1)^{j+1} \operatorname{sign} F_2(z_j, 1) = 1 \quad \text{as } P \text{ contains } L_j^+. \quad (2.4.84)$$

Thus, we deduce from (2.3.2) that

$$e + 2 = 2 \times \text{number of } L_j^+ \text{ with } (-1)^{j+1} \operatorname{sign} F_2(z_j, 1) = 1,$$

$$h = 2 \times \text{number of } L_j^+ \text{ with } (-1)^{j+1} \operatorname{sign} F_2(z_j, 1) = -1,$$

which implies that (2.4.77) holds true.

Thus, the formula (2.4.66) is proved for the case where $z_1 > \dots > z_N$ are simple singular points of $F_1(z, 1)$.

STEP 6. Now, we consider the case where z_j may be multiple singular points of $F_1(z, 1)$. Without loss of generality, we assume the multiplicity of z_1 is $r > 1$, and the others are one, i.e.,

$$z_1 = \dots = z_r < z_{r+1} < \dots < z_N. \quad (2.4.85)$$

Thus, we have that

$$F_1(x) = a_{k0}^1 (x_1 - z_1 x_2)^r (x_1 - z_{r+1} x_2) \cdots (x_1 - z_N x_2) f(x) \quad (2.4.86)$$

where $f(x)$ is a positive definite $(k-m)$ -order polynomial.

When $r=\text{even}$, we see that

$$\sum_{j=1}^N (-1)^{j+1} \operatorname{sign} F_2(z_j, 1) = \sum_{i=1}^{N-r} (-1)^{i+1} \operatorname{sign} F_2(z_{r+i}, 1).$$

In other hand, by (2.4.86) we see that the lines $L_i : x_1 = z_{r+i} x_2 (1 \leq i \leq N-r)$ with $z_{r+1} > \dots > z_{r+N}$ divide the semi-plane $\mathbb{R}_+^2 = \{x \in \mathbb{R}^2 \mid x_2 > 0\}$ into $N-r+1$ sectorial domains D_0, \dots, D_{N-r} as given by (2.4.67), in where (2.4.68) is valid except in D_0 , and

$$\text{sign}F_1(x) = 1 \quad \text{as} \quad x \in D_0/L, \quad L : x_1 = z_1 x_2.$$

Thus, the proof above is valid for the case (2.4.85) with $r=\text{even}$.

When $r=\text{odd}$, we have

$$\sum_{j=1}^N (-1)^{j+1} \text{sign}F_2(z_j, 1) = \sum_{i=1}^{N-r+1} (-1)^{i+1} \text{sign}F_2(z_{r+i-1}, 1).$$

By (2.4.86), the lines $L_i : x_1 = z_{r+i-1}x_2$ ($1 \leq i \leq m-r+1$) with $z_r > \dots > z_N$ divide \mathbb{R}_+^2 into $N-r+2$ domains D_0, \dots, D_{N-r+1} , where (2.4.68) is valid. Thus, the proof above is still valid.

Thus, the proof of the theorem is complete. \square

Remark 2.4.13 When $a_{k0}^1 = 0$, and $a_{k0}^2 \neq 0$, if $z_1 \geq \dots \geq z_N$ are N real zero points of $F_2(z, 1)$, then

$$\text{ind}(F, 0) = \begin{cases} 0 & \text{if } N = 0, \\ \sum_{j=1}^N (-1)^j \text{sign}F_1(z_j, 1) & \text{if } a_{k0}^2 > 0, \quad N \neq 0, \\ \sum_{j=1}^N (-1)^{j+1} \text{sign}F_1(z_j, 1) & \text{if } a_{k0}^2 < 0, \quad N \neq 0. \end{cases} \quad (2.4.87)$$

Remark 2.4.14 The formula (2.4.66) shown that for a k -homogeneous vector field (2.4.64), the index of F at $x = 0$ takes values in the following range:

$$\text{ind}(F, 0) = \begin{cases} 0, \pm 2, \dots, \pm k & \text{if } k = \text{even}, \\ \pm 1, \pm 3, \dots, \pm k & \text{if } k = \text{odd}. \end{cases}$$

2.4.7 Structure of k -th Order Nondegenerate Singularities

In the previous subsections, we studied the transition and its structure from second-order nondegenerate singularities. Now, we shall consider the general k -th order case. For this purpose, in this subsection we investigate the topological structure of a nondegenerate k -homogeneous vector field.

Let $F(x) = (F_1(x), F_2(x))$ be a two-dimensional k -homogeneous nondegenerate vector field given by (2.4.64), and $F_i(z, 1)$ the polynomials induced by $F_i(x_1, x_2)$ ($i = 1, 2$). We introduce another polynomials associated with F :

$$h(z) = a_{k0}^2 z^{k+1} + \sum_{i=1}^k (a_{k-i i}^2 - a_{k-i+1 i-1}^1) z^{k-i+1} - a_{0k}^1. \quad (2.4.88)$$

The polynomial $h(z)$ is derived by

$$z = \frac{F_1(z, 1)}{F_2(z, 1)}, \quad (2.4.89)$$

i.e., $h(z) = zF_2(z, 1) - F_1(z, 1)$. Hence, it is clear that a real zero point z_0 of $h(z)$ represents a straight orbit line L of $F(x)$, given by $L : x_1 = z_0 x_2$.

Hereafter, we always assume that $a_{k0}^1 \neq 0$, i.e., we take the index formula (2.4.66) in our discussion, otherwise we take (2.4.87). Thus, the topological structure of F is completely determined by the zero points of $h(z)$ and $F_1(z, 1)$.

Because the index of F at $x = 0$ is related to only the odd-multiple zero points of $F_1(z, 1)$, we take all zero points of $F_1(z, 1)$ with odd-multiplicity as $z_1 > \dots > z_N$, and all real zero points of $h(z)$ as $\alpha_1 > \dots > \alpha_K$, which correspond to the straight lines as follows

$$\begin{aligned} L_i : x_1 &= z_i x_2 & \forall 1 \leq i \leq N, \\ l_j : x_1 &= \alpha_j x_2 & \forall 1 \leq j \leq K. \end{aligned}$$

Notice that $F(-x) = (-1)^k F(x)$. It suffices to consider the structure of $F(x)$ only in the semi-plane

$$\mathbb{R}_-^2 = \{x \in \mathbb{R}^2 \mid x_1 - \alpha_1 x_2 < 0\}.$$

Thus, the lines l_j ($1 \leq j \leq K$) divide \mathbb{R}_-^2 into K sectorial domains:

$$\begin{aligned} S_i &= \{(x_1, x_2) \in \mathbb{R}_-^2 \mid x_1 - \alpha_{i+1} x_2 > 0, \quad x_1 - \alpha_i x_2 < 0\}, \quad 1 \leq i \leq K-1, \\ S_K &= \{(x_1, x_2) \in \mathbb{R}_-^2 \mid x_1 - \alpha_{K-1} x_2 < 0, \quad x_1 - \alpha_1 x_2 < 0\}. \end{aligned}$$

Since $F(x)$ is homogeneous, we have the following claim.

Claim A *Each sectorial domain S_j ($1 \leq j \leq K$) must be one of the elliptic, hyperbolic, and parabolic regions, as shown in Fig. 2.29. Hence, if we can make sure the type of regions and its orientation in each sectorial domain $S_j \subset \mathbb{R}_-^2$, then the structure of F in \mathbb{R}^2 is completely determined.*

Let S_i be a sectorial domain as above. In the following, we give a method to determine the region type of S_i . By definition,

$$\partial S_i = l_i + l_{i+1}, \quad l_j : x_1 = \alpha_j x_2.$$

We only consider the case where $a_{k0}^2 \neq 0$, i.e., $F_2(x_1, 0) \neq 0$, and for the case where $a_{k0}^2 = 0$, the discussion is the same. Hence, by $\alpha_1 > \alpha_2 > \dots > \alpha_K$, we have

$$x_1 = \alpha_j x_2 \quad (1 \leq j \leq K) \quad \text{and} \quad (x_1, x_2) \in \mathbb{R}_-^2 \Rightarrow x_2 > 0.$$

Let E, H, P be the elliptic, hyperbolic, and parabolic regions. Then

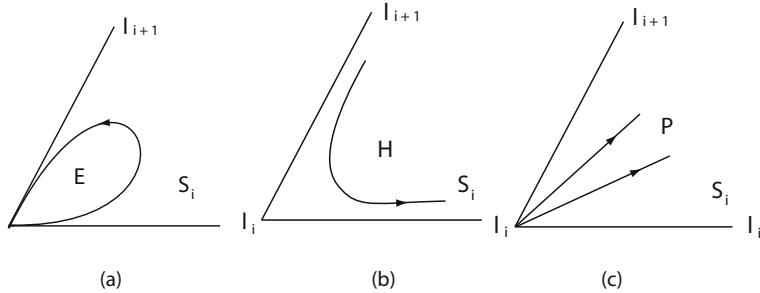


Fig. 2.29 (a) S_i is an elliptic region; (b) S_i is a hyperbolic region; (c) S_i is a parabolic region

$$\begin{cases} S_i = P \Leftrightarrow \text{sign}F_2(\alpha_i, 1) = \text{sign}F_2(\alpha_{i+1}, 1) & \forall i \neq K, \\ S_K = P \Leftrightarrow \text{sign}F_2(\alpha_K, 1) = (-1)^{k+1} \text{sign}F_2(\alpha_1, 1). \end{cases} \quad (2.4.90)$$

$$\begin{cases} S_i = E \Leftrightarrow \text{sign}F_2(z_j, 1) = \text{sign}F_2(\alpha_i, 1) = -\text{sign}F_2(\alpha_{i+1}, 1) \text{ if } \alpha_i > 0, \\ S_i = E \Leftrightarrow \text{sign}F_2(z_j, 1) = \text{sign}F_2(\alpha_{i+1}, 1) = -\text{sign}F_2(\alpha_i, 1) \text{ if } \alpha_i \leq 0, \end{cases} \quad (2.4.91)$$

for $i \neq K$, where $z_1 > \dots > z_K$ are the zero points of $F_1(z, 1)$, and z_j satisfies that $z_j < \alpha_i < z_{j+1}$, and

$$S_K = E \Leftrightarrow -\text{sign}F_2(z_j, 1) = \text{sign}F_2(\alpha_K, 1) = (-1)^K \text{sign}F_2(\alpha_1, 1). \quad (2.4.92)$$

The hyperbolic region H can be determined by ruling out the parabolic and elliptic cases by (2.4.90)–(2.4.92):

$$S_i = H \Leftrightarrow S_i \neq P \text{ and } S_i \neq E. \quad (2.4.93)$$

The properties (2.4.90)–(2.4.93) completely establish the topological structure of a two-dimensional nondegenerate k -homogeneous vector field F , which is crucial to study the transition of (2.1.1) from a k -th order nondegenerate singular points.

From Claim A and (2.4.90)–(2.4.93) we obtain the following theorem.

Theorem 2.4.15 *Let F be given by (2.4.64). Then we have the following assertions:*

- (1) *F has only one parabolic region with $x = 0$ being a focus or a center if and only if the polynomial $h(z)$ given by (2.4.90) has no real zero points.*
- (2) *F has only one parabolic region with $x = 0$ being a node if and only if for all real zero points α_i ($1 \leq i \leq K$) of $h(z)$,*

$$\text{sign}F_2(\alpha_i, 1) = \text{sign}F_2(\alpha_1, 1) \quad \forall 1 \leq i \leq K,$$

and $x = 0$ is stable as $F_2(\alpha_1, 1) < 0$, unstable as $F_2(\alpha_1, 1) > 0$.

- (3) *F has exactly $2(k+1)$ hyperbolic regions and no parabolic and elliptic regions if and only if $\text{ind}(F, 0) = -k$.*
- (4) *F has exactly $2(k-1)$ elliptic regions and no parabolic and hyperbolic regions if and only if $\text{ind}(F, 0) = k$ and $h(z)$ has only $(k-1)$ real solutions.*

2.4.8 Transition from k -th Order Nondegenerate Singularities

We consider the transition of (2.1.1) from the k -th order nondegenerate singular point $u = 0$. Let the nonlinear term $G(u, \lambda)$ of (2.1.1) has the Taylor expansion:

$$G(u, \lambda) = G_k(u, \lambda) + o(\|u\|_{X_1}^k) \quad \text{for some } k \geq 2, \quad (2.4.94)$$

where G_k is a k -multilinear operator.

We set

$$\begin{pmatrix} F_1 \\ F_2 \end{pmatrix} = \begin{pmatrix} (G_k(x_1 e_1 + x_2 e_2, \lambda), e_1^*) \\ (G_k(x_1 e_1 + x_2 e_2, \lambda), e_2^*) \end{pmatrix}, \quad (2.4.95)$$

where e_1, e_2 and e_1^*, e_2^* are as in (2.4.13), and

$$(G_k(x_1 e_1 + x_2 e_2, \lambda), e_l^*) = \sum_{i+j=k} a_{ij}^l x_1^i x_2^j \quad \text{for } l = 1, 2.$$

The polynomials $F_i(z, 1)$ induced by (2.4.95) at $\lambda = \lambda_0$ are as in (2.4.65), and $h(z) = h(z, \lambda_0)$ is as in (2.4.88).

Here we always assume that the conditions (2.1.4) and (2.1.5) with $m = r = 2$ hold true.

Theorem 2.4.16 *Let $F(x, \lambda)$ be given by (2.4.95), which is k -th order nondegenerate. Then we have the following assertions:*

- (1) *If $h(z)$ has no real zero points and $a_{k0}^2 \neq 0$, then (2.1.1) bifurcates from $(u, \lambda) = (0, \lambda_0)$ to a periodic orbit. For the number*

$$\alpha = \int_0^{2\pi} \frac{\cos \theta F_1(\cos \theta, \sin \theta) + \sin \theta F_2(\cos \theta, \sin \theta)}{\cos \theta F_2(\cos \theta, \sin \theta) - \sin \theta F_1(\cos \theta, \sin \theta)} d\theta, \quad (2.4.96)$$

if $\alpha < 0$ the transition from $(0, \lambda_0)$ is an attractor bifurcation, and if $\alpha \geq 0$, the transition is jump.

- (2) *If $h(z)$ has real zero points α_i ($1 \leq i \leq k$), and*

$$\text{sign}F_2(\alpha_i, 1) = \text{sign}F_2(\alpha_1, 1) \quad \forall 1 \leq i \leq k,$$

then, as $F_2(\alpha_1, 1) < 0$, the transition of (2.1.1) from $(0, \lambda_0)$ is an S^1 -attractor bifurcation, and as $F_2(\alpha_1, 1) > 0$, the transition is jump. Moreover, there are at least four singular points bifurcated on $\lambda > \lambda_0$ for $F_2(\alpha_1, 1) < 0$ and on $\lambda < \lambda_0$ for $F_2(\alpha_1, 1) > 0$.

- (3) *If $\text{ind}(F_{\lambda_0}, 0) \leq 0$, then the transition is either jump or mixed, and there are singular points bifurcated on both sides of λ_0 . In particular, if $\text{ind}(F_{\lambda_0}, 0) = -k$, then the transition is jump and there are exactly $(k+1)$ singular points bifurcated on each side of λ_0 .*
- (4) *As $\text{ind}(F_{\lambda_0}, 0) = k$, and $h(z)$ has only $(k-1)$ real solutions, then the transition is continuous. In particular, if $\beta_1(\lambda) = \beta_2(\lambda)$ in (2.1.4), the transition is an*

attractor bifurcation and the bifurcated attractor \mathcal{A}_λ consists of $(k - 1)$ singular points which are minimal attractors.

- (5) As $\text{ind}(F_{\lambda_0}, 0) \geq 2$, the transition is either mixed or continuous, and if $\beta_1(\lambda) = \beta_2(\lambda)$, then (2.1.1) bifurcates in the stable domain to an attractor \mathcal{A}_λ with $\dim \mathcal{A}_\lambda = 0$ or 1, and \mathcal{A}_λ attracts at least n sectorial domains $D(\theta_i)$ with $\sum_{i=1}^n \theta_i \leq 2\pi$, $n = \text{ind}(F, 0) - 1$.

Proof. We only need to consider the reduced equation of (2.1.1) as follows

$$\begin{aligned}\frac{dx_1}{dt} &= \beta_1(\lambda)x_1 + F_1(x, \lambda) + o(|x|^k), \\ \frac{dx_2}{dt} &= \beta_2(\lambda)x_2 + F_2(x, \lambda) + o(|x|^k),\end{aligned}\tag{2.4.97}$$

where $F(x, \lambda) = (F_1(x, \lambda), F_2(x, \lambda))$ is given by (2.4.95).

PROOF OF ASSERTION (1). By Theorem 2.4.15, when $h(z)$ has no real zero points, which implies that at $\lambda = \lambda_0$, $x = 0$ is either a focus of (2.4.97) or there infinite periodic orbits of (2.4.97) in a neighborhood. In the second case, it means a bifurcation of periodic orbits. Hence, we only consider the first case.

When $x = 0$ is a stable focus, by the attractor bifurcation theorem (Theorem 2.2.2), the system (2.4.97) bifurcates from $(0, \lambda_0)$ to an S^1 -attractor. Because $h(z) = zF_2(z, 1) - F_1(z, 1)$ has no real zero points, (2.4.97) has no singular points bifurcated from $(0, \lambda_0)$. Hence, by Theorem 2.6, the bifurcated S^1 -attractor is a periodic orbit. Likewise, when $x = 0$ is an unstable focus, $x = 0$ is a stable focus of $-(F(x) + o(|x|^k))$. Thus, as above, we can prove that the following vector field

$$\begin{pmatrix} \beta_1(\lambda) & 0 \\ 0 & \beta_2(\lambda) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - F(x, \lambda) - o(|x|^k)\tag{2.4.98}$$

bifurcates from $(0, \lambda_0)$ on $\lambda > \lambda_0$ to a periodic orbit, which implies that (2.4.97) bifurcates on $\lambda < \lambda_0$ to a periodic orbit.

It is clear that the transition type of (2.4.96), continuous or jump, depends on if $x = 0$ is a stable or unstable focus of $F(x, \lambda)$. Take the polar coordinate $x_1 = r \cos \theta$, $x_2 = r \sin \theta$, then the equation $dx/dt = F(x, \lambda_0)$ becomes

$$\frac{dr}{rd\theta} = \frac{\cos \theta F_1(\cos \theta, \sin \theta) + \sin \theta F_2(\cos \theta, \sin \theta)}{\cos \theta F_2(\cos \theta, \sin \theta) - \sin \theta F_1(\cos \theta, \sin \theta)}.\tag{2.4.99}$$

In addition, by assumption, $h(z)$ has no zeroes:

$$\cos \theta F_2(\cos \theta, \sin \theta) - \sin \theta F_1(\cos \theta, \sin \theta) \neq 0 \quad \forall 0 \leq \theta \leq 2\pi.$$

Hence, it follows from (2.4.99) that

$$\ln \frac{r(2\pi)}{r(0)} = \alpha,\tag{2.4.100}$$

where α is given by (2.4.96). Thus, we infer from (2.4.100) that if $\alpha < 0$, then $x = 0$ is a stable focus, and if $\alpha > 0$, then $x = 0$ is unstable. Assertion (1) is proved.

PROOF OF ASSERTION (2). It follows from Theorems 2.4.15 and 2.2.2 that as $F_2(\alpha_1, 1) < 0$ the system (2.4.97) has an S^1 -attractor bifurcation, and as $F_2(\alpha_1, 1) > 0$ (2.4.97) has a jump transition.

We consider the case where $x = 0$ is a stable node, and for the case where $x = 0$ is unstable the discussion is the same. It is known that the flows of $F(x, \lambda)$ are inward to $x = 0$, and the flows of $(\beta_1(\lambda)x_1, \beta_2(\lambda)x_2)$ are radially outward from $x = 0$. Hence, for any $\lambda - \lambda_0 > 0$, there a curve $\gamma(\lambda)$ passing through $x = 0$ such that $F(x, \lambda)$ is parallel to $\beta x = (\beta_1(\lambda)x_1, \beta_2(\lambda)x_2)$ on $\gamma(\lambda)$, i.e.,

$$\begin{aligned} \beta(\lambda)x &= \rho(\lambda, x)F(x, \lambda) && \forall x \in \gamma(\lambda), \\ \rho(\lambda, x) &< 0 && \forall x \in \gamma(\lambda) \text{ and } x \neq 0. \end{aligned} \quad (2.4.101)$$

We note that $F(\alpha x, \lambda) = \alpha^k F(x, \lambda)$ ($k \geq 3$). Hence, for any given $\lambda - \lambda_0 > 0$ sufficiently small, there is $r(\lambda) > 0$ such that

$$\begin{aligned} \beta(\lambda)|x| &> |F(x, \lambda)| && \text{if } 0 < |x| < r(\lambda), \\ \beta(\lambda)|x| &< |F(x, \lambda)| && \text{if } |x| > r(\lambda), \\ \lim_{\lambda \rightarrow \lambda_0^+} r(\lambda) &= 0, \end{aligned}$$

yielding, by (2.4.101), that

$$\beta(\lambda)x_\lambda + F(x_\lambda, \lambda) = 0, \quad x_\lambda \in \gamma(\lambda), \quad |x_\lambda| = r(\lambda). \quad (2.4.102)$$

Since $F(x, \lambda) + o(|x|^k)$ is a perturbation of $F(x, \lambda)$ near $x = 0$, the above procedure is also valid for any higher-order perturbation of $F(x, \lambda)$. Hence we have

$$\text{ind}(\beta id + F, x_\lambda) \neq 0.$$

Noting that $\beta id + F$ is odd ($k \geq 3$ is an odd number), hence $\pm x_\lambda$ are the solutions of (2.4.102), and

$$\text{ind}(\beta id + F, x_\lambda) = \text{ind}(\beta id + F, -x_\lambda) \neq 0. \quad (2.4.103)$$

Let $\pm x_\lambda^k$ ($1 \leq k \leq n$) be all solutions of (2.4.102). Then we have

$$\sum_{k=1}^n \text{ind}(\beta id + F, x_\lambda^k) = \text{ind}(F_{\lambda_0}, 0) = 1. \quad (2.4.104)$$

On the other hand, we know

$$\text{ind}(\beta id + F, 0) = 1. \quad (2.4.105)$$

Then, we infer from (2.4.103) to (2.4.105) that there are at least four bifurcated solutions $\pm x_\lambda^1, \pm x_\lambda^2$ of (2.4.102) such that

$$\text{ind}(\beta id + F, \pm x_\lambda^1) = -\text{ind}(\beta id + F, \pm x_\lambda^2) \neq 0. \quad (2.4.106)$$

By the perturbation invariance of index, Assertion (2) follows from (2.4.105).

The proofs of Assertions (3) and (4) are respectively similar to that of Theorems 2.4.4 and 2.4.6.

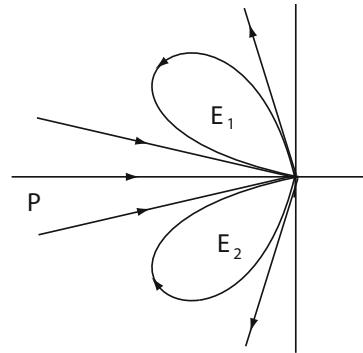
PROOF OF ASSERTION (5). By the Poincaré formula (2.1.9), when the index $\text{ind}(F_{\lambda_0}, 0) \geq 2$, there are at least n ($n=\text{ind}(F_{\lambda_0}, 0) - 1$) pairs of elliptic regions such as E_1, E_2 with a parabolic region P (maybe $P = \emptyset$), as shown in Fig. 2.30, such that

$$\lim_{t \rightarrow \infty} S(t)x = 0 \quad \forall x \in E_1 \cup E_2 \cup P,$$

where $S(t)$ is the semigroup generated by $F(x, \lambda_0)$. Then, by using the same manner as used in Theorem 2.4.6 one can verify Assertion (5).

The proof of the theorem is complete. \square

Fig. 2.30 Illustration of Proof of Assertion 5 for Theorem 2.4.16



2.4.9 Bifurcation to Periodic Orbits

Consider the case where $m = 2, r = 1$ and $k=\text{odd} \geq 3$ in (2.1.4) and (2.1.5), where $r = 1$ is the geometric multiplicity of $\beta_1(\lambda)$ at $\lambda = \lambda_0$. In this case, the two eigenvectors e_1 and e_2 of L_λ at $\lambda = \lambda_0$ satisfy that

$$L_{\lambda_0} e_1 = 0, \quad L_{\lambda_0} e_2 = a e_1, \quad L_{\lambda_0}^* e_2^* = 0, \quad L_{\lambda_0}^* e_1^* = a e_2^*, \quad a \neq 0,$$

$$(e_i, e_j^*) \begin{cases} > 0 & \text{for } i = j, \\ = 0 & \text{for } i \neq j. \end{cases}$$

Let $\alpha \in \mathbb{R}$ be a number defined by

$$\alpha = (G_k(e_1, \lambda_0), e_2^*), \quad (2.4.107)$$

where G_k is as in (2.4.94). Then we have the following transition theorem with the bifurcation of periodic orbits from the real eigenvalues.

Theorem 2.4.17 *Assume the conditions (2.1.4) and (2.1.5) with $m = 2$, $r = 1$, and $k=\text{odd} \geq 3$. Let α be given by (2.4.107). If $\alpha \cdot a < 0$, then the transition of (2.1.1) from $(u, \lambda) = (0, \lambda_0)$ is either continuous (attractor bifurcation) or jump, and (2.1.1) bifurcates from $(0, \lambda_0)$ to a periodic orbit.*

Proof. We proceed in the following four steps.

STEP 1. By the center manifold reduction, it suffices to consider the following reduced equations of (2.1.1):

$$\begin{aligned}\frac{dx_1}{dt} &= \beta_1(\lambda)x_1 + ax_2 + (G(x + h(x, \lambda), \lambda), e_1^*(\lambda)), \\ \frac{dx_2}{dt} &= \beta_2(\lambda)x_2 + (G(x + h(x, \lambda), \lambda), e_2^*(\lambda)),\end{aligned}\tag{2.4.108}$$

where $x = x_1e_1(\lambda) + x_2e_2(\lambda)$, $h(x, \lambda)$ is the center manifold function, and

$$\begin{aligned}L_\lambda e_1(\lambda) &= \beta_1(\lambda)e_1(\lambda), & L_\lambda e_2(\lambda) &= \beta_2(\lambda)e_2(\lambda) + av_1(\lambda), \\ L_\lambda^* e_2^*(\lambda) &= \beta_2(\lambda)e_2^*(\lambda), & L_\lambda^* e_1^*(\lambda) &= \beta_1(\lambda)e_1^*(\lambda) + ae_2^*(\lambda), \\ (e_i(\lambda), e_j^*(\lambda)) &= \delta_{ij}, & a &\neq 0,\end{aligned}$$

where δ_{ij} is the Kronecker symbol.

By (2.4.94) and (2.4.107), equations (2.4.108) at $\lambda = \lambda_0$ reads

$$\frac{dx}{dt} = F(x) = \begin{pmatrix} F_1(x) \\ F_2(x) \end{pmatrix},\tag{2.4.109}$$

where

$$\begin{aligned}F_1(x) &= ax_2 + o(|x_1|^k, |x_2|^k), \\ F_2(x) &= \alpha x_1^k + o(|x_1|^{k+1}, |x_2|^k, |x_2|^{k-1}|x_1|, \dots, |x_2||x_1|^{k-1}).\end{aligned}$$

Since $\alpha \cdot a < 0$ and $k=\text{odd} \geq 3$, we have

$$\text{ind}(F, 0) = 1.\tag{2.4.110}$$

STEP 2. We now prove that the number of elliptic regions of F at $x = 0$ is zero, i.e., $e = 0$. Assume otherwise; then there exists an orbit γ of (2.4.109) connected to $x = 0$, i.e.

$$\lim_{t \rightarrow \infty} S(t)x = 0 \quad \forall x \in \gamma,$$

where $S(t)$ is the operator semigroup generated by (2.4.109). Let γ can be expressed near $x = 0$ as

$$x_2 = f(x_1) \quad \text{for } (x_1, x_2) \in \gamma.$$

It follows from (2.4.109) that for any $(x_1, x_2) \in \gamma$,

$$\frac{dx_2}{dx_1} = \frac{\alpha x_1^k + o(|x_1|^{k+1}, |f|^k, |x_1||f|^{k-1}, \dots, |x_1|^{k-1}|f|)}{af(x) + o(|x_1|^k, |f|^k)}.$$

Thus, we obtain

$$af(x_1)f'(x_1) + o(|x_1|^k, |f|^k)f' = \alpha x_1^k + o(|x_1|^{k+1}, |x_1|^{k-i}|f|^i), \quad (2.4.111)$$

which implies that

$$f(x) = \beta x^m + o(|x|^m) \quad \text{for } 2 \leq m = \frac{k+1}{2} < k, \quad \beta \neq 0. \quad (2.4.112)$$

Therefore, from (2.4.111) and (2.4.112) we get

$$\alpha = am\beta^2,$$

where $a \cdot \alpha < 0, m \geq 2$ and $\beta \neq 0$. It is a contradiction. Hence, $e = 0$.

STEP 3. By (2.4.110) and the Poincaré formula (2.1.9), there are no hyperbolic regions near $x = 0$. Hence, (2.4.109) has only one parabolic region P , and by the proof in Step 2, the orbits in P are circular about $x = 0$. Thus, $x = 0$ is either (a) a stable focus, or (b) an unstable focus, or (c) a singular point having infinite periodic orbits in its neighborhood.

The case (c) implies a bifurcation to periodic orbits for (2.4.108).

For the case (a), $x = 0$ is an asymptotically stable singular point of (2.4.109). Then by Theorem 2.2.2, the equation (2.4.108) bifurcates from $(x, \lambda) = (0, \lambda_0)$ to an S^1 -attractor Σ_λ on $\lambda > \lambda_0$.

For the case (b), $x = 0$ is an asymptotically stable singular point of $-F(x)$, therefore the vector field

$$-\begin{pmatrix} \beta_1(\lambda) & 0 \\ 0 & \beta_2(\lambda) \end{pmatrix} x - F(x, \lambda)$$

bifurcates from $(0, \lambda_0)$ to an S^1 -attractor Σ_λ on $\lambda > \lambda_0$, which implies that (2.4.108) bifurcates from $(0, \lambda_0)$ on $\lambda < \lambda_0$ to an S^1 -repeller Σ_1 , which is an invariant set.

STEP 4. Now, we need to prove that the S^1 -invariant set Σ_λ contains no singular points. Consider the following equations

$$\begin{aligned} \beta_1(\lambda)x_1 + ax_2 + o(|x_1|^k, |x_2|^k) &= 0, \\ \beta_2(\lambda)x_2 + \alpha x_1^k + o(|x_1|^{k+1}, |x_2|^{k-i}|x_i|^i) &= 0. \end{aligned}$$

Hence

$$\begin{aligned} x_2 &= -\beta_1(\lambda)a^{-1}x_1 + o(|x_1|^k, |\beta_1|^k), \\ a\alpha x_1^{k-1} - \beta_1(\lambda)\beta_2(\lambda) + o(|x_1|^k, |x_1|^k|\beta_1|) &= 0. \end{aligned} \quad (2.4.113)$$

By $a\alpha < 0$ and $\beta_1(\lambda)\beta_2(\lambda) > 0$, (2.4.113) has no solutions near $(x, \lambda) = (0, \lambda_0)$. Hence, there are no singular points in Σ_λ , which implies that Σ_λ must contain a periodic orbit.

Finally, as for Theorem 2.3.6, we can prove that the transition of (2.4.108) is either continuous, which is an attractor bifurcation, or jump. Thus the proof is complete. \square

Remark 2.4.18 Theorem 2.4.17 can be considered as a complement of Theorem 4.3 in Ma and Wang (2005b) for $m = 2, r = 1, \alpha < 0$ ($a = 1$), where there are no bifurcation of singular points.

2.4.10 Application to Parabolic Systems

Consider

$$\begin{aligned} \frac{\partial u_1}{\partial t} &= \Delta u_1 + \lambda\alpha_1 u_1 + au_2 + g_1(u, \lambda), \\ \frac{\partial u_2}{\partial t} &= \Delta u_2 + \lambda\alpha_2 u_2 + g_2(u, \lambda), \\ u|_{\partial\Omega} &= 0, \end{aligned} \tag{2.4.114}$$

where $u = (u_1, u_2)^t$, the parameters α_1, α_2 and a are constants, $\Omega \subset \mathbb{R}^n$ ($n \geq 1$) is bounded domain, and

$$g_l(u, \lambda) = \sum_{i+j=k} a_{ij}^l u_1^i u_2^j + o(|u|^k) \quad \text{for } l = 1, 2, \quad k \geq 2. \tag{2.4.115}$$

Let $\lambda_i > 0$ and $h_i(x)$ be the i -th eigenvalue and eigenfunction of $-\Delta$ on Ω with the Dirichlet boundary condition. Let

$$X = L^p(\Omega, \mathbb{R}^2), \quad X_1 = W^{2,p}(\Omega, \mathbb{R}^2) \cap W_0^{1,p}(\Omega, \mathbb{R}^2), \quad \text{for } p \geq n,$$

and define the operators $L_\lambda = -A + B_\lambda : X_1 \rightarrow X$ and $G : X_1 \rightarrow X$ by

$$-Au = \begin{pmatrix} \Delta u_1 \\ \Delta u_2 \end{pmatrix}, \quad B_\lambda u = \begin{pmatrix} \lambda\alpha_1 & a \\ 0 & \lambda\alpha_2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad G(u, \lambda) = \begin{pmatrix} g_1(u, \lambda) \\ g_2(u, \lambda) \end{pmatrix}.$$

Then (2.4.114) can be written in the form (2.1.1), and the conditions (2.1.2) and (2.1.3) are satisfied.

CASE 1. $\alpha_2 > \alpha_1 > 0$. Thus, all eigenvalues of L_λ are $\alpha_1\lambda - \lambda_i$ and $\alpha_2\lambda - \lambda_i$ ($1 \leq i < \infty$), which enjoy the properties (2.1.3) and (2.1.4) with $\lambda_0 = \alpha_1^{-1}\lambda_1$. The first eigenvectors e_1 and e_1^* of L_λ and L_λ^* are given by

$$e_1 = \begin{pmatrix} h_1 \\ 0 \end{pmatrix}, \quad e_1^* = \begin{pmatrix} h_1 \\ -a\alpha_1 h_1 / (\alpha_2 - \alpha_1)\lambda_1 \end{pmatrix}.$$

Let the k -homogeneous term in (2.4.115) be given by

$$G_k(u, \lambda) = \begin{pmatrix} \sum_{i+j=k} a_{ij}^1 u_1^i u_2^j \\ \sum_{i+j=k} a_{ij}^2 u_1^i u_2^j \end{pmatrix},$$

and set

$$\alpha = (G_k(e_1, \lambda_0), e_1^*) = \left[a_{k0}^1 - \frac{\alpha_1 a}{(\alpha_2 - \alpha_1)\lambda_1} a_{k0}^2 \right] \int_{\Omega} h_1^{k+1} dx.$$

Let $\alpha \neq 0$, i.e.

$$a_{k0}^1 - \frac{\alpha_1 a}{(\alpha_2 - \alpha_1)\lambda_1} a_{k0}^2 \neq 0.$$

As k =even, by Theorem 2.3.2, the equations (2.4.114) has a mixed transition from $(u, \lambda) = (0, \alpha_1^{-1}\lambda_1)$, and there is a neighborhood $U \subset X$ of $u = 0$, which can be decomposed into two open sets

$$U_1 = \left\{ (u_1, u_2) \in U \mid u_1 = zh_1 + v, z \cdot \alpha > 0, \int h_1 v dx = 0 \right\},$$

$$U_2 = \left\{ (u_1, u_2) \in U \mid u_1 = zh_1 + v, z \cdot \alpha < 0, \int h_1 v dx = 0 \right\},$$

such that U_1 is an unstable domain, U_2 a stable domain of (2.4.114), and (2.4.114) bifurcate in U_2 to an attractor which is a singular point

$$v_{\lambda} = -\frac{\beta_1(\lambda)}{\alpha} \begin{pmatrix} h_1(x) \\ 0 \end{pmatrix} + o(|\beta_1|) \in U_2 \quad \text{for } \lambda > \lambda_0,$$

and $\{v_{\lambda}\}$ attracts U_2 in the X -norm.

As k =odd, by Theorem 2.3.1, the equations (2.4.114) has a jump transition from $(0, \alpha_1^{-1}\lambda_1)$ for $\alpha > 0$, and have an attractor bifurcation for $\alpha < 0$.

CASE 2: $\alpha_1 = \alpha_2 > 0$ AND $a = 0$. Thus we have $\beta_1(\lambda) = \beta_2(\lambda) = \alpha_1\lambda - \lambda_1$, and the corresponding eigenvectors are given by

$$e_1 = e_1^* = \begin{pmatrix} h_1(x) \\ 0 \end{pmatrix}, \quad e_2 = e_2^* = \begin{pmatrix} 0 \\ h_1(x) \end{pmatrix}. \quad (2.4.116)$$

The k -homogeneous vector field induced by (2.4.114) reads

$$F(x, \lambda) = C \begin{pmatrix} \sum_{i+j=k} a_{ij}^1 u_1^i u_2^j \\ \sum_{i+j=k} a_{ij}^2 u_1^i u_2^j \end{pmatrix},$$

where $C = \int_{\Omega} h_1^{k+1}(x) dx > 0$.

If $F(x, \lambda)$ is k th-order nondegenerate, then the results in this section are valid for (2.4.114).

CASE 3. $\alpha_1 = \alpha_2 > 0, a \neq 0$, AND k =ODD ≥ 3 . In this case, $m = 2, r = 1$, and the first eigenvalues are $\beta_1(\lambda) = \beta_2(\lambda) = \alpha_1\lambda - \lambda_1$, and the corresponding eigenvectors

are given by (2.4.116), which satisfy

$$\begin{aligned} L_\lambda e_1 &= \beta_1(\lambda)e_1, & L_\lambda e_2 &= \beta_1(\lambda)e_2 + ae_1, \\ L_\lambda^* e_2^* &= \beta_1(\lambda)e_2^*, & L_\lambda^* e_1^* &= \beta_1(\lambda)e_1^* + ae_2^*. \end{aligned}$$

Let

$$\alpha = (G_k(e_1, \lambda_0), e_2^*) = a_{k0}^2 \int_{\Omega} h_1^{k+1}(x) dx.$$

By Theorem 2.4.17, if $a \cdot a_{k0}^2 < 0$, then (2.4.114) bifurcate from $(0, \alpha_1^{-1}\lambda_1)$ to a periodic orbit, and the transition is either jump or continuous.

2.5 Singular Separation

2.5.1 General Principle

In this section, we study an important problem associated with the jump and mixed transitions of (2.1.1), which we call the singular separation.

Definition 2.5.1

1. An invariant set Σ of (2.1.1) is called a singular element if Σ is either a singular point or a periodic orbit.
2. Let $\Sigma_1 \subset X$ be a singular element of (2.1.1) and $U \subset X$ a neighborhood of Σ_1 . We say that (2.1.1) has a singular separation of Σ at $\lambda = \lambda_1$ if
 - a. Equation (2.1.1) has no singular elements in U as $\lambda < \lambda_1$ (or $\lambda > \lambda_1$), and generates a singular element $\Sigma_1 \subset U$ at $\lambda = \lambda_1$, and
 - b. there are branches of singular elements Σ_λ , which are separated from Σ_1 for $\lambda > \lambda_1$ (or $\lambda < \lambda_1$), i.e.,

$$\lim_{\lambda \rightarrow \lambda_1} \max_{x \in \Sigma_\lambda} \text{dist}(x, \Sigma_1) = 0.$$

A special case of singular separation is the saddle-node bifurcation:

Definition 2.5.2 Let $u_1 \in X$ be a singular point of (2.1.1) at $\lambda = \lambda_1$ with $u_1 \neq 0$. We say that (2.1.1) has a saddle-node bifurcation at (u_1, λ_1) if

1. the index of $L_\lambda + G$ at (u_1, λ_1) is zero, i.e., $\text{ind}(-(L_{\lambda_1} + G), u_1) = 0$,
2. there are two branches $\Gamma_1(\lambda)$ and $\Gamma_2(\lambda)$ of singular points of (2.1.1), which are separated from u_1 for $\lambda > \lambda_1$ (or $\lambda < \lambda_1$), i.e., for any $u_\lambda \in \Gamma_i(\lambda)$ ($i = 1, 2$) we have

$$u_\lambda \rightarrow u_1 \text{ in } X \quad \text{as } \lambda \rightarrow \lambda_1,$$

and

3. the indices of $u_\lambda^i \in \Gamma_i(\lambda)$ are as follows

$$\text{ind}(-(L_\lambda + G), u_\lambda) = \begin{cases} 1 & \text{if } u_\lambda \in \Gamma_2(\lambda), \\ -1 & \text{if } u_\lambda \in \Gamma_1(\lambda). \end{cases}$$

Intuitively, the saddle-node bifurcation is schematically shown as in Fig. 2.31, where the singular points in $\Gamma_1(\lambda)$ are saddle points and in $\Gamma_2(\lambda)$ are nodes, and the singular separation of periodic orbits is as shown in Fig. 2.32.

Fig. 2.31 Saddle-node bifurcation

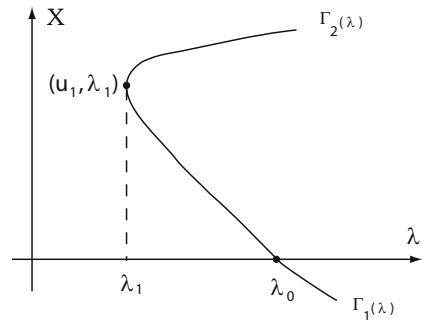
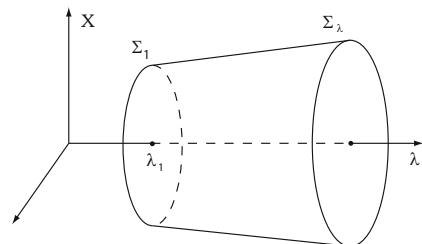


Fig. 2.32 Singular separation of periodic orbits



For the singular separation we can give a general principle as follows, which provides a basis for singular separation theory.

Theorem 2.5.3 Let the conditions (2.1.4) and (2.1.5) hold true, and

$$\operatorname{Re}\beta_j(\lambda) < 0 \quad \forall \lambda < \lambda_0, 1 \leq j < \infty.$$

Then we have the following assertions:

1. If (2.1.1) bifurcates from $(u, \lambda) = (0, \lambda_0)$ to a branch Σ_λ of singular elements on $\lambda < \lambda_0$ which is bounded in $X \times (-\infty, \lambda_0)$, then (2.1.1) has a singular separation of singular elements at some $(\Sigma_0, \lambda_1) \subset X \times (-\infty, \lambda_0)$.
2. If the branch Σ_λ consists of singular points which has index -1 , i.e.,

$$\text{ind}(-(L_\lambda + G), u_\lambda) = -1 \quad \forall u_\lambda \in E_\lambda, \quad \lambda < \lambda_0,$$

then the singular separation is a saddle-node bifurcation from some $(u_1, \lambda_1) \in X \times (-\infty, \lambda_0)$.

Proof. By assumption, for any $\lambda < \lambda_0$, the eigenvalues $\beta_j(\lambda)$ ($1 \leq j < \infty$) of L_λ have negative real parts:

$$\operatorname{Re}\beta_j(\lambda) < 0 \quad \forall \lambda < \lambda_0 \quad 1 \leq j < \infty.$$

Hence, no singular elements are bifurcated from $(u, \lambda) = (0, \lambda)$ for all $\lambda < \lambda_0$. This implies that the branch Σ_λ of singular elements bifurcated from $(0, \lambda_0)$ do not converge to $u = 0$. On the other hand, by assumption Σ_λ is bounded in $X \times (-\infty, \lambda_0)$, it means that Σ_λ separates from some singular element Σ_0 at $\lambda_1 < \lambda_0$. Assertion (1) is proved. Assertion (2) follows from Assertion (1) and Definition 2.5.2. The proof is complete. \square

2.5.2 Saddle-Node Bifurcation

For a nonlinear dissipative system (2.1.1), the singular separation is closely related to the jump and mixed transition, which may lead to chaos. In addition, the singular separation can provide important information for the global topological structure of the jump and mixed transitions.

We consider the equation (2.1.1) defined on two Hilbert spaces $X = H$ and $X_1 = H_1$. Let $L_\lambda = -A + \lambda B$. For L_λ and $G(\cdot, \lambda) : H_1 \rightarrow H$, we assume that $A : H_1 \rightarrow H$ is symmetric, and

$$(Au, u) \geq c\|u\|_{H_{1/2}}^2, \tag{2.5.1}$$

$$(Bu, u) \leq c\|u\|^2, \tag{2.5.2}$$

$$(Gu, u) \leq -c_1\|u\|^p + c_2\|u\|^2, \tag{2.5.3}$$

where $p > 2$, $c, c_1, c_2 > 0$ are constants.

Theorem 2.5.4 Assume the conditions (2.1.3), (2.1.4), and (2.5.1)–(2.5.3), then (2.1.1) has a transition at $(u, \lambda) = (0, \lambda_0)$, and the following assertions hold true:

1. If $u = 0$ is an even-order nondegenerate singular point of $L_\lambda + G$ at $\lambda = \lambda_0$, then (2.1.1) has a singular separation of singular points at some $(u_1, \lambda_1) \in H \times (-\infty, \lambda_0)$.
2. If $m = 1$ and G satisfies (2.3.1) with $\alpha > 0$ if $k=\text{odd}$ and $\alpha \neq 0$ if $k=\text{even}$, then (2.1.1) has a saddle-node bifurcation at some singular point (u_1, λ_1) with $\lambda_1 < \lambda_0$.

Proof. By Theorem 2.1.3, under the conditions (2.1.3) and (2.1.4), the equation (2.1.1) has a transition at $(0, \lambda_0)$. By the conditions (2.5.1)–(2.5.3) we can deduce that for any $\lambda \in \mathbb{R}^1$ there are constants $\alpha_1, \alpha_2 > 0$ such that

$$(L_\lambda u + G(u, \lambda), u) \leq -\alpha_1 \|u\|_{H_{1/2}}^2 + \alpha_2.$$

Then, it is clear that for any $\lambda \in \mathbb{R}^1$ the equation (2.1.1) possesses a global attractor in H .

If $u = 0$ is even-order nondegenerate for $L_\lambda + G$ at $\lambda = \lambda_0$, then by Theorem 4.1 in Ma and Wang (2005b), (2.1.1) bifurcates from $(0, \lambda_0)$ a branch Γ_λ of singular points on $\lambda < \lambda_0$. Since (2.1.1) has a global attractor, for $u_\lambda \in \Gamma_\lambda$ we have

$$\|u_\lambda\| \leq c_\lambda, \quad c_\lambda > 0 \text{ a constant.} \quad (2.5.4)$$

In addition, it follows from (2.5.1) to (2.5.3) that

$$\begin{aligned} (L_\lambda u + G(u, \lambda), u) &\leq -c \|u\|_{H_{1/2}}^2 - c_1 \|u\|^p + (c_2 + \lambda c) \|u\|^2, \\ &\leq -\alpha \|u\|^2, \quad \forall \lambda \leq -c_2 c^{-1}, \quad u \in H, \end{aligned}$$

which implies that as $\lambda \leq -c_2 c^{-1}$, $u = 0$ is a global asymptotically stable singular point of (2.1.1). Hence, it follows from (2.5.4) that Γ_λ is bounded in $H \times (-\infty, \lambda_0)$. Thus, by Theorem 2.5.3, we derive Assertion (1). Likewise, Assertion (2) follows from Theorems 2.3.1, 2.3.2, and 2.5.3.

The proof is complete. \square

The following theorem is motivated for applications in the hydrodynamic stability and transition. For the operators $L_\lambda = -A + B_\lambda$ and G , we assume that

- (A₁) $A : H_1 \rightarrow H$ is positive definite (not necessary symmetric), i.e., the operator A satisfies (2.5.1), and $B_\lambda = 0$ at $\lambda = 0$.
- (A₂) $G : H_1 \rightarrow H$ is bilinear which satisfies

$$(G(u, \lambda), u) = 0 \quad \forall u \in H_1.$$

- (A₃) The conditions (2.1.3) and (2.1.4) with $m = 1$ and $\lambda_0 > 0$ hold true, and for the first eigenvectors e_1 and e_1^* of L_λ and L_λ^* at $\lambda = \lambda_0$, we have

$$(G(e_1, \lambda_0), e_1^*) \neq 0.$$

- (A₄) For any $\lambda \in \mathbb{R}^1$, (2.1.1) possesses a global attractor in H .

Theorem 2.5.5 *Assume the conditions (A₁)–(A₄). Then, (2.1.1) has a saddle-node bifurcation at some (u_1, λ_1) with $0 < \lambda_1 < \lambda_0$. Moreover, the two branches $\Gamma_1(\lambda)$ and $\Gamma_2(\lambda)$ separated from (u_1, λ_1) , as shown in Fig. 2.31, have the following properties:*

1. For $u_\lambda \in \Gamma_1(\lambda)$, $u_\lambda \rightarrow 0$ as $\lambda \rightarrow \lambda_0$, and when $\lambda_1 < \lambda < \lambda_0$, u_λ is a saddle with the Morse index one, and when $\lambda > \lambda_0$, u_λ is an attractor which attracts a sectorial domain $D(\theta)$ with angle $\theta = \pi$.
2. For $u_\lambda \in \Gamma_2(\lambda)$, $u_{\lambda_0} \neq 0$, and there is a $\lambda^* > \lambda_1$ such that when $\lambda_1 < \lambda < \lambda^*$, u_λ is an attractor.

Proof. By Theorem 2.3.2, it follows from (A_3) that (2.1.1) has a mixing transition at $(0, \lambda_0)$, and there is a branch $\Gamma_1(\lambda)$ bifurcated from $(0, \lambda_0)$ on each side of $\lambda = \lambda_0$, and when $\lambda > \lambda_0$, $u_\lambda \in \Gamma_1(\lambda)$ is an attractor, which attracts a sectorial domain $D(\theta)$ with $\lambda < \lambda_1$, $u_\lambda \in \Gamma_1(\lambda)$ is a saddle point with the Morse index one.

By Theorem 2.5.3, we need to prove that $\Gamma_1(\lambda)$ is bounded in $H \times (-\infty, \lambda_0)$ and $\Gamma_1(\lambda) = \emptyset$ at $\lambda = 0$.

In fact, by (A_1) , $B_\lambda = 0$ at $\lambda = 0$. Hence, (2.1.1) at $\lambda = 0$ is in the following form

$$\frac{du}{dt} = -Au + G(u, 0). \quad (2.5.5)$$

By (2.5.1) and (A_2) we can deduce that $u = 0$ is a global asymptotically stable singular point of (2.5.5). Hence, $\Gamma_1(0) = \emptyset$. By (A_4) , $\Gamma_1(\lambda)$ is bounded in $H \times (0, \lambda_0)$. Thus, (2.1.1) has a saddle-node bifurcation at some (u_1, λ_1) with $\lambda_1 < \lambda_0$. By the homotopy invariance of index we can derive Assertion (1).

Assertion (2) can be proved by the fact that the node $u_\lambda \in \Gamma_2(\lambda)$ separated from (u_1, λ_1) is stable as $\lambda_1 < \lambda < \lambda_1 + \varepsilon$ for some $\varepsilon > 0$. Thus the theorem is proved. \square

Remark 2.5.6 The objective of global dynamics of a nonlinear dissipative system is to study the topological structure of the global attractor and its transitions when the system parameter λ varies. Essentially, the singular separation and transition of (2.1.1) provide the key information for the topological structure of the global attractor and its transitions. As we shall see later, the theories of the singular separation and transition are important tools to understand the global behavior of phase transitions in nonlinear sciences. In particular the Type-I transition is essentially characterized by attractor bifurcation, and the Type-II and Type-III transitions often lead to singular separations.

2.5.3 Singular Separation of Periodic Orbits

We consider in this subsection the singular separation of periodic orbits. Let $\beta_1(\lambda) = \overline{\beta_2(\lambda)}$ be a pair of complex eigenvalues of L_λ . Assume that the PES (2.3.6) and (2.3.7) hold true.

For $L_\lambda = -A + B_\lambda$ and $G(\cdot, \lambda) : H_1 \rightarrow H$, we also assume that

$$\begin{aligned} (G(u, \lambda), u) &= 0 & \forall u \in H_1, \\ (L_{\lambda^*} u, u) &\leq -\alpha \|u\|_{H_{1/2}}^2 & \text{for some } \lambda^* < \lambda_0, \end{aligned} \quad (2.5.6)$$

or

$$\begin{aligned} (G(u, \lambda), u) &\leq -c_1 \|u\|^p + c_2 \|u\|^2, \\ (L_\lambda u, u) &\leq -\alpha \|u\|_{H_{1/2}} + c_\lambda \|u\|^2, \end{aligned} \quad (2.5.7)$$

where $\alpha, c_1, c_2 > 0$ are constants, $p > 2$, $c_{\lambda^*} \leq -c_2$ for some $\lambda^* < \lambda_0$, and λ_0 is as in the PES (2.3.6) and (2.3.7).

Theorem 2.5.7 Under the PES (2.3.6) and (2.3.7), the equation (2.1.1) has the Hopf bifurcation at $(0, \lambda_0)$. If the branch Σ_λ of bifurcated periodic orbits is on $\lambda < \lambda_0$, then we have the following assertions.

1. If (2.5.6) holds, and for any λ , (2.1.1) possesses a global attractor in H , then (2.1.1) has a singular separation of periodic orbits at some $(\Sigma_1, \lambda_1) \subset H \times (\lambda^*, \lambda_0)$.
2. If (2.5.7) holds true, then (2.1.1) has a singular separation of periodic orbits at some $(\Sigma_1, \lambda_1) \subset H \times (\lambda^*, \lambda_0)$.
3. The branch Σ_λ of bifurcated periodic orbits converges to Σ_1 as $\lambda \rightarrow \lambda_1$.

The proof of Theorem 2.5.7 is similar to the proofs of Theorems 2.5.4 and 2.5.5, and is omitted here.

Example 2.5.8 As an example, we consider the stability, transition, and singular separation of the 2D Navier–Stokes equations given by

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla)u &= \mu \Delta u - \nabla p + \lambda f, \\ \operatorname{div} u &= 0, \\ u|_{\partial\Omega} &= 0, \end{aligned} \tag{2.5.8}$$

where $\Omega \subset \mathbb{R}^2$ is a bounded and open set. It is known that for any $\lambda \in \mathbb{R}$, the steady-state equations of (2.5.8) are given by

$$\begin{aligned} -\mu \Delta u + (u \cdot \nabla)u + \nabla p &= \lambda f, \\ \operatorname{div} u &= 0, \\ u|_{\partial\Omega} &= 0, \end{aligned}$$

and have a solution $v_\lambda(x) \in H^2(\Omega, \mathbb{R}^2)$, and which is unique for $|\lambda| < \varepsilon$. Let $u = u' + v_\lambda(x)$. Then (2.5.8) can be rewritten as (for simplicity, we omit the prime):

$$\begin{aligned} \frac{\partial u}{\partial t} &= \mu \Delta u - (v_\lambda \cdot \nabla)u - (u \cdot \nabla)v_\lambda - \nabla p - (u \cdot \nabla)u, \\ \operatorname{div} u &= 0, \\ u|_{\partial\Omega} &= 0. \end{aligned} \tag{2.5.9}$$

It is clear then that (2.5.9) can be expressed in the form of (2.1.1). Also, it is well known that (2.5.9) possesses a global attractor; see among others (Foiaş and Temam, 1979; Temam, 1997).

Obviously, as $\lambda > 0$ sufficiently small, all eigenvalues $\beta_j(\lambda)$ of $L_\lambda = -A + B_\lambda$ satisfy that

$$\operatorname{Re}\beta_j(\lambda) < 0 \quad \forall j \geq 1.$$

Based on the theories of transition and singular separation, we can claim that for almost all $f \in L^2(\Omega, \mathbb{R}^2)$, only one of the following three assertions holds true:

1. For any $\lambda \geq 0$, the steady-state $v_\lambda(x)$ of (2.5.8) is locally asymptotically stable.
2. There exists a $\lambda_0 > 0$ such that (2.5.8) has an attractor bifurcation from $(v_{\lambda_0}, \lambda_0)$ on $\lambda > \lambda_0$.
3. (2.5.8) has a singular separation at some $(\Sigma_1, \lambda_1) \subset H \times (0, \lambda_0)$, and there is a branch Σ_λ of singular elements separating from (Σ_1, λ_1) which converges to $(v_{\lambda_0}, \lambda_0)$, i.e.

$$\lim_{\lambda \rightarrow \lambda_0} \max_{u_\lambda \in \Sigma_\lambda} \|u_\lambda - v_{\lambda_0}\| = 0.$$

2.6 Perturbed Systems

2.6.1 General Eigenvalues

We consider the following perturbed equation of (2.1.1):

$$\frac{du}{dt} = (L_\lambda + S_\lambda^\varepsilon)u + G(u, \lambda) + T^\varepsilon(u, \lambda), \quad (2.6.1)$$

where L_λ and G are as in (2.1.1), $S_\lambda^\varepsilon : X_\sigma \rightarrow X$ is a linear perturbed operator, $T^\varepsilon : X_\sigma \rightarrow X$ a C^1 nonlinear perturbed operator, and X_σ the fractional order space, $0 \leq \sigma < 1$. Also assume that G_λ, T^ε are C^3 on u , and

$$\|S_\lambda^\varepsilon\| < \varepsilon, \quad \|T^\varepsilon(\cdot, \lambda)\| < \varepsilon, \quad T^\varepsilon(u, \lambda) = o(\|u\|_\sigma). \quad (2.6.2)$$

By the continuity of eigenvalues of linear operators, under the conditions (2.1.4) and (2.1.5), for any $\varepsilon > 0$ sufficiently small, there exists a λ_0^ε with

$$\lim_{\varepsilon \rightarrow 0} \lambda_0^\varepsilon = \lambda_0,$$

such that the eigenvalues $\{\beta_j^\varepsilon(\lambda)\}$ of $L_\lambda + S_\lambda^\varepsilon$ satisfy

$$\operatorname{Re} \beta_i^\varepsilon(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_0^\varepsilon, \\ = 0 & \text{if } \lambda = \lambda_0^\varepsilon, \\ > 0 & \text{if } \lambda > \lambda_0^\varepsilon \end{cases} \quad \forall 1 \leq i \leq m_1, \quad (2.6.3)$$

$$\operatorname{Re} \beta_j^\varepsilon(\lambda_0^\varepsilon) < 0 \quad \forall j \geq m_1 + 1, \quad (2.6.4)$$

where $1 \leq m_1 \leq m$ and m is as in (2.1.4).

The following theorem determines the transition to an S^m -attractor for the perturbed system (2.6.1) at the critical point $(u, \lambda) = (0, \lambda_0^\varepsilon)$. This theorem is very useful in many nonlinear science problems. In particular, the transition theory for perturbed systems plays an important role in the study of the Taylor problem in fluid dynamics and the Walker circulation in the equatorial atmospheric layer, which will

be discussed later. This is essentially Theorem 6.14 in Ma and Wang (2005b), and we omit the proof here.

Theorem 2.6.1 *Under the hypotheses of Theorem 2.2.11, there exist $\varepsilon > 0$ and $\delta > 0$ such that if S_λ^ε and T_λ^ε satisfy (2.6.2) and $\lambda_0^\varepsilon < \lambda < \lambda_0^\varepsilon + \delta$, then the transition of (2.6.1) at $(0, \lambda_0^\varepsilon)$ enjoys the following properties:*

1. *There is a neighborhood $U \subset X$ of $u = 0$ such that (2.6.1) has an attractor $\Sigma_\lambda^\varepsilon \subset U$, $\Sigma_\lambda^\varepsilon$ attracts an open and dense set U , and*

$$\dim \Sigma_\lambda^\varepsilon \leq m, \quad 0 \notin \Sigma_\lambda^\varepsilon, \quad \lim_{\varepsilon \rightarrow 0} \sup_{x \in \Sigma_\lambda^\varepsilon} \text{dist}(x, \Sigma_\lambda) = 0, \quad (2.6.5)$$

where Σ_λ is the attractor of (2.1.1) bifurcated from $(0, \lambda_0)$.

2. *When $\varepsilon > 0$ is sufficiently small, $\Sigma_\lambda^\varepsilon$ is an $(m-1)$ -dimensional homological sphere for $\lambda > \lambda_0 + \sigma^\varepsilon$ (σ^ε a small number).*
3. *Each $u_\lambda \in \Sigma_\lambda^\varepsilon$ can be expressed as*

$$u_\lambda = v_\lambda + w_\lambda^\varepsilon, \quad v_\lambda \in E_0, \quad \lim_{\lambda \rightarrow \lambda_0} v_\lambda = 0, \quad \lim_{\lambda \rightarrow \lambda_0, \varepsilon \rightarrow 0} \frac{\|w_\lambda^\varepsilon\|}{\|v_\lambda\|} = 0, \quad (2.6.6)$$

where E_0 is the first eigenspace of L_{λ_0} as in Theorem 2.2.11.

4. *If $u = 0$ is a globally asymptotically stable equilibrium point of (2.1.1), and (2.6.1) has a global attractor for any λ near λ_0 , then $\Sigma_\lambda^\varepsilon$ attracts $X \setminus \Gamma$, where Γ is a set with codimension $\text{codim } \Gamma \geq 1$ in X .*

2.6.2 Simple Eigenvalues

In this subsection, we study the transition and singular separation of a perturbed system from simple eigenvalues.

Let the PES (2.1.4) and (2.1.5) with $m = 1$ hold true, $G(u, \lambda) = G_2(u, \lambda) + o(\|u\|_{X_1}^2)$, where $G_2(\cdot, \lambda)$ is a bilinear operator, and

$$b = (G_2(e, \lambda_0), e^*) \neq 0, \quad (2.6.7)$$

where $e \in X$ and $e^* \in X^*$ are the eigenvectors of L_λ and L_λ^* corresponding to $\beta_1(\lambda)$ at $\lambda = \lambda_0$ respectively.

The following result is a stability theorem for perturbed systems.

Theorem 2.6.2 *Assume the PES (2.1.4) and (2.1.5) with $m = 1$, and (2.6.7). Then there is an $\varepsilon > 0$ such that for any S_λ^ε and T_λ^ε satisfying (2.6.2), the transition of the perturbed system (2.6.1) at $(0, \lambda_0^\varepsilon)$ has the same structure as that of (2.1.1) at $(0, \lambda_0)$.*

This theorem is a direct corollary of the structural stability theorem for dynamic bifurcation. In fact, when (2.1.1) is reduced to the center manifold, the number b

given by (2.6.7) is the bifurcation number. Hence, using Theorems 2.2.8 and 2.2.9, one immediately derives Theorem 2.6.2, and we omit the details.

We now consider the transition associated with the saddle-node bifurcation of the perturbed system (2.6.1). Let $h(x, \lambda)$ be the center manifold function of (2.1.1) near $\lambda = \lambda_0$. Assume that

$$(G(xe + h(x, \lambda_0), \lambda_0), e^*) = b_1 x^3 + o(|x|^3), \quad (2.6.8)$$

where $b_1 \neq 0$, and e and e^* are as in (2.6.7).

Then we have the following theorems.

Theorem 2.6.3 *Let the PES (2.1.4) and (2.1.5) with $m = 1$, and (2.6.8) hold true, and $b_1 < 0$. Then there is an $\varepsilon > 0$ such that if S_λ^ε and T_λ^ε satisfy (2.6.2), then the transition of (2.6.1) is either continuous or mixed. If the transition is continuous, then Assertions (2) and (3) of Theorem 2.3.1 are valid for (2.6.1). If the transition is mixed, then the following assertions hold true:*

1. (2.6.1) has a saddle-node bifurcation at some point $(u_1, \lambda_1) \in X \times (-\infty, \lambda_0^\varepsilon)$, and there are exactly two branches

$$\Gamma_i^\lambda = \{(u_i^\lambda, \lambda) \mid \lambda_1 < \lambda < \lambda_0^\varepsilon + \delta\} \quad i = 1, 2,$$

separated from (u_1, λ_1) as shown in Fig. 2.31, which satisfy that

$$\begin{aligned} \|u_2^\lambda\| &\neq 0 & \forall(u_2^\lambda, \lambda) \in \Gamma_2^\lambda, \quad \lambda_1 < \lambda < \lambda_0^\varepsilon + \delta, \\ \lim_{\lambda \rightarrow \lambda_0^\varepsilon} \|u_1^\lambda\| &= 0 & \forall(u_1^\lambda, \lambda) \in \Gamma_1^\lambda. \end{aligned}$$

2. There is a neighborhood $U \subset X$ of $u = 0$, such that for each λ with $\lambda_1 < \lambda < \lambda_0^\varepsilon + \delta$ and $\lambda \neq \lambda_0^\varepsilon$, U contains only two nontrivial singular points u_1^λ and u_2^λ of (2.6.1).
3. For each $\lambda_1 < \lambda < \lambda_0^\varepsilon + \delta$, U can be decomposed into two open sets $\overline{U} = \overline{U}_1^\lambda + \overline{U}_2^\lambda$ with $U_1^\lambda \cap U_2^\lambda = \emptyset$, such that
 - a. if $\lambda_1 < \lambda < \lambda_0^\varepsilon$, then $0 \in U_1^\lambda$, $u_2^\lambda \in U_2^\lambda$, $u_1^\lambda \in \partial U_1^\lambda \cap \partial U_2^\lambda$, with $u = 0$ and u_2^λ being attractors which attract U_1^λ and U_2^λ respectively, and
 - b. if $\lambda_0^\varepsilon < \lambda < \lambda_0^\varepsilon + \delta$, then $u_1^\lambda \in U_1^\lambda$, $u_2^\lambda \in U_2^\lambda$, $0 \in \partial U_1^\lambda \cap \partial U_2^\lambda$, with u_1^λ and u_2^λ being attractors which attract U_1^λ and U_2^λ respectively.
4. Near $(u, \lambda) = (0, \lambda_0^\varepsilon)$, u_1^λ and u_2^λ can be expressed as

$$\begin{aligned} u_1^\lambda &= \alpha_1(\lambda, \varepsilon)e + o(|\alpha_1|), & u_2^\lambda &= \alpha_2(\lambda, \varepsilon)e + o(|\alpha_2|), \\ \lim_{\lambda \rightarrow \lambda_0^\varepsilon} \alpha_1(\lambda, \varepsilon) &= 0, & \alpha_2(\lambda_0^\varepsilon, \varepsilon) &\neq 0, \end{aligned} \quad (2.6.9)$$

where e is as in (2.6.8).

Theorem 2.6.4 *Assume the PES (2.1.4) and (2.1.5) with $m = 1$, and (2.6.8) with $b_1 > 0$. Then, there is an $\varepsilon > 0$ such that when S_λ^ε and T_λ^ε satisfy (2.6.2), the*

transition of (2.6.1) is either jump or mixed. If it is jump transition, then Assertions (1) and (3) of Theorem 2.3.1 are valid for (2.6.1). If it is mixed, then the following assertions hold true:

1. (2.6.1) has a saddle-node bifurcation at some point (u_1, λ_1) with $\lambda_0^\varepsilon < \lambda_1$, and there are exactly two branches

$$\Gamma_i^\lambda = \{(u_i^\lambda, \lambda) \mid \lambda_0^\varepsilon - \delta < \lambda < \lambda_1\} \quad (i = 1, 2),$$

separated from (u_1, λ_1) , which satisfy

$$\begin{aligned} \|u_2^\lambda\| &\neq 0 & \forall (u_2^\lambda, \lambda) \in \Gamma_2^\lambda, \quad \lambda_0^\varepsilon - \varepsilon < \lambda < \lambda_1, \\ \lim_{\lambda \rightarrow \lambda_0^\varepsilon} \|u_1^\lambda\| &= 0 & \forall (u_1^\lambda, \lambda) \in \Gamma_1^\lambda. \end{aligned}$$

2. There is a neighborhood $U \subset X$ of $u = 0$, such that for each λ with $\lambda_0^\varepsilon - \delta < \lambda < \lambda_1$, U contains only two nontrivial singular points u_1^λ and u_2^λ of (2.6.1).
3. For every $\lambda_0^\varepsilon - \delta < \lambda < \lambda_1$, U can be decomposed into three open sets $\overline{U} = \overline{U}_0 + \overline{U}_1 + \overline{U}_2$ with $U_i \cap U_j = \emptyset$ ($i \neq j$) such that
 - a. if $\lambda_0^\varepsilon - \delta < \lambda < \lambda_0^\varepsilon$, then

$$u = 0 \in U_0^\lambda, \quad u_i^\lambda \in \partial U_i^\lambda \cap \partial U_0^\lambda \quad (i = 1, 2),$$

$u = 0$ is an attractor which attracts U_0^λ , and u_i^λ ($i = 1, 2$) are two saddle points with Morse index one; and

- b. if $\lambda_0^\varepsilon < \lambda < \lambda_1$, then

$$u_1^\lambda \in U_1^\lambda, \quad u_2^\lambda \in \partial U_2^\lambda \cap \partial U_1^\lambda, \quad 0 \in \partial U_1^\lambda \cap \partial U_0^\lambda,$$

with u_1^λ being an attractor which attracts U_1^λ and u_2^λ and $u = 0$ being saddle points with the Morse index one.

4. Near $(0, \lambda_0^\varepsilon)$, u_1^λ and u_2^λ can be expressed by (2.6.9).

Remark 2.6.5 Theorems 2.6.3 and 2.6.4 give precisely the relation between transition and singular separation, which is helpful to understand the global structure of transition.

Proof of Theorems 2.6.3 and 2.6.4. We only need to prove Theorem 2.6.3, and the proof of Theorem 2.6.4 is the same.

By the PES (2.1.4) and (2.1.5) with $m = 1$, and by (2.6.8), the reduced equation of (2.6.1) near $(0, \lambda_0^\varepsilon)$ is written in the following form

$$\frac{dx}{dt} = \beta_1^\varepsilon(\lambda)x + b_\lambda^\varepsilon x^3 + g(x, \varepsilon) + o(|x|^3), \quad (2.6.10)$$

where $x \in \mathbb{R}^1$, $\beta_1^\varepsilon(\lambda)$ as in (2.6.3) with $m_1 = 1$, $b_\lambda^\varepsilon \rightarrow b_1$ as $\lambda \rightarrow \lambda_0$ and $\varepsilon \rightarrow 0$, and

$$g(x, \varepsilon) = (T^\varepsilon(xe_\varepsilon(\lambda) + h_\varepsilon(x, \lambda), \lambda), e_\varepsilon^*(\lambda)),$$

where $e_\varepsilon(\lambda)$ and $e_\varepsilon^*(\lambda)$ are the eigenvectors of $L_\lambda + S_\lambda^\varepsilon$ and $L_\lambda^* + S_\lambda^{\varepsilon*}$ corresponding to $\beta_1^\varepsilon(\lambda)$ respectively.

Since G_λ and T_λ^ε are C^3 on u , by (2.6.2) we have

$$\begin{aligned} g(x, \varepsilon) &= a_1(\varepsilon)x^2 + a_2(\varepsilon)x^3 + o(|x|^3), \\ \lim_{\varepsilon \rightarrow 0} a_i(\varepsilon) &= 0 \quad i = 1, 2. \end{aligned} \tag{2.6.11}$$

Notice that

$$b_\lambda^\varepsilon + a_2(\varepsilon) \rightarrow b_1, \quad \text{as } \varepsilon \rightarrow 0, \quad \lambda \rightarrow \lambda_0.$$

Therefore, when $\varepsilon > 0$ is small, if $a_1(\varepsilon) \equiv 0$, then the equation (2.6.10) with (2.6.11) satisfies the conditions of Theorem 2.3.1 with $\alpha < 0$. In this case, the transition of (2.6.1) at $(0, \lambda_0^\varepsilon)$ is continuous, and Assertions (2) and (3) of Theorem 2.3.1 are valid for (2.6.1).

When $a_1(\varepsilon) \neq 0$, the equation (2.6.10) with (2.6.11) satisfies the conditions of Theorem 2.3.2. Hence, we only need to prove that the following equation

$$\beta_1^\varepsilon(\lambda)x + a_1(\varepsilon)x^2 + (b_\lambda^\varepsilon + a_2(\varepsilon))x^3 + o(|x|^3) = 0 \tag{2.6.12}$$

has exactly two solutions x_1^λ and x_2^λ near λ_0^ε with $\lambda \neq \lambda_0^\varepsilon$ satisfying

$$\begin{aligned} x_1^\lambda &\rightarrow 0 \quad \text{as } \lambda \rightarrow \lambda_0^\varepsilon, \\ x_2^\lambda &\neq 0 \quad \forall \lambda_0^\varepsilon - \delta < \lambda < \lambda_0^\varepsilon + \delta, \quad \text{for some } \delta > 0, \end{aligned} \tag{2.6.13}$$

and there exists a $\lambda_1 < \lambda_0^\varepsilon$ such that if $\lambda > \lambda_1$, (2.6.12) has only one solution $x^* \neq 0$ near $x = 0$, with

$$\lim_{\lambda \rightarrow \lambda_1^+} x_i^\lambda \rightarrow x^* \quad i = 1, 2, \tag{2.6.14}$$

and if $\lambda < \lambda_1$, (2.6.12) has no real nontrivial solutions near $x = 0$.

Let $b_\lambda = b_\lambda^\varepsilon + a_2(\varepsilon)$. Then the equation (2.6.12) is rewritten as

$$b_\lambda x^2 + a_1(\varepsilon)x + \beta_1^\varepsilon(\lambda) + o(|x|^2) = 0. \tag{2.6.15}$$

It is clear that (2.6.15) has only two nonzero solutions near $x = 0$:

$$x_{1,2} = \frac{-a_1(\varepsilon) \mp \sqrt{a_1^2(\varepsilon) - 4b_\lambda \beta_1^\varepsilon(\lambda)}}{2b_\lambda} + o(|x_{1,2}|). \tag{2.6.16}$$

By assumption, $b_1 < 0$, hence $b_\lambda < 0$. In addition,

$$\beta_1^\varepsilon(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_0^\varepsilon, \\ = 0 & \text{if } \lambda = \lambda_0^\varepsilon, \\ > 0 & \text{if } \lambda > \lambda_0^\varepsilon. \end{cases}$$

Hence, the solutions x_1^λ and x_2^λ given by (2.6.16) satisfy (2.6.13), and for the number $\lambda_1 < \lambda_0^\varepsilon$ satisfying

$$a_1^2(\varepsilon) = 4b_{\lambda_1}\beta_1^\varepsilon(\lambda_1),$$

(2.6.14) holds true, where $x^* = -a_1(\varepsilon)/2b_{\lambda_1}$. Moreover, if $\lambda < \lambda_1$, $a_1^2(\varepsilon) < 4b_\lambda\beta_1^\varepsilon(\lambda)$. Hence, (2.6.15) has no real solutions.

Thus, the proof is complete. \square

In the following, we consider a special case which can be applied to phase transitions in fluid dynamics. For the perturbed system (2.6.1), we assume that

$$\begin{aligned} X_1 \text{ and } X &\quad \text{are Hilbert spaces,} \\ L_\lambda : X_1 \rightarrow X &\quad \text{is symmetric,} \\ G : X_1 \rightarrow X &\quad \text{is a bilinear operator,} \\ (G(u, v), v) = 0 &\quad \forall u, v \in X_1. \end{aligned} \tag{2.6.17}$$

Theorem 2.6.6 Assume that the PES (2.1.4) and (2.1.5) with $m = 1$, and (2.6.17) hold true. Then the number b_1 defined in (2.6.8) is negative, $b_1 < 0$. Hence the assertions of Theorem 2.6.3 hold true.

Proof. Since L_λ is symmetric, the eigenvectors $e = e^*$. Let $y = h(x, \lambda_0)$ be the center manifold function of (2.1.1) at $\lambda = \lambda_0$. Then by (2.6.17) we find

$$\begin{aligned} (G(xe + y), e) &= (G(xe + y, xe + y), e) \\ &= x(G(e, y), e) + (G(y, y), e) \\ &= -x(G(e), y) + (G(y), e). \end{aligned} \tag{2.6.18}$$

Let $\{e_k \mid k = 1, 2, \dots\}$ be all eigenvectors of L_λ with $e_1 = e$, which constitute an orthogonal basis of X . Hence, y can be expanded as

$$y = \sum_{j=2}^{\infty} y_j e_j.$$

By Theorem A.1.1, the center manifold function $y = y(x)$ can be written as

$$\mathcal{L}_\lambda y = -P_2 G(xe) + o(|x|^2) + O(|\beta_1| |x|^2), \tag{2.6.19}$$

where $\mathcal{L}_\lambda e_j = \beta_j(\lambda) e_j (j = 2, 3, \dots)$, $P_2 : X \rightarrow E_2 = \text{span}\{e_2, e_3, \dots\}$ is the canonical projection. Thus, we derive from (2.6.19) that

$$y_j = -\beta_j^{-1}(\lambda) \alpha_j x^2 + o(|x|^2, |x| |\beta_1|), \tag{2.6.20}$$

where for $j = 2, 3, \dots$,

$$\alpha_j = (G(e), e_j).$$

Putting (2.6.20) in (2.6.18), we find

$$(G(xe + y), e) = - \sum_{j=2}^{\infty} \alpha_j xy_j + O(\|y\|^2) = b_\lambda x^3 + o(|x|^3, |x|^2 |\beta_1|),$$

where $b_\lambda = \sum_{j=2}^{\infty} \alpha_j^2 \beta_j^{-1}(\lambda)$.

By the PES (2.1.4) and (2.1.5) with $m = 1$, $\beta_j(\lambda_0) < 0$ for any $j \geq 2$. Therefore $b_1 = b_{\lambda_0} < 0$. Thus, the theorem is proved. \square

2.6.3 Complex Eigenvalues

Let $\beta_1(\lambda) = \overline{\beta_2(\lambda)}$ be a pair of complex eigenvalues of L_λ , $e_1(\lambda)$, $e_2(\lambda)$ and $e_1^*(\lambda)$, $e_2^*(\lambda)$ be the eigenvectors of L_λ and L_λ^* corresponding to $\beta_1(\lambda)$, $\beta_2(\lambda)$ respectively. Let $h(x, \lambda)$ be the center manifold function of (2.1.1) near $\lambda = \lambda_0$. Under the PES (2.3.6) and (2.3.7), we shall discuss the transition and singular separation of periodic orbits of (2.6.1) at $(0, \lambda_0^\varepsilon)$.

We first consider the perturbation stability of transition. To this end, we take the following expansion

$$(G(x + h(x, \lambda), \lambda), e_l^*) = \sum_{2 \leq i+j \leq 3} a_{ij}^l x_1^i x_2^j + o(|x|^2), \quad (2.6.21)$$

for $l = 1, 2$, $x = x_1 e_1 + x_2 e_2$.

For (2.6.21) and $\beta_1(\lambda) = \alpha + i\sigma$, we can introduce a number b as defined by (2.3.15), which can determine not only the types but also the stability of transition of (2.1.1) at $(0, \lambda_0)$. The following is the transition stability theorem of perturbed system (2.6.1).

Theorem 2.6.7 *Under the PES (2.3.6) and (2.3.7), and (2.6.21), if the number b given by (2.3.15) is nonzero, i.e., $b \neq 0$, then the transition structure of the perturbed system (2.6.1) at $(0, \lambda_0^\varepsilon)$ is the same as that of (2.1.1) at $(0, \lambda_0)$. More precisely, when $b < 0$ the transition of (2.6.1) is continuous, and (2.6.1) bifurcates from $(0, \lambda_0^\varepsilon)$ to an attractor which is a periodic orbit; when $b > 0$ the transition is a jump transition, and (2.6.1) bifurcates on $\lambda < \lambda_0^\varepsilon$ to a unique periodic orbit.*

Proof. Under the PES (2.3.6) and (2.3.7), the reduced equation of (2.1.1) on the center manifold reads

$$\begin{aligned} \frac{dx_1}{dt} &= \alpha(\lambda)x_1 - \sigma(\lambda)x_2 + (G(x + h, \lambda), e_1^*), \\ \frac{dx_2}{dt} &= \sigma(\lambda)x_1 + \alpha(\lambda)x_2 + (G(x + h, \lambda), e_2^*). \end{aligned} \quad (2.6.22)$$

As in the proof of Theorem 2.6.6, from (2.6.21) and (2.6.22) one can derive the number b given by (2.3.15), which is the bifurcation number as in Theorem 2.2.8. Thus, this theorem is deduced from Theorems 2.2.8, 2.2.9, and 2.3.7. The proof is complete. \square

In the following we consider the instability of transitions for the perturbed system (2.6.1), which is associated with the singular separation. As seen in Theorems 2.6.3 and 2.6.4, when the bifurcation number $b = 0$, the transition structure of (2.6.1) may be different from the original equation (2.1.1).

Let $G(x + h, \lambda)$ have the Taylor expansion near $\lambda = \lambda_0$:

$$(G(x + h(x, \lambda), \lambda), e_l^*(\lambda)) = \sum_{3 \leq i+j \leq 5} a_{ij}^l x_1^i x_2^j + o(|x|^5), \quad (2.6.23)$$

for $l = 1, 2$. For (2.6.23) we introduce two numbers as follows

$$b_1 = \frac{3}{4}\pi(a_{30}^1 + a_{03}^2) + \frac{\pi}{4}(a_{12}^1 + a_{21}^2), \quad (2.6.24)$$

$$\begin{aligned} b_2 = & \frac{5\pi}{8}(a_{50}^1 + a_{05}^2) + \frac{\pi}{8}(a_{32}^1 + a_{14}^1 + a_{23}^2 + a_{41}^2) \\ & + \frac{17\pi}{32\sigma}(a_{03}^1 a_{03}^2 - a_{30}^1 a_{30}^2) + \frac{3\pi}{32\sigma}(2a_{30}^1 a_{21}^1 + 2a_{03}^1 a_{12}^1 + a_{30}^1 a_{30}^2 \\ & + a_{03}^1 a_{21}^2 + a_{12}^1 a_{12}^2 + a_{21}^1 a_{03}^2 - 2a_{30}^2 a_{21}^2 - 2a_{03}^2 a_{12}^2 - a_{03}^1 a_{03}^2 \\ & - a_{30}^1 a_{12}^2 - a_{21}^1 a_{21}^2 - a_{12}^1 a_{30}^2) + \frac{3\pi}{64\sigma}(2a_{30}^1 a_{03}^1 + 2a_{12}^1 a_{21}^1 \\ & + a_{12}^1 a_{30}^2 + a_{21}^1 a_{21}^2 + a_{30}^1 a_{12}^2 - 2a_{30}^2 a_{03}^2 - 2a_{12}^2 a_{21}^2 \\ & - a_{21}^1 a_{03}^2 - a_{12}^1 a_{12}^2 - a_{03}^1 a_{21}^2). \end{aligned} \quad (2.6.25)$$

Theorem 2.6.8 *Let the PES (2.3.6) and (2.3.7), and (2.6.23) hold true, the numbers $b_1 = 0$ in (2.6.24), and $b_2 < 0$ in (2.6.25). If the transition at $(0, \lambda_0^\varepsilon)$ is continuous, then (2.6.1) bifurcates on $\lambda_0^\varepsilon < \lambda$ to a unique periodic orbit which is an attractor. If the transition is jump, then the following assertions hold true:*

1. (2.6.1) has a singular separation of periodic orbits at some $(\Sigma_1, \lambda_1) \subset X \times (-\infty, \lambda_0^\varepsilon)$, and there are exactly two branches of periodic orbits $\Sigma_1^\lambda (\lambda_1 < \lambda < \lambda_0^\varepsilon)$ and $\Sigma_2^\lambda (\lambda_1 < \lambda < \lambda_0^\varepsilon + \delta)$, separated from (Σ_1, λ_1) , as shown in Fig. 2.33a, which satisfy

$$\begin{aligned} \lim_{\lambda \rightarrow \lambda_0^\varepsilon} \Sigma_1^\lambda &= \{0\}, & \lim_{\lambda \rightarrow \lambda_1} \Sigma_i^\lambda &= \Sigma_1 \quad i = 1, 2, \\ \text{dist}(\Sigma_2^\lambda, 0) &> 0 & \forall \lambda_1 < \lambda < \lambda_0^\varepsilon + \delta \quad (\delta > 0). \end{aligned}$$

2. There is a neighborhood $U \subset X$ of $u = 0$, such that if $\lambda_1 < \lambda < \lambda_0^\varepsilon$, U can be decomposed into two open sets $\overline{U} = \overline{U}_1^\lambda + \overline{U}_2^\lambda$ with $U_1^\lambda \cap U_2^\lambda = \emptyset$, such that

$$\Sigma_2^\lambda \subset U_2^\lambda, \quad 0 \in U_1^\lambda, \quad \Sigma_1^\lambda \subset \partial U_1^\lambda \cap \partial U_2^\lambda,$$

and Σ_2^λ and $u = 0$ are attractors attracting U_1^λ and U_2^λ respectively.

3. If $\lambda_0^\varepsilon < \lambda < \lambda_0^\varepsilon + \delta$, Σ_2^λ is an attractor attracting $U \setminus \Gamma$, where Γ is the stable manifold of $u = 0$ with $\text{codim } \Gamma = 2$.

Theorem 2.6.9 Let the PES (2.3.6) and (2.3.7), and (2.6.23) hold true, $b_1 = 0$, and $b_2 > 0$. If the transition at $(0, \lambda_0^\varepsilon)$ is jump, then (2.6.1) bifurcates on $\lambda < \lambda_0^\varepsilon$ to a unique periodic orbit which is a repeller. If it is continuous, then the following assertions hold true:

1. (2.6.1) has a singular separation of periodic orbits at some $(\Sigma_1, \lambda_1) \subset X \times (\lambda_0^\varepsilon, +\infty)$, and there are exactly two branches of periodic orbits Σ_1^λ ($\lambda_0^\varepsilon < \lambda < \lambda_1$) and Σ_2^λ ($\lambda_0^\varepsilon - \delta < \lambda < \lambda_1$), separated from (Σ_1, λ_1) , as shown in Fig. 2.33b, which satisfy

$$\lim_{\lambda \rightarrow \lambda_0^\varepsilon} \Sigma_1^\lambda = \{0\}, \quad \lim_{\lambda \rightarrow \lambda_1} \Sigma_i^\lambda = \Sigma_1 \quad i = 1, 2,$$

$$\text{dist}(\Sigma_2^\lambda, 0) > 0 \quad \forall \lambda_0^\varepsilon - \delta < \lambda < \lambda_1.$$

2. When $\lambda_0^\varepsilon < \lambda < \lambda_1$, Σ_1^λ is an attractor which attracts $U_\lambda \setminus \Gamma$, where U_λ is a neighborhood of $u = 0$, Γ the stable manifold of $u = 0$ with $\text{codim } \Gamma = 2$, and $\Sigma_2^\lambda \subset \partial U_\lambda$ with ∂U_λ being a stable manifold of Σ_2^λ .
3. When $\lambda_0^\varepsilon - \delta < \lambda < \lambda_0^\varepsilon$, $u = 0$ is an attractor which attracts a neighborhood U_λ of $u = 0$, and $\Sigma_2^\lambda \subset \partial U_\lambda$ with ∂U_λ being a stable manifold of Σ_2^λ .

Proof of Theorems 2.6.8 and 2.6.9. We only need to prove Theorem 2.6.8, and the proof of Theorem 2.6.9 is the same.

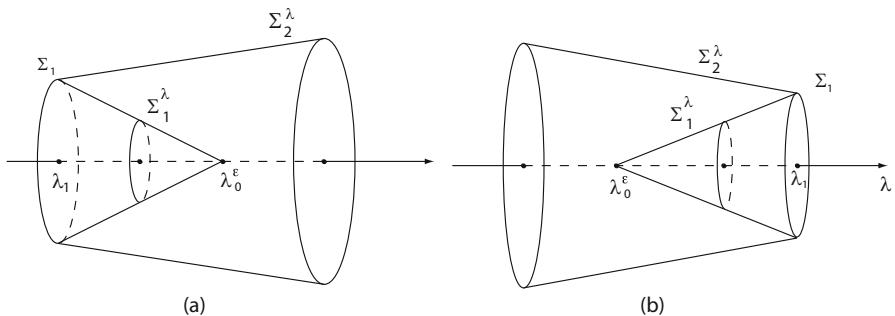


Fig. 2.33 (a) The case of $b_2 < 0$; (b) the case of $b_2 > 0$.

By (2.6.23), the reduced equation of (2.6.1) on the center manifold near $\lambda = \lambda_0^\varepsilon$ can be written as

$$\begin{aligned} \frac{dx_1}{dt} &= \alpha_\varepsilon(\lambda)x_1 - \sigma_\varepsilon x_2 + \sum_{2 \leq i+j \leq 5} a_{ij}^1(\varepsilon)x_1^i x_2^j + o(|x|^5), \\ \frac{dx_2}{dt} &= \sigma_\varepsilon x_1 + \alpha_\varepsilon(\lambda)x_1 + \sum_{2 \leq i+j \leq 5} a_{ij}^2(\varepsilon)x_1^i x_2^j + o(|x|^5), \end{aligned} \tag{2.6.26}$$

where

$$\alpha_\varepsilon(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_0^\varepsilon, \\ = 0 & \text{if } \lambda = \lambda_0^\varepsilon, \\ > 0 & \text{if } \lambda > \lambda_0^\varepsilon, \end{cases} \quad (2.6.27)$$

$$\begin{cases} \lim_{\varepsilon \rightarrow 0} \sigma_\varepsilon = \sigma \neq 0, \\ \lim_{\varepsilon \rightarrow 0} a_{ij}^l(\varepsilon) = 0 \quad \text{for } l = 1, 2 \text{ and } i + j = 2, \\ \lim_{\varepsilon \rightarrow 0} a_{ij}^l(\varepsilon) = a_{ij}^l \quad \text{for } l = 1, 2 \text{ and } 3 \leq i + j \leq 5. \end{cases} \quad (2.6.28)$$

Taking the polar coordinate $x_1 = r \cos \theta, x_2 = r \sin \theta$, we derive then from (2.6.26) that

$$\frac{dr}{d\theta} = \frac{\alpha_\varepsilon(\lambda)r + \sum_{k=2}^5 r^k u_k^\varepsilon(\cos \theta, \sin \theta) + o(r^5)}{\sigma_\varepsilon - \sum_{k=2}^5 r^{k-1} v_k^\varepsilon(\cos \theta, \sin \theta) + o(r^4)}, \quad (2.6.29)$$

where

$$\begin{aligned} u_k^\varepsilon &= \sum_{i+j=k} [a_{ij}^1(\varepsilon) \cos^{i+1} \theta \sin^j \theta + a_{ij}^2(\varepsilon) \cos^i \theta \sin^{j+1} \theta], \\ v_k^\varepsilon &= \sum_{i+j=k} [a_{ij}^1(\varepsilon) \cos^i \theta \sin^{j+1} \theta - a_{ij}^2(\varepsilon) \cos^{i+1} \theta \sin^j \theta]. \end{aligned}$$

By $\sigma_\varepsilon \rightarrow \sigma \neq 0$ ($\varepsilon \rightarrow 0$), near $r = 0$, (2.6.29) can be expressed as

$$\frac{1}{r^2} \frac{dr}{d\theta} = \frac{1}{\sigma_\varepsilon} \left[\frac{\alpha_\varepsilon}{r} + \sum_{k=2}^5 r^{k-2} f_k^\varepsilon(\cos \theta, \sin \theta) + o(r^3) \right], \quad (2.6.30)$$

where

$$\begin{aligned} f_2^\varepsilon &= u_2^\varepsilon + \sigma_\varepsilon^{-1} \alpha_\varepsilon v_2^\varepsilon, \\ f_3^\varepsilon &= u_3^\varepsilon + \sigma_\varepsilon^{-1} \alpha_\varepsilon v_3^\varepsilon + \sigma_\varepsilon^{-1} u_2^\varepsilon v_2^\varepsilon + \alpha_\varepsilon \sigma_\varepsilon^{-2} v_2^{\varepsilon 2}, \\ f_4^\varepsilon &= u_4^\varepsilon + \sigma_\varepsilon^{-1} \alpha_\varepsilon v_4^\varepsilon + \sigma_\varepsilon^{-1} u_2^\varepsilon v_3^\varepsilon + \sigma_\varepsilon^{-1} u_3^\varepsilon v_2^\varepsilon + 2\alpha_\varepsilon \sigma_\varepsilon^{-2} v_2^\varepsilon v_3^\varepsilon \\ &\quad + \sigma_\varepsilon^{-2} u_2^\varepsilon v_2^{\varepsilon 2} + \alpha_\varepsilon \sigma_\varepsilon^{-3} v_2^{\varepsilon 3}, \\ f_5^\varepsilon &= u_5^\varepsilon + \sigma_\varepsilon^{-1} \alpha_\varepsilon v_5^\varepsilon + \sigma_\varepsilon^{-1} u_2^\varepsilon v_4^\varepsilon + \sigma_\varepsilon^{-1} u_3^\varepsilon v_3^\varepsilon + \sigma_\varepsilon^{-1} u_4^\varepsilon v_2^\varepsilon \\ &\quad + 2\alpha_\varepsilon \sigma_\varepsilon^{-2} v_2^\varepsilon v_4^\varepsilon + 2\sigma_\varepsilon^{-2} u_2^\varepsilon v_2^\varepsilon v_3^\varepsilon + \sigma_\varepsilon^{-1} u_3 v_2^{\varepsilon 2} \\ &\quad + 3\alpha_\varepsilon \sigma_\varepsilon^{-3} v_2^{\varepsilon 2} v_3^\varepsilon + \sigma_\varepsilon^{-3} u_2^\varepsilon v_2^{\varepsilon 3} + \alpha_\varepsilon \sigma_\varepsilon^{-4} v_2^{\varepsilon 4}. \end{aligned}$$

The initial condition is

$$r(0, \lambda, a) = a \quad (a > 0). \quad (2.6.31)$$

Let $r(\theta, \lambda, a)$ have the Taylor expansion as follows

$$r(\theta, \lambda, a) = a + r_2(\theta, \lambda)a^2 + r_3(\theta, \lambda)a^3 + \alpha_\varepsilon h + o(a^3), \quad (2.6.32)$$

where $\alpha_\varepsilon h(\theta, \lambda, a)$ is contributed by the term containing α_ε in (2.6.30) which vanishes at $\lambda = \lambda_0^\varepsilon$.

Putting (2.6.32) in (2.6.30), we obtain

$$\frac{dr}{d\theta} = \alpha_\varepsilon H(\theta, \lambda, a) + (a + r_2 a^2)^2 f_2^\varepsilon + a^3 f_3^\varepsilon + o(a^3). \quad (2.6.33)$$

Integrating (2.6.33) on θ , by (2.6.31) we get

$$\begin{aligned} r(\theta, \lambda, a) &= a + a^2 \int_0^\theta f_2^\varepsilon d\theta + a^3 \int_0^\theta (f_3^\varepsilon + 2r_2 f_2^\varepsilon) d\theta \\ &\quad + \alpha_\varepsilon(\lambda) \int_0^\theta H(\theta, \lambda, a) d\theta + o(a^3). \end{aligned} \quad (2.6.34)$$

In comparison of the coefficients of (2.6.32) and (2.6.34), we find

$$\begin{aligned} r_2(\theta, \lambda) &= \int_0^\theta f_2^\varepsilon(\cos \theta, \sin \theta) d\theta, \\ r_3(\theta, \lambda) &= \int_0^\theta (f_3^\varepsilon(\cos \theta, \sin \theta) + 2r_2(\theta, \lambda) f_2^\varepsilon(\cos \theta, \sin \theta)) d\theta. \end{aligned} \quad (2.6.35)$$

Inserting (2.6.35) in (2.6.30), and integrating it from $\theta = 0$ to $\theta = 2\pi$, we deduce that

$$\begin{aligned} \frac{r(2\pi, 0) - r(0, a)}{r(2\pi, a)} &= \frac{1}{\sigma_\varepsilon} \left\{ \rho(a, \lambda) \alpha_\varepsilon(\lambda) + a \int_0^{2\pi} f_2^\varepsilon d\theta \right. \\ &\quad + a^2 \int_0^{2\pi} f_3^\varepsilon d\theta + a^3 \int_0^{2\pi} \left[f_4^\varepsilon + f_3^\varepsilon \left(\int_0^\theta f_2^\varepsilon d\theta \right) \right] d\theta \\ &\quad + a^4 \int_0^{2\pi} \left[f_5^\varepsilon + 2f_4^\varepsilon \left(\int_0^\theta f_2^\varepsilon d\theta \right) + f_3^\varepsilon \left(\int_0^\theta f_3^\varepsilon d\theta \right) \right. \\ &\quad \left. \left. + 2f_3^\varepsilon \left(\int_0^\theta f_2^\varepsilon \left(\int_0^\theta f_2^\varepsilon d\theta \right) d\theta \right) \right] d\theta + \alpha_\varepsilon g(\lambda, a) \right\} + o(a^4), \end{aligned} \quad (2.6.36)$$

where $\rho(a, \lambda) = 2\pi + o(a)$ and $g(\lambda, 0) = 0$.

By a direct computation, we have

$$\begin{aligned} \int_0^{2\pi} f_2^\varepsilon(\cos \theta, \sin \theta) d\theta &= 0, \\ \int_0^{2\pi} f_4^\varepsilon(\cos \theta, \sin \theta) d\theta &= 0, \\ \int_0^{2\pi} f_3^\varepsilon \left(\int_0^\theta f_2^\varepsilon d\theta \right) d\theta &= \frac{2}{3} a_{02}^2(\varepsilon) \int_0^{2\pi} f_3^\varepsilon d\theta + o(\alpha_\varepsilon), \\ \int_0^{2\pi} f_3^\varepsilon \left(\int_0^\theta f_3^\varepsilon d\theta \right) d\theta &= 0. \end{aligned}$$

Hence, (2.6.36) can be rewritten as

$$\frac{r(2\pi, a) - r(0, a)}{r(2\pi, a)} = \varphi(a, \lambda) + \delta(\varepsilon)a^2 + \delta_1(\varepsilon)a^3 + b_\varepsilon(\lambda)a^4 + o(a^5) \quad (2.6.37)$$

where $\varphi(a, \lambda) = \sigma_\varepsilon^{-1}(\rho(a, \lambda) + g(\lambda, a))\alpha_\varepsilon(\lambda)$, by (2.6.27) which satisfies

$$\varphi(\lambda, a) \begin{cases} < 0 & \text{if } \lambda < \lambda_0^\varepsilon, \\ = 0 & \text{if } \lambda = \lambda_0^\varepsilon, \\ > 0 & \text{if } \lambda > \lambda_0^\varepsilon, \end{cases} \quad (2.6.38)$$

and

$$\begin{aligned} \delta(\varepsilon) &= \int_0^{2\pi} f_3^\varepsilon(\cos \theta, \sin \theta) d\theta, \\ \delta_1(\varepsilon) &= \frac{2}{3} a_{02}^2(\varepsilon) \delta(\varepsilon), \\ b_\varepsilon(\lambda) &= \int_0^{2\pi} \left[f_5^\varepsilon + 2f_4^\varepsilon \left(\int_0^\theta f_2^\varepsilon \right) + 2f_3^\varepsilon \left(\int_0^\theta f_2^\varepsilon \left(\int_0^\theta f_2^\varepsilon \right) \right) \right] d\theta \end{aligned}$$

By (2.6.28), we see that

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} (u_2^\varepsilon, v_2^\varepsilon, f_2^\varepsilon) &= 0, \\ \lim_{\varepsilon \rightarrow 0} f_3^\varepsilon &= f_3 = u_3 + \sigma^{-1}\alpha(\lambda)v_3, \\ \lim_{\varepsilon \rightarrow 0} f_5^\varepsilon &= f_5 = u_5 + \sigma^{-1}u_3v_3. \end{aligned}$$

Hence, we have

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \delta(\varepsilon) &= b_1(\lambda) = \int_0^{2\pi} [u_3 + \sigma^{-1}\alpha(\lambda)v_3] d\theta, \\ \lim_{\varepsilon \rightarrow 0} b_\varepsilon(\lambda_0) &= b_2 = \int_0^{2\pi} [u_5 + \sigma^{-1}u_3v_3] d\theta, \end{aligned}$$

where $b_1(\lambda_0) = \int_0^{2\pi} u_3 d\theta$ is the number given by (2.6.24), and b_2 is given by (2.6.25). By the assumption: $b_1(\lambda_0) = 0$, we have

$$\begin{aligned} \delta(\varepsilon), \delta_1(\varepsilon) &\rightarrow 0 & \text{if } \lambda \rightarrow \lambda_0, \varepsilon \rightarrow 0, \\ b_\varepsilon(\lambda) &< -C & \text{for some } C > 0 \text{ near } \lambda = \lambda_0^\varepsilon. \end{aligned} \quad (2.6.39)$$

It is known that each real singular point $a_0 > 0$ of (2.6.37) is associated with a periodic orbit of (2.6.26), and if

$$r(2\pi, a) - r(0, a) \begin{cases} < 0 & \text{if } a_0 < a, \\ > 0 & \text{if } a_0 > a, \end{cases} \quad (2.6.40)$$

then the periodic orbit associated with a_0 is an attractor, if

$$r(2\pi, a) - r(0, a) \begin{cases} > 0 & \text{if } a_0 < a, \\ < 0 & \text{if } a_0 > a, \end{cases} \quad (2.6.41)$$

then the periodic orbit is a repeller.

Consider the equation

$$\varphi(a, \lambda) + \delta(\varepsilon)a^2 + \delta_1(\varepsilon)a^3 + b_\varepsilon(\lambda)a^4 + o(a^5) = 0. \quad (2.6.42)$$

It follows from (2.6.38) and (2.6.39) that when $\delta(\varepsilon) \leq 0$ the equation (2.6.42) has a unique real positive solution $a_0 > 0$ near $(a, \lambda) = (0, \lambda_0^\varepsilon)$ on $\lambda > \lambda_0^\varepsilon$, which satisfies (2.6.40), i.e., the equation (2.6.26) bifurcates from $(x, \lambda) = (0, \lambda_0^\varepsilon)$ to an attractor consisting of one periodic orbit. When $\delta(\varepsilon) > 0$, near $(a, \lambda) = (0, \lambda_0^\varepsilon)$ in $\lambda_1 < \lambda < \lambda_0^\varepsilon$, (2.6.42) has exactly two positive real solutions

$$a_\pm = \left[\frac{\delta(\varepsilon) \pm \sqrt{\delta^2(\varepsilon) - 4|b_\varepsilon|(|\varphi| - \delta_1 a_\pm^3) + o(a_\pm^5)}}{2|b_\varepsilon(\lambda)|} \right]^{1/2},$$

and at $\lambda = \lambda_1$ which satisfies

$$\delta^2(\varepsilon) - 4b_\varepsilon(\lambda_1)(\varphi(a, \lambda_1) + \delta_1(\varepsilon)a^3 + o(a^5)) = 0,$$

the equation (2.6.42) has only one positive real solution

$$a_1 = \left[\frac{-\delta(\varepsilon)}{2b_\varepsilon(\lambda_1)} \right]^{1/2}, \quad (b_\varepsilon < 0),$$

and (2.6.42) has no real solutions on $\lambda < \lambda_1$; however, on $\lambda > \lambda_0^\varepsilon$ (2.6.42) has exactly one positive real solution near $a = 0$

$$a_+ = \left[\frac{\delta(\varepsilon) + \sqrt{\delta^2(\varepsilon) + 4|b_\varepsilon|(\varphi + \delta_1 a_+^3 + o(a_+^5))}}{2|b_\varepsilon(\lambda)|} \right]^{1/2}.$$

Moreover, (2.6.40) holds for $a_0 = a_+$, and (2.6.41) holds for $a_0 = a_-$. Thus Theorem 2.6.8 is proved. \square

Remark 2.6.10 From (2.6.36) we can see that if the condition (2.6.23) is relaxed as

$$(G(x + h(x, \lambda), \lambda), e_l^*) = \sum_{2 \leq i+j \leq 5} a_{ij}^l x_1^i x_2^j + o(|x|^5) \quad \text{for } l = 1, 2,$$

i.e., $G(u, \lambda)$ contains a second-order homogeneous term, then we take the numbers $b_1(\lambda)$ and $b_2(\lambda)$ as follows

$$b_1(\lambda) = \int_0^{2\pi} f_3 d\theta,$$

$$b_2(\lambda) = \int_0^{2\pi} \left[f_5 + 2f_4 \left(\int_0^\theta f_2 d\theta \right) + 2f_3 \int_0^\theta f_2 \left(\int_0^\theta f_2 d\theta \right) d\theta \right] d\theta,$$

where

$$f_k(\theta) = u_k + \sum_{i=1}^{k-2} v_{i+1} v_{k-i},$$

$$u_k = \sum_{i+j=k} a_{ij}^1 \cos^{i+1} \theta \sin^j \theta + a_{ij}^2 \cos^i \theta \sin^{j+1} \theta,$$

$$v_k = \sum_{i+j=k} a_{ij}^1 \cos^i \theta \sin^{j+1} \theta - a_{ij}^2 \cos^{i+1} \theta \sin^j \theta.$$

In this case, when $b_1(\lambda_0) = 0$, for $b_2(\lambda_0) < 0$ and $b_2(\lambda_0) > 0$, Theorems 2.6.8 and 2.6.9 hold true respectively.

2.7 Notes

The dynamic transition theory is developed in the last 20 years by the authors through a sequence of articles and two books based on Philosophy 2 with strong motivations in applications. The theory is partially presented in Ma and Wang (2005c, 2004b, 2005a,b, 2007b), and this chapter gives a comprehensive account on this theory.

As demonstrated in Chap. 1 and in many examples presented in Chaps. 3–6, the dynamic transition theory is a global theory for dynamic transitions of dissipative systems. It is a different theory from the classical bifurcation theory as demonstrated in the introduction and in Chap. 1.

It is worth noting that this book does not touch any of the large literature dealing with the instability problem of nonlinear systems coming from “essential spectrum.” We refer interested readers to, among many others, Johnson et al. (2019) and the references therein.

A combination of mathematical tools used in this chapter includes classical dynamical system, bifurcation theory, topology, and differential equations. For bifurcation theory, see among others (Krasnosel'skii, 1956; Hopf, 1942; Marsden and McCracken, 1976; Rabinowitz, 1971; Crandall and Rabinowitz, 1977; Nirenberg, 2001, 1981; Kuznetsov, 2004; Chow and Hale, 1982; Ma and Wang, 2005b, 2007b). Although not used in the book, the readers are referred to Golubitsky and Schaeffer (1985), Sattinger (1978, 1979, 1980, 1983), and Field (1996) for the bifurcation theory with symmetry. For dynamical systems theory, see, e.g., Henry (1981), Perko (1991), Guckenheimer and Holmes (1990), Wiggins (1990), Temam (1997), Hale. (1988), and Ma and Wang (2005d).

Also, in order to develop the global theory with respect to bifurcation parameters, a systematic computational approach is nowadays indispensable and many software packages such as AUTO have been developed so far. See, for example, Yadome et al. (2014) and Batiste et al. (2006).

Since the publication of the **first edition** of this book, there have been progresses on dynamic transitions of various science problems by many authors, including, e.g., the following:

- fluid problems: Dijkstra et al. (2013) on surface tension driven convection, Özer and Şengül (2016) on the Poiseuille flow, Sengul and Wang (2013), Wang and Yang (2013), and Sengul et al. (2015) on the Rayleigh Benard convection, (Wang, 2014), Sengul and Wang (2014), Han et al. (2018), and Wang and Wang (2016) on MHD convection, (Wang, 2014) and Hou and Ma (2013) on the Taylor problem, and (Wang et al., 2015a) and Luo et al. (2015a) on boundary layer separations,
- geophysical fluid dynamics and climate dynamic problems: (Dijkstra et al., 2015), Kieu et al. (2018), Şengül and Wang (2018), Özer and Şengül (2018), and Li (2017),
- random dynamical systems (Chekroun et al., 2014a,b), and
- various other areas: (Han and Hsia, 2012), Choi et al. (2015), Choi et al. (2017), Hernández and Ong (2018), Li et al. (2018), Ong (2016), Choi et al. (2015), Liu et al. (2012a, 2015), Zhang et al. (2014), Li and Ong (2016), Yarahmadian and Yari (2014), Zhang and Liu (2018), Yari (2015), Peng (2018), Li and Wang (2018), Peres Hari et al. (2013), Zhang and Luo (2013), and You et al. (2013).



Chapter 3

Equilibrium Phase Transitions in Statistical Physics

A principal objective in the study of equilibrium phase transitions is to capture the transitions from one equilibrium to another and to study the nature or order of such transitions. The study of equilibrium phase transitions presented in this book involves a combination of modeling, mathematical analysis, and physical predictions. The goals of phase transition theories are

1. to determine the definition and types of phase transitions,
 2. to derive the critical parameters,
 3. to obtain the critical exponents, and
 4. to understand the mechanism and properties of phase transitions, such as supercooled and superheated states and the Andrews critical points.
1. *Modeling* Recently we have discovered a new potential-descending principle (PDP) (Ma and Wang, 2017a). We show that PDP, together with the principle of equal probability (PEP), leads to the first and second laws of thermodynamics and to the Maxwell–Boltzmann for classical systems, the Fermi–Dirac and Bose–Einstein distributions for quantum systems. Also, PDP serves as the first principle for irreversibility. In this sense, PDP and PEP are the first principles of statistical physics.

Meanwhile, PDP offers a basic dynamical model of thermodynamical systems, given by (3.1.2) and (3.1.3). They are called the standard model of statistical physics. With the standard model, an important issue boils down to finding a better and more accurate account of the thermodynamic potentials, which are dictated by the symmetries of the underlying physical systems. Currently there are two approaches for the derivation of these potentials and Hamiltonian energy functionals. The first one is the classical approach based on the Landau’s mean field theoretical approach. The second is based only on the underlying symmetries and first principles, and is presented in Sects. 7.6, 8.3 and 8.4 in this book.

The general theoretical framework and fluctuation theory in Sects. 3.1 and 3.2 are based on the second approach, a first-principle approach, while applications in the remaining sections of this chapter is using the Landau’s mean field theory

approach. Of course, the same results can also be obtained using the second approach with the functionals in Sects. 7.6, 8.3 and 8.4; see Ma et al. (2018).

2. *Three Basic Theorems* These theorems are first derived in Ma and Wang (2017b). With the standard model, the dynamical transition theory can then be applied, and we obtain in Sect. 3.1 three basic theorems, providing a full theoretical characterization of thermodynamic phase transitions. The first theorem, Theorem 3.1.3, states that as soon as the linear instability occurs, the system always undergoes a dynamical transition, to one of the three types: continuous, catastrophic, and random. This theorem offers the detailed information for the *phase diagram* and the *critical threshold* λ_c in the control parameter $\lambda = (\lambda_1, \dots, \lambda_N) \in \mathbb{R}^N$. They are precisely determined by the following equation:

$$\beta_1(\lambda_1, \dots, \lambda_N) = 0. \quad (3.0.1)$$

Here β_1 is the first eigenvalue of the linearized operator L_λ around the basic equilibrium state \bar{u} .

The second theorem, Theorem 3.1.4, provides the corresponding relationship between the Ehrenfest classification and the dynamical classification. The Ehrenfest classification identifies experimental quantities to determine the transition types, while the dynamical classification provides a full theoretical characterization of the transition.

This theorem establishes a natural bridge between theory and experiments. With the corresponding relationship given in this theorem, we obtain a precise and easy theoretical approach to determine the transition order, the critical exponents, and the transition diagram, which are encoded in the following transition equation:

$$\frac{du_\lambda}{dt} = \beta_1(\lambda)u_\lambda + a(\lambda)u_\lambda^2 + b(\lambda)u_\lambda^3 + \text{h.o.t.} \quad (3.0.2)$$

We remark here that the detailed information on critical exponents is entirely encoded in the potential functionals, and again, our approach is not a mean-field theory.

Also, this theorem shows that there are only first-, second-, and third-order thermodynamic transitions. The third-order transition can hardly be determined by thermodynamic parameters experimentally. The dynamical transition theory, however, offers an easy and clear approach to identify the third-order transition.

The last theorem, Theorem 3.1.5, states that both catastrophic and random transitions lead to saddle-node bifurcations, and latent heat, superheated, and supercooled states always accompany the saddle-node bifurcation associated with the first-order transitions.

Theorem 3.1.5 with the transition equation (3.0.2) provides the needed information on the dynamic phase diagrams; this is important for understanding the dynamical behavior of transition states.

We further emphasize that the three theorems lead to three important diagrams: the phase diagram, the transition diagram, and the dynamical diagram. These diagrams appear only to be derivable by the dynamical transition theory presented

in this paper. In addition, our theory achieves the fourth goal of phase transition theories as stated in the beginning of this Introduction, which is hardly achievable by other existing theories.

3. *Critical Exponents and Fluctuation Theory* This fluctuation theory is based on Liu et al. (2017a). In Sect. 3.2, we derive the dynamical law of fluctuations (3.2.23). Using the dynamic transition theory, we derive theoretical values of critical exponents from the standard model of thermodynamics with or without fluctuations. Then we further show that the standard model, together with the dynamic law of fluctuations, offers correct information for the critical exponents. In a nutshell,

- *the standard model (3.1.2) and (3.1.3), together with the dynamic law of fluctuation (3.2.23), offers correct information for critical exponents; and*
- *this in return validates the standard model of thermodynamics, which is derived based on first principles.*

4. *Applications* Gas-liquid transition is a classical example, and it is understood both experimentally and theoretically that the transition before the Andrews critical point is of first order. By examining the transition behavior near and beyond the critical point, we are able to derive a first example of third-order transitions. Classically, it is hard to observe third- or higher-order transitions, since they involve higher-order derivatives of the free energy; see, for example, the experimental studies by Nishikawa and Morita (1998); see also Fisher (1964). Based on a recent article (Ma and Wang, 2011e), we derive in Sect. 3.3 a new mechanism of the Andrews critical point and obtain that the liquid-gas phase transition going beyond the Andrews point is of third order—a first example of a third-order transition.

Also, by applying dynamic transition theory and the standard model to liquid helium-3, a new prediction on the existence of a new superfluid phase C is made in Sect. 3.8; see also Ma and Wang (2008e). Of course, some of these predictions have yet to be verified experimentally, and it is hoped that our predictions will help to provide new insights into both theoretical and experimental studies for a better understanding of the underlying physical problems.

Extensive studies of the TDGL model of superconductivity can be found in the literature. See, among many others, (de Gennes, 1966; Tinkham, 1996; Aftalion and Du, 2002; Ginzburg, 2004; Du et al., 1992; Ma and Wang, 2005a; Tang and Wang, 1995). Cahn–Hilliard equations are often used to model phase transitions for binary systems as well as in other interface problems; see, e.g., Cahn and Hillard (1957), Novick-Cohen and Segel (1984). The results presented here on superconductivity and on binary systems are based mainly on Ma and Wang (2005a), Ma and Wang (2009b), Ma and Wang (2009c).

In his Nobel lecture (Ginzburg, 2004), Ginzburg expressed the hope of developing a Ginzburg–Landau model for superfluidity. The study presented here can be considered an attempt in the direction of a realization of his vision. It is based on Ma and Wang (2008a,e, 2009d). For other attempts, the reader can consult (Onuki, 2002) and the references therein. However, we would like to mention

that the Ginzburg–Landau theory with only one wave function cannot describe the typical phase transition diagrams for superfluidity, and this demonstrates the necessity for introducing new Ginzburg–Landau models.

Finally, there are many excellent general books on statistical physics, thermodynamics, and condensed-matter physics; see, among others, (Reichl, 1998; Stanley, 1971; Kleman and Laverntovich, 2007) and the references therein.

3.1 Dynamical Theory of Thermodynamic Phase Transitions

3.1.1 Standard Model of Statistical Physics

In this section, we briefly recapitulate potential-descending principle and the standard model of statistical physics, discovered in Ma and Wang (2017a). The fundamental principles of statistical physics will be addressed in detail in Chap. 7.

For a given thermodynamic system, the order parameters (state functions) $u = (u_1, \dots, u_N)$, the control parameters λ , and the thermodynamic potential F are well-defined quantities, fully describing the system. The potential is a functional of the order parameters and is used to represent the thermodynamic state of the system. There are four commonly used thermodynamic potentials: the internal energy, the Helmholtz free energy, the Gibbs free energy, and the enthalpy.

Principle 3.1.1 (Potential-Descending Principle (Ma and Wang, 2017a)) *For each thermodynamic system, there are order parameters $u = (u_1, \dots, u_N)$, control parameters λ , and the thermodynamic potential functional $F(u; \lambda)$. For a nonequilibrium state $u(t; u_0)$ of the system with initial state $u(0, u_0) = u_0$, we have the following properties:*

1. *the potential $F(u(t; u_0); \lambda)$ is decreasing:*

$$\frac{d}{dt} F(u(t; u_0); \lambda) < 0 \quad \forall t > 0;$$

2. *the order parameters $u(t; u_0)$ have a limit*

$$\lim_{t \rightarrow \infty} u(t; u_0) = \bar{u};$$

3. *there is an open and dense set O of initial data in the space of state functions, such that for any $u_0 \in O$, the corresponding \bar{u} is a minimum of F , which is called an equilibrium of the thermodynamic system:*

$$\delta F(\bar{u}; \lambda) = 0. \tag{3.1.1}$$

We have shown that PDP is a more fundamental principle than the first and second laws, provides the first principle for describing irreversibility, and leads all three

distributions: the Maxwell–Boltzmann distribution, the Fermi–Dirac distribution, and the Bose–Einstein distribution in statistical physics. Consequently, the potential-descending principle is the first principle of statistical physics; see Chap. 7 for details.

Also importantly, based on PDP, the dynamic equation of a thermodynamic system in a nonequilibrium state takes the form

$$\frac{du}{dt} = -A\delta F(u, \lambda) \quad \text{for isolated systems,} \quad (3.1.2)$$

$$\begin{cases} \frac{du}{dt} = -A\delta F(u, \lambda) + B(u, \lambda), \\ \int A\delta F(u, \lambda) \cdot B(u, \lambda) = 0 \end{cases} \quad \text{for coupled systems,} \quad (3.1.3)$$

where δ is the derivative operator, B represents coupling operators, and A is a symmetric and positive definite matrix of coefficients. We refer interested readers to (Ma and Wang, 2017a) for details.

Equations (3.1.2) and (3.1.3) are called the *standard model of statistical physics*. They offer a complete description of associated phase transitions and transformation of the system from nonequilibrium states to equilibrium states. With the above model, an important issue boils down to find a better and more accurate account of the thermodynamic potentials, which are dictated by the symmetry of the underlying physical systems. See Sects. 7.6, 8.3, and 8.4 for the systematic derivation of thermodynamic potentials and Hamiltonians based only on the underlying symmetries and first principles.

(2) Remarks on the Standard Model

The order parameters u for conventional thermodynamic systems usually include the entropy density S as a component:

$$u = (\varphi, S), \quad (3.1.4)$$

and for thermodynamic systems of condensates, the order parameters u include also the electromagnetic fields (\mathbf{E}, \mathbf{H}):

$$u = (\varphi, S, \mathbf{E}, \mathbf{H}). \quad (3.1.5)$$

Based on the fluctuation theory in the next section and the dynamic transition theory, the order parameters $(S, \mathbf{E}, \mathbf{H})$, in a sense, play a passive role for the transformation of the system from nonequilibrium states to equilibrium states, as well as for the phase transition behavior. Consequently, by writing

$$F = F(\varphi, \Phi, \lambda), \quad \text{with } \Phi = (S, \mathbf{E}, \mathbf{H}), \quad (3.1.6)$$

in view of (3.1.2) and (3.1.3), the PDP based dynamical equations are given by

$$\begin{aligned}\frac{d\varphi}{dt} &= -A \frac{\delta}{\delta\varphi} F(\varphi, \Phi, \lambda), \\ \frac{\delta}{\delta\Phi} F(\varphi, \Phi, \lambda) &= 0,\end{aligned}\tag{3.1.7}$$

which obey the PDP.

For a quantum system of condensates, the Hamiltonian energy H in general takes the form

$$H = H(\psi, \Phi, \lambda), \quad \Phi = (\mathbf{E}, \mathbf{H}).\tag{3.1.8}$$

With the same reason as before, based on the principle of Lagrangian dynamics, the dynamical equations for (3.1.8) are in the form:

$$\begin{aligned}\hbar \frac{\partial\psi}{\partial t} &= \frac{\delta}{\delta\psi^*} H(\psi, \Phi, \lambda), \\ \frac{\delta}{\delta\Phi} H(\varphi, \Phi, \lambda) &= 0.\end{aligned}\tag{3.1.9}$$

Also, based on the principle of Hamiltonian dynamics, the Hamiltonian equations are

$$\begin{aligned}\hbar \frac{\partial\psi_1}{\partial t} &= \frac{\delta}{\delta\psi_2} H(\psi, \Phi, \lambda), \\ \hbar \frac{\partial\psi_2}{\partial t} &= -\frac{\delta}{\delta\psi_1} H(\psi, \Phi, \lambda), \\ \frac{\delta}{\delta\Phi} H(\varphi, \Phi, \lambda) &= 0,\end{aligned}\tag{3.1.10}$$

where $\psi = \psi_1 + i\psi_2$.

In Ma and Wang (2017e), both models (3.1.9) and (3.1.10) describe faithfully quantum phase transitions in quantum systems of condensates. See also Chaps. 8 and 9 for details.

(3) Standard Model for Deviation Order Parameters

Phase transition is a universal phenomenon in most, if not all, natural systems. In fact, the central problem in statistical physics is on phase transitions. A phase transition refers to the transition of the system from one state to another, as the control parameter crosses certain critical threshold. In physics, a state often refers to a “stable” solution of the mathematical model, and in statistical physics, a state refers to an equilibrium state. Consequently, in thermodynamic phase transitions, we can consider a basic equilibrium state \bar{u} , and study the dynamic transitions of this state \bar{u} . If we write

$$u = \bar{u} + u',$$

then the dynamic law (3.1.2) or (3.1.3) leads to an equation for the deviation order parameter u' . After suppressing the primes, we arrive at the following *standard model for the deviation order parameter* u :

$$\frac{du}{dt} = L_\lambda u + G(u, \lambda). \quad (3.1.11)$$

Again u is the order parameter, λ is the control parameter of the system, L_λ is a linear operator, and $G(u, \lambda)$ is the nonlinear operator.

For all thermodynamic phase transitions, (3.1.11) is a dissipative dynamical system, and $u = 0$ represents the basic equilibrium state \bar{u} .

3.1.2 Ehrenfest Classification

Paul Ehrenfest was the first who gave a definition and a classification of thermodynamic phase transitions in terms of singularities, at the critical threshold, of such thermodynamic observable parameters as heat capacity, magnetic susceptibility, etc., which are observable; see Ehrenfest (1933). Phase transitions are then classified based on the behavior of the thermodynamic potentials, and were labeled by the n -th order derivative of the potential that is discontinuous at the transition.

More precisely, consider a thermodynamic system with potential functional $F(u, \lambda)$.

Definition 3.1.2 (Ehrenfest Classification Scheme) *Let λ_0 be a critical point, and u_λ are equilibrium states near $\lambda = \lambda_0$. A phase transition of the system at $\lambda = \lambda_0$ is of the n -th order if*

$$\frac{\partial^k F(u_\lambda, \lambda)}{\partial \lambda^k}$$

are continuous with respect to λ at $\lambda = \lambda_0$ for any $0 \leq k \leq n - 1$, and

$$\frac{\partial^n F(u_\lambda, \lambda)}{\partial \lambda^n}$$

is discontinuous at $\lambda = \lambda_0$.

Under this scheme, phase transitions are labeled by the lowest-order derivative of the potential that is discontinuous at the transition. First-order phase transitions exhibit a discontinuity in the first derivative of the potential with a thermodynamic variable. In principle, there could be third-, fourth-, and higher-order phase transitions. In fact, we shall prove that there are only first-, second-, and third-order transitions.

Now we recall the physical meaning of the derivatives of $F(u, \lambda)$ on λ in statistical physics. For a thermodynamic system, the first-order derivatives include

- _ the entropy: $S = -\partial F / \partial T$, and
- _ the (observable) phase volume: $V = \partial F / \partial p$ or density.

The second-order derivatives are

- _ the heat capacity in the constant pressure: $C_p = -T \frac{\partial^2 F}{\partial T^2}$,

- _ the compression coefficient $\kappa = -\frac{1}{V} \frac{\partial^2 F}{\partial p^2}$,
- _ the thermal expansion coefficient $\alpha = \frac{1}{V} \frac{\partial^2 F}{\partial T \partial p}$, and
- _ magnetic susceptibility: $\partial^2 F / \partial M \partial H$, where M is the magnetization of the material and H is magnetic field strength.

Thus, for the first-order phase transition, the discontinuity of $\partial F / \partial p$ at critical point $\lambda_0 = (T_0, p_0)$ implies the discontinuity of phase volume:

$$\Delta V = V^2 - V^1 = \frac{\partial F^+}{\partial p} - \frac{\partial F^-}{\partial p}, \quad (3.1.12)$$

and the discontinuity of $\partial F / \partial T$ implies that there is a gap between both phase entropies:

$$\Delta S = S^2 - S^1 = -\frac{\partial F^+}{\partial T} + \frac{\partial F^-}{\partial T}. \quad (3.1.13)$$

Physically, ΔS is a nonmeasurable quantity, hence, we always use latent heat $\Delta Q = T\Delta S$ to determine the first-order phase transition, and ΔQ stand for the absorbing heat for transition from phase \bar{u} to phase u_λ . In the two phase coexistence situation, ΔQ and ΔV are related by the Clausius–Clapeyron equation

$$\frac{dp}{dT} = \frac{\Delta Q}{T\Delta V}.$$

For second-order phase transitions, the variance of heat capacity (or specific heat), compression coefficient, the thermal expansion coefficient, and magnetic susceptibility at the critical value are important measurable physical quantities.

3.1.3 Three Basic Theorems of Thermodynamic Phase Transitions

This section is based on Ma and Wang (2017b), Ma et al. (2018).

(1) First Theorem

The dynamical transition theorem, Theorem 2.1.3, shows that for (3.1.11), as soon as the linear instability occurs, the system always undergoes a dynamical transition. As a direct application of this general transition theorem, we arrive immediately at the following first theorem of thermodynamic phase transitions.

Theorem 3.1.3 Consider a thermodynamic system (3.1.11). Let $\beta_1(\lambda), \beta_2(\lambda), \dots \in \mathbb{R}$ be eigenvalues¹ of the linear operator L_λ . If

¹ The linear operator $L_\lambda = -\left(\frac{\partial^2 F(\bar{u}; \lambda)}{\partial u_i \partial u_j}\right)$ is always self-adjoint, and all eigenvalues of L_λ are real.

$$\beta_i(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_0, \\ = 0 & \text{if } \lambda = \lambda_0, \\ > 0 & \text{if } \lambda > \lambda_0 \end{cases} \quad 1 \leq i \leq m, \quad (3.1.14)$$

$$\beta_j(\lambda_0) < 0 \quad m + 1 \leq j,$$

then the system (3.1.11) always undergoes a dynamic transition to one of the three types of dynamic transitions: continuous, catastrophic, and random, as λ crosses the critical threshold λ_0 .

In the physics literature, (3.1.14) is called the principle of exchange of stabilities. It is important to note that this theorem provides the detailed information for the phase diagram in the control parameter space:

$$\beta_1(\lambda) = 0. \quad (3.1.15)$$

For example, in the Cahn–Hilliard equation describing the binary system in Sect. 3.5, the control parameters are $\lambda = (T, u_0, L)$, where T is the temperature, u_0 is molar density, and L is the length scale of the system. The phase diagram is then obtained by solving (3.1.15). For fixed L , the phase diagram is given by Fig. 3.25. For a fixed molar density u_0 , the TL phase diagram was derived and given in Fig. 3.27.

(2) Second Theorem

We have at our disposal the Ehrenfest classification and our dynamic classification based on Principle 3.1.1; the latter is applicable for both equilibrium and nonequilibrium phase transitions.

The second theorem of thermodynamic phase transitions shows that there are only first-, second-, and third-order transitions, and provides the relationship between the Ehrenfest and dynamic classifications.

Theorem 3.1.4 *For the phase transition of a thermodynamic system, there exist only first-order, second-order, and third-order phase transitions. Moreover the following relations between the Ehrenfest classification and the dynamical classification hold true:*

<i>second order</i>	\iff	<i>continuous</i>
<i>first order</i>	\leftarrow	<i>catastrophic</i>
<i>either first or third order</i>	\leftarrow	<i>random</i>
<i>first order</i>	\rightarrow	<i>either catastrophic or random</i>
<i>third order</i>	\rightarrow	<i>random with asymmetric fluctuations.</i>

Proof. By Theorem 3.1.3, we only have to examine the three types of dynamical transitions. Before we proceed, we observe that the first eigenvalue $\beta_1(\lambda)$ satisfies

$$\beta_1(\lambda) = \alpha(\lambda)(\lambda - \lambda_c), \quad \alpha(\lambda_c) \neq 0. \quad (3.1.16)$$

First, it is clear from the dynamical transition theory in the previous chapter that a catastrophic dynamic transition must be of first order.

Second, consider the case where the dynamical transition at λ_c is continuous. By the dynamical transition theory in the previous chapter, we have shown that the transition states near λ_c are given by

$$u_\lambda = \begin{cases} A(\lambda - \lambda_c)^{1/m} + o((\lambda - \lambda_c)^{1/m}) & \text{for } \lambda > \lambda_c, \\ 0 & \text{for } \lambda < \lambda_c, \end{cases} \quad (3.1.17)$$

where $m \geq 2$ and $A \neq 0$. In fact, physically, $2 \leq m \leq 3$; see among others (Stanley, 1971).

Also, in view of the derivation of the equation (3.1.11) from (3.1.2) and (3.1.3), since u is the deviation order function, we let

$$\tilde{F}(u, \lambda) = F(u + \bar{u}, \lambda) - F(\bar{u}, \lambda), \quad (3.1.18)$$

where F is the potential functional of the thermodynamic system. Then near λ_c , \tilde{F} takes the following form:

$$\tilde{F}(u_\lambda, \lambda) = \frac{1}{2}\beta_1(\lambda)u_\lambda^2 + \text{h.o.t.} \quad (3.1.19)$$

Therefore, in view of (3.1.17), we have

$$\tilde{F}(u_\lambda, \lambda) = \begin{cases} A(\lambda - \lambda_c)^{1+\frac{2}{m}} + \text{h.o.t.} & \text{for } \lambda > \lambda_c, \\ 0 & \text{for } \lambda < \lambda_c. \end{cases} \quad (3.1.20)$$

Hence a continuous dynamical transition leads to a second-order transition in the Ehrenfest sense.

Third, we consider the case where the dynamical transition at λ_c is random. If the fluctuation is symmetric, then the transition must be of first-order. Otherwise, it can either be a first-order transition or the transition with the following transition states:

$$u_\lambda = \begin{cases} A(\lambda - \lambda_c) + \text{h.o.t.} & \text{for } \lambda > \lambda_c, \\ 0 & \text{for } \lambda < \lambda_c, \end{cases} \quad (3.1.21)$$

The potential functional is then written as

$$\tilde{F}(u_\lambda, \lambda) = \begin{cases} A(\lambda - \lambda_c)^3 + \text{h.o.t.} & \text{for } \lambda > \lambda_c, \\ 0 & \text{for } \lambda < \lambda_c. \end{cases} \quad (3.1.22)$$

This gives rise to a third-order transition.

Theorem 3.1.4 shows that there are only three types of dynamical transitions, which lead to only first-, second-, and third-order transitions. Therefore, there are no higher than third-order transitions in thermodynamics and the relations between the

Ehrenfest classification and the dynamical classification given in the theorem hold true as demonstrated above. The proof is complete. \square

A few remarks are now in order.

First, the Ehrenfest transition types are determined by experiments. The dynamical transition theory provides a systematic theory to easily determine the types and transition states. The above theorem establishes a natural bridge between theory and experiments.

Second, the third-order transition can hardly be determined by thermodynamic parameters experimentally. The dynamical transition theory, however, offers a simple approach to identify the third-order transition.

Third, the above theorem leads to transition diagrams, providing an important characteristic of phase transitions. In fact, the dynamical transition theory includes a systematic technical approach in determining the transition types and states, based on the central manifold reduction idea; see the previous chapter and the various applications presented in this and the remaining chapters of this book for details. In particular, for a thermodynamic phase transition, the approach gives rise to the following transition equation, which encodes all the needed information on transition types and states:

$$\frac{du_\lambda}{dt} = \beta_1(\lambda)u_\lambda + a(\lambda)u_\lambda^2 + b(\lambda)u_\lambda^3 + \text{h.o.t.} \quad (3.1.23)$$

Then the transition types are determined by the coefficients $a(\lambda)$ and $b(\lambda)$ as follows:

1. If $a(\lambda_c) \neq 0$, then the transition is random;
2. If $a(\lambda_c) = 0$, then

$$\begin{array}{lll} b(\lambda_c) > 0 & \implies & \text{catastrophic transition} \\ b(\lambda_c) < 0 & \implies & \text{continuous transition.} \end{array}$$

If $a(\lambda) \equiv 0$, the equation

$$b(\lambda) = 0$$

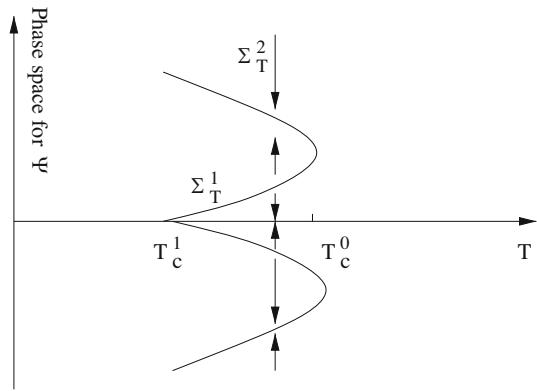
gives the transition diagram; see, e.g., Fig. 3.26 for the Cahn–Hilliard equation in Sect. 3.5.

(3) Third Theorem

For a thermodynamic system, there is no bifurcation point for given pressure control parameter with the temperature higher than the critical threshold. Also, given pressure, at very high temperature, the only stable state is the basic equilibrium state \bar{u} . Consequently, based on the topological degree theory (Ma and Wang, 2005b), both catastrophic and random transitions will give rise to saddle-node bifurcations at a temperature T_c^0 higher than the critical threshold T_c^1 . We refer readers to Fig. 3.33,

reproduced in Fig. 3.1 here for convenience, for a better understanding of superheated and supercooled states.

Fig. 3.1 Catastrophic transition: the basic state $\psi = 0$ in $T_c^1 < T < T_c^0$ is supercooled, the transition states Σ_T^2 are superheated



Also, if the transition is of the first order, the jump of the thermodynamic potential ΔF at $T_c^1 \leq T \leq T_c^0$ corresponds to the latent heat, and the two metastable states are the superheated and supercooled states of the system.

Namely, we obtain the following third theorem for thermodynamic phase transitions.

Theorem 3.1.5 *For a thermodynamic system (3.1.11), the following holds true:*

1. *both catastrophic and random transitions lead to saddle-node bifurcations; and*
2. *latent heat, superheated, and supercooled states always accompany the saddle-node bifurcations associated with the first-order transitions.*

This theorem, together with (3.1.23), offers precise information on the dynamic phase diagrams. For example, Figs. 3.33, 3.5, and 3.6 are such dynamical phase diagrams, derived from the dynamical transition theory.

3.2 Dynamic Theory of Fluctuations and Critical Exponents

This section is to address the classical question on the discrepancy between theoretical and experiment values of critical exponents, and is based on Liu et al. (2017a). In particular, we show in this section that

- *the standard model (3.1.2) and (3.1.3), together with the dynamic law of fluctuation (3.2.23), offers correct information for critical exponents; and*
- *this in return validates the standard model of thermodynamics, which is derived based on first principles.*

3.2.1 Statistical Theory of Fluctuation

(1) Classical Einstein Fluctuation Theory

Thermodynamical quantities in equilibrium are statistical average values. However, from microscopical viewpoint a thermodynamical system state always deviates its average value. The deviation is called a fluctuation of the state.

Let u represent a thermodynamical quantity, such as the temperature T , the pressure p , the volume V , and the molar density ρ . If \bar{u} is its average value, and Δu is the deviation from \bar{u} , called the fluctuation of u at \bar{u} :

$$\Delta u = u - \bar{u}.$$

Since Δu varies in both time and space, we need to know its mean-square value:

$$\overline{(\Delta u)^2} = \text{the mean-square fluctuation of } u. \quad (3.2.1)$$

The classical theory for computing (3.2.1) is the Einstein's fluctuation formula, which we introduce below for convenience.

Let W_{\max} be the maximum state number of a thermodynamical system, and W be a fluctuation of W_{\max} . The core of the Einstein fluctuation theory is to regard W as

$$W = \text{the fluctuation probability distribution.} \quad (3.2.2)$$

This formula defines the state number fluctuation W ; then for any thermodynamical quantity u , its mean-square fluctuation $\overline{(\Delta u)^2}$ can be expressed as

$$\overline{(\Delta u)^2} = \frac{\int_{-\infty}^{\infty} (\Delta u)^2 W(\Delta u) d(\Delta u)}{\int_{-\infty}^{\infty} W(\Delta u) d(\Delta u)}. \quad (3.2.3)$$

To compute the value of (3.2.3), we have to determine the distribution function $W = W(\Delta u)$. To this end, we start with the Boltzmann entropy formula:

$$\overline{S} = k \ln W_{\max},$$

where k is the Boltzmann constant, and \overline{S} is the entropy in the equilibrium state. Let S be the deviation from \overline{S} , which is written as

$$S = k \ln W.$$

Then we obtain the entropy fluctuation as follows:

$$\Delta S = S - \overline{S} = k \ln \frac{W}{W_{\max}}. \quad (3.2.4)$$

We infer from (3.2.4) the fluctuation state number W as

$$W = W_{\max} e^{\Delta S/k}. \quad (3.2.5)$$

In thermodynamics, ΔS is a function of Δu :

$$\Delta S = f(\Delta u). \quad (3.2.6)$$

Hence by inserting (3.2.6) into (3.2.5), we deduce that

$$W = W_{\max} e^{f(\Delta u)/k}. \quad (3.2.7)$$

Let $x = \Delta u$. By (3.2.7), formula (3.2.3) is rewritten as

$$\overline{(\Delta u)^2} = \frac{\int_{-\infty}^{\infty} x^2 e^{f(x)/k} dx}{\int_{-\infty}^{\infty} e^{f(x)/k} dx}. \quad (3.2.8)$$

This is the Einstein fluctuation formula. If we know the function $f(x)$ by the relation (3.2.6), then we can compute the mean square fluctuation $\overline{(\Delta u)^2}$ by (3.2.8).

(2) Revision of Classical Fluctuation Theory

In thermodynamical theory, it is difficult to determine directly the function of (3.2.6). In addition, the formula (3.2.8) is valid if and only if f must be negative definite:

$$f(x) \sim -\alpha|x|^k \quad \text{for } \alpha > 0, k > 0.$$

This is not a natural condition.

In fact, formula (3.2.8) is based on the following definition of thermal energy:

$$kT \ln W = \text{thermal energy}. \quad (3.2.9)$$

Namely $k \ln W$ = entropy. However, in view of the ensemble theory in statistical physics, the general form of a probability distribution for a thermodynamical system is as

$$\rho = \rho_0 e^{-E/kT}, \quad (3.2.10)$$

where E is the energy. Hence, instead of (3.2.9) we should think that

$$kT \ln W = \text{system energy (potential)}. \quad (3.2.11)$$

Thus, in view of (3.2.10) and (3.2.11), it is reasonable to take the fluctuation probability distribution W in the form

$$W = W_0 e^{-|\Delta F|/kT}, \quad (3.2.12)$$

where F is a thermodynamical potential, and ΔF is the fluctuation of F . With the revised formula (3.2.12), we introduce the following postulate, which serves as the foundation for the statistical theory of fluctuations.

Postulate 3.2.1 *Let u be a thermodynamical quantity, and $F(u)$ be the potential functional of u . Then, the fluctuation probability distribution W of the system is given by (3.2.12), and the mean square fluctuation of u is as follows*

$$\overline{(\Delta u)^2} = \frac{\int_{-\infty}^{\infty} x^2 e^{-|\Delta F(x)|/kT} dx}{\int_{-\infty}^{\infty} e^{-|\Delta F(x)|/kT} dx}, \quad (3.2.13)$$

where $x = \Delta u$ is the fluctuation of u .

(3) Some Examples

We use the revised fluctuation formula (3.2.13) to compute a few mean square fluctuations of some thermodynamical quantities.

Let $F = U - ST + PV$ be the Gibbs free energy of a thermodynamic system.

1. Let the entropy S be an order parameter. The second-order approximation of the Taylor expansion of ΔF at the equilibrium \bar{S} is

$$\Delta F = F(S) - F(\bar{S}) = \frac{\partial F(\bar{S})}{\partial S} \Delta S + \frac{1}{2} \frac{\partial^2 F(\bar{S})}{\partial S^2} (\Delta S)^2$$

By the PDP, Principle 3.1.1, for an equilibrium state \bar{S} , we have

$$\frac{\partial F(\bar{S})}{\partial S} = 0.$$

Hence ΔF is expressed as

$$\Delta F(x) = \frac{1}{2} \frac{\partial^2 F(\bar{S})}{\partial S^2} x^2 \quad \text{with } x = \Delta S. \quad (3.2.14)$$

Also, since $F = U - ST + PV$, U is the internal energy and $\frac{\partial U}{\partial S} = T$, we have

$$\frac{\partial^2 F}{\partial S^2} = \frac{\partial^2 U}{\partial S^2} = \frac{\partial T}{\partial S} = \frac{T}{C_V},$$

where C_V is the heat capacity. Thus, (3.2.14) becomes

$$\Delta F(x) = \frac{T}{2C_V} x^2. \quad (3.2.15)$$

Inserting (3.2.15) into (3.2.13), we deduce that

$$\overline{(\Delta S)^2} = kC_V, \quad (3.2.16)$$

where k is the Boltzmann constant.

2. Let V be the order parameter, then we have

$$\Delta F = F(V) - F(\bar{V}) = \frac{1}{2} \frac{\partial^2 F}{\partial V^2} (\Delta V)^2 = -\frac{1}{2} \frac{\partial P}{\partial V} (\Delta V)^2.$$

By the classical theory of thermodynamics,

$$\frac{\partial P}{\partial V} = -\frac{1}{V\alpha_T}, \quad \alpha_T = \text{compression coefficient.}$$

Thus we obtain that

$$\Delta F(x) = \frac{x^2}{2V\alpha_T} \quad \text{with } x = \Delta V. \quad (3.2.17)$$

Then we derive from (3.2.13) and (3.2.17) that

$$\overline{(\Delta V)^2} = kTV\alpha_T. \quad (3.2.18)$$

3. We now derive the mean square fluctuation of temperature T , by using (3.2.16). Let $T = T(S)$, then we have

$$\Delta T = \frac{\partial T}{\partial S} \Delta S \implies \overline{(\Delta T)^2} = \left(\frac{\partial T}{\partial S} \right)^2 \overline{(\Delta S)^2}.$$

By (3.2.16) and $(\partial T / \partial S)^2 = T^2 / C_V^2$, we obtain

$$\overline{(\Delta T)^2} = kT^2 / C_V. \quad (3.2.19)$$

Remark 3.2.2 The results (3.2.16), (3.2.18), and (3.2.19), derived by the revised fluctuation theory, are the same as those by the classical fluctuation theory. In fact, the revised Postulate 3.2.1 is easier to understand and to apply.

3.2.2 Dynamical Theory of Fluctuations

(1) Dynamical Law of Fluctuations

Based on the PDP, an equilibrium state u_0 of a thermodynamic system satisfies the variational equation of the potential functional F :

$$\delta F(u_0) = 0. \quad (3.2.20)$$

Also, the fluctuation of u is the deviation from u_0 :

$$\text{fluctuation } w = u - u_0. \quad (3.2.21)$$

We shall apply (3.1.2) to establish the dynamical equation for w .

It is known that a fluctuation is the process that the system deviates from the equilibrium state u_0 , resulting to a nonequilibrium state $u = u_0 + w$. Therefore, by (3.1.3) the nonequilibrium state u satisfies the following equation:

$$\frac{du}{dt} = -\delta F(u) + \tilde{f}, \quad (3.2.22)$$

where \tilde{f} is the fluctuation of the external force, and for convenience the matrix A and operator B in (3.1.3) are taken as $A = I$, $B = 0$.

Noticing that $du_0/dt = 0$, we see that (3.2.20) can be written in the form as (3.2.22). Hence we arrive at the following general form of the *dynamic law of fluctuations*:

$$\frac{dw}{dt} = -[\delta F(u_0 + w) - \delta F(u_0)] + \tilde{f}. \quad (3.2.23)$$

Furthermore, by the nonlinear operator theory, we know

$$\delta F(u) - \delta F(u_0) = \delta^2 F(u_0)w + \sum_{k \geq 2} \frac{1}{k!} \delta^{k+1} F(u_0)w^k, \quad (3.2.24)$$

where $\delta^m F(u_0)$ is the m -th order variational derivative operator. Since the fluctuation w is naturally small:

$$0 < |w| \ll 1,$$

for a noncritical state u_0 as defined after (3.2.38), we can ignore the higher-order terms with $k \geq 2$ in (3.2.24). Thus, equation (3.2.23) is written as

$$\frac{dw}{dt} = -\delta^2 F(u_0)w + \tilde{f}. \quad (3.2.25)$$

Since $u(0) = u_0$, by (3.2.21) we have the initial value condition

$$w(0) = 0. \quad (3.2.26)$$

The system (3.2.25)–(3.2.26) is also called *dynamical law of fluctuations*, near the noncritical state.

For convenience, we recall the precise definitions of the first- and second-order derivative operators of a functional F at u_0 as follows.

The first-order derivative operator $\delta F(u_0)$ is defined by

$$\langle \delta F(u_0), v \rangle = \left. \frac{d}{dt} F(u_0 + tv) \right|_{t=0}, \quad \forall v \in X, \quad (3.2.27)$$

where X is the space of state functions, and $\langle u, v \rangle$ represents the inner product between two state functions u and v .

The second-order derivative operator $\delta^2 F(u_0)w$ is defined by

$$\langle \delta^2 F(u_0)w, v \rangle = \frac{d}{dt} \langle \delta F(u_0 + tw), v \rangle \Big|_{t=0}, \quad \forall v \in X, \quad (3.2.28)$$

As an example, we compute the first- and second-order derivative operators for the following functional

$$\begin{aligned} F(u) &= \int_{\Omega} \left[\frac{1}{2} |\nabla u|^2 + \alpha u^3 + f(x)u \right] dx, \\ u \in X &= H_0^1(\Omega) = \{u, \nabla u \in L^2(\Omega) \mid u|_{\partial\Omega} = 0\}, \end{aligned} \quad (3.2.29)$$

where $\alpha \neq 0$ is a constant.

For this functional, the left-hand side of (3.2.27) is then written as

$$\langle \delta F(u_0), v \rangle = \int_{\Omega} \delta F(u_0)v dx. \quad (3.2.30)$$

The right-hand side of (3.2.27) reads

$$\begin{aligned} \frac{d}{dt} F(u_0 + tv) \Big|_{t=0} &= \frac{d}{dt} \int_{\Omega} \left[\frac{1}{2} |\nabla(u_0 + tv)|^2 + \alpha(u_0 + tv)^3 \right. \\ &\quad \left. + (u_0 + tv)f \right] dx \Big|_{t=0} \\ &= \int_{\Omega} [\nabla u_0 \cdot \nabla v + 3\alpha u_0^2 v + fv] dx \\ &= \int_{\Omega} [-\Delta u_0 + 3\alpha u_0^2 + f] v dx. \end{aligned}$$

Hence, we infer from (3.2.27) that

$$\delta F(u_0) = -\Delta u_0 + 3\alpha u_0^2 + f. \quad (3.2.31)$$

Consider (3.2.28), then we derive from (3.2.31) that

$$\begin{aligned} \langle \delta^2 F(u_0)w, v \rangle &= \frac{d}{dt} \int_{\Omega} [-\Delta(u_0 + tw) + 3\alpha(u_0 + tw)^2 + f] v dx \Big|_{t=0} \\ &= \int_{\Omega} [-\Delta w + 6\alpha u_0 w] v dx. \end{aligned}$$

It follows that

$$\delta^2 F(u_0)w = -\Delta w + 6\alpha u_0 w. \quad (3.2.32)$$

The above (3.2.31) and (3.2.32) are the first- and second-order derivative operators of (3.2.29).

(2) Noncritical Fluctuation Estimates

We know that the fluctuation equation (3.2.25) near a noncritical state is linear, and rewrite (3.2.25)–(3.2.26) in the form

$$\begin{aligned}\frac{dw}{dt} &= \mathcal{L}w + \tilde{f}, \\ w(0) &= 0,\end{aligned}\tag{3.2.33}$$

where

$$\mathcal{L} = -\delta^2 F(u_0) \quad \text{is a symmetric linear operator.}\tag{3.2.34}$$

We know that for the potential functionals F of all thermodynamic systems, the second-order derivative operator of (3.2.34) is sectorial:

$$\mathcal{L} = -\delta^2 F(u_0) \quad \text{is a sectorial operator.}\tag{3.2.35}$$

Hence, the solution of (3.2.33) can be expressed as

$$w = \int_0^t e^{(t-\tau)\mathcal{L}} \tilde{f}(\tau) d\tau.\tag{3.2.36}$$

In the following we discuss the basic properties of the solution (3.2.36).

1. *Explicit Expression of Solution w* By (3.2.34) and (3.2.35), the eigenvalue equation of \mathcal{L} is

$$\mathcal{L}\varphi_k = \beta_k \varphi_k.\tag{3.2.37}$$

The equilibrium state u_0 is called a noncritical state if the eigenvalues (3.2.37) satisfy

$$0 > \beta_1 \geq \cdots \geq \beta_k \geq \cdots \quad \text{and } \beta_k \rightarrow -\infty \text{ as } k \rightarrow \infty.\tag{3.2.38}$$

If $\beta_1 = 0$, the equilibrium state u_0 is called a critical state. We remark here that for a stable equilibrium state u_0 , $\beta_1 \leq 0$ is always true.

In this section we consider only the noncritical case where $\beta_1 < 0$. Let the corresponding eigenvectors be denoted by $\{\varphi_k \mid k = 1, 2, \dots\}$. Then $\{\varphi_k\}$ form an orthonormal basis of X , where X is the space of state functions. Therefore, $w, \tilde{f} \in X$ can be expanded as

$$w = \sum_{k=1}^{\infty} w_k \varphi_k, \quad \tilde{f} = \sum_{k=1}^{\infty} \tilde{f}_k \varphi_k.\tag{3.2.39}$$

Inserting (3.2.39) into (3.2.36), we get

$$\sum_{k=1}^{\infty} w_k \varphi_k = \sum_{k=1}^{\infty} \int_0^t \tilde{f}_k(\tau) e^{(t-\tau)\mathcal{L}} \varphi_k d\tau.\tag{3.2.40}$$

By (3.2.37) we have

$$e^{(t-\tau)\mathcal{L}}\varphi_k = e^{(t-\tau)\beta_k} \varphi_k. \quad (3.2.41)$$

It follows from (3.2.39) and (3.2.41) that

$$w_k = \int_0^t \tilde{f}_k(\tau) e^{(t-\tau)\beta_k} d\tau.$$

Thus, the solution w of (3.2.33) can be explicitly expressed as

$$w = \sum_{k=1}^{\infty} \left[\int_0^t \tilde{f}_k(\tau) e^{(t-\tau)\beta_k} d\tau \right] \varphi_k, \quad (3.2.42)$$

where \tilde{f}_k are given by $\tilde{f}_k = \langle \tilde{f}, \varphi_k \rangle$.

2. *Estimates of the Fluctuation Radius* Let the fluctuation of the external force $\tilde{f} \in X$ satisfy that

$$\begin{aligned} \sup_t \|\tilde{f}\|_{L^2}^2 &= \sum_{k=1}^{\infty} |\tilde{f}_k|_{C^0}^2 < \infty, \\ |\tilde{f}_k|_{C^0} &= \sup_t |\tilde{f}_k(t)|. \end{aligned} \quad (3.2.43)$$

Under condition (3.2.43), by (3.2.42), the solution w of (3.2.33) has the following estimate

$$\begin{aligned} \sup_t \|w\|_{L^2}^2 &= \sum_{k=1}^{\infty} \sup_t \left[\int_0^t \tilde{f}_k(\tau) e^{(t-\tau)\beta_k} d\tau \right]^2 \\ &\leq \sum_{k=1}^{\infty} |\tilde{f}_k|_{C^0}^2 \left[\sup_t \int_0^t e^{(t-\tau)\beta_k} d\tau \right]^2 \\ &\leq \sum_{k=1}^{\infty} \frac{1}{\beta_k^2} |\tilde{f}_k|_{C^0}^2 \left(\sup_t (1 - e^{t\beta_k}) \right)^2. \end{aligned} \quad (3.2.44)$$

By (3.2.38) we see that

$$0 < (1 - e^{t\beta_k}) < 1 \quad \text{and} \quad 0 < \beta_1^2 \leq \beta_2^2 \leq \dots.$$

Hence from (3.2.44) we obtain

$$\sup_t \|w\|_{L^2} \leq \frac{1}{\beta_1} \|\tilde{f}\|_{L^2}. \quad (3.2.45)$$

In mathematics, for a function u defined in a domain Ω its L^2 -norm is defined as

$$\|u\|_{L^2}^2 = \int_{\Omega} |u|^2 dx.$$

Hence, (3.2.45) represents the estimates for the maximum fluctuation radius at noncritical point u_0 , i.e., $\beta_1 < 0$.

Remark 3.2.3 Since the fluctuation (3.2.21) is at a stable equilibrium u_0 , the operator \mathcal{L} of (3.2.35) is negative definite. Therefore all eigenvalues of \mathcal{L} are non-positive. For a critical state u_0 where $\beta_1 = 0$, the estimate (3.2.45) is no longer valid, and we need to use other methods to estimate the critical fluctuation.

3.2.3 Fluctuation Effect on Critical Parameters

(1) Relation Between Fluctuations and Phase Transitions

Figure 3.2 shows the three types of dynamic phase diagrams of thermodynamic phase transitions; see the three transition Theorems 3.1.3–3.1.5 in the previous section. In this figure, λ_0 represents the critical parameter. Also, if $\lambda > \lambda_0$, the basic equilibrium state $u = 0$ is stable (the double arrows represent stability), and if $\lambda < \lambda_0$ the basic state $u = 0$ losses its stability. However, after $u = 0$ losses its stability, the main source for the system to undergo a transition from $u = 0$ to the new stable state u_λ is fluctuation.

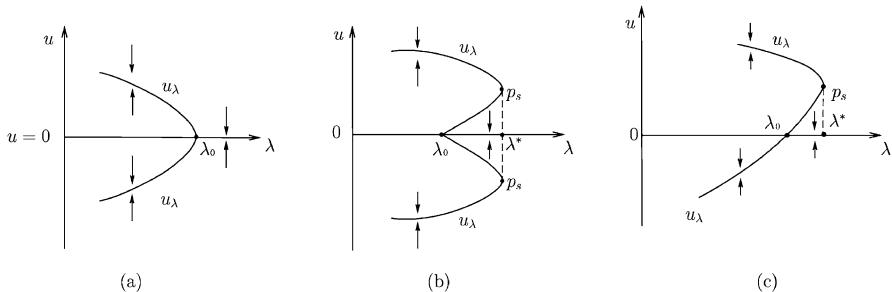


Fig. 3.2 Three types of transitions: (a) continuous, (b) catastrophic, and (c) random. The point p_s in (b) and (c) are saddle-node bifurcation points, and λ_0 is the critical parameter

Consider the initial value problem of (3.1.11):

$$\begin{aligned} \frac{du}{dt} &= L_\lambda u + G(u, \lambda), \\ u(0) &= \varphi. \end{aligned} \tag{3.2.46}$$

If $\lambda > \lambda_0$, $u = 0$ is stable. Let

$$r_s(\lambda) \text{ be the attracting radius of } u = 0. \tag{3.2.47}$$

Namely, for any initial value $\varphi \in X$, if $\|\varphi\|_X < r_s(\lambda)$, then the solution $u(t, \varphi)$ of (3.2.46) with initial value $u(0, \varphi) = \varphi$ satisfies

$$\lim_{t \rightarrow \infty} u(t, \varphi) = 0.$$

We infer from Fig. 3.2 that the attracting radius $r_s(\lambda)$ of $u = 0$ has the following property

$$r_s(\lambda) \begin{cases} > 0 & \text{for } \lambda > \lambda_0, \\ = 0 & \text{for } \lambda < \lambda_0. \end{cases} \quad (3.2.48)$$

For the cases in Fig. 3.2b, c, we have

$$r_s(\lambda) \rightarrow 0 \quad \text{as } \lambda \rightarrow \lambda_0^+. \quad (3.2.49)$$

Since $u = 0$, corresponding to u_0 before taking deviation, is an equilibrium state for all λ , theoretically the system can be in the state $u = 0$ regardless of its stability. However, a spontaneous fluctuation will cause the system deviating from $u = 0$ to a nonequilibrium state. Namely, (3.2.46) becomes

$$\begin{aligned} \frac{du}{dt} &= L_\lambda u + G(u, \lambda), \\ u(0) &= \varphi_f (\neq 0), \end{aligned} \quad (3.2.50)$$

where φ_f is the fluctuation from $u = 0$. If $u = 0$ losses its stability, then with the fluctuation φ_f the solution $u(t, \varphi_f)$ of (3.2.50) will leave $u = 0$.

Hence, we arrive at the following physical conclusion:

It is the spontaneous fluctuations and the external perturbations that cause a system to undergo a transition from a unstable equilibrium state to another stable state at a critical threshold.

(2) Fluctuation Effects on Critical Parameters

Consider the critical parameters λ_0 as in Fig. 3.2b, c. By (3.2.48) and (3.2.49), if $\lambda > \lambda_0$ with $\lambda - \lambda_0 = \varepsilon$ sufficiently small, the attracting radius $r_s(\lambda)$ of $u = 0$ is also small. Therefore, even if $u = 0$ is stable for $\lambda > \lambda_0$, with $\lambda - \lambda_0 > 0$ being small, then under a fluctuation φ_f of $u = 0$ satisfying

$$\|\varphi_f\|_X > r_s(\lambda),$$

system (3.2.50) still undergoes a transition at $\lambda > \lambda_0$. It shows that the observed critical parameter $\bar{\lambda}_0$ is larger than λ_0 :

$$\text{the observed value } \bar{\lambda}_0 > \text{ the theoretic value } \lambda_0, \quad (3.2.51)$$

due to the presence of spontaneous fluctuations.

We now need to know the relation between $\bar{\lambda}_0$ and λ_0 . Let $r_f(\lambda)$ be the fluctuation radius at λ , defined by

$$r_f(\lambda) = \sup_{\Gamma} \|\varphi_f(\lambda)\|_X, \quad (3.2.52)$$

where Γ is the set of all fluctuations of $u = 0$ at λ . It is clear that if $r_f(\lambda) > r_s(\lambda)$, (3.2.50) may undergo a transition. Hence $\bar{\lambda}_0$ is determined by the equation

$$r_f(\lambda) = r_s(\lambda). \quad (3.2.53)$$

We only consider the case where the transition is catastrophic as in Fig. 3.2b to compute $\bar{\lambda}_0$; the case where the transition is random can be addressed in the same fashion. We proceed in a few steps as follows.

1. *Expression of Attracting Radius r_s* For the catastrophic transition, the reduction equation of (3.2.46) on the central manifold is given:

$$\frac{du_1}{dt} = \beta_1(\lambda)u_1 + b_1u_1^3 + o(|u_1|^3), \quad (3.2.54)$$

where $b_1 > 0$ is a constant, called transition number, $\beta_1(\lambda)$ is as in (3.1.14) with $m = 1$, $u_1 = \langle u, \varphi_1 \rangle$, u is the solution of (3.2.46), and φ_1 is the eigenvector corresponding to β_1 . The steady-state solution of (3.2.54) is

$$u_1^\pm(\lambda) = \pm\sqrt{-\beta_1(\lambda)/b_1} + o(|\beta_1|^{1/2}) \quad \text{for } \lambda > \lambda_0. \quad (3.2.55)$$

By the bifurcating theory in Ma and Wang (2005b), the bifurcated solution u_λ^\pm of (3.1.11) near λ_0 can be expressed as

$$u_\lambda^\pm = u_1(\lambda)\varphi_1 + o(|\beta_1|^{1/2}), \quad (3.2.56)$$

where u_1 is as in (3.2.55).

Thus, it follows from (3.2.55)–(3.2.56) that the attracting radius r_s is

$$r_s = \|u_\lambda^\pm\|_{L^2} = [-\beta_1/b_1]^{1/2} \quad \text{for } \lambda > \lambda_0. \quad (3.2.57)$$

2. *Computation of Fluctuation Radius r_f* The fluctuation equation is given by (3.2.33), and its solution w is given by (3.2.36), recalled here for convenience:

$$w = \sum_{k=1}^{\infty} \left[\int_0^t \tilde{f}_k e^{(t-\tau)\beta_k} d\tau \right] \varphi_k. \quad (3.2.58)$$

where the eigenvalues $\beta_k(\lambda)$ for $\lambda > \lambda_0$ satisfy

$$0 > \beta_1 \geq \cdots \geq \beta_k \geq \cdots, \quad \lim_{k \rightarrow \infty} \beta_k = -\infty,$$

where $\beta_1(\lambda)$ is the first eigenvalue of L_λ , the same as in (3.2.57). Physically, β_1 is expressed near λ_0 as

$$\beta_1 = \alpha(\lambda_0 - \lambda) \quad \text{for } \lambda > \lambda_0. \quad (3.2.59)$$

By (3.2.58) we have

$$||\varphi_f||_X = \sup_t ||w||_{L^2} = \sup_t \left[\sum_{k=1}^{\infty} \left(\int_0^t e^{\beta_k(t-\tau)} \tilde{f}_k d\tau \right)^2 \right]^{1/2} = \tilde{f}_0. \quad (3.2.60)$$

Hence the fluctuation radius (3.2.52) is written as

$$r_f = \sup_{\Gamma} \tilde{f}_0 = \tilde{f}, \quad (3.2.61)$$

where \tilde{f}_0 is as in (3.2.60) representing the fluctuation force, and \tilde{f} is the maximum fluctuation force.

3. *Observed Critical Parameters* Then by (3.2.57), (3.2.59), and (3.2.60), we deduce from (3.2.53) that the observed critical parameter $\tilde{\lambda}_0$ is

$$\tilde{\lambda}_0 = \lambda_0 + b_1 \tilde{f}^2 / \alpha, \quad (3.2.62)$$

where b_1 is as in (3.2.54), α is as in (3.2.59), which are known parameters for a given physical system.

For random transitions corresponding to the first-order transitions, we can derive in the same fashion the observed critical parameter as

$$\tilde{\lambda}_0 = \lambda_0 + b_0 \tilde{f} / \alpha, \quad (3.2.63)$$

where $b_0 > 0$ is the coefficient in the following reduced equation of (3.1.11) on the central manifold:

$$\frac{du_1}{dt} = \beta_1(\lambda) u_1 + b_0 u_1^2 + o(|u_1|^2).$$

3.2.4 Theoretical Critical Exponents of Standard Models

(1) Definition of Critical Exponents

Critical exponent theory is an important topic in the field of thermodynamic phase transitions. The concept is only suitable for second-order phase transitions.

Let u be an order parameter, and λ be a control parameter. It is known that u is a function of λ , i.e., $u = u(\lambda)$. For a second-order phase transition, at the critical threshold λ_c , we have

$$\lim_{\lambda \rightarrow \lambda_c} u(\lambda) = u(\lambda_c).$$

This continuity implies that u can be expanded as

$$u(\lambda) = u(\lambda_c) + A|\lambda - \lambda_c|^\theta + o(|\lambda - \lambda_c|^\theta),$$

where A is a parameter independent of λ . The exponent θ is called a critical exponent of the thermodynamic phase transition.

Usually there are six types of critical exponents, denoted by

$$\beta, \delta, \alpha, \gamma, \eta, \nu, \quad (3.2.64)$$

whose precise definitions are given as follows:

1. *β -Exponent (Order Parameter Exponent)* Let u be an order parameter, and the control parameter be the temperature T . Near the critical temperature T_c , $u(T)$ can be expanded as

$$u(T) = A(T_c - T)^\beta + o((T_c - T)^\beta) \quad \text{for } T < T_c. \quad (3.2.65)$$

Here we have used $u(T_c) = 0$. The exponent β in (3.2.65) is the order-parameter exponent as listed in (3.2.64).

2. *δ -Exponent (External Force Exponent)* Let the control parameter be an external force f . Then the order parameter $u = u(f)$ can be expressed near (f_c, T_c) as

$$|u(f) - u(f_c)| = A|f - f_c|^\delta + \text{h.o.t.} \quad \text{with } u(f_c) = 0. \quad (3.2.66)$$

Here h.o.t. refers to the higher-order terms. Conventionally (3.2.66) is rewritten in the form

$$|f - f_c| \sim |u(f) - u(f_c)|^\delta \quad \text{at } T = T_c. \quad (3.2.67)$$

The exponent δ in (3.2.67) is the external force exponent.

3. *α -Exponent (Heat Capacity Exponent)* The heat capacity C is defined by

$$C = -T \frac{\partial^2 F}{\partial T^2} = T \frac{\partial S}{\partial T}, \quad (3.2.68)$$

where F is the Gibbs free energy functional and S is the entropy. Let the heat capacity $C = C(T)$ be expressed near T_c as

$$|C(T) - C(T_c)| = A|T_c - T|^{-\alpha} \quad \text{at } p = p_c. \quad (3.2.69)$$

Then the exponent α in (3.2.69) is the heat capacity exponent.

4. *γ -Exponent (Responding Parameter Exponent)* The compression coefficient κ is given by

$$\kappa = -\frac{\partial^2 F}{\partial \rho^2} = \frac{1}{\rho} \frac{\partial \rho}{\partial p},$$

where ρ is the density.

The magnetic susceptibility χ is defined as

$$\chi = \left. \frac{\partial M}{\partial H} \right|_{H=0},$$

where M is the magnetization, and H is the applied magnetic field.

As functions of temperature T , the exponents γ of κ and χ at T_c , expressed as

$$\begin{aligned}\Delta\kappa(T_c) &= A(T - T_c)^{-\gamma} && \text{at } p = p_c, \\ \Delta\chi(T_c) &= A(T - T_c)^{-\gamma} && \text{at } H = 0,\end{aligned}\quad (3.2.70)$$

are the corresponding parameter exponents.

5. η Exponent (*Coherence Exponent*) The density fluctuation coherence function is defined as

$$R(r) = \overline{\Delta\rho(r)\Delta\rho(0)},$$

where $\Delta\rho(r)$ represents the density fluctuation at r . At the critical temperature T_c , the power exponent on r of $R(r)$ can be expressed as

$$R(r) = Ar^{-n+2-\eta} \quad \text{with } n = \text{the spatial dimension.}$$

The number η is the coherence exponent.

6. ν Exponent (*Coherence Length Exponent*) The coherence length $\xi(T)$ is a function of the temperature as

$$R(r) = \frac{kT}{4\pi b} \frac{1}{r} e^{-r/\xi(T)},$$

where $R(r)$ is the density fluctuation coherence function. At T_c the expression of $\xi(T)$ is as

$$\xi(T) = A(T_c - T)^{-\nu},$$

which defines the coherence length exponent ν .

(2) Critical Exponents of Thermodynamic Potentials

In this section, we compute the exponents β , δ , α , and γ based on the standard model of thermodynamical systems (3.1.2) and (3.1.3). The standard model for the deviation equation from an equilibrium is given by (3.1.11).

1. *Order Parameter Exponent* β We start with the standard model (3.1.11) with the temperature T as the control parameter, i.e.

$$\frac{du}{dt} = L_T u + G(u, T), \quad (3.2.71)$$

where $G(u, T)$ is the higher-order term of u :

$$G(u, T) = o(\|u\|). \quad (3.2.72)$$

Let $\beta_1(T)$ be the first eigenvalue of L_T , and be simple. Based on the dynamic transition theory established in the previous section, for continuous type of transition, the transition solution of (3.2.71)–(3.2.72) near T_c can be written as

$$u(T) - u_c = x(T)\varphi_1 + o(|x|), \quad (3.2.73)$$

where φ_1 is the first eigenvector corresponding to β_1 , and $x(T)$ satisfies the following algebraic equation

$$\begin{aligned}\beta_1(T)x - bx^3 + o(|x|^3) &= 0, \\ \beta_1 &= \alpha(T_c - T),\end{aligned}\quad (3.2.74)$$

where $\alpha, b > 0$ are constants. The solution of (3.2.74) can be written as

$$x(T) = \begin{cases} 0 & \text{for } T_c < T, \\ \pm \sqrt{\frac{\alpha}{b}}(T_c - T)^{1/2} + o(|T_c - T|^{1/2}) & \text{for } T_c > T. \end{cases}$$

Then the solution (3.2.73) is

$$u(T) - u_c = \begin{cases} 0, & T_c \leq T, \\ \pm \sqrt{\frac{\alpha}{b}}(T_c - T)^{1/2}\varphi_1 + o(|T_c - T|^{1/2}), & T_c > T. \end{cases}\quad (3.2.75)$$

In view of definition (3.2.65) for the β -exponent, we derive from (3.2.75) the order-parameter exponent as

$$\beta = \frac{1}{2}. \quad (3.2.76)$$

2. *External Force Exponent* δ For the thermodynamic system with a generalized force f , its dynamical equation is in the form

$$\frac{du}{dt} = L_\lambda u + G(u, \lambda) - f \quad \text{with } \lambda = T. \quad (3.2.77)$$

Let $\lambda_c = T_c$ and f_c be the critical thresholds of (3.2.77), and u_c be the critical equilibrium state:

$$L_{\lambda_c} u_c + G(u_c, \lambda_c) = f_c. \quad (3.2.78)$$

Also, let u_f be the equilibrium solution of (3.2.77) satisfying

$$L_{\lambda_c} u_f + G(u_f, \lambda_c) = f. \quad (3.2.79)$$

We infer then from (3.2.78) and (3.2.79) that

$$L_{\lambda_c}(u_f - u_c) + [G(u_f, \lambda_c) - G(u_c, \lambda_c)] = f - f_c. \quad (3.2.80)$$

Because (T_c, f_c) is a critical point for a continuous transition, in mathematics it implies that (3.2.80) can be expressed in the following form

$$\mathcal{L}\tilde{u} - b\tilde{u}^3 + o(|\tilde{u}|^3) = \tilde{f}, \quad (b > 0), \quad (3.2.81)$$

where $\mathcal{L} = L_{\lambda_c} + DG(u_c, \lambda_c)$, and the first eigenvalue β_1 of \mathcal{L} satisfies $\beta_1(\lambda_c) = 0$, and $\tilde{u} = u_f - u_c$, $\tilde{f} = f - f_c$.

In thermodynamics, the δ -exponent is defined only for a homogeneous system. In this case $\mathcal{L} = \beta_1 = 0$ at $\lambda = \lambda_c$. Hence, (3.2.81) becomes

$$|f_c - f| = b|u_f - u_c|^3 + o(|u_f - u_c|^3). \quad (3.2.82)$$

By (3.2.67) we obtain the external force exponent δ as

$$\delta = 3. \quad (3.2.83)$$

3. Heat Capacity Exponent α

By the heat capacity formula (3.2.68), $C(T)$ is

$$C(T) = T \frac{\partial S}{\partial T}, \quad (3.2.84)$$

where S is the entropy. Let $F(u, S, T)$ be the potential functional, (u, S) be the order parameter, and the temperature T be the control parameter.

By the standard dynamical model of thermodynamic system established in Ma and Wang (2017a), Liu et al. (2019), the equation for $F(u, S, T)$ are in the form

$$\begin{aligned} \frac{du}{dt} &= -\frac{\delta}{\delta u} F(u, S, T), \\ \frac{\delta}{\delta S} F(u, S, T) &= 0. \end{aligned} \quad (3.2.85)$$

In addition, for conventional thermodynamic systems, the general form of F is as follows

$$F = F_1(u, T) + aT(-S^2 + b_1Su^2 + b_2S),$$

where a, b_1, b_2 are system parameters. It follows from the second equation of (3.2.85) that near $T = T_c$ the entropy S is

$$S = \gamma_1 u^2 + \gamma_0, \quad (3.2.86)$$

where γ_0, γ_1 are constants, and $u(T)$ is as in (3.2.75). Hence S in (3.2.86) can be written as

$$S = \begin{cases} 0 & \text{for } T_c \leq T, \\ A(T_c - T)\varphi_1 + A_0 + o(|T_c - T|) & \text{for } T_c < T. \end{cases}$$

Then we derive from (3.2.84) that

$$|\Delta C|_{T_c} = |C(T_c) - C(T)|_{t \rightarrow T_c} = \begin{cases} 0 & \text{as } T \rightarrow T_c^+, \\ A & \text{as } T \rightarrow T_c^-. \end{cases}$$

By (3.2.69) we deduce that

$$\alpha = 0. \quad (3.2.87)$$

4. *Responding parameter exponent γ .* The compression coefficient κ and the magnetic susceptibility χ are responding parameters. In general a responding parameter, denoted by θ , can be expressed as

$$\theta = B(u) \frac{\partial u}{\partial f}, \quad B(u) = 1 \text{ or } \frac{1}{u}, \quad (3.2.88)$$

where u is the order parameter, and f is the generalized force.

Near (T_c, f_c) the relation between u and f is given by (3.2.82). Namely

$$b\tilde{u}^3 = \tilde{f} \quad \text{where } \tilde{u} = u_f(T) - u_c, \quad \tilde{f} = f - f_c. \quad (3.2.89)$$

It follows from (3.2.88) to (3.2.89) that

$$\Delta\theta = B(u_c) \frac{\partial \tilde{u}}{\partial \tilde{f}} = \frac{B(u_c)}{3b\tilde{u}^2}. \quad (3.2.90)$$

By (3.2.75), $\Delta\theta$ is as

$$\Delta\theta = \frac{B(u_c)}{3\alpha||\varphi_1||} \frac{1}{T_c - T} \quad \text{for } T < T_c.$$

Hence we deduce from (3.2.70) that

$$\gamma = 1. \quad (3.2.91)$$

(3) First Theorem of Critical Exponents

The four exponents β , δ , α , and γ given by (3.2.76), (3.2.83), (3.2.87), and (3.2.91) are deduced using the standard model of thermodynamic systems developed in Ma and Wang (2017a), Liu et al. (2019). Although the values of these four exponents are the same as those using the Landau mean field theory, the results by the standard model are derived with a different theoretical foundation and a different viewpoint.

Furthermore there is a discrepancy between the theoretical exponents β , δ , α , γ , and their experimental values. However, the critical fluctuation effect introduced in the next section shows that the discrepancy is *due entirely to* the spontaneous fluctuation. In other words, for the critical exponent theory there are two groups of data:

$$\begin{aligned} &\text{theoretical critical exponents of standard model, and} \\ &\text{fluctuation critical exponents of standard model.} \end{aligned} \quad (3.2.92)$$

Namely, the standard model offers correct theoretical exponents under the no-fluctuation assumption; and for reality, the standard model with fluctuation provides their correct values in agreement with experimental data. Therefore in a nutshell,

- the standard model offers the correct information for critical exponents; and
- this in return validates the standard model of thermodynamics, which is derived based on first principles.

Consider the two groups of critical exponents in (3.2.92), which obey the same relations, called the scaling laws. Write (3.2.89)

$$\Delta f \sim (\Delta u)^\delta,$$

and (3.2.90) is in the form

$$\Delta\theta \sim (\Delta u)^{-(\delta-1)} \quad \text{with} \quad \Delta u \sim (T_c - T)^\beta.$$

Thus we get

$$\Delta\theta \sim (T_c - T)^{-\beta(\delta-1)}.$$

By the definition of the γ -exponent, we have

$$\gamma = \beta(\delta - 1). \quad (3.2.93)$$

The relation (3.2.93) is called the Widom scaling law.

In summary, by using the standard model we derive the following theorem, which we call the first theorem of critical exponents, on theoretical critical exponents.

Theorem 3.2.4 (First Theorem of Critical Exponents) *For a second-order phase transition, near the critical point, there are two groups of critical exponents as in (3.2.92): the theoretical and the fluctuation critical exponents, which satisfy certain common scaling laws as given in (3.2.93). Moreover, the theoretical critical exponents β , δ , α , γ are given by*

$$\beta = \frac{1}{2}, \quad \delta = 3, \quad \alpha = 0, \quad \gamma = 1. \quad (3.2.94)$$

3.2.5 Critical Exponents with Fluctuation Influences

(1) Fluctuation β -Exponent

Theorem 3.2.4 states that there are two groups of critical exponents, which can be derived by the standard model of thermodynamics. The exponents without fluctuation influences are given by (3.2.94), and the exponents with fluctuation influences will be deduced as follows.

Let u be an order parameter of a system, which has a second-order phase transition at T_c . By the dynamical phase transition theory in the previous section, the equation governing the second-order phase transition must be in the form

$$\frac{du}{dt} = L_\lambda u - bu^3 + \text{h.o.t.}, \quad (3.2.95)$$

where $b > 0$, and the control parameter λ is the temperature T . The theoretical exponent $\beta = \frac{1}{2}$ is deduced from (3.2.95). However, in experiments the real β -exponent is detected under the presence of spontaneous fluctuations. Hence, the dynamical equation governing a second-order phase transition with spontaneous fluctuations is given by

$$\frac{du}{dt} = L_\lambda u - bu^3 + \tilde{f}(t), \quad (3.2.96)$$

where \tilde{f} is a fluctuation force. Let $\{\beta_k(\lambda)\}$ be the eigenvalues of L_λ and $\{\varphi_k\}$ be the eigenvectors. Then the solution u of (3.2.96) near the basic state $u = 0$ can be expressed as

$$u = \sum u_k \varphi_k, \quad (3.2.97)$$

$$u_k = \int_0^t e^{\beta_k(t-\tau)} \tilde{f}_k d\tau - b \int_0^t e^{\beta_k(t-\tau)} \langle u^3, \varphi_k \rangle d\tau, \quad (3.2.98)$$

where \tilde{f}_k is the k -th component of \tilde{f} :

$$\tilde{f}_k = \langle \tilde{f}, \varphi_k \rangle.$$

It is known that at the critical threshold T_c ,

$$\beta_1 \simeq 0, \quad \beta_1 > 0, \quad \beta_j < 0, \quad \forall j \geq 2,$$

and $u \simeq 0$, $\tilde{f} \simeq 0$. Hence (3.2.98) can be approximatively written as

$$u_1 = \int_0^t e^{\beta_1(t-\tau)} \tilde{f}_1(\tau) d\tau - bu_1^3 \langle \varphi_1^3, \varphi_1 \rangle \int_0^t e^{\beta_1(t-\tau)} d\tau, \quad (3.2.99)$$

$$u_j = \int_0^t e^{\beta_j(t-\tau)} \tilde{f}_j d\tau, \quad j \geq 2. \quad (3.2.100)$$

Notice that

$$\frac{u_1}{\int_0^t e^{\beta_1(t-\tau)} d\tau} = \frac{\beta_1 u_1}{e^{\beta_1 t} - 1} \rightarrow 0 \quad \text{as } t \rightarrow \infty.$$

We infer from (3.2.99) that as $t \rightarrow \infty$, u_1 in (3.2.99) can be rewritten as

$$u_1^3 = \frac{1}{b \langle \varphi_1^3, \varphi_1 \rangle} \bar{f} \beta_1, \quad (3.2.101)$$

where $\beta_1 = \alpha(T_c - T)$ is as in (3.2.74),

$$\bar{f} = \int_0^\infty e^{-\beta_1 \tau} \tilde{f}_1(\tau) d\tau, \quad \langle \varphi_1^3, \varphi_1 \rangle = \int_{\Omega} \varphi_1^4 dx.$$

Here \bar{f} can be viewed as the weighted average of the fluctuation $\tilde{f}_1(t)$ with an exponential decay weight.

By (3.2.101) near T_c we have

$$u_1 \simeq A(T_c - T)^{1/3}, \quad A = [\alpha \bar{f} / b \langle \varphi_1^3 \rangle]^{1/3}. \quad (3.2.102)$$

Therefore, by (3.2.100) and (3.2.102), the long-time (i.e., $t \rightarrow \infty$) observed equilibrium solution u in (3.2.97) is in the form

$$u = u_c + A(T_c - T)^{1/3} \varphi_1, \quad (3.2.103)$$

where u_c is the mean equilibrium state generated by \bar{f} independent of $(T_c - T)$, expressed as

$$u_c = \sum_{j=2}^{\infty} \bar{f}_j \varphi_j, \quad \bar{f}_j = \lim_{t \rightarrow \infty} \int_0^{\infty} e^{\beta_j(t-\tau)} \tilde{f}_j d\tau.$$

Hence by (3.2.103) the observed β -exponent is derived by

$$|u - u_c| = A |\varphi_1| |(T_c - T)|^{1/3}.$$

Namely

$$\beta = \frac{1}{3}. \quad (3.2.104)$$

(2) Fluctuation δ -Exponent

The equation to determine the theoretical δ -exponent is (3.2.81), or equivalently

$$|u_f - u_c|^3 = \frac{1}{b} |f - f_c|, \quad (3.2.105)$$

where the difference $f - f_c$ of the external force has no fluctuation influence.

In experiments, the operating quantity is $f - f_c$, and the δ -exponent detected satisfies the following relation

$$|u_f - u_c|^{\delta} = A_0 |f - f_c|. \quad (3.2.106)$$

However, the quantity influencing δ in reality is the mean square root of the fluctuation $f - f_c$. Namely (3.2.105) should be replaced by the following equation with the fluctuation effect:

$$|u_f - u_c|^3 = \frac{1}{b} \sqrt{(\Delta(f - f_c))^2}. \quad (3.2.107)$$

Let the relation between $|f - f_c|$ and its fluctuation mean square root be

$$\left[\overline{(\Delta(f - f_c))^2} \right]^{1/2} = A |f - f_c|^\mu, \quad (3.2.108)$$

where μ is a to-be-determined constant. Inserting (3.2.108) into (3.2.109) we deduce that

$$|u_f - u_c|^{3/\mu} = A_0 |f - f_c|.$$

In comparison with (3.2.106) we derive

$$\delta = 3/\mu, \quad (3.2.109)$$

which is the fluctuating δ -exponent.

We now need to compute the value μ in (3.2.109) by using the statistical theory of fluctuations. It suffices to consider $f = p$ the pressure. In this case the fluctuation of pressure is

$$\Delta(p - p_c) = \frac{\partial p}{\partial V} \Delta V - \frac{\partial p_c}{\partial V_c} \Delta V_c. \quad (3.2.110)$$

Taking square on both sides of (3.2.110) and then taking an average of them, we get

$$\begin{aligned} \overline{(\Delta p - \Delta p_c)^2} &= \left(\frac{\partial p}{\partial V} \right)^2 \overline{(\Delta V)^2} + \left(\frac{\partial p_c}{\partial V_c} \right)^2 \overline{(\Delta V_c)^2} \\ &\quad - 2 \left(\frac{\partial p}{\partial V} \right) \left(\frac{\partial p_c}{\partial V_c} \right) \overline{(\Delta V)(\Delta V_c)}. \end{aligned}$$

By (3.2.18), at $T = T_c$ we can write

$$\overline{(\Delta V)^2} = -kT_c \frac{\partial V}{\partial p}, \quad \frac{\partial p}{\partial V} = -\frac{1}{\alpha_T V}.$$

Hence we have

$$\begin{aligned} &\overline{(\Delta p - \Delta p_c)^2} \\ &= \left(\frac{\partial p}{\partial V} \right) \left(\frac{\partial p_c}{\partial V_c} \right) \left[(-kT_c) \frac{\partial V_c}{\partial p_c} + (-kT_c) \frac{\partial V}{\partial p} - 2 \overline{(\Delta V)(\Delta V_c)} \right] \\ &= \left(\frac{\partial p}{\partial V} \right) \left(\frac{\partial p_c}{\partial V_c} \right) \left[\overline{(\Delta V_c)^2} + \overline{(\Delta V)^2} - 2 \overline{(\Delta V)(\Delta V_c)} \right]. \end{aligned}$$

Namely

$$\overline{(\Delta p - \Delta p_c)^2} = \frac{1}{\alpha_T(p)\alpha_T(p_c)VV_c} \overline{\Delta(V - V_c)^2} \quad (3.2.111)$$

Again by (3.2.18):

$$\overline{(\Delta V)^2} = kT_c \alpha_T(p) V,$$

it is reasonable to take

$$\overline{(\Delta(V - V_c))^2} = kT_c (\alpha_T(p) - \alpha_T(p_c))(V - V_c).$$

Putting it into (3.2.111), we obtain

$$\overline{(\Delta(p - p_c))^2} = A_0 |\alpha_T(p) - \alpha_T(p_c)| |p - p_c|, \quad (3.2.112)$$

where we used the equation of state $p = T_c V$, and

$$A_0 = \frac{k}{\alpha_T(p)\alpha_T(p_c)VV_c}.$$

Let

$$|\alpha_T(p) - \alpha_T(p_c)| = A_1 |p - p_c|^{\gamma_p}, \quad (3.2.113)$$

Then it follows from (3.2.112) and (3.2.113) that

$$\left(\overline{(\Delta(p - p_c))^2} \right)^{1/2} = A |p - p_c|^{(1+\gamma_p)/2}.$$

In view of (3.2.108), for $f = p$ we derive that

$$\mu = \frac{1 + \gamma_p}{2}, \quad (3.2.114)$$

where γ_p is the responding parameter exponent defined by (3.2.113), and α_T is the isothermal compression coefficient, which is a new critical exponent. Thus, the fluctuating δ -exponent (3.2.109) becomes

$$\delta = \frac{3}{\mu} = \frac{6}{1 + \gamma_p}. \quad (3.2.115)$$

For a second-order transition,

$$\alpha_T(p) \rightarrow \alpha_T(p_c) \quad \text{as } p \rightarrow p_c.$$

It implies that $\gamma_p > 0$. In addition, physically the mean-square fluctuation $\overline{(\Delta(f - f_c))^2}$ must be larger than $|f - f_c|^2$ for f near f_c , and consequently $\mu < 1$ in (3.2.108). Hence by (3.2.114) we have

$$\frac{1}{2} < \mu < 1. \quad (3.2.116)$$

It follows from (3.2.115) and (3.2.116) that

$$3 < \delta < 6. \quad (3.2.117)$$

This is the estimate of fluctuation δ -exponent deduced from the standard model and the statistical fluctuation theory. The experimental values of the δ -exponent are in good agreement with (3.2.117); see Table 3.1.

(3) Second Theorem of Critical Exponents

From the standard model we deduce two scaling laws (3.2.93) and (3.2.115), rewritten here:

$$\begin{aligned} \text{Widom scaling law : } & \gamma = \beta(\delta - 1), \\ \delta - \gamma_p \text{ scaling law : } & \delta(1 + \gamma_p) = 6. \end{aligned} \quad (3.2.118)$$

In statistical physics, for α , β , γ there is the Rushbrooke scaling law given by

$$\text{Rushbrooke scaling law : } \alpha + 2\beta + \gamma = 2. \quad (3.2.119)$$

In view of the scaling laws (3.2.118) and (3.2.119), by the fluctuating exponents β and δ in (3.2.104) and (3.2.117), we obtain

$$1 \leq \gamma < \frac{5}{3}, \quad 0 \leq \alpha < \frac{2}{3},$$

where $\alpha \geq 0$ is based on the physical fact that for a second-order transition, the increment of the heat capacity $|\Delta C|$ must be finite at T_c . Also, $1 \leq \gamma$ is due to the fact that the value of γ is always bigger than or equal to the theoretical value without fluctuations.

We have now proved the following second theorem of critical exponents.

Theorem 3.2.5 (Second Theorem of Critical Exponents) *The two groups of critical exponents in (3.2.92) satisfy the common scaling laws in (3.2.118) and (3.2.119). Moreover, the fluctuating critical exponents β , δ , α , γ derived from the standard model and the scaling laws in (3.2.118) and (3.2.119) are given by*

$$\beta = \frac{1}{3}, \quad 3 < \delta < 6, \quad 0 < \alpha < \frac{2}{3}, \quad 1 \leq \gamma < \frac{5}{3}. \quad (3.2.120)$$

For comparison, in Table 3.1, we listed the three groups of exponent data for different thermodynamic systems: (1) experimental exponents, (2) theoretical exponents without taking into consideration of fluctuations, and (3) theoretical exponents using the standard model with fluctuations. This table shows clearly the strong agreement of the results using the standard model of thermodynamics with fluctuations.

Table 3.1 Experimental and theoretic data of critical exponents

Exponent	Magnetic system	PVT system	Binary system	Without fluctuation	With fluctuation
β	0.30–0.36	0.32–0.35	0.30–0.34	1/2	1/3
δ	4.2–4.8	4.6–5.0	4.0–5.0	3	3.0–6.0
α	0.0–0.2	0.1–0.2	0.05–0.15	0	0–2/3
γ	1.2–1.4	1.2–1.3	1.2–1.4	1	1–5/3

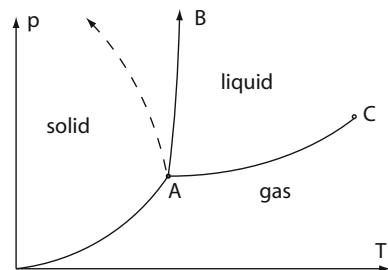
3.3 Third-Order Gas-Liquid Phase Transition

3.3.1 Introduction

A physical-vapor transport (PVT) system is a system composed of one type of molecules, and the interaction between molecules is governed by the Van der Waals law. The molecules generally have a repulsive core and a short-range attraction region outside the core. Such systems have a number of phases: gas, liquid, and solid, and a solid can appear in a few phases. The most typical example of a PVT system is water. In general, the phase transitions of PVT systems mainly refer to the gas-liquid, gas-solid, and liquid-solid phase transitions. These transitions are all first order in the Ehrenfest sense (i.e., discontinuous) and are accompanied by a latent heat and a change in density.

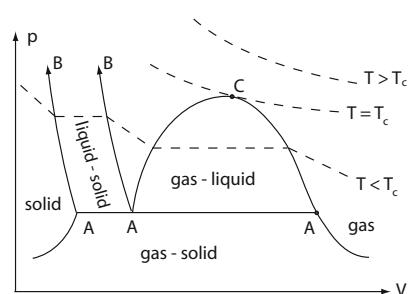
A PT -phase diagram of a typical PVT system is schematically illustrated in Fig. 3.3, where point A is the triple point at which the gas, liquid, and solid phases coexist. Point C is the Andrews critical point at which the gas-liquid coexistence curve terminates (Reichl, 1998; Stanley, 1971).

Fig. 3.3 Coexistence curves of a typical PVT system: A is the triple point, C is the critical point, and the dashed curve is a melting line with negative slope



The phase diagram in the PV -plane is given in Fig. 3.4, where the dashed curves represent lines of constant temperature. In the region of coexistence of phases, the isotherms (dashed lines) are always flat, indicating that in these regions the change in volume (density) occurs for constant pressure p and temperature T .

Fig. 3.4 The PV -phase diagram with the dashed line representing isotherms



Classical view on the termination of the gas-liquid coexistence curve at the critical point is that the system can go continuously from a gaseous state to a liquid state

without ever meeting an observable phase transition. Note that the critical point in the liquid-solid coexistence curve has never been observed.

The main objective of this section is to study the nature of the Andrews critical point in the gas-liquid transition in a physical-vapor transport (PVT) system based on the authors' recent papers (Ma and Wang, 2008b, 2011e). In particular, it is shown that (1) the gas-liquid coexistence curve can be extended beyond the Andrews critical point, and (2) the transition is *first order* before the critical point, *second order* at the critical point, and *third order* beyond the Andrews critical point. This clearly explains why it is hard to observe the gas-liquid phase transition beyond the Andrews critical point. The derived theoretical results are in agreement with the experimental discoveries in Nishikawa and Morita (1998) and the references therein. In their article, the density fluctuations of supercritical CO₂ and CF₃H for various thermodynamic states were measured by small-angle X-ray scattering and also calculated using the empirical state equations. For each sample, the ridge of the density fluctuations forms along the extension of the coexistence curve of gas-liquid in the P-T phase diagram and seems to be the locus of a higher-order phase transition. Here in this section, we shall explicitly calculate the jump of the third-order derivative with respect to the temperature of the Gibbs energy at the critical temperature, and give precise information on the locus of the coexistence curve beyond the Andrews critical point. Also, the derived second-order transition at the critical point is consistent with the result obtained in Fisher (1964).

Finally, the dynamical transition analysis in this section suggests that there are some mathematical states which do not exist in reality. Hence the system appears to avoid the fluctuations in the basin of attraction in these unrealistic states. This leads us to introduce asymmetry principle of fluctuations, and the preferred transition mechanism.

3.3.2 Time-Dependent Models for PVT Systems

Van der Waals Equation and Gibbs Energy The classical and the simplest equation of state which can exhibit many of the essential features of the gas-liquid phase transition is the Van der Waals equation:

$$v^3 - \left(b + \frac{RT}{p} \right) v^2 + \frac{a}{p} v - \frac{ab}{p} = 0, \quad (3.3.1)$$

where v is the molar volume, p is the pressure, T is the temperature, R is the universal gas constant, b is the revised constant of inherent volume, and a is the revised constant of attractive force between molecules. If we adopt the molar density $\rho = 1/v$ to replace v in (3.3.1), then the Van der Waals equation becomes

$$-(bp + RT)\rho + a\rho^2 - ab\rho^3 + p = 0. \quad (3.3.2)$$

Now, we shall apply thermodynamic potentials to investigate the phase transitions of PVT systems, and we shall see later that the Van der Waals equation can be derived as a Euler-Lagrange equation for the minimizers of the Gibbs free energy for PVT systems at gaseous states.

Consider an isothermal-isopiestic process. The thermodynamic potential is taken to be the Gibbs free energy. In this case, the order parameters are the molar density ρ and the entropy density S , and the control parameters are the pressure p and temperature T . The general form of the Gibbs free energy for PVT systems is given as

$$G(\rho, S, T, p) = \int_{\Omega} \left[\frac{1}{2} \mu_1 |\nabla \rho|^2 + \frac{1}{2} \mu_2 |\nabla S|^2 + g(\rho, S, T, p) - ST - \alpha(\rho, T, p)p \right] dx, \quad (3.3.3)$$

where g and α are differentiable with respect to ρ and S , $\Omega \subset \mathbb{R}^3$ is the container, and αp is the mechanical coupling term in the Gibbs free energy given by

$$\alpha(\rho, T, p)p = \rho p - \frac{1}{2} b \rho^2 p, \quad (3.3.4)$$

where $b = b(T, p)$ depends continuously on T and p .

Based on both the physical and mathematical considerations, we take the Taylor expansion of $g(\rho, S, T, p)$ on ρ and S as follows

$$g = \frac{1}{2} \alpha_1 \rho^2 + \frac{1}{2} \beta_1 S^2 + \beta_2 S \rho^2 - \frac{1}{3} \alpha_2 \rho^3 + \frac{1}{4} \alpha_3 \rho^4, \quad (3.3.5)$$

where α_i ($1 \leq i \leq 3$), β_1 , and β_2 depend continuously on T and p , and

$$\beta_1 = \beta_1(T, p) > 0, \quad \alpha_i = \alpha_i(T, p) > 0 \quad i = 2, 3. \quad (3.3.6)$$

Dynamic Equations for PVT Systems In a PVT system, the order parameter is $u = (\rho, S)$,

$$\rho = \rho_1 - \rho_0, \quad S = S_1 - S_0,$$

where ρ_i and S_i ($i = 0, 1$) represent the molar density and entropy density, ρ_0, S_0 are reference points. Hence the conjugate variables of ρ and S are the pressure p and the temperature T . Thus, by the standard model (3.1.2), we derive from (3.3.3) to (3.3.5) the following general form of the time-dependent equations governing a PVT system:

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= \mu_1 \Delta \rho - (\alpha_1 + bp)\rho + \alpha_2 \rho^2 - \alpha_3 \rho^3 - 2\beta_2 \rho S + p, \\ \frac{\partial S}{\partial t} &= \mu_2 \Delta S - \beta_1 S - \beta_2 \rho^2 + T. \end{aligned} \quad (3.3.7)$$

Although the domain Ω depends on T and p , we can still take the Neumann boundary condition

$$\frac{\partial \rho}{\partial n} = 0, \quad \frac{\partial S}{\partial n} = 0 \quad \text{on } \partial \Omega. \quad (3.3.8)$$

An important special case for PVT systems is that the pressure and the temperature functions are homogeneous in Ω . Thus we can assume that ρ and S are independent of $x \in \Omega$, and the free energy (3.3.3) with (3.3.4) and (3.3.5) can be expressed as

$$G(\rho, S, T, p) = \frac{\alpha_1}{2}\rho^2 + \frac{\beta_1}{2}S^2 + \beta_2 S\rho^2 - \frac{\alpha_2}{3}\rho^3 + \frac{\alpha_3}{4}\rho^4 + \frac{bp^2}{2} - \rho p - ST. \quad (3.3.9)$$

We derive then from (3.3.9) the following dynamic equations

$$\begin{aligned} \frac{d\rho}{dt} &= -(\alpha_1 + bp)\rho + \alpha_2\rho^2 - \alpha_3\rho^3 - 2\beta_2S\rho + p, \\ \frac{dS}{dt} &= -\beta_1S - \beta_2\rho^2 + T. \end{aligned} \quad (3.3.10)$$

Because $\beta_1 > 0$ for all T and p , we can replace the second equation of (3.3.10) by

$$S = \beta_1^{-1}(T - \beta_2\rho^2). \quad (3.3.11)$$

Then, (3.3.10) are equivalent to the following equation

$$\frac{d\rho}{dt} = -(\alpha_1 + bp + 2\beta_1^{-1}\beta_2T)\rho + \alpha_2\rho^2 - (\alpha_3 - 2\beta_2^2\beta_1^{-1})\rho^3 + p. \quad (3.3.12)$$

It is clear that the steady-state equation of (3.3.12) should agree with the Van der Waals equation. Hence it is natural to choose

$$\alpha_1 = 0, \quad 2\beta_1^{-1}\beta_2 = R, \quad \alpha_2 = a, \quad (\alpha_3 - 2\beta_2^2\beta_1^{-1}) = ab.$$

3.3.3 Phase Transition Dynamics for PVT Systems

We now use (3.3.12) to discuss dynamical properties of transitions for PVT systems, and remark that similar results can also be derived using the more general form of the time-dependent model (3.3.7).

Let ρ_0 be a steady-state solution of (3.3.12) near the Andrews point $C = (T_c, p_c)$. We take the transformation

$$\rho = \rho_0 + \rho'.$$

Then equation (3.3.12) becomes (drop the prime)

$$\frac{d\rho}{dt} = \lambda\rho + a_2\rho^2 - ab\rho^3. \quad (3.3.13)$$

Here

$$\lambda = 2a\rho_0 - 3ab\rho_0^2 - \alpha_1 - bp - RT, \quad a_2 = a(1 - 3b\rho_0), \quad (3.3.14)$$

where α_1 is close to zero. Here we emphasize that ρ_0 and (λ, a_2) are all functions of the control parameter $\lambda = (T, p)$.

These are two important physical parameters, which are used to fully characterize the dynamic behavior of gas-liquid transition near the Andrews critical point. In fact, from the derivation of the model, we obtain immediately the following physical meaning of these two parameters:

$$\lambda(T, p) = \frac{d^2 G}{dp^2} \Big|_{\rho=\rho_0}, \quad a_2(T, p) = \frac{1}{2} \frac{d^3 G}{dp^3} \Big|_{\rho=\rho_0}, \quad (3.3.15)$$

where $\rho_0 = \rho_0(T, p)$ is the equilibrium state.

In the PT -plane, near the Andrews point $C = (T_c, p_c)$, the critical parameter equation

$$\lambda = \lambda(T, p) = 0 \quad \text{in } |T - T_c| < \delta, \quad |p - p_c| < \delta$$

for some $\delta > 0$, defines a continuous function $T = \phi(p)$, such that

$$\lambda \begin{cases} < 0 & \text{if } T > \phi(p), \\ = 0 & \text{if } T = \phi(p), \\ > 0 & \text{if } T < \phi(p). \end{cases} \quad (3.3.16)$$

Equivalently, this is called the principle of exchange of stabilities (Chandrasekhar, 1981), which, as we have shown in Ma and Wang (2008b), is the necessary and sufficient condition for the gas-liquid phase transition.

One important component of our theory is that the Andrews critical point is determined by the system of equations

$$\begin{aligned} \lambda &= 0, \\ a_2 &= 0, \\ -(bp + RT)\rho_0 + a\rho_0^2 - ab\rho_0^3 + p &= 0. \end{aligned} \quad (3.3.17)$$

Here the first equation is the critical parameter equation, the second equation, as we shall see below, determines the switching point where the phase transition switches types, and the last equation is the Van der Waals equation, which is also the steady-state equation of the dynamic model.

Then by directly solving the system of three algebraic equations in (3.3.17), the critical point C is given by

$$(\rho_c, p_c, T_c) = \left(\frac{1}{3b}, \frac{a}{27b^2}, \frac{8a}{27bR} \right). \quad (3.3.18)$$

This is in agreement with the classical work by Van der Waals. *Here we obtain the Andrews point using a dynamic approach.*

The main result in this section is the following dynamic transition theorem:

Theorem 3.3.1 Let $T_0 = \phi(p_0)$ and $a_3 > 0$. Then the system (3.3.13) has a transition at $(T, p) = (T_0, p_0)$, and the following assertions hold true:

1. If the coefficient $a_2 = a_2(T, p)$ in (3.3.13) is zero at (T_0, p_0) , i.e., $a_2(T_0, p_0) = 0$, then the transition is of Type-I, as schematically shown in Fig. 3.5.
 2. If $a_2(T_0, p_0) \neq 0$, then the transition is of Type-III (i.e., the mixed type), and the following assertions hold true:
 - a. There are two transition solutions near (T_0, p_0) as
- $$\rho^\pm(T, p) = \frac{1}{2a_3} \left(a_2 \pm \sqrt{a_2^2 + 4a_3\lambda} \right). \quad (3.3.19)$$
- b. There is a saddle-node bifurcation at (T_1, p_1) , where $T_1 > T_0$ and $p_1 < p_0$ if $\phi'(p_0) > 0$, and $p_0 > p_1$ if $\phi'(p_0) < 0$.
 - c. For $\phi'(p_0) > 0$, when $a_2(T_0, p_0) > 0$ the transition diagrams are illustrated in Fig. 3.6, where ρ^+ is stable for all (T, p) near (T_0, p_0) , and $\rho = 0$ is stable, ρ^- a saddle for $T_0 < T < T_1$ and $p_1 < p < p_0$, and ρ^- is stable, $\rho = 0$ a saddle for $T < T_0, p > p_0$;
 - d. When $a_2(T_0, p_0) < 0$, the transition diagrams are illustrated in Fig. 3.7, where ρ^- is stable for all (T, p) near (T_0, p_0) , and $\rho = 0$ is stable, ρ^+ a saddle for $T_0 < T < T_1, p_1 < p < p_0$, and ρ^+ is stable, $\rho = 0$ a saddle for $T < T_0, p_0 < p$.

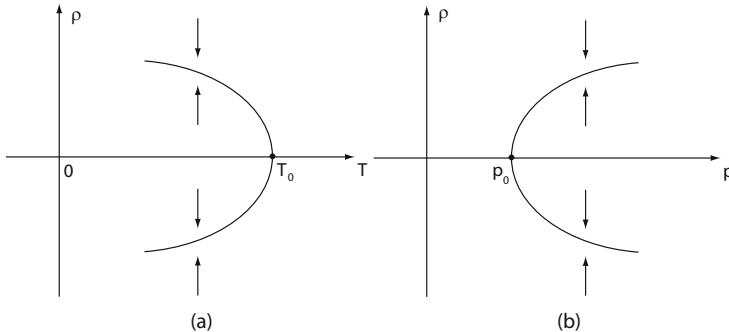


Fig. 3.5 Continuous transition for the case where $\phi'(p_0) > 0$

Proof. By (3.3.13), the theorem follows from the transition theorems (Theorems 2.3.1 and 2.3.2) and the singularity separation theorem (Theorem 2.5.4). We omit the detailed routine analysis. \square

3.3.4 Physical Conclusions

To discuss the physical significance of Theorem 3.3.1, we recall the classical pV -phase diagram given in Fig. 3.4. We take $\rho = 1/V$ to replace volume V , then the gas-liquid coexistence curve in the ρ - P plane is illustrated in Fig. 3.8.

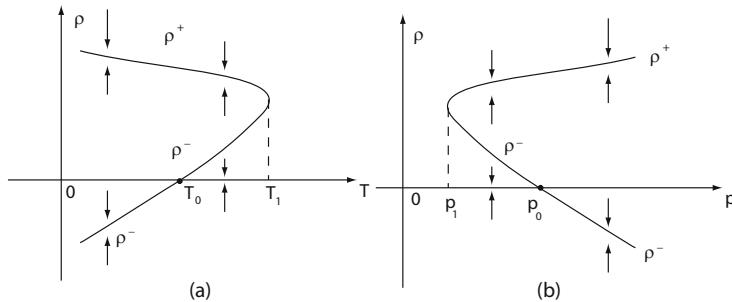


Fig. 3.6 Type-III (mixed) transition for $\alpha_2 > 0$

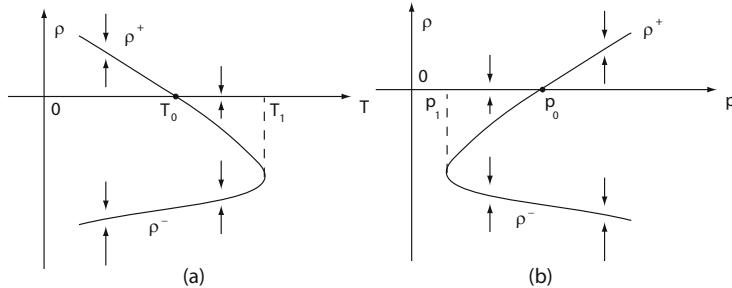
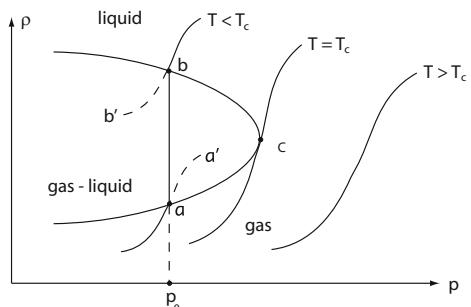


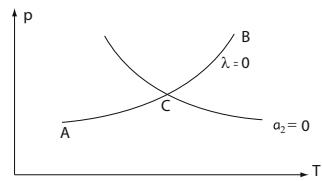
Fig. 3.7 Type-III (mixed) transition for $\alpha_2 < 0$

Fig. 3.8 The gas-liquid coexistence curve in $\rho - p$ plane: point C is the Andrews critical point, the dashed lines aa' and bb' stand for the metastable states



According to the physical experiments, in the gas-liquid coexistence region in the $\rho - p$ phase diagram, there exist metastable states. Mathematically speaking, the metastable states are the attractors which have a small basin of attraction. In Fig. 3.8, the dashed lines aa' and bb' represent the metastable states, and the points in aa' correspond to super-heated liquid, the points in bb' correspond to super-cooled gas. The $\rho - p$ phase diagram shows that along the isothermal line $T > T_C$ where $C = (T_C, P_C)$ is the Andrews critical point, when the pressure p increases, the density ρ varies continuously from gaseous to liquid states. However along the isothermal line $T < T_C$ when the pressure p increases to p_0 the density ρ will undergo an abrupt change, and a transition from gaseous state a to a liquid state b accompanied with an

Fig. 3.9 The point $C = (T_c, p_c)$ is the Andrews critical point



isothermal exothermal process to occur. Likewise, such processes also occur in the gas-solid and liquid-solid transitions.

We now return to discuss Theorem 3.3.1, and explain the gas-liquid transition near the Andrews critical point C .

First, we have shown in (3.3.18) that at the equilibrium point ρ_0 , the two curves given by $\lambda(p, T) = 0$ and $a_2(p, T) = 0$ interact exactly at the critical point C as shown in Fig. 3.9, and the curve segment AB of $\lambda = 0$ is divided into two parts AC and CB by the point C such that

$$\begin{aligned} a_2(T, p) &> 0 \quad \text{for } (T, p) \in AC, \\ a_2(T, p) &< 0 \quad \text{for } (T, p) \in CB. \end{aligned}$$

Here the curve AC is the classical gas-liquid coexistence curve.

Second, on the curve AC , excluding the critical point C , $a_2(T, p) > 0$. The phase transition of the system is a mixed type if we take a pass crossing the AC ; see Fig. 3.6. In Fig. 3.6, ρ is the deviation from the basic gaseous state ρ_0 . We now consider different states given in Fig. 3.6a:

- For $T > T_1$, the gaseous state ρ_0 , corresponding to $\rho = 0$ in the figure, is stable. This is the only stable physical state in this temperature range, and the system is in the gaseous state.
- For $T_0 < T < T_1$, there are two metastable states given by gaseous phase ρ_0 and the liquid phase $\rho_0 + \rho^+$.
- For $T < T_0$, there are three states: the unstable basic gaseous state ρ_0 , and the two metastable states: $\rho_0 + \rho^-$ and $\rho_0 + \rho^+$. One important component of our theory is that the only physical phase here is the liquid phase represented by metastable state: $\rho_0 + \rho^+$. Although, mathematically speaking, the gaseous state $\rho_0 + \rho^-$ is also metastable, it does not appear in nature. The only possible explanation for this exclusion is the asymmetry principle of fluctuations, to be further explored in the next section.
- Hence we have shown that as we lower the temperature, the system undergoes a *first-order* transition from a gaseous state to a liquid state with an abrupt change in density. In fact, there is an energy gap between the gaseous and liquid states:

$$\Delta E = G(\rho_0 + \rho^+) - G(\rho_0) \sim -\frac{\lambda}{4}(\rho^+)^2 - \frac{a_2}{12}(\rho^+)^3 < 0 \quad \text{for } T < T_1.$$

This energy gap $|\Delta E|$ stands for a latent heat, and $\Delta E < 0$ shows that the transition from a gaseous state to a liquid state is an isothermal exothermal process, and from a liquid state to gaseous state is an isothermal endothermal process.

Third, at the critical point C , we have $a_2 = 0$. Then the dynamic transition is as shown in Fig. 3.5:

- As in the previous case, for $T > T_0$, the only physical state is given by the gaseous phase ρ_0 , corresponding to zero deviation shown in the figure.
- As the temperature T is lowered crossing T_0 , the gaseous state losses its stability, leading to two metastable states: one is the liquid phase $\rho_0 + \rho^+$, and the other is the gaseous phase $\rho_0 + \rho^-$. Again, the gaseous phase $\rho_0 + \rho^-$ does not appear, and the asymmetry principle of fluctuations is valid in this situation as well.
- The phase transition here is of the second order, as the energy is continuous at T_0 . In fact, the energy for the transition liquid state is given by

$$G(\rho_0 + \rho^+) = G(\rho_0) - \frac{\alpha^2}{4a_3}(T - T_c)^2 \quad \text{for } T < T_c,$$

for some $\alpha > 0$. Hence the difference of the heat capacity at $T = T_c$ is

$$\Delta C = -T_c \frac{\partial^2}{\partial T^2} (G(\rho_0 + \rho^+) - G(\rho_0)) = \frac{\alpha^2}{2a_3} T_c > 0.$$

Namely the heat capacity has a finite jump at $T = T_c$, therefore the transition at $T = T_c$ is of the *second order*.

Fourth, on the curve BC , $a_2(T, p) < 0$, and the phase transition diagram is given in Fig. 3.7:

- For $T > T_1$, the system is in the gaseous phase, which is stable.
- For $T_0 < T < T_1$, there are two metastable gaseous states given by ρ_0 and $\rho_0 + \rho^-$. As before, although it is metastable, the gaseous state $\rho_0 + \rho^-$ does not appear.
- For $T < T_0$, the gaseous phase ρ_0 losses its stability, and the system undergoes a dynamic transition to the metastable liquid state $\rho_0 + \rho^+$. Mathematically, the gaseous state $\rho_0 + \rho^-$ is also metastable. However, it does not appear either, due to the asymmetry principle of fluctuations.
- The dynamic transition in this case is of the third order. In fact, we have

$$G(\rho_0 + \rho^+) = G(\rho_0(T)) - \frac{\alpha^3}{6|a_2|^2}(T_0 - T)^3 + o(|T_0 - T|^3) \quad \text{for } T < T_0.$$

Namely, the free energy is continuously differentiable up to the second order at $T = T_0$, and the transition is of the *third order*. It implies that as $(T_0, p_0) \in CB$ the third-order transition at (T_0, p_0) cannot be observed by physical experiments.

In summary, we have obtained a precise characterization of the phase transition behavior near the Andrews point, and have derived precisely the nature of the Andrews critical point.

3.4 Ferromagnetism

Ferromagnetism is referred to a spontaneous magnetization of materials such as iron. For ferromagnetic systems, the magnetization is the first derivative of the free energy with respect to the applied magnetic field strength, and the magnetic susceptibility is the second derivative of the free energy with respect to the field. It has been observed that the magnetization increases continuously from zero as the temperature is lowered below the Curie temperature, and the magnetic susceptibility changes discontinuously. Hence, the ferromagnetic phase transition in materials such as iron is regarded as a second-order phase transition (Kleman and Laverntovich, 2007). However, a theoretical understanding of the transition is still lacking. For example, the classical GL free energy leads to interesting results, in agreement with experiments, for both the steady-state and time-dependent models. However, there are obvious discrepancies of the classical theories on both susceptibility and spontaneous magnetization.

The main objective of this section is to identify these discrepancies, to revise the GL free-energy to address these issues using the dynamic transition theory, and to derive a physical prediction on spontaneous magnetization. In addition, the analysis leads naturally to a physical conjecture on asymmetry of fluctuations, which appears in both the ferromagnetic system studied in this section and in PVT systems studied in the previous section; see also Ma and Wang (2008c, 2011e).

3.4.1 Classical Theory of Ferromagnetism

A ferromagnetic material consists of lattices containing particles with a magnetic moment. When no external field is present and the temperature is above some critical value, called the Curie temperature, the magnetic moments are oriented at random and there is no net magnetization. However, as the temperature is lowered, magnetic interaction energy between lattice sites becomes more important than the random thermal energy. Below the Curie temperature, the magnetic moments become ordered in space and a spontaneous magnetization appears. The phase transition from a paramagnetic to a ferromagnetic system takes place at the Curie temperature.

The phase diagrams for magnetic systems are given in Figs. 3.10, 3.11, and 3.12. In Fig. 3.10, below the Curie temperature, the magnetization occurs spontaneously, and the zero magnetic field $H = 0$ separates the two possible orientations of magnetization. Figure 3.11 provides a sketch of the isotherms of magnetic system, and Fig. 3.12 gives the magnetization as a function of temperature; see also Reichl (1998), Stanley (1971) for details.

Based on the classical Ginzburg-Landau theory, for an isotropic system, the Helmholtz free energy can be expressed as

$$A(M, T) = A_0(T) + \frac{1}{2}\alpha_2(T)|M|^2 + \frac{1}{4}\alpha_4(T)|M|^4 + \dots,$$

Fig. 3.10 Below the Curie point the magnetization occurs spontaneously; the curve $H = 0$ separates the two possible orientations of magnetization

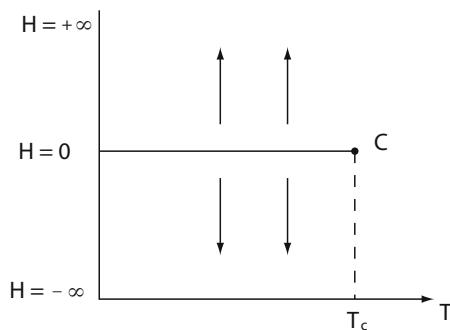


Fig. 3.11 A sketch of the isotherms for a magnetic system

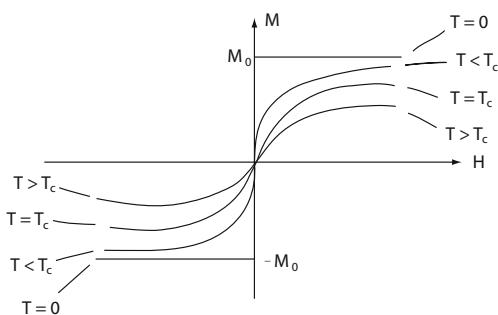
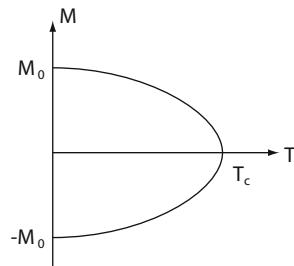


Fig. 3.12 A sketch of the spontaneous magnetization of a magnetic system



where $A_0(T)$ is a magnetization-independent contribution to the free energy, $|M|^2 = M \cdot M$, and $M = (M_1, M_2, M_3)$ is the magnetization of the system. When an external field H is present, the Gibbs free energy is given by

$$\begin{aligned} G(M, H, T) &= A(M, T) - H \cdot M \\ &= A_0(T) - H \cdot M + \frac{1}{2} \alpha_2(T, H)|M|^2 + \frac{1}{4} \alpha_4(T, H)|M|^4 + \dots \end{aligned}$$

For small H , α_2 and α_4 can be considered to be independent of H , and near the Curie point T_c we have

$$\alpha_2(T) = \alpha_0(T - T_c), \quad \alpha_4(T) > 0, \quad \alpha_0 > 0 \text{ a constant.}$$

Usually, $G(M, H, T)$ is called the Ginzburg-Landau free energy. To omit the higher-order terms than $|M|^4$, it is known that the equilibrium state M of the ferromagnetic system satisfies

$$\frac{\delta}{\delta M} G = \alpha_4 |M|^2 M + \alpha_2 M - H = 0. \quad (3.4.1)$$

Thus above the Curie point we obtain from (3.4.1) that

$$M \simeq \frac{1}{\alpha_2} H, \quad \chi = \frac{\partial M}{\partial H} = \frac{1}{\alpha_2(T)} = \frac{1}{\alpha_0(T - T_c)}, \quad (3.4.2)$$

where χ is the isothermal susceptibility, which is a scalar because the system is isotropic. Below the critical point, for $H = 0$, the magnetization M obeys

$$|M| = \sqrt{\frac{\alpha_0(T_c - T)}{\alpha_4}}, \quad \frac{\partial |M|}{\partial T} = -\frac{1}{2} \sqrt{\frac{\alpha_0}{\alpha_4(T_c - T)}}. \quad (3.4.3)$$

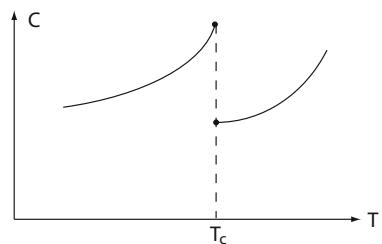
The heat capacity at $T = T_c$ is

$$\Delta C = -T \frac{\partial^2 G}{\partial T^2} = -T_c \frac{\partial^2}{\partial T^2} \left(\frac{1}{2} \alpha_2 |M|^2 + \frac{1}{4} \alpha_4 |M|^4 \right) = \frac{\alpha_0^2 T_c}{2 \alpha_4}. \quad (3.4.4)$$

We infer then from (3.4.2) to (3.4.4) the following classical conclusions for an isotropic magnetic system:

- (1) When an external magnetic field is present, a nonzero magnetization exists above the Curie point T_c , which has the same direction as the applied field H .
- (2) Near the critical point T_c the susceptibility χ tends to infinite with the rate $(T - T_c)^{-1}$, i.e., a very small applied field at $T = T_c$ can yield a large effect on the magnetization.
- (3) In the absence of an external field (i.e., $H = 0$), below the critical point a spontaneous magnetization M appears, which depends continuously on T and tends to zero with the rate $(T - T_c)^{1/2}$; namely the transition is of the second order.
- (4) The heat capacity at $T = T_c$ has a jump at $T = T_c$ with the gap $\Delta C = \frac{\alpha_0^2 T_c}{2 \alpha_4}$, and the jump has the shape of λ , as shown in Fig. 3.13.

Fig. 3.13 The heat capacity at $T = T_c$ has a jump transition at $T = T_c$



Qualitatively, part of the above conclusions is in agreement with experimental results. However these conclusions lead to wrong susceptibility and spontaneous magnetization, whose experimental rates are given by $\chi \propto (T - T_c)^{-r}$ with $r = 1.3$, and by $M \propto (T - T_c)^\beta$ with $\beta = 1/3$.

Free energy G must be a function of the magnetization M . Hence the errors are originated from the fact that the expression of G in the Ginzburg-Landau theory is an approximation. It is difficult to derive a precise formula because G is not analytic on $|M|$, even the differentiability of G on $|M|$ near T_c is very low.

If we study the dynamical properties of ferromagnetic systems by using the classical Ginzburg-Landau free energy, we shall see a more serious error when an external field is present.

To see this, the dynamic equation of classical theory is given by

$$\frac{dM}{dt} = -\alpha_2 M - \alpha_4 |M|^2 M + H. \quad (3.4.5)$$

For simplicity, we take $H = (h, 0, 0)$ with $h > 0$, it is equivalent that we take the x_1 -axis in the direction of H . Then the equation

$$\alpha_4 |M|^2 M + \alpha_2 M - H = 0$$

has a steady-state solution $M_0 = (m_0, 0, 0)$ for $T \geq 0$, which is the magnetization induced by H . Make the transformation

$$M = M' + M_0.$$

Then, the equation (3.4.5) is rewritten as (drop the primes)

$$\begin{aligned} \frac{dM_1}{dt} &= -(\alpha_2 + 3\alpha_4 m_0^2)M_1 - 2\alpha_4 m_0 M_1^2 - \alpha_4 m_0 |M|^2 - \alpha_4 |M|^2 M_1, \\ \frac{dM_2}{dt} &= -(\alpha_2 + \alpha_4 m_0^2)M_2 - 2\alpha_4 m_0 M_1 M_2 - \alpha_4 |M|^2 M_2, \\ \frac{dM_3}{dt} &= -(\alpha_2 + \alpha_4 m_0^2)M_3 - 2\alpha_4 m_0 M_1 M_3 - \alpha_4 |M|^2 M_3. \end{aligned} \quad (3.4.6)$$

Comparing the two critical parameter curves

$$\begin{aligned} \alpha_2 + 3\alpha_4 m_0^2 &= 0 \Rightarrow T_1 = T_c - 3\alpha_4 m_0^2 / \alpha_0, \\ \alpha_2 + \alpha_4 m_0^2 &= 0 \Rightarrow T_2 = T_c - \alpha_4 m_0^2 / \alpha_0, \end{aligned}$$

we find that $T_2 > T_1$. By Theorem 2.1.3, (3.4.6) has the first transition at $T = T_2$, where a new magnetization $M = (M_1, M_2, M_3)$, with $M_2 \neq 0$ and $M_3 \neq 0$, appears. This is unrealistic because any magnetization M of this system must have the same direction as $H = (h, 0, 0)$; see Fig. 3.10.

In fact, when a magnetic field H is applied on an isotropic system the direction of H is a favorable one for magnetization. However, in (3.4.5) this point is not manifested. Therefore, to investigate the phase transition dynamics of ferromagnetic systems we need to revise the *GL* free energy.

3.4.2 Dynamic Transitions in Ferromagnetism

Revised Ginzburg-Landau Free Energy Let the ferromagnetic system be isotropic. When a magnetic field H is present, we introduce a second-order symmetric tensor

$$A(T, H) = (a_{ij}(T, H)), \quad a_{ij} = a_{ji} \quad \text{for } 1 \leq i, j \leq 3,$$

such that $A(T, 0) = 0$, and $A(T, H)$ has eigenvalues

$$\lambda_1 = \lambda_1(T, H), \quad \lambda_2 = \lambda_3 = 0 \quad \text{with } \lambda_1(T, H) > 0 \text{ as } H \neq 0,$$

and H is the eigenvector of A corresponding to λ_1 :

$$AH = \lambda_1 H \quad (\lambda_1 > 0 \text{ as } H \neq 0, \quad \lambda_1 = 0 \text{ as } H = 0). \quad (3.4.7)$$

It is clear that if we take the coordinate system (x_1, x_2, x_3) with x_1 -axis in the H -direction, then $H = (H_1, 0, 0)$ and

$$A = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (3.4.8)$$

Physically, condition (3.4.7) means that H is a favorable direction of magnetization if we add a term

$$-MAM^T = -\sum a_{ij}M_i M_j$$

in the free energy. We also need to consider the nonlinear effect acted by H . To this end we introduce the term $-|M|^2 M \cdot H$ in the free energy.

Thus when the applied field H may vary in $\Omega \subset \mathbb{R}^n$ ($n = 2, 3$), then the *GL* free energy is in the form

$$\begin{aligned} G(M, T, H) = G_0 + \frac{1}{2} \int_{\Omega} \left[\mu |\nabla M|^2 + \alpha_2 |M|^2 + \frac{\alpha_4}{2} |M|^4 \right. \\ \left. - \sum a_{ij} M_i M_j - f(T, H, M) M \cdot H \right] dx, \end{aligned} \quad (3.4.9)$$

where $G_0 = G_0(T)$ is independent of M and H , and f is a scalar function of T, H and M , defined by

$$f(T, H, M) = 2(1 + \beta|M|^2), \quad \beta = \beta(T, H) > 0. \quad (3.4.10)$$

For α_2 and α_4 we assume that

$$\alpha_2 = \alpha_0(H)(T - T_0(H)), \quad T_0(0) = T_c, \quad \alpha_0 > 0, \quad \alpha_4 > 0. \quad (3.4.11)$$

By the standard model (3.1.2), we derive from (3.4.9) and (3.4.10) the following dynamical equations:

$$\begin{aligned} \frac{\partial M_i}{\partial t} = & \mu \Delta M_i - \alpha_2 M_i + \sum_{j=1}^3 a_{ij} M_j - \alpha_4 |M|^2 M_i \\ & + \beta |M|^2 H_i + 2\beta (M \cdot H) M_i + H_i \quad \text{for } 1 \leq i \leq 3, \end{aligned} \quad (3.4.12)$$

supplemented with the Neumann boundary condition $\partial M / \partial n = 0$ on $\partial\Omega$. Obviously, if $H = 0$, (3.4.12) coincide with the classical equations. For simplicity, hereafter we always take

$$H = (h, 0, 0) \quad (h > 0 \text{ is a constant}). \quad (3.4.13)$$

When H is constant, in the study of phase transitions of magnetic systems, (3.4.12) can be replaced by a system of ordinary differential equations as follows:

$$\begin{aligned} \frac{dM_1}{dt} &= (\lambda_1 - \alpha_2) M_1 + \beta h |M|^2 + 2\beta h M_1^2 - \alpha_4 |M|^2 M_1 + h, \\ \frac{dM_2}{dt} &= -\alpha_2 M_2 + 2\beta h M_1 M_2 - \alpha_4 |M|^2 M_2, \\ \frac{dM_3}{dt} &= -\alpha_2 M_3 + 2\beta h M_1 M_3 - \alpha_4 |M|^2 M_3. \end{aligned} \quad (3.4.14)$$

Equations (3.4.14) have a steady-state solution induced by H :

$$M^* = (m_0, 0, 0) \quad \text{with } \lim_{h \rightarrow 0} m_0 = 0.$$

We see in Fig. 3.11 that the magnetization M^* has an upper bound. Namely there is an M_0 such that

$$\begin{aligned} |M^*| &< |M_0| & \forall h \in \mathbb{R}^1, \quad T \geq 0, \\ M^* &\rightarrow M_0 & \text{if } h \rightarrow \infty. \end{aligned} \quad (3.4.15)$$

To satisfy (3.4.15) it is necessary to assume that the coefficients α_0 and α_4 as in (3.4.11) possess the properties

$$\alpha_0(H) \rightarrow +\infty, \quad \alpha_4(T, H) \rightarrow +\infty, \quad \text{as } h \rightarrow \infty. \quad (3.4.16)$$

Both conditions (3.4.11) and (3.4.16) are physical.

Dynamic Transitions To illustrate the main ideas, we only consider the case where H is a constant on Ω . Therefore we shall study phase transition dynamics of the ferromagnetic systems by using equations (3.4.14) for $h > 0$. Analysis for more general case can be carried out in the same fashion, and will be reported elsewhere.

Take the transformation in (3.4.14)

$$M = M^* + M'. \quad (3.4.17)$$

Then equations (3.4.14) are rewritten as (drop the primes)

$$\begin{aligned}\frac{dM_1}{dt} &= \beta_1 M_1 - 2a_2 M_1^2 - a_2 |M|^2 - \alpha_4 |M|^2 M_1, \\ \frac{dM_2}{dt} &= \beta_2 M_2 - 2a_2 M_1 M_2 - \alpha_4 |M|^2 M_2, \\ \frac{dM_3}{dt} &= \beta_2 M_3 - 2a_2 M_1 M_3 - \alpha_4 |M|^2 M_3,\end{aligned}\quad (3.4.18)$$

where

$$\begin{aligned}a_2 &= \alpha_4 m_0 - \beta h, \\ \beta_1 &= \lambda_1 + 6\beta h m_0 - 3\alpha_4 m_0^2 - \alpha_2, \quad \beta_2 = 2\beta h m_0 - \alpha_4 m_0^2 - \alpha_2.\end{aligned}$$

The critical parameter curves $\beta_1 = 0$ and $\beta_2 = 0$ are given by

$$\begin{aligned}\beta_1 = 0 &\Rightarrow T_1 = T_0(H) + \frac{1}{\alpha_0}(\lambda_1 + 6\beta h m_0 - 3\alpha_4 m_0^2), \\ \beta_2 = 0 &\Rightarrow T_2 = T_0(H) + \frac{1}{\alpha_0}(2\beta h m_0 - \alpha_4 m_0^2).\end{aligned}$$

It is clear that $T_1 > T_2$ provided

$$\lambda_1 > 2m_0(\alpha_4 m_0 - 2\beta h), \quad \text{for } h > 0. \quad (3.4.19)$$

Therefore, under condition (3.4.19), the equations (3.4.18) have a transition at $T = T_1$ in the space

$$E = \{(M_1, 0, 0) \mid -\infty < M_1 < \infty\}. \quad (3.4.20)$$

More precisely, we have the following transition theorem.

Theorem 3.4.1 Assume the condition (3.4.19) and $a_2 \neq 0$. Then (3.4.18) has a Type-III (mixed) transition at $T = T_1$, and the transition occurs in the space E . The phase diagram is as shown in Fig. 3.14. Moreover we have the following assertions:

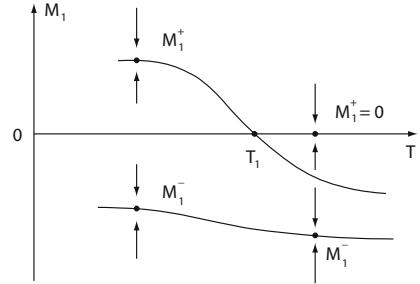
(1) There are two stable equilibrium states near $T = T_1$, which are given by

$$\begin{aligned}M_1^+ &= \begin{cases} 0 & \text{if } T > T_1, \\ \frac{1}{2\alpha_4}[-3a_2 + \sqrt{9a_2^2 + 4\alpha_4\beta_1}] & \text{if } T < T_1, \end{cases} \\ M_1^- &= -\frac{1}{2\alpha_4}[3a_2 + \sqrt{9a_2^2 + 4\alpha_4\beta_1}].\end{aligned}$$

- (2) If $T < T_1$, M_1^+ is stable in the region $0 < M_1 < \infty$, and M_1^- is stable in $-\infty < M_1 < 0$.
- (3) If $T > T_1$, $M_1^+ = 0$ is stable for $-b < M_1 < \infty$ and M_1^- is stable for $-\infty < M_1 < -b$, where

$$b = \frac{1}{2\alpha_4} \left[3a_2 - \sqrt{9a_2^2 + 4\alpha_4\beta_1} \right] > 0 \quad \text{for } T > T_1.$$

Fig. 3.14 Phase diagram of the transitions given by Theorem 3.4.1



Proof. It is clear that if (3.4.19) holds, then

$$\beta_1(T) \begin{cases} < 0 & \text{if } T > T_1, \\ = 0 & \text{if } T = T_1, \\ > 0 & \text{if } T < T_1, \end{cases} \quad \beta_2(T_1) < 0.$$

Hence, by Theorem 2.1.3 the system (3.4.18) has a transition at $T = T_1$. Obviously, the space E defined by (3.4.20) is the center manifold of (3.4.18) near $T = T_1$. Hence, the reduced equation of (3.4.18) on E is expressed as

$$\frac{dM_1}{dt} = \beta_1 M_1 - 3a_2 M_1^2 - \alpha_4 M_1^3. \quad (3.4.21)$$

As $a_2 \neq 0$, by Theorem 2.3.2 we infer from (3.4.21) that this transition is of Type-III, and the transition solutions satisfy

$$\alpha_4 M_1^2 + 3a_2 M_1 - \beta_1 = 0.$$

By a direct compute one obtains Assertions (1)–(3). The proof is complete. \square

3.4.3 Physical Implications

By (3.4.17), the stable steady states of (3.4.14) near $T = T_1$ are

$$M^\pm = (m_0 + M_1^\pm, 0, 0).$$

From the physical point of view, it should be

$$M_1^+ \geq 0, \quad M_1^- < 0, \quad m_0 + M_1^- \geq 0. \quad (3.4.22)$$

The condition (3.4.22) requires that $0 < 3a_2 < \alpha_4 m_0$, which is equivalent to

$$\beta h < \alpha_4 m_0 < \frac{3}{2}\beta h \quad (h > 0) \quad (3.4.23)$$

where $m_0 > 0$ is a solution of the equation near $T = T_1$:

$$\alpha_4 m_0^3 - 3\beta h m_0^2 + (\alpha_2 - \lambda_1) m_0 - h = 0 \quad (h > 0). \quad (3.4.24)$$

Thus, the stable steady states M^+ and M^- of (3.4.14) near $T = T_1$ are physical provided that the coefficients $\alpha_2(T, h), \alpha_4(T, h), \beta(T, h)$ and $\lambda_1(T, h)$ satisfy (3.4.23) and (3.4.24). In this case the temperature T_1 is greater than the Curie temperature T_c :

$$T_1(H) > T_c = T_1(0) \quad \text{for } H \neq 0.$$

The two states M^+ and M^- are mathematically equal, therefore only by Theorem 3.4.1 we cannot determine the magnetization behaviors of ferromagnetic systems near $T = T_1$. However, we see that the magnetization M^+ is stronger than M^- . Physically, it implies that M^+ is favorable in $T < T_1$, and M^- is in $T > T_1$. Thus, from Theorem 3.4.1 there are two possible magnetization behaviors, i.e., two magnetization functions:

$$\begin{aligned} \mu_1(T) &= M^+(T) = (m_0(T) + M_1^+(T), 0, 0), \\ \mu_2(T) &= \begin{cases} M^+(T) = (m_0(T) + M_1^+(T), 0, 0) & \text{if } T < T_1, \\ M^-(T) = (m_0(T) + M_1^-(T), 0, 0) & \text{if } T \geq T_1. \end{cases} \end{aligned}$$

The function $\mu_1(T)$ is continuous on T , as shown in Fig. 3.15a, and its derivative is discontinuous at $T = T_1$:

$$\mu'_1(T_1^-) - \mu'_1(T_1^+) \approx \left(\frac{1}{3} \frac{d}{dT} \frac{\beta_1(T_1)}{\alpha_2(T_1)}, 0, 0 \right).$$

The function $\mu_2(T)$ has a jump at $T = T_1$, as shown in Fig. 3.15b.

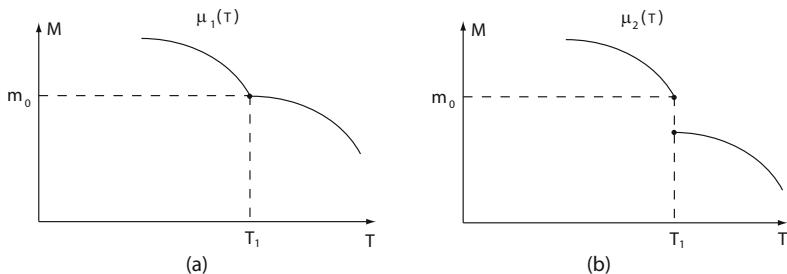


Fig. 3.15 (a) The graph of function $\mu_1(T)$; (b) the graph of function $\mu_2(T)$.

On the other hand, by direct computation, the free energies of M^+ and M^- are shown to be given by

$$G(M^+) = \begin{cases} G(m_0) & \text{if } T \geq T_1, \\ G(m_0) + \frac{1}{4}(M_1^+)^2(a_2 M_1^+ - \beta_1) & \text{if } T < T_1, \end{cases}$$

$$G(M^-) = G(m_0) + \frac{1}{4}(M_1^-)^2(a_2 M_1^- - \beta_1),$$

where m_0 is the magnetization induced by $H = (h, 0, 0)$ satisfying (3.4.24), and $\beta_1(T_1) = 0$. It is clear that

$$G(M^+) > G(M^-) \quad \text{near } T = T_1. \quad (3.4.25)$$

Hence, it follows from (3.4.25) that the magnetization behavior described by $\mu_2(T)$ is prohibited in real world because the free energy cannot abruptly increase (or decrease) in a temperature decreasing (or increasing) process. Thus, by Theorem 3.4.1 and (3.4.25) we can derive the following physical conclusion:

Physical Conclusion 3.4.2 *When an external field H is present, the magnetization $M_H(T)$ of an isotropic ferromagnetic system is continuous on the temperature T , and there is a $T_1(H) > T_c$ (T_c the Curie temperature) with $T_1(H) \rightarrow T_c$ as $H \rightarrow 0$ such that $M_H(T)$ is not differentiable at $T = T_1$, whose derivative has a finite jump*

$$M'_H(T_1^-) - M'_H(T_1^+) = a > 0 \quad (a < \infty).$$

Moreover, the graph of $M_H(T) = \mu_1(T)$ as shown in Fig. 3.15a, and $M_H(T) \rightarrow M_0(T)$ as $H \rightarrow 0$ with

$$M_0(T) = \begin{cases} 0 & \text{if } T \geq T_c, \\ M_s(T) & \text{if } T < T_c, \end{cases}$$

where $M_s(T)$ is the spontaneous magnetization (see Fig. 3.12).

3.4.4 Asymmetry of Fluctuations

The above discussions suggest that for ferromagnetic systems, there are two possible phase transition behaviors near a critical point, and theoretically each of them has some probability to take place, however only one of them can appear in reality. This phenomena is also observed in the previous section for PVT systems.

One explanation of such phenomena is that the symmetry of fluctuation near a critical point is not generally true in equilibrium phase transitions. To make the statement more clear, we first introduce some concepts.

Let $G(u, \lambda)$ be free energy of a thermodynamic system, $u = (u_1, \dots, u_n)$ be the order parameter, and $\lambda = (\lambda_1, \dots, \lambda_N)$ the control parameter ($n, N \geq 1$). Assume that u is defined in the function space $L^2(\Omega, \mathbb{R}^n)$ and $\lambda \in \mathbb{R}^N$. Then the space

$$X = \{(u, \lambda) | u \in L^2(\Omega, \mathbb{R}^n), \lambda \in \mathbb{R}^N\}$$

is called the state space of the system.

Let $(u_0, \lambda_0) \in X$ be a stable equilibrium state of the system; namely (u_0, λ_0) is a locally minimal state of $G(u, \lambda)$. We say that the system has a fluctuation at (u_0, λ_0) if it deviates randomly from (u_0, λ_0) to $(\tilde{u}, \tilde{\lambda})$ with

$$\|\tilde{u} - u_0\| + |\tilde{\lambda} - \lambda_0| > 0.$$

In this case, $(\tilde{u}, \tilde{\lambda})$ is called a state of fluctuation.

The so-called symmetry of fluctuation means that for given $r > 0$, all states $(\tilde{u}, \tilde{\lambda})$ of fluctuation satisfying

$$\|\tilde{u} - u_0\| + |\tilde{\lambda} - \lambda_0| = r, \quad (\tilde{u}, \tilde{\lambda}) \in X,$$

have the same probability to appear in real world. Otherwise, we say that the fluctuation is asymmetric.

The observations in both the PVT systems and the ferromagnetic systems strongly suggest the following physical conjecture, regarding to the uniqueness of transition behaviors.

Physical Principle 3.4.3 (Asymmetry of Fluctuations) *The symmetry of fluctuations for general thermodynamic systems may not be universally true. In other words, in some systems with multi-equilibrium states, the fluctuations near a critical point occur only in one basin of attraction of some equilibrium states, which are the ones that can be physically observed.*

3.5 Phase Separation in Binary Systems

Materials compounded by two components A and B , such as binary alloys, binary solutions, and polymers, are called binary systems. Sufficient cooling of a binary system may lead to phase separations, i.e., at the critical temperature, the concentrations of both components A and B with homogeneous distribution undergo changes, leading to heterogeneous distributions in space. Phase separation of binary systems observed will be in one of two main ways. The first is by nucleation in which sufficiently large nuclei of the second phase appear randomly and grow, and this corresponds to Type-II phase transitions. The second is by spinodal decomposition in which the systems appear to nuclear at once, and periodic or semi-periodic structure is seen, and this corresponds to second-order phase transitions (Reichl, 1998; Pismen, 2006).

Cahn–Hilliard equation describes the process of phase separation, by which the two components of a binary fluid spontaneously separate and form domains pure in each component. There have been extensive studies in the past on the dynamics of the Cahn–Hilliard equations (Cahn and Hilliard, 1957; Novick-Cohen and Segel, 1984; Bates and Fife, 1993; Shen and Yang, 2010; del Pino et al., 2000).

The main objective of this section is to precisely describe the phase separation mechanism and to make a few physical predictions. The material is based largely on

recent articles (Ma and Wang, 2009b,c). In particular, four physical predictions on the phase separation of binary systems are derived. First, the order of phase transitions is precisely determined by the sign of a nondimensional parameter K such that if $K > 0$, the transition is first order with latent heat and if $K < 0$, the transition is second order. Here the parameter K is defined in terms of the coefficients in the quadratic and cubic nonlinear terms of the Cahn–Hilliard equation and the typical length scale of the container. Second, a phase diagram is derived, characterizing the order of phase transitions, and leading in particular to a prediction that there is only second-order transition for molar fraction near 1/2. This is different from the prediction made by the classical phase diagram. Third, a TL -phase diagram is derived, characterizing the regions of both homogeneous and separation phases and their transitions. Fourth, the study demonstrates again the point, mentioned before for the Bénard convection, that the dynamic transition theory opens a new avenue toward to pattern formation, as the patterns and structure of transition states are dictated by solutions in σ_T , which are again determined by eigenvectors of the corresponding linearized problem (with high-order approximations).

3.5.1 Modeling

Since binary systems are conserved, the equations describing the Helmholtz process and the Gibbs process are the same. Hence, without distinction we use the term “free energy” to discuss this problem. Let u_A and u_B be the concentrations of components A and B respectively, then $u_B = 1 - u_A$. In a homogeneous state, $u_B = \bar{u}_B$ is a constant. We now consider the case where the Helmholtz free energy is given by

$$F(u) = F_0 + \int_{\Omega} \left[\frac{\mu}{2} |\nabla u_B|^2 + f(u_B) \right] dx, \quad (3.5.1)$$

and the density function f is given by the Bragg–Williams potential, also called the Hildebrand theory (Reichl, 1998), as follows:

$$f = \mu_A u_A + \mu_B u_B + RT[u_A \ln u_A + u_B \ln u_B] + au_B u_A, \quad (3.5.2)$$

where μ_A, μ_B are the chemical potentials of A and B respectively, R the molar gas constant, $a > 0$ the measure of repulsion action between A and B .

In a homogeneous state, u_B is a constant in space and is equal to its mean value, i.e., $u_B = \bar{u}_B$. Let

$$u = u_B - \bar{u}_B, \quad u_0 = \bar{u}_B.$$

Using the Landau mean field theory, we have the following Cahn–Hilliard equation

$$\frac{\partial u}{\partial t} = -k\Delta^2 u + \Delta[b_1 u^1 + b_2 u^2 + b_3 u^3], \quad (3.5.3)$$

where the coefficients k, b_1, b_2 , and b_3 are given by

$$k = \mu D,$$

$$\begin{aligned} b_1 &= \frac{D}{2} \frac{d^2 f(u_0)}{du^2} = \left[\frac{RT}{u_0(1-u_0)} - 2a \right] \frac{D}{2} \\ b_2 &= \frac{D}{3!} \frac{d^3 f(u_0)}{du^3} = \frac{2u_0 - 1}{6u_0^2(1-u_0)^2} DRT, \\ b_3 &= \frac{D}{4!} \frac{d^4 f(u_0)}{du^4} = \frac{1 - 3u_0 + 3u_0^2}{12u_0^3(1-u_0)^3} DRT, \end{aligned} \quad (3.5.4)$$

where D is the diffusion coefficient.

Physically sound boundary condition for (3.5.3) is either the Neumann boundary condition:

$$\frac{\partial u}{\partial n} = \frac{\partial \Delta u}{\partial n} = 0 \quad \text{on } \partial\Omega, \quad (3.5.5)$$

with $\Omega \subset \mathbb{R}^n$ ($1 \leq n \leq 3$) being a bounded domain, or the periodic boundary condition:

$$u(x + KL) = u(x) \quad (3.5.6)$$

with $\Omega = [0, L]^n$, $K = (k_1, \dots, k_n)$, $1 \leq n \leq 3$.

In deriving the model (3.5.3), we use the idea of Landau mean field theory to retain only the terms up to the third order. We remark here that the model (3.5.3) is a classical model (Novick-Cohen and Segel, 1984), which has been used by many authors. The expansion used here to derive the coefficients is done near critical points of phase transitions.

In the general case, the order parameters consist of the molar concentration function u_B and the entropy S , and the control parameter contains the temperature T , the pressure p , and the u_0 . We refer the interested readers (Ma and Wang, 2009b) for the modeling and mathematical analysis of the general case.

3.5.2 Phase Transition in General Domains

We now discuss the Cahn–Hilliard equation from the mathematical point of view. We start with the nondimensional form of equation. Let

$$\begin{aligned} x &= lx', & t &= \frac{l^4}{k} t', & u &= u_0 u', \\ \lambda &= -\frac{l^2 b_1}{k}, & \gamma_2 &= \frac{l^2 b_2 u_0}{k}, & \gamma_3 &= \frac{l^2 b_3 u_0^2}{k}, \end{aligned}$$

where l is a given length, $u_0 = \bar{u}_B$ is the constant concentration of B , and $\gamma_3 > 0$. Then the equation (3.5.3) can be rewritten as follows (omitting the primes)

$$\begin{aligned}\frac{\partial u}{\partial t} &= -\Delta^2 u - \lambda \Delta u + \Delta(\gamma_2 u^2 + \gamma_3 u^3), \\ \int_{\Omega} u(x, t) dx &= 0, \\ u(x, 0) &= \varphi.\end{aligned}\tag{3.5.7}$$

Let

$$H = \left\{ u \in L^2(\Omega) \mid \int_{\Omega} u dx = 0 \right\}.$$

For the Neumann boundary condition (3.5.5) we define

$$H_1 = \left\{ u \in H^4(\Omega) \cap H \mid \frac{\partial u}{\partial n} = \frac{\partial \Delta u}{\partial n} = 0 \text{ on } \partial\Omega \right\},$$

and for the periodic boundary condition (3.5.6) we define

$$H_1 = \left\{ u \in H^4(\Omega) \cap H \mid u(x + KL) = u(x) \quad \forall K \in \mathbb{Z}^n \right\}.$$

Then we define the operators $L_\lambda = -A + B_\lambda$ and $G : H_1 \rightarrow H$ by

$$Au = \Delta^2 u, \quad B_\lambda u = -\lambda \Delta u, \quad G(u) = \gamma_2 \Delta u^2 + \gamma_3 \Delta u^3. \tag{3.5.8}$$

Thus, (3.5.7) is equivalent to the following operator equation

$$\frac{du}{dt} = L_\lambda u + G(u), \quad u(0) = \varphi \tag{3.5.9}$$

It is clear that the conditions (2.1.2) and (2.1.3) hold true here.

We first consider the case where $\Omega \subset \mathbb{R}^n$ ($1 \leq n \leq 3$) is a general bounded and smooth domain. Let ρ_k and e_k be the eigenvalues and eigenfunctions of the following eigenvalue problem:

$$-\Delta e_k = \rho_k e_k, \quad \frac{\partial e_k}{\partial n}|_{\partial\Omega} = 0, \quad \int_{\Omega} e_k dx = 0. \tag{3.5.10}$$

The eigenvalues of (3.5.10) satisfy $0 < \rho_1 \leq \rho_2 \leq \dots \leq \rho_k \leq \dots$, and $\lim_{k \rightarrow \infty} \rho_k = \infty$. The eigenfunctions $\{e_k\}$ constitute an orthonormal basis of H . Also, $\{e_k\}$ is an orthogonal basis of H_1 with norm

$$\|u\|_1 = \left[\int_{\Omega} |\Delta^2 u|^2 dx \right]^{1/2}.$$

We are now in a position to give a phase transition theorem for the problem (3.5.7) with the following Neumann boundary condition (3.5.5):

Theorem 3.5.1 *Assume that $\gamma_2 = 0$ and $\gamma_3 > 0$ in (3.5.7), then the following assertions hold true:*

1. If the first eigenvalue ρ_1 of (3.5.10) has multiplicity $m \geq 1$, then the problem (3.5.7) with (3.5.5) bifurcates from $(u, \lambda) = (0, \rho_1)$ on $\lambda > \rho_1$ to an attractor Σ_λ , homeomorphic to an $(m - 1)$ -dimensional sphere S^{m-1} , and Σ_λ attracts $H \setminus \Gamma$, where Γ is the stable manifold of $u = 0$ with codimension m .
2. Σ_λ contains at least $2m$ singular points. If $m = 1$, Σ_λ has exactly two steady states $\pm u_\lambda$, and if $m = 2$, $\Sigma_\lambda = S^1$ has at most eight singular points.
3. Each singular point u_λ in Σ_λ can be expressed as

$$u_\lambda = \left(\frac{\lambda - \rho_1}{\gamma_3} \right)^{1/2} e_1 + o(|\lambda - \rho_1|^{1/2}),$$

where e_1 is an eigenfunction corresponding to the first eigenvalue of (3.5.10).

Proof. We proceed in several steps as follows.

STEP 1. It is clear that the eigenfunction $\{e_k\}$ of (3.5.10) are also eigenvectors of the linear operator $L_\lambda = -A + B_\lambda$ defined by (3.5.8) and the eigenvalues of L_λ are given by

$$\beta_k(\lambda) = \rho_k(\lambda - \rho_k), \quad k = 1, 2, \dots. \quad (3.5.11)$$

It is easy to verify the conditions (2.1.4) and (2.1.5) in our case at $\lambda_0 = \rho_1$. We shall prove this theorem using the attractor bifurcation theory introduced in Sect. 2.2.

We need to verify that $u = 0$ is a global asymptotically stable singular point of (3.5.9) at $\lambda = \rho_1$. By $\gamma_2 = 0$, from the energy integration of (3.5.7) we can obtain

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} \int_{\Omega} u^2 dx &= \int_{\Omega} [-|\Delta u|^2 + \rho_1 |\nabla u|^2 - 3\gamma_3 u^2 |\nabla u|^2] dx \\ &\leq -C \int_{\Omega} |\Delta v|^2 dx - 3\gamma_3 \int_{\Omega} u^2 |\nabla u|^2 dx, \end{aligned} \quad (3.5.12)$$

where $C > 0$ is a constant, $u = v + e_1$, and $\int_{\Omega} ve_1 dx = 0$. It follows from (3.5.12) that $u = 0$ is global asymptotically stable. Hence, for Assertion (1), we only have to prove that Σ_λ is homeomorphic to S^{m-1} , as the rest of this assertion follows directly from Theorem 2.2.11.

STEP 2. Now we prove that the bifurcated attractor Σ_λ from $(0, \rho_1)$ contains at least $2m$ singular points.

Let $g(u) = -\Delta u - \lambda u + \gamma_3 u^3$. Then the stationary equation of (3.5.7) is given by

$$\Delta g(u) = 0, \quad \int_{\Omega} u dx = 0,$$

which is equivalent to

$$-\Delta u - \lambda u + \gamma_3 u^3 = \text{constant}, \quad \int_{\Omega} u dx = 0, \quad \frac{\partial u}{\partial n}|_{\partial\Omega} = 0. \quad (3.5.13)$$

By the Lagrange multiplier theorem, (3.5.13) is the Euler equation of the following functional with zero average constraint:

$$F(u) = \int_{\Omega} \left[\frac{1}{2} |\nabla u|^2 - \frac{\lambda}{2} u^2 + \frac{\gamma_3}{4} u^4 \right] dx, \quad (3.5.14)$$

$$u \in \left\{ u \in H^1(\Omega) \cap H \mid \frac{\partial u}{\partial n}|_{\partial\Omega} = 0 \quad \int_{\Omega} u dx = 0 \right\}.$$

Since F is an even functional, by the classical Krasnoselskii bifurcation theorem for even functionals, (3.5.14) bifurcates from $\lambda > \rho_1$ at least to $2m$ mini-max points, i.e., equation (3.5.13) has at least $2m$ bifurcated solutions on $\lambda > \rho_1$. Hence, the attractor Σ_λ contains at least $2m$ singular points.

STEP 3. To complete the proof, we reduce the equation (3.5.9) to the center manifold near $\lambda = \rho_1$. By the center manifold reduction (1.2.9), the reduced equation of (3.5.9) is given by:

$$\frac{dx_i}{dt} = \beta_1(\lambda)x_i - \gamma_3\rho_1 \int_{\Omega} v^3 e_{1i} dx + o(|x|^3) \quad \text{for } 1 \leq i \leq m, \quad (3.5.15)$$

where $\beta_1(\lambda) = \rho_1(\lambda - \rho_1)$, $v = \sum_{i=1}^m x_i e_{1i}$, and $\{e_{11}, \dots, e_{1m}\}$ are the first eigenfunctions of (3.5.10). Equations (3.5.15) can be rewritten as

$$\frac{dx_i}{dt} = \beta_1(\lambda)x_i - \gamma_3\rho_1 \int_{\Omega} \left(\sum_{j=1}^m x_j e_{1j} \right)^3 e_{1i} dx + o(|x|^3). \quad (3.5.16)$$

Let

$$g(x) = \left(\int_{\Omega} v^3 e_{11} dx, \dots, \int_{\Omega} v^3 e_{1m} dx \right), \quad v(x) = \sum_{j=1}^m x_j e_{1j}.$$

Then there is a constant $C > 0$ such that for any $x \in \mathbb{R}^m$,

$$(g(x), x) = \sum_{i=1}^m x_i \int_{\Omega} v^3 e_{1i} dx = \int_{\Omega} v^4 dx \geq C|x|^4. \quad (3.5.17)$$

Thus by Theorem 2.2.5, it follows from (3.5.16) and (3.5.17) that the attractor Σ_λ is homeomorphic to S^{m-1} . Hence, Assertion (1) is proved. Assertions (2) and (3) can then be derived from (3.5.16) and (3.5.17), and the proof is complete. \square

Physically, the coefficients γ_2 and γ_3 depend on $u_0 = \bar{u}_B$, the temperature T , and the pressure p :

$$\gamma_k = \gamma_k(u_0, T, p), \quad k = 2, 3.$$

The set of points satisfying $\gamma_2(u_0, T, p) = 0$ has measure zero in $(u_0, T, p) \in \mathbb{R}^3$. Hence, it is more interesting to consider the case where $\gamma_2 \neq 0$.

For this purpose, let the multiplicity of the first eigenvalue ρ_1 of (3.5.10) be $m \geq 1$, and $\{e_1, \dots, e_m\}$ be the first eigenfunctions. Let $a_{ij}^k = \int_{\Omega} e_i e_j e_k dx$ and consider the following quadratic equations:

$$\sum_{i,j=1}^m a_{ij}^k x_i x_j = 0 \quad \text{for } 1 \leq k \leq m. \quad (3.5.18)$$

Theorem 3.5.2 Let $\gamma_2 \neq 0$, $\gamma_3 > 0$, and $x = 0$ be an isolated singular point of (3.5.18). Then the phase transition of (3.5.7) and (3.5.5) is either Type-II or Type-III. Furthermore, the problem (3.5.7) with (3.5.5) bifurcates to at least one singular point on each side of $\lambda = \rho_1$, and has a saddle-node bifurcation on $\lambda < \rho_1$. In particular, if $m = 1$, then the following assertions hold true:

1. The phase transition is Type-III, and a neighborhood $U \subset H$ of $u = 0$ can be decomposed into two sectorial regions $\overline{U} = \overline{D}_1(\pi) + \overline{D}_2(\pi)$ such that the phase transition in $D_1(\pi)$ is the first order, and in $D_2(\pi)$ is the n -th order with $n \geq 3$.
2. The bifurcated singular point u_λ on $\lambda > \rho_1$ attracts $D_2(\pi)$, which can be expressed as

$$u_\lambda = (\lambda - \rho_1)e_1/\gamma_2 a + o(|\lambda - \rho_1|), \quad (3.5.19)$$

where, by assumption, $a = \int_{\Omega} e_1^3 dx \neq 0$.

3. When $|\gamma_2 a| = \varepsilon$ is small, the assertions in the transition perturbation theorems (Theorems 2.6.3 and 2.6.4) hold true.

Remark 3.5.3 We shall see later that when $\Omega = \prod_{k=1}^n (0, L_k) \subset \mathbb{R}^n$ is a rectangular domain, then $a = \int_{\Omega} e_1^3 dx = 0$. However, for almost all non-rectangular domains Ω , the first eigenvalues are simple and $a \neq 0$. Hence, the Type-III phase transitions for general domains are generic.

Proof of Theorem 3.5.2. Assertions (1)–(3) can be directly proved using Theorems 2.3.2, 2.6.3, and 2.6.4. By assumption, $u = 0$ is a second-order nondegenerate singular point of (3.5.9) at $\lambda = \rho_1$, which implies that $u = 0$ is not locally asymptotically stable. Hence, it follows from Theorem 2.1.6 and Theorem 4.1 in Ma and Wang (2005b) that the phase transition of (3.5.7) with (3.5.5) is either Type-II or Type-III, and there is at least one singular point bifurcated on each side of $\lambda = \rho_1$.

Finally, we shall apply Theorem 2.5.3 to prove that there exists a saddle-node bifurcation on $\lambda < \rho_1$. It is known that

$$\text{ind}(L_\lambda + G, 0) = \begin{cases} \text{even} & \text{if } \lambda = \rho_1, \\ 1 & \text{if } \lambda < \rho_1. \end{cases}$$

Moreover, since $L_\lambda + G$ defined by (3.5.8) is a gradient-type operator, we have

$$\text{ind}(L_{\rho_1} + G, 0) \leq 0.$$

Hence there is a bifurcated branch Σ_λ on $\lambda < \rho_1$ such that

$$\text{ind}(L_\lambda + G, u_\lambda) = -1 \quad \forall u_\lambda \in \Sigma_\lambda, \quad \lambda < \rho_1.$$

It is clear that the eigenvalues (3.5.11) of L_λ satisfy (2.1.4) and (2.1.5), and for any $\lambda \in \mathbb{R}^1$ (3.5.9) possesses a global attractor. Therefore, for bounded λ , $b < \lambda \leq \rho_1$,

the bifurcated branch Σ_λ is bounded:

$$\|u_\lambda\| \leq C \quad \forall u_\lambda \in \Sigma_\lambda, \quad -\infty < b < \lambda \leq \rho_1.$$

We need to prove that there exists $\tilde{\lambda} < \rho_1$ such that for all $\lambda < \tilde{\lambda}$ equation (3.5.9) has no nonzero singular points.

By the energy estimates of (3.5.9), for any $\lambda < \tilde{\lambda} = -\gamma_2^2/2\gamma_3$ and $u \neq 0$ in H ,

$$\begin{aligned} & \int_{\Omega} [|\Delta u|^2 - \lambda |\nabla u|^2 + 2\gamma_2 u |\nabla u|^2 + 3\gamma_3 u^2 |\nabla u|^2] dx \\ & \geq \int_{\Omega} |\Delta u|^2 dx + \int_{\Omega} |\nabla u|^2 (-\lambda - 2|\gamma_2 u| + 3\gamma_3 u^2) dx \\ & \geq \int_{\Omega} |\nabla u|^2 dx + \int_{\Omega} |\nabla u|^2 (-\lambda + \gamma_3 u^2 + 2\gamma_3(u - \frac{\gamma_2}{\gamma_3})^2 - \frac{\gamma_2^2}{2\gamma_3}) dx \\ & \geq \int_{\Omega} |\nabla u|^2 dx + \int_{\Omega} |\nabla u|^2 (-\lambda - \frac{\gamma_2^2}{2\gamma_3}) dx \\ & > 0. \end{aligned}$$

Therefore, when $\lambda < \tilde{\lambda}$, (3.5.9) has no nontrivial singular points in H . Thus we infer from Theorem 2.5.3 that there exists a saddle-node bifurcation on $\lambda < \rho_1$. The proof is complete. \square

3.5.3 Phase Transition in Rectangular Domains

The dynamical properties of phase separation of a binary system in a rectangular container is very different from that in a general container. We see in the previous section that the phase transitions in general domains are Type-III, and we shall show in the following that the phase transitions in rectangular domains are either Type-I or Type-II, which are distinguished by a critical size of the domains.

Let $\Omega = \prod_{k=1}^n (0, L_k) \subset \mathbb{R}^n$ ($1 \leq n \leq 3$) be a rectangular domain. We first consider the case where

$$L = L_1 > L_j \quad \forall 2 \leq j \leq n. \quad (3.5.20)$$

Theorem 3.5.4 *Let $\Omega = \prod_{k=1}^n (0, L_k)$ satisfy (3.5.20). The following assertions hold true:*

1. If

$$\gamma_3 < \frac{2L^2}{9\pi^2} \gamma_2^2,$$

then the phase transition of (3.5.7) and (3.5.5) at $\lambda = \lambda_0 = \pi^2/L^2$ is Type-II. In particular, the problem (3.5.7) with (3.5.5) bifurcates from $(u, \lambda) = (0, \pi^2/L^2)$ on

$\lambda < \pi^2/L^2$ to exactly two equilibrium points which are saddles, and there are two saddle-node bifurcations on $\lambda < \pi^2/L^2$ as shown in Fig. 3.16.

2. If

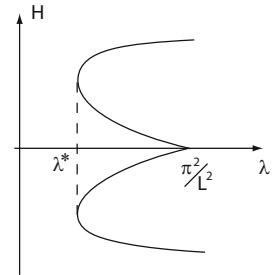
$$\gamma_3 > \frac{2L^2}{9\pi^2} \gamma_2^2,$$

then the transition is Type-I. In particular, the problem bifurcates on $\lambda > \pi^2/L^2$ to exactly two attractors u_1^T and u_2^T which can be expressed as

$$u_{1,2}^T = \pm \frac{\sqrt{2}(\lambda - \frac{\pi^2}{L^2})^{1/2}}{\sigma^{1/2}} \cos \frac{\pi x_1}{L} + o(|\lambda - \frac{\pi^2}{L^2}|^{1/2}), \quad (3.5.21)$$

where $\sigma = \frac{3\gamma_3}{2} - \frac{L^2\gamma_2^2}{3\pi^2}$.

Fig. 3.16 Type-II transition as given by Theorem 3.5.4



Proof. With the spatial domain as given, the first eigenvalue and eigenfunction of (3.5.10) are given by

$$\rho_1 = \pi^2/L^2, \quad e_1 = \cos \frac{\pi x_1}{L}.$$

The eigenvalues and eigenfunctions of $L_\lambda = -A + B_\lambda$ defined by (3.5.8) are as follows:

$$\beta_K = |K|^2(\lambda - |K|^2), \quad (3.5.22)$$

$$e_K = \cos \frac{k_1 \pi x_1}{L_1} \cdots \cos \frac{k_n \pi x_n}{L_n}, \quad (3.5.23)$$

where $k_i \in \mathbb{Z}$,

$$K = \left(\frac{k_1 \pi}{L_1}, \dots, \frac{k_n \pi}{L_n} \right), \quad |K|^2 = \pi^2 \sum_{i=1}^n k_i^2 / L_i^2, \quad |K|^2 \neq 0.$$

By Theorem A.1.1, the reduced equation of (3.5.9) to the center manifold is

$$\frac{dy}{dt} = \beta_1(\lambda)y - \frac{2\pi^2}{(L_1 \cdots L_n)L_1^2} \int_{\Omega} [\gamma_3 y^3 e_1^4 + \gamma_2(y e_1 + \Phi(y))^2 e_1] dx, \quad (3.5.24)$$

where $y \in \mathbb{R}^1$, $\Phi(y)$ is the center manifold function, and

$$\beta_1(\lambda) = \frac{\pi^2}{L^2}(\lambda - \frac{\pi^2}{L^2}). \quad (3.5.25)$$

Direct calculation implies that

$$\int_{\Omega} e_1^3 dx = \int_0^{L_1} \cdots \int_0^{L_n} \cos^3 \frac{\pi x_1}{L} dx = 0, \quad (3.5.26)$$

$$\int_{\Omega} e_1^4 dx = L_2 \cdots L_n \int_0^L \cos^4 \frac{\pi x_1}{L} dx_1 = \frac{3}{8} L_1 \cdots L_n. \quad (3.5.27)$$

By (3.5.26) and $\Phi(y) = O(|y|^2)$ we obtain

$$\int_{\Omega} (ye_1 + \Phi)^2 e_1 dx = 2y \int_{\Omega} \Phi(y) e_1^2 dx + o(|y|^3). \quad (3.5.28)$$

It follows from Theorem A.1.1 that

$$\begin{aligned} \Phi(y) &= \sum_{|K|^2 > \pi^2/L^2}^{\infty} \phi_K(y) e_K + o(|y|^2), \\ \phi_K(y) &= \frac{\gamma_2 y^2}{-\beta_K \|e_K\|^2} \int_{\Omega} \Delta e_1^2 \cdot e_K dx = \frac{|K|^2 \gamma_2 y^2}{\beta_K \|e_K\|^2} \int_{\Omega} e_K e_1^2 dx. \end{aligned}$$

Notice that

$$\int_{\Omega} e_K e_1^2 dx = \begin{cases} 0 & \forall K \neq \left(\frac{2\pi}{L_1}, 0, \dots, 0 \right), \\ L_1 \cdots L_n / 4 & \forall K = \left(\frac{2\pi}{L_1}, 0, \dots, 0 \right). \end{cases}$$

Then we have

$$\Phi(y) = \frac{\gamma_2 y^2}{2(\lambda - 4\pi^2/L^2)} \cos \frac{2\pi x_1}{L} + o(|y|^2).$$

Inserting $\Phi(y)$ into (3.5.28) we find

$$\begin{aligned} \int_{\Omega} (ye_1 + \Phi)^2 e_1 dx &= \frac{\gamma_2 y^3}{\lambda - \frac{4\pi^2}{L^2}} \int_{\Omega} \cos \frac{2\pi x_1}{L} e_1^2 dx + o(|y|^3) \\ &= \frac{L_1 \cdots L_n}{4} \cdot \frac{\gamma_2}{\lambda - \frac{4\pi^2}{L^2}} y^3 + o(|y|^3). \end{aligned} \quad (3.5.29)$$

Finally, by (3.5.26), (3.5.27), and (3.5.29), we derive from (3.5.24) the following reduced equation of (3.5.9):

$$\frac{dy}{dt} = \beta_1(\lambda)y - \frac{\pi^2}{2L^2} \left(\frac{3\gamma_3}{2} + \frac{\gamma_2^2}{\lambda - \frac{4\pi^2}{L^2}} \right) y^3 + o(|y|^3). \quad (3.5.30)$$

Near the critical point $\lambda_0 = \pi^2/L^2$, the coefficient

$$\frac{3\gamma_3}{2} + \frac{\gamma_2^2}{\lambda - \frac{4\pi^2}{L^2}} = \frac{3\gamma_3}{2} - \frac{L^2\gamma_2^2}{3\pi^2}.$$

Thus, by Theorem 2.1.5 we derive from (3.5.30) the assertions of the theorem except the claim for the saddle-node bifurcation in Assertion (1), which can be proved in the same fashion as used in Theorem 3.5.2. The proof is complete. \square

We now consider the case where $\Omega = \prod_{k=1}^n (0, L_k)$ satisfies that

$$L = L_1 = \dots = L_m > L_j \quad \text{for } 2 \leq m \leq 3, m < j \leq n. \quad (3.5.31)$$

Theorem 3.5.5 *Let $\Omega = \prod_{k=1}^n (0, L_k)$ satisfy (3.5.31). Then the following assertions hold true:*

1. If

$$\gamma_3 > \frac{26L^2}{27\pi^2}\gamma_2^2,$$

then the phase transition of the problem (3.5.7) with (3.5.5) at $\lambda_0 = \pi^2/L^2$ is Type-I, satisfying the following properties:

- a. The problem bifurcates on $\lambda > \pi^2/L^2$ to an attractor Σ_λ , containing exactly $3^m - 1$ nondegenerate singular points, and Σ_λ is homeomorphic to an $(m-1)$ -dimensional sphere S^{m-1} .
- b. For $m = 2$, the attractor $\Sigma_\lambda = S^1$ contains four minimal attractors, as shown in Fig. 3.17.
- c. For $m = 3$, $\Sigma_\lambda = S^2$ contains 8 minimal attractors as shown in Fig. 3.18a, if

$$\gamma_3 < \frac{22L^2}{9\pi^2}\gamma_2^2,$$

and contains 6 minimal attractors as shown in Fig. 3.18b if

$$\gamma_3 > \frac{22L^2}{9\pi^2}\gamma_2^2.$$

2. If

$$\gamma_3 < \frac{26L^2}{27\pi^2}\gamma_2^2,$$

then the transition is Type-II. In particular, the problem has a saddle-node bifurcation on $\lambda < \lambda_0 = \pi^2/L^2$, and bifurcates on both side of $\lambda = \lambda_0$ to exactly $3^m - 1$ singular points which are nondegenerate.

Fig. 3.17 For $m = 2$, $\Sigma_\lambda = S^1$ and Z_{2k} ($1 \leq k \leq 4$) are attractors

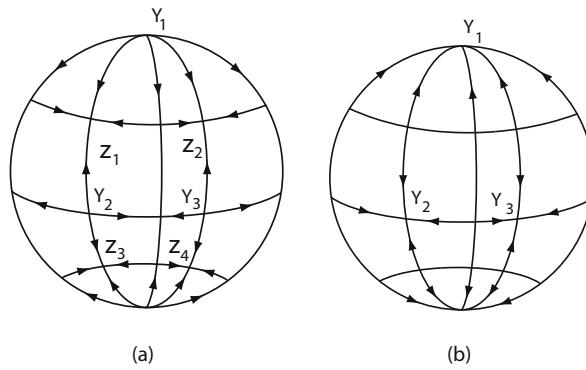
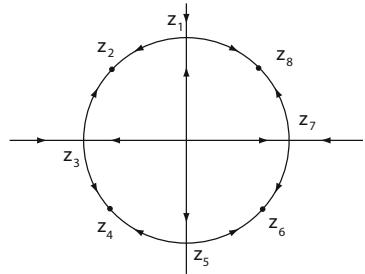


Fig. 3.18 For $m = 3$, $\Sigma_\lambda = S^2$, (a) $\pm Z_k$ ($1 \leq k \leq 4$) are attractors, (b) $\pm Y_k$ ($1 \leq k \leq 3$) are attractors

Proof. We proceed in several steps as follows.

STEP 1. Consider the center manifold reduction. It is known that the eigenvalues and eigenfunctions of $L_\lambda = -A + B_\lambda$ are given by (3.5.22) and (3.5.23) with $L_1 = \dots = L_m$. As before, the reduced equations of (3.5.9) are given by

$$\frac{dy}{dt} = \beta_1(\lambda)y + g(y) + o(|y|^3), \quad (3.5.32)$$

where $y = (y_1, \dots, y_m) \in \mathbb{R}^m$, $\beta_1(\lambda)$ is as in (3.5.25), and

$$\begin{aligned} g(y) &= \frac{2\pi^2}{L_1 \cdots L_n L_1^2} (G_3(y) + G_{23}(y)), \\ G_3(y) &= -\gamma_3 \left(\int_{\Omega} v^3 e_1 dx, \dots, \int_{\Omega} v^3 e_m dx \right), \\ G_{23}(y) &= -\gamma_2 \left(\int_{\Omega} u^2 e_1 dx, \dots, \int_{\Omega} u^2 e_m dx \right). \end{aligned}$$

Here $e_i = \cos \pi x_i / L$ for $1 \leq i \leq m$, L is given by (3.5.31), $v = \sum_{i=1}^m y_i e_i$, $u = v + \Phi(y)$ and Φ is the center manifold function. Direct computation shows that

$$\begin{aligned}\int_{\Omega} v^2 e_i dx &= \int_{\Omega} \left(\sum_{j=1}^m y_j \cos \pi x_j / L \right)^2 \cos \pi x_i / L dx = 0, \\ \int_{\Omega} v^3 e_i dx &= \int_{\Omega} \left(\sum_{j=1}^n y_k \cos \pi x_j / L \right)^3 \cos \pi x_i / L dx \\ &= \frac{3}{4} L_1 \cdots L_n \left(\frac{1}{2} y_i^3 + y_i \sum_{j \neq i} y_j^2 \right),\end{aligned}\tag{3.5.33}$$

$$\begin{aligned}\int_{\Omega} u^2 e_i dx &= \int_{\Omega} \left(\sum_{j=1}^m y_j e_j + \Phi(y) \right)^2 e_i dx \\ &= 2 \sum_{j=1}^m y_j \int_{\Omega} \Phi(y) e_j e_i dx + \int_{\Omega} \Phi^2(y) e_i dx.\end{aligned}\tag{3.5.34}$$

We need to compute the center manifold function $\Phi(y)$. By Theorem A.1.1, we have

$$\Phi(y) = \sum_{|K|>\pi^2/L^2}^{\infty} \phi_K(y) e_K + o(|y|^2) + O(|y|^2 |\beta_1|),\tag{3.5.35}$$

where

$$\begin{aligned}\phi_K(y) &= \frac{\gamma_2}{-\beta_K(\lambda)(e_K, e_K)} \int_{\Omega} \Delta v^2 e_K dx \\ &= \frac{|K|^2 \gamma_2}{\beta_K(\lambda)(e_K, e_K)} \int_{\Omega} v^2 e_K dx \\ &= \frac{|K|^2 \gamma_2}{\beta_K(e_K, e_K)} \sum_{i,j=1}^m y_i y_j \int_{\Omega} e_i e_j e_K dx.\end{aligned}$$

It is clear that

$$\int_{\Omega} e_i e_j e_K dx = 0 \quad \text{if } K \neq K_i + K_j, \quad K_i = \left(\frac{\pi \delta_{1i}}{L_1}, \dots, \frac{\pi \delta_{ni}}{L_n} \right).$$

By (3.5.22) and (3.5.23) we have

$$\phi_K(y) = \frac{\gamma_2 y_i y_j (2 - \delta_{ij})}{(\lambda - |K|^2)(e_K, e_K)} \int_{\Omega} e_i e_j e_K dx = \begin{cases} \frac{\gamma_2 y_i^2}{2(\lambda - \frac{4\pi^2}{L^2})} & \text{if } i = j, \\ \frac{2\gamma_2 y_i y_j}{(\lambda - \frac{2\pi^2}{L^2})} & \text{if } i \neq j,\end{cases}$$

for $K = K_i + K_j$. Thus, by (3.5.35), we obtain

$$\Phi(y) = \sum_{i=1}^m \frac{\gamma_2}{2(\lambda - \frac{4\pi^2}{L^2})} y_i^2 \cos \frac{2\pi x_i}{L} + \sum_{l>r}^m \frac{2\gamma_2}{(\lambda - \frac{2\pi^2}{L^2})} y_l y_r \cos \frac{\pi x_l}{L} \cos \frac{\pi x_r}{L}.$$

Inserting $\Phi(y)$ into (3.5.34) we derive

$$\int_{\Omega} u^2 e_i dx = \frac{\gamma_2}{\lambda - \frac{4\pi^2}{L^2}} \int_{\Omega} \left[y_i^3 \cos^2 \frac{\pi x_i}{L} \cos \frac{2\pi x_i}{L} + 4y_i \sum_{j \neq i} y_j^2 \cos^2 \frac{\pi x_j}{L} \cos^2 \frac{\pi x_i}{L} \right] dx + o(|y|^3).$$

Direct computation gives that

$$\int_{\Omega} u^2 e_i dx = \frac{\gamma_2 L_1 \cdots L_n}{4} \left[\frac{y_i^3}{\lambda - \frac{4\pi^2}{L^2}} + \frac{4y_i}{\lambda - \frac{2\pi^2}{L^2}} \sum_{j \neq i} y_j^2 \right] + o(|y|^3). \quad (3.5.36)$$

Putting (3.5.33) and (3.5.36) into (3.5.32), we obtain the following reduced equations:

$$\frac{dy_i}{dt} = \beta_1(\lambda) y_i - \frac{\pi^2}{2L^2} \left[\sigma_1 y_i^3 + \sigma_2 y_i \sum_{j \neq i} y_j^2 \right] + o(|y|^3) \quad \forall 1 \leq i \leq m, \quad (3.5.37)$$

where

$$\sigma_1 = \frac{3\gamma_3}{2} + \frac{\gamma_2^2}{\lambda - \frac{4\pi^2}{L^2}}, \quad \sigma_2 = 3\gamma_3 + \frac{4\gamma_2^2}{\lambda - \frac{2\pi^2}{L^2}}. \quad (3.5.38)$$

STEP 2. It is known that the transition type of (3.5.9) at the critical point $\lambda_0 = \pi^2/L^2$ is completely determined by (3.5.37), i.e., by the following equations

$$\frac{dy_i}{dt} = -\frac{\pi^2}{2L^2} \left[\sigma_1^0 y_i^3 + \sigma_2^0 y_i \sum_{k \neq i} y_k^2 \right] \quad \forall 1 \leq i \leq m, \quad (3.5.39)$$

where

$$\sigma_1^0 = \frac{3\gamma_3}{2} - \frac{L^2 \gamma_2^2}{3\pi^2}, \quad \sigma_2^0 = 3\gamma_3 - \frac{4L^2 \gamma_2^2}{\pi^2}.$$

It is easy to see that

$$\begin{aligned} \sigma_1^0 + \sigma_2^0 > 0 &\Leftrightarrow \gamma_3 > \frac{26L^2}{27\pi^2} \gamma_2^2, \\ \sigma_1^0 + \sigma_2^0 < 0 &\Leftrightarrow \gamma_3 < \frac{26L^2}{27\pi^2} \gamma_2^2. \end{aligned} \quad (3.5.40)$$

STEP 3. We consider the case where $m = 2$. Thus, the transition type of (3.5.39) is equivalent to that of the following equations

$$\begin{aligned}\frac{dy_1}{dt} &= -y_1[\sigma_1^0 y_1^2 + \sigma_2^0 y_2^2], \\ \frac{dy_2}{dt} &= -y_2[\sigma_1^0 y_2^2 + \sigma_2^0 y_1^2].\end{aligned}\quad (3.5.41)$$

On the straight lines

$$y_1^2 = y_2^2, \quad (3.5.42)$$

equations (3.5.41) satisfy that

$$\frac{dy_2}{dy_1} = \frac{y_2}{y_1} \quad \text{for } \sigma_1^0 + \sigma_2^0 \neq 0, \quad (y_1, y_2) \neq 0.$$

Hence the straight lines (3.5.42) are orbits of (3.5.41) if $\sigma_1^0 + \sigma_2^0 \neq 0$. Obviously, the straight lines

$$y_1 = 0 \text{ and } y_2 = 0 \quad (3.5.43)$$

are also orbits of (3.5.41).

There are four straight lines determined by (3.5.42) and (3.5.43), and each of them contains two orbits. Hence, the system (3.5.41) has at least eight straight line orbits. Hence it is not hard to see that the number of straight line orbits of (3.5.41), if finite, is eight.

Since (3.5.9) is a gradient-type equation, by Lemma A.2.7, there are no elliptic regions at $y = 0$. Hence, when $\sigma_1^0 + \sigma_2^0 > 0$ all the straight line orbits on (3.5.42) and (3.5.43) tend to $y = 0$, as shown in Fig. 3.19a, which implies that the regions are parabolic and stable, therefore $y = 0$ is asymptotically stable for (3.5.41). Accordingly, by Theorem 2.2.2, the transition of (3.5.37) at $\lambda_0 = \pi^2/L^2$ is Type-I.

When $\sigma_1^0 + \sigma_2^0 < 0$ and $\sigma_1^0 > 0$, namely

$$\frac{2}{9} \frac{L^2 \gamma_2^2}{\pi^2} < \gamma_3 < \frac{26}{27} \frac{L^2 \gamma_2^2}{\pi^2},$$

the four straight line orbits on (3.5.42) are outward from $y = 0$, and other four on (3.5.43) are toward $y = 0$, as shown in Fig. 3.19b, which implies that all regions at $y = 0$ are hyperbolic. Hence, by Theorem 2.1.6, the transition of (3.5.37) at $\lambda_0 = \pi^2/L^2$ is Type-II.

When $\sigma_1^0 \leq 0$, then $\sigma_2^0 < 0$ too. In this case, no orbits of (3.5.41) are toward $y = 0$, as shown in Fig. 3.19c, which implies by Theorem 2.1.6 that the transition is Type-II.

Thus by (3.5.40), for $m = 2$ we prove that the transition is Type-I if $\gamma_3 > \frac{26L^2}{27\pi^2} \gamma_2^2$, and Type-II if $\gamma_3 < \frac{26L^2}{27\pi^2} \gamma_2^2$.

STEP 4. Consider the case where $m = 3$. Thus, (3.5.39) are written as

$$\begin{aligned}\frac{dy_1}{dt} &= -y_1[\sigma_1^0 y_1^2 + \sigma_2^0(y_2^2 + y_3^2)], \\ \frac{dy_2}{dt} &= -y_2[\sigma_1^0 y_2^2 + \sigma_2^0(y_1^2 + y_3^2)], \\ \frac{dy_3}{dt} &= -y_3[\sigma_1^0 y_3^2 + \sigma_2^0(y_1^2 + y_2^2)].\end{aligned}\quad (3.5.44)$$

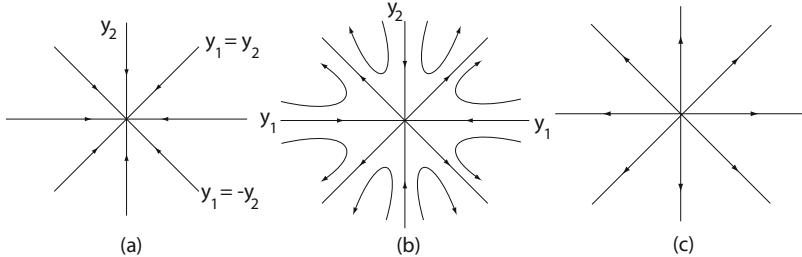


Fig. 3.19 The topological structure of flows of (3.5.37) at $\lambda_0 = \pi^2/L^2$, (a) as $\gamma_3 > \frac{26L^2}{27\pi^2}\gamma_2^2$; (b) as $\frac{2L^2}{9\pi^2}\gamma_2^2 < \gamma_3 < \frac{26L^2}{27\pi^2}\gamma_2^2$; and ((c)) $\gamma_3 \leq \frac{2L^2}{9\pi^2}\gamma_2^2$

It is clear that the straight lines

$$y_i = 0, \quad y_j = 0 \quad \text{for } i \neq j, 1 \leq i, j \leq 3, \quad (3.5.45)$$

$$\begin{cases} y_i^2 = y_j^2, \quad y_k = 0 & \text{for } i \neq j, i \neq k, j \neq k, 1 \leq i, j, k \leq 3, \\ y_1^2 = y_2^2 = y_3^2, \end{cases} \quad (3.5.46)$$

consist of orbits of (3.5.44). There are totally 13 straight lines in (3.5.45) and (3.5.46), each of which consists of two orbits. Thus, (3.5.44) has at least 26 straight line orbits. We shall show that (3.5.44) has just the straight line orbits given by (3.5.45) and (3.5.46). In fact, we assume that the line

$$y_2 = z_1 y_1, \quad y_3 = z_2 y_1 \quad (z_1, z_2 \text{ are real numbers})$$

is a straight line orbit of (3.5.44). Then z_1, z_2 satisfy

$$\begin{aligned}\frac{dy_2}{dy_1} &= z_1 = z_1 \frac{\sigma_1^0 z_1^2 + \sigma_2^0(1 + z_2^2)}{\sigma_1^0 + \sigma_2^0(z_1^2 + z_2^2)}, \\ \frac{dy_3}{dy_1} &= z_2 = z_2 \frac{\sigma_1^0 z_2^2 + \sigma_2^0(1 + z_1^2)}{\sigma_1^0 + \sigma_2^0(z_1^2 + z_2^2)}.\end{aligned}\quad (3.5.47)$$

It is easy to see that when $\sigma_1^0 \neq \sigma_2^0$ the solutions z_1 and z_2 of (3.5.47) take only the values

$$z_1 = 0, \pm 1; \quad z_2 = 0, \pm 1.$$

In the same fashion, we can prove that the straight line orbits of (3.5.44) given by

$$y_1 = \alpha_1 y_3, \quad y_2 = \alpha_2 y_3, \quad \text{and} \quad y_1 = \beta_1 y_2, \quad y_3 = \beta_2 y_2$$

have to satisfy that

$$\alpha_i = 0, \pm 1 \quad \text{and} \quad \beta_i = 0, \pm 1 \quad (i = 1, 2).$$

Thus, we prove that when $\sigma_1^0 \neq \sigma_2^0$, the number of straight line orbits of (3.5.44) is exactly 26.

When $\sigma_1^0 = \sigma_2^0$, we have that $\gamma_3 = \frac{22}{9} \frac{L^2}{\pi^2} \gamma_2^2$ which implies that $\sigma_1^0 = \sigma_2^0 > 0$. In this case, it is clear that $y = 0$ is an asymptotically stable singular point of (3.5.44). Hence, the transition of (3.5.37) at $\lambda_0 = \pi^2/L^2$ is I-type.

When $\sigma_1^0 + \sigma_2^0 > 0$ and $\sigma_1^0 \neq \sigma_2^0$, all straight line orbits of (3.5.44) are toward $y = 0$, which implies that the regions at $y = 0$, are stable, and $y = 0$ is asymptotically stable; see Ma and Wang (2005b). Thereby the transition of (3.5.37) is Type-I.

When $\sigma_1^0 + \sigma_2^0 < 0$ with $\sigma_1^0 > 0$, we can see, as in the case of $m = 2$, that the regions at $y = 0$ are hyperbolic, and when $\sigma_1^0 + \sigma_2^0 < 0$ with $\sigma_1 \leq 0$ the regions at $y = 0$ are unstable. Hence, the transition is Type-II.

STEP 5. We prove Assertion (1). By Steps 3 and 4, if $\gamma_3 > \frac{26L^2}{27\pi^2}\gamma_2^2$, the reduced equation (3.5.37) bifurcates on $\lambda > \lambda_0 = \pi^2/L^2$ to an attractor Σ_λ . All bifurcated equilibrium points of (3.5.9) are one to one correspondence to the bifurcated singular points of (3.5.37). Therefore, we only have to consider the stationary equations:

$$\beta_1(\lambda)y_i - \frac{\pi^2}{2L^2}[\sigma_1 y_i^3 + \sigma_2 y_i \sum_{j \neq i} y_j^2] + o(|y|^3) = 0, \quad 1 \leq i \leq m, \quad (3.5.48)$$

where σ_1 and σ_2 are as in (3.5.38).

Consider the following approximative equations of (3.5.48)

$$\beta_1(\lambda)y_i - y_i(a_1 y_i^2 + a_2 \sum_{j \neq i} y_j^2) = 0 \quad \text{for } 1 \leq i \leq m, \quad (3.5.49)$$

where $a_1 = \pi^2 \sigma_1 / 2L^2$, $a_2 = \pi^2 \sigma_2 / 2L^2$. It is clear that each regular bifurcated solution of (3.5.49) corresponds to a regular bifurcated solution of (3.5.48).

We first prove that (3.5.49) has $3^m - 1$ bifurcated solutions on $\lambda > \lambda_0$. For each k ($0 \leq k \leq m-1$), (3.5.49) has $C_m^k \times 2^{m-k}$ solutions as follows:

$$\begin{aligned} y_{j_1} &= 0, \dots, y_{j_k} = 0 && \text{for } 1 \leq j_l \leq m, \\ y_{r_1}^2 &= \dots = y_{r_{m-k}}^2 = \beta_1(a_1 + (m-k-1)a_2)^{-1} && \text{for } r_i \neq j_l. \end{aligned} \quad (3.5.50)$$

Hence, the number of all bifurcated solutions of (3.5.49) is

$$\sum_{k=0}^{m-1} C_m^k \times 2^{m-k} = (2+1)^m - 1 = 3^m - 1.$$

We need to prove that all bifurcated solutions of (3.5.49) are regular. The Jacobian matrix of (3.5.49) is given by

$$Dv = \begin{pmatrix} \beta_1 - h_1(y) & 2a_2y_1y_2 & \cdots & 2a_2y_1y_m \\ 2a_2y_2y_1 & \beta_1 - h_2(y) & \cdots & 2a_2y_2y_m \\ \vdots & \vdots & & \vdots \\ 2a_2y_my_1 & 2a_2y_my_2 & \cdots & \beta_1 - h_m(y) \end{pmatrix}, \quad (3.5.51)$$

where

$$h_i(y) = 3a_1y_i^2 + a_2 \sum_{j \neq i} y_j^2.$$

For the solutions in (3.5.50), without loss of generality, we take

$$\begin{aligned} y_0 &= (y_1^0, \dots, y_m^0), \\ y_i^0 &= 0 \text{ for } 1 \leq i \leq k, \quad y_{k+1}^0 = \cdots = y_m^0 = \beta_1^{1/2}(a_1 + (m - k - 1)a_2)^{-1/2}. \end{aligned}$$

Inserting them into (3.5.51) we find

$$Dv(y_0) = \begin{pmatrix} \beta I_k & 0 \\ 0 & A_{m-k} \end{pmatrix}, \quad (3.5.52)$$

where

$$\begin{aligned} \beta &= \beta_1 \left(1 - \frac{(m - k)a_2}{a_1 + (m - k - 1)a_2} \right) = \frac{(\sigma_1 - \sigma_2)\beta_1}{\sigma_1 + (m - k - 1)\sigma_2}, \\ A_{m-k} &= \begin{pmatrix} \beta_1 - (3a_1 + (m - k - 1)a_2)(y_{k+1}^0)^2 & \cdots & 2a_2(y_{k+1}^0)^2 \\ \vdots & & \vdots \\ 2a_2(y_{k+1}^0)^2 & \cdots & \beta_1 - (3a_1 + (m - k - 1)a_2)(y_{k+1}^0)^2 \end{pmatrix}. \end{aligned}$$

Direct computation shows that

$$\det A_{m-k} = \frac{\pi^{2m}\beta_1^m}{(a_1 + (m - k - 1)a_2)^m L^{2m}} \det \begin{pmatrix} -\sigma_1 & \sigma_2 & \cdots & \sigma_2 \\ \sigma_2 & -\sigma_1 & \cdots & \sigma_2 \\ \vdots & \vdots & & \vdots \\ \sigma_2 & \sigma_2 & \cdots & -\sigma_1 \end{pmatrix},$$

where σ_1 and σ_2 are given by (3.5.38).

Obviously, there are only finite number of $\lambda > \pi^2/L^2$ satisfying

$$\beta(\lambda) = \frac{(\sigma_1(\lambda) - \sigma_2(\lambda))\beta_1(\lambda)}{\sigma_1(\lambda) + (m - k - 1)\sigma_2(\lambda)} = 0, \quad \det A_{m-k}(\lambda) = 0.$$

Hence, for any $\lambda - \pi^2/L^2 > 0$ sufficiently small the Jacobian matrices (3.5.51) at the singular points (3.5.50) are nondegenerate. Thus, the bifurcated solutions of (3.5.49) are regular.

Since all bifurcated singular points of (3.5.7) with (3.5.5) are nondegenerate, and when Σ_λ is restricted on $x_i x_j$ -plane ($1 \leq i, j \leq m$) the singular points are connected by their stable and unstable manifolds. Hence all singular points in Σ_λ are connected by their stable and unstable manifolds. Therefore, Σ_λ must be homeomorphic to a sphere S^{m-1} .

Assertion (1) is proved.

STEP 6. PROOF OF ASSERTIONS (2) AND (3). When $m = 2$, by Step 5, $\Sigma_\lambda = S^1$ contains eight nondegenerate singular points. By a minimal attractor theorem, Theorem 5.6 in Ma and Wang (2005b), 4 singular points must be attractors and the others are repellers, as shown in Fig. 3.17.

When $m = 3$, we take the six singular points

$$\pm Y_1 = (\pm \beta_1 a_1^{-1}, 0, 0), \pm Y_2 = (0, \pm \beta_1 a_1^{-1}, 0), \pm Y_3 = (0, 0, \pm \beta_1 a_1^{-1}).$$

Then the Jacobian matrix (3.5.51) at Y_i ($1 \leq i \leq 3$) is

$$Dv(\pm Y_i) = \begin{pmatrix} \rho_1 & 0 & 0 \\ 0 & \rho_2 & 0 \\ 0 & 0 & \rho_3 \end{pmatrix},$$

where $\rho_j = \beta_1 \left(1 - \frac{\sigma_2}{\sigma_1}\right)$ as $j \neq i$ and $\rho_i = -2\beta_1$. Obviously, as $\sigma_2 < \sigma_1$, $0 < \rho_j$ ($j \neq i$) and $\rho_i < 0$, in this case, $\pm Y_k$ ($1 \leq k \leq 3$) are repellers in $\Sigma_\lambda = S^2$, which implies that Σ_λ contains 8 attractors $\pm z_k$ ($1 \leq k \leq 4$) as shown in Fig. 3.18a. As $\sigma_2 > \sigma_1$, $\rho_j < 0$ ($1 \leq j \leq 3$), the six singular point $\pm Y_k$ ($1 \leq k \leq 3$) are attractors, which implies that Σ_λ contains only six minimal attractors as shown in Fig. 3.18b. Thus Assertion (2) is proved.

The claim for the saddle-node bifurcation in Assertion (3) can be proved by using the same method as in the proof of Theorem 3.5.2, and the claim for the singular point bifurcation can be proved by the same fashion as used in Step 5.

The proof of the theorem is complete. \square

3.5.4 Spatial Geometry, Transitions, and Pattern Formation

We made a few remarks on the effects of the spatial geometry on phase transition and pattern formation.

First, for the domain $\Omega = [0, L]^m \times D \subset \mathbb{R}^n$ ($1 \leq m < n$), where $n \geq 2$ is arbitrary and $D \subset \mathbb{R}^{n-m}$ a bounded open set, Theorems 3.5.4 and 3.5.5 are also valid provided $\pi^2/L^2 < \lambda_1$, where λ_1 is the first eigenvalue of the equation

$$-\Delta e = \lambda e \quad \text{for } x \in D, \quad \frac{\partial e}{\partial n}|_{\partial D} = 0, \quad \int_D e dx = 0.$$

Second, in Theorem 3.5.5, the minimal attractors in the bifurcated attractor Σ_λ can be expressed as

$$u_\lambda = (\lambda - \pi^2/L^2)^{1/2} e + o(|\lambda - \pi^2/L^2|^{1/2}), \quad (3.5.53)$$

where e is a first eigenfunction of (3.5.10). The expression (3.5.53) can be derived from the reduced equation (3.5.37).

The exponent $\beta = 1/2$ in (3.5.53) is called the critical exponent in physics, and is an important index in the phase transition theory in statistical physics. It is interesting to point out that the critical exponent $\beta = 1$ in (3.5.19) is different from these $\beta = 1/2$ appearing in (3.5.21) and (3.5.53). The first one occurs when the container $\Omega \subset \mathbb{R}^3$ is a non-rectangular region, and the second one occurs when Ω is a rectangle or a cube.

Third, when the sample or container Ω is a loop, or a torus, or bulk in size, then the periodic boundary conditions are necessary. Let $\Omega = S^1 \times (r_1, r_2) \subset \mathbb{R}^2$ be a loop domain, $0 < r_1 < r_2$. Then the boundary condition is given by

$$\begin{aligned} u(\theta + 2k\pi, r) &= u(\theta, r) \quad \text{for } 0 \leq \theta \leq 2\pi, r_1 < r < r_2, k \in \mathbb{Z}, \\ \frac{\partial u}{\partial r} &= 0, \quad \frac{\partial^3 u}{\partial r^3} = 0 \quad \text{at } r = r_1, r_2. \end{aligned} \quad (3.5.54)$$

Assume that the gap $r_2 - r_1$ is small in comparison with the mean radius $r_0 = (r_1 + r_2)/2$, or equivalently, by taking scaling $r_2 - r_1 = 1$:

$$r_0 = (r_1 + r_2)/2 \gg 1, \quad r_2 - r_1 = 1. \quad (3.5.55)$$

Then the Laplacian operator can be approximately expressed as

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r_0^2} \frac{\partial^2}{\partial \theta^2}. \quad (3.5.56)$$

With the boundary condition (3.5.54) and the operator (3.5.56), the eigenvalues and eigenfunctions of the linear operator $L_\lambda = -A + B_\lambda$ defined by (3.5.8) are given by

$$\begin{aligned} \beta_K(\lambda) &= K^2(\lambda - K^2), & K^2 &= \frac{k_1^2}{r_0^2} + k_2^2, \\ e_K^1 &= \cos k_1 \theta \cos k_2 \pi(r - r_1), & e_K^2 &= \sin k_1 \theta \cos k_2 \pi(r - r_1). \end{aligned}$$

By (3.5.55), the first eigenvalue of L_λ is

$$\beta_1(\lambda) = \frac{1}{r_0^2}(\lambda - \frac{1}{r_0^2}),$$

which has multiplicity 2, and the first eigenfunctions are $e_1^1 = \cos \theta$ and $e_1^2 = \sin \theta$.

Theorem 3.5.6 *Let $\Omega = S^1 \times (r_1, r_2)$ satisfy (3.5.55). Then the following assertions hold true:*

1. If

$$\gamma_3 > \frac{2r_0^2}{9}\gamma_2^2,$$

then the phase transition of (3.5.7) with (3.5.54) at $\lambda_0 = 1/r_0^2$ is Type-I. Furthermore, the problem (3.5.7) with (3.5.54) bifurcates on $\lambda > \lambda_0 = r_0^{-2}$ to a cycle attractor $\Sigma_\lambda = S^1$ which consists of singular points, as shown in Fig. 3.20, and the singular points in Σ_λ can be expressed as

$$u_\lambda = \sigma^{-1/2} \left(\lambda - \frac{1}{r_0^2} \right)^{1/2} \cos(\theta + \theta_0) + o\left(\left| \lambda - \frac{1}{r_0^2} \right|^{1/2} \right), \quad \sigma = \frac{3}{4}\gamma_3 - \frac{1}{6}r_0^2\gamma_2^2,$$

where θ_0 is the angle of u_λ in Σ_λ .

2. If

$$\gamma_3 < \frac{2r_0^2}{9}\gamma_2^2,$$

then the transition is Type-II. Moreover, the problem bifurcates on $\lambda < \lambda_0$ to a cycle invariant set $\Gamma_\lambda = S^1$ consisting of singular points, and there is a singularity separation at $\lambda^* < \lambda_0$, where Γ_λ and an S^1 attractor $\Sigma_\lambda = S^1$ are generated such that the system undergoes a transition at $\lambda = \lambda_0$ from $u = 0$ to $u_\lambda \in \Sigma_\lambda$, as shown in Fig. 3.21.

Fig. 3.20 All points in Σ_λ are singular points

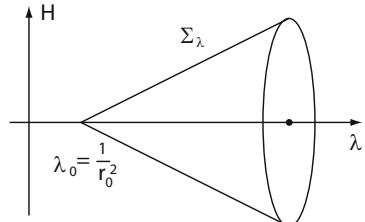
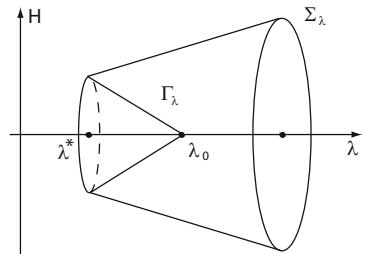


Fig. 3.21 The point λ^* is a singularity separation point from where two invariant sets Σ_λ and Γ_λ are separated with Σ_λ being an attractor, and $\Gamma_\lambda \rightarrow 0$ as $\lambda \rightarrow \lambda_0$



Fourth, consider the problem that the equation (3.5.7) is defined in the whole space $\Omega = \mathbb{R}^n$ ($n \geq 2$), with the periodic boundary condition

$$u(x + 2K\pi) = u(x) \quad \forall K = (k_1, \dots, k_n) \in \mathbb{Z}^n. \quad (3.5.57)$$

In this case, the eigenvalues and eigenfunctions of L_λ are given by

$$\begin{aligned}\beta_K(\lambda) &= |K|^2(\lambda - |K|^2), & K = (k_1, \dots, k_n), \quad |K|^2 = k_1^2 + \dots + k_n^2, \\ e_K^1 &= \cos(k_1 x_1 + \dots + k_n x_n), & e_K^2 = \sin(k_1 x_1 + \dots + k_n x_n).\end{aligned}$$

It is clear that the first eigenvalue $\beta_1(\lambda) = \lambda - 1$ of L_λ has multiplicity $2n$, and the first eigenfunctions are

$$e_j^1 = \cos x_j, \quad e_j^2 = \sin x_j \quad \forall 1 \leq j \leq n.$$

Theorem 3.5.7

1. If

$$\gamma_3 > \frac{14}{27} \gamma_2^2,$$

then the phase transition of (3.5.7) with (3.5.57) at $\lambda_0 = 1$ is Type-I. Moreover,

- a. the problem bifurcates from $(u, \lambda) = (0, 1)$ to an attractor Σ_λ homeomorphic to a $(2n-1)$ -dimensional sphere S^{2n-1} , and
- b. for each k ($0 \leq k \leq n-1$), the attractor Σ_λ contains C_n^k -dimensional tori \mathbb{T}^{n-k} consisting of singular points.

2. If

$$\gamma_3 < \frac{14}{27} \gamma_2^2,$$

then the transition is Type-II.

3.5.5 Phase Diagrams and Physical Conclusions

Criteria of Separation Order For simplicity, we consider the case where the spatial domain $\Omega = \prod_{k=1}^3 (0, L_k) \subset \mathbb{R}^3$ is a rectangular domain. Here we use the Neumann boundary condition, which is physically relevant.

Each 3D rectangular domain is one of the following two cases:

$$\text{Case I: } L = L_1 > L_j \quad \forall j = 2, 3,$$

$$\text{Case II: } L = L_1 = L_2 > L_3 \text{ or } L_1 = L_2 = L_3.$$

We define a nondimensional parameter:

$$K = \begin{cases} \frac{2L^2}{9\pi^2} \gamma_2^2 - \gamma_3 & \text{for Case I,} \\ \frac{26L^2}{27\pi^2} \gamma_2^2 - \gamma_3 & \text{for Case II.} \end{cases} \quad (3.5.58)$$

which is equivalent to the following dimensional parameter:

$$K_d = \begin{cases} \frac{2L_d^2 b_2^2}{9\pi^2 k} - b_3 & \text{for Case I,} \\ \frac{26L_d^2 b_2^2}{27\pi^2 k} - b_3 & \text{for Case II.} \end{cases} \quad (3.5.59)$$

where $L_d = L \cdot l$ is the dimensional length scale.

Theorems 3.5.4 and 3.5.5 give a precise characterization of the order of transitions, which is determined by the sign of this parameter K or K_d . In particular, the following physical conclusion is the direct translation of Theorems 3.5.4 and 3.5.5. Note that the control parameter λ for the transition/bifurcation is naturally linked to the temperature T as

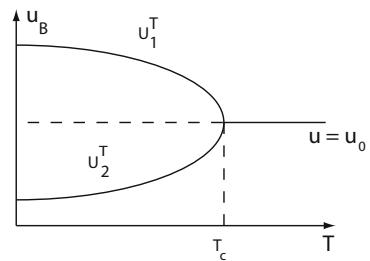
$$\lambda = -\frac{l^2 D}{2k} \left(\frac{RT}{u_0(1-u_0)} - 2a \right).$$

Hence the two critical bifurcation parameters λ^* and λ_c in Theorems 3.5.4 and 3.5.5 lead to two critical temperatures T_c and T^* .

Physical Conclusion 3.5.8 *The order of phase separation is completely determined by the sign of the nondimensional parameter K_d as follows:*

- (1) *If $K_d < 0$, the separation is second order and the dynamic behavior of the Cahn–Hilliard system is as shown in Fig. 3.22.*
- (2) *If $K_d > 0$, the separation is first-order transition with latent heat. In particular, there are two critical temperature $T^* > T_c$ such that if the temperature $T > T^*$, the system is in the homogeneous state, when $T^* > T > T_c$, the system is in metastable state accompanied with hysteresis corresponding to saddle-node bifurcation, and when $T < T_c$, the system is under phase separation state. In addition, the critical temperatures are functions of u_0 and L : $T^* = T^*(u_0, L)$, $T_c = T_c(u_0, L)$. See Fig. 3.23.*

Fig. 3.22 Transition diagram for $K_d < 0$: The state $u_0 = \bar{u}_B$ is stable if $T_c < T$, and the state u_0 is unstable, U_1^T and U_2^T are stable if $T < T_c$



This is in agreement with *part of* the classical phase diagram from the classical thermodynamic theory given in Fig. 3.24; see, among others, (Reichl, 1998; Novick-Cohen and Segel, 1984; Langer, 1971). However, as we shall see below, our result shows that near $u_0 = 1/2$, there is no metastable region; see Fig. 3.25. In other words, the phase diagram given in Fig. 3.25 is different from the one presented in classical studies as shown in Fig. 3.24.

Fig. 3.23 Transition diagram for $K_d > 0$: For fixed u_0 and L , the transition for the case where $K_d > 0$ is first-order separation with latent heat and with hysteresis: U_1^T and U_2^T represent separation states, and u_0 is the homogeneous state. In this case, for $T_c < T < T^*$, all states u_0 , U_1^T , U_2^T are metastable states. For $T < T_c$, u_0 is unstable, and U_1^T and U_2^T are stable states

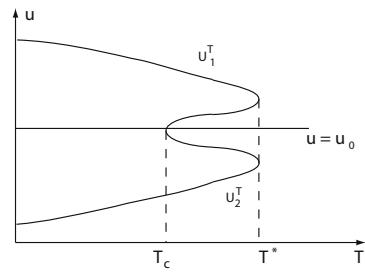
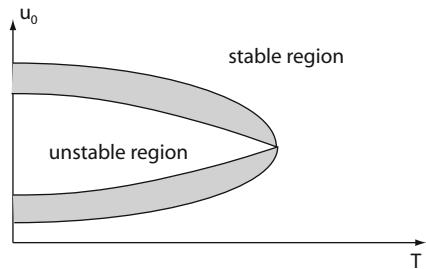


Fig. 3.24 Typical phase diagram from classical thermodynamic theory with shadowed being the metastable region



Critical Parameter Equations To examine the order of separations, we start with two important formulas for the critical parameters. By (3.5.4), solving $K_d = 0$ gives a precise formula for the critical (dimensional) length scale L_d in terms of the critical temperature T_c and the mol fraction u_0 :

$$L_d = \begin{cases} \frac{3\pi\sqrt{k}}{\sqrt{2}} \frac{\sqrt{b_3}}{|b_2|} = \frac{3\sqrt{3k}\pi\sqrt{u_0(1-u_0)(1-3u_0+3u_0^3)}}{2\sqrt{2DRT_c}|u_0 - \frac{1}{2}|} & \text{for Case I,} \\ \frac{3\pi\sqrt{3k}}{\sqrt{26}} \frac{\sqrt{b_3}}{|b_2|} = \frac{3\sqrt{3k}\pi\sqrt{u_0(1-u_0)(1-3u_0+3u_0^3)}}{2\sqrt{26DRT_c}|u_0 - \frac{1}{2}|} & \text{for Case II.} \end{cases} \quad (3.5.60)$$

It is then easy to derive from (3.5.60) the following asymptotic formula for L_d :

$$L_d = \begin{cases} \frac{3\sqrt{3k}\pi}{8\sqrt{2DRT_c}|u_0 - \frac{1}{2}|} + O(1) & \text{for Case I,} \\ \frac{9\sqrt{k}\pi}{8\sqrt{26DRT_c}|u_0 - \frac{1}{2}|} + O(1) & \text{for Case II,} \end{cases} \quad (3.5.61)$$

where T_c is the critical temperature as given in Physical Conclusion 3.5.8.

In fact, the first eigenvalue is given by

$$\beta_1 = -\frac{\pi^2}{L^2} \left(\frac{\pi^2}{L^2} - \lambda \right) = -\frac{\pi^2}{L^2} \left(\frac{\pi^2}{L^2} + \frac{l^2 b_1}{k} \right). \quad (3.5.62)$$

By (3.5.4), the critical parameter curve equation $\beta_1 = 0$ is given by

$$T_c = \frac{u_0(1-u_0)}{RD} \left(a - \frac{k\pi^2}{2l^2 L^2} \right) = \frac{u_0(1-u_0)}{RD} \left(a - \frac{k\pi^2}{2L_d^2} \right). \quad (3.5.63)$$

Phase Diagrams We are ready now to examine the order of separation in terms of the length scale L_d and mol fraction u_0 . First, using (3.5.60) or (3.5.61), we derive the Lu_0 -phase diagram given in Fig. 3.26. Then using Fig. 3.26 and (3.5.63), we derive the Tu_0 -phase diagram given in Fig. 3.25.

Consequently, we have shown the following physical conclusion:

Physical Conclusion 3.5.9

- (1) For a fixed length scale $L = L'$, there are numbers $x_1 < \frac{1}{2} < x_2$ such that the transition is second order if the molar fraction $x_1 < u_0 < x_2$, and the transition is first order if $u_0 > x_2$ or $u_0 < x_1$. See Fig. 3.26.
- (2) The phase diagram Fig. 3.25 is for this fixed length scale L' . The points x_1 and x_2 are the two molar concentrations where there is no metastable region and no hysteresis phenomena for $x_1 < u_0 < x_2$. In other words,

$$T^*(u_0) = T_c(u_0) \quad \text{for } x_1 < u_0 < x_2.$$

Fig. 3.25 Tu_0 -phase diagram for a fixed length scale L' , derived using Fig. 3.26 and (3.5.63); the shadowed region stands for the metastable region, and $T^*(x_0)$ and $T_c(x_0)$ are critical temperatures as shown in Fig. 3.23

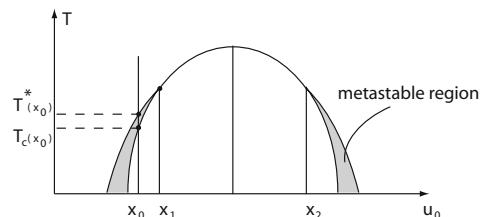
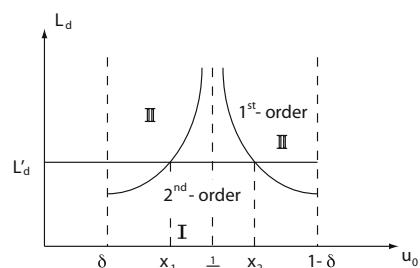
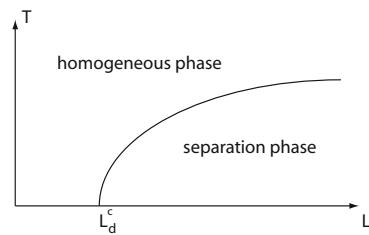


Fig. 3.26 Lu_0 -phase diagram derived using (3.5.60) or (3.5.61); region II is the first-order transition region with latent heat, and region I is the second-order transition region



We now derive the TL -phase diagram. Using (3.5.63), we derive the TL phase diagram given in Fig. 3.27, and the following physical conclusion:

Fig. 3.27 TL phase diagram



Physical Conclusion 3.5.10 For a given molar fraction $0 < u_0 < 1$, the critical parameter curve (3.5.63) divides the TL -plane into two regions: the homogeneous region I, and the separation region II such that the following assertions hold true:

- (1) If (L, T) is region I, then the system is in homogeneous phase, and if (L, T) is in region II, the system is in the separation phase.
- (2) The phase separation takes place as the control parameter (L, T) crosses the critical curve from region I into region II.

In particular, there is a critical (dimensional) length $L_d^c = \sqrt{k\pi^2/2a}$ such that if $L_d < L_d^c$, there is no phase separation for any temperature, and if $L_d > L_d^c$, phase separation occurs at the critical temperature $T = T_c$ given by (3.5.63).

Pattern Formation, Symmetry, and Periodic Structure Physical experiments have shown that in pattern formation via phase separation, periodic or semi-periodic structure appears. From Theorems 3.5.6 and 3.5.7 we see that for the loop domains and bulk domains which can be considered as \mathbb{R}^n or $\mathbb{R}^m \times D$ ($D \subset \mathbb{R}^{n-m}$) the steady-state solutions of the Cahn–Hilliard equation are periodic, and for rectangular domains they are semi-periodic, and the periodicity is associated with the mirror image symmetry.

For $\Omega = (0, L)^m \times D$, Theorem 3.5.4 is also valid. In fact, the following space

$$\tilde{H} = \left\{ u \in L^2(\Omega) \mid u = \sum_{|K|=1}^{\infty} y_k \cos \frac{k_1 \pi}{L} x_1 \cdots \cos \frac{k_m \pi}{L} x_m \right\} \subset H$$

is invariant for the Cahn–Hilliard equations (3.5.7) and (3.5.5). All separated equilibrium states in Theorem 3.5.4 are in \tilde{H} . From the physical viewpoint, all equilibrium states $u(x)$ and their mirror image states $u(L - x', x'')$ are the same to describe the pattern formation. Mathematically, under the mirror image transformation

$$x \rightarrow (L - x', x''), \quad x' = (x_1, \dots, x_m), \quad x'' = (x_{m+1}, \dots, x_n),$$

the Cahn–Hilliard equation (3.5.7) with (3.5.5) is invariant. Hence, the steady-state solutions will appear in pairs. In particular, for Type-I phase transition, there is a remarkable mirror image symmetry. We address this problem as follows.

Let $m = 1$ in $\Omega = (0, L)^m \times D$. By (3.5.21) there are two bifurcated stable equilibrium states, and their projections on the first eigenspace are

$$u_{1,2} = \pm y \cos \frac{\pi x_1}{L}, \quad y = \sqrt{2(\lambda - \pi^2/L^2)} / \left(\frac{3}{2} \gamma_3 - \frac{L^2}{3\pi^2} \gamma_2^2 \right).$$

It is clear that $u_2(x_1) = u_1(L - x_1)$.

Let $m = 2$. By Theorem 3.5.5 the bifurcated attractor Σ_λ contains eight equilibrium states, whose projections are given by

$$u_1^\pm = \pm y_0 e_1, \quad u_2^\pm = \pm y_1(e_1 + e_2), \quad u_3^\pm = \pm y_0 e_2, \quad u_4^\pm = \pm y_1(-e_1 + e_2),$$

where $e_1 = (\cos \frac{\pi x_1}{L}, 0)$ and $e_2 = (0, \cos \frac{\pi x_2}{L})$ form an orthogonal basis in \mathbb{R}^2 , and

$$y_0 = \frac{\sqrt{2(\lambda - \pi^2/L^2)}}{\sqrt{\sigma_1}}, \quad y_1 = \frac{\sqrt{2(\lambda - \pi^2/L^2)}}{\sqrt{\sigma_1 + \sigma_2}} \quad \text{with } \sigma_1, \sigma_2 \text{ as in (3.5.38).}$$

These eight equilibrium states constitute an octagon in \mathbb{R}^2 , as shown in Fig. 3.28, and they are divided into two classes: $\mathcal{A}_1 = \{u_1^\pm, u_3^\pm\}$ and $\mathcal{A}_3 = \{u_2^\pm, u_4^\pm\}$ by the $\frac{\pi}{2}$ -rotation group $G(\frac{\pi}{2})$. Namely, with the action of $G(\frac{\pi}{2})$, \mathcal{A}_i ($i = 1, 2$) are invariant:

$$Bu \in \mathcal{A}_i, \quad \forall u \in \mathcal{A}_i \text{ and } B \in G\left(\frac{\pi}{2}\right),$$

where $G(\frac{\pi}{2})$ consists of the orthogonal matrices

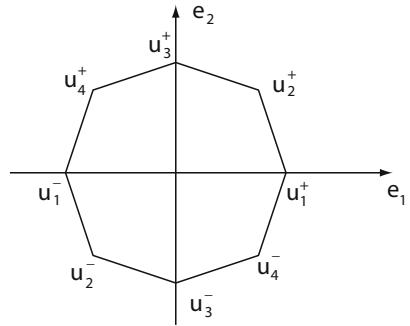
$$B_1^\pm = \pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad B_2^\pm = \pm \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad B_3^\pm = \pm \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad B_4^\pm = \pm \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The stability of the equilibrium states u_k is associated with both classes \mathcal{A}_1 and \mathcal{A}_2 , i.e., either the elements in \mathcal{A}_1 are stable, or those in \mathcal{A}_2 are stable; see Fig. 3.17. By (3.5.52) we can derive the criterion as follows

$$\begin{aligned} u_{2k+1}^\pm \in \mathcal{A}_1 \text{ are stable} &\Leftrightarrow \frac{22}{9} \frac{L^2 \gamma_2^2}{\pi^2} > \gamma_3 > \frac{26}{27} \frac{L^2 \gamma_2^2}{\pi^2} \quad \text{for } k = 0, 1, \\ u_{2k}^\pm \in \mathcal{A}_2 \text{ are stable} &\Leftrightarrow \gamma_3 > \frac{22}{9} \frac{L^2 \gamma_2^2}{\pi^2} \quad \text{for } k = 1, 2, \end{aligned}$$

In Fig. 3.28, we see that elements in \mathcal{A}_1 and \mathcal{A}_2 have a $\frac{\pi}{4}$ difference in their phase angle. However, in their pattern structure, $u_{2K+1}^\pm \in \mathcal{A}_1$ and $u_{2k}^\pm \in \mathcal{A}_2$ also have a $\pi/4$ difference at the angle between the lines of $u_{2K+1}^\pm = 0$ and $u_{2k}^\pm = 0$. In fact, the lines that

Fig. 3.28 The eight equilibrium states for the case where $m = 2$ given by Theorem 3.5.5



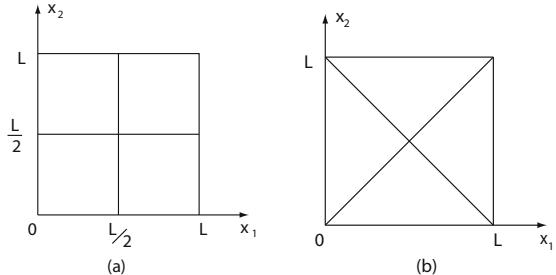
$$u_1^+ = -u_1^- = y_0 \cos \frac{\pi x_1}{L} = 0 \quad \text{and} \quad u_3^+ = -u_3^- = y_0 \cos \frac{\pi x_2}{L} = 0$$

are given by $x_1 = L/2$ and $x_2 = L/2$ respectively, as shown in Fig. 3.29a, and the lines

$$\begin{aligned} u_2^+ = -u_2^- &= y_1 \left(\cos \frac{\pi x_1}{L} + \cos \frac{\pi x_2}{L} \right) = 0, \quad \text{and} \\ u_4^+ = -u_4^- &= y_1 \left(-\cos \frac{\pi x_1}{L} + \cos \frac{\pi x_2}{L} \right) = 0 \end{aligned}$$

are given by $x_2 = L - x_1$ and $x_2 = x_1$ respectively as shown in Fig. 3.29b.

Fig. 3.29 Lines $u_j^\pm = 0$, $j=1,2,3,4$



Let $m = 3$. Then the bifurcated attractor Σ_λ contains 26 equilibrium states which can be divided into three classes by the three-dimensional $(\frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2})$ -rotation group $G(\frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2})$ as follows

$$\begin{aligned} \mathcal{A}_1 &= \{u_1^\pm = \pm y_0 e_1, u_2^\pm = \pm y_0 e_2, u_3^\pm = \pm y_0 e_3\}, \\ \mathcal{A}_2 &= \{u_4^\pm = \pm y_1(e_1 + e_2), u_5^\pm = \pm y_1(e_1 + e_3), u_6^\pm = \pm y_1(e_2 + e_3), \\ &\quad u_7^\pm = \pm y_1(-e_1 + e_2), u_8^\pm = \pm y_1(-e_1 + e_3), u_9^\pm = \pm y_1(-e_2 + e_3)\}, \\ \mathcal{A}_3 &= \{u_{10}^\pm = \pm y_1(e_1 + e_2 + e_3), u_{11}^\pm = \pm y_1(-e_1 + e_2 + e_3), \\ &\quad u_{12}^\pm = \pm y_1(e_1 - e_2 + e_3), u_{13}^\pm = \pm y_1(e_1 + e_2 - e_3)\}. \end{aligned}$$

Only these elements in \mathcal{A}_1 or in \mathcal{A}_3 are stable, and they are determined by the following criterion

$$\begin{aligned} \text{elements in } \mathcal{A}_1 \text{ are stable} &\Leftrightarrow \frac{22}{9} \frac{L^2 \gamma_2^2}{\pi^2} > \gamma_3 > \frac{26}{27} \frac{L^2 \gamma_2^2}{\pi^2}, \\ \text{elements in } \mathcal{A}_3 \text{ are stable} &\Leftrightarrow \gamma_3 > \frac{22}{9} \frac{L^2 \gamma_2^2}{\pi^2}. \end{aligned}$$

3.6 Superconductivity

Superconductivity was first discovered in 1911 by H. Kamerlingh Onnes, who found that Mercury had zero electric resistance when the temperature decreases below some critical value T_c . Since then, one has found that large number of metals and alloys possess the superconducting property. In the superconducting state once a current is set up in a metal ring, it is expected that no change in this current occurs in times more than 10^{10} years; see Tinkham (1996). In 1933, another important superconducting property, called the diamagnetism or the Meissner effect, was discovered by W. Meissner and R. Ochsenfeld. They found that not only a magnetic field is excluded from entering a superconductor, but also that a field in an originally normal sample is expelled as it is cooled below T_c .

One central problem in the theory of superconductivity is to study the phase transition between a normal state, characterized by a complex order parameter ψ that vanishes identically, and a superconducting state, characterized by the order parameter that is not identically zero. The main objective of this section is to address this problem by studying the stability and dynamic transitions of the time dependent Ginzburg-Landau (TDGL) model of superconductivity.

As we know, the TDGL model for superconductivity was first proposed by Schmid (1966) and subsequently validated by Gor'kov and Eliashberg (Gor'kov, 1968) in the context of the microscopic Bardeen-Cooper-Schrieffer (BCS) theory of superconductivity. Over the years, there have been growing physical evidence over the last few decades on the importance of the time-dependent models.

The analysis of the TDGL leads to some interesting theoretical predictions on order of transitions and phase diagrams of superconductivity. The material presented in this section is based on Ma and Wang (2005a), and the physical conclusions are presented here for the first time.

3.6.1 Ginzburg-Landau Model

Ginzburg-Landau Free Energy With the unified model presented in Sect. 3.1, we can derive the model for superconductivity in a much simpler and transparent fashion.

We start with the Ginzburg-Landau free energy G , corresponding to notation H used before for the free energy, given by

$$G = \int_{\Omega} \left[\frac{1}{2m_s} |(-i\hbar\nabla - \frac{e_s}{c}A)\psi|^2 + \frac{a}{2} |\psi|^2 + \frac{b}{4} |\psi|^4 + \frac{H^2}{8\pi} - \frac{HH_a}{4\pi} \right], \quad (3.6.1)$$

where $\Omega \subset \mathbb{R}^n$ ($n = 2, 3$) is a bounded domain, the complex valued function $\psi : \Omega \rightarrow \mathbb{C}$ stands for the order parameter, the vector valued function $H : \Omega \rightarrow \mathbb{R}^n$ stands for the magnetic field, H_a is the applied field, \hbar is the Planck constant, e_s and m_s are the charge and mass of a Cooper pair, c is the speed of light, A is the magnetic potential satisfying $H = \text{curl } A$, and the parameters $a = a(T)$, $b = b(T)$ are coefficients; see de Gennes (1966).

$$b = b(T) > 0, \quad a = a(T) \begin{cases} > 0 & \text{for } T > T_c, \\ < 0 & \text{for } T < T_c, \end{cases} \quad (3.6.2)$$

Here T_c is the critical temperature where incipient superconductivity property may be observed. In the BCS theory, for instance, they are given by (see de Gennes 1966):

$$a(T) = N(0) \frac{T - T_c}{T_c}, \quad b(T) = 0.098 \frac{N(0)}{(k_b T_c)^2},$$

where k_b is the Boltzmann constant, $N(0)$ is the state density at $T = 0$ K.

The order parameter ψ describes the local density n_s of superconducting electrons: $|\psi|^2 = n_s$. In addition, ψ is proportional to the energy gap parameter Δ near T_c , which appears in the BCS theory.

Using the Ginzburg-Landau energy, one can derive naturally the steady-state Euler-Lagrange equations, which have been studied by many authors—both physicists and mathematicians, leading to many interesting physical predictions and mathematical theory.

Derivation of TDGL Model For the time-dependent problem, because of gauge invariance, in addition to the order parameter and the vector potential, a third variable is needed to complete the description of the physical state of the system in a manner consistent with the gauge invariance.

By (3.6.1), the equation (3.1.3) can be written as

$$k_1 \frac{\partial \psi}{\partial t} + \phi_1(\psi, A) = -\frac{1}{2m_s} (i\hbar\nabla + \frac{e_s}{c}A)^2 \psi - a\psi - b|\psi|^2\psi, \quad (3.6.3)$$

$$\begin{aligned} k_2 \frac{\partial A}{\partial t} + \phi_2(\psi, A) &= -\frac{1}{4\pi} \text{curl}^2 A + \frac{1}{4\pi} \text{curl } H_a \\ &\quad - \frac{e_s^2}{m_s c^2} |\psi|^2 A - \frac{e_s \hbar}{2m_s c} i(\psi^* \nabla \psi - \psi \nabla \psi^*), \end{aligned} \quad (3.6.4)$$

where $k_1, k_2 > 0$ are constants, $\phi_i(\psi, A)$ are two operators to be determined by physical laws and the standard model (3.1.3) in several steps as follows:

FIRST, the Maxwell equations give that

$$\frac{\sigma}{c^2} \frac{\partial A}{\partial t} + \sigma \nabla \phi + \frac{1}{4\pi} \operatorname{curl}^2 A - \frac{1}{4\pi} \operatorname{curl} H_a = J_s, \quad (3.6.5)$$

where σ is the conductivity, ϕ is the electric potential, and J_s is the supercurrent density. On the other hand, by Quantum Mechanics, the supercurrent density J_s is given by

$$J_s = -\frac{e_s^2}{m_s c^2} |\psi|^2 A - \frac{e_s \hbar}{2m_s c} i(\psi^* \nabla \psi - \psi \nabla \psi^*). \quad (3.6.6)$$

In comparison with (3.6.4), (3.6.5), and (3.6.6), we have

$$\phi_2(\psi, A) = \sigma \nabla \phi. \quad (3.6.7)$$

SECOND, by direct computation, it is easy to show that by (3.6.3), (3.6.4), and (3.1.3), ϕ_1 is uniquely determined from (3.6.7) by:

$$\phi_1(\psi, A) = i\alpha \phi \psi,$$

where α is a coefficient.

THIRD, according to the physical dimensional balance, also due to P. L. Gork'ov, the coefficients k_1 and α are given by $k_1 = \hbar^2 / 2m_s D$ and $\alpha = he_s / 2m_s D$, where D is the diffusion coefficient. Thus, we deduce the following TDGL equations:

$$\begin{aligned} \frac{\hbar^2}{2m_s D} \left(\frac{\partial}{\partial t} + \frac{ie_s}{\hbar} \phi \right) \psi &= -\frac{1}{2m_s} \left(i\hbar \nabla + \frac{e_s}{c} A \right)^2 \psi - a\psi - b|\psi|^2 \psi, \\ \frac{\sigma}{c^2} \frac{\partial A}{\partial t} + \sigma \nabla \phi &= -\frac{1}{4\pi} \operatorname{curl}^2 A + \frac{1}{4\pi} \operatorname{curl} H_a - \frac{e_s^2}{m_s c^2} |\psi|^2 A \\ &\quad - \frac{e_s \hbar}{2m_s c} i(\psi^* \nabla \psi - \psi \nabla \psi^*). \end{aligned} \quad (3.6.8)$$

Boundary Conditions and Gauge Transformation Let Ω_d be the spatial domain. A physically sound boundary condition for the order parameter is given by

$$C_1(i\hbar \nabla + \frac{e_s}{c} A) \psi \cdot n = -C_2 i \hbar \psi \quad \text{on } \partial \Omega_d, \quad (3.6.9)$$

which means that no current passes through the boundary, where n is the unit outward normal vector at $\partial \Omega_d$, and $C_1, C_2 \geq 0$ are constants depending on the material to which the contact is made. Physically, they satisfy; see Tinkham (1996); de Gennes (1966):

$$\begin{aligned} C_2 = 0, C_1 \neq 0 &\quad \text{for an insulator on } \partial \Omega_d, \\ C_1 = 0, C_2 \neq 0 &\quad \text{for an magnetic material,} \\ 0 < C_2/C_1 < \infty &\quad \text{for a normal metal.} \end{aligned} \quad (3.6.10)$$

We note that the equations (3.6.8) with (3.6.9) are invariant under the following gauge transformation

$$(\psi, A, \phi) \rightarrow \left(\psi e^{i\theta}, A + \frac{hc}{e_s} \nabla \theta, \phi - \frac{h}{e_s} \theta_t \right),$$

where θ is an arbitrary function. If we take θ such that

$$\frac{hc}{e_s} \Delta \theta = \operatorname{div} A \quad \text{in } \Omega_d,$$

$$\frac{hc}{e_s} \frac{\partial \theta}{\partial n} = A \cdot n \quad \text{on } \partial \Omega_d,$$

then we obtain an additional equation and a boundary condition as follows (de Gennes, 1966; Tang and Wang, 1995):

$$\operatorname{div} A = 0 \quad \text{in } \Omega_d, \tag{3.6.11}$$

$$A_n = A \cdot n = 0 \quad \text{on } \partial \Omega_d. \tag{3.6.12}$$

Another boundary condition often imposed for A is

$$\operatorname{curl} A \times n = H_a \times n \quad \text{on } \partial \Omega_d. \tag{3.6.13}$$

Nondimensional Forms of the TDGL Equations From both the mathematical and physical points of view, it is convenient to introduce the nondimensional forms of the TDGL equations. We start with the dimensions of various physical quantities. Let m be the mass, L the length scale, t the time, and E the energy. Then we have

$$\begin{array}{llll} E : L^2 m / t^2, & h : Et, & D : L^2 / t, & e_s^2 : EL, \\ \sigma : 1/t, & c : L/t, & a : E, & b : EL^3, \\ \psi : 1/L^{3/2}, & A : (E/L)^{1/2}, & H : (E/L^3)^{1/2}. \end{array}$$

We introduce some physical parameters, which are classical and physically important:

$$\begin{aligned} \lambda &= \lambda(T) = (m_s c^2 b / 4\pi e_s^2 |a|)^{1/2}, & |\psi_0|^2 &= |a|/b, & H_c &= (4\pi |a|^2/b)^{1/2}, \\ \xi &= \xi(T) = h/(2m_s |a|)^{1/2}, & \kappa &= \lambda/\xi, & \eta &= 4\pi \sigma D/c^2, \\ \tau &= \lambda^2/D. \end{aligned}$$

Physically speaking, $|\psi_0|^2$ stands for the equilibrium density, H_c for the thermodynamic critical field, $\lambda = \lambda(T)$ for the penetration depth, $\xi(T)$ for the coherence length, and τ for the relaxation time. The ratio of the two characteristic length scales $\kappa = \lambda/\xi$ is called the Ginzburg-Landau parameter of the substance. When $0 < \kappa < \frac{1}{\sqrt{2}}$, the material is of the Type-I, and when $\kappa > \frac{1}{\sqrt{2}}$, the material is of the Type-II.

For the purpose of studying phase transition, the nondimensional form below was first introduced in Ma and Wang (2005b,a), and differs slightly from the ones often used in the literature. For this purpose, we set

$$\begin{aligned} l &= \sqrt{b}/e_s, & \tau_0 &= hl/e_s^2, & \phi_0 &= e_s^2/\sqrt{b}, \\ A_0 &= \left(e_s hc^2/D\sqrt{b}\right)^{1/2}, & \alpha &= -2a\sqrt{b}m_s D/e_s^3 h, & \mu &= hD/\sqrt{b}e_s, \\ \beta &= 2m_s D/h, & \zeta &= 4\pi\sigma l e_s^2/c^2 h, & \gamma &= 4\pi e_s^2/m_s c^2 l, \end{aligned}$$

and

$$\begin{aligned} x &= lx', & t &= \tau_0 t', & \psi &= l^{-3/2} \psi', \\ A &= A_0 A', & \phi &= \phi_0 \phi', & H_a &= l^{-1} A_0 H'_a. \end{aligned}$$

Then with the gauge taken such that (3.6.11) and (3.6.12) hold true, we obtain the following nondimensional TDGL equations (suppressing the primes):

$$\begin{aligned} \frac{\partial \psi}{\partial t} + i\phi\psi &= -(i\mu\nabla + A)^2\psi + \alpha\psi - \beta|\psi|^2\psi, \\ \zeta \left[\frac{\partial A}{\partial t} + \mu\nabla\phi \right] &= -\operatorname{curl}^2 A + \operatorname{curl} H_a - \gamma|\psi|^2 A - \frac{\gamma\mu i}{2}(\psi^*\nabla\psi - \psi\nabla\psi^*), \quad (3.6.14) \\ \operatorname{div} A &= 0. \end{aligned}$$

The nondimensional domain is denoted by Ω . The initial conditions are given by

$$\psi(0) = \psi_0, \quad A(0) = A_0.$$

The boundary conditions can be chosen as one of the following:

NEUMANN BOUNDARY CONDITION. For the case where Ω is enclosed by an insulator:

$$\frac{\partial \psi}{\partial n} = 0, \quad A_n = 0, \quad \operatorname{curl} A \times n = H_a \times n \quad \text{on } \partial\Omega.$$

DIRICHLET BOUNDARY CONDITION. For the case where Ω is enclosed by a magnetic material:

$$\psi = 0, \quad A_n = 0, \quad \operatorname{curl} A \times n = H_a \times n \quad \text{on } \partial\Omega.$$

ROBIN BOUNDARY CONDITION. For the case where Ω is enclosed by a normal metal:

$$\frac{\partial \psi}{\partial n} + C\psi = 0, \quad A_n = 0, \quad \operatorname{curl} A \times n = H_a \times n \quad \text{on } \partial\Omega.$$

Remark 3.6.1 If the material is a loop, or a plate $\Omega = \tilde{\Omega} \times (0, h)$, with the height h being small in comparison to the diameter of $\tilde{\Omega}$, then it is reasonable to consider the periodic boundary condition either in the x -direction or in the (x, y) -directions.

Remark 3.6.2 We shall see in later discussions that the parameter α plays a key role in the phase transition of superconductivity, which is given in terms of dimensional quantities by

$$\alpha = \alpha(T) = \frac{2\sqrt{b}m_s DN(0)}{e_s^3 h} \cdot \frac{T_c - T}{T_c}. \quad (3.6.15)$$

3.6.2 TGDL as a Gradient-Type System

Mathematical Setting It is known that for a given applied field H_a with $\operatorname{div} H_a = 0$, there exists a field A_a satisfying

$$\begin{aligned}\operatorname{div} A_a &= 0, & \operatorname{curl} A_a &= H_a \quad \text{in } \Omega, \\ A_a \cdot n &= 0 & & \text{on } \partial\Omega.\end{aligned}$$

Let $A = \mathcal{A} + A_a$. Then (3.6.14) are rewritten as

$$\begin{aligned}\frac{\partial \psi}{\partial t} + i\phi\psi &= -(i\mu\nabla + A_a)^2\psi + \alpha\psi - 2A_a \cdot \mathcal{A}\psi \\ &\quad - 2i\mu\mathcal{A} \cdot \nabla\psi - |\mathcal{A}|^2\psi - \beta|\psi|^2\psi, \\ \zeta \left[\frac{\partial \mathcal{A}}{\partial t} + \mu\nabla\phi \right] &= -\operatorname{curl}^2 \mathcal{A} - \gamma A_a |\psi|^2 - \gamma \mathcal{A} |\psi|^2 \\ &\quad - \frac{\gamma\mu}{2} i(\psi^* \nabla\psi - \psi \nabla\psi^*), \\ \operatorname{div} \mathcal{A} &= 0,\end{aligned}\tag{3.6.16}$$

with the following initial and boundary conditions

$$\psi(0) = \psi_0, \quad \mathcal{A}(0) = \mathcal{A}_0, \tag{3.6.17}$$

$$\mathcal{A}_n|_{\partial\Omega} = 0, \quad \operatorname{curl} \mathcal{A} \times n|_{\partial\Omega} = 0, \tag{3.6.18}$$

and with one of the following three boundary conditions for ψ :

the NEUMANN BOUNDARY CONDITION:

$$\frac{\partial \psi}{\partial n}|_{\partial\Omega} = 0, \tag{3.6.19}$$

the DIRICHLET BOUNDARY CONDITION:

$$\psi|_{\partial\Omega} = 0, \tag{3.6.20}$$

or the ROBIN BOUNDARY CONDITION:

$$\left(\frac{\partial \psi}{\partial n} + C\psi \right)|_{\partial\Omega} = 0, \quad C > 0. \tag{3.6.21}$$

Hereafter, we use $H_B^k(\Omega, \mathbb{C})$ for the Sobolev spaces of complex valued functions defined on Ω , and $H^k(\Omega, \mathbb{R}^n)$ ($n = 2, 3$) the Sobolev spaces of vector valued functions defined on Ω . Let

$$H_B^2(\Omega, \mathbb{C}) = \{\psi \in H^2(\Omega, \mathbb{C}) \mid \psi \text{ satisfy one of (3.6.19)-(3.6.21)}\},$$

$$D^2(\Omega, \mathbb{R}^n) = \{\mathcal{A} \in H^2(\Omega, \mathbb{R}^n) \mid \operatorname{div} \mathcal{A} = 0, \mathcal{A} \text{ satisfy (3.6.18)}\},$$

$$\mathcal{L}^2(\Omega, \mathbb{R}^n) = \{\mathcal{A} \in L^2(\Omega, \mathbb{R}^n) \mid \operatorname{div} \mathcal{A} = 0, \mathcal{A}_n|_{\partial\Omega} = 0\}.$$

We set

$$H = L^2(\Omega, \mathbb{C}) \times \mathcal{L}^2(\Omega, \mathbb{R}^n), \quad H_1 = H_B^2(\Omega, \mathbb{C}) \times D^2(\Omega, \mathbb{R}^n).$$

Let $P : L^2(\Omega, \mathbb{R}^n) \rightarrow \mathcal{L}^2(\Omega, \mathbb{R}^n)$ be the Leray projection. Then it is known that the function ϕ in (3.6.16) is determined uniquely up to constants by the following equation

$$\zeta\mu\nabla\phi = (I - P)\left[\frac{\gamma\mu}{2}i(\psi\nabla\psi^* - \psi^*\nabla\psi) - \gamma|\psi|^2(\mathcal{A} + A_a)\right]. \quad (3.6.22)$$

Namely, for every $u = (\psi, \mathcal{A}) \in H_1$, there is a unique solution of (3.6.22) up to constants. Therefore, we define a nonlinear operator $\Phi : H_1 \rightarrow L^2(\Omega)$ by

$$\Phi(u) = \phi = \text{the solution of (3.6.22) with } \int_{\Omega} \phi dx = 0. \quad (3.6.23)$$

We set the mappings $L_\alpha = -K + B_\alpha$ and $G : H_1 \rightarrow H$ by

$$\begin{aligned} Ku &= \begin{pmatrix} (i\mu\nabla + A_a)^2\psi \\ \zeta^{-1}\operatorname{curl}^2\mathcal{A} \end{pmatrix}, & B_\alpha u &= \begin{pmatrix} \alpha\psi \\ 0 \end{pmatrix}, \\ G(u) &= -\begin{pmatrix} i\psi\Phi(u) + 2A_a \cdot \mathcal{A}\psi + 2i\mu\mathcal{A} \cdot \nabla\psi + |\mathcal{A}|^2\psi + \beta|\psi|^2\psi \\ P[\gamma\zeta^{-1}A_a|\psi|^2 + \gamma\zeta^{-1}\mathcal{A}|\psi|^2 + \frac{\gamma\mu\zeta^{-1}}{2}i(\psi^*\nabla\psi - \psi\nabla\psi^*)] \end{pmatrix}, \end{aligned}$$

where $u = (\psi, \mathcal{A})$, $\Phi(u)$ is defined by (3.6.23), and $P : L^2(\Omega, \mathbb{R}^n) \rightarrow \mathcal{L}(\Omega, \mathbb{R}^n)$ the Leray projection. Thus, the problems (3.6.16)–(3.6.18) with one of the boundary conditions (3.6.19)–(3.6.21) can be rewritten in the following operator form

$$\frac{du}{dt} = L_\alpha u + G(u), \quad u(0) = u_0. \quad (3.6.24)$$

Gradient-Type Structure Now we show that the Ginzburg-Landau equation has the gradient-type structure. The Ginzburg-Landau energy of (3.6.16) is given by

$$\begin{aligned} E(\psi, \mathcal{A}) &= \frac{1}{2} \int_{\Omega} \left[|(i\mu\nabla + \mathcal{A} + A_a)\psi|^2 + \frac{\beta}{2}|\psi|^4 - \alpha|\psi|^2 + \gamma^{-1}|\operatorname{curl}\mathcal{A}|^2 \right] dx \\ &\quad \left(+ \frac{1}{2}\mu^2 C \int_{\partial\Omega} |\psi|^2 ds \text{ for BC (3.6.21)} \right). \end{aligned} \quad (3.6.25)$$

The derivative operator of E is given by

$$DE(u) = \begin{pmatrix} (i\mu\nabla + \mathcal{A} + A_a)^2\psi - \alpha\psi + \beta|\psi|^2\psi \\ \gamma^{-1}\operatorname{curl}^2\mathcal{A} + |\psi|^2(\mathcal{A} + A_a) + \frac{\mu}{2}i(\psi^*\nabla\psi - \psi\nabla\psi^*) \end{pmatrix}.$$

Let

$$DE_1 = (i\mu\nabla + \mathcal{A} + A_a)^2\psi - \alpha\psi + \beta|\psi|^2\psi,$$

$$\begin{aligned} DE_2 &= \gamma^{-1} \operatorname{curl}^2 \mathcal{A} + |\psi|^2 (\mathcal{A} + A_a) + \frac{\mu}{2} i(\psi^* \nabla \psi - \psi \nabla \psi^*), \\ \tilde{L}u + \tilde{G}(u) &= -\left(\frac{\zeta DE_1}{\gamma DE_2} \right) - \Psi(u), \\ \Psi(u) &= \begin{pmatrix} \zeta i \phi \psi \\ \zeta \mu \nabla \phi \end{pmatrix}. \end{aligned}$$

Let $v = (\zeta \psi, \zeta \mathcal{A})$. Then, the equation (3.6.24) is equivalently rewritten as

$$\frac{dv}{dt} = \tilde{L}v + \tilde{G}(v), \quad v(0) = \zeta \varphi. \quad (3.6.26)$$

It suffices to prove that (3.6.26) is of the gradient-type structure. We see that

$$-(\tilde{L}u + \tilde{G}(u), DE(u)) \leq C \|DE\|^2 + (\Psi, DE). \quad (3.6.27)$$

By the boundary conditions and

$$\operatorname{div} A_a = 0, \quad \operatorname{div} \mathcal{A} = 0, \quad \operatorname{div} \operatorname{curl} \mathcal{A} = 0,$$

we find

$$\begin{aligned} (DE, \Psi) &= \zeta \operatorname{Re} \int_{\Omega} \left[\left((i\mu \nabla + \mathcal{A} + A_a)^2 \psi - \alpha \psi + \beta |\psi|^2 \psi \right) (i\phi \psi)^* \right. \\ &\quad \left. + \left(|\psi|^2 (\mathcal{A} + A_a) + \frac{\mu}{2} i(\psi^* \nabla \psi - \psi \nabla \psi^*) \right) \mu \nabla \phi \right] dx \\ &= \zeta \operatorname{Re} \int_{\Omega} \left[-i(i\mu \nabla + \mathcal{A} + A_a)^2 \psi \cdot \psi^* \phi + \right. \\ &\quad \left. + \left(|\psi|^2 (\mathcal{A} + A_a) + \frac{\mu}{2} i(\psi^* \nabla \psi - \psi \nabla \psi^*) \right) \mu \nabla \phi \right] dx \\ &= \zeta \operatorname{Re} \int_{\Omega} \left[i\mu^2 \phi \psi^* \Delta \psi + \mu \phi (\mathcal{A} + A_a) \cdot \nabla |\psi|^2 \right. \\ &\quad \left. + \mu \nabla \phi \cdot (\mathcal{A} + A_a) |\psi|^2 + \mu^2 \operatorname{Re}(i\psi^* \nabla \psi) \nabla \phi \right] dx \\ &= 0. \end{aligned}$$

Hence we derive from (3.6.27) that

$$-(\tilde{L}u + \tilde{G}(u), DE(u)) \leq C \|DE(u)\|^2.$$

It is clear that

$$DE(u_0) = 0 \Leftrightarrow Lu_0 + G(u_0) = 0.$$

Thus (3.6.26) is a gradient-type equation with the energy functional $E(u)$.

It is not hard to establish the following theorem on existence of solutions and global attractors, and we omit the proof. Note also that the existence of strong solutions and global attractor in H^2 -Sobolev space were first derived in Tang and Wang (1995).

Theorem 3.6.3 Consider the problems (3.6.16)–(3.6.18) with either (3.6.19), or (3.6.20), or (3.6.21).

1. For every initial value $\varphi = (\tilde{\psi}_0, \tilde{\mathcal{A}}_0) \in H_{1/2}$ and $A_a \in H_1$, the problem has a unique classical solution

$$(\psi, \mathcal{A}) \in C^1((0, \infty), H) \cap C^0((0, \infty), H_1), \quad \phi \in C^0((0, \infty), H^1(\Omega)).$$

2. If $\Omega \subset \mathbb{R}^n$ is C^∞ and $A_a \in C^\infty(\Omega, \mathbb{R}^n)$, then the solution is C^∞ :

$$(\psi, \mathcal{A}) \in C^\infty((0, \infty) \times \overline{\Omega}, \mathbb{C}) \times C^\infty((0, \infty) \times \overline{\Omega}, \mathbb{R}^n), \\ \phi \in C^\infty((0, \infty) \times \overline{\Omega}).$$

3. For any $k \geq 1/2$, the problem has a global attractor in $H_k \subset H^{2k}(\Omega, \mathbb{C}) \times H^{2k}(\Omega, \mathbb{R}^n)$, which attracts H_k in the H^{2k} -norm. Moreover, the set S consisting of all singular points of (3.6.16)–(3.6.18) in the attractor attracts H_k in the H^{2k} -norm.

3.6.3 Phase Transition Theorems

Eigenvalue Problems In order to describe the dynamical properties of the TDGL equations, it is necessary to consider the eigenvalue problems of the linearized equations of (3.6.16).

Let α_1 be the first eigenvalue of the following equation

$$(i\mu\nabla + A_a)^2\psi = \alpha\psi \quad \forall x \in \Omega, \quad (3.6.28)$$

with one of the boundary conditions (3.6.19)–(3.6.21). It is clear that (3.6.28) can be equivalently expressed as

$$\begin{aligned} -\mu^2\Delta\psi_1 + |A_a|^2\psi_1 - 2\mu A_a \cdot \nabla\psi_2 &= \alpha\psi_1, \\ -\mu^2\Delta\psi_2 + |A_a|^2\psi_2 + 2\mu A_a \cdot \nabla\psi_1 &= \alpha\psi_2, \end{aligned} \quad (3.6.29)$$

where $\psi = \psi_1 + i\psi_2$.

It is ready to verify that (3.6.29) with one of the boundary conditions (3.6.19)–(3.6.21) is symmetric. Therefore, eigenvalues of (3.6.28) are real:

$$\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_k \leq \cdots, \quad \lim_{k \rightarrow \infty} \alpha_k = +\infty, \quad (3.6.30)$$

and the corresponding eigenvector sequence

$$\{e_k \in H_B^2(\Omega, \mathbb{C}) \mid k = 1, 2, \dots\} \quad (3.6.31)$$

forms an orthogonal basis of $L^2(\Omega, \mathbb{C})$.

The eigenvalues of (3.6.28) always have even multiplicity, i.e., if ψ is an eigenvector of (3.6.28), then $e^{i\theta}\psi(\theta \in \mathbb{R}^1)$ are also eigenvectors corresponding to the same eigenvalue. Let the first eigenvalue α_1 have multiplicity $2m(m \geq 1)$ with eigenvectors

$$e_{2k-1} = \psi_{k1} + i\psi_{k2}, \quad e_{2k} = -\psi_{k2} + i\psi_{k1} \quad \forall 1 \leq k \leq m. \quad (3.6.32)$$

We know that α_1 enjoys the following properties

$$\begin{aligned} \alpha_1 &= \alpha_1(A_a) && \text{depends continuous on } A_a, \\ \alpha_1(A_a) &> 0 && \text{for } A_a \neq 0, \\ \alpha_1(0) &= 0 && \text{for b.c. (3.6.19),} \\ \alpha_1(0) &> 0 && \text{for b.c. (3.6.20) or (3.6.21).} \end{aligned} \quad (3.6.33)$$

Another crucial eigenvalue problem for (3.6.16) is

$$\begin{aligned} \operatorname{curl}^2 A + \nabla\phi &= \rho A, \\ \operatorname{div} A &= 0, \\ A_n|_{\partial\Omega} &= 0, \quad \operatorname{curl} A \times n|_{\partial\Omega} = 0. \end{aligned} \quad (3.6.34)$$

Here, we point out that the boundary condition in (3.6.34), i.e., (3.6.18), is the free boundary condition, which can be expressed as

$$A_n|_{\partial\Omega} = 0, \quad \frac{\partial A_\tau}{\partial n}|_{\partial\Omega} = 0, \quad (3.6.35)$$

where τ is the tangent vector on $\partial\Omega$; see Ma and Wang (2005a).

The eigenvalues of (3.6.34) are real:

$$0 < \rho_1 \leq \rho_2 \leq \cdots \leq \rho_k \leq \cdots, \quad \lim_{k \rightarrow \infty} \rho_k = +\infty, \quad (3.6.36)$$

and the corresponding eigenvectors

$$\{a_k \in D^2(\Omega, \mathbb{R}^n) | k = 1, 2, \dots\}, \quad (3.6.37)$$

form an orthogonal basis of $\mathcal{L}^2(\Omega, \mathbb{R}^n)$.

Dynamic Transition Theorems In superconductivity, the parameter α in (3.6.16) cannot exceed a maximal value $\alpha(T) \leq \alpha(0)$ because $T \geq 0$. Hence, we have to impose a basic hypothesis:

$$\alpha_1 < \alpha(0) = \frac{2\sqrt{bm_s DN_0}}{e_s^3 h}, \quad (3.6.38)$$

where α_1 is the first eigenvalue of (3.6.28), and N_0 is the density of states at the Fermi level.

We mainly consider the case where the first eigenvalue α_1 of (3.6.28) has multiplicity two, i.e., α_1 is complex simple eigenvalue. We start with the introduction of a

crucial physical parameter, which completely determines the dynamical properties of the phase transition of the Ginzburg-Landau equations.

Let $e \in H_B^2(\Omega, \mathbb{C})$ be a first eigenvector of (3.6.28). Then there is a unique solution for the following equations

$$\begin{aligned} \operatorname{curl}^2 \mathcal{A}_0 + \nabla \phi &= |e|^2 A_a + \frac{\mu}{2} i(e^* \nabla e - e \nabla e^*), \\ \operatorname{div} \mathcal{A}_0 &= 0, \\ \mathcal{A}_0 \cdot n|_{\partial\Omega} &= 0, \quad \operatorname{curl} \mathcal{A}_0 \times n|_{\partial\Omega} = 0. \end{aligned} \tag{3.6.39}$$

Then, we define a physical parameter R as follows

$$R = -\frac{\beta}{\gamma} + \frac{2 \int_{\Omega} |\operatorname{curl} \mathcal{A}_0|^2 dx}{\int_{\Omega} |e|^4 dx}. \tag{3.6.40}$$

From (3.6.39) and (3.6.40) it is easy to see that the parameter R is independent of the choice of the first eigenvectors of (3.6.28) and $H_0 = \operatorname{curl} \mathcal{A}_0$ given by (3.6.39) depend on the applied magnetic potential A_a and the geometric properties of Ω , the parameter R is essentially a function of A_a, Ω and physical parameters β, γ, μ .

The main results in this section are the following theorems.

Theorem 3.6.4 *Assume that the first eigenvalue α_1 of (3.6.28) with one of the boundary conditions (3.6.19)–(3.6.21) is complex simple, and the condition (3.6.38) holds true. If the parameter R defined by (3.6.40) satisfies $R < 0$, then the phase transition of (3.6.16)–(3.6.18) with one of (3.6.19)–(3.6.21) at $\alpha = \alpha_1$ is of Type-I. Moreover, the following assertions hold true.*

1. *The problem bifurcates from $((\psi, \mathcal{A}), \alpha) = (0, \alpha_1)$ to an attractor Σ_α for $\alpha > \alpha_1$, which is homeomorphic to S^1 , and consists of steady-state solutions.*
2. *There is a neighborhood $U \subset H$ of $(\psi, \mathcal{A}) = 0$ such that the attractor Σ_α attracts $U \setminus \Gamma$ in H , where Γ is the stable manifold of $(\psi, \mathcal{A}) = 0$ with codimension two in H .*
3. *Each $(\psi, \mathcal{A}) \in \Sigma_\alpha$ can be expressed as*

$$\begin{aligned} \psi &= \left| \frac{\alpha - \alpha_1}{R_1} \right|^{1/2} e + o(|\alpha - \alpha_1|^{1/2}), \\ \operatorname{curl}^2 \mathcal{A} &= -\gamma \left| \frac{\alpha - \alpha_1}{R_1} \right| \cdot [|e|^2 A_a + \mu I_m(e \nabla e^*)] + o(|\alpha - \alpha_1|), \end{aligned} \tag{3.6.41}$$

where e is the first eigenvector of (3.6.28) and

$$R_1 = \frac{\gamma R \int_{\Omega} |e|^4 dx}{\int_{\Omega} |e|^2 dx}.$$

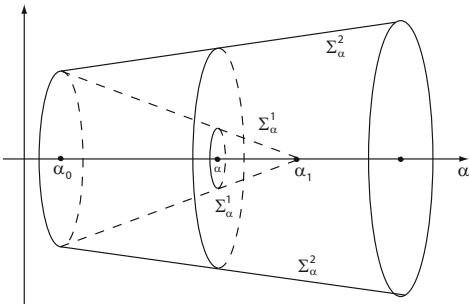
Theorem 3.6.5 *Assume that the first eigenvalue α_1 of (3.6.28) with one of the boundary conditions (3.6.19)–(3.6.21) is complex simple, and the condition (3.6.38) holds*

true. If $R > 0$, then for the problem (3.6.16)–(3.6.18) with one of (3.6.19)–(3.6.21), we have the following assertions:

1. The phase transition at $\alpha = \alpha_1$ is of Type-II.
2. The problem bifurcates from $((\psi, \mathcal{A}), \alpha) = (0, \alpha_1)$ to an invariant set Σ_α on $\alpha < \alpha_1$ and has no bifurcation on $\alpha > \alpha_1$.
3. $\Sigma_\alpha = S^1$ is a circle consisting of steady states, and has a two-dimensional unstable manifold.

By Theorem 3.6.3, the TDGL equations possess a global attractor. It is clear that $(\psi, \mathcal{A}) = 0$ is a unique minimal value point of (3.6.25) for $\alpha = 0$, which implies that if $\alpha = 0$, then $(\psi, \mathcal{A}) = 0$ is globally asymptotically stable for (3.6.16)–(3.6.18) with one of the boundary conditions (3.6.19)–(3.6.21). Thus, the following theorem follows from Theorems 2.5.3 and 3.6.5.

Fig. 3.30 For each $\alpha_0 < \alpha < \alpha_1$, Σ_α is a repeller, containing unstable superconducting states. For each $\alpha_0 < \alpha$, Σ_α^2 is an attractor, containing stable superconducting states



Theorem 3.6.6 Under the assumptions of Theorem 3.6.5. For the case where $R > 0$, there exists a saddle-node bifurcation point $\alpha_0 (0 < \alpha_0 < \alpha_1)$ for the TDGL equations, such that the following statements hold true, which are described schematically in Figs. 3.30 and 3.31:

1. at $\alpha = \alpha_0$, there is a cycle $\Sigma_0 = \Sigma_{\alpha_0}$ of singular points with $0 \notin \Sigma_0$.
2. For $\alpha < \alpha_0$, there are no invariant sets near Σ_0 .
3. For $\alpha_0 < \alpha < \alpha_1$ there are two connected branches of cycles Σ_α^1 and Σ_α^2 of singular points, and Σ_α^2 extends to $\alpha \geq \alpha_1$.
4. For each $\alpha > \alpha_1$, Σ_α^2 is an attractor with $\text{dist}(\Sigma_\alpha^2, 0) > 0$ at $\alpha = \alpha_1$.
5. For $\alpha_0 < \alpha < \alpha_1$, Σ_α^1 is a repeller with $0 \notin \Sigma_\alpha^1$ and $\Sigma_\alpha^1 \rightarrow \{0\}$ as $\alpha \rightarrow \alpha_1^-$.

Remark 3.6.7 The parameter R defined by (3.6.40) can be equivalently expressed as follows

$$R = -\frac{\beta}{\gamma} + \frac{2 \sum_{k=1}^{\infty} \frac{1}{\rho_k} \left[\int_{\Omega} (|e|^2 A_a + 2\mu e_2 \nabla e_1) \cdot a_k \right]^2}{\int_{\Omega} |e|^4 dx} \quad (3.6.42)$$

where $e = e_1 + ie_2$, ρ_k are the eigenvalues of (3.6.34) given by (3.6.36) and $\{a_k\}$ the normalized eigenvectors given by (3.6.37).

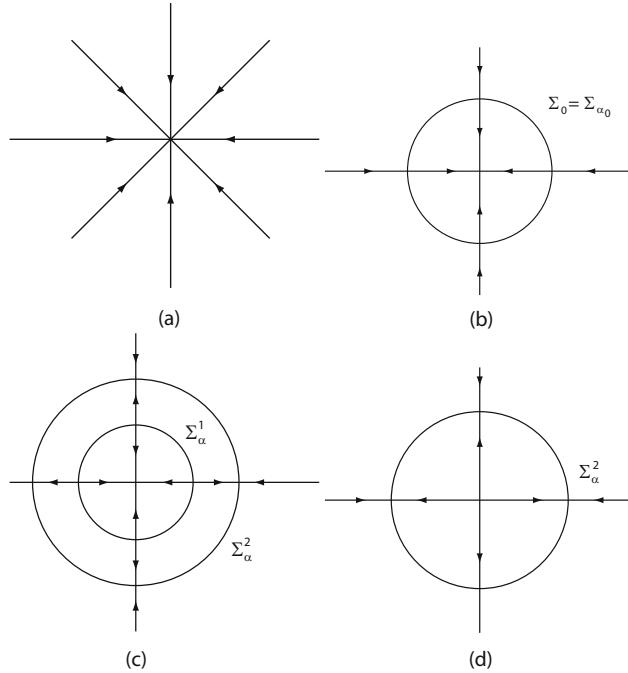


Fig. 3.31 Phase diagram on the center manifold for various α : (a) case $\alpha < \alpha_0$, (b) case $\alpha = \alpha_0$, (c) case $\alpha_0 < \alpha < \alpha_1$, (d) case $\alpha_1 \leq \alpha$

Proof of Theorems 3.6.4 and 3.6.5. We proceed in several steps as follows.

STEP 1. It is easy to see that $L_\alpha : H_1 \rightarrow H$ is a sectorial operator, and the eigenvalues of L_α satisfy that

$$\beta_1(\alpha) = \beta_2(\alpha) = \alpha - \alpha_1 \begin{cases} < 0 & \text{if } \alpha < \alpha_1, \\ = 0 & \text{if } \alpha = \alpha_1, \\ > 0 & \text{if } \alpha > \alpha_1, \end{cases}$$

and for $j \geq 3$,

$$\begin{aligned} \beta_j(\alpha_1) &= \alpha_1 - \alpha_k, \quad \text{or } \beta_j(\alpha_1) = -\zeta^{-1} \rho_l, \\ \beta_j(\alpha_1) &< 0, \end{aligned}$$

for all $k > 1$ and $l \geq 1$. Namely, the conditions (2.1.2), (2.1.4), and (2.1.5) are satisfied by (3.6.24). It is clear that the operator $\Phi : H_1 \rightarrow L^2(\Omega, C)$ defined by (3.6.23) is C^∞ , and it is easy to derive the following estimate:

$$\begin{aligned} \int_{\Omega} |\Phi(u)\psi|^2 dx &\leq \left[\int_{\Omega} |\Phi(u)|^3 dx \right]^{2/3} \left[\int_{\Omega} |\psi|^6 dx \right]^{1/3} \\ &\leq C \left(\|u\|_{H_{1/2}}^2 + \|u\|_{H_{1/2}} \|\psi\|_{L^4}^2 \right)^2 \|\psi\|_{L^6}^2, \end{aligned}$$

where $H_{1/2}$ is the closure of H_1 for the H^1 -norm. Hence, it is not difficult to verify that there is a number $1/2 < \sigma < 1$ such that $G : H_\sigma \rightarrow H$ is C^∞ . Thus, the condition (2.1.3) holds true.

STEP 2. It is known that the transition of (3.6.24) is determined by its reduced equation to the center manifold. Let

$$\psi_0 \in E = \{ze| z \in \mathbb{C}, e \text{ the first eigenvector of (3.6.28)}\}.$$

Then the reduced equation of (3.6.24) reads

$$\frac{d\psi_0}{dt} = \beta_1(\alpha)\psi_0 - P_1 G(\psi_0 + \tilde{\psi}(\psi_0), \tilde{\mathcal{A}}(\psi_0)), \quad (3.6.43)$$

where $P_1 : H \rightarrow E_1$ is the canonical projection, and $\tilde{\Phi}(\psi_0) = (\tilde{\psi}(\psi_0), \tilde{\mathcal{A}}(\psi_0)) \in H_1$ is the center manifold function.

The k multi-linear operators ($k = 2, 3$) in G are given by

$$\begin{aligned} G_2(u) &= -\left(\frac{2A_a \cdot \mathcal{A}\psi + 2i\mu\mathcal{A} \cdot \nabla\psi}{\zeta^{-1}\gamma A_a |\psi|^2 + \frac{\gamma\mu}{2\zeta} i(\psi^* \nabla\psi - \psi \nabla\psi^*)} \right), \\ G_3(u) &= -\left(\frac{i\psi \Phi_2(u) + |\mathcal{A}|^2\psi + \beta|\psi|^2\psi}{\gamma\zeta^{-1}\mathcal{A}|\psi|^2} \right), \end{aligned}$$

where $\Phi_2(u)$ is the bilinear operator in $\Phi(u)$.

By the first approximation of the center manifold reduction, the center manifold function $\tilde{\Phi} = (\tilde{\psi}(\psi_0), \tilde{\mathcal{A}}(\psi_0))$ satisfies that

$$\begin{aligned} \operatorname{curl}^2 \tilde{\mathcal{A}} + \mu \nabla\phi &= -\gamma A_a |\psi_0|^2 - \frac{\gamma\mu}{2} i(\psi_0^* \nabla\psi_0 - \psi_0 \nabla\psi_0^*) \\ &\quad + o(\|\psi_0\|^2, |\beta_1(\alpha)| \cdot \|\psi_0\|), \end{aligned} \quad (3.6.44)$$

$$\tilde{\psi} = O(\|\tilde{\mathcal{A}}\| \cdot \|\psi_0\|) = O(\|\psi_0\|^3). \quad (3.6.45)$$

Based on (3.6.44) and (3.6.45), (3.6.43) can be expressed as

$$\frac{d\psi_0}{dt} = \beta_1(\alpha)\psi_0 - g_3(\psi_0) + o(\|\psi_0\|^3) + O(\|\psi_0\|^3 \cdot |\beta_1(\alpha)|), \quad (3.6.46)$$

where

$$g_3(\psi_0) = P_1[\beta|\psi_0|^2\psi_0 + 2A_a \cdot \tilde{\mathcal{A}}_2\psi_0 + 2i\mu\tilde{\mathcal{A}}_2 \cdot \nabla\psi_0 + i\Phi_2\psi_0], \quad (3.6.47)$$

$$\begin{cases} \operatorname{curl}^2 \tilde{\mathcal{A}}_2 + \nabla\phi = -\gamma A_a |\psi_0|^2 - \frac{\gamma\mu}{2} i(\psi_0^* \nabla\psi_0 - \psi_0 \nabla\psi_0^*), \\ \operatorname{div} \tilde{\mathcal{A}}_2 = 0, \\ \tilde{\mathcal{A}}_2 \cdot n|_{\partial\Omega} = 0, \quad \operatorname{curl} \tilde{\mathcal{A}}_2 \times n|_{\partial\Omega} = 0. \end{cases} \quad (3.6.48)$$

Equations (3.6.46)–(3.6.48) are the second-order approximation of the reduction of (3.6.24) to the center manifold.

STEP 3. Let $\psi_0 = \psi_1^0 + i\psi_2^0$. We infer from (3.6.47) that

$$\begin{aligned} & (g_3(\psi_0), \psi_0) \\ &= \operatorname{Re} \int_{\Omega} g_3(\psi_0) \cdot \psi_0^* dx \\ &= \int_{\Omega} \left[\beta |\psi_0|^4 + 2|\psi_0|^2 A_a \cdot \tilde{\mathcal{A}}_2 + 2\mu \tilde{\mathcal{A}}_2 (\psi_2^0 \nabla \psi_1^0 - \psi_1^0 \nabla \psi_2^0) \right] dx \\ &= \int_{\Omega} \left[\beta |\psi_0|^4 + 2|\psi_0|^2 A_a \cdot \tilde{\mathcal{A}}_2 + 4\mu \psi_2^0 \tilde{\mathcal{A}}_2 \cdot \nabla \psi_1^0 \right] dx. \end{aligned} \quad (3.6.49)$$

Let $\tilde{\mathcal{A}}_2$ have the Fourier expansion for the basis (3.6.37) of $\mathcal{L}^2(\Omega, \mathbb{R}^n)$ as $\tilde{\mathcal{A}} = \sum_{k=1}^{\infty} y_k a_k$. Then, for (3.6.48) we can derive the solutions y_k :

$$y_k = -\frac{\gamma}{\rho_k} \int_{\Omega} [|\psi_0|^2 A_a \cdot a_k + 2\mu \psi_2^0 \nabla \psi_1^0 \cdot a_k] dx.$$

Inserting y_k into (3.6.49) we find

$$\begin{aligned} (g_3(\psi_0), \psi_0) &= \beta \int_{\Omega} |\psi_0|^4 dx - 2\gamma \sum_{k=1}^{\infty} \frac{1}{\rho_k} \left[\left(\int_{\Omega} |\psi_0|^2 A_a \cdot a_k dx \right)^2 \right. \\ &\quad + 4\mu \left(\int_{\Omega} |\psi_0|^2 A_a \cdot a_k dx \right) \left(\int_{\Omega} \psi_2^0 \nabla \psi_1^0 \cdot a_k dx \right) \\ &\quad \left. + 4\mu^2 \left(\int_{\Omega} \psi_2^0 \nabla \psi_1^0 \cdot a_k dx \right)^2 \right]. \end{aligned} \quad (3.6.50)$$

Let $\psi_0 = x_1 e_1 + x_2 e_2$, where $(x_1, x_2) \in \mathbb{R}^2$, and e_1 and e_2 are as in (3.6.32); then

$$\psi_0 = \psi_1^0 + i\psi_2^0 = (x_1 \psi_{11} - x_2 \psi_{12}) + i(x_1 \psi_{12} + x_2 \psi_{11}).$$

Hence

$$\begin{aligned} \int_{\Omega} |\psi_0|^4 dx &= \int_{\Omega} (|\psi_1^0|^2 + |\psi_2^0|^2)^2 dx = (x_1^2 + x_2^2)^2 \int_{\Omega} |e_1|^4 dx, \\ \int_{\Omega} |\psi_0|^2 A_a \cdot a_k dx &= (x_1^2 + x_2^2) \int_{\Omega} |e_1|^2 A_a \cdot a_k dx, \\ \int_{\Omega} \psi_2^0 \nabla \psi_1^0 \cdot a_k dx &= (x_1^2 + x_2^2) \int_{\Omega} \psi_{12} \nabla \psi_{11} \cdot a_k dx. \end{aligned} \quad (3.6.51)$$

Putting (3.6.51) into (3.6.50) we obtain

$$(g_3(\psi_0), \psi_0) = -\gamma R(x_1^2 + x_2^2)^2 \int_{\Omega} |e_1|^4 dx, \quad (3.6.52)$$

where R is as in (3.6.42). It is easy to see that both numbers in (3.6.40) and (3.6.42) are the same.

STEP 4. PROOF OF THEOREM 3.6.4. Since the TDGL equations are of the gradient-type structure, the Krasnoselskii bifurcation theorem for the potential operators, Theorem 1.12 in Ma and Wang (2005b), is valid for (3.6.16). Hence the problem (3.6.16)–(3.6.18) with one of the boundary conditions (3.6.19)–(3.6.21) must bifurcate from $((\psi, \mathcal{A}), \alpha) = (0, \alpha_1)$ to at least one steady-state solution.

When $R < 0$, by (3.6.46) and (3.6.52) we infer from Theorem 2.1.6 that the transition at $\alpha = \alpha_1$ is Type-I, and from Theorem 2.2.4 that the Ginzburg-Landau equations bifurcate from $(0, \alpha_1)$ to a cycle of attractor Σ_α for $\alpha > \alpha_1$, which contains at least one singular point. Thanks to the invariance of the TDGL equations for the gauge transformation $\psi \rightarrow \psi e^{i\theta}$ for $\theta \in \mathbb{R}^1$, the steady-state solutions of the Ginzburg-Landau equations appear as a circle S^1 . Therefore the attractor $\Sigma_\alpha = S^1$ consists of steady-state solutions. Assertion (1) is proved.

Assertion (2) follows from Theorem 2.2.11, and Assertion (3) can be directly derived from the equations (3.6.46) and (3.6.52). Thus, Theorem 3.6.4 is proved.

STEP 5. PROOF OF THEOREM 3.6.5. When $R > 0$, by (3.6.46) and (3.6.52) we infer from Theorem 2.1.6 that the transition is II-type. In addition, the time-reversed semigroup $S_\alpha(-t)$ generated by (3.6.46) has the same dynamic properties as the following equation

$$\frac{d\psi_0}{dt} = (\alpha_1 - \alpha)\psi_0 + g_3(\psi_0) + o(\|\psi_0\|^3, |\alpha - \alpha_0|\|\psi_0\|^2). \quad (3.6.53)$$

In the same fashion as used in Step 4, from (3.6.52) and (3.6.53) we deduce that the semigroup $\tilde{S}_\alpha(t)$ generated by (3.6.53) bifurcates from $(\psi_0, \alpha) = (0, \alpha_1)$ to a S^1 attractor Σ_α for $\alpha < \alpha_1$, which consists of singular points of (3.6.53). Hence, Assertions (1)–(3) hold true for the semigroup $S_\alpha(t)(= \tilde{S}_\alpha(-t))$ generated by (3.6.46) with $R > 0$.

Thus Theorem 3.6.5 is proved. \square

Remark 3.6.8 We know that the first eigenvalues α_1 with complex simplicity is generic. Namely, for each bounded open set $\Omega \subset \mathbb{R}^n$ there is an open and dense set $U \subset L^2(\Omega, \mathbb{R}^n)$ such that for any $A_a \in U$ the first eigenvalue $\alpha_1(A_a)$ of (3.6.28) with one of the boundary conditions (3.6.19)–(3.6.21) is complex simple. In particular, when A_a is small, $\alpha_1(A_a)$ must be complex simple.

3.6.4 Model Coupled with Entropy

According to the thermodynamic theory, the first-order phase transitions (i.e., the Type-II and Type-III phase transitions) release latent heat. Therefore, it is necessary to take the entropy into account in the Ginzburg-Landau model. To this end, we revise the Ginzburg-Landau free energy (3.6.25) by

$$E = \int_{\Omega} \left[\frac{1}{2} |(i\mu\nabla + \mathcal{A} + A_a)\psi|^2 - \frac{\alpha}{2} |\psi|^2 + \frac{\beta}{4} |\psi|^4 + \frac{\rho}{6} |\psi|^6 + \frac{1}{2} k_1 |\nabla S|^2 + \frac{1}{2} k_2 |S|^2 + k_3 S |\psi|^2 + \gamma^{-1} |\operatorname{curl} \mathcal{A}|^2 \right] dx, \quad (3.6.54)$$

where $\rho > 0$ and $k_i > 0$ are nondimensional constants, S is the entropy density deviation. As before, we can derive easily the TDGL corresponding to the free energy (3.6.54). Replace the parameter R defined in (3.6.40) by

$$\tilde{R} = R + \frac{k_3^2}{\gamma} \frac{\int_{\Omega} S_1 |e|^2 dx}{\int_{\Omega} |e|^4 dx}, \quad (3.6.55)$$

where S_0 satisfies the following equation

$$-k_1 \Delta S_1 + k_2 S_1 = |e|^2, \quad \frac{\partial S_1}{\partial n} |_{\partial\Omega} = 0. \quad (3.6.56)$$

Then we have the following dynamic transition theorem for the TDGL corresponding to the free energy (3.6.54):

Theorem 3.6.9 *If the parameter \tilde{R} given by (3.6.55) satisfies $\tilde{R} < 0$, then the statements of Theorem 3.6.4 hold true, and if $\tilde{R} > 0$, then the statements of Theorems 3.6.5 and 3.6.6 hold true. In particular, the entropy has a jump at the critical temperature T_c for $\tilde{R} > 0$:*

$$\delta S = S(T_c^-) - S(T_c^+) = -k_3(-k_1 \Delta + k_2)^{-1} |\psi(T_c)|^2,$$

where $\psi(T_c)$ is the stable superconducting state at $T = T_c$.

Physically, the entropy gradient δS implies that the first-order transition from the normal state to superconducting states (resp. from the superconducting states to normal state) is exothermic (resp. is endothermic), with the latent heat given by

$$\delta H = T_c \delta S = -T_c k_3 (-k_1 \Delta + k_2)^{-1} |\psi(T_c)|^2.$$

3.6.5 Physical Conclusions

General Remarks The permanent current, called supercurrent, is expressed in the Ginzburg-Landau equations by (3.6.6). In the steady-state case, the supercurrent in the second type of nondimensional form is written as

$$J_s = J_s(\psi, \mathcal{A}) = -\gamma(A_a + \mathcal{A})|\psi|^2 - \frac{\gamma\mu}{2} i(\psi^* \nabla \psi - \psi \nabla \psi^*). \quad (3.6.57)$$

We take the Meissner effect into account in the Ginzburg-Landau equations. Mathematically speaking, in the normal state, the magnetic field H in a sample

should be $H = H_a + \mathcal{H}$, $H_a = \operatorname{curl} A_a$ is the applied field, and $\mathcal{H} = \operatorname{curl} \mathcal{A}$ the nonequilibrium fluctuation, and in the superconducting state $H = \operatorname{curl} \mathcal{A}$. In the both cases, \mathcal{A} satisfy the Ginzburg-Landau equations (3.6.16) and boundary condition (3.6.18). Namely, we can express the magnetic field H in a sample Ω in the following form

$$H = \operatorname{curl} A, \quad A = \begin{cases} A_a + \mathcal{A} & \text{in the normal state,} \\ \mathcal{A} & \text{in the superconducting state,} \end{cases}$$

and the supercurrent J_s in the nondimensional form is given by

$$J_s = \operatorname{curl}^2 \mathcal{A}, \quad (3.6.58)$$

where \mathcal{A} is the superposition of the magnetic potentials associated with supercurrent and nonequilibrium fluctuation, which satisfies (3.6.16) and (3.6.18).

An equilibrium state $(\tilde{\psi}, \tilde{\mathcal{A}})$ of the TDGL equation (3.6.16) is called in the normal state if $\tilde{\psi} = 0$, and it is called in the superconducting state if $\tilde{\psi} \neq 0$. A solution (ψ, \mathcal{A}) of (3.6.16) is said in the normal state if (ψ, \mathcal{A}) is in a domain of attraction of a normal equilibrium state, otherwise (ψ, \mathcal{A}) is said in the superconducting state.

We consider the simplest case where the applied field vanishes $A_a = 0$. In this case, the eigenvalue equation (3.6.28) becomes

$$-\mu\Delta\psi = \alpha\psi. \quad (3.6.59)$$

The first eigenvalue α_1 of (3.6.59) with one of the boundary conditions (3.6.19)–(3.6.21) is complex simple, and the eigenvector is as $e^{i\theta}\psi_0$ with ψ_0 a real function. Therefore, the parameter R defined by (3.6.40) reads

$$R = -\frac{\beta}{\gamma} < 0.$$

By Theorem 3.6.4, when the parameter $\alpha(T) \leq \alpha_1$ the solutions (ψ, \mathcal{A}) of (3.6.58) are in the normal states, and when $\alpha(T) > \alpha_1$, (ψ, \mathcal{A}) with initial values (ψ_0, \mathcal{A}_0) in $U \setminus \Gamma$ are in the superconducting states.

As $A_a = 0$, the steady-state solutions $(\tilde{\psi}, \tilde{\mathcal{A}})$ of (3.6.58) are real, i.e., $\tilde{\psi} = e^{i\theta}\psi$, $\operatorname{Im}\psi = 0$. Hence we have

$$\tilde{\psi}^* \nabla \tilde{\psi} - \tilde{\psi} \nabla \tilde{\psi}^* = 0,$$

which implies that $\tilde{\mathcal{A}} = 0$. Thus, the supercurrent (3.6.58) vanishes: $J_s = 0$. This shows that with zero applied field $H_a = 0$, there is no supercurrent in a superconductor.

Implications of (3.6.38) For the case with the Neumann boundary condition (3.6.19), e.g., for a sample enclosed by an insulator, the first eigenvalue $\alpha_1 = 0$ for (3.6.59), which is independent of Ω , the geometry of the sample. Therefore, the condition (3.6.38) always holds true.

However for the Dirichlet and the Robin boundary conditions (3.6.20) and (3.6.21), the situation is different. We know that the first eigenvalue α_1 of (3.6.59) depends on Ω . In particular,

$$\alpha_1 = \alpha_1(\Omega) \rightarrow \infty \quad \text{if } |\Omega| \rightarrow 0.$$

Hence, the condition (3.6.38) implies that for the cases where the sample is enclosed by a magnetic material or a normal metal, the volume of a sample must be greater than some critical value $|\Omega| > V_c > 0$. Otherwise no superconducting state occurs at any temperature. This property also holds true for the case where there is an applied magnetic field H_a present. Of course, in this case, the critical volume V_c depends on H_a as well. In superconductivity, it again shows that it is a universal physical phenomenon that the dynamical properties of phase transitions depend on the geometry of domain Ω .

Physical theory and experiments show that there is a critical applied magnetic field H_c by which a superconducting state will be destroyed, and H_c satisfies the following approximative equation near the critical temperature T_c which is given by (3.6.2)

$$H_c(T) = H_0(1 - T^2/T_c^2) \quad (3.6.60)$$

Equation (3.6.60) is an empirical formula. A related equation is the equation of critical parameter (1.2.6) for superconductivity, which reads

$$\alpha(T) = \alpha_1(H_a, \Omega) \quad (3.6.61)$$

where $\alpha(T)$ is the parameter in (3.6.16) and $\alpha_1 = \alpha_1(H_a, \Omega)$ the first eigenvalue of (3.6.28), $H_a = \operatorname{curl} A_a$. It is expected that the applied magnetic field H_a satisfying (3.6.61) is the critical field H_c .

Transitions with $R < 0$ As mentioned earlier, there are two phase transitions: Type-I and Type-II, determined by a simple parameter R defined by (3.6.40). This parameter R links the superconducting behavior with the geometry of the material, the applied field, and the physical parameters.

Let a magnetic field $H_a = \operatorname{curl} A_a$ be applied. By the phase transition theorems, the critical temperature T_c^1 of superconducting transition satisfies $T_c^1 < T_c$ where T_c is as in (3.6.2) and $T = T_c^1$ satisfies (3.6.61). It is known that

$$\alpha_1(H_a, \Omega) \rightarrow \infty \quad \text{if } |A_a| \rightarrow \infty.$$

It implies that the applied magnetic field H_a cannot be very strong for superconductivity as required by condition (3.6.38).

From Theorems 3.6.4–3.6.6, we see that the number R defined by (3.6.40) is an important parameter to distinguish two different types of superconducting transitions. We first examine the case where $R < 0$.

By Theorem 3.6.4, when $\alpha > \alpha_1$, the equations (3.6.16) bifurcate from $((\psi, \mathcal{A}), \alpha) = (0, \alpha_1)$ to a cycle of steady-state solutions $(\psi_\alpha, \mathcal{A}_\alpha)$ which is an attrac-

tor attracting an open set $U \setminus \Gamma \subset H$. From the physical point of view, this theorem leads to the following properties of superconducting transitions in the case where $R < 0$.

Physical Conclusion 3.6.10 *Let $R < 0$ and $T = T_c^1$ satisfy (3.6.61). Then the following physical properties hold true.*

1. *When the control temperature decreases (resp. increases) and crosses the critical temperature T_c^1 , there will be a phase transition of the sample from the normal to superconducting states (resp. from the superconducting to normal state).*
2. *(Stability) When the control temperature $T > T_c^1$, under a fluctuation deviating the normal state, the sample will be soon restored to the normal state. In addition, when $T < T_c^1$, under a fluctuation deviating both the normal and superconducting states, the sample will be soon restored to the superconducting states.*
3. *In general, the supercurrent given by*

$$J_s(\alpha) = -\gamma(A_\alpha + \mathcal{A}_\alpha)|\psi_\alpha|^2 - \frac{\gamma\mu}{2}i(\psi_\alpha^*\nabla\psi_\alpha - \psi_\alpha\nabla\psi_\alpha^*)$$

is nonzero, i.e., $J_s \neq 0$ for $T < T_c^1 (\alpha_1 < \alpha)$.

4. *(Continuity) The order parameter ψ_α and supercurrent $J_s(\alpha)$ depend continuously on the control temperature T (or the parameter α), namely*

$$\psi_\alpha \rightarrow 0, \quad J_s(\alpha) \rightarrow 0, \quad \text{if } T \rightarrow (T_c^1)^- \quad (\text{or } \alpha \rightarrow \alpha_1^+).$$

5. *The superconducting state of the system is dominated by the lowest energy eigenfunction of (3.6.28) in the sense given by (3.6.41).*
6. *The phase transition is of second order in the Ehrenfest sense with the critical exponent $\beta = \frac{1}{2}$, and its phase diagram is as shown in Fig. 3.32.*

Transitions with $R > 0$ Phase transitions in this case are precisely described by Theorems 3.6.5–3.6.9, as shown in Figs. 3.30 and 3.31. In particular, we have the following conclusions, which recapitulate some superconducting properties obtained in Theorems 3.6.5–3.6.9 in physical terms.

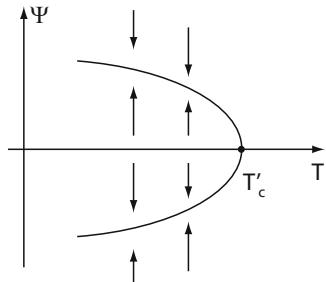
Physical Conclusion 3.6.11 *Consider a material described by the TDGL model with $R > 0$. There are two transition temperature T_c^0 and $T_c^1 (T_c^0 > T_c^1)$ such that*

$$\alpha(T_c^i) = \alpha_i \quad (i = 0, 1) \quad \text{with} \quad \alpha_0 < \alpha_1,$$

and the following phase transition properties hold true:

1. *When the control temperature T decreases and crosses T_c^1 or equivalently α increases and crosses α_1 , the stability of the normal state changes from stable to unstable.*
2. *When $T_c^1 < T < T_c^0 (\alpha_0 < \alpha < \alpha_1)$, physically observable states consist of the normal state and the superconducting states in Σ_α^2 (see Fig. 3.30). When $T < T_c^1 (\alpha_1 < \alpha)$, physically observable states are in Σ_α^1 .*

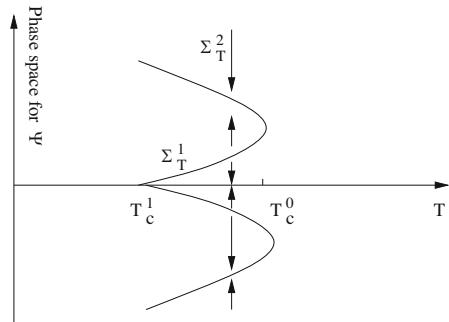
Fig. 3.32 Type-I transition in the case where $R < 0$



3. (Instability) When the control temperature T is in the interval $T_c^1 < T < T_c^0$ (or $\alpha_0 < \alpha < \alpha_1$), the superconducting states in Σ_α^1 are unstable, i.e., with a fluctuation deviating a superconducting state in Σ_α^1 , transition to either the normal state or a superconducting state in Σ_α^2 will occur.
4. (Discontinuity) At the critical temperature T_c^0 (resp. at T_c^1) of the phase transitions, there is a jump from the superconducting states to the normal state (resp. from the normal state to superconducting states).
5. The other energy-level eigenfunctions possibly have a stronger influence for the superconducting states.
6. In the temperature interval $T_c^1 < T < T_c^0$, phase transitions occur and are accompanied with the latent heat to appear.

Based on the conclusions (1)–(4) above, we can draw the phase diagram in Fig. 3.33, where the superconducting states in Σ_T^2 and the normal state $\psi = 0$ are metastable for $T_c^1 < T < T_c^0$, and Σ_T^2 is stable and $\psi = 0$ is unstable for $T < T_c^1$. Physically speaking, the metastable superconducting phases are the superheated states, and the metastable normal phase is the supercooled state.

Fig. 3.33 The normal state $\psi = 0$ in $T_c^1 < T < T_c^0$ is supercooled, the superconducting states Σ_T^2 is superheated



Physical Significance of R Physically, it is considered that the phase transition is of the second order if the applied magnetic field $H_a = 0$, and is of the first order if $H_a \neq 0$. However, by Theorems 3.6.4–3.6.6 we know that the two types of continuous and discontinuous transitions should be distinguished by the parameter

R. From (3.6.39) we can deduce that for given material and Ω there is a critical magnetic field intensity $\tilde{H}_c^2 > 0$ such that

$$R = R(H_a) \begin{cases} < 0 & \text{if } H_a^2 < \tilde{H}_c^2, \\ > 0 & \text{if } H_a^2 > \tilde{H}_c^2. \end{cases} \quad (3.6.62)$$

Hence, when $H_a^2 < \tilde{H}_c^2$ the phase transition is of the second order, and when $H_a^2 > \tilde{H}_c^2$ it is of the zero order, i.e., the Ginzburg-Landau free energy E in (3.6.25) has a jump at critical temperature.

We now consider the influence of physical parameters to R . It is noted that the parameter β/γ can be characterized by the Ginzburg-Landau number κ and the parameter μ :

$$\frac{\beta}{\gamma} = \kappa^2 \mu^2, \quad \kappa^2 = \frac{m_s^2 C^2}{2\pi e_s^2 h^2} b, \quad \mu^2 = \frac{h^2 D^2}{e_s^2 b}. \quad (3.6.63)$$

In the Ginzburg-Landau energy, the term

$$E_0 = \int_{\Omega} |e|^4 dx \quad (3.6.64)$$

represents the nonlinear part of the energy of the superconducting electrons in the lowest-energy state, and the term

$$E_m = \int_{\Omega} H_0^2 dx, \quad H_0 = \operatorname{curl} \mathcal{A}_0 \text{ satisfies (3.6.39)}, \quad (3.6.65)$$

is the energy contributed by the magnetic field associated with the supercurrent

$$\operatorname{curl} H_0 = |e|^2 A_a + \frac{\mu}{2} i(e^* \nabla e - e \nabla e^*),$$

which is generated by the applied magnetic potential A_a and the superconducting electrons in the lowest energy states.

By (3.6.40), we obtain from (3.6.63) to (3.6.65) that

$$R = -\kappa^2 \mu^2 + \frac{2E_m}{E_0}.$$

Therefore, the type of superconducting phase transitions among the two described above for a given material depends on the “competition” between the two energy E_0 and E_m :

$$R \begin{cases} < 0 & \text{if } \frac{\kappa^2}{2} E_0 > \frac{1}{\mu^2} E_m, \\ > 0 & \text{if } \frac{\kappa^2}{2} E_0 < \frac{1}{\mu^2} E_m. \end{cases} \quad (3.6.66)$$

According to the Abrikosov theory, the materials with $\kappa^2 < \frac{1}{2}$ and $\kappa^2 > \frac{1}{2}$ are of types I and II respectively. From (3.6.66), we infer that for given geometric shape of sample and applied magnetic field, a type I material favors more to the jumped

phase transition (i.e., the case ($R > 0$)), and a type II material favors the continuous phase transition (i.e., the case ($R < 0$)).

Finally, we need to consider the action of entropy. By (3.6.55) and (3.6.56), the parameter \tilde{R} associated with entropy is given by

$$\tilde{R} = -\kappa^2 \mu^2 + \frac{2E_m}{E_0} + \frac{k_3^2}{\gamma} \frac{\int_{\Omega} [|e|^2(-k_1 \Delta + k_2)^{-1}|e|^2] dx}{E_0} \quad (3.6.67)$$

If we neglect the term $-k_1 \Delta$ in (3.6.67), then we have

$$\tilde{R} = -\kappa^2 \mu^2 + \frac{k_3^2}{\gamma k_2} + \frac{2E_m}{E_0}, \quad (3.6.68)$$

where the parameters $k_2, k_3 > 0$ depend on the materials.

It is known that $E_m = 0$ if $H_a = 0$. For general materials physical experiments showed that when $H_a = 0$ the superconducting phase transitions are continuous or I-type, which implies that for most materials the parameters in \tilde{R} satisfy that

$$\frac{k_3^2}{\gamma k_2} - \kappa^2 \mu^2 \leq 0.$$

However, we cannot exclude the existence of such materials in which the parameters satisfy

$$\frac{k_3^2}{\gamma k_2} - \kappa^2 \mu^2 > 0. \quad (3.6.69)$$

For the materials possessing property (3.6.69), by (3.6.68) and Theorem 3.6.9, the superconducting phase transition is Type-II.

Critical Length Scale and Strength of Applied Field Let $H_a = KH_a^0 = K\operatorname{curl} A_a^0$ with K being the strength of the applied field. Then the parameter R can be written as

$$R = -\kappa^2 \mu^2 + 2(K^2 q_2 + K q_1 + q_0). \quad (3.6.70)$$

Here the parameters in (3.6.70) are given by

$$q_2 = \frac{\int_{\Omega} |\operatorname{curl} \mathcal{A}_{01}|^2 dx}{\int_{\Omega} |e|^4 dx}, \quad q_1 = \frac{2 \int_{\Omega} \operatorname{curl} \mathcal{A}_{01} \cdot \operatorname{curl} \mathcal{A}_{02} dx}{\int_{\Omega} |e|^4 dx}, \quad q_0 = \frac{\int_{\Omega} |\operatorname{curl} \mathcal{A}_{02}|^2 dx}{\int_{\Omega} |e|^4 dx},$$

and \mathcal{A}_{01} and \mathcal{A}_{02} solve the following Stokes problems:

$$\begin{cases} \operatorname{curl}^2 \mathcal{A}_{01} + \nabla \phi = |e|^2 A_a^0, \\ \operatorname{div} \mathcal{A}_{01} = 0, \\ \mathcal{A}_{01} \cdot n|_{\partial\Omega} = 0, \quad \operatorname{curl} \mathcal{A}_{01} \times n|_{\partial\Omega} = 0. \end{cases} \quad (3.6.71)$$

$$\begin{cases} \operatorname{curl}^2 \mathcal{A}_{02} + \nabla \phi = \frac{\mu}{2} i(e^* \nabla e - e \nabla e^*), \\ \operatorname{div} \mathcal{A}_{02} = 0, \\ \mathcal{A}_{02} \cdot n|_{\partial\Omega} = 0, \quad \operatorname{curl} \mathcal{A}_{02} \times n|_{\partial\Omega} = 0. \end{cases} \quad (3.6.72)$$

It is easy to see that q_0, q_1 and q_2 satisfy

$$\begin{aligned} 0 < C_1 \leq q_2 \leq C_2, \quad 0 \leq q_1, q_0 \leq C_2, \\ q_0 = q_1 = 0 \quad &\text{if } K = 0 \end{aligned}$$

Hence, it is easy to verify that there is a unique solution $K_c > 0$ solving

$$R = -\kappa^2 \mu^2 + 2(K^2 q_2 + K q_1 + q_0) = 0.$$

Hence by (3.6.38), we derive the following physical prediction:

Physical Conclusion 3.6.12 (Prediction of Critical Magnetic Strength) *There is a critical magnetic strength K_c such that for $0 \leq K < K_c$, the transition is of second order, and for $K > K_c$, the transition is of first order, provided that*

$$\alpha_1(K_c \tilde{A}_a) < \frac{2\sqrt{b}m_s DN_0}{e_s^3 h}.$$

Formula of R in Terms of Sample Size In a case to be defined below, the formula for R given by (3.6.40) reduces to

$$R(L, H) = -\kappa^2 \mu^2 + 2L^2(p_3 L^4 + p_2 L^2 + p_1). \quad (3.6.73)$$

Here the notations are given as follows:

First, for any $0 < L < \infty$ and $h > 0$, let

$$\begin{aligned} \Omega_0 &= D_0 \times (0, h) \subset \mathbb{R}^3, \quad x' = (x_1, x_2) \in D_0, \quad 0 \in D_0, \\ \Omega(L) &= \{(Lx', x_3) | x' \in D_0, 0 < x_3 < h\} = (LD_0) \times (0, h). \end{aligned}$$

Second, the applied field is $H_a = H_a(x') = (0, 0, H(x'))$, where $x' = (x_1, x_2) \in D_0$. Then H_a induces an applied field \tilde{H}_a on $\Omega(L)$ by $\tilde{H}(y) = H(y/L)$ for any $y = Lx'$ with $x' \in D_0$. Also, we let $H_a = \operatorname{curl} A_a$, i.e., $H = \frac{\partial A_1}{\partial x_2} - \frac{\partial A_2}{\partial x_1}$, where

$$A_a = (A_1(x'), A_2(x'), 0). \quad (3.6.74)$$

Third the parameters in (3.6.73) are given by

$$\begin{cases} p_3 = \frac{\int_{D_0} |\operatorname{curl} A_0|^2 dx'}{\int_{D_0} |e|^4 dx'} > \delta > 0, \\ \operatorname{curl}^2 A_0 + \nabla \phi = |e|^2 A_a \quad \text{in } x' \in D_0, \\ p_1 = \frac{4\mu^2 \int_{D_0} |\operatorname{curl} B_0|^2 dx'}{\int_{D_0} |e|^4 dx'}, \\ \operatorname{curl}^2 B_0 + \nabla \phi = e_{12} \nabla e_{11} \quad \text{in } x' \in D_0, \end{cases}$$

$$p_2 = \frac{4\mu \int_{D_0} \operatorname{curl} A_0 \cdot \operatorname{curl} B_0 dx'}{\int_{D_0} |e|^4 dx'}.$$

Here $\delta > 0$ is independent of L , and $e(x') = e_{11} + ie_{12}$ is the first eigenfunction of

$$\begin{aligned} (i\mu\nabla + L^2 A_a)^2 e &= L^2 \alpha e && \text{in } D_0, \\ \frac{\partial e}{\partial n} &= 0 && \text{on } \partial D_0. \end{aligned} \quad (3.6.75)$$

It is easy to show that there is a unique positive solution L_c such that

$$p_3 L_0^6 + p_2 L_0^4 + p_1 L_0^2 = \frac{1}{2} \kappa^2 \mu^2.$$

Hence we derive the following physical prediction:

Physical Conclusion 3.6.13 (Physical Prediction of Critical Length Scale)

Given a superconducting material, and an applied field $H_a \neq 0$, there is a critical scale $L_c > 0$ such that the phase transition in $\Omega(L)$ is a second-order transition if $L < L_c$, and is a first-order transition if $L > L_c$.

Abrikosov Vortices and Topological Structure of Supercurrents In 1957, A. Abrikosov predicted that in the mixed state of type II superconductors, there is a square or triangular array of vortices of supercurrents, called the Abrikosov vortices, and this vortex array was confirmed by experiments later. In Abrikosov theory, for type II superconductors, the first and the second critical magnetic fields H_{c1} and H_{c2} are given by

$$H_{c1} = \frac{H_c}{\sqrt{2}\kappa} (\ln \kappa + 0.081), \quad H_{c2} = \sqrt{2}\kappa H_c,$$

where H_c is the thermodynamic critical magnetic field. When $H_{c1} < H < H_{c2}$ the mixed state has two kinds of vortex structures: the square vortex lattice, as shown in Fig. 3.34a, and the triangular vortex lattice, as shown in Fig. 3.34b.

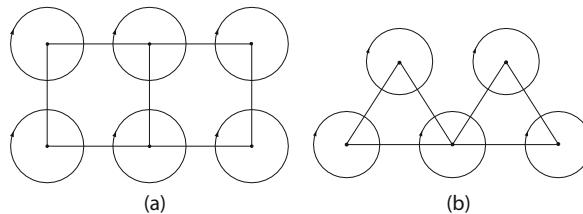
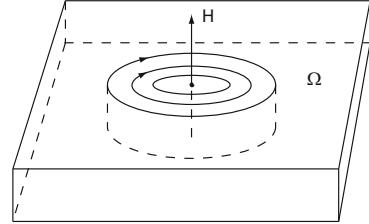


Fig. 3.34 (a) The square vortex lattice, (b) the triangular vortex lattice

In each vortex center N , the order parameter vanishes $\psi(N) = 0$, which corresponds to a normal region, the magnetic field H penetrates the sample in a neighborhood of N , and the supercurrents J_s form vortex around N in the reversed direction of H , as shown in Fig. 3.35. Each vortex is called vortex line, or a magnetic flux line.

Fig. 3.35 Structure of vortex lines



We investigate this problem from the mathematical viewpoint. The supercurrent J_s satisfies that

$$J_s = -\gamma|\psi|^2(\mathcal{A} + A_a) - \frac{\gamma\mu}{2}i(\psi^*\nabla\psi - \psi\nabla\psi^*) \quad (3.6.76)$$

$$\operatorname{div} J_s = 0, \quad J_s \cdot n|_{\partial\Omega} = 0. \quad (3.6.77)$$

Based on the geometric theory for two-dimensional incompressible flows introduced in Ma and Wang (2005d), by (3.6.77) the topological structure of J_s corresponding to the square and triangular vortex lattices in mathematics are as shown in Fig. 3.36a and b respectively. However, physically the supercurrent J_s is of the counter-magnetic property, which can be seen by the first term in (3.6.76). Thus, the paramagnetic vortices (i.e., the counterclockwise vortices) in Fig. 3.36a, b do not appear in real world. Therefore, combining with physical version, the real topological structure of J_s should be as shown in Fig. 3.37a, b, wherein the shadowed regions corresponding to the counterclockwise vortex domains in Fig. 3.36 represent the vanishing supercurrent $J_s = 0$, or the surface supercurrents.

In the previous discussion, we see that if the applied magnetic $H_a = (0, 0, H)$ is weak, then the superconductivity is in the Meissner state. In this case, $\psi \approx \text{constant}$ and the supercurrent J_s in (3.6.76) is approximatively expressed by

$$J_s = -\gamma|\psi|^2 A, \quad (A = \mathcal{A} + A_a).$$

This supercurrent is completely counter-magnetic, which causes the Meissner effect. Furthermore, in the Meissner state, super-current J_s only exists in the surface $\partial\Omega$ of sample Ω with depth λ , where $\lambda = \lambda(T) = (m_s c^2 b / 4\pi e_s^2 |a|)^{1/2}$ is the penetration depth. Thus, J_s is essentially a two-dimensional incompressible flow, and the index formula (6.1.1) in Ma and Wang (2005d) is valid, i.e.,

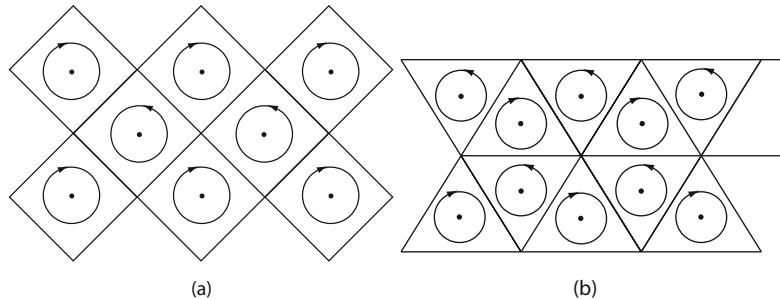


Fig. 3.36 Mathematical vortex structure of J_s

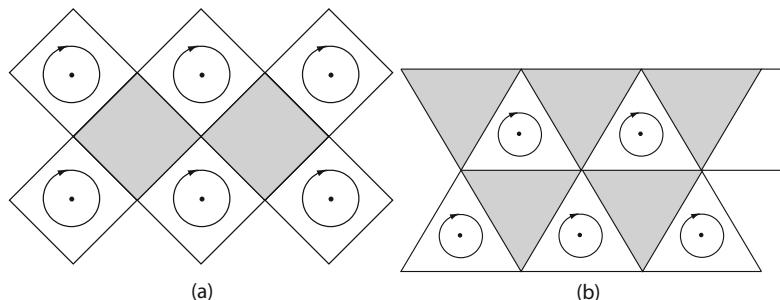


Fig. 3.37 Abrikosov vortex structure of J_s

$$C - S - \frac{B}{2} = 1 - k,$$

where C is the number of centers, S the number of interior saddles, and B the number of boundary saddles.

In addition, by the structural stability theorems, Theorems 2.1.2 and 2.2.9 in Ma and Wang (2005d), the supercurrents J_s observable in experiments are structurally stable.

3.7 Liquid Helium-4

Superfluidity is a phase of matter in which “unusual” effects are observed when liquids, typically of helium-4 or helium-3, overcome friction by surface interaction when at a stage, known as the “lambda point” for helium-4, at which the liquid’s viscosity becomes zero. Also known as a major facet in the study of quantum hydrodynamics, it was discovered by Pyotr Leonidovich Kapitsa, John F. Allen, and Don Misener in 1937 and has been described through phenomenological and microscopic theories.

Atoms helium have two stable isotopes ^4He and ^3He . ^4He consists of two electrons, two protons, and two neutrons, which are six fermions. Therefore, ^4He has an integral spin and obey the Bose–Einstein statistics. Liquid ^4He , called the Bose liquid, displays a direct transition from the normal liquid state (liquid He I) to the superfluid state (liquid He II) at temperature $T = 2.19\text{K}$, which can be considered as the condensation of particles at simple quantum state.

In the late 1930s, Ginzburg-Landau proposed a mean field theory of continuous phase transitions. With the successful application of the Ginzburg-Landau theory to superconductivity, it is nature to transfer something similar to the superfluidity case, as the superfluid transitions in liquid ^3He and ^4He are of similar quantum origin as superconductivity. Unfortunately, the classical Ginzburg-Landau free energy is poorly applicable to ^4He in a quantitative sense, as described in Ginzburg (2004).

The main objectives of this section are (1) to establish a time-dependent Ginzburg-Landau model for liquid ^4He , and (2) to study its dynamic phase transitions. The results obtained in this section predict

1. the existence of an unstable region H , where both solid and liquid He II states appear randomly depending on fluctuations, and
2. the existence of a switch point M on the λ curve, where the transitions changes from first order (Type-II with the dynamic classification scheme) to second order (Type-I).

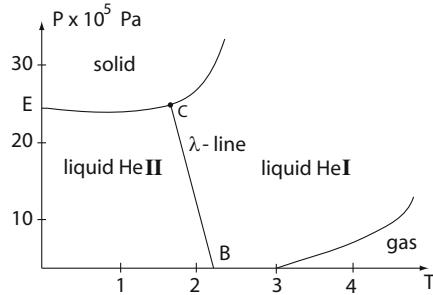
An important difference between the dynamic models introduced here and the classical ones is based on the separation of superfluid and normal fluid densities as described in (3.7.1) and their interactions in the Ginzburg-Landau free energies; see among others (Onuki, 2002). The results presented in this section were first proved in Ma and Wang (2008a).

3.7.1 Dynamic Model for Liquid Helium-4

^4He was first liquidized at $T = 4.215\text{K}$ and $p = 1 \times 10^5(\text{Pa})$ by Kamerlingh Onnes in 1908. In 1938, P.L. Kapitza found that when the temperature T decreases below $T_C = 2.17\text{K}$, the liquid ^4He will transit from normal liquid state to superfluid state, in which the fluid possesses zero viscosity, i.e., the viscous coefficient $\eta = 0$. The liquids with $\eta = 0$ are called the superfluids, and the flow without drag is called the superfluidity. The superfluid transition is called λ -phase transition, and its phase diagram is illustrated in Fig. 3.38.

Ginzburg-Landau Free Energy The order parameter describing superfluidity is characterized by a nonvanishing complex valued function $\psi : \Omega \rightarrow \mathbb{C}$ as in the superconductivity, originating from quantum Bose–Einstein condensation. In the two-fluid hydrodynamic theory of super-fluidity, the density ρ of ^4He is given by

Fig. 3.38 Classical PT phase diagram of ${}^4\text{He}$



$$\rho = \rho_s + \rho_n, \quad (3.7.1)$$

where ρ_s is the superfluid density and ρ_n the normal fluid density. The square $|\psi|^2$ is proportional to ρ_s , and without loss of generality, we take ψ as

$$|\psi|^2 = \rho_s. \quad (3.7.2)$$

Based on the quantum mechanics, $-ih\nabla\psi$ represents the momentum associated with the Bose-Einstein condensation. Hence, the free energy density contains the term

$$\frac{1}{2m}(-ih\nabla\psi)^2 = \frac{h^2}{2m}|\nabla\psi|^2,$$

where h is the Planck constant, and m the mass of atom ${}^4\text{He}$. Meanwhile the superfluid state does not obey the classical thermodynamic laws, which the normal liquid state obeys. Therefore, in the free energy density, ψ satisfies the Ginzburg-Landau expansion

$$\frac{k_1 h^2}{2m}|\nabla\psi|^2 + \frac{\gamma_1}{2}|\psi|^2 + \frac{\gamma_2}{4}|\psi|^4,$$

and ρ_n has the expansion as in the free energy (3.3.5) for PVT systems. For simplicity we ignore the entropy, and consider the coupling action of ψ and ρ_n , i.e., add the term $\frac{1}{2}\mu\rho_n|\psi|^2$ in the free energy density. Thus the Ginzburg-Landau free energy for liquid ${}^4\text{He}$ near the superfluid transition is given by

$$\begin{aligned} G(\psi, \rho_n) = & \int_{\Omega} \left[\frac{k_1 h^2}{2m} |\nabla\psi|^2 + \frac{\gamma_1}{2} |\psi|^2 + \frac{\gamma_2}{4} |\psi|^4 + \frac{\gamma_3}{2} \rho_n |\psi|^2 \right. \\ & \left. + \frac{k_2}{2} |\nabla\rho_n|^2 + \frac{\mu_1}{2} \rho_n^2 + \frac{\mu_2}{3} \rho_n^3 + \frac{\mu_3}{4} \rho_n^4 - p(\rho_n + \frac{\mu_0}{2} \rho_n^2) \right] dx. \end{aligned} \quad (3.7.3)$$

Dynamic model governing the superfluidity. By the standard model (3.1.2), we derive from (3.7.3) the following time-dependent Ginzburg-Landau equations governing the superfluidity of liquid ${}^4\text{He}$:

$$\begin{aligned}\frac{\partial\psi}{\partial t} &= \frac{k_1 h^2}{m} \Delta\psi - \gamma_1\psi - \gamma_2|\psi|^2\psi - \gamma_3\rho_n\psi, \\ \frac{\partial\rho_n}{\partial t} &= k_2\Delta\rho_n - (\mu_1 - p\mu_0)\rho_n - \mu_2\rho_n^2 - \mu_3\rho_n^3 - \frac{\gamma_3}{2}|\psi|^2 + p.\end{aligned}\quad (3.7.4)$$

It is known that the following problem has a solution $\rho_n^0 \in H^2(\Omega) \cap H_0^1(\Omega)$ standing for the density of liquid He I for any $p \in L^2(\Omega)$:

$$\begin{aligned}-k_2\Delta\rho_n^0 + (\mu_1 - p\mu_0)\rho_n^0 + \mu_2(\rho_n^0)^2 + \mu_3(\rho_n^0)^3 &= p, \\ \left.\frac{\partial\rho_n^0}{\partial n}\right|_{\partial\Omega} &= 0.\end{aligned}\quad (3.7.5)$$

To derive the nondimensional form of (3.7.4), let

$$\begin{aligned}(x, t) &= (lx', \tau t'), & (\psi, \rho_n) &= (\psi_0\psi', \rho_0\rho_n' + \rho_n^0), \\ \tau &= \frac{ml^2}{h^2k_1}, & \mu &= \frac{k_2\tau}{l^2}, \\ a_1 &= \gamma_3\rho_0\tau, & a_2 &= \gamma_2|\psi_0|^2\tau, \\ b_1 &= \frac{\gamma_3|\psi_0|^2}{2\rho_0}\tau, & b_2 &= \tau\rho_0(3\rho_n^0\mu_3 + \mu_2), \\ b_3 &= \rho_0^2\mu_3\tau, & \lambda_1 &= -\tau(\gamma_1 + \gamma_3\rho_n^0), \\ \lambda_2 &= \tau(3(\rho_n^0)^2\mu_3 + 2\rho_n^0\mu_2 + \mu_0p - \mu_1),\end{aligned}$$

where ρ_n^0 is the solution of (3.7.5).

Thus, suppressing the primes, the equations (3.7.4) are rewritten as

$$\begin{aligned}\frac{\partial\psi}{\partial t} &= \Delta\psi + \lambda_1\psi - a_1\rho_n\psi - a_2|\psi|^2\psi, \\ \frac{\partial\rho_n}{\partial t} &= \mu\Delta\rho_n + \lambda_2\rho_n - b_1|\psi|^2 - b_2\rho_n^2 - b_3\rho_n^3.\end{aligned}\quad (3.7.6)$$

The boundary conditions associated with (3.7.6) are

$$\frac{\partial\psi}{\partial n} = 0, \quad \frac{\partial\rho_n}{\partial n} = 0 \quad \text{on } \partial\Omega. \quad (3.7.7)$$

When the pressure p is independent of $x \in \Omega$, then the problems (3.7.6) and (3.7.7) can be approximatively replaced by the following systems of ordinary differential equations for superfluid transitions:

$$\begin{aligned}\frac{d\psi}{dt} &= \lambda_1\psi - a_1\rho_n\psi - a_2|\psi|^2\psi, \\ \frac{d\rho_n}{dt} &= \lambda_2\rho_n - b_1|\psi|^2 - b_2\rho_n^2 - b_3\rho_n^3.\end{aligned}\quad (3.7.8)$$

By multiplying ψ^* to both sides of the first equation of (3.7.8) and by (3.7.2), the equations (3.7.8) are reduced to:

$$\begin{aligned}\frac{d\rho_s}{dt} &= \lambda_1 \rho_s - a_1 \rho_n \rho_s - a_2 \rho_s^2, \\ \frac{d\rho_n}{dt} &= \lambda_2 \rho_n - b_1 \rho_s - b_2 \rho_n^2 - b_3 \rho_n^3, \\ (\rho_s(0), \rho_n(0)) &= (x_0, y_0).\end{aligned}\tag{3.7.9}$$

for $\rho_s \geq 0$ and $x_0 \geq 0$.

We need to explain the physical properties of the coefficients in (3.7.4) and (3.7.6). It is known that the coefficients γ_i ($1 \leq i \leq 3$) and μ_j ($0 \leq j \leq 3$) depend continuously on the temperature T and the pressure p :

$$\gamma_i = \gamma_i(T, p), \quad \mu_j = \mu_j(T, p) \quad \forall 1 \leq i \leq 3, \quad 0 \leq j \leq 3.$$

From the both mathematical and physical points of view, the following conditions are required:

$$\gamma_2 > 0, \quad \gamma_3 > 0, \quad \mu_3 > 0 \quad \forall T, p.\tag{3.7.10}$$

In addition, by the Landau mean field theory we have

$$\gamma_1 \begin{cases} > 0 & \text{if } (T, p) \in A_1, \\ < 0 & \text{if } (T, p) \in A_2, \end{cases}\tag{3.7.11}$$

$$\mu_1 \begin{cases} > 0 & \text{if } (T, p) \in B_1, \\ < 0 & \text{if } (T, p) \in B_2, \end{cases}\tag{3.7.12}$$

where A_i, B_i ($i = 1, 2$) are connected open sets such that $\overline{A}_1 + \overline{A}_2 = \overline{B}_1 + \overline{B}_2 = \mathbb{R}_+^2$, and $\overline{A}_1 \cap \overline{A}_2, \overline{B}_1 \cap \overline{B}_2$ are two simple curves in \mathbb{R}_+^2 ; see Fig. 3.39a, b. In particular, in the PT-plane,

$$\frac{\partial \gamma_1}{\partial T} > 0, \quad \frac{\partial \gamma_1}{\partial p} > 0, \quad \frac{\partial \mu_1}{\partial T} > 0, \quad \frac{\partial \mu_1}{\partial p} > 0.\tag{3.7.13}$$

By (3.7.10) the following nondimensional parameters are positive

$$a_1 > 0, \quad a_2 > 0, \quad b_1 > 0, \quad b_3 > 0.\tag{3.7.14}$$

By (3.7.11)–(3.7.13), the following two critical parameter equations

$$\lambda_1 = \lambda_1(T, p) = 0, \quad \lambda_2 = \lambda_2(T, p) = 0\tag{3.7.15}$$

define two simple curves l_1 and l_2 in the PT-plane \mathbb{R}_+^2 ; see Fig. 3.39c, d.

The parameters μ_2 and b_2 depend on the physical properties of the atom He and satisfy the following relations:

$$\begin{aligned}b_2(T, p) &= \tau \rho_0 (\mu_2 + 3\rho_n^0 \mu_3) < 0 & \text{iff } \rho_{sol} > \rho_l \text{ at } \lambda_2(T, p) = 0, \\ b_2(T, p) &= \tau \rho_0 (\mu_2 + 3\rho_n^0 \mu_3) > 0 & \text{iff } \rho_{sol} < \rho_l \text{ at } \lambda_2(T, p) = 0,\end{aligned}\tag{3.7.16}$$

where ρ_{sol} and ρ_l are the densities of solid and liquid, and ρ_n^0 the solution of (3.7.5) representing the liquid density. These relations in (3.7.16) can be deduced by the dynamic transition theorem of (2.1.1) from a simple critical eigenvalue.

3.7.2 Dynamic Phase Transition for Liquid 4He

In order to illustrate the main ideas, we discuss only the case where the pressure p is independent of $x \in \Omega$, i.e., we only consider the equations (3.7.8) and (3.7.9); the general case can be studied in the same fashion and will be reported elsewhere.

PT-Phase Diagram Based on physical experiments together with (3.7.11)–(3.7.13), the curves of $\gamma_1(T, p) = 0$ and $\mu_1(T, p) = 0$ in the PT -plane are given in Fig. 3.39a, b respectively. By the formulas

$$\begin{aligned}\lambda_1(T, p) &= -\tau(\gamma_1(T, p) + \gamma_3\rho_n^0), \\ \lambda_2(T, p) &= -\tau(\mu_1(T, p) - \mu_0p - 2\rho_n^0\mu_2 - 3\rho_n^{02}\mu_3),\end{aligned}$$

together with (3.7.10) and (3.7.16), the curves l_1 and l_2 given by (3.7.15) in the PT -plane are illustrated in Fig. 3.39c, d.

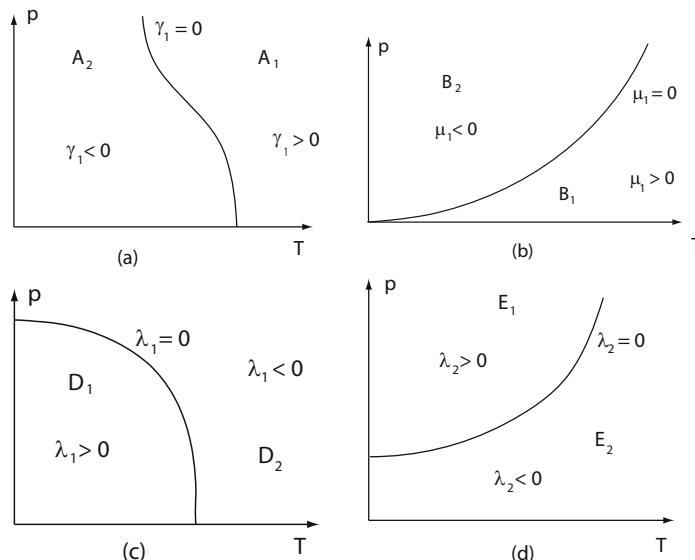
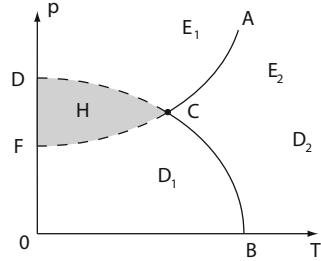


Fig. 3.39 (a) The curve of $\gamma_1 = 0$, (b) the curve of $\mu_1 = 0$, (c) the curve of $\lambda_1 = 0$, (d) the curve of $\lambda_2 = 0$

Fig. 3.40 ACF is the curve of $\lambda_2 = 0$, and BCD is the curve of $\lambda_1 = 0$



Let

$$\begin{aligned} D_1 &= \{(T, p) \in \mathbb{R}_+^2 \mid \lambda_1(T, p) > 0\}, & D_2 &= \{(T, p) \in \mathbb{R}_+^2 \mid \lambda_1(T, p) < 0\}, \\ E_1 &= \{(T, p) \in \mathbb{R}_+^2 \mid \lambda_2(T, p) > 0\}, & E_2 &= \{(T, p) \in \mathbb{R}_+^2 \mid \lambda_2(T, p) < 0\}. \end{aligned}$$

Let the curve of $\lambda_1(T, p) = 0$ intersect with the curve of $\lambda_2(T, p) = 0$ at a point C ; see Fig. 3.40.

When λ_1 crosses the curve segment CB to enter into D_1 from D_2 (see Fig. 3.40), we have $\lambda_2 < 0$ and λ_1 satisfies

$$\lambda_1(T, p) \begin{cases} < 0 & \text{if } (T, p) \in D_2, \\ = 0 & \text{if } (T, p) \in CB, \\ > 0 & \text{if } (T, p) \in D_1. \end{cases}$$

In this case, by Theorem 2.1.3, the equations (3.7.9) have a phase change which describes the transition from liquid He-I to liquid He-II, and the second equation of (3.7.9) can be equivalently rewritten as

$$\frac{d\tilde{\rho}_n}{dt} = (\lambda_2 - 2b_2\rho_n^* - 3b_3(\rho_n^*)^2)\tilde{\rho}_n - (b_2 + 3b_3\rho_n^*)\tilde{\rho}_n^2 - b_3\tilde{\rho}_n^3, \quad (3.7.17)$$

where $\tilde{\rho}_n = \rho_n - \rho_n^*$, ρ_n^* satisfies the equation

$$\lambda_2\rho_n - b_2\rho_n^2 - b_3\rho_n^3 = b_1\rho_s,$$

and $\rho_s > 0$ is the transition solution of (3.7.9). Since ρ_n^* is the density deviation of normal liquid, by (3.7.1) we have

$$\rho_n^* = \rho - \rho_n^0 - \rho_s.$$

On the other hand, ρ_n^0 ($\simeq \rho$) represents the density of liquid He I. Thus, $\rho_n^* \simeq -\rho_s < 0$. From the PT -phase diagram of ${}^4\text{He}$ (Fig. 3.38), we see that $\rho_{sol} \geq \rho_l$ for ${}^4\text{He}$ near $T = 0$, therefore by (3.7.16), $b_2 \leq 0$. Hence we derive that

$$\lambda_2 - 2b_2\rho_n^* - 3b_3(\rho_n^*)^2 < 0, \quad (3.7.18)$$

for (T, p) in the region $D_1 \setminus H$, as shown in Fig. 3.40.

It follows from (3.7.17) and (3.7.18) that when (T, p) is in the region $D_1 \setminus H$, the liquid ^4He is in the superfluid state.

When λ_2 crosses the curve segment CA , as shown in Fig. 3.40, to enter into E_1 from E_2 , then $\lambda_1 < 0$ and

$$\lambda_2(T, p) \begin{cases} < 0 & \text{if } (T, p) \in E_2, \\ = 0 & \text{if } (T, p) \in CA, \\ > 0 & \text{if } (T, p) \in E_1. \end{cases}$$

In this case, the equations (3.7.9) characterize the liquid-solid phase transition, and the first equation of (3.7.9) is equivalently expressed as

$$\frac{d\rho_s}{dt} = (\lambda_1 - a_1 \tilde{\rho}_n) \rho_s - a_2 \rho_s^2, \quad (3.7.19)$$

where $\tilde{\rho}_n > 0$ is the transition state of solid ^4He . By (3.7.14) we have

$$\lambda_1(T, p) - a_1 \tilde{\rho}_n < 0, \quad (3.7.20)$$

for any (T, p) in the region $E_1 \setminus H$, as shown in Fig. 3.40.

From (3.7.19) and (3.7.20) we can derive the conclusion that as (T, p) in $E_1 \setminus H$, $\rho_s = 0$ is stable, i.e., ^4He is in the solid state.

However, the shadowed region H in Fig. 3.40 is a metastable domain for the solid and liquid He II states, where any of these two phases may appear depending on the random fluctuations. Thus, from the discussion above, we can derive the theoretical PT -phase diagram given in Fig. 3.41, based on equations (3.7.9). In comparison with the experimental PT -phase diagram (Fig. 3.38), a slight difference in Fig. 3.41 is that there exists a metastable region H , where the solid phase and the He II phase are possible to occur. This metastable region in Fig. 3.41 corresponding to the coexistence curve CE in Fig. 3.38 is a theoretical prediction, which need to be verified by experiments.

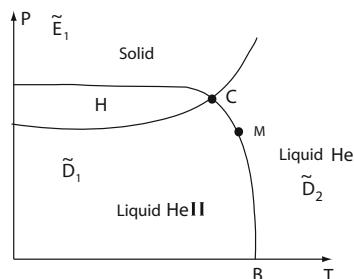


Fig. 3.41 Theoretical PT -phase diagram: H is a metastable region where both solid and liquid He II states appear randomly depending on fluctuations. The point M is a point where the transitions, between superfluid state (liquid He II) and the normal fluid state (liquid He I), changes from Type-II to Type-I; namely, the transition crossing CM is Type-II and the transition crossing MB is Type-I

States in Metastable Region We consider the dynamical properties of transitions for (3.7.9) in the unstable region. It is clear that at point $C = (T_C, p_C)$,

$$\lambda_1(T_C, p_C) = 0, \quad \lambda_2(T_C, p_C) = 0, \quad (3.7.21)$$

and the unstable region H satisfies that

$$H = \{(T, p) \in \mathbb{R}_+^2 \mid \lambda_1(T, p) > 0, \lambda_2(T, p) > 0\}.$$

To study the structure of flows of (3.7.9) for $(T, p) \in H$ it is necessary to consider the equations (3.7.9) at the point $C = (T_C, p_C)$, which, by (3.7.21), are given by

$$\begin{aligned} \frac{d\rho_s}{dt} &= -a_1\rho_n\rho_s - a_2\rho_s^2, \\ \frac{d\rho_n}{dt} &= -b_1\rho_s - b_2\rho_n^2 - b_3\rho_n^3. \end{aligned} \quad (3.7.22)$$

Due to (3.7.16) and $\rho_{sol} > \rho_l$, we have

$$b_2 < 0 \quad \text{for } (T, p) \subset H. \quad (3.7.23)$$

Under the condition (3.7.23), equations (3.7.22) have the following two steady-state solutions:

$$\begin{aligned} Z_1 = (\rho_s, \rho_n) &= (0, |b_2|/b_3), & Z_2 = (\rho_s, \rho_n) &= \left(\frac{a_1}{a_2}\alpha, -\alpha \right), \\ \alpha &= \frac{|b_2|}{2b_3}(\sqrt{1 + 4a_1b_1b_3/a_2|b_2|^2} - 1). \end{aligned}$$

By direct computation, we can prove that the eigenvalues of the Jacobian matrices of (3.7.22) at Z_1 and Z_2 are negative. Hence, Z_1 and Z_2 are stable equilibrium points of (3.7.22). Physically, Z_1 stands for solid state, and Z_2 for superfluid state. The topological structure of (3.7.22) is schematically illustrated in Fig. 3.42a, the two regions R_1 and R_2 divided by curve AO in Fig. 3.42b are the basins of attraction of Z_1 and Z_2 respectively.

We note that in H , λ_1 and λ_2 is small, i.e.,

$$0 < \lambda_1(T, p), \quad \lambda_2(T, p) \ll 1, \quad \text{for } (T, p) \in H,$$

and (3.7.9) can be consider as a perturbed system of (3.7.22).

Thus, for $(T, p) \in H$ the system (3.7.9) has four steady-state solutions $\tilde{Z}_i = \tilde{Z}(T, p)$ ($1 \leq i \leq 4$) such that

$$\lim_{(T, p) \rightarrow (T_C, p_C)} (\tilde{Z}_1(T, p), \tilde{Z}_2(T, p), \tilde{Z}_3(T, p), \tilde{Z}_4(T, p)) = (Z_1, Z_2, 0, 0),$$

and \tilde{Z}_1 and \tilde{Z}_2 are stable, representing solid state and liquid He-II state respectively, \tilde{Z}_3 and \tilde{Z}_4 are two saddle points. The topological structure of (3.7.9) for $(T, p) \in H$

is schematically shown in Fig. 3.42c, and the basins of attraction of \tilde{Z}_1 and \tilde{Z}_2 are \tilde{R}_1 and \tilde{R}_2 as illustrated in Fig. 3.42d.

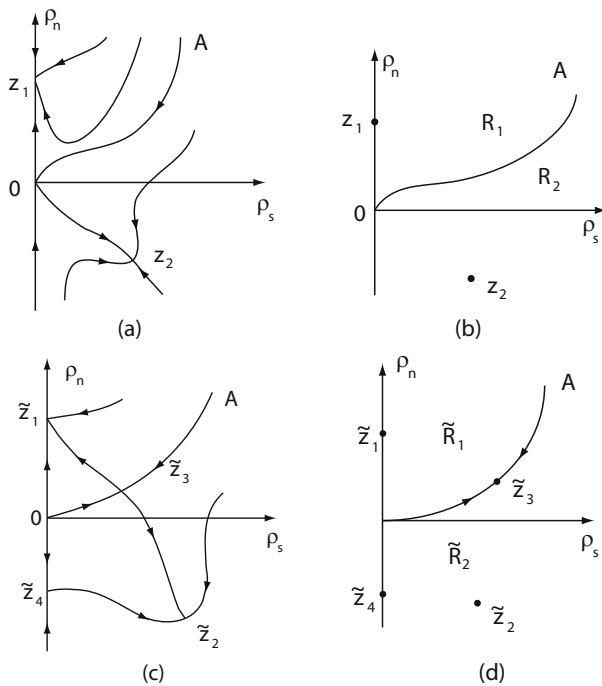


Fig. 3.42 Dynamic behavior of (3.7.9)

In summary, with the above analysis and the dynamic transition theory, we arrive at the following theorem:

Theorem 3.7.1 *There exist four regions \tilde{E}_1 , \tilde{D}_1 , \tilde{D}_2 , and H in the PT-plane (see Fig. 3.41), which are defined by*

$$\begin{aligned}\tilde{E}_1 &= \{(T, p) \in \mathbb{R}_+^2 \mid \lambda_1(T, p) < 0, \lambda_2(T, p) > 0\}, \\ \tilde{D}_1 &= \{(T, p) \in \mathbb{R}_+^2 \mid \lambda_1(T, p) > 0, \lambda_2(T, p) < 0\}, \\ \tilde{D}_2 &= \{(T, p) \in \mathbb{R}_+^2 \mid \lambda_1(T, p) < 0, \lambda_2(T, p) < 0\}, \\ H &= \{(T, p) \in \mathbb{R}_+^2 \mid \lambda_1(T, p) > 0, \lambda_2(T, p) > 0\},\end{aligned}$$

such that the following conclusions hold true:

- (1) If $(T, p) \in \tilde{E}_1$, the phase of ${}^4\text{He}$ is in solid state.
- (2) If $(T, p) \in \tilde{D}_1$, the phase is in superfluid state.
- (3) If $(T, p) \in \tilde{D}_2$, the phase is in the normal fluid state.
- (4) If $(T, p) \in H$, there are two regions \tilde{R}_1 and \tilde{R}_2 in the state space (ρ_s, ρ_n) such that, under a fluctuation which is described by the initial value (x_0, y_0) in (3.7.9),

if $(x_0, y_0) \in \tilde{R}_1$, then the phase is in solid state, and if $(x_0, y_0) \in \tilde{R}_2$, then it is in superfluid state.

In general, the observed superfluid transitions of ${}^4\text{He}$ are of the second order, i.e., Type-I (continuous) with the dynamic classification. But, from (3.7.9) we can prove the following theorem, which shows that for a higher pressure the superfluid transitions may be of zeroth order, i.e., the Type-II (jump) transition with the dynamic classification scheme.

Theorem 3.7.2 Let (T_0, p_0) satisfy that $\lambda_1(T_0, p_0) = 0, \lambda_2(T_0, p_0) < 0$. Then, (3.7.9) have a superfluid transition at (T_0, p_0) from D_2 to D_1 . In particular, the following assertions hold true:

(1) Let

$$A = \frac{a_1 b_1}{|\lambda_2|} - a_2 \quad \text{at } (T, p) = (T_0, p_0). \quad (3.7.24)$$

Then if $A < 0$ the superfluid transition is Type-I, and if $A > 0$ the superfluid transition is Type-II.

(2) The equation

$$A = \frac{a_1 b_1}{|\lambda_2|} - a_2 = 0$$

determines a point M on the CB coexistence curve in Fig. 3.41, where the transition changes type from Type-II to Type-I.

Proof. STEP 1. By the assumption, it is clear that

$$\lambda_1(T, p) = \begin{cases} < 0 & \text{if } (T, p) \in D_2, \\ = 0 & \text{if } (T, p) = (T_0, p_0), \\ > 0 & \text{if } (T, p) \in D_1, \end{cases}$$

where D_1, D_2 are as in Fig. 3.39c. Hence, by Theorem 2.1.3, (3.7.9) have a transition at (T_0, p_0) . By $\lambda_2(T_0, p_0) < 0$, the first eigenvalue of (3.7.9) is simple, and its eigenvector is given by $e = (e_s, e_n) = (1, 0)$.

The reduced equation of (3.7.9) on the center manifold reads

$$\frac{dx}{dt} = \lambda_1 x - a_1 x h(x) - a_2 x^2, \quad x > 0,$$

where $h(x)$ is the center manifold function. By (3.7.9), h can be expressed as

$$h(x) = \frac{b_1}{\lambda_2} x + o(x^2), \quad x > 0.$$

Thus the reduced equation of (3.7.9) is given by

$$\frac{dx}{dt} = \lambda_1 x + Ax^2 + o(x^2) \quad \text{for } x > 0, \quad (3.7.25)$$

where A is as in (3.7.24). The theorem follows from (3.7.25) and Theorem 2.1.6.

STEP 2. By the nondimensional form, we see that

$$a_1 b_1 = \frac{1}{2} \gamma_3^2 |\psi_0|^2 \tau^2,$$

where γ_3 is the coupled coefficient of ρ_n and ρ_s . Physically, γ_3 is small in comparison with γ_2 ; namely

$$0 < a_1 b_1 \ll a_2.$$

On the other hand, we know that

$$\lambda_2(T_0, p_0) \rightarrow 0 \quad \text{as } (T_0, p_0) \rightarrow (T_C, p_C),$$

where $C = (T_C, p_C)$ is as in (3.7.21). Therefore we deduce that there exists a pressure p^* ($b < p^* < p_C$) such that

$$A = \frac{a_1 b_1}{|\lambda_2(T_0, p_0)|} - a_2 \begin{cases} < 0 & \text{if } 0 \leq p_0 < p^*, \\ > 0 & \text{if } p^* < p_0 < p_C. \end{cases}$$

Thus, when the transition pressure p_0 is below some value $p_0 < p^*$ the superfluid transition is of the second order, i.e., is continuous with the dynamic classification scheme, and when $p_0 > p^*$ the superfluid transition is the first order, i.e., is jump in the dynamic classification scheme. The proof is complete. \square

3.8 Superfluidity of Helium-3

The superfluidity of liquid ${}^3\text{He}$ was found in 1971 by D. M. Lee, D. D. Osheroff, and R. C. Richardson, and its transition temperature is $T \approx 10^{-3}\text{K}$ under $p = 1\text{ atm}$ (10^5Pa). ${}^3\text{He}$ is a fermion and obeys the Fermi-Dirac statistics, and its atoms have to be paired to form the superfluid phase, leading to the existence of multiple superfluid phases; see Fig. 3.43. In particular, there are two superfluid phases A and B if there is no applied magnetic field.

The main objective of this section is

- (1) to establish some time-dependent Ginzburg-Landau models for ${}^3\text{He}$ with or without applied magnetic fields, and
- (2) to study superfluid dynamic transitions and their physical significance.

The model is obtained by a careful examination of the classical phase transition diagrams, together with the insights from the dynamical transition theory, and with the unified Ginzburg-Landau model presented early in this chapter.

Note that there are four possible pairs for ${}^3\text{He}$ atoms:

$$\uparrow\uparrow, \quad \downarrow\downarrow, \quad \uparrow\downarrow, \quad \downarrow\uparrow, \quad (3.8.1)$$

where \uparrow represents the $1/2$ -spin, and \downarrow stands for the $-1/2$ -spin. Consequently, all paired atoms in (3.8.1) have integral spins and become bosons. Therefore, for the case without applied magnetic field, the following are possible states for superfluid phases of ${}^3\text{He}$:

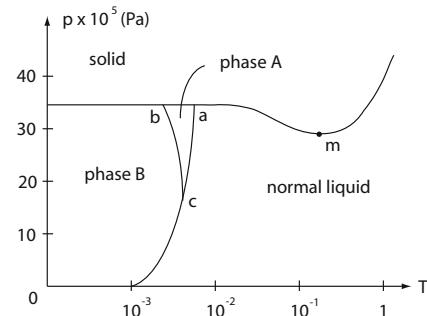
$$\begin{aligned}\text{State A: } & (a + ib)| \uparrow\uparrow \rangle + (a - ib)| \downarrow\downarrow \rangle, \\ \text{State B: } & (a + ib)| \uparrow\uparrow \rangle + (a - ib)| \downarrow\downarrow \rangle + c(| \uparrow\downarrow \rangle + | \downarrow\uparrow \rangle), \\ \text{State C: } & | \uparrow\downarrow \rangle + | \downarrow\uparrow \rangle.\end{aligned}$$

The state A is called the Anderson-Brinkman-Morel (ABM) state, corresponding to the superfluid phase A, the state B is called the Balian-Werthamer (BW) state, corresponding to superfluid phase B. Although physical experiments have only illustrated the existence of phases A and B, the dynamic theory for ${}^3\text{He}$ in this section predicts the existence of superfluid phase C, characterized by the state C above.

The dynamic theory for superfluidity of ${}^3\text{He}$ in this section includes the following three main ingredients:

FIRST, a dynamic model is established following the main ideas of the unified dynamic model for equilibrium phase transitions. The crucial point for the modeling is the introduction of three wave functions ψ_0, ψ_1, ψ_2 to characterize the superfluidity of ${}^3\text{He}$, with ψ_0 for the state $| \uparrow\uparrow \rangle$, ψ_1 for the state $| \downarrow\downarrow \rangle$, and ψ_2 for the state $| \uparrow\downarrow \rangle + | \downarrow\uparrow \rangle$. The introduction of these wave functions is based on insights from the dynamic transition theory and from the classical phase diagram given in Fig. 3.43; see among many others (Ginzburg, 2004; Reichl, 1998).

Fig. 3.43 The coexistence curve of ${}^3\text{He}$ without an applied magnetic field



SECOND, the model is then analyzed using the dynamical transition theory, leading to some interesting physical predictions, synthesized in a theoretical PT -phase diagram of ${}^3\text{He}$ given in Fig. 3.44. The analysis of these models leads to some interesting physical predictions. For the case without the applied magnetic field, the main results obtained are synthesized in a theoretical PT -phase diagram of ${}^3\text{He}$ given in Fig. 3.44.

One prediction is the existence of a metastable region $H = H_1 \cup H_2$, in which the solid state and the superfluid states A and B appear randomly depending on

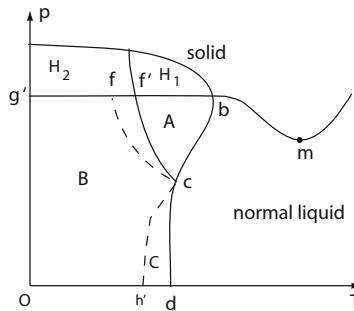


Fig. 3.44 Derived theoretical PT -phase diagram of ${}^3\text{He}$: The region $H = H_1 \cup H_2$ is the metastable domain, where the solid state and the superfluid state appear randomly depending on fluctuations. The curve \widehat{bcd} is the first critical curve where phase transition between normal fluid and superfluid states occur

fluctuations. In particular, in H_1 , phase A superfluid state and the solid state may appear, and in H_2 the phase B superfluid state and the solid state may appear.

Another prediction is the possible existence of a phase C superfluid state, which is characterized by the wave function ψ_2 , representing $|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle$. However, phase C region is very narrow, which may be the reason why it is hard to be observed experimentally.

Also, the results predict that near the two triple points b and c , there is a possibility of the existence of two switch points, where the transition on the corresponding coexistence curve switches types at each switch point. The existence of such switch points depends on the physical parameters.

THIRD, in his Nobel lecture (Ginzburg, 2004), Ginzburg was hoping to develop a Ginzburg-Landau model for superfluidity. The study presented here can be considered as an attempt toward to his vision. The study indicates that, on one hand, in order to have a Ginzburg-Landau theory for superfluidity consistent with the classical phase diagrams, multiple wave functions appear to be crucial, and, on the other hand, such a model with multiple wave functions will always lead to new predictions, as outlined above. We are confident that these predictions will be useful for designing better physical experiments and lead to better understanding of the physical mechanism of superfluidity, as evidenced by the success of the ideas and methods in other related nonlinear problems.

We would like to mention that the Ginzburg-Landau theory with only one wave function cannot describe this phase transition diagram, and this demonstrates the necessity for introducing the new Ginzburg-Landau models. The results in this section were first derived in Ma and Wang (2008e).

3.8.1 Dynamic Model for Liquid ${}^3\text{He}$ with Zero Applied Field

We introduce three complex valued functions ψ_0, ψ_1, ψ_2 to characterize the superfluidity of ${}^3\text{He}$, in which ψ_0 to the state $|\uparrow\uparrow\rangle$, ψ_1 to the state $|\downarrow\downarrow\rangle$, and ψ_2 to the

state $|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle$. Let ρ_n be the normal fluid density, ρ_a , ρ_b , and ρ_c represent the densities of superfluid phases A, B, and C respectively. Then we have

$$\begin{aligned}\rho_a &= \tau_0|\psi_0|^2 + \tau_1|\psi_1|^2 & (\tau_0 > 0, \tau_1 > 0), \\ \rho_b &= \tau_2|\psi_0|^2 + \tau_3|\psi_1|^2 + \tau_4|\psi_2|^2 & (\tau_2, \tau_3 > 0, \tau_4 > 0), \\ \rho_c &= \tau_5|\psi_2|^2 & (\tau_5 > 0).\end{aligned}$$

The total density of ${}^3\text{He}$ is given by

$$\rho = \begin{cases} \rho_n & \text{in the normal state,} \\ \rho_n + \rho_a & \text{in the phase A state,} \\ \rho_n + \rho_b & \text{in the phase B state,} \\ \rho_n + \rho_c & \text{in the phase C state.} \end{cases}$$

Physically, the states ψ_0 , ψ_1 , and ψ_2 are independent, and consequently there are no coupling terms $|\nabla(\psi_i + \psi_j)|^2$ and $|\psi_i + \psi_j|^{2k}$ ($i \neq j$) in the free energy density. Since in the case without applied magnetic field ψ_0 and ψ_1 are equal, their coefficients in the free energy should be the same. Thus, we have the following Ginzburg-Landau free energy for ${}^3\text{He}$ with $H = 0$:

$$\begin{aligned}G(\psi_0, \psi_1, \psi_2, \rho_n) &= \frac{1}{2} \int_{\Omega} \left[\frac{k_1 h^2}{m} |\nabla \psi_0|^2 + \alpha_1 |\psi_0|^2 + \alpha_2 \rho_n |\psi_0|^2 + \frac{\alpha_3}{2} |\psi_0|^4 \right. \\ &\quad + \frac{k_1 h^2}{m} |\nabla \psi_1|^2 + \alpha_1 |\psi_1|^2 + \alpha_2 \rho_n |\psi_1|^2 + \frac{\alpha_3}{2} |\psi_1|^4 + \alpha_3 |\psi_0|^2 |\psi_1|^2 \\ &\quad + \frac{k_2 h^2}{m} |\nabla \psi_2|^2 + \beta_1 |\psi_2|^2 + \frac{\beta_2}{4} |\psi_2|^4 + \beta_3 |\psi_0|^2 |\psi_2|^2 \\ &\quad + \beta_3 |\psi_1|^2 |\psi_2|^2 + \beta_4 \rho_n |\psi_2|^2 + k_3 |\nabla \rho_n|^2 + \mu_1 |\rho_n|^2 \\ &\quad \left. + \frac{2\mu_2}{3} \rho_n^3 + \frac{\mu_3}{2} \rho_n^3 - p \left(\rho_n + \frac{\mu_0}{2} \rho_n^2 \right) \right] dx, \end{aligned} \quad (3.8.2)$$

where the coefficients depend on T and p , and for $1 \leq i \leq 3$, $j = 2, 3, 4$,

$$k_i > 0, \quad \beta_j > 0, \quad \alpha_2 > 0, \quad \alpha_3 > 0, \quad \mu_3 > 0, \quad \mu_2 < 0. \quad (3.8.3)$$

For α_1, β_1 and μ_1 , there are regions A_i, B_i, C_i ($i = 1, 2$) in the PT -plane \mathbb{R}_+^2 such that $\overline{A}_1 + \overline{A}_2 = \overline{B}_1 + \overline{B}_2 = \overline{C}_1 + \overline{C}_2 = \mathbb{R}_+^2$, and

$$\alpha_1 = \alpha_1(T, p) \begin{cases} > 0 & \text{if } (T, p) \in A_1, \\ < 0 & \text{if } (T, p) \in A_2, \end{cases} \quad (3.8.4)$$

$$\beta_1 = \beta_1(T, p) \begin{cases} > 0 & \text{if } (T, p) \in B_1, \\ < 0 & \text{if } (T, p) \in B_2, \end{cases} \quad (3.8.5)$$

$$\mu_1 = \mu_1(T, p) \begin{cases} > 0 & \text{if } (T, p) \in C_1, \\ < 0 & \text{if } (T, p) \in C_2. \end{cases} \quad (3.8.6)$$

It is known that for ${}^3\text{He}$, $\mu_1 = \mu_1(T, p)$ is not monotone on T . In fact, at $T_m = 0.318K, p_m = 29.31 \times 10^5 \text{ Pa}$, we have

$$\mu_1(T_m, p_m) = 0, \quad \frac{\partial \mu_1(T_m, p_m)}{\partial T} = 0, \quad (3.8.7)$$

where $m = (T_m, p_m)$ is as shown in Fig. 3.43. Near the point m the famous Pomeranchuk effect takes place, i.e., when pressure increases, the liquid ${}^3\text{He}$ will absorb heat to undergo a transition to solid state.

By the standard model (3.1.2), we infer from (3.8.2) the following time-dependent GL equations for the superfluidity of liquid ${}^3\text{He}$:

$$\begin{aligned} \frac{\partial \psi_0}{\partial t} &= \frac{k_1 h^2}{m} \Delta \psi_0 - \alpha_1 \psi_0 - \alpha_2 \rho_n \psi_0 \\ &\quad - \beta_3 |\psi_2|^2 \psi_0 - \alpha_3 |\psi_0|^2 \psi_0 - \alpha_3 |\psi_1|^2 \psi_0, \\ \frac{\partial \psi_1}{\partial t} &= \frac{k_1 h^2}{m} \Delta \psi_1 - \alpha_1 \psi_1 - \alpha_2 \rho_n \psi_1 \\ &\quad - \beta_3 |\psi_2|^2 \psi_1 - \alpha_3 |\psi_1|^2 \psi_1 - \alpha_3 |\psi_0|^2 \psi_1, \\ \frac{\partial \psi_2}{\partial t} &= \frac{k_2 h^2}{m} \Delta \psi_2 - \beta_1 \psi_2 - \beta_3 |\psi_0|^2 \psi_2 \\ &\quad - \beta_3 |\psi_1|^2 \psi_2 - \beta_4 \rho_n \psi_2 - \beta_2 |\psi_2|^2 \psi_2, \\ \frac{\partial \rho_n}{\partial t} &= k_3 \Delta \rho_n - (\mu_1 - \mu_0 p) \rho_n - \mu_2 \rho_n^2 \\ &\quad - \mu_3 \rho_n^3 - \frac{\alpha_2}{2} |\psi_0|^2 - \frac{\alpha_2}{2} |\psi_1|^2 - \frac{\beta_4}{2} |\psi_2|^2 - p. \end{aligned} \quad (3.8.8)$$

The nondimensional form of (3.8.8) can be written as

$$\begin{aligned} \frac{\partial \psi_0}{\partial t} &= \Delta \psi_0 + \lambda_1 \psi_0 - a_1 \rho_n \psi_0 - a_2 |\psi_2|^2 \psi_0 - a_3 |\psi_0|^2 \psi_0 - a_3 |\psi_1|^2 \psi_0, \\ \frac{\partial \psi_1}{\partial t} &= \Delta \psi_1 + \lambda_1 \psi_1 - a_1 \rho_n \psi_1 - a_2 |\psi_2|^2 \psi_1 - a_3 |\psi_1|^2 \psi_1 - a_3 |\psi_0|^2 \psi_1, \\ \frac{\partial \psi_2}{\partial t} &= \kappa_1 \Delta \psi_2 + \lambda_2 \psi_2 - b_1 \rho_n \psi_2 - b_2 |\psi_0|^2 \psi_2 - b_2 |\psi_1|^2 \psi_2 - b_3 |\psi_2|^2 \psi_2, \\ \frac{\partial \rho_n}{\partial t} &= \kappa_2 \Delta \rho_n + \lambda_3 \rho_n - c_1 |\psi_0|^2 - c_1 |\psi_1|^2 - c_2 |\psi_2|^2 + c_3 \rho_n^2 - c_4 \rho_n^3, \end{aligned} \quad (3.8.9)$$

where

$$\begin{aligned} \lambda_1 &= -\frac{ml^2}{h^2 k_1} (\alpha_1 + \alpha_2 \rho_n^0), \\ \lambda_2 &= -\frac{ml^2}{h^2 k_1} (\beta_1 + b_1 \rho_n^0), \\ \lambda_3 &= -\frac{ml^2}{h^2 k_1} (\mu_1 - \mu_0 p - 2\rho_n^0 \mu_2 - 3(\rho_n^0)^2 \mu_3), \end{aligned} \quad (3.8.10)$$

and ρ_n^0 is determined by the steady-state solution of the form $(\psi_0, \psi_1, \psi_2, \rho_n) = (0, 0, 0, \rho_n^0)$ of (3.8.8). By (3.8.3) the coefficients in (3.8.9) satisfy

$$a_i > 0, \quad b_i > 0, \quad c_j > 0 \quad \forall 1 \leq i \leq 3, \quad 1 \leq j \leq 4.$$

The boundary conditions are the Neumann conditions

$$\frac{\partial}{\partial n}(\psi_0, \psi_1, \psi_2, \rho_n) = 0 \quad \text{on } \partial\Omega. \quad (3.8.11)$$

When p is a constant, the problem (3.8.9) with (3.8.11) can be reduced to the following system of ordinary differential equations:

$$\begin{aligned} \frac{d\rho_1}{dt} &= \lambda_1\rho_1 - a_1\rho_n\rho_1 - a_2\rho_2\rho_1 - a_3\rho_1^2, \\ \frac{d\rho_2}{dt} &= \lambda_2\rho_2 - b_1\rho_n\rho_2 - b_2\rho_1\rho_2 - b_3\rho_2^2, \\ \frac{d\rho_n}{dt} &= \lambda_3\rho_n - c_1\rho_1 - c_2\rho_2 + c_3\rho_n^2 - c_4\rho_n^3, \end{aligned} \quad (3.8.12)$$

where $\rho_1 = |\psi_0|^2 + |\psi_1|^2$ and $\rho_2 = |\psi_2|^2$.

3.8.2 Critical Parameter Curves and PT-Phase Diagram

Critical Parameter Curves Critical parameter curves in the PT -plane are given by

$$l_i = \{(T, p) \in \mathbb{R}_+^2 \mid \lambda_i(T, p) = 0\}, \quad i = 1, 2, 3,$$

where $\lambda_i = \lambda_i(T, p)$ are defined by (3.8.10).

It is clear that the critical parameter curves l_i are associated with the PT -phase diagram of ${}^3\text{He}$. As in the last section if we can determine the critical parameter curves l_i ($1 \leq i \leq 3$), then we obtain the PT -phase diagram.

Phenomenologically, according to the experimental PT -phase diagram (Fig. 3.43), the parameter curves l_i ($i = 1, 2, 3$) in the PT -plane should be as schematically illustrated in Fig. 3.45a–c. The combination of the diagrams (a)–(c) in Fig. 3.45 gives Fig. 3.46, in which the real line \widehat{bm} stands for the coexistence curve of the solid and liquid phases, and \widehat{bcd} for the coexistence curve of superfluid and normal liquid phases.

Now we rigorously examine phase transitions in different regimes determined by the equations and the critical parameter curves.

States in the Metastable Region We consider the dynamical properties of transitions for (3.8.12) in the metastable region. It is clear that at point $b = (T_b, p_b)$,

$$\lambda_1(T_b, p_b) = \lambda_3(T_b, p_b) = 0, \quad \lambda_2(T_b, p_b) < 0, \quad (3.8.13)$$

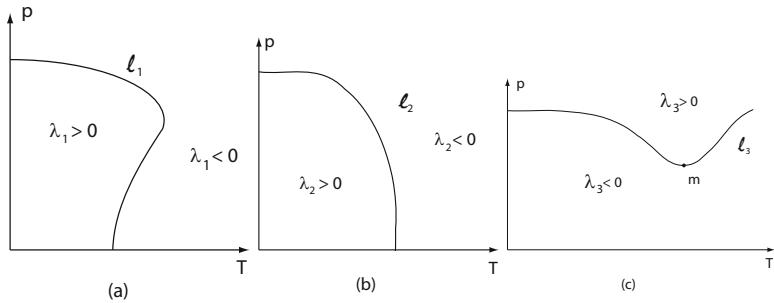
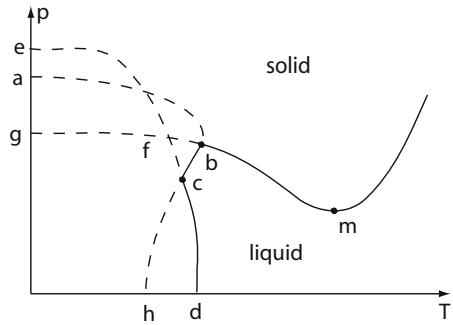


Fig. 3.45 (a) The curve l_1 ($\lambda_1 = 0$), (b) the curve l_2 ($\lambda_2 = 0$), and (c) the curve l_3 ($\lambda_3 = 0$)

Fig. 3.46 The curve \widehat{abch} is l_1 , \widehat{efcd} is l_2 , and \widehat{gfbm} is l_3



and the metastable region H_1 near the triple point b is defined by

$$H_1 = \{(T, p) \in \mathbb{R}_+^2 \mid (\lambda_1, \lambda_2, \lambda_3)(T, p) = (+, -, +)\}.$$

To study the structure of flows of (3.8.12) for $(T, p) \in H$ it is necessary to consider the equations (3.8.12) at the point $b = (T_b, p_b)$, which, by (3.8.13), are given by

$$\begin{aligned} \frac{d\rho_1}{dt} &= -a_1\rho_n\rho_1 - a_3\rho_1^2, \\ \frac{d\rho_n}{dt} &= -c_1\rho_1 + c_3\rho_n^2 - c_4\rho_n^3, \end{aligned} \quad (3.8.14)$$

We know that

$$c_3 > 0 \quad \text{for} \quad (T, p) \subset H_1. \quad (3.8.15)$$

Equations (3.8.14) have the following two steady-state solutions:

$$\begin{aligned} Z_1 &= (\rho_1, \rho_n) = (0, c_3/a_4), \\ Z_2 &= (\rho_1, \rho_n) = \left(\frac{a_1}{a_3}\alpha, -\alpha \right), \\ \alpha &= \frac{c_3}{2c_4} \left(\sqrt{1 + \frac{4c_1a_1c_4}{a_3c_3^2}} - 1 \right). \end{aligned}$$

By direct computation, we can prove that the eigenvalues of the Jacobian matrices of (3.8.14) at Z_1 and Z_2 are negative. Hence, Z_1 and Z_2 are stable equilibrium points of (3.8.14). Physically, Z_1 stands for solid state, and Z_2 for superfluid state. The topological structure of (3.8.14) is schematically illustrated in Fig. 3.47a, and the two regions R_1 and R_2 divided by curve AO in Fig. 3.47b are the basins of attraction of Z_1 and Z_2 respectively.

We note that in H_1 , λ_1 and λ_3 are small, i.e.,

$$0 < \lambda_1(T, p), \quad \lambda_2(T, p) \ll 1 \quad \text{for } (T, p) \in H,$$

and (3.8.14) can be considered as a perturbed system of (3.8.12).

Thus, for $(T, p) \in H_1$ the system (3.8.12) have four steady-state solutions $\tilde{Z}_i = \tilde{Z}(T, p)$ ($1 \leq i \leq 4$) such that

$$\lim_{(T, p) \rightarrow (T_C, p_C)} (\tilde{Z}_1, \tilde{Z}_2, \tilde{Z}_3, \tilde{Z}_4)(T, p) = (Z_1, Z_2, 0, 0),$$

and \tilde{Z}_1 and \tilde{Z}_2 are stable, representing solid state and liquid He-3 state respectively, \tilde{Z}_3 and \tilde{Z}_4 are two saddle points. The topological structure of (3.8.14) for $(T, p) \in H_1$ is schematically shown in Fig. 3.47c, and the basins of attraction of \tilde{Z}_1 and \tilde{Z}_2 are \tilde{R}_1 and \tilde{R}_2 as illustrated in Fig. 3.47d.

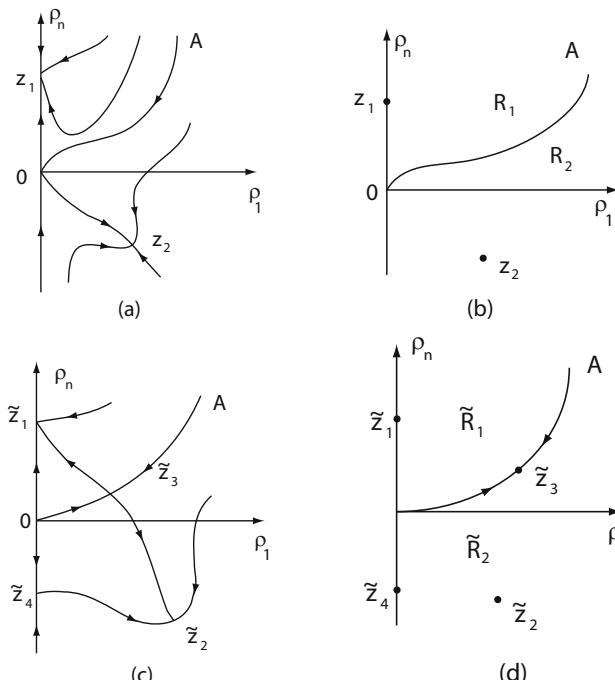


Fig. 3.47 Dynamical properties of transitions for (3.8.12)

First Phase Transition On the coexistence curve \widehat{bc} in Fig. 3.44,

$$\lambda_1 = 0, \lambda_2 < 0, \lambda_3 < 0.$$

Hence the first phase transition crossing \widehat{bc} is between normal fluid state and phase A superfluid state. On the coexistence curve \widehat{cd} ,

$$\lambda_1 < 0, \quad \lambda_2 = 0, \quad \lambda_3 < 0.$$

In this case, the first phase transition crossing this coexistence curve is between the normal fluid state and the phase C superfluid state.

Second Phase Transitions When (T, p) crosses the curve segment \widehat{bcd} , (3.8.12) will undergo a second transition. We need to consider two cases.

CASE 1. SECOND TRANSITION CROSSING $\widehat{f'c}$: If (T, p) passes through this curve segment \widehat{bc} , then the first transition solution is given by

$$(\rho_1, \rho_2, \rho_n) = (\rho_1^*, 0, \rho_n^*), \quad \rho_1^* = \rho_1^*(T, p) > 0, \quad \rho_n^* = \rho_n^*(T, p) < 0.$$

Take a transformation

$$\rho'_1 = \rho_1 - \rho_1^*, \quad \rho'_2 = \rho_2, \quad \rho'_n = \rho_n - \rho_n^*.$$

Then, the system (3.8.12) is in the following form (for simplicity, we drop the primes):

$$\begin{aligned} \frac{d\rho_1}{dt} &= \tilde{\lambda}_1 \rho_1 - a_1 \rho_1^* \rho_n - a_2 \rho_1^* \rho_2 - a_2 \rho_1 \rho_2 - a_3 \rho_1^3, \\ \frac{d\rho_2}{dt} &= \tilde{\lambda}_2 \rho_2 - b_1 \rho_n \rho_2 - b_2 \rho_1 \rho_2 - b_3 \rho_2^2, \\ \frac{d\rho_n}{dt} &= \tilde{\lambda}_3 \rho_n - c_1 \rho_1 - c_2 \rho_2 + (c_3 - 3c_4 \rho_n^*) \rho_n^2 - c_4 \rho_n^3, \end{aligned} \quad (3.8.16)$$

where

$$\begin{aligned} \tilde{\lambda}_1 &= \lambda_1 + a_1 |\rho_n^*| - 2a_3 \rho_1^*, \\ \tilde{\lambda}_2 &= \lambda_2 + b_1 |\rho_n^*| - b_2 \rho_1^*, \\ \tilde{\lambda}_3 &= \lambda_3 - 2c_3 |\rho_n^*| - 3c_4 \rho_n^{*2}. \end{aligned} \quad (3.8.17)$$

The linear operator in (3.8.16) reads

$$L = \begin{pmatrix} \tilde{\lambda}_1 & -a_2 \rho_1^* & -a_1 \rho_1^* \\ 0 & \tilde{\lambda}_2 & 0 \\ -c_1 & -c_2 & \tilde{\lambda}_3 \end{pmatrix}.$$

The three eigenvalues of L are

$$\beta_1 = \tilde{\lambda}_2, \quad \beta_{\pm} = \frac{1}{2} \left[\tilde{\lambda}_1 + \tilde{\lambda}_3 \pm \sqrt{(\tilde{\lambda}_3 - \tilde{\lambda}_1)^2 + 4a_1 c_1 \rho_1^*} \right]. \quad (3.8.18)$$

It is known that the transition solution $(\rho_1^*, 0, \rho_n^*)$ is stable near \widehat{bc} . Therefore the eigenvalues of L satisfy

$$\beta_1(T, p) < 0, \quad \beta_{\pm}(T, p) < 0 \text{ for } (T, p) \text{ near } \widehat{bc}.$$

However, near \widehat{fc} there is a curve segment $\widehat{f'c}$ such that

$$\widehat{f'c} = \{(T, p) \in \mathbb{R}_+^2 \mid \beta_1(T, p) = 0\}.$$

Thus, system (3.8.16) has a transition on $\widehat{f'c}$, which is called the second transition of (3.8.12), and $\widehat{f'c}$ is the coexistence curve of phases A and B; see Fig. 3.44.

CASE 2. SECOND TRANSITION CROSSING $\widehat{ch'}$: If (T, p) passes through this curve segment \widehat{bc} , then the first transition solution is given by

$$(\rho_1, \rho_2, \rho_n) = (0, \eta_2, \eta_n), \quad \eta_2 = \eta_2(T, p) > 0, \quad \eta_n = \eta_n(T, p) < 0.$$

Take a transformation

$$\rho'_1 = \rho, \quad \rho'_2 = \rho_2 - \eta_2, \quad \rho'_n = \rho_n - \eta_n.$$

Then, the system (3.8.12) is in the following form (for simplicity, we drop the primes):

$$\begin{aligned} \frac{d\rho_1}{dt} &= \tilde{\tilde{\lambda}}_1 \rho_1 - a_1 \rho_1 \rho_n - a_2 \rho_1 \rho_2 - a_3 \rho_1^2, \\ \frac{d\rho_2}{dt} &= \tilde{\tilde{\lambda}}_2 \rho_2 - b_1 \eta_2 \rho_n - b_2 \eta_2 \rho_1 - b_1 \rho_2 \rho_n - b_2 \rho_1 \rho_2 - b_3 \rho_2^2, \\ \frac{d\rho_n}{dt} &= \tilde{\tilde{\lambda}}_3 \rho_n - c_1 \rho_1 - c_2 \rho_2 + (c_3 - 3c_4 \eta_n) \rho_n^2 - c_4 \rho_n^3, \end{aligned} \quad (3.8.19)$$

where

$$\begin{aligned} \tilde{\tilde{\lambda}}_1 &= \lambda_1 + a_1 |\eta_n| - a_2 \eta_2, \\ \tilde{\tilde{\lambda}}_2 &= \lambda_2 + b_1 |\eta_n| - 2b_2 \eta_2, \\ \tilde{\tilde{\lambda}}_3 &= \lambda_3 - 2c_3 |\eta_n| - 3c_4 \eta_n^2. \end{aligned} \quad (3.8.20)$$

The linear operator L in (3.8.19) has three eigenvalues:

$$\beta_1 = \tilde{\tilde{\lambda}}_1, \quad \beta_{\pm} = \frac{1}{2} \left[\tilde{\tilde{\lambda}}_1 + \tilde{\tilde{\lambda}}_3 \pm \sqrt{(\tilde{\tilde{\lambda}}_3 - \tilde{\tilde{\lambda}}_1)^2 + 4b_1 c_2 \eta_2} \right]. \quad (3.8.21)$$

It is known that the transition solution $(\rho_1^*, 0, \rho_n^*)$ is stable near \widehat{bc} . Therefore the eigenvalues of L satisfy

$$\beta_1(T, p) < 0, \quad \beta_{\pm}(T, p) < 0 \text{ for } (T, p) \text{ near } \widehat{cd}.$$

However, near \widehat{hc} there is a curve segment $\widehat{h'c}$ such that

$$\widehat{h'c} = \{(T, p) \in \mathbb{R}_+^2 \mid \beta_1(T, p) = \tilde{\lambda}_1 = 0\}.$$

Thus, system (3.8.19) has a transition on $\widehat{h'c}$, which is called the second transition of (3.8.12), and $\widehat{h'c}$ is the coexistence curve of phases C and B ; see Fig. 3.44.

In summary, with the above analysis and the dynamic transition theory, we arrive at the following transition theorem:

Theorem 3.8.1 Define a few regions in the PT-plane (see Fig. 3.44) by

$$\begin{aligned} E_1 &= \{(T, p) \in \mathbb{R}_+^2 \mid (\lambda_1, \lambda_2, \lambda_3)(T, p) = (-, -, +)\}, \\ E_2 &= \{(T, p) \in \mathbb{R}_+^2 \mid (\lambda_1, \lambda_2, \lambda_3)(T, p) = (-, -, -)\}, \\ H_1 &= \{(T, p) \in \mathbb{R}_+^2 \mid (\lambda_1, \tilde{\lambda}_2, \lambda_3)(T, p) = (+, -, +)\}, \\ H_2 &= \{(T, p) \in \mathbb{R}_+^2 \mid (\lambda_1, \tilde{\lambda}_2, \lambda_3)(T, p) = (+, +, +)\}, \\ \text{Region } f'bc &= \{(T, p) \in \mathbb{R}_+^2 \mid (\lambda_1, \lambda_2, \tilde{\lambda}_2, \lambda_3)(T, p) = (+, -, -, -)\}, \\ \text{Region } cdh' &= \{(T, p) \in \mathbb{R}_+^2 \mid (\lambda_1, \tilde{\lambda}_2, \lambda_2, \lambda_3)(T, p) = (-, -, +, -)\}, \end{aligned}$$

and let the Region $0g'f'ch'$ be the complement of the sum of the above regions. Then the following conclusions hold true:

- (1) If $(T, p) \in E_1$, the phase of ${}^3\text{He}$ is in solid state.
- (2) If $(T, p) \in E_2$, the phase is in normal liquid state.
- (3) If $(T, p) \in \text{Region } f'bc$, the phase is in phase A superfluid state.
- (4) If $(T, p) \in \text{Region } cdh'$, the phase is in phase C superfluid state.
- (5) If $(T, p) \in \text{Region } 0g'f'ch'$, the phase is in phase B superfluid state.
- (6) If $(T, p) \in H_1$, there are two regions R_1 and R_2 in the state space (ρ_1, ρ_2, ρ_n) such that, under a fluctuation which is described by the initial value (x_0, y_0, z_0) in (3.8.12): If $(x_0, y_0, z_0) \in R_1$, then the phase is in solid state, and if $(x_0, y_0, z_0) \in R_2$, then it is in phase A superfluid state.
- (7) If $(T, p) \in H_2$, there are two regions K_1 and K_2 in the state space (ρ_1, ρ_2, ρ_n) such that if $(x_0, y_0, z_0) \in K_1$, then the phase is in solid state, and if $(x_0, y_0, z_0) \in K_2$, then it is in phase B superfluid state.

3.8.3 Classification of Superfluid Transitions

We now classify the superfluid transitions of (3.8.12) crossing various coexistence curves. First we consider the transitions crossing curve segments \widehat{cd} and \widehat{bc} in Fig. 3.44. Obviously, we have

$$\begin{aligned} \widehat{cd} &= \{(T, p) \in \mathbb{R}_+^2 \mid (\lambda_1, \lambda_2, \lambda_3)(T, p) = (-, 0, -)\}, \\ \widehat{bc} &= \{(T, p) \in \mathbb{R}_+^2 \mid (\lambda_1, \lambda_2, \lambda_3)(T, p) = (0, -, -)\}. \end{aligned}$$

Let

$$\begin{aligned} A_1 &= a_1 c_1 - a_3 |\lambda_3|, & A_2 &= a_1 c_2 - a_2 |\lambda_3|, \\ B_1 &= b_1 c_2 - b_3 |\lambda_3|, & B_2 &= b_1 c_1 - b_2 |\lambda_3|. \end{aligned}$$

In fact, in (3.8.8) and (3.8.9), we see that

$$a_2 = b_2, \quad a_1 = c_1, \quad b_1 = c_2.$$

Therefore $A_2 = B_2$.

Theorem 3.8.2 *For the system (3.8.12) we have the following assertions:*

- (1) *As $(T_0, p_0) \in \widehat{cd}$, the transition of (3.8.12) at (T_0, p_0) is between the phase B and normal liquid. Furthermore if $B_1 \leq 0$, then it is Type-I, and if $B_1 > 0$, then it is Type-II.*
- (2) *As $(T_0, p_0) \in \widehat{bc}$, the transition is between the phase A and normal liquid. Moreover, if $A_1 \leq 0$, then it is Type-I, and if $A_1 > 0$, then it is Type-II.*

Theorem 3.8.2 provides conditions for the first transition of (3.8.12). The following theorem gives sufficient conditions for the second transition near \widehat{fc} in Fig. 3.44.

Obviously, the curve \widehat{fc} is given by

$$\widehat{fc} = \{(T, p) \in \mathbb{R}_+^2 \mid (\lambda_1, \lambda_2, \lambda_3)(T, p) = (+, 0, -)\}.$$

To set up the second transition theorem, we need to assume the following conditions. Let $\varepsilon > 0$ be small. Suppose that

$$B_2 > -\varepsilon, \quad b_1 a_3 - b_2 a_1 > -\varepsilon \quad \text{for } (T, p) \in \widehat{fc}, \quad (3.8.22)$$

and the gap between \widehat{bc} and \widehat{fc} is small, i.e.,

$$|T_2 - T_3| = O(\varepsilon) \quad \forall (T_2, p) \in \widehat{bc}, \quad (T_3, p) \in \widehat{fc}. \quad (3.8.23)$$

We also assume that

$$(a_1, c_1) = O(\varepsilon), \quad (b_1, c_2, c_3, a_3, b_3) = O(1), \quad (3.8.24)$$

$$a_3 B_1 - b_2 A_2 > 0 \quad \text{in } \widehat{fc} \text{ with } A_1 \leq 0. \quad (3.8.25)$$

Theorem 3.8.3 *Under conditions (3.8.22) and (3.8.23), there exists a curve segment \widehat{fc} near \widehat{fc} as shown in Fig. 3.44 such that (3.8.12) has the second transition from the first transition solution $(\rho_1^*, 0, \rho_n^*)$, i.e., (3.8.16) has a transition from $(\rho_1, \rho_2, \rho_n) = 0$ in \widehat{fc} , and the transition solutions $(\tilde{\rho}_1, \tilde{\rho}_2, \tilde{\rho}_n)$ satisfy that $\tilde{\rho}_2 > 0$. In addition, if (3.8.24) and (3.8.25) hold true, then this transition is Type-II.*

Physical experiments display that the superfluid transition of liquid ${}^3\text{He}$ between the normal liquid and superfluid phase B is continuous. Hence, it is necessary to give the conditions of Type-I transition of (3.8.12) at the intersecting point C of two curves $\lambda_1 = 0$ and $\lambda_2 = 0$.

Theorem 3.8.4 Let (T_0, p_0) be the point C that $\lambda_1(T_0, p_0) = 0$ and $\lambda_2(T_0, p_0) = 0$. Then the transition of (3.8.12) at (T_0, p_0) is Type-I if and only if one of the following two conditions hold true:

- (i) $A_1 \leq 0, B_1 \leq 0, A_2 = B_2 < 0,$
- (ii) $A_1 \leq 0, B_1 \leq 0, A_2 = B_2 \geq 0$ and $A_1 B_1 > A_2 B_2.$

In particular, if the transition is Type-I, then for $\lambda_1 > 0, \lambda_2 > 0$ near (p_0, T_0) , there are four types of topological structure of the transition on center manifold, which are classified as follows:

- (1) This transition is of the structure as shown in Fig. 3.48a, if

$$\lambda_1|B_1| + \lambda_2 A_2 > 0 \quad \text{and} \quad \lambda_2|A_1| + \lambda_1 B_2 > 0.$$

- (2) The transition is of the structure as shown in Fig. 3.48b, if

$$\lambda_1|B_1| + \lambda_2 A_2 < 0 \quad \text{and} \quad \lambda_2|A_1| + \lambda_1 B_2 < 0.$$

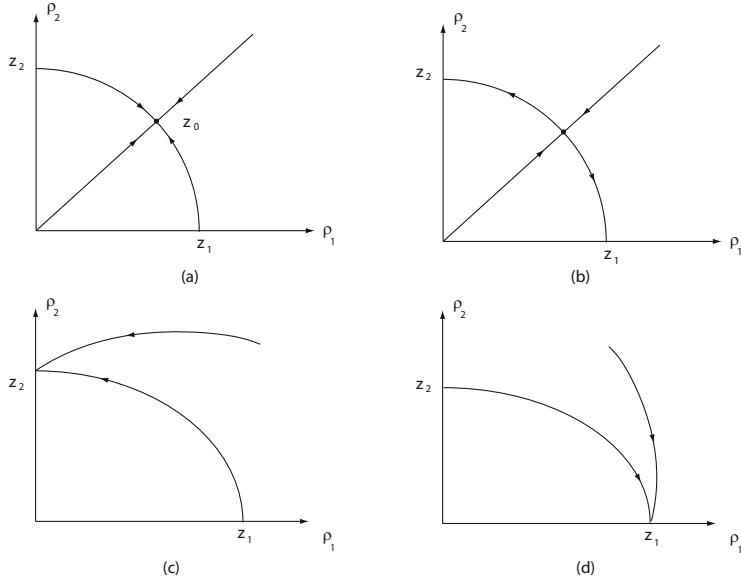


Fig. 3.48 Topological structure of Type-I transition near the intersection point $C = (T_0, p_0)$ of $\lambda_1 = 0$ and $\lambda_2 = 0$. Here ρ_1 and ρ_2 stands for the densities of phases A and C respectively

- (3) The transition is of the structure as shown in Fig. 3.48c, if

$$\lambda_1|B_1| + \lambda_2 A_2 < 0 \quad \text{and} \quad \lambda_2|A_1| + \lambda_1 B_2 > 0.$$

- (4) The transition has the structure as shown in Fig. 3.48d, if

$$\lambda_1|B_1| + \lambda_2 A_2 > 0 \quad \text{and} \quad \lambda_2|A_1| + \lambda_1 B_2 < 0.$$

Remark 3.8.5 Physically, the transition between normal liquid and superfluid for ${}^3\text{He}$ is generally Type-I, the transition between superfluid phases A and B is Type-II, and the region of phase A is narrow. Therefore, under the conditions (3.8.22)–(3.8.25) and

$$\begin{aligned} B_1 &< 0 & \text{in } \widehat{cd}, \\ B_1 \leq 0, A_1 &< 0 & \text{near point } C = \overline{\Gamma}_1 \cap \overline{\Gamma}_2 = (T_0, p_0), \end{aligned}$$

the above theorems (Theorems 3.8.2–3.8.4) provide a precise mathematical proof for superfluid transitions of liquid ${}^3\text{He}$ with no applied magnetic field.

By condition (3.8.24) we see that

$$a_1 \cdot c_1 = O(\varepsilon^2), \quad a_3 = O(1).$$

Assertion (1) of Theorem 3.8.2 implies that only in a very small range of (T, p) near the change point of superfluid and solid, the transition between normal liquid and superfluid is II-type, and this range is

$$0 > \lambda_3(T, p) > -\frac{a_1 c_1}{a_3} = -O(\varepsilon^2).$$

Moreover, the superfluid density of phase A near the solid phase is in the quantitative order ε^3 , i.e.,

$$\rho_1 = \frac{a_1^2 c_1}{a_3^2 c_3} = O(\varepsilon^3).$$

Hence, the difference between the Type-I and the Type-II phase transitions in experiments is very small.

Note that Theorem 3.8.3 is also valid if condition (3.8.22) is replaced by that

$$\frac{\partial \lambda_2}{\partial T} \gg 1 \quad \text{for } (T, p) \text{ near } \widehat{fc}.$$

Proof of Theorem 3.8.2. As $(T_0, p_0) \in \widehat{bc}$, $\lambda_2(T_0, p_0) < 0$ and the space of $(\rho_1, 0, \rho_n)$ is invariant for (3.8.12). Therefore, the transition equations of (3.8.12) at (T_0, p_0) are referred to the following form

$$\begin{aligned} \frac{d\rho_1}{dt} &= \lambda_1 \rho_1 - a_1 \rho_n \rho_1 - a_3 \rho_1^2, \\ \frac{d\rho_n}{dt} &= \lambda_3 \rho_n - c_1 \rho_1 + c_3 \rho_n^2 - c_4 \rho_n^3. \end{aligned} \tag{3.8.26}$$

The second-order approximation of the center manifold function ρ_n of (3.8.26) satisfies the equation

$$\lambda_3 \rho_n + c_3 \rho_n^2 = c_1 \rho_1.$$

Its solution is

$$\rho_n = -\frac{c_1 \rho_1}{|\lambda_3|} + \frac{c_1^2 c_3}{|\lambda_3|^3} \rho_1^2 + o(\rho_1^2).$$

Putting ρ_n in the first equation of (3.8.26) we get the reduced equation of (3.8.12) on center manifold as follows

$$\frac{d\rho_1}{dt} = \lambda_1 \rho_1 + \frac{1}{|\lambda_3|} A_1 \rho_1^2 - \frac{a_1 c_1 c_3}{|\lambda_3|^3} \rho_1^3 + o(\rho_1^3). \quad (3.8.27)$$

Assertion (2) follows from (3.8.27).

Likewise, if $(T_0, p_0) \in \widehat{cd}$, $\lambda_1(T_0, p_0) < 0$ and the space of $(0, \rho_2, \rho_n)$ is invariant for (3.8.12), therefore in the same fashion as above we can prove Assertion (1). The proof is complete. \square

Proof of Theorem 3.8.3. We proceed in the following two cases.

CASE 1: $A_1 \leq 0$ IN \widehat{bc} . In this case, by Theorem 3.8.2, the transition of (3.8.12) in \widehat{bc} is Type-I, and the transition solution $(\rho_1^*, 0, \rho_n^*)$ satisfies that

$$\rho_n^* = -\frac{c_1}{|\lambda_3|} \rho_1^*.$$

The equations describing the second transition are given by (3.8.16) and the eigenvalues in (3.8.17) and (3.8.18) are rewritten as

$$\beta_1 = \lambda_2 + b_1 |\rho_n^*| - b_2 \rho_1^* = \lambda_2 + \frac{1}{|\lambda_3|} B_2 \rho_1^*, \quad (3.8.28)$$

$$\beta_+ = \frac{1}{2} [\tilde{\lambda}_1 + \tilde{\lambda}_3 + \sqrt{(\tilde{\lambda}_1 - \tilde{\lambda}_3)^2 + 4a_1 c_1 \rho_1^*}] \quad (3.8.29)$$

= (by (3.8.23) and ρ_1^* , $\tilde{\lambda}_1$ being small)

$$= -\lambda_1 + o(|\lambda_1|),$$

$$\beta_1 < \beta_+. \quad (3.8.30)$$

In addition, we know that

$$\begin{aligned} \lambda_1 &= 0, & \lambda_2 &< 0 & \text{on } \widehat{bc}, \\ \lambda_2(T - \delta, p) &> 0 & & \text{for } (T, p) \in \widehat{fc} \text{ and } \delta > 0. \end{aligned}$$

Hence, by assumptions (3.8.22) and (3.8.23), from (3.8.28) to (3.8.30) we can infer that there exists a curve segment $\widehat{f'c}$ near \widehat{fc} such that for $(T_2, p) \in \widehat{bc}$ and $(T_0, p) \in \Gamma'_3$ we have

$$\beta_1(T, p) \begin{cases} < 0 & \text{if } T_0 < T \leq T_2, \\ = 0 & \text{if } T = T_0, \\ > 0 & \text{if } T < T_0, \end{cases} \quad (3.8.31)$$

and $\beta_- < \beta_+ = -\lambda_1 + o(|\lambda_1|) < 0$. Hence, by Theorem 2.1.3, the system (3.8.16) has a transition on Γ'_3 .

To determine the transition type, we consider the center manifold function of (3.8.16), which satisfies

$$\begin{pmatrix} \tilde{\lambda}_1 & -a_1\rho_1^* \\ -c_1 & \tilde{\lambda}_3 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_n \end{pmatrix} = \begin{pmatrix} a_2\rho_1^*\rho_2 \\ c_2\rho_2 \end{pmatrix} + o(\rho_2). \quad (3.8.32)$$

The solution of (3.8.32) is

$$\rho_1 = \frac{a_2\tilde{\lambda}_3 + c_2a_1}{\tilde{\lambda}_3\tilde{\lambda}_1 - c_1a_1\rho_1^*}\rho_1^*\rho_2 + o(\rho_2), \quad \rho_n = \frac{\tilde{\lambda}_1c_2 + c_1a_2\rho_1^*}{\tilde{\lambda}_3\tilde{\lambda}_1 - c_1a_1\rho_1^*}\rho_2 + o(\rho_2).$$

From (3.8.17), (3.8.23), and (3.8.24) we can obtain

$$\rho_1 \simeq \frac{A_2\rho_2}{a_3|\lambda_3|}, \quad \rho_n \simeq -\frac{c_2}{|\lambda_3|}\rho_2. \quad (3.8.33)$$

Inserting the center manifold function (3.8.33) into the second equation of (3.8.16), we get the reduced equation as

$$\frac{d\rho_2}{dt} = \tilde{\lambda}_2\rho_2 + \frac{1}{|\lambda_3|}(B_1 - \frac{b_2A_2}{a_3})\rho_2^2$$

By (3.8.25), the transition of (3.8.16) is Type-II.

CASE 2. $A_1 > 0$ in \widehat{bc} . In this case, the transition of (3.8.12) in \widehat{bc} is Type-II, and the transition solution in \widehat{bc} is

$$\rho_1^* = \frac{a_1}{a_3}|\rho_n^*|, \quad \rho_n^* = -\frac{c_3}{2c_4} \left(\sqrt{1 + \frac{4c_4c_1a_1}{a_3c_3^2}} - 1 \right).$$

The eigenvalue β_1 in (3.8.18) reads

$$\beta_1 = \lambda_2 + \frac{1}{a_3}(b_1a_3 - b_2a_1)|\rho_n^*|.$$

By (3.8.22) and (3.8.23) it implies that (3.8.31) holds. Hence (3.8.12) has a second transition in Γ'_3 for $A_1 > 0$ in \widehat{bc} .

Under the condition (3.8.24), we have

$$\rho_n^* \simeq -\frac{a_1c_1}{a_3c_3}, \quad \tilde{\lambda}_1 \simeq -a_3\rho_1^*, \quad \tilde{\lambda}_3 \simeq \lambda_3 - \frac{2a_1c_1}{a_3}.$$

By $A_1 > 0$ we get that $|\lambda_3| \leq 0(\varepsilon^2)$. Thus, the solutions of (3.8.32) can be rewritten as

$$\rho_1 \simeq \frac{a_1c_1}{\tilde{\lambda}_1\tilde{\lambda}_3 - c_1a_1\rho_1^*}\rho_1^*\rho_2 \simeq \frac{c_2}{c_1}\rho_2, \quad \rho_2 \simeq -\frac{a_3c_2}{a_1c_1}\rho_2.$$

Putting ρ_1 and ρ_n into the second equation of (3.8.16), we obtain reduced equation on the center manifold as

$$\frac{d\rho_2}{dt} = \tilde{\lambda}_2 \rho_2 + \left(\frac{b_1 c_2 a_3}{a_1 c_1} - \frac{c_2 b_2}{c_1} - b_3 \right) \rho_2^2.$$

Due to (3.8.24) we see that

$$\frac{b_1 c_2 a_3}{a_1 c_1} - \frac{c_2 b_2}{c_1} - b_3 > 0.$$

Therefore, this transition of (3.8.16) is Type-II.

It is clear that the second transition solutions $(\tilde{\rho}_1, \tilde{\rho}_2, \tilde{\rho}_n)$ satisfy that $\tilde{\rho}_2 > 0$. Thus, the theorem is proved. \square

Proof of Theorem 3.8.4. At point $C = (T_0, p_0)$, $\lambda_1(T_0, p_0) = 0$, $\lambda_2(T_0, p_0) = 0$. Hence, the center manifold function of (3.8.12) at (T_0, p_0) reads

$$\rho_n = -\frac{c_1}{|\lambda_3|} \rho_1 - \frac{c_2}{|\lambda_3|} \rho_2 + \frac{c_3}{|\lambda_3|^3} (c_1 \rho_1 + c_2 \rho_2)^2.$$

Putting ρ_n in the first two equations of (3.8.12) we get the reduced equations on the center manifold as

$$\begin{aligned} \frac{d\rho_1}{dt} &= \lambda_1 \rho_1 + \frac{1}{|\lambda_3|} (A_1 \rho_1^2 + A_2 \rho_1 \rho_2) - \frac{c_3 a_1}{|\lambda_3|^3} (c_1 \rho_1 + c_2 \rho_2)^2 \rho_1, \\ \frac{d\rho_2}{dt} &= \lambda_2 \rho_2 + \frac{1}{|\lambda_3|} (B_1 \rho_2^2 + B_2 \rho_1 \rho_2) - \frac{c_3 b_1}{|\lambda_3|^3} (c_1 \rho_1 + c_2 \rho_2)^2 \rho_2. \end{aligned} \quad (3.8.34)$$

To verify the Type-I transition, by the attractor bifurcation theorem, it suffices to consider the following equations:

$$\begin{aligned} \frac{d\rho_1}{dt} &= A_1 \rho_1^2 + A_2 \rho_1 \rho_2 - \frac{c_3 a_1}{|\lambda_3|^2} (c_1 \rho_1 + c_2 \rho_2)^2 \rho_1, \\ \frac{d\rho_2}{dt} &= B_1 \rho_2^2 + B_2 \rho_1 \rho_2 - \frac{c_3 b_1}{|\lambda_3|^2} (c_1 \rho_1 + c_2 \rho_2)^2 \rho_2. \end{aligned} \quad (3.8.35)$$

Since (3.8.9) have variational structure, the flows of (3.8.35) are of gradient type. Therefore, $(\rho_1, \rho_2) = 0$ has no elliptic region for (3.8.35). Hence, in the same fashion as used in Section 6.3 in Ma and Wang (2005b), one can prove that the region

$$S = \{(\rho_1, \rho_2) \in \mathbb{R}^2 \mid \rho_1 > 0, \rho_2 > 0\}$$

is a stable parabolic region. Namely, $(\rho_1, \rho_2) = 0$ is an asymptotically stable singular point of (3.8.35) if and only if one of the two conditions (i) and (ii) holds true. Thus, we only need to prove Assertions (1)–(4).

For Type-I transition at point $C = (T_0, p_0)$, by conditions (i) and (ii), $A_1 < 0$ and $B_1 < 0$. Hence, as $\lambda_1 > 0$, $\lambda_2 > 0$ there are bifurcated solutions of (3.8.34) in the ρ_1 -axis and ρ_2 -axis as

$$z_1 = (\rho_1^*, 0) = \left(\frac{|\lambda_3|}{|A_1|} \lambda_1, 0 \right), \quad z_2 = (0, \rho_2^*) = \left(0, \frac{|\lambda_3|}{|B_1|} \lambda_2 \right).$$

The Jacobian matrices of (3.8.34) at z_1 and z_2 are given by

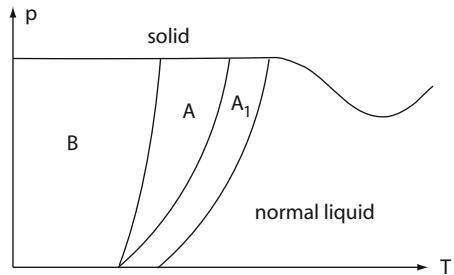
$$J(z_1) = \begin{pmatrix} -\lambda_1 & * \\ 0 & \lambda_2 + \frac{B_2}{|A_1|}\lambda_1 \end{pmatrix}, \quad J(z_2) = \begin{pmatrix} \lambda_1 + \frac{A_2}{|B_1|}\lambda_2 & 0 \\ * & -\lambda_2 \end{pmatrix}.$$

Since (3.8.34) has at most three bifurcated singular points in region \bar{S} , there are only four types of Type-I transitions, as shown in Fig. 3.48a–d, and each type is completely determined by the signs of the eigenvalues of $J(z_1)$ and $J(z_2)$. Thus, from the eigenvalues $\beta(z_1) = \frac{1}{|A_1|}(\lambda_2|A_1| + \lambda_1 B_2)$ of $J(z_1)$ and $\beta(z_2) = \frac{1}{|B_1|}(\lambda_1|B_1| + \lambda_2 A_2)$ of $J(z_2)$ it is ready to derive Assertions (1)–(4). The proof is complete. \square

3.8.4 Liquid ^3He with Nonzero Applied Field

When liquid ^3He is placed in a magnetic field H , the superfluid transition is different from that with $H = 0$. Experiments show that as a magnetic field is applied, a new superfluid phase A_1 appears, and the region of phase A can be extended to the bottom at $p = 0$. The PT -phase diagram is schematically illustrated in Fig. 3.49.

Fig. 3.49 PT -phase diagram of ^3He in a magnetic field



As a magnetic field H is applied, there is a new pairing state to appear, in which the spin of pairing atoms is parallel to the magnetic field. This new state corresponds to the phase A_1 and is expressed by

$$\sqrt{2}|\Phi\rangle = 2a|\uparrow\uparrow\rangle.$$

We introduce the complex valued functions ψ_0 to the state $|\uparrow\uparrow\rangle$, ψ_1 to the state $|\downarrow\downarrow\rangle$, and ψ_2 to this state $|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle$. Let ρ_0, ρ_a, ρ_b stand for the densities of superfluid phases A_1, A, B respectively. Then we have

$$\begin{aligned} \rho_0 &= |\psi_0|^2, \\ \rho_a &= \tau_0|\psi_0|^2 + \tau_1|\psi_1|^2 \quad (\tau_0 > 0, \tau_1 > 0), \end{aligned}$$

$$\rho_b = \tau_2|\psi_0|^2 + \tau_3|\psi_1|^2 + \tau_4|\psi_2|^2 \quad (\tau_2, \tau_3 \geq 0, \tau_4 > 0),$$

and the total density of liquid ${}^3\text{He}$ in a magnetic field is given by

$$\rho = \begin{cases} \rho_n + \rho_0 & \text{at state } A_1, \\ \rho_n + \rho_\alpha & \text{at state } A, \\ \rho_n + \rho_b & \text{at state } B. \end{cases}$$

Thus, similar to (3.8.2), for liquid ${}^3\text{He}$ with $H \neq 0$ we give the Ginzburg-Landau free energy in the following form. For simplicity we take the nondimensional form:

$$\begin{aligned} G(\psi_0, \psi_1, \psi_2, \rho_n) = & \frac{1}{2} \int_{\Omega} \left[\kappa_0 |\nabla \psi_0|^2 - \lambda_0 |\psi_0|^2 + \alpha_0 \rho_n |\psi_0|^2 \right. \\ & + \alpha_1 |\psi_0|^2 |\psi_1|^2 + \alpha_2 |\psi_0|^2 |\psi_2|^2 + \frac{\alpha_3}{2} |\psi_0|^4 \\ & + \kappa_1 |\nabla \psi_1|^2 - \lambda_1 |\psi_1|^2 + a_1 \rho_n |\psi_1|^2 \\ & + a_2 |\psi_1|^2 |\psi_2|^2 + \frac{a_3}{2} |\psi_1|^4 \\ & + \kappa_2 |\nabla \psi_2|^2 - \lambda_2 |\psi_2|^2 + b_1 \rho_1 |\psi_2|^2 + \frac{b_3}{2} |\psi_2|^4 \\ & \left. + \kappa_3 |\nabla \rho_n|^2 - \lambda_3 |\rho_n|^2 - \frac{c_3}{3} \rho_n^3 - \frac{c_4}{4} \rho_n^4 \right] dx. \end{aligned} \quad (3.8.36)$$

The equations describing liquid ${}^3\text{He}$ with $H \neq 0$ read

$$\begin{aligned} \frac{\partial \psi_0}{\partial t} &= \kappa_0 \Delta \psi_0 + \lambda_0 \psi_0 - \alpha_0 \rho_n \psi_0 - \alpha_1 |\psi_1|^2 \psi_0 - \alpha_2 |\psi_2|^2 \psi_0 - \alpha_3 |\psi_0|^2 \psi_0, \\ \frac{\partial \psi_1}{\partial t} &= \kappa_1 \Delta \psi_1 + \lambda_1 \psi_1 - a_1 \rho_n \psi_1 - \alpha_1 |\psi_0|^2 \psi_1 - a_2 |\psi_2|^2 \psi_1 - a_3 |\psi_1|^2 \psi_1, \\ \frac{\partial \psi_2}{\partial t} &= \kappa_2 \Delta \psi_2 + \lambda_2 \psi_2 - b_1 \rho_n \psi_2 - \alpha_2 |\psi_0|^2 \psi_2 - a_2 |\psi_1|^2 \psi_2 - b_3 |\psi_2|^2 \psi_2, \\ \frac{\partial \rho_n}{\partial t} &= \kappa_3 \Delta \rho_n + \lambda_3 \rho_n - \frac{\alpha_0}{2} |\psi_0|^2 - \frac{\alpha_1}{2} |\psi_1|^2 - \frac{b_1}{2} |\psi_2|^2 + c_2 \rho_n^2 - c_3 \rho_n^3, \\ \frac{\partial}{\partial n}(\psi_0, \psi_1, \psi_2, \rho_n) &= 0 \quad \text{on } \partial \Omega, \end{aligned} \quad (3.8.37)$$

where the coefficients satisfy that for any $0 \leq i \leq 3$ and $1 \leq j \leq 3$,

$$\alpha_i > 0, \quad a_j > 0, \quad b_1, b_3, c_2, c_3 > 0.$$

Equations (3.8.37) should be the same as (3.8.9) for $H = 0$. Therefore we assume that when $H = 0$,

$$\kappa_0 = \kappa_1, \quad \lambda_0 = \lambda_1, \quad \alpha_0 = a_1, \quad \alpha_1 = 0, \quad \alpha_2 = a_2, \quad \alpha_3 = a_3. \quad (3.8.38)$$

Based on the physical facts, we also assume that

$$\begin{aligned}\lambda_0 &= \lambda_1(T, p) + \tilde{\lambda}(T, p, H), \\ \tilde{\lambda}(T, p, H) &> 0 \quad \text{if } H \neq 0, \\ \tilde{\lambda}(T, p, H) &\rightarrow 0 \quad \text{if } H \rightarrow 0.\end{aligned}\tag{3.8.39}$$

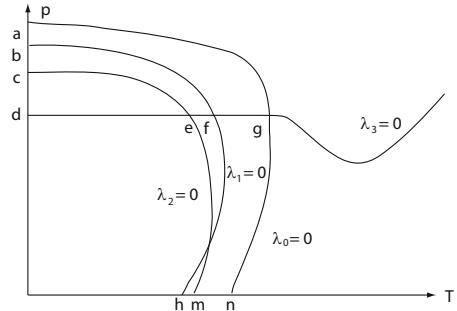
When the magnetic field H and the pressure p are homogeneous on Ω , the problem (3.8.37) can be reduced to

$$\begin{aligned}\frac{d\rho_0}{dt} &= \lambda_0\rho_0 - \alpha_0\rho_n\rho_0 - \alpha_1\rho_1\rho_0 - \alpha_2\rho_2\rho_0 - \alpha_3\rho_0^2, \\ \frac{d\rho_1}{dt} &= \lambda_1\rho_1 - a_1\rho_n\rho_1 - a_1\rho_0\rho_1 - a_2\rho_2\rho_1 - a_3\rho_1^2, \\ \frac{d\rho_2}{dt} &= \lambda_2\rho_2 - b_1\rho_n\rho_2 - a_2\rho_0\rho_2 - a_2\rho_2\rho_1 - b_3\rho_2^2, \\ \frac{d\rho_n}{dt} &= \lambda_3\rho_n - \frac{a_0}{2}\rho_0 - \frac{a_1}{2}\rho_1 - \frac{b_1}{2}\rho_2 + c_2\rho_n^2 - c_3\rho_n^3,\end{aligned}\tag{3.8.40}$$

where $\rho_i = |\psi_i|^2$ ($i = 0, 1, 2$), and λ_i ($i = 1, 2, 3$) are as in (3.8.12).

Let λ_1 , λ_2 , and λ_3 be that as shown in Fig. 3.45a–c respectively. Then, due to (3.8.39) the curves $\lambda_j(T, p) = 0$ ($0 \leq j \leq 3$) in PT -plane are schematically illustrated in Fig. 3.50.

Fig. 3.50 Curve \widehat{agn} is $\lambda_0 = 0$, \widehat{bfh} is $\lambda_1 = 0$, \widehat{cem} is $\lambda_2 = 0$, and \widehat{defg} is $\lambda_3 = 0$



Let the applied magnetic field $H \neq 0$ such that

$$0 < \tilde{\lambda}(T, p, H) < \varepsilon, \quad \text{for } \varepsilon > 0 \text{ small},\tag{3.8.41}$$

where $\tilde{\lambda}$ is as in (3.8.39). Then, under the conditions (3.8.22), (3.8.23), (3.8.38), (3.8.39), and (3.8.41), by using the same fashion as in Theorems 3.8.2 and 3.8.3, we can prove the following transition theorem for (3.8.40).

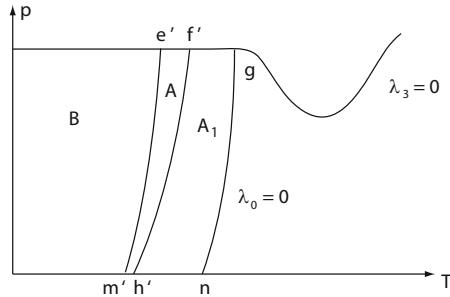
Theorem 3.8.6 *Assume the conditions (3.8.22), (3.8.23), (3.8.38), (3.8.39), and (3.8.41), then for $H \neq 0$ there exist two curve segments $\widehat{f'h'}$ near $\lambda_1 = 0$ and $\widehat{e'm'}$ near $\lambda_2 = 0$ in the PT -plane as shown in Fig. 3.51 such that the following assertions hold true:*

- (1) The system (3.8.40) has a transition in curve segment $\lambda_0 = 0$ with $\lambda_3 < 0$ (i.e., the curve segment \widehat{gn} in Fig. 3.51), which is Type-I for $\alpha_0^2 - 2|\lambda_3|\alpha_3 \leq 0$, and is Type-II for $\alpha_0^2 - 2|\lambda_3|\alpha_3 > 0$. The transition solution is given by $(\rho_0^*, 0, 0, \rho_n^*)$ with $\rho_0^* > 0, \rho_n^* < 0$.
- (2) The system has a second transition from $(\rho_0^*, 0, 0, \rho_n^*)$ in the curve segment $\widehat{f'h'}$ (i.e., $\lambda_1 - a_1\rho_n^* - a_1\rho_0^* = 0$), and the transition solution is as $(\rho'_0, \rho'_1, 0, \rho'_n)$ with $\rho'_0 > 0, \rho'_1 > 0$ and $\rho'_n < 0$.
- (3) The system has a third transition from $(\rho'_0, \rho'_1, 0, \rho'_n)$ in the curve segment $\widehat{e'm'}$ (i.e., $\lambda_2 - b_1\rho'_n - a_2\rho'_0 - a_2\rho'_1 = 0$), and the transition solution is $(\rho''_0, \rho''_1, \rho''_2, \rho''_n)$ with $\rho''_i > 0 (0 \leq i \leq 2)$ and $\rho'_n < 0$.

Remark 3.8.7 The first transition of (3.8.40) in curve segment \widehat{gn} corresponds to the phase transition of ${}^3\text{He}$ in a magnetic field between the normal liquid and superfluid phase A_1 , and the second transition in $\widehat{f'h'}$ corresponds to the phase transition between superfluid phase A_1 and A , and the third transition in $\widehat{e'm'}$ corresponds to the phase transition between superfluid phases A and B ; see Fig. 3.51.

Remark 3.8.8 The transition theorems, Theorems 3.7.2–3.8.3 and 3.8.6, provide theoretical foundation to explain the PT -phase diagrams of superfluidity, meanwhile they support these models of liquid He which are based on the phenomenology.

Fig. 3.51 Phase transition diagram



3.8.5 Physical Remarks

By carefully examining the classical phase transition diagrams and with both mathematical and physical insights offered by the dynamical transition theory, we derived two new models for superfluidity of ${}^3\text{He}$ with or without applied field. A crucial component of these two models is the introduction of three wave functions to represent Anderson-Brinkman-Morel (ABM) and the Balian-Werthamer (BM) states.

Then we have obtained a theoretical PT -phase diagram of ${}^3\text{He}$ as shown in Fig. 3.44, based on the models and the dynamic phase transition analysis. A few main characteristics of the results are as follows.

First, the analysis shows the existence of an unstable region $H = H_1 \cup H_2$, in which the solid state and the superfluid states A and B appear randomly depending on fluctuations. In particular, in H_1 , phase B superfluid state and the solid may appear, and in H_2 the phase A superfluid state and the solid state may appear.

Second, theoretical analysis suggests the existence of phase C superfluid state, which is characterized by the wave function ψ_2 , representing $|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle$. However, phase C region is very narrow, which may be the reason why it is hard to be observed in experiments.

Third, the curve \widehat{bcd} is the first critical curve where phase transition between normal fluid and superfluid states occur. The curve $\widehat{f'c}$ is the coexistence curve between phases A and B . The curve \widehat{bc} is the coexistence curve between normal fluid state and the phase A superfluid state, the curve \widehat{cd} is the coexistence curve between normal fluid state and the phase C superfluid state, the curve $\widehat{ch'}$ is the coexistence curve between the phases B and C superfluid states.

Fourth, Theorems 3.8.2–3.8.4 imply that near the two triple points b and c , there is a possibility of the existence of two switch points, where the transition on the corresponding coexistence curve switches types at each switch point. The existence of such switch points depends on the physical parameters.

In comparison to classical results as shown in Fig. 3.43, our results lead to the predictions of the existence of (1) an unstable region H , (2) a new phase C in a narrow region, and (3) switch points. It is hoped that these predictions will be useful for designing better physical experiments and lead to better understanding of the physical mechanism of superfluidity.

3.9 Mixture of He-3 and He-4

The main objectives of this section are to study λ -phase transitions of liquid helium-4 and phase separations between liquid helium-3 and liquid helium-4 from both the modeling and the analysis points of view. This section is based on a recent paper (Ma and Wang, 2009d).

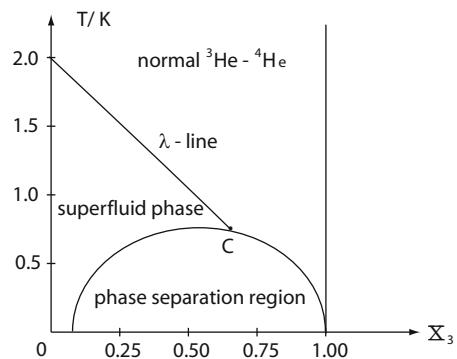
The main ingredient in the modeling is that we use an order parameter ψ for the phase transition of liquid ${}^4\text{He}$ between the normal and superfluid states, and the mol fraction u for liquid ${}^3\text{He}$. As u is a conserved quantity, a Cahn–Hilliard type equation is needed for u , and a Ginzburg–Landau type equation is needed for the order parameter.

The analysis leads to three critical length scales $L_1 < L_2 < L_3$ and the corresponding λ -transition and phase separation diagrams. The derived theoretical phase diagrams based on our analysis agree with classical phase diagram, as shown, e.g., in Reichl (1998), Onuki (2002), and it is hoped that the study here will lead to a better understanding of mature of superfluids.

3.9.1 Model for Liquid Mixture of ${}^3\text{He}$ - ${}^4\text{He}$

Liquid ${}^3\text{He}$ and ${}^4\text{He}$ can be dissolved into each other. When ${}^3\text{He}$ -atoms are dissolved in liquid ${}^4\text{He}$ and the density of ${}^3\text{He}$ increases, the λ -transition temperature T_λ decreases; see the liquid mixture phase diagram of ${}^3\text{He}$ - ${}^4\text{He}$ (Fig. 3.52), where $X = n_3/(n_3 + n_4)$, n_3 , and n_4 are the atom numbers of ${}^3\text{He}$ and ${}^4\text{He}$ respectively. When $X = 0$, $T_\lambda = 2.17\text{K}$, where λ -phase transition takes place and liquid ${}^4\text{He}$ undergoes a transition to superfluid phase from the normal liquid phase. When $X = 0.67$ and temperature decreases to $T = 0.87\text{K}$, i.e., at the triple point C in Fig. 3.52, the liquid mixture of ${}^3\text{He}$ - ${}^4\text{He}$ has a phase separation.

Fig. 3.52 Liquid mixture phase diagram of ${}^3\text{He}$ - ${}^4\text{He}$



Let the complex valued function $\psi = \psi_1 + i\psi_2$ describe the superfluidity of ${}^4\text{He}$, and u be the density of ${}^3\text{He}$, which is conserved, i.e.,

$$\int_{\Omega} u dx = c \quad (c > 0 \text{ is a fixed number}). \quad (3.9.1)$$

The Ginzburg-Landau (Gibbs) free energy is taken in the following form:

$$\begin{aligned} G(\psi, \rho_n, u) = & \int_{\Omega} \left[\frac{k_1 h^2}{2m} |\nabla \psi|^2 + \frac{\gamma_1}{2} |\psi|^2 + \frac{\gamma_2}{4} |\psi|^4 \right. \\ & \left. + \frac{k_2}{2} |\nabla u|^2 + \frac{\nu_1}{2} u^2 - \frac{\nu_2}{3} u^3 + \frac{\nu_3}{4} u^4 + \nu_4 u |\psi|^2 \right] dx. \end{aligned} \quad (3.9.2)$$

where h is the Planck constant, m is the mass of helium-4 atom, and the coefficients satisfy

$$k_1, k_2, \gamma_1, \gamma_2, \nu_1, \nu_2, \nu_3, \nu_4 > 0. \quad (3.9.3)$$

When $\psi = 0$, the liquid mixture of helium-3 and helium-4 is a binary system, and $G(0, u)$ stands for the Cahn-Hilliard free energy. By the standard model (3.1.2) and (3.1.3), from (3.9.1) and (3.9.2), the equations governing liquid mixture of ${}^3\text{He}$ - ${}^4\text{He}$ are given as follows²:

² We used the classical Cahn-Hilliard modeling here for the conserved state function u ; see (7.1.10) and the remark given there.

$$\begin{aligned}\frac{\partial\psi}{\partial t} &= \frac{k_1 h^2}{m} \Delta\psi - \gamma_1\psi - \gamma_2|\psi|^2\psi - 2\nu_4 u\psi, \\ \frac{\partial u}{\partial t} &= -k_2 \Delta^2 u + \Delta[\nu_1 u - \nu_2 u^2 + \nu_3 u^3 + \nu_4 |\psi|^2].\end{aligned}\quad (3.9.4)$$

These equations have a physically sound constant steady-state solution given by

$$(\psi, u) = (0, u^0),$$

where $u^0 > 0$ is the density of helium-3 in a homogeneous state. For simplicity, we assume the total density $\rho = 1$. Then the system control parameter X , the mol fraction of ${}^3\text{He}$, becomes

$$X = u^0 \quad \text{with } 0 \leq X \leq 1.$$

Now we consider the derivations from this basic state:

$$(\psi, u) = (\psi', X + u'),$$

and we derive the following equations (drop the primes):

$$\begin{aligned}\frac{\partial\psi}{\partial t} &= \mu_1 \Delta\psi - \lambda_1\psi - \gamma_2|\psi|^2\psi - 2\nu_4 u\psi, \\ \frac{\partial u}{\partial t} &= -\mu_2 \Delta^2 u + \lambda_2 \Delta u + \Delta[(-\nu_2 + 3\nu_3 X)u^2 + \nu_3 u^3 + \nu_4 |\psi|^2], \\ \int_{\Omega} u dx &= 0,\end{aligned}\quad (3.9.5)$$

with the Neumann boundary condition

$$\frac{\partial}{\partial n}(u, \Delta u, \psi) = 0 \quad \text{on } \partial\Omega, \quad (3.9.6)$$

where $\mu_1 = k_1 h^2 / m$, $\mu_2 = k_2$, and

$$\lambda_1 = \gamma_1 + 2\nu_4 X, \quad \lambda_2 = \nu_1 - 2\nu_2 X + 3\nu_3 X^2. \quad (3.9.7)$$

3.9.2 Critical Parameter Curves

For simplicity we only consider the case where $\Omega = (0, L) \times (0, l)^2 \subset \mathbb{R}^3$ is a rectangle with $L > l$, and the control parameters are the temperature T and the mol fraction X , and the length scale L of the container Ω .

Physically, by the Hildebrand theory (Reichl, 1998), in the lower temperature region and at $p = 1$ atm, the critical parameter curve $\lambda_2 = 0$ is equivalent to

$$T = \frac{2a}{R}(1 - X)X - \sigma_0, \quad (3.9.8)$$

where R is the molar gas constant and $a > 0$ is a constant. Here $\sigma_0 > 0$ is small correction term. The original Hildebrand theory leads to the case where $\sigma_0 = 0$. However, as we can see from the classical phase separation of a binary system, the Hildebrand theory fails when the molar fraction is near 0 or 1, and the correction term added here agrees with the experimental phase diagram as shown, e.g., in Figure 4.13 in Reichl (1998).

Furthermore, by (3.9.7) and (3.9.8), we have

$$\nu_1 \simeq \theta_1(T + \sigma_0), \quad 3\nu_3 \simeq 2\nu_2, \quad \frac{2\nu_2}{\theta_1} \simeq \frac{2a}{R}. \quad (3.9.9)$$

Consider eigenvalue problem of the linear operator in (3.9.5):

$$\begin{pmatrix} \mu_1\Delta - \lambda_1 & 0 & 0 \\ 0 & \mu_1\Delta - \lambda_1 & 0 \\ 0 & 0 & -\Delta(\mu_2\Delta - \lambda_2) \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \\ u \end{pmatrix} = \beta \begin{pmatrix} \psi_1 \\ \psi_2 \\ u \end{pmatrix}. \quad (3.9.10)$$

Here $\psi = \psi_1 + \psi_2$. It is known that the first eigenvalue and eigenvector of the Laplacian operator with the Neumann condition and zero average are $\lambda = \pi^2/L^2$ and $u = \cos \pi x_1/L$. Thus, the first two eigenvalues and their corresponding eigenvectors of (3.9.10) are given by

$$\begin{cases} \beta_1 = -\lambda_1 = -(\gamma_1 + 2\nu_4 X), \\ (\psi_1, \psi_2, u) = (1, 0, 0) \text{ and } (0, 1, 0), \end{cases} \quad (3.9.11)$$

$$\begin{cases} \beta_2 = -\frac{\pi^2}{L^2} \left(\frac{\mu_2\pi^2}{L^2} + \lambda_2 \right) = -\frac{\pi^2}{L^2} \left(\frac{\mu_2\pi^2}{L^2} + \nu_1 - 2\nu_2 X + 3\nu_3 X^2 \right), \\ (\psi_1, \psi_2, u) = (0, 0, \cos \pi x_1/L). \end{cases} \quad (3.9.12)$$

As for ν_1 , the parameter γ_1 is approximately a linear function of T . Phenomenologically, we take

$$\gamma_1 = -\sigma_1 + \theta_2 T \quad (\theta_2, \sigma_1 > 0). \quad (3.9.13)$$

Then, by (3.9.9) and (3.9.11)–(3.9.13), the critical parameter curves in the TX -plane are as follows:

$$\begin{aligned} l_1 : \beta_1 = 0 &\iff T_{c1} = \frac{\sigma_1 - 2\nu_4 X}{\theta_2} = \frac{\sigma_1}{\theta_2}(1 - X), \\ l_2 : \beta_2 = 0 &\iff T_{c2} = \frac{2a}{R}X(1 - X) - \sigma_0 - \frac{\mu_2\pi^2}{\theta_1 L^2}. \end{aligned} \quad (3.9.14)$$

Here we have assumed that $\sigma_1 = 2\nu_4$. This assumption is true at $X = 1$ as required by $T_{c1} = 0$ at $X = 1$, and is true as well at other values of X as we approximately take σ_1 and ν_4 as constants.

The critical parameter curve l_1 ($\beta_1 = 0$) is as shown in Fig. 3.53a. Let

$$L_0 = \sqrt{\frac{\mu_1\pi^2}{\theta_1 \left(\frac{a}{2R} - \sigma_0 \right)}} \quad \left(\sigma_0 < \frac{a}{2R} \right), \quad (3.9.15)$$

then when $L > L_0$ the critical parameter curve l_2 ($\beta_2 = 0$) is as shown in Fig. 3.53b.

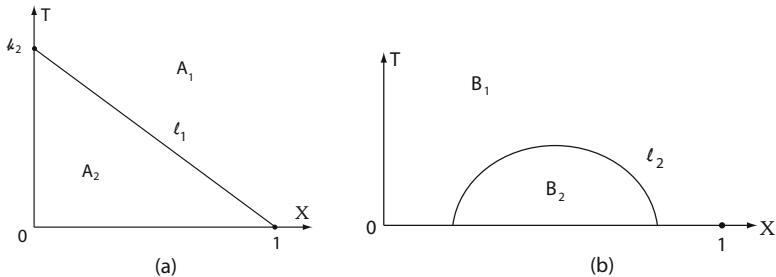


Fig. 3.53 (a) $\beta_1 < 0$ in region A_1 , and $\beta_1 > 0$ in A_2 , (b) $\beta_2 < 0$ in region B_1 , and $\beta_2 > 0$ in B_2

3.9.3 Transition Theorems

The following is the transition theorem for liquid mixture ^3He - ^4He . For this purpose, we introduce three length scales:

$$L_1 = \left(\frac{\mu_2 \pi^2}{\theta_1} \right)^{1/2} \cdot \frac{1}{\left(\frac{a}{2R} + \frac{2\nu_4^2}{\theta_1 \gamma_2} - \sigma_0 \right)^{1/2}}, \quad (3.9.16)$$

$$L_2 = \begin{cases} \left(\frac{\mu_2 \pi^2}{\theta_1} \right)^{1/2} \cdot \frac{1}{A^{1/2}} & \text{if } A > 0 \\ \infty & \text{otherwise,} \end{cases} \quad (3.9.17)$$

$$L_3 = \begin{cases} \left(\frac{\mu_2 \pi^2}{\theta_1} \right)^{1/2} \cdot \frac{1}{B^{1/2}} & \text{if } B > 0 \\ \infty & \text{otherwise,} \end{cases} \quad (3.9.18)$$

where

$$A = \frac{R}{8a} \left(\frac{2a}{R} - \frac{\sigma_1}{\theta_2} \right)^2 + \frac{2\nu_4^2}{\theta_1 \gamma_2} - \sigma_0, \quad B = \frac{R}{8a} \left(\frac{2a}{R} - \frac{\sigma_1}{\theta_2} \right)^2 - \sigma_0.$$

We remark here again that σ_0 is small as a correction term in the Bragg-Williams theory mentioned earlier in (3.9.8).

Theorem 3.9.1 Let T_{c1} and T_{c2} be given in (3.9.14). For equations (3.9.5), we have the following assertions:

- (1) When $L < L_1$ the system has only the superfluid phase transition (i.e., the λ -phase transition) at $T = T_c^1$, and the TX-phase diagram is as shown in Fig. 3.53a.

- (2) Let $\sigma_1/\theta_2 < 4a/R$. Then $L_1 < L_2$, and if $L_1 < L < L_2$, then there are two numbers $0 < X_1 < \frac{1}{2} < X_2 < 1$ given by

$$X_{1,2} = \frac{1}{2} \left[1 \pm \sqrt{1 + \frac{2R}{a} \left(\frac{2\nu_4^2}{\theta_1\gamma_2} - \sigma_0 - \frac{\mu_2\pi^2}{\theta_1L^2} \right)} \right] \quad (3.9.19)$$

such that the following holds true:

- (a) if $0 \leq X < X_1$ or $X_2 < X < 1$, the system has only the λ -phase transition at $T = T_c$;
 - (b) if $X_1 < X < X_2$, then the system has the λ -phase transition at $T = T_{c1}$, and has the phase separation at $T^* = T_{c2} + \frac{2\nu_4^2}{\theta_1\gamma_2} < T_{c1}$. Moreover, the TX -phase diagram is as shown in Fig. 3.54.
- (3) Let $\sigma_1/\theta_2 < 4a/R$. For any $L_2 < L < L_3$, the phase diagram is as shown in Fig. 3.55. For any $L_3 < L$, the phase diagram is as shown in Fig. 3.56.
- (4) Let $\sigma_1/\theta_2 \geq 4a/R$. Then for any $L \geq L_1$, Assertion (2) holds true.
 - (5) The λ -phase transition at $T = T_c^1$ is Type-I, which corresponds to second-order transition.

It is worth mentioning that the order of second transitions crossing $\tilde{\beta}_1 = 0$ is an interesting and open problem.

Fig. 3.54 Region D_1 is normal ${}^3\text{He}$ - ${}^4\text{He}$, D_2 is the superfluid phase, and D_3 is the phase separation region

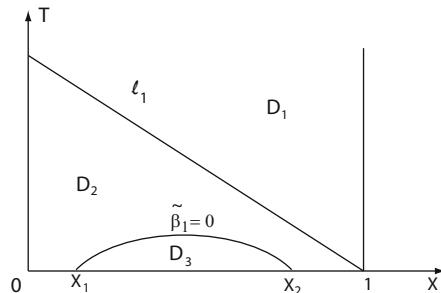


Fig. 3.55 Region D_1 is normal ${}^3\text{He}$ - ${}^4\text{He}$, D_2 is the superfluid phase, and D_3 is the phase separation region

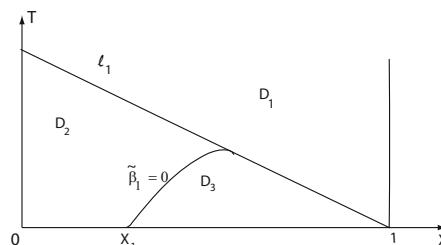
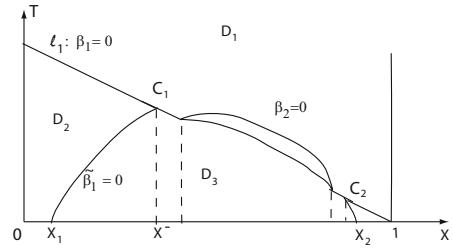


Fig. 3.56 Region D_1 is normal ${}^3\text{He} - {}^4\text{He}$, D_2 is the superfluid phase, and D_3 is the phase separation region



Proof of Theorem 3.9.1. We proceed in several steps as follows.

STEP 1. It is easy to see that the space

$$E = \{(u, \psi) = (0, \psi) \mid \psi \in \mathbb{C}\}$$

is invariant for (3.9.5). Therefore, the transition solutions of (3.9.5) from the critical parameter curve l_1 ($\lambda_1 = 0$) must be in E , which corresponds to the superfluid transition. On the other hand, restricted on E , equations (3.9.4) are equivalent to the following ordinary differential equation:

$$\frac{d\psi}{dt} = -\lambda_1\psi - \gamma_2|\psi|^2\psi. \quad (3.9.20)$$

It is then easy to see that Assertion (5) holds true, and the bifurcated solutions consist of a circle given by

$$\{\sqrt{-\lambda_1/\gamma_2} e^{i\phi} \mid \phi \in [0, 2\pi]\}.$$

STEP 2. We now consider the second transition. For this purpose, let the transition solution of (3.9.4) from $T = T_c^1$ (i.e., from $\lambda_1 = 0$) be given by

$$(0, \tilde{\psi}(T)) = (0, \sqrt{-\lambda_1/\gamma_2} e^{i\phi}) \in E. \quad (3.9.21)$$

Take the transition

$$(u, \psi) \rightarrow (u', \psi' + \tilde{\psi}).$$

Then the equations (3.9.5) are rewritten as (drop the primes)

$$\begin{aligned} \frac{\partial \psi}{\partial t} &= \mu_1 \Delta \psi + \lambda_1 \psi + \lambda_1 \psi^* - 2\nu_4 \sqrt{\frac{-\lambda_1}{\gamma_2}} u - 2\gamma_2 \sqrt{\frac{-\lambda_1}{\gamma_2}} |\psi|^2 - 2\nu_4 u \psi - \gamma_2 |\psi|^2 \psi, \\ \frac{\partial u}{\partial t} &= -\mu_2 \Delta^2 u + \lambda_2 \Delta u + 2\nu_4 \sqrt{\frac{-\lambda_1}{\gamma_2}} \Delta \psi_1 + \Delta[(-\nu_2 + 3\nu_3 X)u^2 + \nu_3 u^3 + \nu_4 |\psi|^2], \\ \int_{\Omega} u dx &= 0, \end{aligned} \quad (3.9.22)$$

where $\psi = \psi_1 + i\psi_2$

We consider the transition of (3.9.22) beyond E . The linear operator of (3.9.22) is given by

$$B = \begin{pmatrix} 2\lambda_1 & 0 & -2\nu_4\sqrt{\frac{-\lambda_1}{\gamma_2}} \\ 0 & \lambda_1 & 0 \\ 2\nu_4\sqrt{\frac{-\lambda_1}{\gamma_2}}\Delta & 0 & \Delta(-\mu_2\Delta + \lambda_2) \end{pmatrix}. \quad (3.9.23)$$

Restricted to its first eigenspace

$$E_1 = \left\{ (\psi_1, \psi_2, u) = (y_1, y_2, y_3) \cos \frac{\pi x_1}{L} \mid (y_1, y_2, y_3) \in \mathbb{R}^3 \right\},$$

the linear B is given by

$$B|_{E_1} = \begin{pmatrix} 2\lambda_1 & 0 & -2\nu_4\sqrt{\frac{-\lambda_1}{\gamma_2}} \\ 0 & \lambda_1 & 0 \\ -\frac{2\nu_4\pi^2}{L^2}\sqrt{\frac{-\lambda_1}{\gamma_2}} & 0 & -\frac{\pi^2}{L^2}(\frac{\mu_2\pi^2}{L^2} + \lambda_2) \end{pmatrix}. \quad (3.9.24)$$

The eigenvalues $\tilde{\beta}_1$, $\tilde{\beta}_2$, and $\tilde{\beta}_3$ of $B|_{E_1}$ are given by

$$\begin{aligned} \tilde{\beta}_{1,2} &= \lambda_1 - \frac{\pi^2}{2L^2} \left[\frac{\mu_2\pi^2}{L^2} + \lambda_2 \right] \\ &\pm \sqrt{\left[\lambda_1 - \frac{\pi^2}{2L^2} \left(\frac{\mu_2\pi^2}{L^2} + \lambda_2 \right) \right]^2 + \frac{2\lambda_1\pi^2}{L^2} \left(\frac{\mu_2\pi^2}{L^2} + \lambda_2 - \frac{2\nu_4^2}{\gamma_2} \right)}, \\ \tilde{\beta}_3 &= \lambda_1 < 0. \end{aligned}$$

We know that

$$\tilde{\beta}_3 < 0, \quad \tilde{\beta}_2 < \tilde{\beta}_1 \quad \forall T < T_{c1}.$$

By (3.9.9), we have

$$\begin{aligned} \lambda_2 &= \nu_1 - 2\nu_2X + 3\nu_3X^2 \\ &= \theta_1(T + \sigma_0) - 2\nu_2(1 - X)X \\ &= \theta_1 \left(T + \sigma_0 - \frac{2\nu_2}{\theta_1}X(1 - X) \right) \\ &= \theta_1(T - T_{c2}) - \frac{\nu_2\pi^2}{L^2}. \end{aligned}$$

Hence the transition curve $\tilde{\beta}_1 = 0$ is given by

$$\frac{\mu_2\pi^2}{L^2} + \lambda_2 = \frac{2\nu_4^2}{\gamma_2},$$

which is equivalent to

$$\theta_1(T - T_{c2}) = \frac{2\nu_4^2}{\gamma_2}.$$

Hence the transition curve $\tilde{\beta}_1 = 0$ is given by

$$T = T_{c2} + \frac{2\nu_4^2}{\theta_1\gamma_2} = \frac{2a}{R}X(1-X) - \sigma_0 - \frac{\mu_2\pi^2}{\theta_1L^2} + \frac{2\nu_4^2}{\theta_1\gamma_2}. \quad (3.9.25)$$

Then L_1 defined by (3.9.16) is the critical length scale to make the transition curve $\tilde{\beta}_1 = 0$ achieving its maximum at $(T, X) = (0, 1/2)$.

By definition, it is then easy to see that if $L < L_1$, no phase separation occurs at any temperature although phase transition for ${}^4\text{He}$ does occur as the temperature decreases below certain critical temperature.

STEP 3. Now we calculate the length scale L_2 where the critical curves $\beta_1 = 0$ and $\tilde{\beta}_1 = 0$ are tangent to each other, i.e., they interact at exactly one point. By definition, we need to solve L such that the equation

$$\frac{\sigma_1}{\theta_2}(1-X) = \frac{2a}{R}X(1-X) - \sigma_0 - \frac{\mu_2\pi^2}{\theta_1L^2}$$

has exactly one solution. Hence, it is easy to derive that the formula L_2 is as given by (3.9.17).

By comparing L_1 and L_2 , we have

$$L_2 > L_1 \quad \text{if and only if} \quad \frac{\sigma_1}{\theta_2} < \frac{4a}{R}. \quad (3.9.26)$$

Now we calculate the length scale L_3 where the critical curves $\beta_1 = 0$ and $\beta_2 = 0$ are tangent to each other, i.e., we need to solve L such that the equation

$$\frac{\sigma_1}{\theta_2}(1-X) = \frac{2a}{R}X(1-X) - \sigma_0$$

has exactly one solution. The formula L_3 is as given by (3.9.18).

STEP 4. Now we return to prove Assertions (1)–(4). First, we consider the case where $\frac{\sigma_1}{\theta_2} < \frac{4a}{R}$, i.e., $L_2 > L_1$. In this case, for $L_1 < L < L_2$, second transition curve $\tilde{\beta}_1 = 0$ is as shown in Fig. 3.54. Namely, as one decreases the temperature T entering from region D_2 into D_3 through the curve $\tilde{\beta}_1 = 0$, phase separation occurs.

By (3.9.25), solving

$$\frac{2a}{R}X(1-X) - \sigma_0 - \frac{\mu_2\pi^2}{\theta_1L^2} + \frac{2\nu_4^2}{\theta_1\gamma_2} = 0$$

gives the least and biggest mol fractions $X_1 < X_2$ defined by (3.9.19), and Assertion (2) follows. Other assertions can be proved in the same fashion. \square

3.9.4 Physical Conclusions

We have introduced a dynamical Ginzburg-Landau phase transition/separation model for the mixture of liquid helium-3 and helium-4.

The analysis of the model established enables us to give a detailed study on the λ -phase transition and the phase separation between liquid ^3He and ^4He . In particular, we derived three critical length scales $L_1 < L_2 < L_3$ with the following conclusions:

- (1) For $L < L_1$, there is only a λ -phase transition for ^4He and no phase separation between ^3He and ^4He , as shown in Fig. 3.53a.
- (2) For $L_1 < L < L_2$, there is no triple points, and phase separation occurs as a second phase transition after the λ -transition when the mol fraction is between two critical values, as shown in Fig. 3.54.
- (3) For $L_2 < L < L_3$, the λ -transition is always the first transition, as shown in Fig. 3.55.
- (4) For $L_3 < L$, both the λ -transition and the phase separation can appear as either the first transition or the second transition depending on the mol fractions, as shown in Fig. 3.56. In this case, when the phase separation is the first transition, the separation mechanism is the same as a typical binary system as described in the early section for binary systems.

Also, the λ -transition is always second order. The derive theoretical phase diagram based on our analysis agrees with classical phase diagram, and it is hoped that the study here will lead to a better understanding of the nature of superfluidity. Finally, we remark that the order of second transition is mathematically more challenging and will be reported elsewhere.



Chapter 4

Fluid Dynamics

One important source of dynamic transitions and pattern formation is transition and stability problems in fluid dynamics, including in particular Rayleigh–Bénard convection, the Couette–Taylor problem, the Couette–Poiseuille–Taylor problem, and the parallel shear flow problem. The study of these basic problems leads to a better understanding of turbulent behavior in fluid flows, which in turn often leads to new insights and methods in the solution of other problems in science and engineering.

Fluid flow is modeled by Navier–Stokes-type equations, which can be viewed as an infinite-dimensional dissipative system. The dissipative structure of Navier–Stokes equations is mathematically described by global absorbing sets and global attractors; see, among many others, (Foiaş and Temam, 1979; Ladyzenskaja, 1991; Vishik, 2012; Temam, 1997). The existence and properties of solutions of the Navier–Stokes equations are classical topics; see, among others, (Leray, 1933; Hopf, 1951; Kiselev and Ladyženskaya, 1957; Ladyženskaya, 1958; Caffarelli et al., 1982), (Temam, 1984, Appendix III), and the references therein.

Also, since Rayleigh’s pioneering work on classical Bénard convection, there have been intensive studies of different fluid problems from the point of view of linear analysis and local bifurcation points; see, among many others, (Chandrasekhar, 1981; Drazin and Reid, 1981; Yudovich, 1967a,b; Rabinowitz, 1968; Kirchgässner, 1975; Busse, 1978; Cross and Hohenberg, 1993; Palm, 1975; Getling, 1997; Golubitsky et al., 1988; Swinney, 1988; Friedlander et al., 2000; Iooss and Adelmeyer, 1998; Koschmieder, 1993; Lappa, 2009; Ma and Wang, 2005d,b).

In addition, the classical point of view on Lagrangian dynamics and its applications has been explored by many authors; see, among many others, (Surana et al., 2006; Samelson and Wiggins, 2006; Wiggins, 1990; Rogerson et al., 1999; Mezić, 2001) and the references therein. For the study of Lagrangian dynamics in conjunction with vortex dynamics, we refer the interested reader to Newton (2001) and the references therein.

The main focus of this chapter is dynamic transitions and pattern-formation mechanisms of classical Bénard convection, the Taylor problem, Taylor–Poiseuille flow, and the rotating convection problem. Structural transition and pattern formation

are only briefly touched on, due to the scope and limitations of this book. A crucial component of our study is based on our Principle 1, following Philosophy 2. Namely, we classify transitions into their types—continuous, catastrophic, and random—and determine patterns by searching for a complete set of transition states. The physical implications of our rigorous analysis thereby become transparent. The results presented are based mainly on recent work of the authors Ma and Wang (2004a,b, 2006, 2007a, 2009a).

For many problems in fluid dynamics, our study involves two types of transitions. The first is the dynamic transition of the solutions of the partial differential equations governing the motion and states of fluids. The second is structural transitions in physical space. Our analysis is based on the dynamic transition theory presented in this book and the geometric theory of incompressible flows synthesized in Ma and Wang (2005d).

The study of structural transitions in physical space was originally motivated by a desire to understand the dynamics of ocean currents. Our approach is to classify the topological structure of the *instantaneous* velocity field (i.e., streamlines in Eulerian coordinates) and its transitions, treating the time variable as a parameter. Based on this approach, our project has two goals: the development of a global geometric theory of divergence-free vector fields on general two-dimensional compact manifolds with or without boundaries, and the establishment of connections between the solutions of the Navier–Stokes (or Euler) equations and the dynamics of velocity fields in physical space. This work is synthesized in Ma and Wang (2005d).

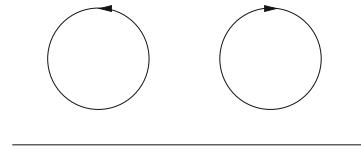
This point of view is explored throughout this chapter and in systematically in Chap. 9. This is a difficult topic, associated with topological phase transitions (TPTs) to be studied in Chap. 9. The notion and theory of TPTs have many potential applications in many areas of nonlinear sciences as demonstrated in this chapter, and in Chaps. 5 and 9.

4.1 Rayleigh–Bénard Convection

4.1.1 Bénard Problem

Convection is a well-known phenomenon of fluid motion induced by buoyancy when a fluid is heated from below. It is familiar as the driving force in atmospheric and oceanic phenomena. The Rayleigh–Bénard convection problem was originated in the famous experiments conducted by H. Bénard in 1900, who investigated a liquid in a horizontal dish with the diameter being much larger than the height, keeping the upper surface free touch with air and heating from below. He noticed that as the lower surface temperature T is less than some critical temperature T_c the liquid is in a static state, and as T is greater than T_c the liquid suddenly breaks into a rather regular cellular pattern of hexagonal convection cells. The cells have the convection roll structure in their vertical section as illustrated in Fig. 4.1.

Fig. 4.1 The vertical section of thermal convection



In 1916, Lord Rayleigh in his paper (Rayleigh, 1916) developed a linear theory to interpret the phenomena of the Bénard experiments. He chose the Boussinesq equations with some boundary conditions to model the Bénard experiments, and linearized these equations using normal modes. He then showed that the convection would occur only when the nondimensional parameter, called the Rayleigh number,

$$R = \frac{g\alpha\beta}{\kappa\nu} h^4 \quad (4.1.1)$$

exceeds a certain critical value, where g is the gravity constant, α is the coefficient of thermal expansion of the fluid,

$$\beta = |dT/dz| = (\bar{T}_0 - \bar{T}_1)/h$$

is the vertical temperature gradient with \bar{T}_0 the temperature on the lower surface and \bar{T}_1 on the upper surface, h is the depth of the layer of the fluid, κ is the thermal diffusivity, and ν is the kinematic viscosity.

As mentioned in Sect. 1.3.1, the dynamic transition and pattern formation mechanism of the Bénard convection can be thoroughly explored using the dynamic transition theory and the geometric theory of 2D incompressible flows by Ma and Wang (2005d); see, e.g., Ma and Wang (2004b, 2007a). This section is devoted to the analysis and the physical implications of this study.

4.1.2 Boussinesq Equations

The Bénard problem can be modeled by the Boussinesq equations; see among others (Rayleigh, 1916; Drazin and Reid, 1981; Chandrasekhar, 1981). They read

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla) u - \nu \Delta u + \rho_0^{-1} \nabla p &= -gk[1 - \alpha(T - \bar{T}_0)], \\ \frac{\partial T}{\partial t} + (u \cdot \nabla) T - \kappa \Delta T &= 0, \\ \operatorname{div} u &= 0, \end{aligned} \quad (4.1.2)$$

where ν, κ, α, g are the constants defined in (4.1.1), $u = (u_1, u_2, u_3)$ is the velocity field, p is the pressure function, T is the temperature function, \bar{T}_0 is a constant representing the lower surface temperature at $x_3 = 0$, and $k = (0, 0, 1)$ the unit vector in the x_3 -direction.

To make the equations nondimensional, let

$$\begin{aligned} (x, t) &= (hx', h^2 t' / \kappa), \\ (u, T) &= (\kappa u' / h, \beta h(T' / \sqrt{R}) + \bar{T}_0 - \beta h x'_3), \\ p &= \rho_0 \kappa^2 p' / h^2 + p_0 - g \rho_0 (h x'_3 + \alpha \beta h^2 (x'_3)^2 / 2), \\ \text{Pr} &= \nu / \kappa. \end{aligned}$$

Here the Rayleigh number R is defined by (4.1.1), β is as in (4.1.1), and Pr is the Prandtl number.

Omitting the primes, the equations (4.1.2) can be rewritten as

$$\begin{aligned} \frac{1}{\text{Pr}} \left[\frac{\partial u}{\partial t} + (u \cdot \nabla) u + \nabla p \right] - \Delta u - \sqrt{R} T k &= 0, \\ \frac{\partial T}{\partial t} + (u \cdot \nabla) T - \sqrt{R} u_3 - \Delta T &= 0, \\ \operatorname{div} u &= 0. \end{aligned} \tag{4.1.3}$$

The nondimensional domain is $\Omega = D \times (0, 1) \subset \mathbb{R}^3$, and $D \subset \mathbb{R}^2$ is an open set. The coordinate system is given by $x = (x_1, x_2, x_3) \in \Omega$.

The Boussinesq equations (4.1.3) are basic equations to study the Rayleigh–Bénard problem. They are supplemented with the following initial value conditions

$$(u, T) = (u_0, T_0) \quad \text{at} \quad t = 0. \tag{4.1.4}$$

Boundary conditions are needed at the top and bottom boundaries and at the lateral boundary $\partial D \times (0, 1)$. At the top and bottom boundaries ($x_3 = 0, 1$), either the Dirichlet boundary condition or the free boundary condition is given

$$(u, T) = 0 \quad \text{(Dirichlet)}, \tag{4.1.5}$$

$$T = 0, \quad u_3 = 0, \quad \frac{\partial(u_1, u_2)}{\partial x_3} = 0 \quad \text{(free)}. \tag{4.1.6}$$

On the lateral boundary $\partial D \times [0, 1]$, one of the following boundary conditions are usually used:

DIRICHLET BOUNDARY CONDITION:

$$u = 0, \quad T = 0 \quad \left(\text{or} \frac{\partial T}{\partial n} = 0 \right). \tag{4.1.7}$$

FREE BOUNDARY CONDITION:

$$u_n = 0, \quad \frac{\partial u_\tau}{\partial n} = 0, \quad T = 0 \quad \left(\text{or } \frac{\partial T}{\partial n} = 0 \right), \quad (4.1.8)$$

where n and τ are the unit normal and tangent vector on $\partial D \times [0, 1]$ respectively, and $u_n = u \cdot n, u_\tau = u \cdot \tau$.

PERIODIC CONDITION:

$$(u, T)(x_1 + k_1 L_1, x_2 + k_2 L_2, x_3) = (u, T)(x_1, x_2, x_3) \quad \text{for } k_1, k_2 \in \mathbb{Z}. \quad (4.1.9)$$

A physically sound boundary condition is a combination of (4.1.5) or (4.1.6) at the top and bottom boundary ($x_3 = 0, 1$) with one of the lateral boundary conditions (4.1.7)–(4.1.9). For simplicity, we study the phase transition problem with the following Dirichlet boundary conditions

$$(u, T) = 0 \quad \text{on } \partial\Omega. \quad (4.1.10)$$

Of course, all results obtained in this section hold true as well for other combinations of boundary conditions. For the problem (4.1.3) with (4.1.10) we define the function spaces as

$$\begin{aligned} H &= \{(u, T) \in L^2(\Omega)^4 \mid \operatorname{div} u = 0, u \cdot n = 0 \text{ on } \partial\Omega\}, \\ H_1 &= \{(u, T) \in H^2(\Omega)^4 \mid \operatorname{div} u = 0, (u, T) = 0 \text{ on } \partial\Omega\}. \end{aligned} \quad (4.1.11)$$

4.1.3 Dynamic Transition Theorems

The linearized equations of (4.1.3) are given by

$$\begin{aligned} -\Delta u + \nabla p - \sqrt{R}Tk &= 0, \\ -\Delta T - \sqrt{R}u_3 &= 0, \\ \operatorname{div} u &= 0. \end{aligned} \quad (4.1.12)$$

It is clear that this eigenvalue problem (4.1.12) with (4.1.10) for the Rayleigh number R is symmetric. Hence, we know that all R_k with multiplicities m_k of (4.1.12) with (4.1.10) are real numbers, and

$$0 < R_1 < R_2 < \cdots < R_k < R_{k+1} < \cdots.$$

The first eigenvalue R_1 , also denoted by $R_c = R_1$, is called the critical Rayleigh number. Let the multiplicity of R_c be m (≥ 1), and the first eigenvectors be $\psi_1 = (e_1, T_1), \dots, \psi_m = (e_m, T_m)$, which satisfy

$$(\psi_i, \psi_j) = \int_{\Omega} [e_i \cdot e_j + T_i \cdot T_j] dx = \delta_{ij},$$

where δ_{ij} is the Kronecker symbol. Let E_0 be the first eigenspace of (4.1.12) with (4.1.10):

$$E_0 = \left\{ \sum_{k=1}^m \alpha_k \psi_k \quad | \quad \alpha_k \in \mathbb{R}^1, 1 \leq k \leq m \right\}.$$

The following theorems provide general phase transition results for the Rayleigh–Bénard convection.

Theorem 4.1.1 *For the Bénard problem (4.1.3) with (4.1.10), the following assertions hold true:*

- (1) *The problem has a transition at the critical Rayleigh number $R = R_c$, and this transition is of Type-I.*
- (2) *When $R < R_c$ the steady-state $(u, T) = 0$ is globally asymptotically stable for (4.1.3) and (4.1.10).*
- (3) *When $R > R_c$ the problem bifurcates from $((u, T), R) = (0, R_c)$ to an attractor \mathcal{A}_R , which is an $(m - 1)$ -dimensional homological sphere.*
- (4) *For any $(u, T) \in \mathcal{A}_R$, the velocity field u can be expressed as*

$$u = \sum_{k=1}^m \alpha_k e_k + o \left(\sum_{k=1}^m |\alpha_k| \right), \quad \lim_{R \rightarrow R_c} (\alpha_1, \dots, \alpha_m) = 0, \quad (4.1.13)$$

where e_k are the velocity fields of the first eigenvectors in E_0 .

- (5) *The attractor \mathcal{A}_R attracts $H \setminus \Gamma$, where Γ is the stable manifold of $(u, T) = 0$ with codimension m .*

Theorem 4.1.2 *If the first eigenvalue of (4.1.12) is simple, i.e., $\dim E_0 = 1$, then the following assertions hold true:*

1. *The bifurcated attractor \mathcal{A}_R of (4.1.3) with (4.1.10) consists of exactly two points $\phi_1, \phi_2 \in H_1$ given by*

$$\phi_{1,2} = \pm \alpha \left(\sqrt{R} - \sqrt{R_c} \right)^{1/2} \psi + o(|\sqrt{R} - \sqrt{R_c}|^{1/2}), \quad (4.1.14)$$

where $\alpha > 0$ is a constant, and $\psi \in E_0$ is the first eigenvector.

2. *The phase space H can be decomposed into two open sets:*

$$H = \overline{U}_1 + \overline{U}_2, \quad U_1 \cap U_2 = \emptyset,$$

such that for $i = 1, 2$ and $\varphi_i \in U_i$,

$$\lim_{t \rightarrow \infty} \|S_R(t)\varphi_0 - \varphi_i\| = 0 \quad \forall \varphi_0 = (u_0, T_0) \in U_i,$$

where $S_R(t)\varphi_0$ is the solution of (4.1.3) with (4.1.4) and (4.1.10).

A few remarks are now in order.

Remark 4.1.3 As mentioned earlier, both theorems hold true for the Boussinesq equations (4.1.3) with different combinations of boundary conditions given by (4.1.5)–(4.1.9).

Remark 4.1.4 As we shall see later, (4.1.13) and (4.1.14) are crucial for studying the topological structure of the Rayleigh–Bénard convection in the physical space.

Remark 4.1.5 Theorem 4.1.2 corresponds to the classical pitchfork bifurcation, and is much stronger in the sense that Assertion two here provides a global asymptotic stability of the bifurcated steady-state solutions, which does not appear to be derived by any other existing methods.

Proof of Theorem 4.1.1. We shall apply Theorem 2.2.11 to prove Theorem 4.1.1, and proceed in the following steps.

STEP 1. For simplicity, we assume the Prandtl number $\text{Pr} = 1$; otherwise we only have to consider the following form of (4.1.3), and the proof is the same:

$$\begin{aligned}\frac{\partial u}{\partial t} + (u \cdot \nabla)u + \nabla p - \Delta u - \sqrt{R}\theta k &= 0, \\ \frac{\partial \theta}{\partial t} + (u \cdot \nabla)\theta - \sqrt{R}\theta u_3 - \Delta \theta &= 0, \\ \operatorname{div} u &= 0,\end{aligned}$$

where $\theta = \sqrt{\text{Pr}}T$.

Let H and H_1 be the function space defined by (4.1.11). Then let $L_\lambda = -A + B_\lambda : H_1 \rightarrow H$, and $G : H_1 \rightarrow H$ be defined by

$$\begin{aligned}A\phi &= (-P(\Delta u), -\Delta T), & B_\lambda\phi &= \lambda(P[Tk], u_3), \\ G(\phi, \phi_1) &= (-P[(u \cdot \nabla)u_1], -(u \cdot \nabla)T_1), & G(\phi) &= G(\phi, \phi)\end{aligned}\quad (4.1.15)$$

for any $\phi = (u, T), \phi_1 = (u_1, T_1) \in H_1$. Here $\lambda = \sqrt{R}$, and $P : L^2(\Omega)^3 \rightarrow H$ is the Leray projection. These operators enjoy the following properties:

- (a) The linear operators A, B_λ and L_λ are all symmetric;
- (b) The nonlinear operator G is orthogonal, i.e.,

$$(G(\phi_1, \phi), \phi) = - \int_{\Omega} [(u_1 \cdot \nabla)u \cdot u + (u_1 \cdot \nabla)T \cdot T] dx = 0; \quad (4.1.16)$$

- (c) The conditions (2.1.2) and (2.1.3) hold true for the operators defined in (4.1.15).

Then the Boussinesq equations (4.1.3) can be rewritten in the following operator form

$$\frac{d\phi}{dt} = L_\lambda\phi + G(\phi). \quad (4.1.17)$$

STEP 2. VERIFICATION OF CONDITIONS (2.1.4) AND (2.1.5). Consider the eigenvalue problem

$$L_\lambda \phi = \beta(\lambda) \phi, \quad (4.1.18)$$

which is equivalent to

$$\begin{aligned} \Delta u - \nabla p + \lambda T k &= \beta(\lambda) u, \\ \Delta T + \lambda u_3 &= \beta(\lambda) T, \\ \operatorname{div} u &= 0, \end{aligned} \quad (4.1.19)$$

supplemented with (4.1.10). For each λ fixed, it is known that the eigenvalues β_k ($k = 1, 2, \dots$) of (4.1.18) are real numbers satisfying

$$\beta_1(\lambda) \geq \beta_2(\lambda) \geq \dots \geq \beta_k(\lambda) \geq \dots \rightarrow -\infty, \quad (4.1.20)$$

and the first eigenvalue $\beta_1(\lambda)$ of (4.1.18) and the first eigenvalue $\lambda_1 = \sqrt{R_c}$ of (4.1.12) with (4.1.10) have the relation

$$\beta_1(\lambda) \begin{cases} < 0 & \text{if } 0 \leq \lambda < \lambda_1, \\ = 0 & \text{if } \lambda = \lambda_1. \end{cases} \quad (4.1.21)$$

By Theorem 2.1 in Ma and Wang (2008d) (see also Ma and Wang (2005b) for a different proof), the first eigenvalue $\beta_1(\lambda)$ is differentiable on λ , and the derivative of $\beta_1(\lambda)$ at $\lambda = \lambda_1$ is given by

$$\frac{d}{d\lambda} \beta_1(\lambda_1) = \int_{\Omega} [T_1^2 + e_{13}^2] dx > 0, \quad (4.1.22)$$

where $\psi_1 = (e_1, T_1)$ with $e_1 = (e_{11}, e_{12}, e_{13})$ is the eigenvector of (4.1.18) corresponding to $\beta_1(\lambda_1) = 0$.

Thus it follows from (4.1.20)–(4.1.22) that the conditions (2.1.4) and (2.1.5) hold true for the eigenvalues of (4.1.18).

STEP 3. Finally, in order to use Theorem 2.2.11 to prove Theorem 4.1.1, we need to show that $(u, T) = 0$ is a globally asymptotically stable equilibrium point of (4.1.3) with (4.1.10) at the critical Rayleigh number $\lambda_1 = \sqrt{R_c}$. By Theorem 3.16 in Ma and Wang (2005b), it suffices to prove that the equations (4.1.3) with (4.1.10) have no invariant sets except the steady-state $(u, T) = 0$ in the first eigenspace E_0 of (4.1.12).

We know that the Boussinesq equations (4.1.3) with (4.1.10) have a bounded absorbing set in H . Assume that (4.1.3) with (4.1.10) have an invariant set $B \subset E_0$ with $B \neq \{0\}$ at $\lambda_1 = \sqrt{R_c}$. Then restricted in B , which contains eigenfunctions of (4.1.18) corresponding to $\beta_1(\lambda_1) = 0$, the equation (4.1.17) can be rewritten as

$$\frac{d\phi}{dt} = G(\phi) \quad \forall \phi = (u, T) \in B \subset E_0, \quad (4.1.23)$$

which is equivalent to

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla) u + \nabla p &= 0, \\ \frac{\partial T}{\partial t} + (u \cdot \nabla) T &= 0, \end{aligned} \quad (4.1.24)$$

for any $(u, T) \in B$. It is easy to see that for the solutions $(u, T) \in B$ of (4.1.24), $(\tilde{u}, \tilde{T}) = \alpha(u(\alpha t), T(\alpha t)) \in \alpha B \subset E_0$ are also solutions of (4.1.24). Namely for any real number $\alpha \in R$, the set $\alpha B \subset E_0$ is an invariant set of (4.1.24). Thus, we infer that (4.1.3) with (4.1.10) have an unbounded invariant set, which is a contradiction to the existence of bounded absorbing sets. Therefore, the invariant set B consists only of $(u, T) = 0$. The proof is complete. \square

Proof of Theorem 4.1.2. By Theorem 4.1.1, it suffices to prove that the steady-state equations of (4.1.3) bifurcate to exactly two singular points in H_1 as $R > R_c$. We proceed with the Lyapunov–Schmidt method.

Since the operator $L_\lambda : H_1 \rightarrow H$ defined by (4.1.15) is symmetric, all eigenvectors of L_λ

$$\{\psi_k \mid k = 1, 2, \dots\} \subset H_1 \quad (4.1.25)$$

constitute an orthogonal basis of H , and H can be decomposed into

$$\begin{aligned} H &= E_0^\lambda \bigoplus E_1^\lambda, \\ E_0^\lambda &= \{x\psi_1 \mid x \in \mathbb{R}^1, \psi_1 \text{ is the first eigenvector of } L_\lambda\}, \\ E_1^\lambda &= \left\{ \sum_{k=2}^{\infty} y_k \psi_k \mid y_k \in \mathbb{R}^1, \psi_k \text{ are as in (4.1.25)} \right\}. \end{aligned}$$

In addition, E_0^λ and E_1^λ are invariant subspaces of L_λ .

For any $\phi \in H$,

$$\phi = x\psi_1 + \sum_{k=2}^{\infty} y_k \psi_k, \quad x, y_k \in \mathbb{R}^1.$$

Then the equation $L_\lambda \phi + G(\phi) = 0$ can be decomposed into

$$\beta_1(\lambda)x + (G(\phi), \psi_1) = 0, \quad (4.1.26)$$

$$\beta_k(\lambda)y_k + (G(\phi), \psi_k) = 0 \quad \forall k = 2, 3, \dots, \quad (4.1.27)$$

where $\beta_k(\lambda)$ are the eigenvalues of L_λ satisfying

$$\beta_1(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_1, \\ = 0 & \text{if } \lambda = \lambda_1, \\ > 0 & \text{if } \lambda > \lambda_1, \end{cases} \quad (4.1.28)$$

$$\beta_k(\lambda_1) < 0 \quad \forall k \geq 2. \quad (4.1.29)$$

We infer from (4.1.26) and (4.1.16) that

$$\beta_1(\lambda)x - \sum_{k=2}^{\infty} xy_k(G(\psi_1), \psi_k) + O(\|y\|^2) = 0. \quad (4.1.30)$$

We obtain then from (4.1.27) and (4.1.16) that

$$y_k = -\frac{x^2}{\beta_k(\lambda)}(G(\psi_1), \psi_k) + O(|x|^2) \quad \forall k \geq 2.$$

Putting y_k into (4.1.30) we obtain

$$\beta_1(\lambda)x - ax^3 + o(|x|^3) = 0. \quad (4.1.31)$$

Here, by (4.1.29), we have

$$a = \sum_{k=2}^{\infty} \frac{-1}{\beta_k(\lambda)} |(G(\psi_1), \psi_k)|^2 > 0 \quad \text{near } \lambda = \lambda_1.$$

Equation (4.1.31) is the Lyapunov-Schmidt reduction of (4.1.18), which determines the bifurcated singular points of the Boussinesq equations (4.1.3) with (4.1.10). By (4.1.28) the equation (4.1.31) bifurcates from $(x, \lambda) = (0, \lambda_1)$ to exactly two solutions for $\lambda > \lambda_1$:

$$x_{1,2} = \pm(\beta_1(\lambda)/a)^{1/2} + o(|\beta_1|^{1/2}).$$

Therefore the problem (4.1.3) with (4.1.10) bifurcates from $(\phi, \lambda) = (0, \lambda_1)$ to exactly two singular points for $\lambda > \lambda_1$ ($R > R_c$)

$$\phi_1 = b\beta_1^{1/2}(\lambda)\psi_1 + o(|\beta_1|^{1/2}), \quad \phi_2 = -b\beta_1^{1/2}(\lambda)\psi_1 + o(|\beta_1|^{1/2}), \quad (4.1.32)$$

where $b = a^{-1/2} > 0$. By (4.1.32), to prove (4.1.14), it suffices to verify that

$$\beta_1(\lambda) = c(\lambda - \lambda_1) + o(|\lambda - \lambda_1|), \quad (4.1.33)$$

where $\lambda = \sqrt{R}$ and $\lambda_1 = \sqrt{R_c}$. The formula (4.1.33) follows from (4.1.22). \square

4.1.4 Topological Structure and Pattern Formation

As seen in the expressions (4.1.13) and (4.1.14), the topological structure of the first eigenvectors of the linearized equations plays an important role for studying the onset of the Rayleigh–Bénard convection, which essentially determines the thermal convection structure in the physical space.

We now examine in detail the first eigenvectors of the linearized problem with the free boundary conditions in a rectangular domain: $\Omega = (0, L_1) \times (0, L_2) \times (0, 1)$, and consider the eigenvalue problem (4.1.12) with the free boundary condition:

$$\begin{aligned} u_n &= 0, & \frac{\partial u_\tau}{\partial n} &= 0 \quad \text{on } \partial\Omega, \\ T &= 0 & \text{at } x_3 &= 0, 1, \\ \frac{\partial T}{\partial n} &= 0 & \text{at } x_1 &= 0, L_1 \text{ or } x_2 = 0, L_2. \end{aligned} \quad (4.1.34)$$

Solutions of the Eigenvalue Problem For the eigenvalue problem (4.1.12) with (4.1.34), we take the separation of variables as follows:

$$(u_1, u_2, u_3, T) = \left(\frac{1}{a^2} \nabla_2 f(x_1, x_2) \frac{dH(x_3)}{dx_3}, fH(x_3), f\theta(x_3) \right), \quad (4.1.35)$$

where $a^2 > 0$ is an arbitrary constant. It follows from (4.1.12) and (4.1.34) that the functions $f = f(x_1, x_2)$, $H = H(x_3)$, $\theta = \theta(x_3)$ satisfy

$$\begin{cases} -\Delta_2 f = a^2 f, \\ \frac{\partial f}{\partial x_1} \Big|_{x_1=0, L_1} = 0, \quad \frac{\partial f}{\partial x_2} \Big|_{x_2=0, L_2} = 0, \end{cases} \quad (4.1.36)$$

$$\begin{cases} \left(\frac{d^2}{dx_3^2} - a^2 \right)^2 H = a^2 \lambda \theta, \\ \left(\frac{d^2}{dx_3^2} - a^2 \right) \theta = -\lambda H, \\ (\theta(x_3), H(x_3), H''(x_3)) = 0 \quad \text{at } x_3 = 0, 1. \end{cases} \quad (4.1.37)$$

It is clear that the solutions of (4.1.36) are given by

$$\begin{aligned} f(x_1, x_2) &= \cos a_1 x_1 \cos a_2 x_2, \\ (a_1, a_2) &= (k_1 \pi / L_1, k_2 \pi / L_2), \quad a^2 = a_1^2 + a_2^2 \quad \forall k_1, k_2 = 0, 1, \dots. \end{aligned} \quad (4.1.38)$$

For each a^2 , the first eigenvalue $\lambda_0(a)$ and eigenvectors of (4.1.37) are given by

$$\lambda_0(a) = \frac{(\pi^2 + a^2)^{3/2}}{a}, \quad (H, \theta) = \left(\sin \pi x_3, \frac{1}{a} \sqrt{\pi^2 + a^2} \sin \pi x_3 \right). \quad (4.1.39)$$

The first eigenvalue $\lambda_1 = \sqrt{R_c}$ of (4.1.12) with (4.1.34) is the minimum of $\lambda_0(a)$:

$$R_c = \min_{a^2=a_1^2+a_2^2} \lambda_0^2(a) = \min_{k_1, k_2 \in \mathbb{Z}} \left[\pi^4 \left(1 + \frac{k_1^2}{L_1^2} + \frac{k_2^2}{L_2^2} \right)^3 / \left(\frac{k_1^2}{L_1^2} + \frac{k_2^2}{L_2^2} \right) \right]. \quad (4.1.40)$$

Thus the first eigenvectors of (4.1.12) with (4.1.34) are immediately derived from (4.1.35), (4.1.38), and (4.1.39):

$$(u, T) = \left(\frac{\pi}{a^2} \nabla_2 f \cdot \cos \pi x_3, f \sin \pi x_3, \frac{\sqrt{\pi^2 + a^2}}{a} f \sin \pi x_3 \right), \quad (4.1.41)$$

where $f = \cos a_1 x_1 \cos a_2 x_2$, and $a^2 = a_1^2 + a_2^2$ achieves the minimum in (4.1.40).

By Theorems 4.1.1 and 4.1.2, the topological structure of the bifurcated solutions of the Bénard problem (4.1.3) with (4.1.34) in a rectangular region is determined

by that of (4.1.41), and consequently depends, by (4.1.40), on the horizontal length scales L_1 and L_2 . Namely the pattern of the Rayleigh–Bénard convection depends on the size and form of the containers of the fluid. In fact, the convection pattern in a rectangular region is completely determined by the wave numbers $(k_1, k_2) \in \mathbb{Z}^2$ satisfying (4.1.40), and we now address this issue.

Horizontal Scales and Wave Numbers By (4.1.40), the wave numbers $(k_1, k_2) \in \mathbb{Z}^2$ of convection depend on the horizontal length scales L_1 and L_2 . Let L_1 and L_2 have the (horizontal) aspect ratio α given by

$$L = L_1 = \alpha L_2, \quad (4.1.42)$$

where L is called the horizontal scale, and $\alpha > 0$ is the ratio constant. Thus the critical Rayleigh number (4.1.40) can be rewritten as

$$R_c = \min_{k_1, k_2 \in \mathbb{Z}} \frac{\pi^4(L^2 + k_1^2 + \alpha^2 k_2^2)^3}{L^4(k_1^2 + \alpha^2 k_2^2)}. \quad (4.1.43)$$

Let

$$b = k_1^2 + \alpha^2 k_2^2 \quad (4.1.44)$$

achieve the minimum in (4.1.43). Then

$$\begin{aligned} \frac{dR_c}{db} &= \frac{\pi^4}{L^4 b^2} (L^2 + b)^2 (2b - L^2) = 0, \\ \frac{dR_c}{dL} &= \frac{2\pi^4}{L^5 b} (L^2 + b)^2 (L^2 - 2b) = 0. \end{aligned}$$

Hence we derive the formula

$$b = L^2/2. \quad (4.1.45)$$

Namely, the module of wave numbers is proportional to the square of the horizontal scale. Notice that the wave numbers k_1 and k_2 are integers. When the aspect ratio $\alpha > 0$ is fixed, for each given $(k_1, k_2) \in \mathbb{Z}^2$, the corresponding horizontal scale L , with which (k_1, k_2) satisfies (4.1.43), is in an interval.

To see this, let the integer pairs (k_{i1}, k_{i2}) ($i = 1, 2, 3$) be given so that the corresponding $b_i = k_{i1}^2 + \alpha^2 k_{i2}^2$ satisfy

$$b_1 < b_2 < b_3, \quad (4.1.46)$$

and there are no integers $(k_1, k_2) \in \mathbb{Z}^2$ such that $b = k_1^2 + \alpha^2 k_2^2$ is between b_j and b_{j+1} , i.e., $b \notin (b_j, b_{j+1})$ ($j = 1, 2$) for all $(k_1, k_2) \in \mathbb{Z}^2$.

The following theorem is a direct consequence of (4.1.43)–(4.1.46).

Theorem 4.1.6 *Let l_1 and l_2 satisfy*

$$b_{i+1}(l_i^2 + b_i)^3 = b_i(l_i^2 + b_{i+1})^3 \quad \text{for } i = 1, 2. \quad (4.1.47)$$

If the horizontal scale L satisfies

$$l_1 < L < l_2, \quad (4.1.48)$$

then the wave numbers of the Rayleigh–Bénard convection take the integer pair of (k_{21}, k_{22}) .

This theorem provides a precise relationship between the horizontal scale L of a rectangular container and the wave numbers (k_1, k_2) of thermal convection. In the following, we discuss the convection pattern using Theorem 4.1.6.

Single Roll Structure We note that the horizontal scale $L = L_1$ also represents the aspect ratio between the horizontal scale and the vertical scale of the rectangular container for the case of $\alpha \geq 1$.

We see from (4.1.44) and (4.1.45) that for $\alpha > 1$, if the aspect ratio L is small, then the wave numbers are given by $(k_1, k_2) = (1, 0)$. More precisely, by Theorem 4.1.6, when

$$0 < L^2 < \frac{1 + \alpha^2 - (1 + \alpha^2)^{1/3}}{(1 + \alpha^2)^{1/3} - 1} \quad \text{for } \alpha > 1,$$

the convection pattern is a single roll structure; see Fig. 4.2a, b.

Consider the case where $\alpha = 1$, i.e.,

$$L = L_1 = L_2 \quad \text{and} \quad 0 < L^2 < \frac{2 - 2^{1/3}}{2^{1/3} - 1} \simeq 3. \quad (4.1.49)$$

In this case, the wave numbers (k_1, k_2) are given by

$$(k_1, k_2) = (1, 0) \quad \text{and} \quad (0, 1).$$

Then the multiplicity of the first eigenvalue is two, and the first eigenspace E_0 of (4.1.12) with (4.1.34) is given by

$$E_0 = \{\alpha_1 \psi_1 + \alpha_2 \psi_2 \mid \alpha_1, \alpha_2 \in \mathbb{R}^1\},$$

where for $i = 1, 2$, $\psi_i = (e_i, T_i)$ are given by

$$\begin{aligned} (e_1, T_1) &= \left(-L \sin \frac{\pi x_1}{L} \cos \pi x_3, 0, \cos \frac{\pi x_1}{L} \sin \pi x_3, \sqrt{L^2 + 1} \cos \frac{\pi x_1}{L} \sin \pi x_3 \right), \\ (e_2, T_2) &= \left(0, -L \sin \frac{\pi x_2}{L} \cos \pi x_3, \cos \frac{\pi x_2}{L} \sin \pi x_3, \sqrt{L^2 + 1} \cos \frac{\pi x_2}{L} \sin \pi x_3 \right). \end{aligned}$$

When $\alpha_1^2 + \alpha_2^2 = 1$ and $\alpha_1 \neq \alpha_2$, the topological structure of $u = \alpha_1 e_1 + \alpha_2 e_2$ is equivalent to the roll structure as shown in Fig. 4.2a, b.

In addition, the critical Rayleigh number is

$$R_c = \pi^4 (1 + L^2)^3 / L^4. \quad (4.1.50)$$

Remark 4.1.7 The roll structure of $u = \alpha_1 e_1 + \alpha_2 e_2 (\alpha_1 \neq \alpha_2)$ has a certain stability, although it is not the structural stability, i.e., under a perturbation the roll trait remains

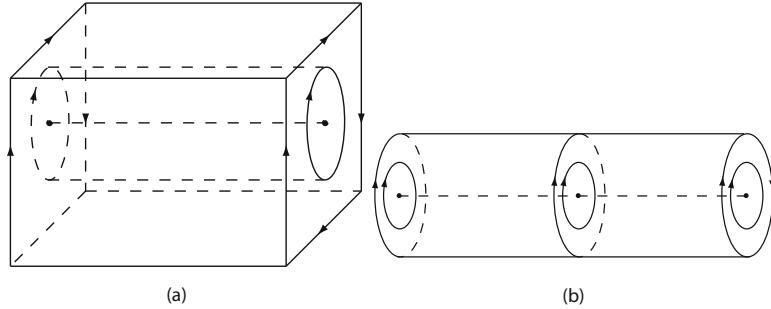


Fig. 4.2 Roll structure: (a) an elevation of the flow, and (b) flow structure in the interior of the cube

invariant. Hence, by Theorem 4.1.1, the solutions $u = \alpha_1 e_1 + \alpha_2 e_2 + o(|\alpha_1| + |\alpha_2|)$ with $\alpha_1 \neq \alpha_2$ in the bifurcated attractor \mathcal{A}_R have also the single roll structure.

Remark 4.1.8 Since the eigenspace E_0 has dimension two, the classical bifurcation theories cannot be applied to this case for the Rayleigh–Bénard convection. In later subsection, we shall continue to discuss this case by using the construction of center manifold functions introduced in Sect. 1.2, and obtain the precise structure of the bifurcated attractor \mathcal{A}_R .

Remark 4.1.9 The formula (4.1.50) shows that the critical Rayleigh number R_c depends on the aspect ratio; see also Remark 4.1.10 later.

Double Roll Structure We first consider the case

$$L_1 = L_2 = L, \quad \frac{2 - 2^{1/3}}{2^{1/3} - 1} < L^2 < \frac{2(2 - 2^{1/3})}{2^{1/3} - 1}. \quad (4.1.51)$$

By Theorem 4.1.6, the wave numbers are $(k_1, k_2) = (1, 1)$, the eigenvalue is simple, the critical Rayleigh number is

$$R_c = \frac{\pi^4 (2 + L^4)^3}{2L^4},$$

and the eigenspace is

$$E_0 = \{\beta \psi_1 \mid \beta \in \mathbb{R}^1, \psi_1 = (e_1, T_1)\}, \quad (4.1.52)$$

where

$$(e_1, T_1) = \left(\frac{L^2}{2\pi} \nabla_2 f \cos \pi x_3, f \sin \pi x_3, \sqrt{\frac{L^2}{2} + 1} f \sin \pi x_3 \right), \quad f = \cos \frac{\pi x_1}{L} \cos \frac{\pi x_2}{L}.$$

The topological structure of (4.1.52) consists of two rolls with reverse orientations, and the axes of both rolls are given by (see Fig. 4.3):

$$\begin{aligned} & \left\{ \left(\frac{L}{2}, x_2, \frac{1}{2} \right) \mid 0 \leq x_2 \leq \frac{L}{2} \right\} \cup \left\{ \left(x_1, \frac{L}{2}, \frac{1}{2} \right) \mid 0 \leq x_1 \leq \frac{1}{2} \right\}, \\ & \left\{ \left(x_1, \frac{L}{2}, \frac{1}{2} \right) \mid \frac{L}{2} \leq x_1 \leq L \right\} \cup \left\{ \left(\frac{L}{2}, x_2, \frac{1}{2} \right) \mid \frac{L}{2} \leq x_2 \leq L \right\}. \end{aligned}$$

The structure given in Fig. 4.3 is unstable, under a perturbation it will lead to the structures as shown in Fig. 4.4.

Fig. 4.3 Topological structure of (4.1.52) with (unstable) rolls

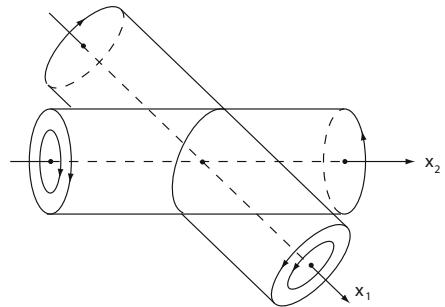
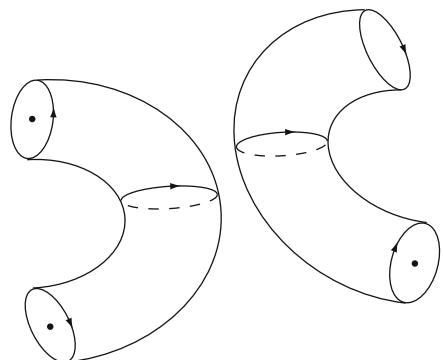


Fig. 4.4 Two stable rolls approximating the structure of (4.1.52)



We now consider the case

$$L_1 = L_2 = L, \quad \frac{2(2 - 2^{1/3})}{2^{1/3} - 1} \leq L^2 \leq \frac{5 - 4(1 + \frac{1}{4})^{1/3}}{(1 + \frac{1}{4})^{1/3} - 1}. \quad (4.1.53)$$

By Theorem 4.1.6, in this case the wave numbers are

$$(k_1, k_2) = (2, 0) \quad \text{and} \quad (0, 2),$$

and the first eigenspace E_0 is two dimension given by

$$E_0 = \{\alpha_1\psi_1 + \alpha_2\psi_2 \mid \alpha_1, \alpha_2 \in \mathbb{R}^1, \psi_i = (e_i, T_i), i = 1, 2\},$$

$$(e_1, T_1) = \left(-\frac{L}{2} \sin \frac{2\pi x_1}{L} \cos \pi x_3, 0, \cos \frac{2\pi x_1}{L} \sin \pi x_3, \frac{\sqrt{4+L^2}}{2} \cos \frac{2\pi x_1}{L} \sin \pi x_3 \right),$$

$$(e_2, T_2) = \left(0, -\frac{1}{2} \sin^2 \frac{2\pi x_1}{L} \cos \pi x_3, \cos \frac{2\pi x_1}{L} \sin \pi x_3, \frac{1}{\sqrt{4+L^2}} \cos \frac{2\pi x_2}{L} \sin \pi x_3 \right),$$

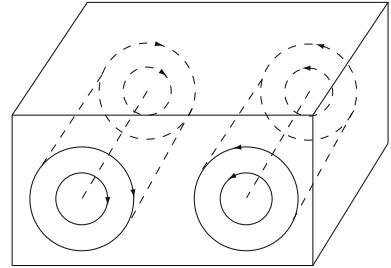
and the critical Rayleigh number is

$$R_c = \frac{\pi^4(4+L^4)^2}{4L^4}.$$

The topological structure of $u = \alpha_1 e_1 + \alpha_2 e_2$ with $\alpha_1 \neq \alpha_2$ is topologically equivalent to the double roll structure as shown in Fig. 4.5.

Remark 4.1.10 These cases of (4.1.49), (4.1.51), and (4.1.53) are consistent with experiments. As we boil water in a container, when the ratio between the diameter and the height is smaller than $\sqrt{3}$ (Condition (4.1.43)), the convection of heating water takes the single roll pattern as illustrated in Fig. 4.2. When the ratio is between $\sqrt{3}$ and $\sqrt{6}$ (Condition (4.1.51)), the convection takes the double roll structure as shown in Fig. 4.4. If the ratio is between $\sqrt{6}$ and $\sqrt{8}$ (Condition (4.1.53)), then the convection takes the pattern given in Fig. 4.5.

Fig. 4.5 Double rolls structure



Honeycomb Structure As in the Bénard experiments, if the horizontal length scales L_1 and L_2 are sufficiently large, then it is reasonable to consider the periodic boundary condition in the (x_1, x_2) -plane as follows:

$$(u, T)(x_1 + L_1, x_2 + L_2, x_3) = (u, T)(x),$$

$$T = 0, u_3 = 0, \frac{\partial(u_1, u_2)}{\partial x_3} = 0 \quad \text{at } x_3 = 0, 1. \quad (4.1.54)$$

In this case, the critical Rayleigh number R_c takes the minimum of $\lambda_0^2(a)$ defined by (4.1.40), which is

$$R_c = \min_a \lambda_0^2(a) = 657.5, \quad (4.1.55)$$

where $a_c = \frac{\pi}{\sqrt{2}}$ is the critical wave number, representing the size of the cells in the Bénard convection. Hence the number $r = 2^{3/2}\pi h/a_c$ can be regarded as the radius of the cells. The first eigenspace E_0 of (4.1.12) with (4.1.54) is generated by eigenvectors of the following type:

$$(e, T) = \left(\frac{\pi}{a_c^2} \nabla_2 f \cos \pi x_3, f(x_1, x_2) \sin \pi x_3, \frac{\sqrt{\pi^2 + a_c^2}}{a_c} f(x_1, x_2) \sin \pi x_3 \right), \quad (4.1.56)$$

where $f(x_1, x_2)$ is any one of the following functions

$$\begin{aligned} \cos \frac{2\pi k_1 x_1}{L_1} \cos \frac{2\pi k_2 x_2}{L_2}, & \quad \cos \frac{2\pi k_1 x_1}{L_1} \sin \frac{2\pi k_2 x_2}{L_2}, \\ \sin \frac{2\pi k_1 x_1}{L_1} \cos \frac{2\pi k_2 x_2}{L_2}, & \quad \sin \frac{2\pi k_1 x_1}{L_1} \sin \frac{2\pi k_2 x_2}{L_2}, \end{aligned}$$

with the periods L_1 and L_2 satisfying

$$4\pi^2 \left(\frac{k_1^2}{L_1^2} + \frac{k_2^2}{L_2^2} \right) = a_c^2 = \frac{\pi^2}{2} \quad \text{for } k_1, k_2 \in \mathbb{Z};$$

namely

$$\frac{k_1^2}{L_1^2} + \frac{k_2^2}{L_2^2} = \frac{1}{8} \quad \text{for } k_1, k_2 \in \mathbb{Z}. \quad (4.1.57)$$

It is clear that the dimension of the first eigenspace E_0 is determined by the given periods L_1 and L_2 satisfying (4.1.57), and

$$\dim E_0 = \text{even} \geq 4.$$

Various solutions having the honeycomb structure are found in E_0 . For convenience, we list two examples as follows.

SQUARE CELLS: The solution of (4.1.12) with (4.1.54) given by

$$(e_1, T_1) = \left(\frac{2}{\pi} \nabla_2 f \cos \pi x_3, f \sin \pi x_3, \sqrt{3} f \sin \pi x_3 \right), \quad f = \cos \frac{2\pi x_1}{L_1} \cos \frac{2\pi x_2}{L_2},$$

is a rectangular cell with sides of lengths L_1 and L_2 , where $\frac{1}{L_1^2} + \frac{1}{L_2^2} = \frac{1}{8}$.

HEXAGONAL CELLS: A solution in E_0 having the hexagonal pattern was found by Christopherson in 1940, and is given by Chandrasekhar (1981):

$$\begin{aligned}\psi &= (e, T) = (u_1, u_2, u_3, T), \\ u_1 &= -\frac{2}{\sqrt{6}} \sin \frac{3\pi x_1}{2\sqrt{6}} \cos \frac{\pi x_2}{2\sqrt{2}} \cos \pi x_3, \\ u_2 &= -\frac{2}{3\sqrt{2}} \left(\cos \frac{3\pi x_1}{2\sqrt{6}} + 2 \cos \frac{\pi x_2}{2\sqrt{2}} \right) \sin \frac{\pi x_2}{2\sqrt{2}} \cos \pi x_3, \\ u_3 &= \frac{1}{3} \left(2 \cos \frac{3\pi x_1}{2\sqrt{6}} \cos \frac{\pi x_2}{2\sqrt{2}} + \cos \frac{\pi x_2}{\sqrt{2}} \right) \sin \pi x_3, \\ T &= \frac{1}{\sqrt{3}} \left(2 \cos \frac{3\pi x_1}{2\sqrt{6}} \cos \frac{\pi x_2}{2\sqrt{2}} + \cos \frac{\pi x_2}{\sqrt{2}} \right) \sin \pi x_3.\end{aligned}$$

This solution is for the case where the periods are taken as $L_2 = \sqrt{3}L_1$ and $L_1 = 4\sqrt{6}/3$, and the wave numbers are $(k_1, k_2) = (1, 1)$ and $(0, 1)$.

In summary, for any fixed periods L_1 and L_2 , the first eigenspace E_0 of (4.1.12) with (4.1.54) has its dimension determined by

$$\dim E_0 = \begin{cases} 6 & \text{if } L_2 = \sqrt{k^2 - 1} L_1, \quad k = 2, 3, \dots, \\ 4 & \text{otherwise.} \end{cases}$$

The attractor bifurcation theorem, Theorem 4.1.1, provides general dynamical properties for the Rayleigh–Bénard convection. However by using the construction of center manifold functions, for the problem (4.1.3) with (4.1.54) we can prove the following more refined results:

- When the Rayleigh number $R_c < R < R_c + \varepsilon$ (R_c as in (4.1.55)) for some $\varepsilon > 0$, the Bénard problem bifurcates from the trivial state $\phi = 0$ to an attractor \mathcal{A}_R homeomorphic to a sphere, i.e.,

$$\mathcal{A}_R = \begin{cases} S^5 & \text{if } L_2 = \sqrt{k^2 - 1} L_1, \quad k = 2, 3, \dots, \\ S^3 & \text{otherwise.} \end{cases}$$

- The bifurcated attractor \mathcal{A}_R consists of singular points and their stable, unstable manifold and the singular points constitute finite number of tori. We remark here that with the bifurcation with symmetry theory, one can only prove that there are finite number of 2-dimensional tori of steady states, which can never cover all the solutions in a higher dimensional attractor \mathcal{A}_R . In fact, there are many transient states, occupying the same dimensional space as \mathcal{A}_R , which are physically important. These transient states cannot be derived by any of the classical bifurcation methods including the bifurcation with symmetry theory.
- The minimal attractors in \mathcal{A}_R consist of singular points, and are small perturbations of the eigenvectors in E_0 , which have the honeycomb structure.
- \mathcal{A}_R attracts $H \setminus \Gamma$, where Γ is the stable manifold of $\phi = 0$. Consequently, for any initial value $\phi_0 \in H \setminus \Gamma$, the solution $S_R(t)\phi_0$ of (4.1.3) with (4.1.54) converges to the minimal attractors in \mathcal{A}_R , which approximates the honeycomb structure.

4.1.5 Asymptotic Structure of Solutions for the Bénard problem

In this subsection, we study the structural stability of bifurcated solutions and asymptotic structure of solutions for the 2D Bénard problem. It is easy to see that both Theorems 4.1.1 and 4.1.2 hold true for the 2D Boussinesq equation with any combination of boundary conditions as described in (4.1.5)–(4.1.9). Hence, we focus on justifying the roll pattern formation generated by the thermal convection as illustrated in Fig. 4.1, and on theoretically verifying the phase transition pattern consistent with the observations in physical experiments.

Technically speaking, we see from (4.1.43) that if $L_2/L_1 = \alpha^{-1}$ is small, the wave number $k_2 = 0$. In this case the 3D Bénard problem is reduced to the two-dimensional one. Furthermore, due to the symmetry on the x_1x_2 -plane of the honeycomb structure of the Rayleigh–Bénard convection, from the viewpoint of a cross section, the 3D convection can be well understood by the two-dimensional version.

Let $\Omega = (0, L) \times (0, 1)$ with coordinate system $x = (x_1, x_3)$. Thus, the Boussinesq equations take the form

$$\begin{aligned} \frac{1}{\text{Pr}} \left[\frac{\partial u_1}{\partial t} + (u \cdot \nabla) u_1 + \frac{\partial p}{\partial x_1} \right] - \Delta u_1 &= 0, \\ \frac{1}{\text{Pr}} \left[\frac{\partial u_3}{\partial t} + (u \cdot \nabla) u_3 + \frac{\partial p}{\partial x_3} \right] - \Delta u_3 - \sqrt{RT} &= 0, \\ \frac{\partial T}{\partial t} + (u \cdot \nabla) T - \sqrt{R} u_3 - \Delta T &= 0, \\ \frac{\partial u_1}{\partial x_1} + \frac{\partial u_3}{\partial x_3} &= 0, \end{aligned} \quad (4.1.58)$$

where $u = (u_1, u_3)$ is the velocity field in the $x = (x_1, x_2)$ coordinate system, and $\nabla = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2} \right)$. For simplicity, we consider here only the free boundary conditions:

$$\begin{aligned} u_1 &= 0, & \frac{\partial u_3}{\partial x_1} &= 0, & \frac{\partial T}{\partial x_1} &= 0 & \text{at } x_1 = 0, L, \\ u_3 &= 0, & \frac{\partial u_1}{\partial x_3} &= 0, & T &= 0 & \text{at } x_3 = 0, 1. \end{aligned} \quad (4.1.59)$$

The initial value condition is

$$(u, T) = \phi_0 = (u_0, T_0) \quad \text{at } t = 0. \quad (4.1.60)$$

In this case, the function space H defined by (4.1.11) is replaced by

$$H = \{(u, T) \in L^2(\Omega)^3 \mid \text{div}u = 0, u_3|_{x_3=0,1} = 0, u_1|_{x_1=0,L} = 0\}.$$

By (4.1.41) and (4.1.43), for the problem (4.1.58)–(4.1.60), the wave number k and the critical Rayleigh number are

$$k \simeq L/\sqrt{2} \quad \text{for } L \text{ large,}$$

$$R_c = \pi^4(k^2 + L^2)^3/(L^4 k^2),$$

and the first eigenspace E_0 is one dimensional, given by

$$E_0 = \text{span}\{\psi_1 = (e_1, T_1)\},$$

$$(e_1, T_1) = \left(-\frac{L}{k} \sin \frac{k\pi x_1}{L} \cos \pi x_3, \cos \frac{k\pi x_1}{L} \sin \pi x_3, \frac{\sqrt{L^2 + k^2}}{k} \cos \frac{k\pi x_1}{L} \sin \pi x_3 \right). \quad (4.1.61)$$

The topological structure of e_1 in (4.1.61) consists of k vortices as shown in Fig. 4.6a, b.

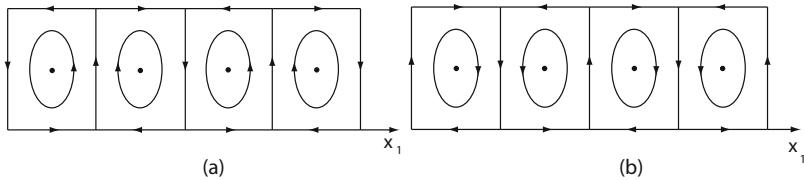


Fig. 4.6 Rolls in (a) and (b) have reverse orientations

The following theorem shows that the asymptotic structure of solutions of the 2D Bénard problem (4.1.58)–(4.1.60) is topologically equivalent to the roll structure as shown in Fig. 4.6a or b, which is consistent with the experiment results.

Theorem 4.1.11 *There is a $\delta > 0$ such that if the Rayleigh number satisfies $R_c < R < R_c + \delta$, H can be decomposed into two open sets $H = \overline{U}_1 + \overline{U}_2$ with $U_1 \cap U_2 = \emptyset$ such that the following assertions hold:*

- (1) *For any initial value $\phi_0 = (u_0, T_0) \in U_i$ ($i = 1, 2$), there exists a time $t_0 > 0$ such that when $t > t_0$, the velocity field $u(t, \phi_0)$ of the solution $S_R(t)\phi_0$ of (4.1.58)–(4.1.60) is topologically equivalent to the structure as shown in either Fig. 4.6a or b, depending on if $\phi_0 \in U_1$ or $\phi_0 \in U_2$.*
- (2) *When $t > t_0$, the velocity $u(t, \phi_0)$ has k (k is the wave number) vortices. In particular, if L is large, then $k \simeq L/\sqrt{2}$.*

Proof. It is easy to check that the velocity fields $\pm e_1$ as in (4.1.61) are of the topological structure as shown in Fig. 4.6a, b, and have k number of vortices. In particular, by Theorem 4.1.6, if $L \gg 1$ is large, from the formula

$$(k+1)^2(L^2 + k^2)^3 = k^2(L^2 + (k+1)^2)^3$$

one can obtain that the wave number $k \simeq L/\sqrt{2}$.

Direct computation implies that the singular points of e_1 are nondegenerate. Hence based on the structural stability theorem, Theorem 2.2.9 in Ma and Wang (2005d), the velocity fields $\pm e_1$ are structurally stable.

In addition, by Theorem 4.1.2 and expressions (4.1.14), there exists a $\delta > 0$ such that if $R_c < R < R_c + \delta$, then both velocity fields u_1 and u_2 of the two bifurcated solutions ϕ_1 and ϕ_2 are topologically equivalent to e_1 and $-e_1$ respectively. Furthermore, we know that for any initial value $\phi_0 = (u_0, T_0) \in H$ there is a time $\tau > 0$ such that the solution $S_R(t)\phi_0$ of (4.1.58)–(4.1.60) is C^∞ for $t > \tau$. Therefore, this theorem immediately follows from Theorem 4.1.2 and the following lemma on H^k -uniform boundedness of solutions of (4.1.58)–(4.1.60). Thus, the proof is complete. \square

The following lemma is necessary in the proof of the asymptotically structural stability of solutions for the Bénard problem and the Taylor problem discussed in the next section. Also, we refer the interested readers to (Ma and Wang, 2006) for its proof. Let

$$H_B^k = \{(u, T) \in H^k(\Omega)^3 \cap H \mid (u, T) \text{ satisfy (4.1.59)}\}.$$

Lemma 4.1.12 *Let $k \geq 1$ be any integer. If $\phi_0 = (u_0, T_0) \in H_B^k$, then there exists a number $C > 0$ depending on ϕ_0 such that*

$$\|S_R(t)\phi_0\|_{H^k} \leq C \quad \forall t \geq 0,$$

where $S_R(t)\phi_0 = (u(t, \phi_0), T(t, \phi_0))$ is the solution of (4.1.58)–(4.1.60).

Remark 4.1.13 Lemma 4.1.12 and Theorem 4.1.11 hold true as well for (4.1.88) with the following boundary conditions:

$$\begin{aligned} u &= T = 0 && \text{at } x_3 = 0, 1, \\ u_1 &= 0, \quad \frac{\partial u_3}{\partial x_1} = 0, \quad \frac{\partial T}{\partial x_1} = 0 && \text{at } x_1 = 0, L. \end{aligned}$$

4.1.6 Structure of Bifurcated Attractors

As an example we discuss the topological structure of the bifurcated attractor for the Bénard problem (4.1.3) with the free boundary condition (4.1.34) in a rectangular region $\Omega = (0, L)^2 \times (0, 1)$. For simplicity, we consider the case where

$$0 < L^2 < \frac{2 - 2^{1/3}}{2^{1/3} - 1}. \quad (4.1.62)$$

We proceed by first considering the eigenvalues and eigenvectors of the linearized equations of (4.1.3):

$$\begin{aligned} \Delta u - \nabla p + \sqrt{R}kT &= \beta(R)u, \\ \Delta T + \sqrt{R}u_3 &= \beta(R)T, \\ \operatorname{div} u &= 0, \end{aligned} \quad (4.1.63)$$

supplemented with the boundary condition (4.1.34).

For $\psi = (u_1, u_2, u_3, T)$, we take the separation of variables as follows:

$$\psi = \left(\nabla_2 f \ h'(x_3), a^2 f h(x_3), f \ \theta(x_3) \right), \quad f = \cos a_1 x_1 \cos a_2 x_2,$$

where as before $a_i = k_i \pi / L$ for $i = 1, 2$, and $a^2 = a_1^2 + a_2^2$.

We infer from the horizontal velocity equations of (4.1.63) that

$$p = \cos a_1 x_1 \cos a_2 x_2 \left(\left(\frac{d^2}{dx_3^2} - a^2 \right) h' - \beta h' \right).$$

Inserting p into the third component of the velocity equation and the temperature in (4.1.63), we then derive that

$$\begin{aligned} \left(\frac{d^2}{dx_3^2} - a^2 \right)^2 h - \sqrt{R} \theta &= \beta \left(\frac{d^2}{dx_3^2} - a^2 \right) h, \\ \left(\frac{d^2}{dx_3^2} - a^2 \right) \theta + \sqrt{R} a^2 h &= \beta \theta, \end{aligned} \quad (4.1.64)$$

supplemented with the following boundary conditions

$$(h, h'', \theta) = 0 \quad \text{at} \quad x_3 = 0, 1. \quad (4.1.65)$$

For given Rayleigh number $R > 0$ and a pair of integers k_1 and k_2 , the eigenvectors of (4.1.64) and (4.1.65) are given by

$$(h, \theta) = (\sin n\pi x_3, \alpha_{k_1 k_2}^{nm} \sin m\pi x_3) \quad n \geq 1, \quad m \geq 0, \quad (4.1.66)$$

and the associated eigenvalue $\beta_{k_1 k_2}^{nm}(R)$ and the coefficient $\alpha_{k_1 k_2}^{nm}$ satisfy

$$\begin{aligned} (n^2 \pi^2 + a^2)^2 - \sqrt{R} a^2 \alpha_{k_1 k_2}^{nm} &= -\beta_{k_1 k_2}^{nm} (n^2 \pi^2 + a^2), \\ -(m^2 \pi^2 + a^2) \alpha_{k_1 k_2}^{nm} + \sqrt{R} &= \beta_{k_1 k_2}^{nm} \alpha_{k_1 k_2}^{nm}, \end{aligned}$$

for $a^2 = (k_1^2 \pi^2 + k_2^2 \pi^2) / L^2 \neq 0$. Consequently,

$$\begin{aligned} \alpha_{k_1 k_2}^{nm} &= \frac{1}{a^2 \sqrt{R}} \left[(n^2 \pi^2 + a^2)^2 - (m^2 \pi^2 + a^2)(n^2 \pi^2 + a^2) + \frac{\sqrt{R}(n^2 \pi^2 + a^2)}{\alpha_{k_1 k_2}^{nm}} \right], \\ \beta_{k_1 k_2}^{nm} &= -(m^2 \pi^2 + a^2) + \sqrt{R} / \alpha_{k_1 k_2}^{nm}, \quad \forall k_1^2 + k_2^2 \geq 1, m \geq 1, n \geq 1. \end{aligned} \quad (4.1.67)$$

As $a^2 = 0$, i.e., $k_1 = k_2 = 0$, the eigenvalues are

$$\beta_0^m = -m^2 \pi^2, \quad m = 1, 2, \dots, \quad n = 0. \quad (4.1.68)$$

Then all eigenvalues of (4.1.63) with (4.1.34) are given by (4.1.67) and (4.1.68), and by (4.1.66) the eigenvectors are

$$\begin{cases} \psi_{k_1 k_2 nm} = \left(\nabla_2 f_{k_1 k_2} \cos n\pi x_3, a^2 f_{k_1 k_2} \sin n\pi x_3, a_{k_1 k_2}^{nm} f_{k_1 k_2} \sin m\pi x_3 \right), \\ f_{k_1 k_2} = \cos \frac{k_1 \pi x_2}{L} \cos \frac{k_2 \pi x_2}{L}, \end{cases} \quad (4.1.69)$$

$$\psi_{0m} = (0, 0, 0, \sin m\pi x_3). \quad (4.1.70)$$

Under the condition (4.1.62), it is known from (4.1.50) that the critical Rayleigh number $R_c = \pi^4(1 + L^2)^3/L^4$, and the wave numbers are $(k_1, k_2) = (1, 0)$ and $(0, 1)$. Namely the first eigenvectors of (4.1.63) with (4.1.34) near $R = R_c$ are given by

$$\psi_{10} = \left(-L \sin \frac{\pi x_1}{L} \cos \pi x_3, 0, \cos \frac{\pi x_1}{L} \sin \pi x_3, \sqrt{1 + L^2} \cos \frac{\pi x_1}{L} \sin \pi x_3 \right) \quad (4.1.71)$$

$$\psi_{01} = \left(0, -L \sin \frac{\pi x_2}{L} \cos \pi x_3, \cos \frac{\pi x_2}{L} \sin \pi x_3, \sqrt{1 + L^2} \cos \frac{\pi x_2}{L} \sin \pi x_3 \right) \quad (4.1.72)$$

and by (4.1.67) the first eigenvalue $\beta_1(R)$ of (4.1.63) with (4.1.34) near $R = R_c$ satisfies the PES:

$$\beta_1(R) \begin{cases} < 0 & \text{if } R < R_c, \\ = 0 & \text{if } R = R_c, \\ > 0 & \text{if } R > R_c. \end{cases} \quad (4.1.73)$$

The following is the main result in this subsection.

Theorem 4.1.14 *Assume the condition (4.1.62). Then for the Bénard problem (4.1.3) with (4.1.34), the following assertions hold true:*

- (1) *For $R > R_c (= \pi^4(1 + L^2)^3/L^4)$, the problem bifurcates from the trivial solution $((u, T), R) = (0, R_c)$ to an attractor Σ_R , homeomorphic to a cycle S^1 .*
- (2) *The attractor Σ_R consists of exactly eight singular points Φ_i ($1 \leq i \leq 8$) and eight heteroclinic orbits connecting the singular points, as shown in Fig. 4.7, where $\Phi_1, \Phi_3, \Phi_5, \Phi_7$ are minimal attractors, and $\Phi_2, \Phi_4, \Phi_6, \Phi_8$ are saddle points.*
- (3) *The bifurcated minimal attractors can be expressed as*

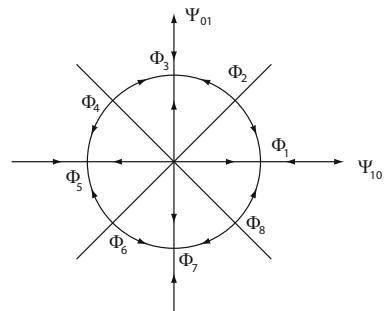
$$\Phi_{1,5} = \pm 4\beta_1^{1/2}(R)\psi_{10} + o(|\beta_1|), \quad \Phi_{3,7} = \pm 4\beta_1^{1/2}(R)\psi_{01} + o(|\beta_1|),$$

where β_1 is the first eigenvalue of (4.1.63) with (4.1.34) satisfying (4.1.73), and ψ_{10} and ψ_{01} are the first eigenvectors given by (4.1.71) and (4.1.72) respectively.

A few remarks are now in order.

Remark 4.1.15 Theorem 4.1.11 can be proved only by using the construction of center manifold functions together with the attractor bifurcation theory. The classical theories and methods are unavailable for the Bénard problem (4.1.3) and (4.1.34) in a rectangular region where the first eigenvalue has even multiplicity.

Fig. 4.7 The singular points $\Phi_1, \Phi_3, \Phi_5, \Phi_7$ are the minimal attractors, and the other singular points are the saddles



Remark 4.1.16 We know that the multiplicity of the first eigenvalue of (4.1.63) with (4.1.34) in a square region $\Omega = (0, L)^2 \times (0, 1)$ is two for many horizontal lengths $L > 0$. In the same fashion one can prove that the assertions (1) and (2) of Theorem 4.1.11 are also valid for these cases.

Remark 4.1.17 It is clear that the velocity fields in ψ_{10} and ψ_{01} have the roll structure as shown in Fig. 4.2. The roll axis of ψ_{10} is the x_2 -axis, and that of ψ_{01} is the x_1 -axis. The roll structure is of certain stability, therefore the bifurcated minimal attractors ϕ_{2k-1} ($1 \leq k \leq 4$) given by Assertion (3) of Theorem 4.1.11 have similar roll structure as well.

Proof of Theorem 4.1.14. To prove this theorem, we need to reduce (4.1.3) with (4.1.34) to the center manifold.

Let H be the space as defined in (4.1.11). It is known that all eigenvectors (4.1.69) and (4.1.70) constitute an orthogonal basis of H . Hence for any $\psi \in H$, ψ can be written as

$$\psi = \sum_{K,J} x_{KJ} \psi_{KJ}, \quad K = (k_1, k_2), \quad J = (j_1, j_2).$$

For simplicity, we take $\text{Pr} = 1$ in (4.1.3). Then the eigenvalue problem (4.1.63) with (4.1.34) is written as

$$L_\lambda \psi = \beta \psi, \quad (4.1.74)$$

and β_1 is the first eigenvalue of (4.1.74) satisfying (4.1.73), and ψ_{10} and ψ_{01} are the first eigenvectors given by (4.1.71) and (4.1.72). Let

$$\begin{aligned} \psi &= x\psi_{10} + y\psi_{01} + \Phi, \\ \Phi : E_0 &\rightarrow E_0^\perp \quad \text{the center manifold function,} \\ E_0 &= \text{span}\{\psi_{10}, \psi_{01}\}, \\ E_0^\perp &= \{\psi \in H \mid (\psi, \psi_{10}) = 0, (\psi, \psi_{01}) = 0\}. \end{aligned}$$

The reduced equations of (4.1.3) with (4.1.34) to the center manifold are given by

$$\begin{aligned}\frac{dx}{dt} &= \beta_1 x + \frac{1}{\|\psi_{10}\|^2} (G(x\psi_{10} + y\psi_{01} + \Phi), \psi_{10}), \\ \frac{dy}{dt} &= \beta_1 y + \frac{1}{\|\psi_{01}\|^2} (G(x\psi_{10} + y\psi_{01} + \Phi), \psi_{01}).\end{aligned}\quad (4.1.75)$$

By Theorem A.1.1, the center manifold function Φ satisfies

$$-L_\lambda \Phi = P_2 G(x\psi_{10} + y\psi_{01}) + o(x^2 + y^2), \quad (4.1.76)$$

where $P_2 : H \rightarrow E_0^\perp$ is the canonical projection.

For any $\psi_1 = (u, T_1)$ and $\psi_2 = (v, T_2)$ we denote

$$G(\psi_1, \psi_2) = (P[-(u \cdot \nabla)v], (u \cdot \nabla)T_2).$$

Then (4.1.76) can be rewritten as

$$\begin{aligned}-L_\lambda \Phi &= P_2[x^2 G(\psi_{10}, \psi_{10}) + y^2 G(\psi_{01}, \psi_{01}) + xy G(\psi_{10}, \psi_{01}) \\ &\quad + xy G(\psi_{01}, \psi_{10})] + o(x^2 + y^2).\end{aligned}\quad (4.1.77)$$

Notice that $\|\psi_{10}\| = \|\psi_{01}\|$. Hence the equations (4.1.75) are expressed as

$$\begin{aligned}\frac{dx}{dt} &= \beta_1 x + \frac{1}{\|\psi_{10}\|^2} \left[x(G(\psi_{10}, \Phi) + G(\Phi, \psi_{10}), \psi_{10}) \right. \\ &\quad \left. + y(G(\psi_{01}, \Phi) + G(\Phi, \psi_{01}), \psi_{10}) \right] + o(|x|^3 + |y|^3), \\ \frac{dy}{dt} &= \beta_1 y + \frac{1}{\|\psi_{10}\|^2} \left[y(G(\psi_{01}, \Phi) + G(\Phi, \psi_{01}), \psi_{01}) \right. \\ &\quad \left. + x(G(\psi_{10}, \Phi) + G(\Phi, \psi_{10}), \psi_{01}) \right] + o(|x|^3 + |y|^3).\end{aligned}\quad (4.1.78)$$

Using $(G(\psi_1, \psi_2), \psi_3) = -(G(\psi_1, \psi_3), \psi_2)$, we derive from (4.1.78) that

$$\begin{aligned}\frac{dx}{dt} &= \beta_1 x - \frac{1}{\|\psi_{10}\|^2} \left[x(G(\psi_{10}, \psi_{10}), \Phi) + y(G(\psi_{01}, \psi_{10}), \Phi) \right. \\ &\quad \left. - y(G(\Phi, \psi_{01}), \psi_{10}) \right] + o(|x|^3 + |y|^3), \\ \frac{dy}{dt} &= \beta_1 y - \frac{1}{\|\psi_{10}\|^2} \left[y(G(\psi_{01}, \psi_{01}), \Phi) + x(G(\psi_{10}, \psi_{01}), \Phi) \right. \\ &\quad \left. - x(G(\Phi, \psi_{01}), \psi_{10}) \right] + o(|x|^3 + |y|^3).\end{aligned}\quad (4.1.79)$$

Direct computation yields that

$$G(\psi_{10}, \psi_{10}) = P \begin{cases} -L\pi \sin \frac{\pi x_1}{L} \cos \frac{\pi x_1}{L}, \\ 0, \\ -\pi \sin \pi x_3 \cos \pi x_3, \\ -\pi \sqrt{1+L^2} \sin \pi x_3 \cos \pi x_3, \end{cases} \quad (4.1.80)$$

$$G(\psi_{01}, \psi_{01}) = P \begin{cases} 0, \\ -L\pi \sin \frac{\pi x_2}{L} \cos \frac{\pi x_2}{L}, \\ -\pi \sin \pi x_3 \cos \pi x_3, \\ -\pi \sqrt{1+L^2} \sin \pi x_3 \cos \pi x_3, \end{cases} \quad (4.1.81)$$

$$G(\psi_{10}, \psi_{01}) = P \begin{cases} 0, \\ -L\pi \cos \frac{\pi x_1}{L} \sin \frac{\pi x_2}{L} \sin^2 \pi x_3, \\ -\pi \cos \frac{\pi x_1}{L} \cos \frac{\pi x_2}{L} \sin \pi x_3 \cos \pi x_3, \\ -\pi \sqrt{1+L^2} \cos \frac{\pi x_1}{L} \cos \frac{\pi x_2}{L} \sin \pi x_3 \cos \pi x_3, \end{cases} \quad (4.1.82)$$

$$G(\psi_{01}, \psi_{10}) = P \begin{cases} -L\pi \sin \frac{\pi x_1}{L} \cos \frac{\pi x_2}{L} \sin^2 \pi x_3, \\ 0, \\ -\pi \cos \frac{\pi x_1}{L} \cos \frac{\pi x_2}{L} \sin \pi x_3 \cos \pi x_3, \\ -\pi \sqrt{1+L^2} \cos \frac{\pi x_1}{L} \cos \frac{\pi x_2}{L} \sin \pi x_3 \cos \pi x_3. \end{cases} \quad (4.1.83)$$

Under the action of the Leray projection $P : L^2(\Omega, \mathbb{R}^3) \rightarrow H$ in G , we find that

$$P \left(\sin \frac{\pi x_1}{L} \cos \frac{\pi x_1}{L}, 0, \sin \pi x_3 \cos \pi x_3 \right) = 0,$$

$$P \left(0, \sin \frac{\pi x_2}{L} \cos \frac{\pi x_2}{L}, \sin \pi x_3 \cos \pi x_3 \right) = 0.$$

Hence we have

$$G(\psi_{10}, \psi_{10}) = G(\psi_{01}, \psi_{01}) = (0, 0, 0, -\frac{\pi}{2} \sqrt{1+L^2} \sin 2\pi x_3). \quad (4.1.84)$$

Now we return to compute the center manifold function. Let

$$\Phi(x, y) = \sum_{K \neq (1, 0), (0, 1)} \Phi_{KJ}(x, y) \psi_{KJ} \quad \text{for } K = (k_1, k_2), J = (j_1, j_2).$$

We infer from (4.1.74) and (4.1.77) that

$$\begin{aligned}
-L_\lambda \Phi &= \sum_{K \neq (1,0), (0,1)} \Phi_{KJ} (-\beta_K^J) \psi_{KJ} \\
&= x^2 G(\psi_{10}, \psi_{10}) + y^2 G(\psi_{01}, \psi_{01}) + xy G(\psi_{10}, \psi_{01}) \\
&\quad + xy G(\psi_{01}, \psi_{10}) + o(x^2 + y^2).
\end{aligned} \tag{4.1.85}$$

On the other hand, from (4.1.80) to (4.1.84), we see that

$$\begin{aligned}
(G(\psi_{10}, \psi_{10}), \psi_{KJ}) &= (G(\psi_{01}, \psi_{01}), \psi_{KJ}) = 0 & \forall \psi_{KJ} \neq \psi_{02}, \\
(G(\psi_{10}, \psi_{01}), \psi_{KJ}) &= 0 & \forall \psi_{KJ} \neq \psi_{1122}, \\
(G(\psi_{01}, \psi_{10}), \psi_{KJ}) &= 0 & \forall \psi_{KJ} \neq \psi_{1122}.
\end{aligned}$$

Hence, by (4.1.85), we derive that

$$\begin{aligned}
\Phi(x, y) &= \Phi_{02}(x, y)\psi_{02} + \Phi_{1122}(x, y)\psi_{1122} + o(x^2 + y^2), \\
\Phi_{02}(x, y) &= \frac{x^2 + y^2}{-\beta_0^2 \|\psi_{02}\|^2} (G(\psi_{10}, \psi_{10}), \psi_{02}), \\
\Phi_{1122}(x, y) &= \frac{xy}{-\beta_{11}^{22} \|\psi_{1122}\|^2} (G(\psi_{10}, \psi_{01}) + G(\psi_{01}, \psi_{10}), \psi_{1122}) \\
&= \frac{2xy}{-\beta_{11}^{22} \|\psi_{1122}\|^2} (G(\psi_{10}, \psi_{01}), \psi_{1122}).
\end{aligned} \tag{4.1.86}$$

Here ψ_{02} and ψ_{1122} are given by

$$\psi_{02} = (0, 0, 0, \sin 2\pi x_3), \quad \psi_{1122} = \begin{cases} -L \sin \frac{\pi x_1}{L} \cos \frac{\pi x_2}{L} \cos 2\pi x_3, \\ -L \cos \frac{\pi x_1}{L} \sin \frac{\pi x_2}{L} \cos 2\pi x_3, \\ \cos \frac{\pi x_1}{L} \cos \frac{\pi x_2}{L} \sin 2\pi x_3, \\ \alpha_{11}^{22} \cos \frac{\pi x_1}{L} \cos \frac{\pi x_2}{L} \sin 2\pi x_3. \end{cases}$$

By (4.1.67), $\alpha_{11}^{22} = \sqrt{1 + 2L^2}$.

Putting the center manifold function (4.1.86) in (4.1.79), we obtain that

$$\begin{aligned}
\frac{dx}{dt} &= \beta_1 x - \alpha x(x^2 + y^2) - (b + c)xy^2 + o(|x|^3 + |y|^3), \\
\frac{dy}{dt} &= \beta_1 y - \alpha y(x^2 + y^2) - (b + c)yx^2 + o(|x|^3 + |y|^3),
\end{aligned} \tag{4.1.87}$$

where $b > c$ and

$$\begin{aligned}\alpha &= \frac{|(G(\psi_{10}, \psi_{10}), \psi_{02})|^2}{-\beta_0^2 \|\psi_{10}\|^2 \|\psi_{02}\|^2} = \frac{1}{16}, \\ b &= \frac{2|(G(\psi_{10}, \psi_{01}), \psi_{1122})|^2}{-\beta_{11}^{22} \|\psi_{1122}\|^2} = \frac{\pi^2 L^2 \left[1 + L^2 + \sqrt{(1 + L^2)(1 + 2L^2)} \right]^2}{32(-\beta_{11}^{22})(1 + 2L^2)}, \\ c &= \frac{2(G(\psi_{1122}, \psi_{01}), \psi_{10})(G(\psi_{10}, \psi_{01}), \psi_{1122})}{-\beta_{11}^{22} \|\psi_{1122}\|^2} \\ &= \frac{\pi^2 L^2 (1 + L^2)(1 + L^2 + \sqrt{(1 + L^2)(1 + 2L^2)})}{32(-\beta_{11}^{22})(1 + 2L^2)}, \\ -\beta_{11}^{22} &= 2\pi^2(2 + \frac{1}{L^2}) - \frac{\sqrt{R}}{\sqrt{1 + 2L^2}} > 0 \quad \text{at} \quad R = R_c = \frac{\pi^4(1 + L^2)^3}{L^4}.\end{aligned}$$

Define a vector field $g(x, y)$ by

$$g(x, y) = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} \alpha x(x^2 + y^2) + (b + c)xy^2 + o(|x|^3 + |y|^3) \\ \alpha y(x^2 + y^2) + (b + c)yx^2 + o(|x|^3 + |y|^3) \end{pmatrix}.$$

Then we have

$$xg_1 + yg_2 = \alpha(x^2 + y^2)^2 + 2(b + c)x^2y^2 + o(|x|^4 + |y|^4). \quad (4.1.88)$$

By the S^1 -attractor bifurcation theorem (Theorem 2.2.4), it follows from (4.1.73) and (4.1.88) that the reduced equation (4.1.87) bifurcate from $((x, y), R) = (0, R_c)$ to an attractor Σ_R homeomorphic to a cycle S^1 , i.e., $\Sigma_R = S^1$. Thus Assertion (1) is proved.

In addition, it is easy to see that the equations (4.1.87) have eight singular points:

$$\begin{aligned}Z_1 &= (4\beta_1^{1/2} + o(|\beta_1|^{1/2}), 0), & Z_3 &= (0, 4\beta_1^{1/2} + o(|\beta_1|^{1/2})), \\ Z_5 &= (-4\beta_1^{1/2} + o(|\beta_1|^{1/2}), 0), & Z_7 &= (0, -4\beta_1^{1/2} + o(|\beta_1|^{1/2})), \\ Z_2 &= (y_0, y_0), & Z_4 &= (-y_0, y_0), \\ Z_6 &= (-y_0, -y_0), & Z_8 &= (y_0, -y_0),\end{aligned}$$

where

$$y_0 = 2\sqrt{\frac{1}{1/8 + b + c}\beta_1^{1/2} + o(|\beta_1|^{1/2})}.$$

The Jacobian matrices of (4.1.87) at Z_{2k-1} ($1 \leq k \leq 4$) are given by

$$J(Z_{1,5}) = \begin{pmatrix} -2\beta_1 & 0 \\ 0 & -1 - 16(b + c)\beta_1 \end{pmatrix}, \quad J(Z_{3,7}) = \begin{pmatrix} -1 - 16(b + c)\beta_1 & 0 \\ 0 & -2\beta_1 \end{pmatrix}.$$

Namely, the eigenvalues at Z_{2k-1} ($1 \leq k \leq 4$) are negative. Therefore, Z_{2k-1} ($1 \leq k \leq 4$) are the minimal attractors of (4.1.87). Then it follows from Theorem 5.6 in Ma and Wang (2005b) that the other singular points Z_{2k} ($1 \leq k \leq 4$) are saddle points, and the topological structure of the bifurcated attractor Σ_R is as shown in Fig. 4.7. Thus Assertions (2) and (3) are proved. The proof is complete. \square

4.1.7 Physical Remarks

As mentioned earlier, the Rayleigh–Bénard convection together with the Taylor problem has become a paradigm of nonequilibrium phase transitions with its instability, bifurcations, pattern formation, and chaotic behaviors in fluid dynamics as well as in physical systems in general.

We have shown that as the Rayleigh number R crosses the critical parameter R_c , the system always under a continuous (Type-I) transition to an $(m - 1)$ -dimensional sphere $\mathcal{A}_R = S^{m-1}$, where m is the dimension of the eigenspace for the linearized problem, corresponding to the critical Rayleigh number R_c .

The structure of the set $\mathcal{A}_R = S^{m-1}$ of transition states and the related pattern formation mechanism depend on the physical boundary conditions and the spatial geometry (of the domain). In most cases studies, the high dimensional attractor \mathcal{A}_R consists of not only steady-state solutions, but also transients, which has full measure in \mathcal{A}_R . The Key Philosophy 2 and Principle 1 play crucial role in deriving these results, which cannot be directly derived by any existing methods. In fact, without the Key Philosophy 2 and Principle 1, one may not have any motivation to envision these results. Also, these results are global in nature.

Finally, the dynamic transition theory opens a new avenue toward to pattern formation, as the patterns and structure of transition states are dictated by solutions in \mathcal{A}_R , which are again determined by eigenvectors of the corresponding linearized problem (with high-order approximations). With the theory, for example, we can easily examine the formation and stability of different structures such as roles, rectangles, and hexagons.

4.2 Taylor–Couette Flow

In the Rayleigh–Bénard convection we have seen that the thermal motion is an important cause, leading to the instability of fluid flows. In this section we shall investigate the instability of the Taylor–Couette flow caused by the centrifugal forces of curved fluid flows.

The study of hydrodynamic instability caused by the centrifugal forces originated from the famous experiments conducted by Taylor (1923), in which he observed and

studied the stability of an incompressible viscous fluid between two rotating coaxial cylinders. In his experiments, Taylor investigated the case where the gap between the two cylinders is small in comparison with the mean radius, and the two cylinders rotate in the same direction. He found that when the Taylor number T is smaller than a critical value $T_c > 0$, called the critical Taylor number, the basic flow, called the Couette flow, is stable, and when the Taylor number crosses the critical value, the Couette flow breaks out into a cellular pattern that is radially symmetric; see Fig. 4.8.

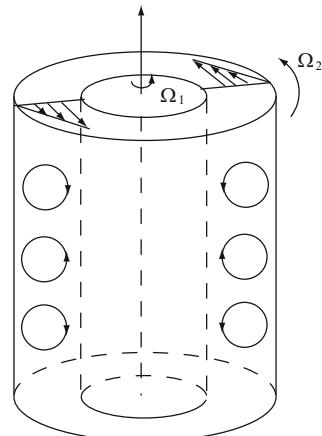
The main objective of this section is to explore the dynamic transitions and its physical implications of the Taylor problem. The results presented here are based on Ma and Wang (2006, 2010a).

4.2.1 Taylor Problem

Consider an incompressible viscous fluid between two coaxial cylinders. Let r_1 and r_2 ($r_2 > r_1$) be the radii of the two cylinders, Ω_1 and Ω_2 the angular velocities of the inner and the outer cylinders respectively, and

$$\mu = \Omega_2 / \Omega_1, \quad \eta = r_1 / r_2. \quad (4.2.1)$$

Fig. 4.8 The Couette flow and the Taylor vortices



The nondimensional Taylor number is defined by

$$T = \frac{4h^4\Omega_1^2}{\nu^2}, \quad (4.2.2)$$

where $\nu > 0$ is the kinematic viscosity, and h is the vertical length scale.

There exists a basic steady-state flow, called the Couette flow. In the cylindrical polar coordinate (r, θ, z) , the Couette flow is defined by

$$(u_r, u_\theta, u_z, p) = \left(0, V(r), 0, \rho \int \frac{1}{r} V^2(r) dr\right), \quad V(r) = ar + \frac{b}{r}, \quad (4.2.3)$$

where (u_r, u_θ, u_z) is the velocity field, p is the pressure, and a, b are constants. It follows from the boundary conditions that

$$V(r_1) = \Omega_1 r_1, \quad V(r_2) = \Omega_2 r_2, \quad a = -\Omega_1 \eta^2 \frac{1 - \mu/\eta^2}{1 - \eta^2}, \quad b = \Omega_1 \frac{r_1^2(1 - \mu)}{1 - \eta^2}.$$

Here μ and η are given by (4.2.1).

Based on the Rayleigh criterion, when $\mu > \eta^2$, the Couette flow is always stable at a distribution of angular velocities

$$\Omega(r) = a + b/r^2 \quad \text{for } r_1 < r < r_2.$$

However, when $\mu < \eta^2$, the situation is different. As in the Taylor experiments, consider the case where the gap $r_2 - r_1$ is much smaller than the mean radius:

$$r_2 - r_1 \ll r_0 = (r_1 + r_2)/2,$$

and the two cylinders rotate in the same direction. If the Taylor number T in (4.2.2) satisfies $T < T_c$, then the Couette flow (4.2.3) is stable, and if $T_c < T < T_c + \varepsilon$ for some $\varepsilon > 0$, a street of vortices along the z -axis, called the Taylor vortices, emerge abruptly from the basic flow, as shown in Fig. 4.8, and the corresponding flow pattern is radically symmetric and structurally stable.

When the gap $r_2 - r_1$ is not small than r_0 , or when the cylinders rotate in the opposite directions, the phenomena one observes are much more complex; see Chandrasekhar (1981) for details. Hence we always assume the condition

$$\eta^2 > \mu \geq 0. \quad (4.2.4)$$

4.2.2 Governing Equations

The hydrodynamic equations governing an incompressible viscous fluid between two coaxial cylinders are the Navier–Stokes equations. In the cylindrical polar coordinates (r, θ, z) , they are given by

$$\begin{aligned}
\frac{\partial u_r}{\partial t} + (u \cdot \nabla) u_r - \frac{u_\theta^2}{r} &= \nu \left(\Delta u_r - \frac{2}{r^2} \frac{\partial u_\theta}{\partial \theta} - \frac{u_r}{r^2} \right) - \frac{1}{\rho} \frac{\partial p}{\partial r}, \\
\frac{\partial u_\theta}{\partial t} + (u \cdot \nabla) u_\theta + \frac{u_r u_\theta}{r} &= \nu \left(\Delta u_\theta + \frac{2}{r^2} \frac{\partial u_r}{\partial \theta} - \frac{u_\theta}{r^2} \right) - \frac{1}{r \rho} \frac{\partial p}{\partial \theta}, \\
\frac{\partial u_z}{\partial t} + (u \cdot \nabla) u_z &= \nu \Delta u_z - \frac{1}{\rho} \frac{\partial p}{\partial z}, \\
\frac{\partial(r u_r)}{\partial r} + \frac{\partial u_\theta}{\partial \theta} + \frac{\partial(r u_z)}{\partial z} &= 0,
\end{aligned} \tag{4.2.5}$$

where ν is the kinematic viscosity, ρ is the density, $u = (u_r, u_\theta, u_z)$ is the velocity field, p is the pressure function, and

$$u \cdot \nabla = u_r \frac{\partial}{\partial r} + \frac{u_\theta}{r} \frac{\partial}{\partial \theta} + u_z \frac{\partial}{\partial z}, \quad \Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2}.$$

Then it is easy to see that the Couette flow (4.2.3) is a steady-state solution of (4.2.5). In order to investigate its stability and transitions, we need to consider the perturbed state of (4.2.3):

$$u_r, u_\theta + V(r), u_z, p + \rho \int \frac{1}{r} v^2(r) dr.$$

The perturbed equations read

$$\begin{aligned}
\frac{\partial u_r}{\partial t} + (u \cdot \nabla) u_r - \frac{u_\theta^2}{r} &= \nu \left(\Delta u_r - \frac{2}{r^2} \frac{\partial u_\theta}{\partial \theta} - \frac{u_r}{r^2} \right) - \frac{1}{\rho} \frac{\partial p}{\partial r} \\
&\quad + \frac{2V(r)}{r} u_\theta - \frac{V(r)}{r} \frac{\partial u_r}{\partial \theta}, \\
\frac{\partial u_\theta}{\partial t} + (u \cdot \nabla) u_\theta + \frac{u_r u_\theta}{r} &= \nu \left(\Delta u_\theta + \frac{2}{r^2} \frac{\partial u_r}{\partial \theta} - \frac{u_\theta}{r^2} \right) \\
&\quad - \frac{1}{r \rho} \frac{\partial p}{\partial \theta} - (V' + \frac{V}{r}) u_r - \frac{V}{r} \frac{\partial u_\theta}{\partial \theta}, \\
\frac{\partial u_z}{\partial t} + (u \cdot \nabla) u_z &= \nu \Delta u_z - \frac{1}{\rho} \frac{\partial p}{\partial z} - \frac{V}{r} \frac{\partial u_z}{\partial \theta}, \\
\frac{\partial(r u_z)}{\partial z} + \frac{\partial(r u_r)}{\partial r} + \frac{\partial u_\theta}{\partial \theta} &= 0.
\end{aligned} \tag{4.2.6}$$

To derive the nondimensional form of equations (4.2.6), let

$$\begin{aligned}
(x, t) &= (hx', h^2 t' / \nu) & (x = (r, r\theta, z)), \\
(u, p) &= (\nu u' / h, \rho \nu^2 p' / h^2) & (u = (u_r, u_\theta, u_z)).
\end{aligned}$$

Omitting the primes, we obtain the nondimensional form of (4.2.6) as follows:

$$\begin{aligned} \frac{\partial u_r}{\partial t} &= \Delta u_r - \frac{2}{r^2} \frac{\partial u_\theta}{\partial \theta} - \frac{u_r}{r^2} - (u \cdot \nabla) u_r + \frac{u_\theta^2}{r} - \frac{\partial p}{\partial r} \\ &\quad - \sqrt{T} \left(\frac{\eta^2 - \mu}{1 - \eta^2} - \frac{1 - \mu}{1 - \eta^2} \frac{r_1^2}{r^2} \right) \left(u_\theta - 1/2 \frac{\partial u_r}{\partial \theta} \right), \\ \frac{\partial u_\theta}{\partial t} &= \Delta u_\theta + \frac{2}{r^2} \frac{\partial u_r}{\partial \theta} - \frac{u_\theta}{r^2} - (u \cdot \nabla) u_\theta - \frac{u_\theta u_r}{r} - \frac{1}{r} \frac{\partial p}{\partial \theta} \\ &\quad + \sqrt{T} \frac{\eta^2 - \mu}{1 - \eta^2} u_r + \frac{\sqrt{T}}{2} \left(\frac{\eta^2 - \mu}{1 - \eta^2} - \frac{1 - \mu}{1 - \eta^2} \frac{r_1^2}{r^2} \right) \frac{\partial u_\theta}{\partial \theta}, \\ \frac{\partial u_z}{\partial t} &= \Delta u_z - (u \cdot \nabla) u_z - \frac{\partial p}{\partial z} + \frac{\sqrt{T}}{2} \left(\frac{\eta^2 - \mu}{1 - \eta^2} - \frac{1 - \mu}{1 - \eta^2} \frac{r_1^2}{r^2} \right) \frac{\partial u_z}{\partial \theta}, \\ \frac{\partial(r u_z)}{\partial z} + \frac{\partial(r u_r)}{\partial r} + \frac{\partial u_\theta}{\partial \theta} &= 0, \end{aligned} \tag{4.2.7}$$

where T is the Taylor number as defined in (4.2.2).

The nondimensional domain for (4.2.7) is

$$\Omega = (l_1, l_2) \times (0, 2\pi) \times (0, L),$$

where $l_i = r_i/h$ ($i = 1, 2$), and L is the height of the fluid between the two cylinders. The initial value condition for (4.2.7) is given by

$$u(r, \theta, z, 0) = u_0(r, \theta, z). \tag{4.2.8}$$

There are different physically sound boundary conditions. In the θ -direction it is periodic

$$u(r, \theta + 2k\pi, z) = u(r, \theta, z), \quad \forall k \in \mathbb{Z}. \tag{4.2.9}$$

In the radial direction, there is the rigid boundary condition

$$u = (u_z, u_r, u_\theta) = 0, \quad \text{at } r = l_1, l_2. \tag{4.2.10}$$

At the top and bottom in the z -direction ($z = 0, L$), either the free boundary condition or the rigid boundary condition or the periodic boundary condition can be used:

DIRICHLET BOUNDARY CONDITION:

$$u = (u_r, u_\theta, u_z) = 0 \quad \text{at } z = 0, L; \tag{4.2.11}$$

FREE-SLIP BOUNDARY CONDITION:

$$u_z = 0, \quad \frac{\partial u_r}{\partial z} = \frac{\partial u_\theta}{\partial z} = 0 \quad \text{at } z = 0, L; \tag{4.2.12}$$

FREE-RIGID BOUNDARY CONDITION:

$$\begin{aligned} u_z = 0, \frac{\partial u_r}{\partial z} = \frac{\partial u_\theta}{\partial z} = 0 & \quad \text{at } z = L, \\ u = (u_z, u_r, u_\theta) = 0 & \quad \text{at } z = 0; \end{aligned} \tag{4.2.13}$$

PERIODIC BOUNDARY CONDITION:

$$u(r, \theta, z + 2kL) = u(r, \theta, z) \quad \forall k \in \mathbb{Z}. \tag{4.2.14}$$

Remark 4.2.1 From the physical point of view, the free boundary condition is usually used for the case where the friction on the surface of the boundary is negligible; otherwise the rigid boundary condition is used. Moreover, if the height L of the fluid is sufficiently large in comparison with the gap $r_2 - r_1$ between the two cylinders, it is reasonable to use the periodic boundary condition.

4.2.3 Narrow-gap Case with Axisymmetric Perturbations

We consider here the case where the gap $r_2 - r_1$ is small compared to the mean radius $r_0 = (r_1 + r_2)/2$ with $\mu \geq 0$ and with axisymmetric perturbations. This case is the situation investigated by Taylor (1923) in 1923.

We take the length scale $h = r_2 - r_1$. Then the narrow gap condition is given by

$$1 = r_2 - r_1 \ll (r_1 + r_2)/2. \tag{4.2.15}$$

Under the assumption (4.2.15), we can neglect the terms containing r^{-n} ($n \geq 1$) in (4.2.7). In addition, by (4.2.15) we have

$$\begin{aligned} -\sqrt{T} \left(\frac{\eta^2 - \mu}{1 - \eta^2} - \frac{1 - \mu}{1 - \eta^2} \frac{r_1^2}{r^2} \right) &= \sqrt{T} \left(1 - \frac{1 - \mu}{1 - \eta^2} \frac{r^2 - r_1^2}{r^2} \right) \\ &\simeq \sqrt{T}(1 - (1 - \mu)(r - r_1)). \end{aligned}$$

Let

$$\alpha = \frac{\eta^2 - \mu}{1 - \eta^2}. \tag{4.2.16}$$

Replacing u_θ by $\sqrt{\alpha}u_\theta$, and assuming the perturbations are axisymmetric and are independent of θ , we obtain from (4.2.7):

$$\begin{aligned}\frac{\partial u_z}{\partial t} + (\tilde{u} \cdot \nabla) u_z &= \Delta u_z - \frac{\partial p}{\partial z}, \\ \frac{\partial u_r}{\partial t} + (\tilde{u} \cdot \nabla) u_r &= \Delta u_r - \frac{\partial p}{\partial r} + \sqrt{\alpha T}(1 - (1 - \mu)(r - r_1))u_\theta, \\ \frac{\partial u_\theta}{\partial t} + (\tilde{u} \cdot \nabla) u_\theta &= \Delta u_\theta + \sqrt{\alpha T}u_r, \\ \frac{\partial u_r}{\partial r} + \frac{\partial u_z}{\partial z} &= 0,\end{aligned}\tag{4.2.17}$$

where

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{\partial^2}{\partial z^2}, \quad (\tilde{u} \cdot \nabla) = u_r \frac{\partial}{\partial r} + u_z \frac{\partial}{\partial z}.$$

In this case, the spatial domain is $M = (r_1, r_1 + 1) \times (0, L)$. For convenience, we consider here the Dirichlet boundary condition

$$u|_{\partial M} = 0.\tag{4.2.18}$$

The initial value condition is axisymmetric, and given by

$$u = u_0(r, z) \quad \text{at} \quad t = 0.\tag{4.2.19}$$

The linearized equations of (4.2.17) read

$$\begin{aligned}-\Delta u_z + \frac{\partial p}{\partial z} &= 0, \\ -\Delta u_r + \frac{\partial p}{\partial r} &= \lambda u_\theta - \lambda(1 - \mu)(r - r_1)u_\theta, \\ -\Delta u_\theta &= \lambda u_r, \\ \frac{\partial u_r}{\partial r} + \frac{\partial u_z}{\partial z} &= 0,\end{aligned}\tag{4.2.20}$$

where $\lambda = \sqrt{\alpha T}$, and T is the Taylor number given by (4.2.2).

Let $\lambda_1 > 0$ be the first eigenvalue of (4.2.20) with (4.2.18). We call

$$T_c = \lambda_1^2 / \alpha,\tag{4.2.21}$$

the critical Taylor number, where α is given by (4.2.16).

As $\mu \rightarrow 1$, equations (4.2.20) are reduced to the following symmetric linear equations:

$$\begin{aligned}-\Delta u_z + \frac{\partial p}{\partial z} &= 0, \\ -\Delta u_r + \frac{\partial p}{\partial r} &= \lambda u_\theta, \\ -\Delta u_\theta &= \lambda u_r, \\ \frac{\partial u_z}{\partial z} + \frac{\partial u_r}{\partial r} &= 0.\end{aligned}\tag{4.2.22}$$

Let the first eigenvalue $\lambda_0 > 0$ of (4.2.22) with (4.2.18) have multiplicity $m \geq 1$, the corresponding eigenfunctions be v_i ($i = 1, \dots, m$), and the corresponding eigenspace be

$$E_0 = \text{span}\{v_i \mid 1 \leq i \leq m\}.$$

We remark here that under conditions (4.2.4) and (4.2.15), the condition $\mu \rightarrow 1$ can be equivalently replaced by

$$r_1 = (2 + \delta)/(1 - \mu), \quad (4.2.23)$$

for some $\delta > 0$. In this case the parameter α in (4.2.16) is

$$\alpha = (\eta^2 - \mu)/(1 - \eta^2) \simeq \delta/2.$$

Let $\tilde{u} = (u_z, u_r)$ and $u = (\tilde{u}, u_\theta)$. For the Taylor problems (4.2.17)–(4.2.19), we set the function spaces as follows:

$$\begin{aligned} H &= \{u = (\tilde{u}, u_\theta) \in L^2(M)^3 \mid \text{div} \tilde{u} = 0, \tilde{u} \cdot n|_{\partial M} = 0\}, \\ H_1 &= \{u \in H^2(M)^3 \cap H \mid u \text{ satisfy (4.2.18)}\}. \end{aligned}$$

The following theorems provide the nonequilibrium phase transitions at the critical Taylor number T_c (T_c as in (4.2.21)) for the Taylor problems (4.2.17)–(4.2.19) in the narrow-gap case with $\mu > 0$ and with axisymmetric perturbations.

Theorem 4.2.2 *Let the conditions (4.2.15) and (4.2.23) hold true. Then the Taylor problems (4.2.17)–(4.2.19) have a transition at $(u, T) = (0, T_c)$. More precisely, we have the following assertions:*

- (1) *The problem undergoes a transition to an attractor \mathcal{A}_T on $T > T_c$ with $0 \notin \mathcal{A}_T$, and \mathcal{A}_T is an $(m - 1)$ -dimensional homological sphere.*
- (2) *The bifurcated attractor \mathcal{A}_T attracts $H \setminus \Gamma$ where $\Gamma \subset H$ is the stable manifold of $u = 0$ with codimension m in H .*
- (3) *Any $u_T \in \mathcal{A}_T$ can be expressed as*

$$u_T = v_T + w_T^\mu, \quad v_T \in E_0, \quad \lim_{\mu \rightarrow 1, \lambda \rightarrow \lambda_0} \|w_T^\mu\| / \|v_T\| = 0, \quad (4.2.24)$$

where $\lambda = \sqrt{\alpha T}$, and λ_0 is the first eigenvalue of (4.2.22) with (4.2.18).

Theorem 4.2.3 *Under the hypotheses of Theorem 4.2.2, if the first eigenvalue λ_0 of (4.2.22) with (4.2.18) is simple, then the transition of (4.2.17)–(4.2.19) at $(0, T_c)$ is of either Type-I (continuous) or Type-III (mixed). Moreover, we have the following assertions:*

- (1) *The bifurcated attractor \mathcal{A}_T in Theorem 4.2.2 consists of exactly two equilibrium points of (4.2.17)–(4.2.19), i.e., $\mathcal{A}_T = \{u_1^T, u_2^T\}$, such that*

$$\begin{aligned} u_1^T &= \alpha_1(T, \mu)v_0 + w_1(T, \mu), & u_2^T &= -\alpha_2(T, \mu)v_0 + w_2(T, \mu), \\ w_i(T, \mu) &= o(|\alpha_i|), & v_0 &\in E_0, \end{aligned} \quad (4.2.25)$$

where $\alpha_i > 0$ and $\alpha_i(T, \mu) \rightarrow 0$ as $\mu \rightarrow 1$ and $T \rightarrow T_0 = \sqrt{\lambda_0}/\alpha$.
(2) H can be decomposed into two open sets $H = \overline{U}_1^T + \overline{U}_2^T$ such that

$$\begin{aligned} U_1^T \cap U_2^T &= \emptyset, & 0 &\in \partial U_1^T \cap \partial U_2^T, \\ u_i^T &\in U_i^T, & \lim_{t \rightarrow \infty} \|u(t, u_0) - u_i^T\| &= 0 \quad \text{if } u_0 \in U_i^T, \end{aligned}$$

for $i=1, 2$, where $u(t, u_0)$ is the solution of (4.2.17)–(4.2.19).

- (3) If the transition is of Type-III, then the problem (4.2.17)–(4.2.19) has a saddle-node bifurcation at some point (u^*, T^*) with $u^* \neq 0$ and $0 < T^* < T_c$, where T_c is given in (4.2.21).

Remark 4.2.4 For (4.2.17) with other physically sound boundary conditions, given in (4.2.10)–(4.2.14), the results of Theorems 4.2.2 and 4.2.3 are also valid.

Remark 4.2.5 The expressions (4.2.24) and (4.2.25) are very useful for the Taylor problem, and show that the asymptotic topological structure of the problem (4.2.17)–(4.2.19) is governed by the first eigenfunctions of the symmetric linearized equation (4.2.22). In the next subsection, we shall see that the first eigenfunctions of (4.2.22) with the free boundary condition have the Taylor vortex type of structure, and these eigenfunctions are structurally stable.

Proof of Theorem 4.2.2. We shall apply Theorem 2.6.1 to prove this theorem, and divide the proof into the following three steps.

STEP 1. Let $G, S_\lambda^\mu, L_\lambda = -A + B_\lambda : H_1 \rightarrow H$ be mappings defined by

$$\begin{aligned} -Au &= \{P\Delta\tilde{u}, \Delta u_\theta\}, & B_\lambda u &= \{\lambda P(0, u_\theta), \lambda u_r\}, \\ S_\lambda^\mu u &= \{\lambda P(0, -(1-\mu)(r-r_1)u_\theta), 0\}, \\ G(u) &= \{-P(\tilde{u} \cdot \nabla)\tilde{u}, -(\tilde{u} \cdot \nabla)u_\theta\}, \end{aligned} \quad (4.2.26)$$

where $\lambda = \sqrt{\alpha T}$, $u = (u_z, u_r, u_\theta) \in H_1$, and the operator P is the Leray projection. Thus (4.2.17) can be written in the abstract form:

$$\frac{du}{dt} = L_\lambda u + S_\lambda^\mu u + G(u). \quad (4.2.27)$$

It is well known that the operators (4.2.26) satisfy the conditions (2.1.2) and (2.1.3), and L_λ is symmetric. In addition, the operator S_λ^μ satisfies that

$$\|S_\lambda^\mu\| \leq \lambda(1-\mu).$$

By (4.2.15) and (4.2.23), $1 - \mu \rightarrow 0$ as $r_1 \rightarrow \infty$. Thus the condition (2.6.2) is verified.

STEP 2. Consider the eigenvalue problem

$$L_\lambda u = \beta(\lambda)u \quad \text{for } u = (u_z, u_r, u_\theta) \in H_1. \quad (4.2.28)$$

By definition, the abstract form (4.2.28) is equivalent to the following eigenvalue equations with boundary condition (4.2.18):

$$\begin{aligned} \Delta u_z - \frac{\partial p}{\partial z} &= \beta(\lambda)u_z, \\ \Delta u_r - \frac{\partial p}{\partial r} + \lambda u_\theta &= \beta(\lambda)u_r, \\ \Delta u_\theta + \lambda u_r &= \beta(\lambda)u_\theta, \\ \operatorname{div} \tilde{u} &= 0, \end{aligned} \quad (4.2.29)$$

where $\tilde{u} = (u_z, u_r)$. It is known that the eigenvalues β_k ($k = 1, 2, \dots$) of (4.2.29) with (4.2.18) are real numbers satisfying

$$\beta_1(\lambda) \geq \beta_2(\lambda) \geq \dots \geq \beta_k(\lambda) \geq \dots, \quad \lim_{k \rightarrow \infty} \beta_k(\lambda) = -\infty. \quad (4.2.30)$$

At the first eigenvalue λ_0 of (4.2.22) with (4.2.18), the eigenvalues $\beta_k(\lambda)$ ($k = 1, 2, \dots$) in (4.2.30) satisfy

$$\begin{aligned} \beta_i(\lambda_0) &= 0 \quad \forall 1 \leq i \leq m, \\ \beta_j(\lambda_0) &< 0 \quad \forall j \geq m+1, \end{aligned} \quad (4.2.31)$$

where $m \geq 1$ is the multiplicity of the first eigenvalue $\beta_1(\lambda_0)$.

By Theorem 2.1 in Ma and Wang (2008d), it is ready to verify that the first eigenvalues $\beta_1(\lambda) = \dots = \beta_m(\lambda)$ are differentiable at $\lambda = \lambda_0$, and have the following Taylor expansions:

$$\beta_i(\lambda) = \beta'_i(\lambda_0)(\lambda - \lambda_0) + o(|\lambda - \lambda_0|), \quad \beta'_i(\lambda_0) = 2 \int_M u_{i\theta} u_{ir} dx > 0, \quad (4.2.32)$$

for $1 \leq i \leq m$, where $u_i = (u_{iz}, u_{ir}, u_{i\theta})$ are the first eigenfunctions.

Hence it follows from (4.2.31) and (4.2.32) that the conditions (2.1.4) and (2.1.5) hold true here.

STEP 3. It is well known that the Taylor problems (4.2.17)–(4.2.19) possess a global attractor for all $(\lambda, \mu) \in \mathbb{R}^2$. Hence we only need to prove that $u = 0$ is globally asymptotically stable for the following equation at $\lambda = \lambda_0$:

$$\frac{du}{dt} = L_\lambda u + G(u), \quad (4.2.33)$$

which corresponds to

$$\begin{aligned}\frac{\partial u_z}{\partial t} + (\tilde{u} \cdot \nabla) u &= \Delta u_z - \frac{\partial p}{\partial z}, \\ \frac{\partial u_r}{\partial t} + (\tilde{u} \cdot \nabla) u_r &= \Delta u_r - \frac{\partial p}{\partial r} + \lambda u_\theta, \\ \frac{\partial u_\theta}{\partial t} + (\tilde{u} \cdot \nabla) u_\theta &= \Delta u_\theta + \lambda u_r, \\ \operatorname{div} \tilde{u} &= 0.\end{aligned}$$

These equations have the same form as that of the two-dimensional Boussinesq equations. We note that L_λ is symmetric, and $G(u)$ is orthogonal. Hence, as in the proof of Theorem 4.1.1, by using Theorem 3.16 in Ma and Wang (2005b), we obtain that $u = 0$ is globally asymptotically stable for (4.2.33) at $\lambda = \lambda_0$.

Thus the proof of the theorem is complete. \square

Proof of Theorem 4.2.3. The proof follows from Theorem 2.6.6 and the proof of Theorem 4.2.2, which has verified the conditions that appeared in Theorem 2.6.6. \square

4.2.4 Asymptotic Structure of Solutions and the Taylor vortices

The main objective of this subsection is to study the structure and its stability in the physical space of the solutions associated with the secondary flows for the Taylor problem. In particular, we shall provide a rigorous confirmation of the presence of the Taylor vortex structure as observed by Taylor in his experiments (Taylor, 1923).

We mainly consider the equations (4.2.17) with one of the following two boundary conditions:

DIRICHLET BOUNDARY CONDITION:

$$\begin{aligned}u_z &= u_r = u_\theta = 0 && \text{at } r = r_1, r_1 + 1, \\ u_z &= 0, \frac{\partial u_r}{\partial z} = 0, \frac{\partial u_\theta}{\partial z} = 0 && \text{at } z = 0, L;\end{aligned}\tag{4.2.34}$$

FREE BOUNDARY CONDITION:

$$\begin{aligned}u_r &= 0, \frac{\partial u_z}{\partial r} = 0, \frac{\partial u_\theta}{\partial r} = 0 && \text{at } r = r_1, r_1 + 1, \\ u_z &= 0, \frac{\partial u_r}{\partial z} = 0, \frac{\partial u_\theta}{\partial z} = 0 && \text{at } z = 0, L.\end{aligned}\tag{4.2.35}$$

Eigenvectors of the Linearized Taylor Problem By the expressions (4.2.24) and (4.2.25) of transition solutions, we see that the first eigenvectors of linearized equations play a crucial role to determine the asymptotic structure of solutions for the Taylor problem. Hence start with the computation of the first eigenvalue and eigenvectors of (4.2.22) with either the Dirichlet boundary condition (4.2.34) or the free boundary condition (4.2.35), following (Chandrasekhar, 1981).

For the eigenvalue equation (4.2.22), we take the separation of variables as follows

$$(u_z, u_r, u_\theta) = \left(\frac{1}{a^2} \frac{dh(z)}{dz} \frac{dR(r)}{dr}, h(z)R(r), h(z)\varphi(r) \right), \quad (4.2.36)$$

where $a^2 > 0$ is an arbitrary constant.

By (4.2.22) and the boundary conditions $u_z = 0$ at $z = 0, L$ in (4.2.34) and (4.2.35), we obtain

$$\begin{aligned} \frac{d^2 h}{dz^2} &= -a^2 h, \\ h'(0) &= h'(L) = 0. \end{aligned} \quad (4.2.37)$$

Furthermore, the functions R and φ satisfy

$$\begin{aligned} \left(\frac{d^2}{dr^2} - a^2 \right)^2 R &= a^2 \lambda \varphi, \\ \left(\frac{d^2}{dr^2} - a^2 \right) \varphi &= -\lambda R. \end{aligned} \quad (4.2.38)$$

With the Dirichlet boundary condition (4.2.34), we find

$$(\varphi(r), R(r), R'(r)) = 0 \quad \text{at } r = r_1, r_2, \quad (4.2.39)$$

while with the free boundary (4.2.35), we have

$$(\varphi(r), R(r), R''(r)) = 0 \quad \text{at } r = r_1, r_2. \quad (4.2.40)$$

The solutions of (4.2.37) are given by

$$h(z) = \cos az \quad \text{with } a = k\pi/L, \quad k = 1, 2, \dots. \quad (4.2.41)$$

In the following we discuss the eigenvalue problems of (4.2.38) with one of the two boundary conditions (4.2.39) and (4.2.40).

DIRICHLET BOUNDARY CONDITION CASE: The eigenvalue problem of (4.2.38) with (4.2.39) is equivalent to

$$\begin{aligned} \left(\frac{d^2}{dr^2} - a^2 \right)^3 R &= -a^2 \lambda^2 R, \\ (R, R') &= 0, \quad \left(\frac{d^2}{dr^2} - a^2 \right)^2 R = 0 \quad \text{at } r = r_1, r_2. \end{aligned} \quad (4.2.42)$$

When the height L is large, the first eigenvalue and eigenvector of (4.2.42) are given in Chapter II-15 of (Chandrasekhar, 1981) by

$$\begin{aligned} \lambda_0^2 &\simeq 1700, \\ R(r) &\simeq \cos \alpha_0 \xi - 0.06 \cosh(\alpha_1 \xi) \cos \alpha_2 \xi + 0.1 \sinh(\alpha_1 \xi) \sin \alpha_2 \xi, \end{aligned} \quad (4.2.43)$$

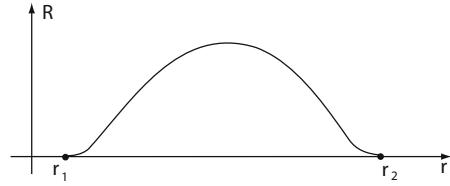
where $\xi = r - r_1 - 1/2$, $\alpha_0 \approx 3.97$, $\alpha \approx 5.2$, and $\alpha_3 \approx 2.1$. This function is illustrated in Fig. 4.9.

Thus, from (4.2.36), (4.2.38), and (4.2.41) we obtain the first eigenvector of (4.2.22) with (4.2.34) as follows

$$(u_z, u_r, u_\theta) = \left(-\frac{\sin az}{a} R'(r), \cos az R(r), \frac{\cos az}{a^2 \lambda_0} \left[\frac{d^2}{dz^2} - a^2 \right] R(r) \right), \quad (4.2.44)$$

where $R(r)$ is given by (4.2.43).

Fig. 4.9 Graph of the function $R(r)$



FREE BOUNDARY CONDITION CASE: For each given $a^2 = k^2\pi^2/L^2$, the first eigenvalue $\lambda_0(a)$ and eigenvectors of (4.2.38), with (4.2.40), are given by

$$\lambda_0(a) = \frac{(\pi^2 + a^2)^{3/2}}{a}, \quad (R, \varphi) = \sin \pi(r - r_1) \left(1, \frac{\sqrt{a^2 + \pi^2}}{a} \right). \quad (4.2.45)$$

It is clear that the first eigenvalue λ_0 of (4.2.22) with (4.2.35) is the minimum of $\lambda_0(a)$:

$$\lambda_0^2 = \min_a \lambda_0^2(a) = \min_{k \in \mathbb{N}} \frac{\pi^4 L^2 \left(1 + \frac{k^2}{L^2} \right)^3}{k^2}. \quad (4.2.46)$$

The corresponding first eigenvectors of (4.2.22) with (4.2.35) are derived from (4.2.36), (4.2.41), and (4.2.45), and are given by

$$\begin{aligned} u_z &= -\frac{\pi}{a} \sin az \cos(\pi(r - r_1)), \\ u_r &= \cos az \sin \pi(r - r_1), \\ u_\theta &= \frac{1}{a} \sqrt{\pi^2 + a^2} \cos az \sin \pi(r - r_1), \end{aligned} \quad (4.2.47)$$

where $a = k\pi/L$ satisfies (4.2.46).

Structural Stability of the First Eigenvectors It is easy to see that the first eigenvectors $\tilde{u} = (u_z, u_r)$ given by (4.2.44) and (4.2.47) have the structure of the Taylor vortices, and by Theorems 2.1.2 and 2.2.9 in Ma and Wang (2005d), we readily verify the structural stability of the vector fields $\tilde{u} = (u_z, u_r)$ of (4.2.44) and (4.2.47) in $B^r(M, \mathbb{R}^2)$ ($r \geq 1$) and $B_0^r(M, \mathbb{R}^2)$ respectively. Here

$$B^r(M, \mathbb{R}^2) = \{\tilde{u} \in C^r(M, \mathbb{R}^2) \mid \operatorname{div}\tilde{u} = 0, \tilde{u} \text{ satisfy (4.2.35)}\},$$

$$B_0^r(M, \mathbb{R}^2) = \{\tilde{u} \in C^r(M, \mathbb{R}^2) \mid \operatorname{div}\tilde{u} = 0, \tilde{u} \text{ satisfy (4.2.34)}\}.$$

The following two lemmas establish the structure and its stability for the first eigenvectors given in (4.2.44) and (4.2.47).

Lemma 4.2.6 *For the vector fields given by (4.2.44) and $\alpha \neq 0$,*

$$\tilde{u} = (u_r, u_z) = \left(\alpha \cos \frac{k\pi z}{L} R(r), -\frac{\alpha L}{k\pi} \sin \frac{k\pi z}{L} R'(r) \right) \quad (4.2.48)$$

are structurally stable in $B_0^r(M, \mathbb{R}^2)$, and are topologically equivalent to the flow structure as shown in Fig. 4.10.

Proof. By (4.2.43) simple calculation shows that

$$R''(r_i) \neq 0 \quad \text{for } i = 0, 1, 2,$$

where $r_0 = (r_1 + r_2)/2$ satisfies that $R'(r_0) = 0$.

It is easy to see that all the interior singular points of the vector field in (4.2.48) are given by

$$(r_0, z_j) = \left(r_0, \frac{jL}{2k} \right) \quad \forall j = 1, 3, \dots, 2m-1 \text{ and } m \leq k.$$

By direct computation, we find

$$\det \begin{pmatrix} \frac{\partial u_z}{\partial z} & \frac{\partial u_z}{\partial r} \\ \frac{\partial u_r}{\partial z} & \frac{\partial u_r}{\partial r} \end{pmatrix}_{(r_0, z_j)} = \frac{\alpha^2 k \pi}{L} R(r_0) R''(r_0) \neq 0.$$

The boundary singular points of (4.2.48) read

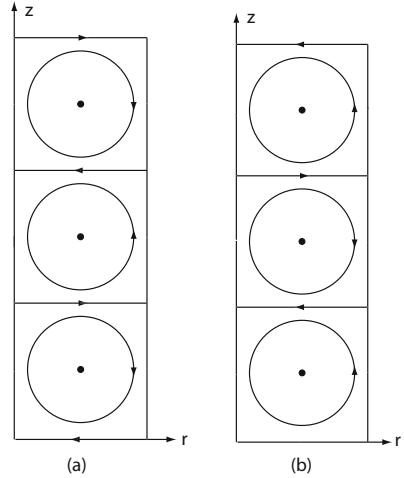
$$(r_i, \tilde{z}_j) = \left(r_i, \frac{jL}{k} \right) \quad \text{for } i = 1, 2 \text{ and } j = 1, 2, \dots, k,$$

and the determinant

$$\det \begin{pmatrix} \frac{\partial^2 u_z}{\partial z \partial r} & \frac{\partial^2 u_z}{\partial r^2} \\ \frac{\partial^2 u_r}{\partial z \partial r} & \frac{\partial^2 u_r}{\partial r^2} \end{pmatrix}_{(r_i, \tilde{z}_j)} = -\alpha^2 R'^{''2}(r_i) \neq 0.$$

Hence, the vector fields (4.2.48) are D -regular for $\alpha \neq 0$. Thus, this lemma follows from Theorem 2.2.9 in Ma and Wang (2005d). The proof is complete. \square

Fig. 4.10 Flow structure of (u_r, u_z) with boundary condition (4.2.34): (a) $\alpha < 0$, (b) $\alpha > 0$



Lemma 4.2.7 For the vector fields of (4.2.47) and $\alpha \neq 0$, the following vector fields

$$(u_r, u_z) = \left(\alpha \cos \frac{k\pi z}{L} \sin \pi(r - r_1), -\frac{L\alpha}{k} \sin \frac{k\pi z}{L} \cos \pi(r - r_1) \right) \quad (4.2.49)$$

are structurally stable in $B^k(M, \mathbb{R}^2)$, and are topologically equivalent to the flow structure as shown in Fig. 4.11.

Proof. The interior singular points of (4.2.49) are

$$(r_0, z_j) = \left(\frac{r_1 + r_2}{2}, \frac{jL}{2k} \right) \quad \forall j = 1, 3, \dots, 2m - 1, m \leq k,$$

and the boundary singular points read

$$(r_i, \tilde{z}_j) = \left(r_i, \frac{jL}{k} \right) \quad \forall i = 1, 2, j = 1, 2, \dots, k.$$

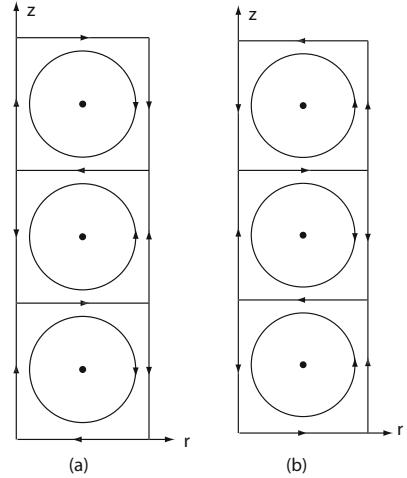
It is clear that the determinants of the Jacobian matrix $D\tilde{u}$ at the singular points (r_0, z_j) and (r_i, \tilde{z}_j) are nonzero:

$$\det D\tilde{u}(r, z) \neq 0 \quad \forall (r, z) = (r_0, z_j) \text{ and } (r_i, \tilde{z}_j).$$

Hence the vector fields (4.2.49) are regular. Thus this lemma follows from Theorem 2.1.2 in Ma and Wang (2005d). The proof is complete. \square

Taylor Vortex Structure of Solutions The following theorem shows that the solutions of the Taylor problem with the axisymmetric perturbation have the Taylor vortices as their asymptotic structure.

Fig. 4.11 Flow structure of (u_r, u_z) with boundary condition (4.2.35): (a) $\alpha < 0$, (b) $\alpha > 0$



Theorem 4.2.8 There are discrete numbers L_i ($i = 1, 2, \dots$) such that for the height $L \neq L_i$, if the Taylor number T satisfies $0 < T - T_c < \delta$ for some $\delta > 0$, where T_c is the critical Taylor number as in (4.2.21), then the following assertions hold true:

- (1) The space H can be decomposed into two open sets $H = \overline{U}_1^T + \overline{U}_2^T$ satisfying

$$U_1^T \cap U_2^T = \emptyset, \quad 0 \in \partial U_1^T \cap \partial U_2^T.$$

- (2) For any $\varphi \in U_1^T$, there is a time $t_0 \geq 0$ such that for the solution $u(t, \varphi) = (\tilde{u}, u_\theta)$ of (4.2.17) and (4.2.34) (resp. of (4.2.17) and (4.2.35)) with $u(0, \varphi) = \varphi, \tilde{u}(t, \varphi) = (u_r, u_z)$ is topologically equivalent to the structure as shown in Fig. 4.10a (resp. in Fig. 4.11a) for any $t > t_0$. Namely, the solution $u(t, \varphi)$ of (4.2.17) and (4.2.34) (resp. of (4.2.17) and (4.2.35)) has the Taylor vortex structure as $t > t_0$.
- (3) For any $\varphi \in U_2^T$, there is a $t_0 \geq 0$ such that the solution $\tilde{u}(t, \varphi)$ of (4.2.17) and (4.2.34) (resp. of (4.2.17) and (4.2.35)) is topologically equivalent to the structure as shown in Fig. 4.10b (resp. in Fig. 4.11b) for any $t > t_0$.

Proof. There are discrete countable numbers L_i ($i = 1, 2, \dots$) of the height L such that when $L \neq L_i$ the first eigenvalue of (4.2.22) with (4.2.34) is simple; see Yudovich (1966); Chandrasekhar (1981). When $L^2 \neq l^2$, where l^2 satisfy that for any integers k_1 and k_2 with $k_1 \neq k_2$,

$$k_2^2(l^2 + k_1^2)^3 = k_1^2(l^2 + k_2^2)^3,$$

we infer from (4.2.46) that the first eigenvalue of (4.2.22) with boundary condition (4.2.35) is also simple. Thus the theorem follows immediately from Theorem 4.2.3, Remark 4.2.4, and Lemmas 4.1.12–4.2.7. The proof is complete. \square

Remark 4.2.9 For any height $L > 0$, the multiplicity of the first eigenvalue of (4.2.22) with (4.2.34) or with (4.2.35) is either one or two and if it is two, then the two corresponding eigenvectors have the wave numbers $k_1 = k$ and $k_2 = k + 1$ in (4.2.41) for some $k \geq 1$ respectively.

In fact, in the method of construction of the center manifold functions as used in Theorems 4.1.14 and 4.2.12 in next subsection, we can prove the following results, and omit here the details of the proof.

Theorem 4.2.10 *For the Taylor problem (4.2.17) with boundary condition (4.2.34) or with (4.2.35) the following assertions hold true.*

- (1) *The transition of the Taylor problem at $T = T_c$ is of Type-I, i.e., the problem has an attractor bifurcation on $T > T_c$.*
- (2) *For any $L > 0$, the assertions of Theorem 4.2.8 hold true.*
- (3) *When $L = L_i$ ($i = 1, 2, \dots$) such that the first eigenvalue has multiplicity two, the Taylor problem bifurcates from $(u, T) = (0, T_c)$ to an attractor \mathcal{A}_T homeomorphic to S^1 , and $\mathcal{A}_T = S^1$ contains exactly eight singular points, four of which are minimal attractors, and other four of which are saddle points. The structure of the attractor \mathcal{A}_T is topological equivalent to that as shown in Fig. 4.7.*

4.2.5 Taylor Problem with z -Periodic Boundary Condition

We now study the Taylor problem (4.2.7) with the z -periodic boundary condition (4.2.14) and with axisymmetric perturbations. Assuming that the equations (4.2.7) are independent of θ , and taking the length scale $h = r_2$ in the nondimensional form, we obtain

$$\begin{aligned} \frac{\partial u_z}{\partial t} &= \Delta u_z - \frac{\partial p}{\partial z} - (\tilde{u} \cdot \nabla) u_z, \\ \frac{\partial u_r}{\partial t} &= (\Delta - \frac{1}{r^2}) u_r + \lambda \left(\frac{1}{r^2} - \kappa \right) u_\theta - \frac{\partial p}{\partial r} + \frac{u_\theta^2}{r} - (\tilde{u} \cdot \nabla) u_r, \\ \frac{\partial u_\theta}{\partial t} &= \left(\Delta - \frac{1}{r^2} \right) u_\theta + \lambda \kappa u_r - \frac{u_r u_\theta}{r} - (\tilde{u} \cdot \nabla) u_\theta, \\ \frac{\partial(r u_z)}{\partial z} + \frac{\partial(r u_r)}{\partial r} &= 0, \end{aligned} \tag{4.2.50}$$

where $(\tilde{u} \cdot \nabla)$ is as in (4.2.17), $\lambda = \sqrt{T}$, T is the Taylor number, and

$$\begin{aligned} T &= \frac{4r_2^4 \Omega_1^2 (1 - \mu)^2 \eta^4}{\nu^2 (1 - \eta^2)^2}, & \eta^2 &= r_1^2 / r_2^2, \\ \kappa &= \frac{1 - \mu / \eta^2}{1 - \mu}, & \Delta &= \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{\partial^2}{\partial z^2}. \end{aligned}$$

The nondimensional domain is $M = (\eta, 1) \times (0, L)$, and the boundary conditions take (4.2.10) and (4.2.14), i.e.,

$$\begin{aligned} u = (u_z, u_r, u_\theta) &= 0 \quad \text{at } r = \eta, 1, \\ u &\text{ is periodic with period } L \text{ in the } z\text{-direction.} \end{aligned} \quad (4.2.51)$$

The initial value condition is

$$u = u_0(r, z) \quad \text{at } t = 0. \quad (4.2.52)$$

Functional Setting For the Taylor problems (4.2.50)–(4.2.52), we set

$$\begin{aligned} H &= \left\{ u = (\tilde{u}, u_\theta) \in L^2(M)^3 \left| \begin{array}{l} \operatorname{div}(r\tilde{u}) = 0, u_r = 0 \text{ at } r = \eta, 1, \text{ and} \\ u \text{ is L-periodic in the } z\text{-direction} \end{array} \right. \right\} \\ H_1 &= \left\{ u \in H^2(M)^3 \bigcap H \mid u \text{ satisfies (4.2.51)} \right\}, \end{aligned}$$

and the inner product of H is defined by

$$(u, v) = \int_M u \cdot v r dz dr.$$

Let the linear operator $L_\lambda = -A + \lambda B : H_1 \rightarrow H$ and nonlinear operator $G : H_1 \rightarrow H$ be defined by

$$\begin{aligned} Au &= -P \left(\Delta u_z, \left(\Delta - \frac{1}{r^2} \right) u_r, \left(\Delta - \frac{1}{r^2} \right) u_\theta \right), \\ Bu &= P \left(0, \left(\frac{1}{r^2} - \kappa \right) u_\theta, \kappa u_r \right), \\ G(u) &= -P \left((\tilde{u} \cdot \nabla) u_z, (\tilde{u} \cdot \nabla) u_r - \frac{u_\theta^2}{r}, (\tilde{u} \cdot \nabla) u_\theta + \frac{u_\theta u_r}{r} \right), \end{aligned} \quad (4.2.53)$$

where $P : L^2(M)^3 \rightarrow H$ is the Leray projection. Thus the Taylor problems (4.2.50)–(4.2.52) are rewritten in the abstract form

$$\begin{aligned} \frac{du}{dt} &= L_\lambda u + G(u), \\ u(0) &= u_0. \end{aligned} \quad (4.2.54)$$

As before, we also use G to denote the corresponding bilinear operator.

Eigenvalue Problem To study the phase transition of the Taylor problems (4.2.50)–(4.2.52) it is necessary to consider the eigenvalue problem of its linearized equation.

Let L_λ be the linear operator defined by (4.2.53) and L_λ^* be its conjugate operator. The associated eigenvalue equation of (4.2.54) reads

$$L_\lambda u = -Au + \lambda Bu = \beta(\lambda)u, \quad (4.2.55)$$

and the conjugate equation of (4.2.55) is given by

$$L_\lambda^* u^* = -A^* u^* + \lambda B^* u^* = \beta(\lambda) u^*. \quad (4.2.56)$$

The equations corresponding to (4.2.55) are as follows

$$\begin{aligned} \Delta u_z - \frac{\partial p}{\partial z} &= \beta(\lambda) u_z, \\ \left(\Delta - \frac{1}{r^2} \right) u_r + \lambda \left(\frac{1}{r^2} - \kappa \right) u_\theta - \frac{\partial p}{\partial r} &= \beta(\lambda) u_r, \\ \left(\Delta - \frac{1}{r^2} \right) u_\theta + \lambda \kappa u_r &= \beta(\lambda) u_\theta, \\ \operatorname{div}(r \tilde{u}) &= 0. \end{aligned} \quad (4.2.57)$$

The equations corresponding to (4.2.56) are given by

$$\begin{aligned} \Delta u_z^* - \frac{\partial p^*}{\partial z} &= \beta(\lambda) u_z^*, \\ \left(\Delta - \frac{1}{r^2} \right) u_r^* + \lambda \kappa u_\theta^* - \frac{\partial p^*}{\partial r} &= \beta(\lambda) u_r^*, \\ \left(\Delta - \frac{1}{r^2} \right) u_\theta^* + \lambda \left(\frac{1}{r^2} - \kappa \right) u_r^* &= \beta(\lambda) u_\theta^*, \\ \operatorname{div}(r \tilde{u}^*) &= 0. \end{aligned} \quad (4.2.58)$$

Both (4.2.57) and (4.2.58) are supplemented with the boundary condition (4.2.51).

By Theorem 2.1.3, the phase transition of hydrodynamic equations require the validity of (2.1.4) and (2.1.5), also called in physical literature the principle of exchange of stability (PES). In the following, we shall verify the PES for the Taylor problem.

It is known that for each given period L , there is a $\lambda_0^* = \lambda_0(L)$ such that the eigenvalues $\beta_j(\lambda)$ ($j = 1, 2, \dots$) of (4.2.57) with (4.2.51) near $\lambda = \lambda_0^*$ satisfy that $\beta_1(\lambda), \dots, \beta_m(\lambda)$ ($m \geq 1$) are real, and

$$\begin{aligned} \beta_i(\lambda) &\begin{cases} < 0 & \text{if } \lambda < \lambda_0^*, \\ = 0 & \text{if } \lambda = \lambda_0^* \end{cases} \quad \text{for } 1 \leq i \leq m, \\ \operatorname{Re}\beta_j(\lambda_0^*) &< 0 \quad \text{for } j \geq m+1. \end{aligned} \quad (4.2.59)$$

In addition, there is a period $L' > 0$ such that

$$\lambda_0 = \lambda_0(L') = \min_{L>0} \lambda_0(L). \quad (4.2.60)$$

Thanks to (Yudovich, 1966; Velte, 1966), for $\mu = \Omega_1/\Omega_2 \geq 0$ the multiplicity $m = 2$ in (4.2.59) at $\lambda_0 = \lambda_0(L')$; see also Kirchgässner (1975); Temam (1984).

In this subsection, we always take L' as the period given by (4.2.60), and define the following number as the critical Taylor number:

$$T_c = \lambda_0^2(L').$$

For simplicity, omitting the prime we denote L' by L .

By (4.2.59) and (4.2.60), to verify the PES it suffices to prove that for $\lambda > \lambda_0$,

$$\beta_i(\lambda) > 0 \quad \forall 1 \leq i \leq m. \quad (4.2.61)$$

To this end, by Theorem 2.1 in Ma and Wang (2008d), we need to derive the eigenvectors of (4.2.57) and (4.2.58) at $\beta_i(\lambda_0) = 0$ ($i = 1, 2$).

It is ready to check that the eigenvectors of (4.2.57) with (4.2.51) corresponding to $\beta_i(\lambda_0) = 0$ ($i = 1, 2$) are given by

$$\psi_1 = (\psi_z, \psi_r, \psi_\theta) = (-\sin az D_* h(r), a \cos az h(r), \cos az \varphi(r)), \quad (4.2.62)$$

$$\tilde{\psi}_1 = (\tilde{\psi}_z, \tilde{\psi}_r, \tilde{\psi}_\theta) = (\cos az D_* h(r), a \sin az h(r), \sin az \varphi(r)), \quad (4.2.63)$$

where $(h(r), \varphi(r))$ satisfies

$$\begin{aligned} (DD_* - a^2)^2 h &= a^2 \lambda_0 \left(\frac{1}{r^2} - \kappa \right) \varphi, \\ (DD_* - a^2) \varphi &= -\lambda_0 \kappa h, \\ (h, Dh, \varphi) &= 0 \quad \text{at } r = \eta, 1, \end{aligned} \quad (4.2.64)$$

and

$$D = \frac{d}{dr}, \quad D_* = \frac{d}{dr} + \frac{1}{r}, \quad a = \frac{2\pi}{L}.$$

The dual eigenvectors of (4.2.58) with (4.2.51) read

$$\psi_1^* = (\psi_z^*, \psi_r^*, \psi_\theta^*) = (-\sin az D_* h^*(r), a \cos az h^*(r), \cos az \varphi^*(r)), \quad (4.2.65)$$

$$\tilde{\psi}_1^* = (\tilde{\psi}_z^*, \tilde{\psi}_r^*, \tilde{\psi}_\theta^*) = (\cos az D_* h^*(r), a \sin az h^*(r), \sin az \varphi^*(r)), \quad (4.2.66)$$

where (h^*, φ^*) satisfies that

$$\begin{aligned} (DD_* - a^2)^2 h^* &= \lambda_0 \kappa \varphi^*, \\ (DD_* - a^2) \varphi^* &= -a^2 \lambda_0 \left(\frac{1}{r^2} - \kappa \right) h^*, \\ (h^*, Dh^*, \varphi^*) &= 0 \quad \text{at } r = \eta, 1. \end{aligned} \quad (4.2.67)$$

The following lemma shows that the PES is valid for the Taylor problems (4.2.50)–(4.2.52) with $\mu \geq 0$.

Lemma 4.2.11 *If $\mu \geq 0$, then the first eigenvalues $\beta_i(\lambda)$ ($1 \leq i \leq m$) of (4.2.57) are real with multiplicity $m = 2$ near $\lambda = \lambda_0 = \sqrt{T_c}$, and the first eigenvectors at $\lambda = \lambda_0$ are given by (4.2.62) and (4.2.63). Moreover, the eigenvalues $\beta_j(\lambda)$ ($j = 1, 2, \dots$)*

satisfy the conditions (5.4) and (5.5) at $\lambda = \lambda_0$, i.e., the PES holds true at the critical Taylor number T_c .

Proof. We only need to prove (4.2.61). By Theorem 2.1 in Ma and Wang (2008d), it suffices to verify that

$$(B\psi_1, \psi_1^*) \neq 0, \quad (B\tilde{\psi}_1, \tilde{\psi}_1^*) \neq 0. \quad (4.2.68)$$

We infer from (4.2.53), (4.2.62), (4.2.63), (4.2.65), and (4.2.66) that

$$\begin{aligned} (B\psi_1, \psi_1^*) &= (B\tilde{\psi}_1, \tilde{\psi}_1^*) \\ &= \int_0^L \int_\eta^1 r \left[\left(\frac{1}{r^2} - \kappa \right) \psi_\theta \psi_r^* + \kappa \psi_r \psi_\theta^* \right] dz dr \\ &= \frac{La}{2} \int_\eta^1 r \left[\left(\frac{1}{r^2} - \kappa \right) h\varphi^* + \kappa\varphi h^* \right] dr. \end{aligned} \quad (4.2.69)$$

Since $\mu \geq 0$, by (4.2.4), we have $0 < \kappa < 1$ and $\frac{1}{r^2} - \kappa > 0$ for $\eta < r < 1$. On the other hand, we know that the first eigenvectors $(h(r), \varphi(r))$ of (4.2.64) and $(h^*(r), \varphi^*(r))$ of (4.2.67) at $\lambda = \lambda_0$ are positive; see Velte (1966); Kirchgässner (1975); Temam (1984):

$$h(r) > 0, \quad \varphi(r) > 0, \quad h^*(r) > 0, \quad \varphi^*(r) > 0 \quad \forall \eta < r < 1. \quad (4.2.70)$$

Thus (4.2.68) follows from (4.2.69) and (4.2.70). The proof is complete. \square

Phase Transition Theorems Here we always assume that the first eigenvalue of (4.2.57) with (4.2.51) is real with multiplicity $m = 2$, i.e., the first eigenvalue λ_0 of (4.2.64) is simple, and the PES holds true. By Lemma 4.2.11, this assumption is valid for all $\mu \geq 0$ and $0 < \eta < 1$.

Let ψ_1 and ψ_1^* be given by (4.2.62) and (4.2.65). We define a number R by

$$R = \frac{1}{(\psi_1, \psi_1^*)} [(G(\Phi, \psi_1), \psi_1^*) + (G(\psi_1, \Phi), \psi_1^*)], \quad (4.2.71)$$

where $\Phi \in H_1$ is defined by

$$(A - \lambda_0 B)\Phi = G(\psi_1, \psi_1). \quad (4.2.72)$$

Here the operator A, B and G are as in (4.2.53). The solution Φ of (4.2.72) exists because $G(\psi, \psi_1)$ is orthogonal with ψ_1^* and $\tilde{\psi}_1^*$ in H .

The following results characterize the dynamical properties of phase transitions for the Taylor problem with the z -periodic boundary condition.

Theorem 4.2.12 *If the number $R < 0$ in (4.2.71), then the Taylor problems (4.2.50)–(4.2.52) have a Type-I (continuous) transition at the critical Taylor number $T = T_c$ or $\lambda = \lambda_0$, and the following assertions hold true:*

- (1) When the Taylor number $T \leq T_c$ or $\lambda \leq \lambda_0$, the steady-state $u = 0$ is locally asymptotically stable.
- (2) The problem bifurcates from $(u, \lambda) = (0, \lambda_0)$ (or from $(u, T) = (0, T_c)$) to an attractor \mathcal{A}_λ homeomorphic to a circle S^1 on $\lambda_0 < \lambda$, which consists of steady states of this problem.
- (3) Any $u \in \mathcal{A}_\lambda$ can be expressed as

$$u = |\beta_1(\lambda)/R|^{1/2}v + o(|\beta_1|^{1/2}), \quad v = x\psi_1 + y\tilde{\psi}_1, \quad x^2 + y^2 = 1,$$

where $\psi_1, \tilde{\psi}_1$ are given in (4.2.62) and (4.2.65).

- (4) There is an open set $U \subset H$ with $0 \in U$ such that \mathcal{A}_λ attracts $U \setminus \Gamma$, where Γ is the stable manifold of $u = 0$ with codimension two in H .
- (5) When $1 - \mu > 0$ is small, for any $u_0 \in U \setminus (\Gamma \cup \tilde{H})$, there exists a time $t_0 \geq 0$ such that for any $t > t_0$, the vector field $\tilde{u}(t, u_0) = (u_z, u_r)$ is topologically equivalent to one of the patterns shown in Fig. 4.12, where $u = (\tilde{u}(t, u_0), u_\theta(t, u_0))$ is the solution of (4.2.50)–(4.2.52), and

$$\tilde{H} = \{u = (u_z, u_r, u_\theta) \in H \mid \int_0^L \int_\eta^1 r u_z dr dz = 0\}.$$

- (6) When $1 - \mu > 0$ is small, for any $u_0 \in (U \cap \tilde{H}) \setminus \Gamma$, there exists a time $t_0 \geq 0$ such that for any $t > t_0$, $\tilde{u}(t, u_0) = (u_z, u_r)$ is topologically equivalent to the structure as shown in Fig. 4.13.

Fig. 4.12 Taylor vortices with a cross-channel flow

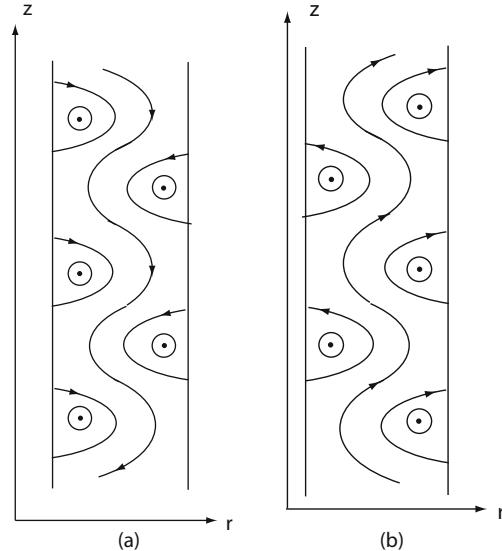
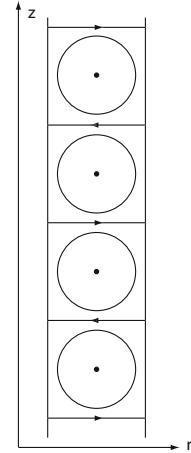


Fig. 4.13 Taylor vortices without a cross-channel flow



Theorem 4.2.13 For the case where $R > 0$, the transition of the Taylor problems (4.2.50)–(4.2.52) at $T = T_c$ is of Type-II. Moreover, the Taylor problem has a singularity separation at $T^* < T_c$ ($\lambda^* < \lambda_0$). More precisely we have the following assertions:

- (1) There exists a number λ^* ($0 < \lambda^* < \lambda_0$) such that the problem generates a circle $\Sigma^* = S^1$ at $\lambda = \lambda^*$ consisting of singular points, and bifurcates from (Σ^*, λ^*) on $\lambda^* < \lambda$ to at least two branches of circles Σ_1^λ and Σ_2^λ , each consisting of steady states satisfying (see Fig. 4.14):

$$\lim_{\lambda \rightarrow \lambda_0} \Sigma_\lambda^1 = \{0\}, \quad \text{dist}\left(\Sigma_2^\lambda, 0\right) = \min_{u \in \Sigma_2^\lambda} \|u\| > 0 \quad \text{at} \quad \lambda = \lambda_0.$$

- (2) For each $\lambda^* < \lambda < \lambda_0$, the space H can be decomposed into two open sets U_1^λ and U_2^λ : $H = \overline{U_1^\lambda} + \overline{U_2^\lambda}$ with $U_1^\lambda \cap U_2^\lambda = \emptyset$, $\Sigma_1^\lambda \subset \partial U_1^\lambda \cap \partial U_2^\lambda$ such that the problem has two disjoint attractors \mathcal{A}_1^λ and \mathcal{A}_2^λ :

$$\mathcal{A}_1^\lambda = \{0\} \subset U_1^\lambda, \quad \Sigma_2^\lambda \subset \mathcal{A}_2^\lambda \subset U_2^\lambda,$$

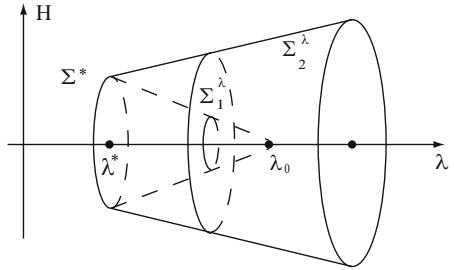
- and \mathcal{A}_i^λ attracts U_i^λ ($i = 1, 2$).
- (3) For $\lambda_0 \leq \lambda$, the problem has an attractor \mathcal{A}^λ satisfying

$$\lim_{\lambda \rightarrow \lambda_0} \mathcal{A}_2^\lambda = \mathcal{A}^{\lambda_0}, \quad \text{dist}(\mathcal{A}^\lambda, 0) > 0 \quad \forall \lambda \geq \lambda_0,$$

and \mathcal{A}^λ attracts $H \setminus \Gamma_\lambda$, where Γ_λ is the stable manifold of $u = 0$ with codimension $m_\lambda \geq 2$ in H .

Proof of Theorem 4.2.12. We proceed as follows.

Fig. 4.14 Singularity separation of circles consisting of steady states at $\lambda = \lambda^*$



STEP 1. We claim that the problems (4.2.50)–(4.2.52) bifurcate from $(u, \lambda) = (0, \lambda_0)$ to a circle S^1 which consists of steady states.

It is easy to see that the problem (4.2.50) with (4.2.51) is invariant for the transition in the z -direction: $u(z, r, t) \rightarrow u(z + z_0, r, t)$ for $z_0 \in \mathbb{R}^1$. Therefore, if u_0 is a steady-state solution of (4.2.50) with (4.2.51), then for any $z_0 \in \mathbb{R}^1$ the function $u_0(z + z_0, r)$ is also a steady-state solution. We can see that the set

$$\Sigma = \{u_0(z + z_0, r) \mid z_0 \in \mathbb{R}^1\}$$

is homeomorphic to a circle S^1 in H_1 for any $u_0 \in H_1$. Hence, the singular points of (4.2.50) with (4.2.51) appear as a circle.

It is known in Yudovich (1966), Velte (1966) that there exist singular points bifurcated from $(u, \lambda) = (0, \lambda_0)$. Thus this claim is proved.

STEP 2. REDUCTION TO THE CENTER MANIFOLD. By the spectral theorem (Theorem 3.4 in Ma and Wang (2005b)), the reduced equations of (4.2.54) to the first eigenspace are given by

$$\begin{aligned} \frac{dx}{dt} &= \beta_1(\lambda)x + \frac{(G(u), \psi_{1\lambda}^*)}{(\psi_{1\lambda}, \psi_{1\lambda}^*)}, \\ \frac{dy}{dt} &= \beta_1(\lambda)y + \frac{(G(u), \tilde{\psi}_{1\lambda}^*)}{(\tilde{\psi}_{1\lambda}, \tilde{\psi}_{1\lambda}^*)}, \end{aligned} \quad (4.2.73)$$

where $\psi_{1\lambda}$ and $\tilde{\psi}_{1\lambda}$ are the eigenvectors of (4.2.57) corresponding to $\beta_1(\lambda)$ near $\lambda = \lambda_0$ with

$$\begin{aligned} \lim_{\lambda \rightarrow \lambda_0} \psi_{1\lambda} &= \psi_1 \quad (\psi_1 \text{ as in (4.2.62)}), \\ \lim_{\lambda \rightarrow \lambda_0} \tilde{\psi}_{1\lambda} &= \tilde{\psi}_1 \quad (\tilde{\psi}_1 \text{ as in (4.2.63)}), \end{aligned} \quad (4.2.74)$$

and $\psi_{1\lambda}^*$ and $\tilde{\psi}_{1\lambda}^*$ are the dual eigenvectors of $\psi_{1\lambda}$ and $\tilde{\psi}_{1\lambda}$ satisfying

$$\begin{aligned} \lim_{\lambda \rightarrow \lambda_0} \psi_{1\lambda}^* &= \psi_1^* \quad (\psi_1^* \text{ as in (4.2.65)}), \\ \lim_{\lambda \rightarrow \lambda_0} \tilde{\psi}_{1\lambda}^* &= \tilde{\psi}_1^* \quad (\tilde{\psi}_1^* \text{ as in (4.2.66)}). \end{aligned} \quad (4.2.75)$$

Let $\Psi : E_0 \rightarrow E_0^\perp$ be the center manifold function of (4.2.54) at $\lambda = \lambda_0$, where

$$E_0 = \text{span}\{\psi_1, \tilde{\psi}_1\}, \quad E_0^\perp = \{u \in H \mid (u, \psi_1^*) = 0, (u, \tilde{\psi}_1^*) = 0\}.$$

Let $u_0 = x\psi_1 + y\tilde{\psi}_1 \in E_0$. Then it is easy to check that $G(u_0) \in E_0^\perp$. Hence, by Theorem A.1.1 we find that

$$\begin{aligned} \Psi &= \phi(x, y) + o(|x|^2 + |y|^2) + O(\beta_1(\lambda)(|x|^2 + |y|^2)), \\ -L_\lambda\phi &= G(\psi_1, \psi_1)x^2 + G(\tilde{\psi}_1, \tilde{\psi}_1)y^2 + (G(\psi_1, \tilde{\psi}_1) + G(\tilde{\psi}_1, \psi_1))xy. \end{aligned} \quad (4.2.76)$$

On the center manifold, $u = u_0 + \Phi(u_0)$, therefore from (4.2.73)–(4.2.76) we obtain the reduced equations of (4.2.54) to the center manifold as follows:

$$\begin{aligned} \frac{dx}{dt} &= \beta_1(\lambda)x + \frac{x}{\rho}(G(\phi, \psi_1) + G(\psi_1, \phi), \psi_1^*) \\ &\quad + \frac{y}{\rho}(G(\phi, \tilde{\psi}_1) + G(\tilde{\psi}_1, \phi), \psi_1^*) \\ &\quad + o(|x|^3 + |y|^3) + \varepsilon_1(\lambda)O(|x|^3 + |y|^3), \\ \frac{dy}{dt} &= \beta_1(\lambda)y + \frac{x}{\rho}(G(\phi, \psi_1) + G(\psi_1, \phi), \tilde{\psi}_1^*) \\ &\quad + \frac{y}{\rho}(G(\phi, \tilde{\psi}_1) + G(\tilde{\psi}_1, \phi), \tilde{\psi}_1^*) \\ &\quad + o(|x|^3 + |y|^3) + \varepsilon_2(\lambda)O(|x|^3 + |y|^3), \end{aligned} \quad (4.2.77)$$

where $\rho = (\psi_1, \psi_1^*) = (\tilde{\psi}_1, \tilde{\psi}_1^*)$, and $\lim_{\lambda \rightarrow \lambda_0} \varepsilon_i(\lambda) = 0$ for $i = 1, 2$.

Furthermore, direct calculation shows that

$$\begin{aligned} G(\psi_1, \psi_1) &= P\psi_0 + P\psi_2, & G(\tilde{\psi}_1, \tilde{\psi}_1) &= P\psi_0 - P\psi_2, \\ G(\psi_1, \tilde{\psi}_1) &= P\tilde{\psi}_0 + P\tilde{\psi}_2, & G(\tilde{\psi}_1, \psi_1) &= -P\tilde{\psi}_0 + P\tilde{\psi}_2. \end{aligned}$$

where $P : L^2(M)^3 \rightarrow H$ is the Leray projection, and

$$\begin{aligned} \psi_0 &= -\left(0, \frac{1}{2}(a^2 h D_* h + a^2 h D h - \frac{1}{r} \varphi^2), \frac{a}{2}(\varphi D_* h + h D \varphi + \frac{1}{r} h \varphi)\right), \\ \psi_2 &= -\left(\frac{a}{2} \sin 2az((D_* h)^2 - h D D_* h), \frac{1}{2} \cos 2az(a^2 h D h - a^2 h D_* h - \frac{1}{r} \varphi^2), \right. \\ &\quad \left.\frac{a}{2} \cos 2az(h D \varphi - \varphi D_* h + \frac{1}{r} \varphi h)\right), \\ \tilde{\psi}_0 &= -\left(\frac{a}{2}((D_* h)^2 + h D D_* h), 0, 0\right), \\ \tilde{\psi}_2 &= -\left(-\frac{a}{2} \cos 2az((D_* h)^2 - h D D_* h), \frac{1}{2} \sin 2az(a^2 h D h - a^2 h D_* h - \frac{1}{r} \varphi^2), \right. \\ &\quad \left.\frac{a}{2} \sin 2az(h D \varphi - \varphi D_* h + \frac{1}{r} \varphi h)\right), \end{aligned}$$

$$\frac{a}{2} \sin 2az(hD\varphi - \varphi D_* h + \frac{1}{r}\varphi h) \Big).$$

Thus, (4.2.76) is rewritten as

$$(A - \lambda_0 B)\phi = (x^2 + y^2)P\psi_0 + (x^2 - y^2)P\psi_2 + 2xyP\tilde{\psi}_2. \quad (4.2.78)$$

Let

$$\phi = -[(x^2 + y^2)\phi_0 + (x^2 - y^2)\phi_2 + 2xy\tilde{\phi}_2], \quad (4.2.79)$$

$$\begin{cases} \phi_0 = (0, 0, \varphi_0), \\ \phi_2 = (-1/2 \sin 2az\varphi_z, \cos 2az\varphi_r, \cos 2az\varphi_\theta), \\ \tilde{\phi}_2 = (1/2 \cos 2az\tilde{\varphi}_z, \sin 2az\tilde{\varphi}_r, \sin 2az\tilde{\varphi}_\theta). \end{cases} \quad (4.2.80)$$

Then we deduce from (4.2.78) and (4.2.79) that

$$\varphi_z = \tilde{\varphi}_z, \quad \varphi_r = \tilde{\varphi}_r, \quad \varphi_\theta = \tilde{\varphi}_\theta,$$

and $(\varphi_z, \varphi_r, \varphi_\theta)$ satisfies

$$\begin{aligned} (DD_* - 4a^2)\varphi_r + 4a^2\lambda_0(\frac{1}{r^2} - \kappa)\varphi_\theta &= 4a^2H_2 + 2aDH_1, \\ (DD_* - 4a^2)\varphi_\theta + \lambda_0\kappa\varphi_r &= H_3, \\ \varphi_z &= \frac{1}{2a}D_*\varphi_r, \\ \varphi_z &= 0, \varphi_\theta = 0, \varphi_r = D\varphi_r = 0 \quad \text{at } r = \eta, 1, \end{aligned}$$

where H_1, H_2 , and H_3 are given by

$$\begin{aligned} H_1 &= a((D_*h)^2 - hDD_*h), \\ H_2 &= a^2hDh - a^2hD_*h - \frac{1}{r}\varphi^2, \\ H_3 &= a(hD\varphi - \varphi D_*h + \frac{1}{r}\varphi h). \end{aligned} \quad (4.2.81)$$

Based on (4.2.80) we find

$$\begin{aligned} (G(\tilde{\psi}_1, \phi_i) + G(\phi_i, \tilde{\psi}_1), \psi_1^*) &= 0 \quad \text{for } i = 0, 2, \\ (G(\psi_1, \phi_i) + G(\phi_i, \psi_1), \tilde{\psi}_1^*) &= 0 \quad \text{for } i = 0, 2, \\ (G(\tilde{\psi}_1, \tilde{\phi}_2) + G(\tilde{\phi}_2, \tilde{\psi}_1), \tilde{\psi}_1^*) &= 0, \\ (G(\psi_1, \tilde{\phi}_2) + G(\tilde{\phi}_2, \psi_1), \psi_1^*) &= 0. \end{aligned}$$

Then, putting (4.2.79) into (4.2.77), we deduce that

$$\begin{aligned}
\frac{dx}{dt} &= \beta_1 x - \frac{1}{\rho} x(x^2 + y^2)(G(\phi_0, \psi_1) + G(\psi_1, \phi_0), \psi^*)_1 \\
&\quad - \frac{1}{\rho} x(x^2 - y^2)(G(\phi_2, \psi_1) + G(\psi_1, \phi_2), \psi_1^*) \\
&\quad - \frac{2}{\rho} xy^2(G(\tilde{\phi}_2, \tilde{\psi}_1) + G(\tilde{\psi}_1, \tilde{\phi}_2), \psi_1^*) \\
&\quad + o(|x|^3 + |y|^3) + \varepsilon_1(\lambda)O(|x|^3 + |y|^3), \\
\frac{dy}{dt} &= \beta_1 y - \frac{1}{\rho} y(x^2 + y^2)(G(\phi_0, \tilde{\psi}_1) + G(\tilde{\psi}_1, \phi_0), \tilde{\psi}_1^*) \\
&\quad - \frac{1}{\rho} y(x^2 - y^2)(G(\phi_2, \tilde{\psi}_1) + G(\tilde{\psi}_1, \phi_2), \tilde{\psi}_1^*)_h \\
&\quad - \frac{2}{\rho} yx^2(G(\tilde{\phi}_2, \psi_1) + G(\psi_1, \tilde{\phi}_2), \tilde{\psi}_1^*) \\
&\quad + o(|x|^2 + |y|^2) + \varepsilon_2(\lambda)O(|x|^3 + |y|^3).
\end{aligned} \tag{4.2.82}$$

Direct computation yields

$$\begin{aligned}
(G(\phi_2, \psi_1) + G(\psi_1, \phi_2), \psi_1^*) &= (G(\tilde{\phi}_2, \tilde{\psi}_1) + G(\tilde{\psi}_1, \tilde{\phi}_2), \psi_1^*) \\
&= (G(\phi_2, \tilde{\psi}_1) + G(\tilde{\psi}_1, \phi_2), \tilde{\psi}_1^*) \\
&= (G(\tilde{\phi}_2, \psi_1) + G(\psi_1, \tilde{\phi}_2), \tilde{\psi}_1^*).
\end{aligned}$$

Hence (4.2.82) can be rewritten as

$$\begin{aligned}
\frac{dx}{dt} &= \beta_1 x + Rx(x^2 + y^2) + o(|x|^3 + |y|^3) + \varepsilon_1(\lambda)O(|x|^3 + |y|^3), \\
\frac{dy}{dt} &= \beta_1 y + Ry(x^2 + y^2) + o(|x|^3 + |y|^3) + \varepsilon_2(\lambda)O(|x|^3 + |y|^3),
\end{aligned} \tag{4.2.83}$$

where

$$R = -\frac{1}{\rho}(G(\phi_0 + \phi_2, \psi_1) + G(\psi_1, \phi_2 + \phi_2), \psi_1^*). \tag{4.2.84}$$

On the other hand, we infer from (4.2.78) and (4.2.80) that

$$(A - \lambda_0 B)\phi_i = P\psi_i \quad \text{for } i = 0, 2.$$

Hence we find

$$\Phi = -(\phi_0 + \phi_2), \quad (A - \lambda_0 B)\Phi = G(\psi_1, \psi_1) = P\psi_0 + P\psi_2.$$

Thus the number (4.2.84) is the same as that in (4.2.71).

STEP 3. PROOF OF ASSERTIONS (1)–(4). When $R < 0$, $(x, y) = 0$ is locally asymptotically stable for (4.2.83) at $\lambda = \lambda_0$. Therefore $u = 0$ is a locally asymptotically stable singular point of (4.2.54). By Theorem 2.2.11, the problem (4.2.50)–(4.2.52) bifurcates from $(u, \lambda) = (0, \lambda_0)$ to an attractor \mathcal{A}_λ which attracts an open set $U \setminus \Gamma$, and Assertions (1), (3), and (4) hold true.

In addition, the nonlinear terms in (4.2.83) satisfy the condition (2.2.6). Hence by the S^1 -attractor bifurcation theorem (Theorem 2.2.4) and the conclusion in Step 1, we derive Assertion (2).

STEP 4. ATTRACTION IN C^r -NORM. It is known that for any initial value $u_0 \in H$ there is a time $t_0 > 0$ such that the solution $u(t, u_0)$ of (4.2.50)–(4.2.52) is analytic for $t > t_0$, and uniformly bounded in C^r -norm for any $r \geq 1$; see also Lemma 4.1.12. Hence, by Assertion (4), for any $u_0 \in U \setminus \Gamma$ we have

$$\lim_{t \rightarrow \infty} \min_{v_0 \in \mathcal{A}_\lambda} \|u(t, u_0) - v_0\|_{C^r} = 0. \quad (4.2.85)$$

STEP 5. STRUCTURE OF SOLUTIONS IN \mathcal{A}_λ . By Assertion (3), for any steady-state solution $u_0 = (u_z, u_r, u_\theta) \in \mathcal{A}_\lambda$, the vector field $\tilde{u} = (u_z, u_r)$ of u_0 can be expressed as

$$\begin{aligned} u_z &= \gamma \cos a(z + z_0) D_* h(r) + w_1(z, r, \beta_1), \\ u_r &= a\gamma \sin a(z + z_0) h(r) + w_2(z, r, \beta_1), \end{aligned} \quad (4.2.86)$$

for some $z_0 \in \mathbb{R}^1$, where

$$\gamma = |\beta_1(\lambda)/R|^{1/2}, \quad w_i = o(|\beta_1|^{1/2}) \quad \text{for } i = 1, 2.$$

When $1 - \mu$ is small, $r_1 \gg 1$. In this case, we have

$$D_* h(r) \simeq h'(r), \quad h(r) \simeq h_0(r),$$

where $h_0(r)$, as $R(r)$, satisfies (4.2.38) with (4.2.39). Therefore, as in the proof of Lemma 4.2.6, we deduce that the vector field (4.2.86) is D -regular for all $0 < \lambda - \lambda_0 < \varepsilon$ for some $\varepsilon > 0$. Moreover, the first-order vector field in (4.2.86)

$$(v_z, v_r) = (\gamma \cos a(z + z_0) D_* h(r), a\gamma \sin a(z + z_0) h(r)) \quad (4.2.87)$$

has the topological structure as shown in Fig. 4.13.

Furthermore, it is easy to check that the space

$$\tilde{H} = \left\{ u = (u_z, u_r, u_\theta) \in H \mid \int_M r u_z dr dz = 0 \right\}$$

is invariant for the operator $L_\lambda + G$ defined by (4.2.53). To see this, since u is z -periodic and $u = 0$ at $r = 1, \eta$, for any $u \in H$,

$$\begin{aligned} \int_0^L \int_\eta^1 r(\tilde{u} \cdot \nabla) u_z dr dz &= \int_0^L \int_\eta^1 r u_r \frac{\partial u_z}{\partial r} dr dz \\ &= - \int_0^L \int_\eta^1 \frac{\partial(r u_r)}{\partial r} u_z dr dz = \int_0^L \int_\eta^1 r \frac{\partial u_z}{\partial z} u_z dr dz = 0. \end{aligned}$$

Also,

$$\int_0^L \int_\eta^1 r \Delta u_z dr dz = 0 \quad \forall u \in \tilde{H}.$$

Thus we see that \tilde{H} is invariant for $L_\lambda + G$.

Therefore, for the vector field (4.2.85), $\int_M r u_z dr dz = 0$. By the connection lemma and the orbit-breaking method in Ma and Wang (2005d), the vector field (4.2.86) is topologically equivalent to its first-order field (4.2.87) for $0 < \lambda - \lambda_0 < \varepsilon$.

STEP 6. PROOF OF ASSERTIONS (5) AND (6). For any initial value $u_0 \in U \setminus (\Gamma \cup \tilde{H})$, we have

$$u_0 = \sum_{k=1}^{\infty} \alpha_k e_k + w_0, \quad (4.2.88)$$

where $w_0 \in \tilde{H}$, and for any $k = 1, 2, \dots$,

$$e_k = (\tilde{e}_k(r), 0, 0), \quad \int_\eta^1 \tilde{e}_k(r) dr \neq 0,$$

and $\tilde{e}_k(r)$ satisfies that

$$D_* D \tilde{e}_k = -\rho_k \tilde{e}_k, \quad \tilde{e}_k|_{r=\eta, 1} = 0, \quad 0 < \rho_1 < \rho_2 < \dots$$

Let

$$\begin{aligned} H_1 &= E \bigoplus \tilde{H}_1, & \tilde{H}_1 &= H_1 \bigcap \tilde{H}, \\ H &= E \bigoplus \tilde{H}, & E &= \text{span}\{e_1, e_2, \dots\}. \end{aligned}$$

Then the equation (4.2.54) can be decomposed into

$$\begin{aligned} \frac{de}{dt} &= L_\lambda e & e \in E, \\ \frac{dw}{dt} &= L_\lambda w + G(w) & w \in \tilde{H}_1, \\ (e(0), w(0)) &= \left(\sum_k \alpha_k e_k, w_0 \right). \end{aligned} \quad (4.2.89)$$

It is obvious that

$$L_\lambda e_k = -\rho_k e_k.$$

Hence for the initial value (4.2.88), the solution $u(t, u_0)$ of (4.2.89) can be expressed as

$$u(t, u_0) = \sum_k \alpha_k e^{-\rho_k t} e_k + w(t, u_0), \quad \int_\eta^1 e_k(r) dr \neq 0. \quad (4.2.90)$$

By (4.2.85) we have

$$\lim_{t \rightarrow \infty} \|w(t, u_0) - v_0\|_{C^r} = 0, \quad v_0 \in \mathcal{A}_\lambda,$$

which implies by Step 5 that $w(t, u_0)$ is topologically equivalent to (4.2.87) for $t > 0$ sufficiently large, i.e., $w(t, u_0)$ has the topological structure as shown in Fig. 4.13.

By the structural stability theorem, Theorem 2.2.9 and Lemmas 2.3.1 and 2.3.3 (connection lemmas) in Ma and Wang (2005d), we infer from (4.2.90) that the vector field in (4.2.90) is topologically equivalent to either the structure as shown in Fig. 4.12a or the structure as shown in Fig. 4.12b, dictated by the sign of α_{k_0} in (4.2.88) with $k_0 = \min\{k | \alpha_k \neq 0\}$. Thus Assertion (5) is proved.

Assertion (6) can be derived by the invariance of H under the operator $L_\lambda + G$ and the structural stability theorem with perturbation in \tilde{H} , in the same fashion as in the proof of Theorem 2.2.9 in Ma and Wang (2005d) by using the Connection Lemma.

The proof of Theorem 4.2.12 is complete. \square

Proof of Theorem 4.2.13. When $R > 0$, by Theorem 2.1.5 we infer from the reduced equation (4.2.83) that the transition of (4.2.50)–(4.2.52) is of Type-II. In the following, we shall use the saddle-node bifurcation theorem, Theorem 2.5.4, to prove this theorem. Let

$$H^* = \{(u_z, u_r, u_\theta) \in H \mid u_z(-z, r) = -u_z(z, r)\}, \quad H_1^* = H_1 \cap H^*.$$

It is easy to see that the space H^* is invariant under the action of the operator $L_\lambda + G$ defined by (4.2.53):

$$L_\lambda + G : H_1^* \rightarrow H^*, \quad (4.2.91)$$

and the first eigenvalue $\beta_1(\lambda)$ of $L_\lambda : H_1^* \rightarrow H^*$ at $\lambda = \lambda_0(T = T_c)$ is simple, with the first eigenvector ψ_1 given by (4.2.62). Hence, the number R in (4.2.84) is valid for the mapping (4.2.90), i.e.,

$$(G(x\psi_1 + \Phi(x)), \psi_1^*) = Rx^3 + o(|x|^3).$$

Thus, it is ready to check that all conditions in Theorem 2.5.4 are fulfilled by the operator (4.2.91). By Step 1 in the proof of Theorem 4.2.12, each singular point of (4.2.91) generates a singularity circle for $L_\lambda + G$ in H . Therefore, Theorem 4.2.13 follows from Theorem 2.5.4.

The proof of Theorem 4.2.13 is complete. \square

Explicit Expression of the Parameter R The parameter R defined by (4.2.71) and (4.2.72) can be explicitly expressed in the following integral formula:

$$\begin{aligned} R = & -\frac{1}{(\psi_1, \psi_1^*)} \left[\frac{\pi}{2} \int_\eta^1 rh\varphi^* \frac{d\phi_0}{dr} dr \right. \\ & + \int_0^L \int_\eta^1 r \left((\tilde{\phi} \cdot \nabla) \psi_z \psi_z^* + (\tilde{\psi} \cdot \nabla) \phi_z \psi_z^* + (\tilde{\phi} \cdot \nabla) \psi_r \psi_r^* \right. \\ & + (\tilde{\psi} \cdot \nabla) \phi_r \psi_r^* + (\tilde{\phi} \cdot \nabla) \psi_\theta \psi_\theta^* + (\tilde{\psi} \cdot \nabla) \phi_\theta \psi_\theta^* \\ & \left. \left. + \frac{\phi_\theta \psi_r \psi_\theta^*}{r} + \frac{\psi_\theta \phi_r \psi_\theta^*}{r} - 2 \frac{\psi_\theta \phi_\theta \psi_r^*}{r} \right) dr dz \right], \end{aligned}$$

where $\tilde{\psi} = (\psi_z, \psi_r)$, $\psi_1 = (\psi_z, \psi_r, \psi_\theta)$, $\psi_1^* = (\psi_z^*, \psi_r^*, \psi_\theta^*)$ are given by (4.2.62) and (4.2.65), and $\phi_0, \phi = (\phi_z, \phi_r, \phi_\theta)$ satisfy

$$\begin{cases} DD_*\phi_0 = a(\varphi D_*h + hD\varphi + \frac{1}{r}\varphi h), \\ \phi_0|_{r=\eta,1} = 0, \\ -\Delta\phi_z + \frac{\partial p}{\partial z} = -1/2 \sin 2azH_1(r), \\ -\left(\Delta - \frac{1}{r^2}\right)\phi_r - \lambda_0\left(\frac{1}{r^2} - \kappa\right)\phi_\theta + \frac{\partial p}{\partial r} = -1/2 \cos 2azH_2(r), \\ -\left(\Delta - \frac{1}{r^2}\right)\phi_\theta - \lambda_0\kappa\phi_r = -1/2 \cos 2azH_3(r), \\ \operatorname{div}(r\tilde{\phi}) = 0, \quad \tilde{\phi} = (\phi_z, \phi_r), \\ \phi|_{r=\eta,1} = 0. \end{cases}$$

Here $H_i(r), i = 1, 2, 3$, are as in (4.2.81).

Parameter R in the Narrow-Gap Case When the conditions (4.2.15) and (4.2.23) hold true, $\mu \rightarrow 1$ and $r_1 \rightarrow \infty$. In this case the equations (4.2.50) are replaced by (4.2.17), and the linearized equations of (4.2.17) reduce to the symmetric linear system (4.2.22). For the approximate problem (4.2.22) with (4.2.51), we use R_0 to denote the number R defined by (4.2.71) and (4.2.72):

$$R_0 = \frac{1}{\|\psi_1\|^2} [(G(\Phi, \psi_1), \psi_1) + G(\psi_1, \Phi), \psi_1].$$

Here ψ_1 is given by (4.2.62) with (h, φ) satisfying

$$\begin{aligned} (D^2 - a^2)^2 h &= \lambda_0 \varphi, \\ (D^2 - a^2)\varphi &= -\lambda_0 h, \\ h = Dh &= 0, \varphi = 0 \quad \text{at } r = 1, \eta, \end{aligned}$$

and Φ is defined by

$$(A - \lambda_0 B_0)\Phi = G(\psi_1, \psi_1), \quad B_0\Phi = P(0, \Phi_\theta, \Phi_r). \quad (4.2.92)$$

We infer from (4.2.92) that

$$R_0 = -\frac{1}{\|\psi_1\|^2} (G(\psi_1, \psi_1), \Phi) = -\frac{1}{\|\psi_1\|^2} ((A - \lambda_0 B_0)\Phi, \Phi).$$

We see that $A - \lambda_0 B_0$ is symmetric and semi-positive definite, and

$$G(\psi_1, \psi_1) \perp \operatorname{Ker}(A - \lambda_0 B_0), \quad \Phi \perp \operatorname{Ker}(A - \lambda_0 B_0).$$

Therefore it follows that

$$R_0 = -\frac{1}{\|\psi_1\|^2}((A - \lambda_0 B_0)^{1/2}\Phi, (A - \lambda_0 B_0)^{1/2}\Phi) < 0.$$

On the other hand, $R(\mu)$ in (4.2.71) is continuous on μ , and $\lim_{\mu \rightarrow 0} R(\mu) = R_0$. Hence we derive the following conclusion.

Theorem 4.2.14 *For the Taylor problems (4.2.50)–(4.2.52) there exist $\mu_0 < 1$ and $0 < \eta_0 < 1$ such that for any $\mu_0 < \mu < 1$ and $\eta_0 < \eta < 1$ with $\mu < \eta^2$, the parameter $R = R(\mu, \eta)$ defined by (4.2.71) is negative, i.e.,*

$$R(\mu, \eta) < 0, \quad \forall \mu_0 < \mu < 1, \quad \eta_0 < \eta < 1.$$

Consequently, the conclusions in Theorem 4.2.12 hold true.

Asymptotic Structure of Solutions Assertions (5) and (6) in Theorem 4.2.12 provide an asymptotic structure of the solutions in the physical space when the gap $r_2 - r_1$ is small, as observed in the experiments. However, for general parameters η and μ we cannot give the precisely theoretic results, and only present some qualitative description. Here we consider two general cases as follows.

CASE: $\mu \geq 0$. Following (Yudovich, 1966; Velte, 1966), for the eigenvector (h, φ) of (4.2.64) the function h can be taken as positive and has a unique maximum point in the interval $(\eta, 1)$. Therefore, for the eigenvectors defined by (4.2.62) and (4.2.63), the vector fields (ψ_z, ψ_r) and $(\tilde{\psi}_z, \tilde{\psi}_r)$ are divergence-free and have the topological structure as shown in Fig. 4.13. Hence to obtain Assertions (5) and (6) in Theorem 4.2.12 for any $\mu \geq 0$ it suffices to prove that

$$h''(r) \neq 0 \quad \text{at} \quad r = r_0, 1, \eta, \quad (4.2.93)$$

where $r_0 \in (\eta, 1)$ is the maximum point of h . We conjecture that the property (4.2.93) for the eigenvector (h, φ) of (4.2.64) is valid for all $0 \leq \mu < 1$ and $0 < \eta < 1$.

CASE: $\mu < 0$. In this case, the situation is different. Numerical results show that the vector field (ψ_z, ψ_r) in (4.2.62) has $k \geq 2$ vortices in the radial direction, called the Taylor vortices, which has the topological structure as shown in Fig. 4.15; see Chandrasekhar (1981). This type of structure is structurally unstable. However, as discussed in Ma and Wang (2000), under a perturbation either in space H or in

$$\tilde{H} = \left\{ (u_z, u_r, u_\theta) \in H \mid \int_0^L \int_\eta^1 r u_z dz dr = 0 \right\},$$

there are only finite types of stable structures. In particular, if the vector field (ψ_z, ψ_r) in (4.2.62) is D -regular, i.e., $h(r)$ satisfies (4.2.93), then there is only one class of stable structures regardless of the orientation. For example, when $u_0 \in H/(\Gamma \cup \tilde{H})$, the asymptotic structure of the solution $u(t, u_0)$ of (4.2.50)–(4.2.52) is as shown in Fig. 4.16, and when $u_0 \in \tilde{H} \setminus \Gamma$, the asymptotic structure of the solution $u(t, u_0)$ is as shown in Fig. 4.17. It is clear that the class of structures illustrated in Fig. 4.16 is different from that illustrated in Fig. 4.17. The first one has a cross in the channel

Fig. 4.15 (ψ_z, ψ_r) has k vortices in r -direction

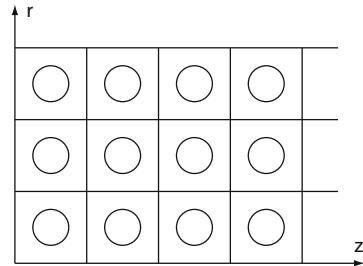


Fig. 4.16 The stable structure with a perturbation in space $H/(\Gamma \cup \bar{H})$

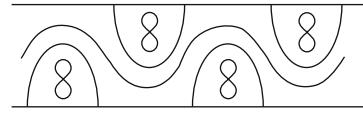
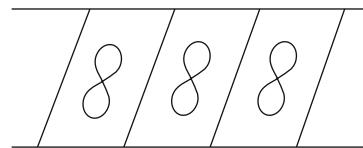


Fig. 4.17 The stable structure with a perturbation in space $\bar{H} \setminus \Gamma$



traveling flow in the z -direction and the second one does not have such a cross the channel flow.

4.2.6 Other Boundary Conditions

Free Boundary Conditions in the z -Direction We consider the Taylor problem (4.2.50) with the boundary condition (4.2.34), and the results obtained here are also valid for the boundary condition (4.2.35). Let

$$\begin{aligned} H_0 &= \{u = (\tilde{u}, u_\theta) \in L^2(M)^3 \mid \operatorname{div}(r\tilde{u}) = 0, u_z|_{z=0,L} = 0, u_r|_{r=r_1,r_2} = 0\}, \\ H_F &= \{u \in H^2(M)^3 \cap H_0 \mid u \text{ satisfies (4.2.34)}\}. \end{aligned}$$

The mappings $L_\lambda + G : H_F \rightarrow H_0$ and $L_\lambda = -A + \lambda B$ are defined as in (4.2.53).

The eigenvalue problem for the free boundary condition is essentially the same as the case for the periodic boundary condition. Their common characteristic is that the eigenvectors can be obtained by the separation of variables.

We consider the eigenvalue problem as follows

$$\begin{aligned}
& -\Delta u_z + \frac{\partial p}{\partial z} = 0, \\
& -\left(\Delta - \frac{1}{r^2}\right) u_r + \frac{\partial p}{\partial r} = \lambda \left(\frac{1}{r^2} - \kappa\right) u_\theta, \\
& -\left(\Delta - \frac{1}{r^2}\right) u_\theta = \lambda \kappa u_r, \\
& \operatorname{div}(r\tilde{u}) = 0,
\end{aligned} \tag{4.2.94}$$

where $\tilde{u} = (u_z, u_r)$. The adjoint equations of (4.2.94) are given by

$$\begin{aligned}
& -\Delta u_z^* + \frac{\partial p^*}{\partial z} = 0, \\
& -\left(\Delta - \frac{1}{r^2}\right) u_r^* + \frac{\partial p^*}{\partial z} = \lambda \kappa u_\theta^*, \\
& -\left(\Delta - \frac{1}{r^2}\right) u_\theta^* = \lambda \left(\frac{1}{r^2} - \kappa\right) u_r^*, \\
& \operatorname{div}(r\tilde{u}^*) = 0.
\end{aligned} \tag{4.2.95}$$

Both (4.2.94) and (4.2.95) are supplemented with the boundary condition (4.2.34).

For the eigenvalue equation (4.2.94) with (4.2.34) we take the separation of variables as follows

$$\psi = (\psi_z, \psi_r, \psi_\theta) = (-\sin a_k z D_* h(r), a_k \cos a_k z h(r), \cos a_k z \varphi(r)) \tag{4.2.96}$$

where $a_k = k\pi/L$, $k \geq 1$, and $(h(r), \varphi(r))$ satisfies (4.2.64) with a_k replacing a . The eigenvectors of (4.2.95) with (4.2.34) are given by

$$\psi^* = (\psi_z^*, \psi_r^*, \psi_\theta^*) = (-\sin a_k z D_* h^*, a_k \cos a_k z h^*, \cos a_k z \varphi^*), \tag{4.2.97}$$

where $(h^*(r), \varphi^*(r))$ satisfies (4.2.67) with a_k replacing a , and ψ^* is the conjugate eigenvector of ψ in (4.2.96).

When $\mu \geq 0$, it is known that there is a sequence of eigenvalues λ_n ($n = 0, 1, \dots$) of (4.2.94) such that λ_0 is real (Temam, 1984), and

$$|\lambda_1| < \dots < |\lambda_n| < |\lambda_{n+1}| < \dots$$

Moreover, for all $L > 0$ except discrete countable values, the first eigenvalue λ_0 is simple, and the first eigenvector ψ_0 is given by (4.2.96) with some wave number $k \geq 1$. We can prove in the same fashion as the proof of Lemma 4.2.11 that for the first eigenvalue λ_0 , the PES is valid for $\mu \geq 0$. Hence, in this section we always consider the case where the first eigenvalue λ_0 is simple, and the PES holds true.

We are now in a position to state and prove the phase transition theorems for the Taylor problem (4.2.50) with (4.2.34).

Let ψ_0 be the first eigenvector of (4.2.94) with (4.2.34), and ψ_0^* be the conjugate eigenvector. We define a parameter R_0 by

$$R_0 = \frac{1}{(\psi_0, \psi_0^*)} (G(\Phi, \psi) + G(\psi, \Phi), \psi_0^*), \quad (4.2.98)$$

where $\Phi \in H_F$ satisfies

$$(A - \lambda_0 B)\Phi = G(\psi_0, \psi_0),$$

and λ_0 is the first eigenvalue of (4.2.94) with (4.2.34).

Theorem 4.2.15 *When the number R_0 defined by (4.2.98) is negative, i.e., $R_0 < 0$, the Taylor problem (4.2.50) with (4.2.34) has a phase transition with Type-I at $\lambda = \lambda_0$. Moreover we have the following assertions:*

- (1) *When $\lambda \leq \lambda_0$, the steady-state $u = 0$ is locally asymptotically stable.*
- (2) *The problem bifurcates from $(u, \lambda) = (0, \lambda_0)$ to an attractor \mathcal{A}_λ consisting of exactly two singular points $\mathcal{A}_\lambda = \{u_1^\lambda, u_2^\lambda\} \subset H_F$ on $\lambda_0 < \lambda$.*
- (3) *There is a neighborhood $U \subset H_0$ of $u = 0$, such that U can be decomposed into two open sets: $\overline{U} = \overline{U}_1^\lambda + \overline{U}_2^\lambda$ satisfying that*

$$U_1^\lambda \cap U_2^\lambda = \emptyset, \quad 0 \in \partial U_1^\lambda \cap \partial U_2^\lambda,$$

and for $i = 1, 2$,

$$u_i^\lambda \in U_i^\lambda, \quad \lim_{t \rightarrow \infty} \|u(t, u_0) - u_i^\lambda\| = 0 \quad \forall u_0 \in U_i^\lambda,$$

where $u(t, u_0)$ is the solution of the problem with $u(0, u_0) = u_0$.

- (4) *If $1 - \mu$ is small, then for any $u_0 \in U_1^\lambda + U_2^\lambda$ there is a time $t_0 \geq 0$ such that $\tilde{u}(t, u_0)$ is topologically equivalent to one of the pattern as shown in Fig. 4.10, where $u(t, u_0) = (\tilde{u}(t, u_0), u_\theta(t, u_0))$ is the solution of the Taylor problem.*

Theorem 4.2.16 *If $R_0 > 0$, then the Taylor problem (4.2.50) with (4.2.34) has a phase transition with Type-II at $\lambda = \lambda_0$, and the following assertions hold true:*

- (1) *The problem has exactly two saddle-node bifurcation points $(u_1^*, \lambda^*), (u_2^*, \lambda^*) \in H_F$ with $0 < \lambda^* < \lambda_0$, and there are at least four branches $\Gamma_1^\lambda, \Gamma_2^\lambda, \tilde{\Gamma}_1^\lambda$, and $\tilde{\Gamma}_2^\lambda$ emanating from (u_1^*, λ^*) and (u_2^*, λ^*) on $\lambda^* < \lambda$; see Fig. 4.18.*
- (2) *The branches Γ_i^λ and $\tilde{\Gamma}_i^\lambda$ satisfy that for $i = 1, 2$,*

$$\lim_{\lambda \rightarrow \lambda_0} \Gamma_i^\lambda = \{0\}, \quad \text{dist}(\tilde{\Gamma}_i^\lambda, 0) = \min_{u \in \tilde{\Gamma}_i^\lambda} \|u\| > 0 \quad \text{at } \lambda = \lambda_0.$$

- (3) *For each $\lambda \in (\lambda^*, \lambda_0)$, the space H_0 can be decomposed into two disjoint open sets $H_0 = \overline{U}_1^\lambda + \overline{U}_2^\lambda$ with $\Gamma_1^\lambda, \Gamma_2^\lambda \subset \partial U_1^\lambda \cap \partial U_2^\lambda$, and the problem has two disjoint attractors \mathcal{A}_1^λ and \mathcal{A}_2^λ , such that*

$$\mathcal{A}_1^\lambda = \{0\} \subset U_1^\lambda, \quad \tilde{\Gamma}_1^\lambda, \tilde{\Gamma}_2^\lambda \subset \mathcal{A}_2^\lambda \subset U_2^\lambda,$$

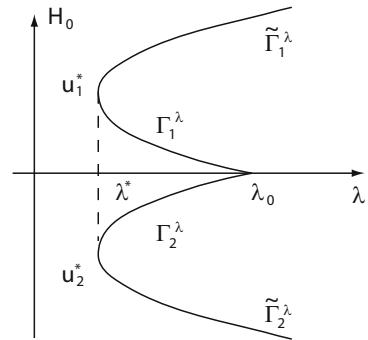
and \mathcal{A}_i^λ attracts U_i^λ ($i = 1, 2$).

(4) When $\lambda_0 \leq \lambda$, the problem has an attractor \mathcal{A}^λ such that $0 \notin \mathcal{A}^\lambda$,

$$\lim_{\lambda \rightarrow \lambda_0^-} \mathcal{A}_2^\lambda = \mathcal{A}^{\lambda_0},$$

and \mathcal{A}^λ attracts $H_0 \setminus \Gamma_\lambda$, where Γ_λ is the stable manifold of $u = 0$ with codimension $m_\lambda \geq 1$.

Fig. 4.18 Saddle node bifurcation and jump transition



Remark 4.2.17 Similar to the case with periodic boundary conditions, if $1 - \mu$ is small, the number given by (4.2.98) is negative: $R_c < 0$, as in Theorem 4.2.14. We expect that Assertion (4) in Theorem 4.2.15 is valid for any $\mu \geq 0$.

Proof of Theorems 4.2.15 and 4.2.16. The proof of Theorem 4.2.15 is similar to that of Theorem 4.2.12, we here omit the details.

Theorem 4.2.16 can be obtained by the same method as used in Theorem 4.2.13 and the fact that the singular points of the Taylor problem (4.2.50) with (4.2.34) appear in pairs. To see this, for any $u \in H_F$,

$$u = \left(\sum_{k=1}^{\infty} \phi_{zk}(r) \sin \frac{k\pi z}{L}, \sum_{k=1}^{\infty} \phi_{rk}(r) \cos \frac{k\pi z}{L}, \sum_{k=1}^{\infty} \phi_{\theta k}(r) \cos \frac{k\pi z}{L} \right),$$

which implies that the problem is z_2 -invariant. Namely, under the transformation: $z \rightarrow z + L$, the equations (4.2.50) and boundary condition (4.2.34) are invariant. Hence the singular points appear in pairs. \square

Rigid and Free-Rigid Boundary Conditions We know that it is much harder to get explicit expression of the eigenvectors for both the rigid and the free-rigid boundary conditions. For simplicity, we only consider the rigid boundary condition

$$u|_{\partial M} = 0, \quad (4.2.99)$$

and the discussion holds true as well for the free-rigid condition.

When $\mu \geq 0$, it is reasonable to conjecture that there is an open and dense set $S \subset (0, +\infty) \times (0, 1)$ such that for any $(L, \eta) \in S$, the first eigenvalue λ_0 of (4.2.94) with (4.2.99) on $M = (0, L) \times (\eta, 1)$ is real and simple, and the PES is valid. Moreover the first eigenvector ψ and its conjugate eigenvector ψ^* satisfy

$$R = (G(\psi, \psi), \psi^*) \neq 0. \quad (4.2.100)$$

Under the assumption above, from Theorems 2.3.2 and 2.5.5 we immediately derive the following theorem.

Theorem 4.2.18 *Let the first eigenvalue λ_0 of (4.2.94) with (4.2.99) be real and simple, and the PES be valid. If (4.2.100) holds true, then the Taylor problem (4.2.50) with (4.2.99) undergo a phase transition with Type-III at $\lambda = \lambda_0$. Moreover, we have the following assertions:*

- (1) *The problem has a saddle-node bifurcation from (u^*, λ^*) with $0 < \lambda^* < \lambda_0$ on $\lambda^* < \lambda$; see Fig. 4.19, and there are two branches Γ_i^λ ($i = 1, 2$) emanating from (u^*, λ^*) satisfying*

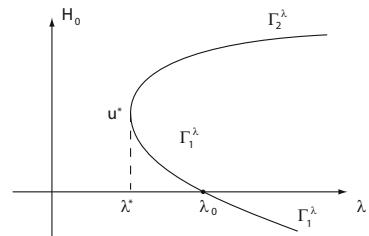
$$\lim_{\lambda \rightarrow \lambda_0} \Gamma_1^\lambda = \{0\} \quad \text{and} \quad 0 \notin \Gamma_2^{\lambda_0}.$$

- (2) *When $\lambda^* < \lambda < \lambda_0$, the space H_0 can be decomposed into two disjoint open sets $H_0 = \overline{U}_1^\lambda + \overline{U}_2^\lambda$ with $\Gamma_1^\lambda \subset \partial U_1^\lambda \cap \partial U_2^\lambda$, and the problem has two disjoint attractors $\mathcal{A}_i^\lambda \subset U_i^\lambda$ ($i = 1, 2$) with $\mathcal{A}_1^\lambda = \{0\}$, $\Gamma_2^\lambda \subset \mathcal{A}_2^\lambda$ such that \mathcal{A}_i^λ attracts U_i^λ .*
- (3) *When $\lambda_0 < \lambda < \lambda_0 + \delta$ for some $\delta > 0$, Γ_1^λ consists of a bifurcated singular point $\Gamma_1^\lambda = \{u^\lambda\}$, and H_0 is decomposed into two disjoint open set $H_0 = \overline{V}_1^\lambda + \overline{V}_2^\lambda$ with $0 \in \partial V_1^\lambda \cap \partial V_2^\lambda$ such that the problem has two disjoint attractors $\Sigma_i^\lambda \subset V_i^\lambda$ ($i = 1, 2$) with $\Sigma_1^\lambda = \{u^\lambda\}$, $\Gamma_2^\lambda \subset \Sigma_2^\lambda$, and Σ_i^λ attracts V_i^λ .*
- (4) *The bifurcated attractor u^λ for $\lambda_0 < \lambda < \lambda_0 + \delta$ can be expressed as*

$$u^\lambda = -R^{-1}\beta_1(\lambda)\psi + o(|\beta_1(\lambda)|),$$

where R is as in (4.2.100).

Fig. 4.19 Type-III transition



4.2.7 Three-Dimensional Perturbation for Narrow-Gap Case

In the previous subsections, we always discuss the phase transitions of the Taylor problems under axisymmetric perturbations. In this subsection we make some remarks on the narrow-gap case with three-dimensional perturbations.

We assume the conditions (4.2.15) and (4.2.23) hold true. When the perturbation is three dimensional and depends on θ , then the equations (4.2.7) are rewritten as

$$\begin{aligned} \frac{\partial u_r}{\partial t} &= \Delta u_r - (u \cdot \nabla) u_r - \frac{\partial p}{\partial r} + A(r) \frac{\partial u_r}{\partial \theta} \\ &\quad + \lambda(1 - (1 - \mu)(r - r_1)) u_\theta, \\ \frac{\partial u_\theta}{\partial t} &= \Delta u_\theta - (u \cdot \nabla) u_\theta - \frac{1}{r} \frac{\partial p}{\partial \theta} + A(r) \frac{\partial u_\theta}{\partial \theta} + \lambda u_r, \\ \frac{\partial u_z}{\partial t} &= \Delta u_z - (u \cdot \nabla) u_z - \frac{\partial p}{\partial z} + A(r) \frac{\partial u_z}{\partial \theta}, \\ \frac{\partial u_r}{\partial r} + \frac{1}{r} \frac{\partial u_\theta}{\partial \theta} + \frac{\partial u_z}{\partial \theta} &= 0, \end{aligned} \quad (4.2.101)$$

where $\lambda = \sqrt{T}$, T is the Taylor number, and

$$\begin{aligned} T &= \frac{4(r_2 - r_1)^4 \Omega_1^2 (\eta^2 - \mu)}{(1 - \eta^2) \nu^2}, & A(r) &= \frac{\sqrt{T}}{2} \left(\frac{\eta^2 - \mu}{1 - \eta^2} - \frac{1 - \mu}{1 - \eta^2} \frac{r_1^2}{r^2} \right), \\ \Delta &= \frac{\partial^2}{\partial r^2} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2}, & u \cdot \nabla &= u_r \frac{\partial}{\partial r} + \frac{u_\theta}{r} \frac{\partial}{\partial \theta} + u_z \frac{\partial}{\partial z}. \end{aligned}$$

The nondimensional domain is $\Omega = (r_1, r_2) \times (0, 2\pi) \times (0, L)$ for $x = (r, \theta, z)$. Equation (4.2.101) is supplemented with the following boundary conditions

$$\begin{aligned} u(r, \theta + 2k\pi, z) &= u(r, \theta, z) & \forall k \in \mathbb{Z}, \\ u &= 0 & \text{at } r = r_1, r_2, \quad \text{or} \quad z = 0, L. \end{aligned} \quad (4.2.102)$$

The initial condition is non-axisymmetric, given by

$$u = u_0(r, \theta, z) \quad \text{at} \quad t = 0. \quad (4.2.103)$$

Consider the eigenvalue problem

$$\begin{aligned} \Delta u_r - \frac{\partial p}{\partial r} + A(r) \frac{\partial u_r}{\partial \theta} + \lambda(1 - (1 - \mu)(r - r_1)) u_\theta &= \beta(\lambda) u_\theta, \\ \Delta u_\theta - \frac{1}{r} \frac{\partial p}{\partial \theta} + A(r) \frac{\partial u_\theta}{\partial \theta} + \lambda u_r &= \beta(\lambda) u_r, \\ \Delta u_z - \frac{\partial p}{\partial z} + A(r) \frac{\partial u_z}{\partial \theta} &= \beta(\lambda) u_z, \\ \frac{\partial u_r}{\partial r} + \frac{1}{r} \frac{\partial u_\theta}{\partial \theta} + \frac{\partial u_z}{\partial z} &= 0, \end{aligned} \quad (4.2.104)$$

with the boundary condition (4.2.102).

Let all eigenvalues (counting the multiplicities) of (4.2.104) be given by

$$\operatorname{Re}\beta_1(\lambda) \geq \operatorname{Re}\beta_2(\lambda) \geq \cdots \geq \operatorname{Re}\beta_k(\lambda) \geq \cdots, \quad \lim_{k \rightarrow \infty} \operatorname{Re}\beta_k(\lambda) = -\infty. \quad (4.2.105)$$

It is clear that if λ^* is an eigenvalue of (4.2.20) with (4.2.102), then there is a real eigenvalue β_k in (4.2.105) satisfying $\beta_k(\lambda^*) = 0$.

Let λ_1 be the first eigenvalue of (4.2.20) with (4.2.102) with multiplicity $m \geq 1$. Then there exists $\beta_{k+1}(\lambda), \dots, \beta_{k+m}(\lambda)$ for some $k \geq 0$, such that $\beta_{k+j}(\lambda)$ ($1 \leq j \leq m$) are real near $\lambda = \lambda_1$, and

$$\beta_{k+j}(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_1, \\ = 0 & \text{if } \lambda = \lambda_1, \\ > 0 & \text{if } \lambda > \lambda_1 \end{cases} \quad \forall 1 \leq j \leq m, \quad (4.2.106)$$

$$\operatorname{Re}\beta_i(\lambda_1) \neq 0 \quad \forall i \neq k + j, \quad 1 \leq j \leq m. \quad (4.2.107)$$

It is clear that $0 > \operatorname{Re}\beta_1(0)$. Hence it follows from (4.2.105) to (4.2.107) that there exists a λ_0 with $0 < \lambda_0 \leq \lambda_1$ such that for some $m_1 \geq 1$

$$\operatorname{Re}\beta_i(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_0, \\ = 0 & \text{if } \lambda = \lambda_0, \\ > 0 & \text{if } \lambda > \lambda_0 \end{cases} \quad \forall 1 \leq i \leq m_1, \quad (4.2.108)$$

$$\operatorname{Re}\beta_j(\lambda_0) < 0 \quad \forall j \geq m_1 + 1. \quad (4.2.109)$$

Thus by the transition theorem, Theorem 2.1.3, we obtain from (4.2.108) and (4.2.109) the following dynamic transition theorem for the Taylor problems (4.2.101)–(4.2.103):

Theorem 4.2.19 *Let λ_1 be the first eigenvalue of (4.2.20) with (4.2.18). Then there exists a number λ_0 with $0 < \lambda_0 \leq \lambda_1$ such that the properties (4.2.108) and (4.2.109) hold true. Therefore the Taylor problems (4.2.101)–(4.2.103) have a phase transition at $\lambda = \lambda_0$. In particular, if $\lambda_0 = \lambda_1$, then the assertions of Theorems 4.2.2 and 4.2.3 are valid for this problem.*

This theorem is a corollary of Theorems 2.1.3 and 2.6.1, and Theorem 4.1 in Ma and Wang (2008d), and we omit details of the proof.

In the three-dimensional perturbation problem, we conjecture that the critical parameter λ_0 satisfying (4.2.108) and (4.2.109) is the first eigenvalue λ_1 of (4.2.20) with (4.2.18), i.e., $\lambda_0 = \lambda_1$.

4.2.8 Physical Remarks

As the Rayleigh–Bénard convection, the Taylor problem has become a paradigm of nonequilibrium phase transitions with its instability, bifurcations, structure formation, and chaotic behaviors in fluid dynamics as well as in physical systems in general.

The classical linear theory explains well the instability at the critical Taylor number, and the known nonlinear theory set by Yudovich (1966); Velte (1966) could only deal with the steady-state bifurcations for the z -periodic and the free boundary conditions.

Here in this section we mainly introduce a dynamical theory for the phase transitions of the Taylor–Couette flow, by using the transition and its classification theorems, the constructive method of center manifold functions, the transition perturbation theory, and the singularity separation results, together with the geometric theory of incompressible flows, established by Ma and Wang (2005d). In the following we shall give some remarks on the dynamical characterization for the Taylor problem with various physical sound boundary conditions.

Rigid and Free-Rigid Boundary Conditions Both the rigid and free-rigid boundary conditions are the most popular conditions used in physics. In this case, Theorem 4.2.18 provided a general description for the phase transitions of the Taylor–Couette flow. Mathematically Theorem 4.2.18 is not a good result because it is difficult to check its conditions. However this theorem is very useful in physics to determine the transition types because the conditions, in fact, are generic and most realizable in experiments.

Based on Theorems 4.2.2, 4.2.3, and 4.2.18, for the boundary conditions we can provide some interesting and useful physical predictions as follows:

- (1) The transition at the critical Taylor number $T_c = \lambda_0^2$ is of Type-III, i.e., the transition is mixed.
- (2) There are two critical Taylor numbers $T_c^* = \lambda^{*2}$ and $T_c = \lambda_0^2$ with $T_c^* < T_c$ at which the phase transitions may occur.
- (3) When $T_c^* < T < T_c$, there is only one type of secondary flows σ_1^T , and the system may have a jump transition between the basic flow σ_0 (the Couette flow) and the secondary flow σ_1^T .
- (4) When $T_c < T < T_c + \varepsilon$ for some $\varepsilon > 0$, there are exactly two types of secondary flows σ_1^T and σ_2^T , and the system may have a continuous transition between the basic flow σ_0 and the secondary flow σ_2^T , or have a jump transition between either σ_0 and σ_1^T or σ_1^T and σ_2^T , depending randomly on the external perturbation.
- (5) In the case where the gap $r_2 - r_1$ is small, $T_c^1 \simeq T_c^2$, and the two types of secondary flows σ_1^T and σ_2^T are controlled by the first eigenvectors, and cannot be distinguished in the experiments.

The z -Periodic and Free Boundary Conditions As mentioned before, the z -periodic boundary condition is an approximate description for the case where the ratio $L/(r_2 - r_1)$ between the height L and the gap $r_2 - r_1$ is sufficiently large, and

the free boundary condition is suitable to the case where the upper and the lower surfaces of boundary are very smooth, or the viscosity of fluid is very low.

By Theorems 4.2.2–4.2.16 we obtain the following dynamical properties:

- (1) The phase transition at the critical Taylor number $T_c = \lambda_0^2$ is of either Type-I or Type-II.
- (2) The two transitions are dictated precisely by the sign of a nondimensional parameter R determined completely by the first eigenvectors and the parameters μ and η .
- (3) When $R < 0$, the transition is continuous, i.e., Type-I, and when $R > 0$, it is jump, i.e., Type-II.
- (4) For the continuous transition case, i.e., $R < 0$, the critical exponent of the phase transition, i.e., the exponent in the expression of bifurcated solutions, is $\beta = 1/2$. Moreover, there is only one critical Taylor number T_c such that the secondary flow $\sigma^T \rightarrow 0$ as $T \rightarrow T_c$.
- (5) When $R > 0$, there are two critical Taylor numbers $T_c = \lambda_0^2$ and $T_c^* = \lambda^{*2}$ with $T_c^* < T_c$, and there is only one type of secondary flow σ^T . When $T_c^* < T < T_c$ the system may have a jump transition between the Couette flow σ_0 and the secondary flow σ^T , and when $T_c \leq T$ the system undergoes a transition from σ_0 to σ^T .
- (6) For the narrow-gap case, the parameter R defined by (4.2.71) is negative: $R < 0$, provided the two coaxial cylinders rotating in the same direction. In fact, even the outer cylinder is static, under the drive of rotation of the inner cylinder, in the narrow gap case, the fluid on $r = r_2$ also has a rotation with an angular velocity $\Omega_2 > 0$, and

$$\mu = \Omega_2/\Omega_1 \rightarrow 1 \quad \text{as} \quad r_1/(r_2 - r_1) \rightarrow \infty,$$

which implies that $R < 0$.

- (7) Theoretic analysis shows that a street of vortices appear in the secondary flow for the narrow-gap case with $\mu \rightarrow 1$. Thus the theoretic results are consistent with the Taylor experiments.

In summary, the dynamical theory established here provides some interesting and useful physical predictions on the types of phase transitions with various physically sound boundary conditions, on the critical exponent of the transition with type-I, on the globally dynamical structure of phase transitions. From both the mathematical and physical points of view, the dynamical theories and methods are valuable to understand the nonequilibrium phase transitions in fluid dynamics.

4.3 Boundary-Layer and Interior Separations in the Taylor–Couette–Poiseuille Flow

Consider a viscous fluid between two coaxial rotating cylinders. As in the Taylor problem, the basic (Couette) flow becomes unstable as soon as the rotation speed of the inner cylinder exceeds a critical value. This instability gives rise to a stationary

axisymmetric counter-rotating vortices that fill the whole annular region. The associated flow is referred to as Taylor–Couette (TC) flow. When a through flow driven by a pressure gradient along the rotating axis is added, the resulting system can exhibit both convective and absolute instabilities. The base flow consists of a superposition of circular Couette flow and annular Poiseuille flow, called Couette–Poiseuille (CP) flow. The axial through-flow suppresses the basic stationary instability, and as the axial pressure gradient increases while the rotation speed of the inner cylinder is held fixed, the first bifurcation gives rise to a traveling train of axisymmetric Taylor vortices, commonly referred to as propagating Taylor vortices (PTV). Henceforth, the term Taylor–Couette–Poiseuille (TCP) flow is used to refer to all hydrodynamic phenomena pertaining to the open system described above; see among others (Raghu and Georgiadis, 2004) and the references therein.

The main objective of this section is to study dynamic transitions of the Couette–Poiseuille flow and formation of the PTV. This study involves two types of transitions: the dynamic transition and the structure transition in the physical space. The analysis is based on the dynamic transition theory presented in this book, and the geometric theory of incompressible flows synthesized in Ma and Wang (2005d).

The results presented here are based on Ma and Wang (2009a). The results show in particular that, contrary to what is commonly believed, the PTV do not appear after the first dynamical bifurcation, and they appear only when the Taylor number is further increased to cross another critical value so that a structural bifurcation occurs. This structural bifurcation corresponds to the boundary-layer and interior separations of the flow structure in the physical spaces.

4.3.1 Model for the Taylor–Couette–Poiseuille Problem

We consider the viscous flow between two rotating coaxial cylinders with an axial pressure gradient. The instability of this flow was first discussed by Goldstein (1937) in 1933, and a further investigation was made by Chandrasekhar (1981).

Let r_1 and r_2 ($r_1 < r_2$) be the radii of two coaxial cylinders, Ω_1 and Ω_2 the angular velocities with which the inner and the outer cylinders rotate respectively. The hydrodynamical equations governing an incompressible viscous fluid between two coaxial cylinders are the Navier–Stokes equations in the cylindrical polar coordinates (r, θ, z) , given by (4.2.5). Let a constant pressure gradient

$$p_0 = \partial p / \partial z$$

be applied in the z -direction, the cylinders rotate, and there exist no radial motions. Under these conditions, (4.2.5) admits a steady-state solution:

$$\mathcal{U}^{CP} = (u_z, u_r, u_\theta) = (W, 0, V), \quad p = p(r, z), \quad (4.3.1)$$

where

$$\frac{1}{\rho} \frac{dp}{dr} = V^2/r, \quad (4.3.2)$$

$$\nu \frac{d}{dr} \left(\frac{d}{dr} + \frac{1}{r} \right) V = 0, \quad (4.3.3)$$

$$\nu \left(\frac{d}{dr} + \frac{1}{r} \right) \frac{d}{dr} W = \frac{p_0}{\rho}, \quad (4.3.4)$$

supplemented with the following boundary conditions

$$(u_r, u_z) = \left(0, -\frac{p_0}{4\rho\nu} W_0 \right) \quad \text{at } r = r_1, r_2, \quad (4.3.5)$$

$$u_\theta = r_i \Omega_i \quad \text{at } r = r_i \text{ (i=1, 2).}$$

Here $W_0 \geq 0$ is a constant.

We derive then from (4.3.1) to (4.3.5) that

$$V(r) = \frac{\Omega_1}{1 - \eta^2} \left(r_1^2 (1 - \mu) \frac{1}{r} - (\eta^2 - \mu)r \right), \quad (4.3.6)$$

$$W(r) = -\frac{p_0}{4\rho\nu} \left(r_1^2 - r^2 + \frac{2r_1 d + d^2}{\ln(1 + d/r_1)} \ln \left(\frac{r}{r_1} \right) + W_0 \right), \quad (4.3.7)$$

$$p(r, z) = p_0 z + \rho \int \frac{1}{r} V^2(r) dr, \quad (4.3.8)$$

where $d (= r_2 - r_1)$ is the gap width, μ and η are as in (4.2.1).

Classically, the flow described by $\mathcal{U}^c = (0, 0, V(r))$ is called the Couette flow, and the flow by $\mathcal{U}^p = (W(r), 0, 0)$ is called the Poiseuille flow. Therefore the flow described by (4.3.1) is a super-position of the Couette flow and the Poiseuille flow, and is usually called the Couette–Poiseuille (CP) flow.

Consider the perturbed state

$$W + u_z, \quad u_r, \quad V + u_\theta, \quad p + p_0 z + \rho \int \frac{1}{r} V^2 dr.$$

Assume that the perturbation is axisymmetric and independent of θ , we derive from (4.2.5) the perturbed equations as follows:

$$\begin{aligned}
\frac{\partial u_z}{\partial t} + (\tilde{u} \cdot \nabla) u_z &= v \Delta u_z - W \frac{\partial u_z}{\partial z} - u_r \frac{dW}{dr} - \frac{1}{\rho} \frac{\partial p}{\partial z}, \\
\frac{\partial u_r}{\partial t} + (\tilde{u} \cdot \nabla) u_r - \frac{u_\theta^2}{r} &= v \left(\Delta u_r - \frac{u_r}{r^2} \right) - \frac{1}{\rho} \frac{\partial p}{\partial r} \\
&\quad - W \frac{\partial u_r}{\partial z} + \frac{2V}{r} r_\theta, \\
\frac{\partial u_\theta}{\partial t} + (\tilde{u} \cdot \nabla) u_\theta + \frac{u_\theta u_r}{r} &= v \left(\Delta u_\theta - \frac{u_\theta}{r^2} \right) \\
&\quad - \left(V' + \frac{1}{r} V \right) u_r - W \frac{\partial u_\theta}{\partial z}, \\
\frac{\partial(ru_z)}{\partial z} + \frac{\partial(ru_r)}{\partial r} &= 0,
\end{aligned} \tag{4.3.9}$$

where $\tilde{u} = (u_z, u_r)$, and

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{\partial^2}{\partial z^2}, \quad (\tilde{u} \cdot \nabla) = u_z \frac{\partial}{\partial z} + u_r \frac{\partial}{\partial r}.$$

The spatial domain for (4.3.9) is $M = (0, L) \times (r_1, r_2)$. The physically sound boundary conditions are as follows:

$$\begin{aligned}
u = (u_z, u_r, u_\theta) &= 0 && \text{at } r = r_1, r_2, \\
u_z = 0, \quad \frac{\partial u_r}{\partial z} &= \frac{\partial u_\theta}{\partial z} = 0 && \text{at } z = 0, L.
\end{aligned} \tag{4.3.10}$$

We consider the narrow-gap case with $\mu \geq 0$, and assume the condition (4.2.15). Then we deduce from (4.3.6) and (4.3.7) that

$$\begin{aligned}
V' + \frac{1}{r} V &= -2\alpha\Omega_1, \\
\frac{2V}{r} &= 2\Omega_1 \left(1 - \frac{1-\mu}{1-\eta^2} \frac{r^2 - r_1^2}{r^2} \right) \simeq 2\Omega_1 \left(1 - \frac{(1-\mu)(r - r_1)}{d} \right), \\
W &\simeq -\frac{p_0}{4\rho v} [(r - r_1)(r_2 - r) + W_0],
\end{aligned}$$

where α is as in (4.2.16).

Replacing u_θ by $\sqrt{\alpha}u_\theta$ in (4.3.9), we obtain the following approximation equations describing the flow between two cylinders with a narrow gap:

$$\begin{aligned}
\frac{\partial u_z}{\partial t} + (\tilde{u} \cdot \nabla) u_z &= v \Delta u_z + \frac{p_0}{4\rho v} (d - 2\bar{r}) u_r \\
&\quad + \frac{p_0}{4\rho v} (W_0 + \bar{r}(d - \bar{r})) \frac{\partial u_z}{\partial z} - \frac{1}{\rho} \frac{\partial p}{\partial z}, \\
\frac{\partial u_r}{\partial t} + (\tilde{u} \cdot \nabla) u_r &= v \Delta u_r + 2\sqrt{\alpha} \Omega_1 \left(1 - \frac{(1-\mu)\bar{r}}{d}\right) u_\theta \\
&\quad + \frac{p_0}{4\rho v} (W_0 + \bar{r}(d - \bar{r})) \frac{\partial u_r}{\partial z} - \frac{1}{\rho} \frac{\partial p}{\partial r}, \\
\frac{\partial u_\theta}{\partial t} + (\tilde{u} \cdot \nabla) u_\theta &= v \Delta u_\theta + 2\sqrt{\alpha} \Omega_1 u_r + \frac{p_0}{4\rho v} (W_0 + \bar{r}(d - \bar{r})) \frac{\partial u_\theta}{\partial z}, \\
\operatorname{div} \tilde{u} &= 0,
\end{aligned} \tag{4.3.11}$$

where $\bar{r} = r - r_1$, $\Delta = \partial^2/\partial z^2 + \partial^2/\partial r^2$.

We need to consider the nondimensionalized form of (4.3.11). To this end, let

$$\begin{aligned}
(x, t) &= (x'd, t'd^2/v) & (x = (z, r, r\theta)), \\
(u, p, W_0) &= (u'v/d, p'\rho v^2/d^2, W'_0 d^2) & (u = (u_z, u_r, u_\theta)).
\end{aligned}$$

Omitting the primes, the equations can be rewritten as

$$\begin{aligned}
\frac{\partial u_z}{\partial t} &= \Delta u_z + \gamma(1 - 2\bar{r}) u_r \\
&\quad + \gamma(W_0 + \bar{r}(1 - \bar{r})) \frac{\partial u_z}{\partial z} - \frac{\partial p}{\partial z} - (\tilde{u} \cdot \nabla) u_z, \\
\frac{\partial u_r}{\partial t} &= \Delta u_r + \lambda(1 - (1 - \mu)\bar{r}) u_\theta \\
&\quad - \gamma(W_0 + \bar{r}(1 - \bar{r})) \frac{\partial u_r}{\partial z} - \frac{\partial p}{\partial r} - (\tilde{u} \cdot \nabla) u_r, \\
\frac{\partial u_\theta}{\partial t} &= \Delta u_\theta + \lambda u_r + \gamma(W_0 + \bar{r}(1 - \bar{r})) \frac{\partial u_\theta}{\partial z} - (\tilde{u} \cdot \nabla) u_\theta, \\
\operatorname{div} \tilde{u} &= 0,
\end{aligned} \tag{4.3.12}$$

where $\lambda = \sqrt{T}$. Here the Taylor number T and the nondimensional parameter γ are given by

$$T = \frac{4\alpha\Omega_1^2 d^4}{v^2}, \quad \gamma = \frac{p_0 d^3}{4\rho v^2}. \tag{4.3.13}$$

We remark here that the nondimensional parameter γ , proportional to the Reynolds number Re squared times the small gap d , is a small parameter.

When the gap $d = r_2 - r_1$ is small in comparison with the mean radius, we need not distinguish the two equations (4.3.9) and (4.3.12) to discuss their dynamic bifurcation. Therefore, we always consider the problem (4.3.12) instead of (4.3.9). Equation (4.3.12) is supplemented with the following initial condition

$$u = \varphi \quad \text{at} \quad t = 0. \quad (4.3.14)$$

4.3.2 Phase Transition of the TCP Problem

Eigenvalue and Eigenvectors of the Linearized Equations The linearized equations of (4.3.12) with (4.3.10) read

$$\begin{aligned} -\Delta u_z - \gamma(1 - 2\bar{r})u_r + \gamma(W_0 + \bar{r}(1 - \bar{r}))\frac{\partial u_z}{\partial z} + \frac{\partial p}{\partial z} &= 0, \\ -\Delta u_r + \lambda(1 - \mu)\bar{r}u_\theta + \gamma(W_0 + \bar{r}(1 - \bar{r}))\frac{\partial u_r}{\partial z} + \frac{\partial p}{\partial r} &= \lambda u_\theta, \\ -\Delta u_\theta - \gamma(W_0 + \bar{r}(1 - \bar{r}))\frac{\partial u_\theta}{\partial z} &= \lambda u_r, \\ \frac{\partial u_z}{\partial z} + \frac{\partial u_r}{\partial r} &= 0, \end{aligned} \quad (4.3.15)$$

with the boundary conditions

$$\begin{aligned} u &= 0 && \text{at } r = r_1, r_1 + 1, \\ u_z &= 0, \quad \frac{\partial u_r}{\partial z} = \frac{\partial u_\theta}{\partial z} = 0 && \text{at } z = 0, L, \end{aligned} \quad (4.3.16)$$

where $\lambda = \sqrt{T}$, T is the Taylor number, T and γ are given by (4.3.13), and μ is defined by (4.2.1). We denote

$$\varepsilon = (\gamma, 1 - \mu) \in (0, \infty) \times (0, \infty).$$

Let $\lambda_0^\varepsilon > 0$ be the first eigenvalue of (4.3.15) and (4.3.16). The critical Taylor number is given by

$$T_c = (\lambda_0^\varepsilon)^2. \quad (4.3.17)$$

When $\varepsilon \rightarrow 0$, (4.3.15) become symmetric as follows::

$$\begin{aligned} -\Delta u_z + \frac{\partial p}{\partial z} &= 0, \\ -\Delta u_r + \frac{\partial p}{\partial r} &= \lambda u_\theta, \\ -\Delta u_\theta &= \lambda u_r, \\ \frac{\partial u_z}{\partial z} + \frac{\partial u_r}{\partial r} &= 0. \end{aligned} \quad (4.3.18)$$

The eigenvalue problem (4.3.18) with (4.3.16) has the first eigenvectors:

$$\psi = \left(\frac{1}{a} \sin az R'(r), -\cos az R(r), -\frac{1}{a^2 \lambda_0} \cos az \left(\frac{d^2}{dz^2} - a^2 \right)^2 R(r), \right), \quad (4.3.19)$$

where $a = k\pi/L > 0$ is the wave length, for some integer $k \geq 1$ with k depending on L . When L is large, k is taken to be the minimizer of

$$\min_k \left| \frac{k\pi}{L} - 3.117 \right|.$$

The number λ_0 is the first eigenvalue of (4.3.18) with (4.3.16), and when $a \approx 3.117$, $\lambda_0^2 \approx 1700$.

The function $R(r)$ in (4.3.19) satisfies

$$\begin{aligned} \left(\frac{d^2}{dz^2} - a^2 \right)^3 R &= -a^2 \lambda_0^2 R, \\ R = 0, \quad R' = 0, \quad \left(\frac{d^2}{dz^2} - a^2 \right)^2 R &= 0 \text{ at } r = r_1, r_1 + 1. \end{aligned}$$

Its solution can be expressed as (see Chapter II–15 of Chandrasekhar (1981)):

$$R(r) = \cos \alpha_0 x - \beta_1 \cosh \alpha_1 x + \cos \alpha_2 x + \beta_2 \sinh \alpha_1 x \sin \alpha_2 x, \quad (4.3.20)$$

where $x = \bar{r} - \frac{1}{2}$, $\bar{r} = r - r_1$,

$$\begin{aligned} \beta_1 &= 0.06151664, \quad \beta_2 = 0.10388700, \\ \alpha_0 &= 3.973639, \quad \alpha_1 = 5.195214, \quad \alpha_2 = 2.126096. \end{aligned} \quad (4.3.21)$$

In addition, the first eigenvalue λ_0 of (4.3.18) with (4.3.16) is simple except for the following values of L :

$$L_k = \frac{k\pi}{b}, \quad k = 1, 2, \dots, \quad b \approx 4.2. \quad (4.3.22)$$

Phase Transition Theorems We are now in a position to state and prove the phase transition theorems for the Taylor–Couette–Poiseuille problem in the narrow-gap case. For this purpose we set the function spaces as

$$\begin{aligned} H &= \{u = (\tilde{u}, u_\theta) \in L^2(M)^3 \mid \operatorname{div} \tilde{u} = 0, \tilde{u} \cdot n|_{\partial M} = 0\}, \\ H_1 &= \{u \in H^2(M, \mathbb{R}^3) \cap H \mid u \text{ satisfies (4.3.16)}\}. \end{aligned} \quad (4.3.23)$$

Theorem 4.3.1 *Assume the conditions (4.2.15) and (4.2.23). Then there is a $\delta > 0$ such that for any $|\gamma| < \delta$ the problem (4.3.12) with (4.3.10) has a transition at $(u, T) = (0, T_c)$, where $T = \lambda^2$ and T_c is the critical Taylor number given by (4.3.17). Moreover we have*

- (1) The problem undergoes a transition to an attractor \mathcal{A}_T on $T > T_c$ with $0 \notin \mathcal{A}_T$, such that the dimension of \mathcal{A}_T is either zero or one, and \mathcal{A}_T attracts $H \setminus \Gamma$, where Γ is the stable manifold of $u = 0$ with codimension m ($1 \leq m \leq 2$) in H .
- (2) Any $u_T \in \mathcal{A}_T$ can be expressed as

$$u_T = v_T + w_T^\varepsilon, \quad \lim_{\varepsilon \rightarrow 0, T \rightarrow T_0} \|w_T^\varepsilon\|/\|v_T\| = 0,$$

where $T_0 = \lambda_0^2$, λ_0 and v_T are the first eigenvalue and eigenvector of (4.3.18) with (4.3.16).

Theorem 4.3.2 Under the hypotheses of Theorem 4.3.1, there are $\tilde{L}_K \simeq L_k$, where L_k are given by (4.3.22) such that when $L \neq \tilde{L}_k$ the following assertions hold true:

- (1) The bifurcated attractor \mathcal{A}_T in Theorem 4.3.1 consists of exactly two steady-state solutions, $\mathcal{A}_T = \{u_1^T, u_2^T\}$, of (4.3.12) with (4.3.10), such that for $i = 1, 2$, u_i^T can be expressed as

$$\begin{aligned} u_1^T &= \alpha_1(T, \varepsilon)\psi + v_1(T, \varepsilon), \\ u_2^T &= -\alpha_2(T, \varepsilon)\psi + v_2(T, \varepsilon), \\ v_i(\lambda, \varepsilon) &= o(|\alpha_i(\lambda, \varepsilon)|). \end{aligned} \quad (4.3.24)$$

Here ψ is the first eigenvector given by (4.3.19), and

$$\begin{aligned} \alpha_{1,2} &= \frac{\sqrt{b^2(\varepsilon) + 4\beta_1(\lambda, \varepsilon)} \mp |b(\varepsilon)|}{2c} + o(|b|, |\beta_1|), \\ \beta_1(\lambda, \varepsilon) &= \sigma(\varepsilon)(\lambda - \lambda_0^\varepsilon) + o(|\lambda - \lambda_0^\varepsilon|), \\ \sigma(0) &> 0, \quad \lambda = T^2, \\ b(\varepsilon) &= - \int_M (\tilde{\psi}^\varepsilon \cdot \nabla) \psi^\varepsilon \cdot \psi^{\varepsilon*} dx, \end{aligned} \quad (4.3.25)$$

where $\psi^\varepsilon = (\tilde{\psi}^\varepsilon, \psi_\theta^\varepsilon)$, with $\tilde{\psi}^\varepsilon = (\psi_z^\varepsilon, \psi_r^\varepsilon)$, is the eigenvector of (4.3.15) and (4.3.16) corresponding to λ_0^ε , $\psi^{\varepsilon*}$ is the dual eigenvector, and $c > 0$ a constant.

- (2) The space H can be decomposed into two disjoint open sets $H = \overline{U}_1^T + \overline{U}_2^T$ with $0 \in \Gamma = \partial U_1^T \cap \partial U_2^T$, such that for $i = 1, 2$, $u_i^T \in U_i^T$, and

$$\lim_{t \rightarrow \infty} \|u(t, u_0) - u_i^T\| = 0 \quad \text{for } u_0 \in U_i^T,$$

where $u(t, u_0)$ is the solution of (4.3.12) with (4.3.10) and (4.3.14).

Remark 4.3.3 It is known that (see Kato 1995)

$$\psi^\varepsilon \rightarrow \psi, \quad \psi^{\varepsilon*} \rightarrow \psi \quad \text{as } \varepsilon \rightarrow 0,$$

where ψ is given by (4.3.19). Consequently the number $b(\varepsilon)$ in (4.3.25) satisfies

$$b(\varepsilon) \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0.$$

Remark 4.3.4 The transition theorems (Theorems 4.3.1 and 4.3.2) provide three dynamical ingredients different from any classical bifurcation theories as follows:

- (1) The basin of attraction of the bifurcated attractor \mathcal{A}_T . As seen in Assertion (1) of Theorem 4.3.1, \mathcal{A}_T attracts $H \setminus \Gamma$, i.e., the stable domain of the TCP flow for the external perturbation is almost whole phase space.
- (2) The explicit expressions (4.3.24) and (4.3.25) of the bifurcated solutions. These expressions are crucial to determine, in the next two subsections, the asymptotic structure in the physical space of solutions of the problem, leading to the justification of boundary-layer and interior separation for the TCP flow.
- (3) Global transition structure. We can see that the explicit forms (4.3.24) and (4.3.25) provide a precise graph of the saddle-node bifurcation. In fact, if we take $\Gamma_i^\lambda = \{u_i^T\}$ with $\lambda = \sqrt{T}$, where $u_i^T (i = 1, 2)$ are as in (4.3.24) and (4.3.25), then $\Gamma_i^\lambda (i = 1, 2)$ depict in the $\lambda - H$ space a graph as shown in Fig. 4.19. Here λ_0^ε represent λ_0 , and $\lambda^* (\lambda^* < \lambda_0^\varepsilon)$ satisfies that

$$b^2(\varepsilon) + 4\beta_1(\lambda^*, \varepsilon) = 0, \quad \lambda^* = \sqrt{T^*}. \quad (4.3.26)$$

Proof of Theorems 4.3.1 and 4.3.2. We shall apply Theorems 2.6.1 and 2.6.6 to prove these two theorems. Let H and H_1 be defined by (4.3.23). We define the mappings $L_\lambda = A + \lambda B : H_1 \rightarrow H$, $S_\lambda^\varepsilon : H_1 \rightarrow H$, and $G : H_1 \rightarrow H$ by

$$\begin{aligned} Au &= \{P\Delta\tilde{u}, \Delta u_\theta\}, \quad Bu = \{P(0, u_\theta), u_r\}, \\ S_\lambda^\varepsilon u &= \left\{ P \left(\gamma(1 - 2\bar{r})u_r + \gamma(W_0 + \bar{r}(1 - \bar{r})) \frac{\partial u_z}{\partial z}, \lambda(1 - \mu)\bar{r}u_\theta \right. \right. \\ &\quad \left. \left. + \gamma(W_0 + \bar{r}(1 - \bar{r}(1 - \bar{r})) \frac{\partial u_r}{\partial z} \right), \gamma(W_0 + \bar{r}(1 - \bar{r})) \frac{\partial u_\theta}{\partial z} \right\}, \\ G(u) &= \{-P(\tilde{u} \cdot \nabla)\tilde{u}, -(\tilde{u} \cdot \nabla)u_\theta\}. \end{aligned}$$

where $u = (\tilde{u}, u_\theta) \in H_1$ with $\tilde{u} = (u_z, u_r)$, and the operator $P : L^2(M)^3 \rightarrow H$ is the Leray projection. Thus the equations (4.3.12) with (4.3.10) can be written in the abstract form

$$\frac{du}{dt} = L_\lambda u + S_\lambda^\varepsilon u + G(u).$$

It is well known that the operators L_λ , S_λ^ε , and G satisfy the conditions (2.1.2), (2.1.3), and (2.6.2) in Theorems 2.6.1 and 2.6.6. It is clear that L_λ is symmetric, and the condition (2.6.18) in Theorem 2.6.6 holds true.

It is clear that the eigenvalues $\beta_k(\lambda) (k = 1, 2, \dots)$ of the equation:

$$L_\lambda u = \beta(\lambda)u$$

satisfy the conditions (2.2.11) and (2.2.12) in Theorems 2.6.1 and 2.6.6. In particular, by Theorem 3.16 in Ma and Wang (2005b), we see that $u = 0$ is a globally asymptotically stable steady-state solution of the following equation

$$\frac{du}{dt} = L_\lambda u + G(u).$$

Thus Theorem 4.3.1 follows from Theorem 2.6.1, and Theorem 4.3.2 follows from Theorem 2.6.6. We note that the expression (4.3.25) is derived from (2.6.16) in the proof of Theorems 2.6.3 and 2.6.6.

The proofs of Theorems 4.3.1 and 4.3.2 are complete. \square

4.3.3 Boundary-Layer Separation from the Couette–Poiseuille Flow

In this and the next subsections, we always assume that the conditions in Theorem 4.3.2 hold true. The main objective of these two subsections is to study the asymptotic structure and its transition in the physical space in the CP flow and the TCP flow regimes.

We shall see that the results obtained here provide a rigorous characterization on how, when and where the Taylor vortices are generated. In particular, contrary to what is commonly believed, we show that the propagating Taylor vortices (PTV) do not appear after the first phase transition, and they appear only after further increase the Taylor number so that the boundary-layer or interior separation occurs.

We shall prove that the type of separations (boundary-layer or interior) is dictated by the (vertical) through flow driven by a pressure gradient along the rotating axis. Hence we consider two cases. The first is the case where the through flow vanishes at the cylinder walls, i.e., $W_0 = 0$ in (4.3.5). This leads to boundary-layer separation and will be studied in this subsection.

The second is the case where $W_0 \neq 0$, leading to interior separations. This case will be addressed in the next subsection.

We now consider the case where $W_0 = 0$, and the main results are the following two theorems. The first gives a precise characterization on how, when and where the propagating Taylor vortices (PTV) are generated for the bifurcated solutions, and the second theorem described time evolution to the PTV for the solutions of the equations.

Theorem 4.3.5 *Assume that the conditions in Theorem 4.3.2 hold true, and assume that the constant velocity in boundary condition (4.3.5) is zero, i.e., $W_0 = 0$. Then there is a $\gamma_0 > 0$ such that for any $0 < \gamma < \gamma_0$ where γ is defined by (4.3.13), the following assertions hold true for the two (bifurcated) steady-state solutions $v_i^T = \mathcal{U}^{CP} + u_i^T$ ($i = 1, 2$) of (4.2.5) with (4.3.5), where \mathcal{U}^{CP} and u_i^T are given by (4.3.1) and (4.3.24):*

1. *For v_1^T , there is a number $T_1 > T_c$ (T_c as in (4.3.17)) such that*
 - a. *for $T_c < T < T_1$, the vector field $\tilde{v}_1^T = (W + u_{1z}^T, u_{1r}^T)$ is topologically equivalent to the vertical shear flow $(W, 0)$ as shown in Fig. 4.20, and*

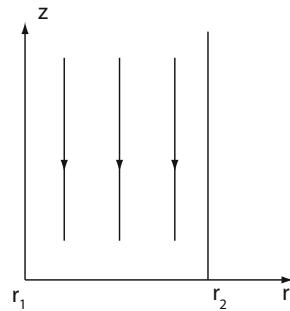
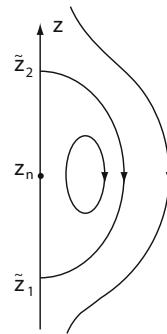
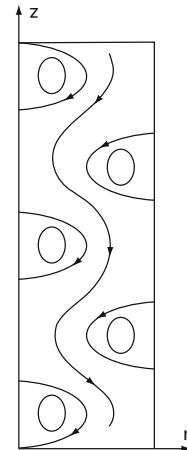
- b. for $T_1 < T$, there is a unique center (vortex) of \tilde{v}_1^T separated from each point $(z_n, r_1) \in \partial M$, where $z_n \simeq (4n + 1)\pi/2a$ (resp. $(\bar{z}_n, r_2) \in \partial M$, $\bar{z}_n = (4n + 3)\pi/2a$) for $n = 0, 1, \dots, k_0$, as shown in Fig. 4.21, and $a = k\pi/L$ as in (4.3.19).
2. For \tilde{v}_2^T , there is a number $b_0 > 0$ such that only one of the following two assertions holds true:
- If $|b(\varepsilon)| < b_0$ ($b(\varepsilon)$ as in (4.3.25)), there is a $T_c > T_c$ such that
 - for $T_c < T < T_2$, the vector field $\tilde{v}_2^T = (W + u_{2z}^T, u_{2r}^T)$ has the topological structure as shown in Fig. 4.20, and
 - for $T_2 < T$, there is a unique center of \tilde{v}_2^T separated from each point $(\bar{z}_n, r_1) \in \partial M$ (resp. $(z_n, r_2) \in \partial M$) for $n = 0, 1, \dots, k_0$, as shown in Fig. 4.21.
 - If $|b(\varepsilon)| > b_0$, then for $T > T_c$, the vector field \tilde{v}_2^T is topologically equivalent to the structure as shown in Fig. 4.22.
3. There exists a $T_3 > T_1$ ($T_1 \geq T_2$) such that for $T_i < T < T_3$, the vector field \tilde{v}_i^T ($i = 1, 2$) is topologically equivalent to the structure as shown in Fig. 4.22, and T_3 is independent of γ .

From Theorems 4.3.2 and 4.3.5, together with the structural stability theorem, Theorem 2.1.2 in Ma and Wang (2005d), we immediately derive the following theorem which links the dynamics with the structure in the physical space.

Theorem 4.3.6 Assume that the conditions in Theorem 4.3.2 hold true, $W_0 = 0$ in (4.3.5), and $0 < \gamma < \gamma_0$. Then there are \tilde{T}_j ($1 \leq j \leq 3$) with $\tilde{T}_1 = T_1$, $T^* \leq \tilde{T}_2 \leq T_2$, $\tilde{T}_3 = T_3$, where T_j ($1 \leq j \leq 3$) are as in Theorem 4.3.5 and T^* satisfies (4.3.26), such that for each $T < \tilde{T}_3$ the space H can be decomposed into two disjoint open sets $H = \overline{U}_1^T + \overline{U}_2^T$ with $0 \in \partial U_1^T \cap \partial U_2^T$, and the following assertions hold true:

- (1) For $u_0 \in U_i^T$ there is a time $t_0 \geq 0$ such that if $T < \tilde{T}_i$ ($i = 1, 2$) for the solution $u(t, u_0) = (\tilde{u}, u_0)$ of (4.2.5) with (4.3.5) and (4.2.104), the vector field \tilde{u} is topologically equivalent to the structure as shown in Fig. 4.20, and if $\tilde{T}_i < T < \tilde{T}_3$, \tilde{u} is topologically equivalent to the structure as shown in Fig. 4.22 for all $t > t_0$.
- (2) In particular, for the initial value $u_0 \in U_i^T$ near the Couette–Poiseuille flow (4.3.1), if $T > \tilde{T}_i$ there is a time $t_0 > 0$ such that the solution $u(t, u_0)$ of (4.2.5) with (4.3.5) and (4.3.14) has a boundary-layer separation at $t = t_0$, i.e., the vector field \tilde{u} of $u(t, u_0)$ is topologically equivalent to the structure as shown in Fig. 4.20 for $t < t_0$, and \tilde{u} is topologically equivalent to the structure near boundary ∂M as shown in Fig. 4.21 for $t_0 < t$.

Remark 4.3.7 The critical value \tilde{T}_i ($i = 1, 2$) in Theorem 4.3.6 which determines the boundary-layer separation for the Taylor–Couette–Poiseuille problem depends continuously on the nondimensional parameter γ given by (4.3.13). The expansion is as follows

Fig. 4.20 Vertical shear flow**Fig. 4.21** Boundary separation and reattachment**Fig. 4.22** Propagating Taylor Vortices after boundary-layer separations

$$\tilde{T}_i = \alpha_i \gamma + o(\gamma), \quad \alpha_i > 0 \text{ a constant, } i = 1, 2.$$

In addition, the condition $0 < \gamma < \gamma_0$ in the above theorems is the consequence of the narrow-gap assumption (i.e., $d = r_2 - r_1$ is small).

Remark 4.3.8 The height L has an integer number times the length of π/a , i.e., $L = K \cdot \frac{\pi}{a}$ for some integer K . Hence the number k_0 of vortices separated from $r = r_1$ and r_2 in Theorems 4.3.5 and 4.3.6 satisfies

$$k_0 = \begin{cases} \left[\frac{K+1}{2} \right] & \text{for } r = r_1 \text{ (or } r = r_2\text{),} \\ \left[\frac{K}{2} \right] & \text{for } r = r_2 \text{ (or } r = r_1\text{).} \end{cases}$$

where $[\alpha]$ is the integer part of α .

Proof of Theorem 4.3.5. We proceed in a few steps as follows.

STEP 1. By Theorem 4.3.2, as the Taylor number $T = \lambda^2$ satisfies $T_c < T$ the equations (4.2.5) with (4.3.5) generate from the basic flow (4.3.1) two steady-state solutions

$$v_i^T = (W(r) + u_{iz}^T, u_{ir}^T, V(r) + u_{i\theta}^T), \quad i = 1, 2,$$

which are globally asymptotically stable under the axisymmetric perturbation.

As seen in Remark 4.3.3, the number $b(\varepsilon)$ in (4.3.25) satisfies

$$\lim_{\varepsilon \rightarrow 0} b(\varepsilon) = 0 \quad (\varepsilon = (1 - \mu, r)).$$

Hence, to prove this theorem, it suffices to consider only the following two vector fields:

$$\tilde{V}_1^T = (W(r), 0) + \frac{\sqrt{b^2 + 4\alpha(\lambda - \lambda_0^\varepsilon)} - b}{2c} \tilde{u}_0, \quad (4.3.27)$$

$$\tilde{V}_2^T = (W(r), 0) + \frac{\sqrt{b^2 + 4\alpha(\lambda - \lambda_0^\varepsilon)} + b}{2c} \tilde{u}_0, \quad (4.3.28)$$

where $\lambda = \sqrt{T}$, $b = |b(\varepsilon)|$, $\alpha = \alpha(\varepsilon) > 0$, and $c > 0$ a constant. In addition,

$$\begin{aligned} W(r) &= -\gamma \bar{r}(1 - \bar{r}) & (\bar{r} = r - r_1), \\ \tilde{u}_0 &= \left(\frac{1}{a} \sin azR'(r), -\cos azR(r) \right), \end{aligned}$$

and $R(r)$ is given by (4.3.20).

STEP 2. CLAIM: \tilde{V}_1^T HAS A BOUNDARY-LAYER SEPARATION ON $T_1 < T$ FOR SOME $T_1 > T_c = (\lambda_0^\varepsilon)^2$. We shall apply the structural bifurcation theorems, Theorems 5.3.6 and 5.3.7 in Ma and Wang (2005d), to prove this claim. To this end, let

$$\Lambda = \left(\sqrt{b^2 + 4\alpha(\lambda - \lambda_0^\varepsilon)} - b \right) / 2ac,$$

which is an increasing function of T with $\Lambda = 0$ at $T_c = (\lambda_0^\varepsilon)^2$. Thus, the vector field (4.3.27) can be rewritten as

$$\tilde{V}_1^T = (\Lambda \sin azR'(r) - \gamma \bar{r}(1 - \bar{r}), -a\Lambda \cos azR(r)). \quad (4.3.29)$$

We discuss the structural bifurcation of (4.3.29) only on the portion $r = r_1$ of ∂M , and the bifurcation on the portion $r = r_2$ can be proved in the same fashion.

Since $R(r_1) = R'(r_1) = 0$, the zero points of $\partial \tilde{V}_1^T / \partial n = \partial \tilde{V}_1^T / \partial r$ on $r = r_1$ are given by the solutions of the equation

$$\Lambda \sin azR''(r_1) - \gamma = 0. \quad (4.3.30)$$

Let $\Lambda_0 = \gamma/R''(r_1)$. Then we infer from (4.3.30) that for $r = r_1$,

$$\frac{\partial \tilde{V}_1^T}{\partial n} \begin{cases} \neq 0 & \text{if } \Lambda < \Lambda_0 \\ = 0 & \text{if } \Lambda = \Lambda_0 \text{ and } z_n = \frac{(4n+1)\pi}{2a}. \end{cases} \quad (4.3.31)$$

Therefore \tilde{V}_1^T has the Taylor expansion at $\Lambda = \Lambda_0$ as follows:

$$\begin{aligned} \tilde{V}_1^T &= V^0 + (\Lambda - \Lambda_0)V^1, \\ V^0 &= (V_z^0, V_r^0) = (\Lambda_0 \sin azR' - \gamma \bar{r}(1 - \bar{r}), -a\Lambda \cos azR) \\ V^1 &= (V_z^1, V_r^1) = (\sin azR', -a \cos azR). \end{aligned} \quad (4.3.32)$$

We derive from (4.3.20) that

$$\begin{aligned} R''(r) &= -a_0^2 \cos \alpha_0 x + [\beta_1(\alpha_2^2 - \alpha_1^2) + 2\beta_2\alpha_1\alpha_2] \cosh(\alpha_1 x) \cos \alpha_2 x \\ &\quad + [\beta_2(\alpha_1^2 - \alpha_2^2) + 2\beta_1\alpha_1\alpha_2] \sinh(\alpha_1 x) \sin \alpha_2 x, \end{aligned} \quad (4.3.33)$$

where $x = \bar{r} - \frac{1}{2}$. Inserting (4.3.21) into (4.3.33) we find

$$R''(r_1) \simeq 8 + 1.7e^{5/2} \simeq 28. \quad (4.3.34)$$

Thus we derive from (4.3.31), (4.3.32), and (4.3.34) that at $(z_n, r_1) = (4n+1/(2a)\pi, r_1)$,

$$\begin{aligned} \frac{\partial V^0}{\partial n} &= \frac{\partial V^0}{\partial r} = 0, \\ \frac{\partial V^1}{\partial n} &= \frac{\partial V^1}{\partial r} \simeq (28, 0), \\ \frac{\partial^3 V_z^0}{\partial \tau^2 \partial n} &= \frac{\partial^3 V_z^0}{\partial z^2 \partial r} \simeq -28a^2\Lambda_0. \end{aligned}$$

Namely, the conditions (5.3.4), (5.3.6), and (5.3.7) in Theorems 5.3.6 and 5.3.7 in Ma and Wang (2005d) are valid.

Now we verify the condition (5.3.5) in Ma and Wang (2005d), i.e., we need to prove that the index of the vector field $\partial \tilde{V}_1^T / \partial r$ vanishes at $\Lambda = \Lambda_0$ and (z_n, r_1) :

$$\text{ind} \left(\frac{\partial V^{\Lambda_0}}{\partial r}, (z_n, r_1) \right) = 0. \quad (4.3.35)$$

To this end, we consider the equation

$$\frac{\partial V_z^T}{\partial r} = \Lambda \sin az R''(r) - \gamma(1 - 2\bar{r}) = 0. \quad (4.3.36)$$

The function $R''(r)$ has the Taylor expansion at $r = r_1$ as follows:

$$R''(r) = R''(r_1) + R'''(r_1)\bar{r} + o(|\bar{r}|). \quad (4.3.37)$$

We deduce from (4.3.20) and (4.3.21) that

$$R'''(r_1) \simeq -64 \times \frac{1.7}{2} - 10 \times e^{5/2} \simeq -175. \quad (4.3.38)$$

On the other hand, we note that

$$\Lambda = \frac{\gamma\Lambda}{R''(r_1)\Lambda_0} = \frac{\gamma\Lambda}{28\Lambda_0}.$$

Thus, by (4.3.34), (4.3.37), and (4.3.38), the equation (4.3.36) near $(z, r) = (z_n, r_1)$ can be rewritten in the following form

$$\left(1 - \frac{\Lambda}{\Lambda_0} \sin az\right) + \left(\frac{175\Lambda \sin az}{28\Lambda_0} - 2\right)\bar{r} + o(\bar{r}) = 0. \quad (4.3.39)$$

Obviously, there is a $\delta > 0$ such that the equation (4.3.39) has no solutions in $0 < r - r_1 = \bar{r} < \delta$, $|z - z_n| < \delta$, and $\Lambda < \Lambda_0$. Namely, in a neighborhood of (z_n, r_1) , the vector field $\partial \tilde{V}_1^T / \partial r$ has no singular points for $\Lambda < \Lambda_0$. Hence, by the invariance of the index sum of singular points in a domain, one can derive (4.3.35).

Let T_1 be the number that $\Lambda(T_1) = \Lambda_0$. It is clear that $T_1 > T_c = (\lambda_0^\varepsilon)^2$. Thus the claim follows from Theorems 5.3.6 and 5.3.7 in Ma and Wang (2005d).

STEP 3. The vector field \tilde{V}_1^T has no singular points in M for $T_c < T < T_1$, and has only one singular point in each domain $(z_k, z_{k+1}) \times (r_1, r_2)$, for $T > T_1$, where $z_k = \frac{k\pi}{a}$ ($0 \leq k \leq K = \frac{aL}{\pi} - 1$).

From (4.3.29) we find that the singular points of \tilde{V}_1^T must be in the lines $l_k = \left\{ \left(\frac{k\pi}{a} + \frac{\pi}{2a}, r \right) \mid r_1 < r < r_2 \right\} \subset M$, and

$$\tilde{V}_{1r}^T = 0, \quad \tilde{V}_{1z}^T = (-1)^k \Lambda R'(r) - \gamma \bar{r}(1 - \bar{r}) \quad \text{on } l_k.$$

Therefore, to prove this claim we only have to consider

$$(-1)^k \frac{R'(r)}{R''(r_1)} - \frac{\Lambda_0}{\Lambda} \bar{r}(1 - \bar{r}) = 0, \quad 0 < \bar{r} < 1. \quad (4.3.40)$$

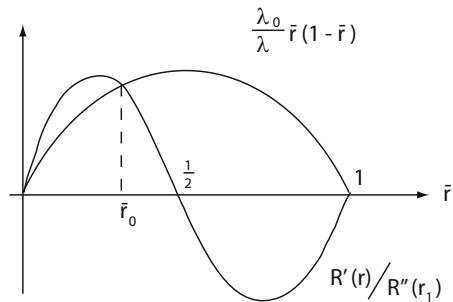
It is known that $R'(r)$ is odd and $\bar{r}(1-\bar{r})$ is an even function on the variable $x = \bar{r} - \frac{1}{2}$. These functions are illustrated in Fig. 4.23.

Direct computations show that (4.3.40) has no zero points in $0 < \bar{r} < 1$ for $\Lambda \leq \Lambda_0$, and there is a unique zero point $\bar{r}_0 \in (0, 1)$ for all $\Lambda_0 < \Lambda$. Furthermore, $0 < \bar{r}_0 < \frac{1}{2}$ as $k=\text{even}$, and $\frac{1}{2} < \bar{r}_0 < 1$ as $k=\text{odd}$. Thus, the claim is proved.

Finally, by Steps 1–3 one readily derives Assertion (1) of the theorem. For the vector field V_2^T given by (4.3.28), we can also prove Assertion (2) in the same fashion. Obviously Assertion (3) holds true.

The proof of the theorem is complete. \square

Fig. 4.23 \bar{r}_0 is a zero point of (4.3.40)



4.3.4 Interior Separation from the Couette–Poiseuille Flow

In this subsection, we study the structural transition of the Couette–Poiseuille flow for the case where the z -direction boundary velocity $W_0 \neq 0$ in (4.3.5). We show that the secondary flow from the Couette–Poiseuille flow (4.3.1) will have interior separation as the Taylor number T exceeds some critical value $T_1 (> T_c)$ which is an increasing function of the nondimensional parameter $\gamma = p_0 d^3 / 4\rho v$.

Let the constant velocity $W_0 \neq 0$ in the boundary condition (4.3.5). Then we have the following interior separation theorems for the TCP flow problem. The results obtained are in agreement with the numerical results obtained by Raguin and Georgiadis; see, e.g., Figure 8 in Raguin and Georgiadis (2004).

Theorem 4.3.9 *Assume that the conditions in Theorem 4.3.2 hold true, and $W_0 \neq 0$ in (4.3.5). Then there exists $\gamma_0 > 0$ such that if $0 < \gamma < \gamma_0$, then for the two (bifurcated) steady-state solutions $v_i^T = \mathcal{U}^{cp} + u_i^T$ ($i = 1, 2$) of (4.2.5) with (4.3.5), the following assertions hold true:*

(1) *For v_1^T , there is a $T_1 > T_c$ such that*

(a) *for $T_c < T < T_1$, the vector field $\tilde{v}_1^T = (W + u_{1z}^T, u_{1r}^T)$ is topologically equivalent to the structure as shown in Fig. 4.24, and*

- (b) for $T_1 < T$, there is exactly one pair of center and saddle points separated from a point in each domain $(\tilde{z}_k, \tilde{z}_{k+1}) \times (r_1, r_2) \subset M$, with $\tilde{z}_k = \frac{k\pi}{a}$ ($0 \leq k \leq K$), as shown in Fig. 4.25.
- (2) For v_2^T there is a number $b_0 > 0$ such that only one of the following two assertions holds true:
- If $|b(\varepsilon)| < b_0$, where $b(\varepsilon)$ is given by (4.3.25), then there is a $T_2 > T_c$ ($T_1 \geq T_2$) such that the same conclusions as Assertion (1) hold true for v_2^T with T_2 replacing T_1 .
 - If $|b(\varepsilon)| > b_0$, then as $T > T_c$, the vector field $\tilde{v}_i^T = (W + u_{2z}^T, u_{2r}^T)$ is topologically equivalent to the structure as shown in Fig. 4.26.
- (3) There exists a $T_3 > T_1$ ($T_1 \geq T_2$), with T_3 independent of γ , such that if $T_i < T < T_3$, the vector field \tilde{v}_i^T ($i = 1, 2$) is topologically equivalent to the structure as shown in Fig. 4.26.

The following theorem is a direct corollary of Theorems 4.3.2, 4.3.9, and the structural stability theorem, Theorem 2.1.2 in Ma and Wang (2005d), and provides a link between the dynamics and the structure in the physical space.

Theorem 4.3.10 Assume that the conditions in Theorem 4.3.2 hold true, $W_0 \neq 0$ in (4.3.5), and $0 < \gamma < \gamma_0$. Then there are \tilde{T}_j ($1 \leq j \leq 3$) with $\tilde{T}_1 = T_1$, $T^* \leq \tilde{T}_2 \leq T_2$, $\tilde{T}_3 = T_3$, and for each $T < \tilde{T}_3$ the space H can be decompose into two disjoint open sets $H = \overline{U}_1^T + \overline{U}_2^T$, $0 \in \partial U_1^T \cap \partial U_2^T$, such that the following assertions hold true:

- For $u_0 \in U_i^T$, there is a time $t_0 > 0$ such that when $t > t_0$, for the solution $u(t, u_0) = (\tilde{u}, u_\theta)$ of (4.2.5) with (4.3.5) and (4.3.14), as $T < \tilde{T}_i$, the vector field \tilde{u} is topologically equivalent to the structure as shown in Fig. 4.24, and as $\tilde{T}_i < T < \tilde{T}_3$ ($i = 1, 2$), \tilde{u} is topologically equivalent to the structure as shown in Fig. 4.26.
- In particular, for the initial value $u_0 \in U_i^T$ near the Couette–Poiseuille flow (4.3.1), and for $T > \tilde{T}_i$ ($i = 1, 2$), there is a time $t_0 > 0$ such that for the solution $u(t, u_0) = (\tilde{u}, u_\theta)$ of (4.2.5) with (4.3.5) and (4.3.14), \tilde{u} has an interior separation at $t = t_0$ from a point in each domain $(\tilde{z}_k, \tilde{z}_{k+1}) \times (r_1, r_2) \subset M$ with $\tilde{z}_k = k\pi/a$, i.e., \tilde{u} is topologically equivalent to the structure as shown in Fig. 4.24 for $t < t_0$, and equivalent to the structure as shown in Fig. 4.26 for $t > t_0$.

Proof of Theorem 4.3.9. The proof is similar to that of Theorem 4.3.5. Here we only need to prove the interior structural bifurcation in Assertion (1) for the following vector field

Fig. 4.24 Vertical shear flow with a constant velocity $W_0 \neq 0$ on boundary

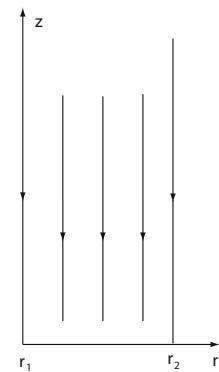


Fig. 4.25 Interior separation

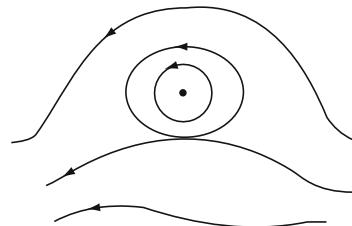
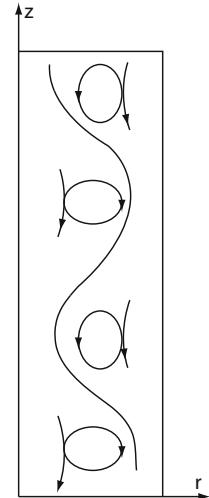


Fig. 4.26 Propagating Taylor vortices separated from the interior



$$\tilde{V}_1^T = (\Lambda \sin azR'(r) - \gamma\bar{r}(1 - \bar{r}) - \gamma W_0, -a\Lambda \cos azR(r)). \quad (4.3.41)$$

We proceed by applying Theorems 5.6.9 and 5.6.10 in Ma and Wang (2005d). It is clear that the zero points of \tilde{V}_1^T must be in the lines

$$l_k = \{(z_k, r) | z_k = \frac{k\pi}{a} + \frac{\pi}{2a}, r_1 < r < r_2\}.$$

Hence for the solutions r_0 of the following equations

$$\pm \Lambda R'(r) - \gamma \bar{r}(1 - \bar{r}) - \gamma W_0 = 0. \quad (4.3.42)$$

we can get the singular points $(z_k, r_0) \in l_k$ of \tilde{V}_1^T .

In addition, we see that for given γ ($0 < \gamma < \gamma_0$) and $W_0 > 0$, there is a $\Lambda_0 > 0$ such that when $\Lambda < \Lambda_0$, the equation (4.3.42) has no solutions, and when $\Lambda = \Lambda_0$, (4.3.42) has exactly two solutions r_0^\pm . In particular, r_0^\pm are the extreme points of the functions:

$$f^\pm(r) = \pm \Lambda_0 R'(r) - \gamma \bar{r}(1 - \bar{r}) - \gamma W_0.$$

Therefore, to find the structural bifurcation points of \tilde{V}_1^T , consider the equations

$$\begin{aligned} \pm \Lambda R'(r) - \gamma \bar{r}(1 - \bar{r}) - \gamma W_0 &= 0, \\ \pm \Lambda R''(r) - \gamma(1 - 2\bar{r}) &= 0. \end{aligned} \quad (4.3.43)$$

By a direct computation, we obtain from (4.3.20) the two unique solutions (Λ_0, r_0^\pm) of (4.3.43) which satisfy that

$$r_0^+ - r_1 = r_2 - r_0^-, \quad r_0^+ < r_*^+ < r_*^- < r_0^-,$$

where r_*^\pm are the extreme points of $R'(r)$, i.e.,

$$R''(r_*^\pm) = 0 \quad (r_*^+ \simeq r_1 + \frac{1}{5}, r_*^- \simeq r_1 + \frac{4}{5}). \quad (4.3.44)$$

The vector field (4.3.41) has the following expression at $\Lambda = \Lambda_0$:

$$\begin{aligned} \tilde{V}_1^T &= V^0 + (\Lambda - \Lambda_0)V^1, \\ V^0 &= (V_z^0, V_r^0) = (\Lambda_0 \sin azR' - \gamma \bar{r}(1 - \bar{r}) - \gamma W_0, -a\Lambda_0 \cos azR), \\ V^1 &= (V_z^1, V_r^1) = (\sin azR', -a \cos azR). \end{aligned}$$

It is easy to see that for the solutions (Λ_0, r_0^\pm) of (4.3.38) the points (z_{2k}, r_0^+) and $(z_{2k+1}, r_0^-) \in M$ are singular points of V^0 , where $z_m = \frac{m\pi}{a} + \frac{\pi}{2a}$.

For simplicity, we only consider the structural bifurcation of \tilde{V}_1^T at the singular point $(z_0, r_0) = (\frac{\pi}{2a}, r_0^+)$, and the proof for the other singular points this proof is the same.

In the previous discussion, we have known that as $\Lambda < \Lambda_0$, the vector field \tilde{V}_1^T given by (4.3.41) has no singular point in M , and (z_0, r_0) is an isolated singular point of \tilde{V}_1^T at $\Lambda = \Lambda_0$. Therefore we have

$$\text{ind}(V^0, (z_0, r_0)) = 0. \quad (4.3.45)$$

In addition, we see that

$$DV^0(z_0, r_0) = \begin{pmatrix} \frac{\partial V_z^0}{\partial z} & \frac{\partial V_z^0}{\partial r} \\ \frac{\partial V_r^0}{\partial z} & \frac{\partial V_r^0}{\partial r} \end{pmatrix}_{(z_0, r_0)} = \begin{pmatrix} 0 & 0 \\ a^2 \Lambda_0 R(r_0) & 0 \end{pmatrix} \neq 0, \quad (4.3.46)$$

and the eigenvector e_1 and e_2 satisfying

$$\begin{aligned} DV^0(z_0, r_0)e_2 &= \alpha e_1 \quad (\alpha = a^2 \Lambda_0 R(r_0)), \\ DV^0(z_0, r_0)e_1 &= 0, \end{aligned}$$

are given by $e_1 = (0, 1)$, $e_2 = (1, 0)$. Hence we have

$$V^1(z_0, r_0) \cdot e_2 = V_z^1(z_0, r_0) = R'(r_0) \neq 0. \quad (4.3.47)$$

From (4.3.45) to (4.3.47) we find that the conditions (5.6.16)–(5.6.18) in Theorems 5.6.9 and 5.6.10 in Ma and Wang (2005d) hold true for the vector field \tilde{V}_1^T at $\Lambda = \Lambda_0$ and $(z, r) = (z_0, r_0)$.

Finally, we verify the condition (5.6.19) in Ma and Wang (2005d). By (4.3.43) we have

$$\frac{\partial^k (V^0 \cdot e_2)}{\partial e_1^k} \Big|_{(z_0, r_0)} = \frac{\partial^k V_z^0(z_0, r_0)}{\partial r^k} = 0 \quad \text{for } k = 0, 1, \quad (4.3.48)$$

$$\frac{\partial^2 V_z^0(z_0, r_0)}{\partial r^2} = \Lambda_0 R'''(r_0) + 2\gamma. \quad (4.3.49)$$

By (4.3.43) we see that

$$\frac{\gamma}{\Lambda_0} = \frac{R''(r_0)}{1 - 2(r_0 - r_1)}. \quad (4.3.50)$$

On the other hand, it is easy to check that

$$-\frac{R'''(r)}{R''(r)} > \frac{2}{1 - 2(r - r_1)} \quad \forall r_1 \leq r < r_*^+, \quad (4.3.51)$$

where r_*^+ satisfies (4.3.44). It follows from (4.3.49)–(4.3.51) that

$$\frac{\partial^2 V_z^0(z_0, r_0)}{\partial r^2} \neq 0. \quad (4.3.52)$$

Thus we derive from (4.3.48) and (4.3.52) the condition (5.6.19) in Ma and Wang (2005d) with $m = 2$. By the interior separation theorem of incompressible flows, Theorem 5.6.10 in Ma and Wang (2005d), the vector field \tilde{V}_1^T has an interior separation from each point (y_m, r_0) at $\Lambda = \Lambda_0$, where $r_0 = r_0^+$ for $y_m = \frac{2k\pi}{a} + \frac{\pi}{2a}$, and $r_0 = r_0^-$ for $y_m = \frac{(2k+1)\pi}{a} + \frac{\pi}{2a}$.

The proof of the theorem is complete. \square

4.3.5 Nature of Boundary-Layer and Interior Separations

The boundary-layer and interior separations of incompressible viscous fluid flow play a fundamental role in many hydrodynamic problems, and often lead to more complicated turbulent behaviors. Hence it is very important to understand in both dynamical and kinematic aspects the nature of flow's separations from the boundary and the interior of fluid. These results established in this section give a basic example in fluid dynamics to show when, where, and how the separations occur. Moreover they also provide a link between the dynamics and the kinematic theories of the boundary-layer and interior separation developed by the authors, in collaboration in part with Ghil; see Ma and Wang (2005d, 2001, 2004a,c), Ghil et al. (2001, 2005).

Theorems 4.3.5–4.3.10 reveal an important phenomenon that the structural transitions of a fluid flow undergo in general two bifurcations: the first is the dynamic bifurcation in the hydrodynamic equations, and the second is the structural bifurcation for the bifurcated solutions of the dynamical equations. Meanwhile, the proofs of these theorems also exhibit the mechanism of boundary-layer and interior separations generated by the two types of bifurcations. In the following, we shall briefly illustrate this mechanism in the context of the Navier–Stokes equations.

Consider the Navier–Stokes equations

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla) u &= \mu \Delta u - \nabla p, \\ \operatorname{div} u &= 0, \end{aligned} \quad (4.3.53)$$

on a two-dimensional channel: $M = \{(x_1, x_2) \in \mathbb{R}^2 \mid 0 < x_2 < L\}$ with the width $L > 0$. Equation (4.3.53) admits the following steady-state solutions, which describe parallel shear flows in M ,

$$u_0 = (V(x_2), 0), \quad (4.3.54)$$

where V satisfies $d^2V/dx_2^2 = \text{constant}$. Therefore V is given by

$$V_\lambda(x_2) = \lambda x_2^2 + \alpha_1(\lambda)x_2 + \alpha_2(\lambda).$$

Here $\alpha_1(\lambda)$ and $\alpha_2(\lambda)$ are two arbitrary functions of λ .

Consider the translation

$$u = v + u_0.$$

Then the equations (4.3.53) become

$$\begin{aligned} \frac{\partial v_1}{\partial t} + (v \cdot \nabla)v_1 &= \mu \Delta v_1 - (2\lambda x_2 + \alpha_1(\lambda))v_2 - V_\lambda \frac{\partial v_1}{\partial x_1} - \frac{\partial p}{\partial x_1}, \\ \frac{\partial v_2}{\partial t} + (v \cdot \nabla)v_2 &= \mu \Delta v_2 - V_\lambda \frac{\partial v_2}{\partial x_1} - \frac{\partial p}{\partial x_2}, \\ \operatorname{div} v &= 0, \end{aligned} \quad (4.3.55)$$

supplemented with the following boundary conditions:

$$\begin{aligned} v &= 0 && \text{at } x_2 = 0, L, \\ v &\text{ is periodic} && \text{in the } x_1\text{-direction.} \end{aligned} \quad (4.3.56)$$

Consider the eigenvalue problem

$$\begin{aligned} -\mu\Delta u_1 + (2\lambda x_2 + \alpha_1(\lambda))u_2 + V_\lambda \frac{\partial u_1}{\partial x_1} + \frac{\partial p}{\partial x_1} &= \beta(\lambda)u_1, \\ -\mu\Delta u_2 + V_\lambda \frac{\partial u_2}{\partial x_1} + \frac{\partial p}{\partial x_2} &= \beta(\lambda)u_2, \\ \operatorname{div} u &= 0. \end{aligned} \quad (4.3.57)$$

Let the eigenvalues $\beta_j(\lambda)$ of (4.3.57) with (4.3.56) at $\lambda = \lambda_0$ satisfy the PES (1.2.4) and (1.2.5).

By the transition theorem, Theorem 2.1.3, and the existence of global attractors for the Navier–Stokes equations (Temam, 1997), the problem (4.3.55) with (4.3.56) undergoes a transition to a minimal attractor Σ_λ . It is clear that each solution $v^\lambda \in \Sigma_\lambda$ of (4.3.55) with (4.3.56) is not shear flow, hence v^λ has vortices in M , as shown in Fig. 4.27.

Thus, as $\lambda > \lambda_0$ the equations (4.3.53) possess new stable states

$$u^*(x, \lambda) = u_0 + v^\lambda = (V_\lambda(x_2) + v_1^\lambda, v_2^\lambda),$$

where $v^\lambda = (v_1^\lambda, v_2^\lambda)$. In other hand, we can see in Fig. 4.27 that there are curves $l_k \subset \overline{M}$ such that

$$\operatorname{sign} V_\lambda = -\operatorname{sign} v_1^\lambda, \quad \text{and} \quad v_2^\lambda = 0, \quad \text{on } l_k.$$

It implies that if

$$\begin{aligned} \left| \frac{\partial v_1^{\lambda_0}}{\partial x_2} \right| &< \left| \frac{\partial V_{\lambda_0}}{\partial x_2} \right| && \text{at } l_k \cap \partial M, \\ \frac{d}{d\lambda} \left| \frac{\partial v_1^\lambda}{\partial x_2} \right| &> \frac{d}{d\lambda} \left| \frac{\partial V_\lambda}{\partial x_2} \right| && \text{at } l_k \cap \partial M, \text{ for } \lambda > \lambda_0, \end{aligned} \quad (4.3.58)$$

then there exists a $\lambda_1 > \lambda_0$ such that

$$\begin{aligned} \frac{\partial}{\partial x_2} u^*(x, \lambda) &\neq 0 && \forall x \in \partial M \text{ for } \lambda_0 \leq \lambda < \lambda_1, \\ \frac{\partial}{\partial x_2} u^*(x_0, \lambda_1) &= 0 && \text{for some } x_0 \in l_k \cap \partial M. \end{aligned} \quad (4.3.59)$$

Thus, (4.3.58) and (4.3.59) mean that

$$\text{ind}\left(\frac{\partial u^*(\cdot, \lambda_1)}{\partial x_2}, x_0\right) = 0, \quad \frac{\partial}{\partial \lambda} \frac{\partial u^*(x_0, \lambda_1)}{\partial x_2} \neq 0.$$

Hence, by Theorems 5.3.6 and 5.3.7 in Ma and Wang (2005d), a boundary-layer separation at $x_0 \in \partial M$ on $\lambda > \lambda_1$ for $u^*(x, \lambda)$ will occur.

Instead of (4.3.58) if

$$\begin{aligned} |v_1^{\lambda_0}| &< |V_{\lambda_0}| && \text{on } l_k \cap M, \\ \frac{d}{d\lambda} |v_1^\lambda| &> \frac{d}{d\lambda} |V_\lambda| && \text{on } l_k \cap M \text{ for } \lambda > \lambda_0, \end{aligned} \quad (4.3.60)$$

then there is $\lambda_1 > \lambda_0$ such that

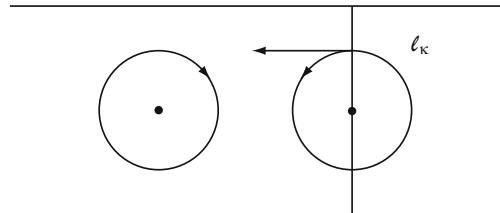
$$\begin{aligned} u^*(x, \lambda) &\neq 0 && \forall x \in M \text{ for } \lambda_0 \leq \lambda < \lambda_1, \\ u^*(x_0, \lambda_1) &= 0 && \text{for some } x_0 \in l_k \cap M. \end{aligned} \quad (4.3.61)$$

We infer from (4.3.60) and (4.3.61) that

$$\text{ind}(u^*(\cdot, \lambda_1), x_0) = 0, \quad \frac{\partial}{\partial \lambda} u^*(x_0, \lambda_1) \neq 0.$$

Thus, by Theorems 5.6.9 and 5.6.10 in Ma and Wang (2005d), the stable state $u^*(x, t)$ of (4.3.53) has an interior separation at $x_0 \in M$ on $\lambda > \lambda_1$.

Fig. 4.27 The vortex structure of the transited solution $v^\lambda = (v_1^\lambda, v_2^\lambda)$



We note that there are no external forces in the equations (4.2.5) and (4.3.53) to directly act on the fluid flows. As seen in the previous discussion, in this case the cause leading to the boundary-layer and interior separation is the transitions between the stable steady-state solutions of hydrodynamic equations.

However, if there exists an applied force, the mechanism to generate a structural transition is different. To demonstrate this point, in the following we introduce a structural bifurcation theorem for this case, which was proved by Ma and Wang (2004a, 2005d).

Consider the following problem

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla) u - \mu \Delta u + \nabla p &= \lambda f, \\ \operatorname{div} u &= 0, \\ u|_{\partial M} &= 0, \\ u(x, 0) &= u_0, \end{aligned} \tag{4.3.62}$$

where $u = (u_1, u_2)$, $M \subset \mathbb{R}^2$ is an open set, and $u_0|_{\partial\Omega} = 0$.

The following theorem provides a sufficient condition for structural bifurcation near boundary of the solutions of (4.3.62) generated by the initial date and forcing.

Theorem 4.3.11 *Let $u_0 \in C^{2+\alpha}(M, \mathbb{R}^2)$ with $u_0|_{\partial M} = 0$, $f \in C^\alpha(M, \mathbb{R}^2)$ ($0 < \alpha < 1$) and $u \in C^1([0, \infty), C^{2+\alpha}(M, \mathbb{R}^2))$ be a solution of (4.3.62). Assume that $\Gamma \subset \partial M$ is a connected component of ∂M , and*

$$\frac{\partial(f \cdot \tau)}{\partial n} \cdot \frac{\partial(u_0 \cdot \tau)}{\partial n} < 0 \text{ on } \Gamma, \tag{4.3.63}$$

where n and τ are the normal and tangent vectors at ∂M . Then, there is a $\lambda_0 > 0$, depending on $\|f\|_{C^1}$, $\|u_0\|_{C^3}$ and $\mu > 0$, such that for any $\lambda > \lambda_0$ there exist a time $t_0 > 0$ and point $\bar{x} \in \Gamma$, the solution $u(x, t, u_0)$ has a structural bifurcation at (\bar{x}, t_0) in its global structure, and the index $\operatorname{ind}(u(\cdot, t_0), \bar{x}) = -k$, $k \geq 0$ an integer.

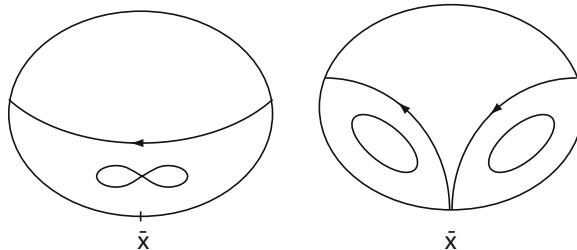


Fig. 4.28 An example of flow transitions for the case with index -1

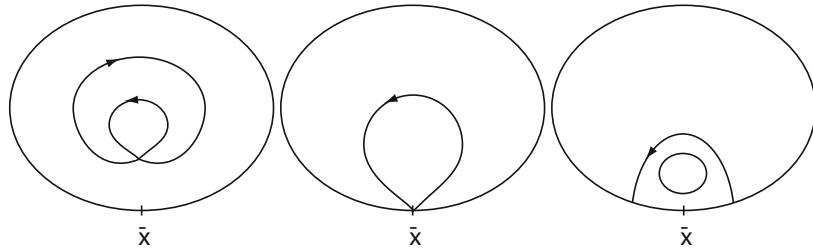


Fig. 4.29 An example of flow transitions for the case with index -1

Remark 4.3.12 Condition (4.3.63) implies that the direction of the external force f is inverse to that of the initial velocity u_0 near the boundary Γ . In fact, only two kinds of structural bifurcations of 2D divergence free vector fields are generic: the indices are 0 and -1 , namely in the real world we can only observe the structural bifurcation of fluid flows in which the index is either 0 or -1 ; see Ma and Wang (2005d). The structural bifurcation from index zero is the well-known boundary-layer separation, and from index -1 has two types of flow transitions, which are shown respectively in Figs. 4.28 and 4.29.

4.4 Rotating Convection Problem

The stability and transition of a rotating flow are of basic importance in geophysical fluid dynamics, and in particular, in the study of the atmospheric and oceanic circulations (Pedlosky, 1987; Ghil and Childress, 1987; Salby, 1996). As we know, both the atmospheric and oceanic flows are flows under the rotation of the earth. In fact, fast rotation and small aspect ratio are two main characteristics of the large scale atmospheric and oceanic motion.

In this section, we address the thermal instability of a rotating fluid, and study the stability and transition of basic flows governed by the rotating Boussinesq equations. This study is based on recent articles (Hsia et al., 2007, 2010).

We show in particular that there are only two types of transitions, Type-I (continuous) and Type-II (jump) transitions, as the Rayleigh number crosses the first real or complex eigenvalues. Specific criteria are given to determine the type of transitions as well. The detailed analysis is mainly based on (1) the study of eigenvalue problem for the linearized equations around the basic state, and (2) the reduction to the center manifold. In comparison to the classical Rayleigh–Bénard convection (Ma and Wang, 2004b, 2007a), the linear problem here is non-self-adjoint, leading to much more complicated spectrum, and more complicated dynamical behaviors. We derive in particular two critical Rayleigh numbers R_{c_1} and R_{c_2} such that R_{c_1} leads to the onset of the transition of steady-state solutions, and R_{c_2} leads to the onset of the transition of periodic solutions (Hopf bifurcation). Both parameters are explicitly given in terms of the physical parameters. In addition, we derive two parameters b_1 and b_2 characterized by the physical parameters to determine the types of transitions of the steady-state solutions and the periodic solutions respectively. It is proved that the transition of steady-state solutions is Type-I (continuous) if $b_1 < 0$, and is of Type-II (jump) if $b_1 > 0$. The transition of periodic solutions is Type-I if $b_2 < 0$, is Type-II (jump) if $b_2 > 0$.

4.4.1 Rotating Boussinesq Equations

The rotating Boussinesq equations are basic equations in the geophysical fluid dynamics, and their nondimensional form is given by

$$\begin{aligned}\frac{\partial u}{\partial t} &= \text{Pr } \Delta u + \text{Pr } RT k - \omega \bar{k} \times u - (u \cdot \nabla) u - \text{Pr } \nabla p, \\ \frac{\partial T}{\partial t} &= \Delta T + u_3 - (u \cdot \nabla) T, \\ \text{div } u &= 0,\end{aligned}\tag{4.4.1}$$

for (x_1, x_2, x_3) in the nondimensional domain $\Omega = \mathbb{R}^2 \times (0, 1)$, where $u = (u_1, u_2, u_3)$ is the velocity field, $k = (0, 0, 1)$ the unite vector in the x_3 -direction, Pr the Prandtl number, R the Rayleigh number given by (4.1.1), ω a constant proportional to the rotating angular velocity, and T the temperature. In particular, the term $\omega k \times u$ stands for the Coriolis force in the atmosphere-ocean dynamics.

At the top and bottom boundaries, we impose the free-free boundary conditions:

$$(T, u_3) = 0, \quad \frac{\partial u}{\partial x_3} = \frac{\partial u_2}{\partial x_3} = 0 \quad \text{at } x_3 = 0, 1.\tag{4.4.2}$$

In the x_1 and x_2 directions we use the periodic boundary condition

$$(u, T)(x_1, x_2, x_3) = (u, T)(x_1 + 2jL_1\pi, x_2 + 2kL\pi, x_3) \quad \forall j, k \in \mathbb{Z}.\tag{4.4.3}$$

For the problems (4.4.1)–(4.4.3), one of the following two constrained conditions is adopted:

$$\int_{\Omega} u_1 dx = \int_{\Omega} u_2 dx = 0, \quad \text{and}\tag{4.4.4}$$

$$u_1(0, 0, x_3) = u_2(0, 0, x_3) = 0.\tag{4.4.5}$$

The constraint (4.4.5) means that the fluid flow rises or descends along the rotating axis, which is a natural condition in many fluid dynamical problems. Physically, this condition indicates the rotation center lies at the point $x = 0$. In particular, we can see this characteristic in the Polar cell of the latitudinal circulation in the atmospheric physics. For (4.4.4), we set the spaces

$$H = \{(u, T) \in L^2(\Omega, \mathbb{R}^4) \mid \text{div } u = 0, u|_{x_3=0,1} = 0, u \text{ satisfies (4.4.3)}\},$$

$$H_1 = \{(u, T) \in H^2(\Omega, \mathbb{R}^4) \cap H \mid (u, T) \text{ satisfies (4.4.2)–(4.4.4)}\},$$

and for the condition (4.4.5) let

$$\tilde{H} = \{(u, T) \in H \mid (u_1, u_2, u_3, T)(-x_1, -x_2, x_3) = (-u_1, -u_2, u_3, T)(x)\},$$

$$\tilde{H}_1 = H_1 \cap \tilde{H}.$$

It is clear that $(u, T) \in H$ satisfies (4.4.5) if and only if $(u, T) \in \tilde{H}$.

Let $L_R = A + B_R : H_1 \rightarrow H$ (resp., $\tilde{H}_1 \rightarrow \tilde{H}$) and $G : H_1 \rightarrow H$ (resp., $\tilde{H}_1 \rightarrow \tilde{H}$) be defined by

$$\begin{aligned} A\psi &= P(\operatorname{Pr} \Delta u, \Delta T), & B_R\psi &= P(\operatorname{Pr} RTk - \omega k \times u, u_3), \\ G(\psi, \tilde{\psi}) &= P(-(u \cdot \nabla) \tilde{u}, -(u \cdot \nabla) \tilde{T}), & G(\psi) &= G(\psi, \psi), \end{aligned}$$

for any $\psi = (u, T) \in H_1$ (resp. \tilde{H}_1), where $P : L^2(\Omega, \mathbb{R}^4) \rightarrow H$ is the Leray projection.

We note that \tilde{H}_1 and \tilde{H} are invariant under the bilinear operator G . Hence, \tilde{H}_1 and \tilde{H} are invariant under the operator $L_R + G$. Thus, the problems (4.4.1)–(4.4.3) with (4.4.4) (or with (4.4.5)) can be written in the following operator form

$$\frac{d\psi}{dt} = L_R\psi + G(\psi) \quad \text{for } \psi = (u, T) \in H_1 \text{ (or } \tilde{H}_1\text{).} \quad (4.4.6)$$

The existence and basic properties of solutions for rotating Boussinesq equations are known; see Lions et al. (1992b).

4.4.2 Eigenvalue Problem

To study the transition of (4.4.1)–(4.4.3), consider the following eigenvalue problem:

$$L_R\psi = \beta(R)\psi \quad \text{for } \psi \in H_1 \text{ (or } \tilde{H}_1\text{).} \quad (4.4.7)$$

The equations associated with (4.4.7) are given by

$$\begin{aligned} \operatorname{Pr} \Delta u + \operatorname{Pr} RTk - \omega k \times u - \operatorname{Pr} \nabla p &= \beta u, \\ \Delta T + u_3 &= \beta T, \\ \operatorname{div} u &= 0, \end{aligned} \quad (4.4.8)$$

supplemented with (4.4.2) and (4.4.3).

Under the constraint (4.4.5), for $\psi = (u_1, u_2, u_3, T) \in \tilde{H}_1$, the eigenvectors of (10.67) have the separation of variables as follows

$$\psi = \begin{cases} u_{jk}(x_3) \sin(j\alpha_1 x_1 + k\alpha_2 x_2), \\ v_{jk}(x_3) \sin(j\alpha_1 x_1 + k\alpha_2 x_2), \\ w_{jk}(x_3) \cos(j\alpha_1 x_1 + k\alpha_2 x_2), \\ T_{jk}(x_3) \cos(j\alpha_1 x_1 + k\alpha_2 x_2) \end{cases} \quad \text{for } j, k \in \mathbb{Z} \quad (4.4.9)$$

$$p = p_{jk}(x_3) \cos(j\alpha_1 x_1 + k\alpha_2 x_2) \quad \text{for } j, k \in \mathbb{Z} \quad (4.4.10)$$

$$\alpha_1 = L_1^{-1}, \quad \alpha_2 = L_2^{-1}. \quad (4.4.11)$$

If the condition (4.4.5) is replaced by (4.4.4), then for each $(j, k) \in \mathbb{Z}^2$ with $(j, k) \neq (0, 0)$ we have another eigenvector given by

$$\tilde{\psi} = \begin{cases} -u_{jk}(x_3) \cos(j\alpha_1 x_1 + k\alpha_2 x_2), \\ -v_{jk}(x_3) \cos(j\alpha_1 x_1 + k\alpha_2 x_2), \\ w_{jk}(x_3) \sin(j\alpha_1 x_1 + k\alpha_2 x_2), \\ T_{jk}(x_3) \sin(j\alpha_1 x_1 + k\alpha_2 x_2), \end{cases} \quad (4.4.12)$$

$$\tilde{p} = p_{jk} \sin(j\alpha_1 x_1 + k\alpha_2 x_2). \quad (4.4.13)$$

Putting (ψ, p) and $(\tilde{\psi}, \tilde{p})$ into (4.4.8) respectively, we deduce from (4.4.9) to (4.4.3) that $(u_{jk}, v_{jk}, w_{jk}, T_{jk})$ satisfies the following system of ODEs:

$$\begin{aligned} \text{Pr}(D_{jk}u_{jk} + j\alpha_1 p_{jk}) + \omega v_{jk} &= \beta u_{jk}, \\ \text{Pr}(D_{jk}v_{jk} + k\alpha_2 p_{jk}) - \omega u_{jk} &= \beta v_{jk}, \\ \text{Pr}(D_{jk}w_{jk} - Dp_{jk} + RT_{jk}) &= \beta w_{jk}, \\ D_{jk}T_{jk} + w_{jk} &= \beta T_{jk}, \\ j\alpha_1 u_{jk} + k\alpha_2 v_{jk} + Dw_{jk} &= 0, \\ Du_{jk} = Dv_{jk} = w_{jk} = T_{jk} &= 0 \quad \text{at } x_3 = 0, 1, \end{aligned} \quad (4.4.14)$$

for any $j, k \in \mathbb{Z}$ with $(j, k) \neq 0$, where

$$D = d/dx_3, \quad D_{jk} = D^2 - \alpha_{jk}^2, \quad \alpha_{jk}^2 = j^2\alpha_1^2 + k^2\alpha_2^2.$$

We infer from (4.4.14) that w_{jk} and \tilde{w}_{jk} satisfy the equation

$$\begin{aligned} &\left[(\text{Pr } D_{jk} - \beta)(\text{Pr } D_{jk} - \beta)^2 D_{jk} + \omega^2(D_{jk} - \beta)D^2 \right. \\ &\quad \left. + \text{Pr } Ra_{jk}^2(\text{Pr } D_{jk} - \beta) \right] w_{jk} = 0 \end{aligned} \quad (4.4.15)$$

$$w_{jk} = D^2 w_{jk} = D^4 w_{jk} = D^6 w_{jk} = 0 \quad \text{at } x_3 = 0, 1. \quad (4.4.16)$$

By (4.4.16), w_{jk} are the sine functions

$$w_{jk}(x_3) = \sin l\pi x_3 \quad \text{for } l = 1, 2, \dots. \quad (4.4.17)$$

In addition, we can deduce from (4.4.14) that

$$(\text{Pr } D_{jk} - \beta)u_{jk} = \frac{-j\alpha_1}{\alpha_{jk}^2}(\text{Pr } D_{jk} - \beta - \frac{k\alpha_2}{j\alpha_1}\omega)Dw_{jk}, \quad (4.4.18)$$

$$(\text{Pr } D_{jk} - \beta)v_{jk} = \frac{-k\alpha_2}{\alpha_{jk}^2}(\text{Pr } D_{jk} - \beta + \frac{j\alpha_1}{k\alpha_2}\omega)Dw_{jk}, \quad (4.4.19)$$

$$(D_{jk} - \beta)T_{jk} = -w_{jk}. \quad (4.4.20)$$

Substituting (4.4.17) into (4.4.15), we see that the eigenvalues β of (4.4.7) satisfy the cubic equations

$$\gamma_{jkl}^2(\beta + \gamma_{jkl}^2)(\beta + Pr\gamma_{jkl}^2)^2 + l^2\pi^2\omega^2(\beta + \gamma_{jkl}^2) - Pr Ra_{jk}^2(\beta + Pr\gamma_{jkl}^2) = 0, \quad (4.4.21)$$

for $(j, k, l) \in \mathbb{Z}^2 \times \mathbb{N}$ with $(j, k) \neq (0, 0)$, where

$$\gamma_{jkl}^2 = \alpha_{jk}^2 + l^2\pi^2 = j^2\alpha_1^2 + k^2\alpha_2^2 + l^2\pi^2.$$

Consider the following index sets

$$I_1 = \{(j, k, l) \mid (j, k) \in \mathbb{Z}^2, j \geq 0, (j, k) \neq (0, 0), l = 1, 2, \dots, \},$$

$$I_2 = \{(j, k, 0) \mid (j, k) \in \mathbb{Z}^2, j \geq 0, (j, k) \neq (0, 0)\},$$

$$I_3 = \{(0, 0, l) \mid l = 1, 2, \dots, \}.$$

FIRST, for $(j, k, 0) \in I_2$, we have the following eigenvalues and eigenvectors of (4.4.8) with (4.4.2) and (4.4.3).

$$\beta_{jk0} = -Pr\alpha_{jk}^2 = -Pr(j^2\alpha_1^2 + k^2\alpha_2^2),$$

$$\psi_{jk0} = (k\alpha_2 \sin(j\alpha_1 x_1 + k\alpha_2 x_2), -j\alpha_1 \sin(j\alpha_1 x_1 + k\alpha_2 x_2), 0, 0),$$

$$\tilde{\psi}_{jk0} = (-k\alpha_2 \cos(j\alpha_1 x_1 + k\alpha_2 x_2), j\alpha_1 \cos(j\alpha_1 x_1 + k\alpha_2 x_2), 0, 0);$$

namely

$$L_R\psi_{jk0} = \beta_{jk0}\psi_{jk0}, \quad L_R\tilde{\psi}_{jk0} = \beta_{jk0}\tilde{\psi}_{jk0}.$$

We denote

$$\mathcal{B}_2 = \{\beta_{jk0} = -Pr\alpha_{jk} \mid (j, k, 0) \in I_2\},$$

$$E_2 = \{\psi_{jk0}, \tilde{\psi}_{jk0} \mid (j, k, 0) \in I_2\},$$

$$\tilde{E}_2 = \{\psi_{jk0} \mid (j, k, 0) \in I_2\}.$$

Then, \mathcal{B}_2 is the set of all eigenvalues corresponding to index set I_2 , E_2 is the set of all eigenvectors associated with \mathcal{B}_2 for the condition (4.4.4), and \tilde{E}_2 the set of eigenvectors for the condition (4.4.5).

SECOND, for $(0, 0, l) \in I_3$, the eigenvalues and eigenvectors are given by

$$\beta_{00l}^1 = -l^2\pi^2, \quad \beta_{00l}^2 = -Pr l^2\pi^2 + i\omega, \quad \beta_{00l}^3 = -Pr l^2\pi^2 - i\omega,$$

$$\psi_{00l}^1 = (0, 0, 0, \sin l\pi x_3),$$

$$\psi_{00l}^2 = (\cos l\pi x_3, 0, 0, 0),$$

$$\psi_{00l}^3 = (0, \cos l\pi x_3, 0, 0),$$

which satisfy

$$\begin{aligned} L_R \psi_{00l}^1 &= \beta_{00l}^1 \psi_{00l}^1, \\ L_R \psi_{00l}^2 &= -\text{Pr } l^2 \pi^2 \psi_{00l}^2 + \omega \psi_{00l}^3, \\ L_R \psi_{00l}^3 &= -\omega \psi_{00l}^2 - \text{Pr } l^2 \pi^2 \psi_{00l}^3. \end{aligned}$$

We denote the sets of eigenvalues and eigenvectors corresponding to I_3 for the conditions (4.4.4) and (4.4.5) respectively as follows

$$\begin{aligned} \mathcal{B}_3 &= \{\beta_{00l}^1, \beta_{00l}^2, \beta_{00l}^3 \mid l = 1, 2, \dots\}, & E_3 &= \{\psi_{00l}^1, \psi_{00l}^2, \psi_{00l}^3 \mid l = 1, 2, \dots\}, \\ \tilde{\mathcal{B}}_3 &= \{\beta_{00l}^1 \mid l = 1, 2, \dots\}, & \tilde{E}_3 &= \{\psi_{00l}^1 \mid l = 1, 2, \dots\}. \end{aligned}$$

THIRD, for $(j, k, l) \in I_1$, it is known that all solutions β_{jkl} of the cubic equations (4.4.21) are eigenvalues of (4.4.7). Let $\beta_{jkl}^i (R) (1 \leq i \leq 3)$ be three zeros of (4.4.21) with

$$\text{Re}\beta_{jkl}^1 \geq \text{Re}\beta_{jkl}^2 \geq \text{Re}\beta_{jkl}^3.$$

Then, all eigenvalues corresponding to I_1 are given by

$$\mathcal{B}_1 = \{\beta_{jkl}^i(R) \in D \mid \beta_{jkl}^i \text{ satisfies (4.4.21), } (j, k, l) \in I_1, 1 \leq i \leq 3\}.$$

Let $\psi_{jkl}^i, \tilde{\psi}_{jkl}^i \in H_1 (\psi_{jkl}^i \in \tilde{H}_1)$ be the eigenvectors corresponding to $\beta_{jkl}^i \in \mathcal{B}_1$. By (4.4.9)–(4.4.13) and (4.4.17)–(4.4.20) with β being replaced by β_{jkl}^i , we obtain the expressions of ψ_{jkl}^i and $\tilde{\psi}_{jkl}^i (1 \leq i \leq 3)$ as follows:

$$\psi_{jkl}^i = \begin{cases} u_{jkl}^i \sin(j\alpha_1 x_1 + k\alpha_2 x_2) \cos l\pi x_3, \\ v_{jkl}^i \sin(j\alpha_1 x_1 + k\alpha_2 x_2) \cos l\pi x_3, \\ w_{jkl}^i \cos(j\alpha_1 x_1 + k\alpha_2 x_2) \sin l\pi x_3, \\ T_{jkl}^i \cos(j\alpha_1 x_1 + k\alpha_2 x_2) \sin l\pi x_3, \end{cases} \quad (4.4.22)$$

$$\tilde{\psi}_{jkl}^i = \begin{cases} -u_{jkl}^i \cos(j\alpha_1 x_1 + k\alpha_2 x_2) \cos l\pi x_3, \\ -v_{jkl}^i \cos(j\alpha_1 x_1 + k\alpha_2 x_2) \cos l\pi x_3, \\ w_{jkl}^i \sin(j\alpha_1 x_1 + k\alpha_2 x_2) \sin l\pi x_3, \\ T_{jkl}^i \sin(j\alpha_1 x_1 + k\alpha_2 x_2) \sin l\pi x_3, \end{cases} \quad (4.4.23)$$

where

$$\begin{aligned} u_{jkl}^i &= -\frac{l\pi}{\alpha_{jk}^2} (j\alpha_1 + k\alpha_2 \omega A_1(\beta_{jkl}^i)), & v_{jkl}^i &= -\frac{l\pi}{\alpha_{jk}^2} (k\alpha_2 - j\alpha_1 \omega A_1(\beta_{jkl}^i)), \\ w_{jkl}^i &= 1, & T_{jkl}^i &= A_2(\beta_{jkl}^i), \\ A_1(\beta) &= \frac{1}{\text{Pr } \gamma_{jkl}^2 + \beta}, & A_2(\beta) &= \frac{1}{\gamma_{jkl}^2 + \beta}, \end{aligned}$$

When $\beta_{jkl}^q = \bar{\beta}_{jkl}^{q+1} = \sigma + i\rho$ are complex numbers, ψ_{jkl}^i and ψ_{jkl}^i ($i = q$) are complex functions. Therefore, the eigenvectors ϕ_{jkl}^q , ϕ_{jkl}^{q+1} and $\tilde{\phi}_{jkl}^q$, $\tilde{\phi}_{jkl}^{q+1}$ corresponding to β_{jkl}^q and β_{jkl}^{q+1} are expressed as

$$\begin{aligned}\phi_{jkl}^q &= Re\psi_{jkl}^i, & \phi_{jkl}^{q+1} &= I_m\psi_{jkl}^i, \\ \tilde{\phi}_{jkl}^q &= Re\tilde{\psi}_{jkl}^i, & \tilde{\phi}_{jkl}^{q+1} &= I_m\tilde{\psi}_{jkl}^i,\end{aligned}\quad (4.4.24)$$

for $i = q$. In this case, ϕ_{jkl}^q and ϕ_{jkl}^{q+1} satisfy

$$\begin{aligned}L_r\phi_{jkl}^q &= \sigma\phi_{jkl}^q - \rho\phi_{jkl}^{q+1}, \\ L_R\phi_{jkl}^{q+1} &= \rho\phi_{jkl}^q + \sigma\phi_{jkl}^{q+1}.\end{aligned}\quad (4.4.25)$$

Thus, the sets of all eigenvectors corresponding to I_1 are given by

$$\begin{aligned}E_1 &= \{\psi_{jkl}^i, \tilde{\psi}_{jkl}^i \mid (j, k, l) \in I_1, 1 \leq i \leq 3\}, \\ \widetilde{E}_1 &= \{\psi_{jkl}^i \mid (j, k, l) \in I_1, 1 \leq i \leq 3\}.\end{aligned}$$

In view of the Fourier expansion, it is clear that $E = E_1 \cup E_2 \cup E_3$ is a basis of H , and $\widetilde{E} = \widetilde{E}_1 \cup \widetilde{E}_2 \cup \widetilde{E}_3$ is a basis of \widetilde{H} , and we have the following lemma:

- Lemma 4.4.1(a)** *The set $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3$ consists of all eigenvalues of (4.4.8) with (4.4.2)–(4.4.4), and the (generalized) eigenvectors in E form a basis of H .*
- (b) *The set $\widetilde{\mathcal{B}} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3$ consists of all eigenvalues of (4.4.8) with (4.4.2), (4.4.3) and (4.4.5), and the (generalized) eigenvectors in \widetilde{E} form a basis of \widetilde{H} .*
 - (c) *$Re\beta < 0$ for each $\beta \in \mathcal{B}_2 \cup \mathcal{B}_3$.*
 - (d) *Only the eigenvalues $\beta \in \mathcal{B}_1$ depend on the Rayleigh number.*

Now we are ready to examine the dual eigenvectors. For $(j, k, l) \in I_1$, the dual eigenvectors ψ_{jkl}^{i*} and $\tilde{\psi}_{jkl}^{i*}$ of ψ_{jkl}^i and $\tilde{\psi}_{jkl}^i$ are given by

$$\psi_{jkl}^{i*} = \begin{cases} u_{jkl}^{i*} \sin(j\alpha_1 x_1 + k\alpha_2 x_2) \cos l\pi x_3, \\ v_{jkl}^{i*} \sin(j\alpha_1 x_1 + k\alpha_2 x_2) \cos l\pi x_3, \\ w_{jkl}^{i*} \cos(j\alpha_1 x_1 + k\alpha_2 x_2) \sin l\pi x_3, \\ T_{jkl}^{i*} \cos(j\alpha_1 x_1 + k\alpha_2 x_2) \sin l\pi x_3, \end{cases} \quad (4.4.26)$$

$$\tilde{\psi}_{jkl}^{i*} = \begin{cases} -u_{jkl}^{i*} \cos(j\alpha_1 x_1 + k\alpha_2 x_2) \cos l\pi x_3, \\ -v_{jkl}^{i*} \cos(j\alpha_1 x_1 + k\alpha_2 x_2) \cos l\pi x_3, \\ w_{jkl}^{i*} \sin(j\alpha_1 x_1 + k\alpha_2 x_2) \sin l\pi x_3, \\ T_{jkl}^{i*} \sin(j\alpha_1 x_1 + k\alpha_2 x_2) \sin l\pi x_3, \end{cases} \quad (4.4.27)$$

where

$$\begin{aligned} u_{jkl}^{i*} &= -\frac{l\pi}{\alpha_{jk}^2}(j\alpha_1 - k\alpha_2\omega A_1(\beta_{jkl}^i)), & v_{jkl}^{i*} &= -\frac{l\pi}{\alpha_{jk}^2}(k\alpha_2 + j\alpha_1\omega A_1(\beta_{jkl}^i)), \\ w_{jkl}^{i*} &= 1, & T_{jkl}^{i*} &= \text{Pr } RA_2(\beta_{jkl}^i). \end{aligned}$$

By the spectral theorem, Theorem 3.4 in Ma and Wang (2005b), the dual eigenvectors satisfy

$$(\psi_{jkl}^q, \psi_{jkl}^{p*}) = 0, \quad (\tilde{\psi}_{jkl}^q, \tilde{\psi}_{jkl}^{p*}) = 0 \quad \text{for } q \neq p. \quad (4.4.28)$$

In particular, if $\beta_{jkl}^q = \bar{\beta}_{jkl}^{q+1} = \sigma + i\rho$, then

$$\phi_{jkl}^{q*} = \text{Re}\psi_{jkl}^{i*}(\bar{\beta}_{jkl}^q), \quad \phi_{jkl}^{q+1*} = \text{Im}\psi_{jkl}^{i*}(\bar{\beta}_{jkl}^q) \quad \text{for } i = q \quad (4.4.29)$$

and for any $\psi_J^i \neq \phi_{jkl}^q \phi_{jkl}^{q+1}$,

$$L_R^* \phi_{jkl}^{q*} = \sigma \phi_{jkl}^{q*} + \rho \phi_{jkl}^{q+1*}, \quad L_R^* \phi_{jkl}^{q+1*} = -\rho \phi_{jkl}^{q*} + \sigma \phi_{jkl}^{q+1*}, \quad (4.4.30)$$

$$(\psi_J^i, \phi_{jkl}^{q*}) = (\psi_J^i, \phi_{jkl}^{q+1*}) = 0. \quad (4.4.31)$$

4.4.3 Principle of Exchange of Stabilities

It is known that only the eigenvalues in \mathcal{B}_1 depend on the Rayleigh number R . Hence to study the critical-crossing of the first eigenvalues of (4.4.7), it suffices to focus this problem on the set \mathcal{B}_1 , i.e., on the solutions β of (4.4.21), which can be equivalently written in the following form

$$\begin{aligned} \beta^3 + (2\text{Pr} + 1)\gamma_{jkl}^2\beta^2 + \left[(\text{Pr}^2 + 2\text{Pr})\gamma_{jkl}^4 + \frac{l^2\pi^2\omega}{\gamma_{jlk}^2} - \text{Pr } R \frac{\alpha_{jlk}^2}{\gamma_{jkl}^2} \right] \beta + \\ + \text{Pr}^2\gamma_{jkl}^6 - \text{Pr}^2R\alpha_{jk}^2 + l^2\pi^2\omega^2 = 0. \end{aligned} \quad (4.4.32)$$

CASE 1. $\beta = 0$ is a zero of (4.4.32) if and only if

$$\text{Pr}^2\gamma_{jkl}^6 - \text{Pr}^2R\alpha_{jk}^2 + l^2\pi^2\omega^2 = 0.$$

In this case, we have

$$R = \frac{\gamma_{jkl}^6}{\alpha_{jkl}^2} + \frac{l^2\pi^2\omega^2}{\alpha_{jlk}^2\text{Pr}^2} \geq \frac{(\alpha_{jkl}^2 + \pi^2)^3}{\alpha_{jkl}^2} + \frac{\pi^2\omega^2}{\alpha_{jkl}^2\text{Pr}^2}.$$

Hence the critical Rayleigh number R_{c_1} associated with real eigenvalues is given by

$$R_{c_1} = \min_{(j,k,l) \in I_1} \left(\frac{\gamma_{jkl}^6}{\alpha_{jkl}^2} + \frac{l^2\pi^2\omega^2}{\alpha_{jlk}^2\text{Pr}^2} \right) = \frac{(\alpha_{j_1 k_1}^2 + \pi^2)^3}{\alpha_{j_1 k_1}^2} + \frac{\pi^2\omega^2}{\alpha_{j_1 k_1}^2\text{Pr}^2}, \quad (4.4.33)$$

for some $(j_1, k_1, 1) \in I_1$.

CASE 2. Let a purely imaginary number $\beta = i\rho$ ($\rho \neq 0$) be a zero of (4.4.32). Then

$$\begin{aligned}\rho^2 &= (\Pr^2 + 2\Pr)\gamma_{jkl}^4 + \frac{l^2\pi^2\omega^2}{\gamma_{jkl}^2} - \Pr R \frac{\alpha_{jk}^2}{\gamma_{jkl}^2}, \\ \rho^2 &= \frac{1}{(2\Pr + 1)\gamma_{jkl}^2} \left[\Pr^2\gamma_{jkl}^6 - \Pr^2 R \alpha_{jk}^2 + l^2\pi^2\omega^2 \right].\end{aligned}\quad (4.4.34)$$

We infer from (4.4.34) that the equation (4.4.32) has a pair of purely imaginary solutions $\beta = \pm i\rho$ if and only if the following conditions hold true:

$$\begin{aligned}(\Pr^2 + 2\Pr)\gamma_{jkl}^4 + \frac{l^2\pi^2\omega^2}{\gamma_{jkl}^2} - \Pr R \frac{\alpha_{jk}^2}{\gamma_{jkl}^2} &> 0, \\ (2\Pr + 1)\gamma_{jkl}^2 \left[(\Pr^2 + 2\Pr)\gamma_{jkl}^4 + \frac{l^2\pi^2\omega^2}{\gamma_{jkl}^2} - \Pr R \frac{\alpha_{jk}^2}{\gamma_{jkl}^2} \right] \\ &= \Pr^2\gamma_{jkl}^2 - \Pr^2 R \alpha_{jk}^2 + l^2\pi^2\omega^2.\end{aligned}$$

In this case, we have

$$R = \frac{2(\Pr + 1)\gamma_{jkl}^6}{\alpha_{jk}^2} + \frac{2l^2\pi^2\omega^2}{(\Pr + 1)\alpha_{jk}^2}, \quad (4.4.35)$$

$$R < \frac{(\Pr + 2)\gamma_{jkl}^6}{\alpha_{jk}^2} + \frac{l^2\pi^2\omega^2}{\Pr \alpha_{jk}^2}. \quad (4.4.36)$$

We derive, from (4.4.35) and (4.4.36), a lower bound for the rotating angular velocity ω :

$$\omega^2 > \frac{\Pr^2(1 + \Pr)\gamma_{jkl}^6}{(1 - \Pr)^2\pi^2}, \quad (4.4.37)$$

which could only hold true when the Prandtl number $\Pr < 1$.

As in Case 1, the minimum of the right-hand side of (4.4.35) is always obtained at $l = 1$. Thus the critical Rayleigh number R_{c2} associated with complex eigenvalues is given by

$$\begin{aligned}R_{c2} &= \min_{(j,k,l) \in \Lambda} \left[\frac{2(\Pr + 1)\gamma_{jkl}^6}{\alpha_{jk}^2} + \frac{2l^2\pi^2\omega^2}{(\Pr + 1)\alpha_{jk}^2} \right] \\ &= \frac{2(\Pr + 1)(\alpha_{j_2k_2}^2 + \pi^2)^3}{\alpha_{j_2k_2}^2} + \frac{2\pi^2\omega^2}{(\Pr + 1)\alpha_{j_2k_2}^2},\end{aligned}\quad (4.4.38)$$

for some $(j_2, k_2, 1) \in \Lambda \subset I_1$, where Λ is the set of (j, k, l) satisfying (4.4.37).

Lemma 4.4.2 Let $\alpha_0 = \min\{\alpha_1, \alpha_2\}$. Assume either $\text{Pr} \geq 1$ or

$$\frac{(1 - \text{Pr})\pi^2}{\text{Pr}^2(1 + \text{Pr})}\omega^2 < (\pi^2 + \alpha_0^2)^3 \quad \text{for } \text{Pr} < 1. \quad (4.4.39)$$

Let $(j, k, l) = (j_1, k_1, 1) \in I_1$ minimize the right-hand side of (4.4.33). Then the zero $\beta_{j_1 k_1 1}^1(R)$ of (4.4.32) is a real eigenvalue of (4.4.7) near R_{c_1} , and satisfies that

$$\beta_{j_1 k_1 1}^1(R) \begin{cases} < 0 & \text{if } R < R_{c_1}, \\ = 0 & \text{if } R = R_{c_1}, \\ > 0 & \text{if } R > R_{c_1}, \end{cases} \quad (4.4.40)$$

$$\operatorname{Re}\beta(R_{c_1}) < 0 \quad \forall \beta(R) \in \mathcal{B}_1 \bigcup \mathcal{B}_2 \bigcup \mathcal{B}_3 \text{ with } \beta(R_{c_1}) \neq 0. \quad (4.4.41)$$

Proof. We first prove that there is a $\delta > 0$ such that

$$\operatorname{Re}\beta_{jkl}^q(R) < 0 \quad \forall \beta_{jkl}^q \in \mathcal{B}_1, \quad 0 \leq R < \delta. \quad (4.4.42)$$

In fact, for each $(j, k, l) \in I_1$, as $R = 0$ the solutions of (4.4.21) are

$$\beta_{jkl}^1 = -\gamma_{jkl}^2, \quad \beta_{jkl}^2 = \bar{\beta}_{jkl}^3 = -\text{Pr} \gamma_{jkl}^2 + i \frac{l\pi\omega}{\gamma_{jkl}}.$$

Namely, $\operatorname{Re}\beta_{jkl}^q(0) < 0$, which implies that (4.4.42) holds true because $\beta_{jkl}^q(R)$ are continuous on R .

By the above discussion, we only need to show that the first eigenvalue crosses the imaginary axis. Let

$$\begin{aligned} g_{jkl}(\beta) &= \gamma_{jkl}^2(\beta + \gamma_{jkl}^2) \left[(\beta + \text{Pr} \gamma_{jkl}^2)^2 + l^2 \pi^2 \omega^2 \right], \\ h_{jkl}(\beta) &= \text{Pr} R \alpha_{jk}^2 (\beta + \text{Pr} \gamma_{jkl}^2), \\ f_{jkl}(\beta) &= g_{jkl}(\beta) - h_{jkl}(\beta). \end{aligned}$$

Thus, the equation (4.4.21) is referred to $f_{jkl}(\beta) = 0$.

We note that $f_{j_1 k_1 1}(\beta) = 0$ is equivalent to $g_{j_1 k_1 1}(\beta) = h_{j_1 k_1 1}(\beta)$, i.e.

$$\gamma_{j_1 k_1 1}^2(\beta + \gamma_{j_1 k_1 1}^2) \left[(\beta + \text{Pr} \gamma_{j_1 k_1 1}^2)^2 + l^2 \pi^2 \omega^2 \right] = \text{Pr} R \alpha_{j_1 k_1}^2 (\beta + \text{Pr} \gamma_{j_1 k_1 1}^2). \quad (4.4.43)$$

When $\text{Pr} > 1$, we see that both $g_{j_1 k_1 1}(\beta)$ and $h_{j_1 k_1 1}(\beta)$ are strictly increasing for $\beta > -\gamma_{j_1 k_1 1}^2$. Let Γ_1 be the graph of $\eta = g_{j_1 k_1 1}(\beta)$ and Γ_2 the graph of $\eta = h_{j_1 k_1 1}(\beta)$ as shown in Fig. 4.30. When $R = R_{c_1}$, the point S_0 , the intersecting point of Γ_1 and Γ_2 corresponding to $\beta_{j_1 k_1 1}(R)$ (i.e., the β coordinate of S_0 is $\beta_{j_1 k_1 1}(R)$), is on the η axis. When R increases (resp., decreases), so become S_1 (resp., S_2). This proves (4.4.40).

When $\text{Pr} = 1$, the solution $\beta_{j_1 k_1 1}^1(R)$ of (4.4.43) is given by

$$\beta_{j_1 k_1 1}^1(R) = -\gamma_{j_1 k_1 1}^2 + \gamma_{j_1 k_1 1}^{-1} \sqrt{R \alpha_{j_1 k_1}^2 - \pi^2 \omega^2},$$

which satisfies (4.4.40).

When $\text{Pr} < 1$, by (4.4.37), the condition (4.4.39) implies that no complex eigenvalues $\beta_{jkl}^q(R)$ satisfying that

$$\lim_{R \rightarrow R_{c_1}} \operatorname{Re} \beta_{jkl}^q(R) = 0, \quad \lim_{R \rightarrow R_{c_1}} \operatorname{Im} \beta_{jkl}^q(R) = 0.$$

In addition, we note that R_{c_1} is the first critical Rayleigh number. Hence, when $R = R_{c_1}$, along the β -increasing direction Γ_2 crosses Γ_1 at the intersecting point S_0 from the left-hand side to the right-hand side of Γ_2 , as shown in Fig. 4.31. As proved by (4.4.43) and Fig. 4.30, this implies that (4.4.40) is valid. \square

Remark 4.4.3 In the proof of Lemma 4.4.2, as shown by (4.4.43), Figs. 4.30 and 4.31, we see that, near $R = R_{c_1}$, the first eigenvalue $\beta_{j_1 k_1 1}^1(R)$ is a simple zero of $f_{j_1 k_1 1}(\beta)$. We have seen in (4.4.22) and (4.4.23) that there are two eigenvectors $\psi_{j_1 k_1 1}^1$ and $\tilde{\psi}_{j_1 k_1 1}^1 \in H_1$ corresponding to $\beta_{j_1 k_1 1}^1$. Therefore, the multiplicity of the first eigenvalue of $L_R|_{H_1}$ (resp., $L_R|_{\tilde{H}_1}$) is $m_{H_1} = 2m$ (resp., $m_{\tilde{H}_1} = m$), where m is the number of (j, k, l) 's ($\in I_1$) satisfying $\alpha_{jk}^2 = \alpha_{j_1 k_1 1}^2$.

Fig. 4.30 There are at most three intersections of Γ_1 and Γ_2 , and as $\text{Pr} > 1$, Γ_2 must cross Γ_1 at S_0 from the left-hand side to the right-hand side of Γ_2 along the β -increasing direction

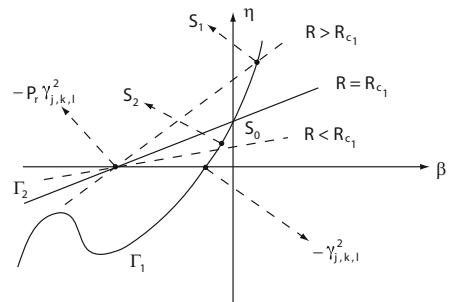
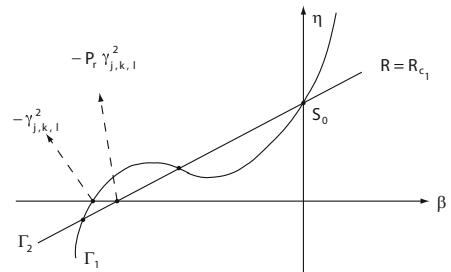


Fig. 4.31 As $\text{Pr} < 1$, the condition (4.4.39) implies that the straight line Γ_2 at $R = R_{c_1}$ must cross Γ_1 at S_0 from the left-hand side to the right-hand side of Γ_1 along the β -increasing direction



Remark 4.4.4 For the classical Bénard problem without rotation, the second term on the right-hand side of (4.4.33) is not presented. Therefore, the first critical Rayleigh number of the Bénard problem dependent only on the aspect of ratio; while the

first critical Rayleigh number of the rotating problem depends on the aspect ratio: $\alpha = L_2/L_1$, the Prandtl number and the rotating angular velocity. And it is clear that the first critical Rayleigh number of rotating flows is larger than that of the Bénard problem. This indicates that the rotating flows are much more stable than the nonrotating flows; see Sect. 4.1.

When the condition (4.4.39) is not valid, (4.4.37) holds true for some indices $(j, k, l) \in I_2$, and there are complex eigenvalues with real part crossing the imaginary axis. In this case, the smaller one of both the critical Rayleigh numbers R_{c_1} and R_{c_2} defined respectively by (4.4.33) and (4.4.38) is the first critical Rayleigh number.

Let $(j_1, k_1, 1)$ and $(j_2, k_2, 1)$ satisfy (4.4.33) and (4.4.38) respectively. It is ready to check that if

$$\frac{(1 - \text{Pr})\pi^2}{\text{Pr}^2(1 + \text{Pr})}\omega^2 > \max\{(\alpha_{j_1 k_1}^2 + \pi^2)^3, (\alpha_{j_2 k_2}^2 + \pi^2)^3\}, \quad (4.4.44)$$

then R_{c_2} is less than R_{c_1} , therefore R_{c_2} is the first critical Rayleigh number. However, if

$$(\alpha_{j_2 k_2}^2 + \pi^2)^3 < \frac{(1 - \text{Pr})\pi^2}{\text{Pr}^2(1 + \text{Pr})}\omega^2 < (\alpha_{j_1 k_1}^2 + \pi^2)^3, \quad (4.4.45)$$

then we need additional methods to identify them.

For any $(j, k, l) \in I_1$, there are at most three zeros of $f_{jkl}(\beta)$ therefore each pair of complex eigenvalues $\beta_{jkl}^q = \bar{\beta}_{jkl}^{q+1}$ are simple zero of $f_{jkl}(\beta)$. The following lemma shows that the complex eigenvalues $\beta_{j_2 k_2 1}^q(R) = \bar{\beta}_{j_2 k_2 1}^{q+1}(R)$ are critical-crossing at the critical Rayleigh number R_{c_2} regardless of if it is the first one.

Lemma 4.4.5 *Let $\text{Pr} < 1$ and $\omega^2 > \text{Pr}^2(1 + \text{Pr})(\pi^2 + \alpha_{j_2 k_2}^2)^3/(1 - \text{Pr})\pi^2$. Then the pair of complex eigenvalues $\beta_{j_2 k_2 1}^1(R) = \bar{\beta}_{j_2 k_2 1}^2(R)$ are critical-crossing at the critical Rayleigh number R_{c_2} , i.e.*

$$\text{Re}\beta_{j_2 k_2 1}^1(R) \begin{cases} < 0 & \text{if } R < R_{c_2}, \\ = 0 & \text{if } R = R_{c_2}, \\ > 0 & \text{if } R > R_{c_2}. \end{cases}$$

Proof. Since $\beta_{j_2 k_2 1}^1(R), \beta_{j_2 k_2 1}^2(R)$ and $\beta_{j_2 k_2 1}^3(R)$ (real) are zeros of (4.4.32), we obtain that

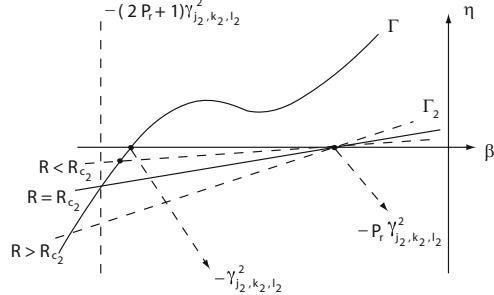
$$\beta_{j_2 k_2 1}^3(R) = -2\text{Re}\beta_{j_2 k_2 1}^1(R) - (2\text{Pr} + 1)(\alpha_{j_2 k_2}^2 + \pi^2).$$

Hence, the critical-crossing at R_{c_2} is equivalent to

$$\beta_{j_2 k_2 1}^3(R) \begin{cases} > -(2\text{Pr} + 1)(\alpha_{j_2 k_2}^2 + \pi^2) & \text{if } R < R_{c_2}, \\ = -(2\text{Pr} + 1)(\alpha_{j_2 k_2}^2 + \pi^2) & \text{if } R = R_{c_2}, \\ < -(2\text{Pr} + 1)(\alpha_{j_2 k_2}^2 + \pi^2) & \text{if } R > R_{c_2}, \end{cases}$$

which is true as shown in Fig. 4.32. This proof is complete. \square

Fig. 4.32 Critical-crossing for the complex pair of eigenvalues, based on the third real eigenvalue



In fact, for the case of (4.4.45), if R_{c_1} is the first critical Rayleigh number, then (4.4.40) and (4.4.41) are also valid. Thus, in view of Lemmas 4.4.2 and 4.4.5, we have the following critical-crossing theorem for the first eigenvalues of (4.4.7).

Theorem 4.4.6 Let $(j_1, k_1, 1)$ and $(j_2, k_2, 1)$ satisfy (4.4.33) and (4.4.38) respectively. Assume that

$$\frac{(1 - Pr)\pi^2}{Pr^2(1 + Pr)}\omega^2 \neq \begin{cases} (\alpha_{j_1 k_1}^2 + \pi^2)^3, \\ (\alpha_{j_2 k_2}^2 + \pi^2)^3 \end{cases} \quad \text{if } R_{c_2} < R_{c_1}. \quad (4.4.46)$$

Then if R_{c_1} is the first critical Rayleigh number, (4.4.40) and (4.4.41) hold true, and if $R_{c_2} < R_{c_1}$ we have

$$Re\beta_{j_2 k_2 l}^1(R) = Re\beta_{j_2 k_2 l}^2(R) \begin{cases} < 0 & \text{if } R < R_{c_2}, \\ = 0 & \text{if } R = R_{c_2}, \\ > 0 & \text{if } R > R_{c_2}. \end{cases} \quad (4.4.47)$$

$$Re\beta(R_{c_2}) < 0 \quad \forall \beta \in \mathcal{B}_1 \bigcup \mathcal{B}_2 \bigcup \mathcal{B}_3 \text{ with } Re\beta(R_{c_2}) \neq 0. \quad (4.4.48)$$

In particular, if $Pr \geq 1$ or (4.4.39) holds true, then R_{c_1} is the first critical Rayleigh number, and if (4.4.44) holds true, then $R_{c_2} < R_{c_1}$.

In the following we consider the multiplicity of the first critical-crossing eigenvalues. It is easy to see that the first eigenvalues of (4.4.7) depend on the periods L_1 and L_2 (i.e., $\alpha_{jk}^2 = j^2\alpha_1^2 + k^2\alpha_2^2$, α_1 and α_2 are as in (4.4.11)). Let

$$L = L_1 = \alpha L_2. \quad (4.4.49)$$

Then, by (4.4.33) and (4.4.38), we have

$$\begin{aligned} R_{c_1} &= \min_{(j, k, 1) \in I_1} \left(\frac{(\alpha_{jk}^2 + \pi^2)^3}{\alpha_{jk}^2} + \frac{\pi^2 \omega^2}{Pr^2 \alpha_{jk}^2} \right) \\ &= \frac{(j_1^2 + k_1^2 \alpha^2 + \pi^2 L^2)^3}{L^4 (j_1^2 + k_1^2 \alpha^2)} + \frac{\pi^2 \omega^2 L^2}{Pr^2 (j_1^2 + k_1^2 \alpha^2)}, \end{aligned} \quad (4.4.50)$$

$$\begin{aligned} R_{c_2} &= \min_{(j,k,l) \in \Lambda} 2(\Pr + 1) \left[\frac{(\alpha_{jk}^2 + \pi^2)^3}{\alpha_{jk}^2} + \frac{\pi^2 \omega^2}{(\Pr + 1)^2 \alpha_{jk}^2} \right] \\ &= 2(\Pr + 1) \left[\frac{(j_2^2 + k_2^2 \alpha^2 + \pi^2 L^2)^3}{L^4 (j_2^2 + k_2^2 \alpha^2)} + \frac{\pi^2 \omega^2 L^2}{(\Pr + 1)^2 (j_2^2 + k_2^2 \alpha^2)} \right], \end{aligned} \quad (4.4.51)$$

where $(j, k, l) \in \Lambda$ satisfies (4.4.37).

It is known that the multiplicity of the first eigenvalues $\beta_{j_1 k_1 l}^1(R)$ of $L_R|_{H_1}$ (resp., $L_R|_{\tilde{H}_1}$) is $m_{H_1} = 2n$ (resp., $m_{\tilde{H}_1} = n$), where n is the number of $(j, k, l) \in I_1$ satisfying

$$j^2 + k^2 \alpha^2 = j_1^2 + k_1^2 \alpha^2,$$

which yields

$$\alpha^2 = \frac{j^2 - j_1^2}{k_1^2 - k^2}, \quad (j, k) \neq (0, 0). \quad (4.4.52)$$

Likewise, the multiplicity of the first complex eigenvalues $\beta_{j_2 k_2 l}^1 = \bar{\beta}_{j_2 k_2 l}$ is $m_{H_1} = 4n$, and $m_{\tilde{H}_1} = 2n$.

By (4.4.52) we see that if α^2 is an irrational number, then $n = 1$. Namely, we have the following lemma.

Lemma 4.4.7 *Let α , defined by (4.4.49), be an irrational number, then the first real eigenvalue $\beta_{j_1 k_1 l}^1(R)$ at R_{c_1} has the multiplicity $m = 2$ in H_1 and $m = 1$ in \tilde{H}_1 , and the first complex eigenvalues $\beta_{j_2 k_2 l}^1(R) = \beta_{j_2 k_2 l}^2(R)$ at R_{c_2} has multiplicity $m = 4$ in H_1 and $m = 2$ in \tilde{H}_1 .*

Now, we look for conditions to make R_{c_2} the first critical Rayleigh number. For $x > 0, b > 0$, we define

$$f_b(x) = \frac{(x + \pi^2)^3 + b}{x}.$$

Let $x = \alpha_{jk}^2$, then the right-hand side of (4.4.33) can be expressed as $f_{b_1}(x)$ with $b_1 = \pi^2 \omega^2 / \Pr^2$, and the second line of (4.4.38) could be expressed as $2(\Pr + 1)f_{b_2}(x)$ with $b_2 = \pi^2 \omega^2 / (\Pr + 1)^2$. Consider

$$f'_b(x) = \frac{(2x - \pi^2)(x + \pi^2)^2 - b}{x^2}. \quad (4.4.53)$$

Let x_i be the zero of $f'_{b_i}(x)$ ($i = 1, 2$). Then we infer from (4.4.53) that

$$x_1 \sim \left(\frac{1}{\Pr} \right)^{2/3} \rightarrow \infty, \quad x_2 < \frac{\pi^2}{2}(1 + \omega^2) \quad \text{for } \Pr \rightarrow 0,$$

which implies that

$$R_{c_1} \sim f_{b_1}(x_1) \rightarrow +\infty, \quad R_{c_2} \sim f_{b_2}(x_2) < +\infty \quad \text{if } \text{Pr} \rightarrow 0.$$

Hence the following lemma holds true.

Lemma 4.4.8 *Given $L, \alpha, \omega > 0$, there exists a number $\delta = \delta(L, \alpha, \omega) \in (0, 1)$ such that if $0 < \text{Pr} < \delta$, R_{c_2} is the first critical Rayleigh number, i.e., $R_{c_2} < R_{c_1}$.*

4.4.4 Transition from First Real Eigenvalues

In this section, we consider the transition of (4.4.1)–(4.4.3) from the first real eigenvalues of (4.4.7) with multiplicities $m = 2$ in H_1 and $m = 1$ in \tilde{H}_1 , which are called (generalized) simple eigenvalues. By Lemma 4.4.7, for almost all $(L_1, L_2) = (L, \alpha L) \in \mathbb{R}_+^2$, the first eigenvalues of (4.4.7) are simple. Here we always assume that the number R_{c_1} defined by (4.4.33) is the first critical Rayleigh number, and the first eigenvalues of (4.4.7) are simple near R_{c_1} .

Theorem 4.4.9 *If the Prandtl number Pr and the rotating angular velocity ω satisfy*

$$b_1 = \text{Pr}^4(\alpha_{j_1 k_1}^2 + \pi^2)^3 \alpha_{j_1 k_1}^2 - \pi^2(\pi^2 - \text{Pr}^2 \alpha_{j_1 k_1}^2) \omega^2 > 0, \quad (4.4.54)$$

where (j_1, k_1) satisfies (4.4.50), then the transition of Problems (4.4.1)–(4.4.3) at $R = R_{c_1}$ is Type-I (continuous), and the following assertions hold true:

- (1) For the constraint (4.4.4), this problem bifurcates from $((u, T), R) = (0, R_{c_1})$ to an attractor $\Sigma_R = S^1 \subset H_1$, consisting of only steady-state solutions.
- (2) For the constraint (4.4.5), the problem bifurcates to two steady-state solutions $\Sigma_R = \{(u_1^R, T_1^R), (u_2^R, T_2^R)\} \subset \tilde{H}_1$, which are attractors.
- (3) The steady-state solutions $(u, T) \in \Sigma_R = S^1 \subset H_1$ can be expressed as

$$(u, T) = a(\beta_{j_1 k_1 1}^1(R))^{1/2} (x \psi_{j_1 k_1 1}^1 + y \tilde{\psi}_{j_1 k_1 1}^1) + o(|\beta_{j_1 k_1 1}^1|^{1/2}),$$

where $a > 0$ is a constant, $\psi_{j_1 k_1 1}^1$ and $\tilde{\psi}_{j_1 k_1 1}^1$ are the first eigenvectors given by (4.4.22) and (4.4.23), and $x^2 + y^2 = 1$.

- (4) The two bifurcated steady-state solutions (u_i^R, T_i^R) ($i = 1, 2$) in \tilde{H}_1 can be written as

$$(u_i^R, T_i^R) = (-1)^{i+1} a(\beta_{j_1 k_1 1}^1(R))^{1/2} \psi_{j_1 k_1 1}^1 + o(|\beta_{j_1 k_1 1}^1|^{1/2}).$$

Theorem 4.4.10 *If Pr and ω satisfy*

$$b_1 = \text{Pr}^4(\alpha_{j_1 k_1}^2 + \pi^2)^3 \alpha_{j_1 k_1}^2 - \pi^2(\pi^2 - \text{Pr}^2 \alpha_{j_1 k_1}^2) \omega^2 < 0, \quad (4.4.55)$$

then the transition of Problem (4.4.1)–(4.4.3) at $R = R_{c_1}$ is a Type-II (jump) transition. Moreover we have the following assertions:

- (1) Under the constraint (4.4.4) there is a singular separation of a cycle of steady-state solutions of (Σ^*, R^*) with $\Sigma^* = S^1$, $R^* < R_{c_1}$, i.e., there are two branches of cycles Σ_R^1 and Σ_R^2 of steady-state solutions bifurcated from Σ^* on $R^* < R$ such that

$$\lim_{R \rightarrow R_{c_1}} \Sigma_R^1 = \{0\}, \quad \Sigma_{R_{c_1}}^2 \neq \{0\} \quad (= S^1),$$

and $\Sigma_R^2 = S^1$ is an attractor for all $R^* < R < R_{c_1} + \delta$ for some $\delta > 1$.

- (2) Under the constraint (4.4.5) there is a saddle-node bifurcation of steady-state solutions from (ψ_0^*, R^*) and (ψ_1^*, R^*) with $\psi_0^*, \psi_1^* \neq 0$, $R^* < R_{c_1}$, i.e., there are four branches of steady-state solutions ψ_R^i ($1 \leq i \leq 4$) with ψ_R^1 and ψ_R^2 bifurcated from ψ_0^* , ψ_R^3 and ψ_R^4 from ψ_1^* on $R^* < R$, such that

$$\lim_{R \rightarrow R_{c_1}} \psi_R^1 = 0, \quad \lim_{R \rightarrow R_{c_1}} \psi_R^3 = 0, \quad \psi_{R_{c_1}}^2, \psi_{R_{c_1}}^4 \neq 0,$$

and ψ_R^2 and ψ_R^4 are attractors for $R^* < R < R_{c_1} + \delta$ with some $\delta > 0$.

Remark 4.4.11 Theorems 4.4.9 and 4.4.10 show that the transition of Problems (4.4.1)–(4.4.3) is possibly jumping (type-II) only under the condition $\pi^2 > \text{Pr}^2 \alpha_{j_1 k_1}^2$, which provides an upper bound for the Prandtl number Pr leading to the jump transition at R_{c_1} .

Proof of Theorems 4.4.9 and 4.4.10. We first reduce the equations of (4.4.1)–(4.4.3) with (4.4.5) to the center manifold in \tilde{H} .

Let $J_1 = (j_1, k_1, 1)$, $\psi_{J_1} = \psi_{j_1 k_1 1}^1$. The reduced equations of (4.4.1)–(4.4.3) with (4.4.5) in \tilde{H} read

$$\frac{dx}{dt} = \beta_{J_1}^1(R)x + \frac{1}{(\psi_{J_1}, \psi_{J_1}^*)}(G(\psi, \psi), \psi_{J_1}^*), \quad (4.4.56)$$

where $\psi \in \tilde{H}_1$ is written as

$$\psi = x\psi_{J_1} + \Phi, \quad (4.4.57)$$

Φ is the center manifold function, and $G(\psi_1, \psi_2) = -P((U_1 \cdot \nabla)U_2, (U_1 \cdot \nabla)T_2)$ for $\psi_i = (U_i, T_i)$ ($i = 1, 2$).

Hereafter, we make the following convention

$$o(m) = o(|x|^m) + O(|\beta_J^1| \cdot |x|^m) \quad \text{for } x \in \mathbb{R}^n \text{ near 0.}$$

By $\Phi(x) = O(2)$, inserting (4.4.57) into (4.4.56) we find

$$\frac{dx}{dt} = \beta_{J_1}^1(R)x - \frac{x}{(\psi_{J_1}, \psi_{J_1}^*)}[(G(\psi_{J_1}, \psi_{J_1}^*), \Phi) + (G(\Phi, \psi_{J_1}^*), \psi_{J_1})] + o(3). \quad (4.4.58)$$

Let the center manifold function be in the form

$$\Phi = \sum_{(J,i) \neq (J_1,1)} y_J^i \psi_J^i. \quad (4.4.59)$$

Then, by the approximation formula for center manifold functions in Theorem A.1.1, we obtain

$$L_R \Phi = \sum_{(J,i) \neq (J_1,1)} y_J^i \beta_J^i \psi_J^i = -x^2 G(\psi_{J_1}, \psi_{J_1}) + o(2).$$

It follows that

$$y_J^i = \frac{x^2}{-\beta_J^i(\psi_J^i, \psi_J^{i*})} (G(\psi_{J_1}, \psi_{J_1}), \psi_J^{i*}) + o(2). \quad (4.4.60)$$

Putting (4.4.59) and (4.4.60) into (4.4.58) we deduce that

$$\frac{dx}{dt} = \beta_{J_1}^1(R)x - \frac{\delta x^3}{(\psi_{J_1}, \psi_{J_1}^*)} + o(3), \quad (4.4.61)$$

where

$$\begin{aligned} \delta = \sum_{(J,i) \neq (J_1,1)} & \left[\frac{(G(\psi_{J_1}, \psi_{J_1}), \psi_J^{i*})(G(\psi_{J_1}, \psi_{J_1}^*), \psi_J^i)}{-\beta_J^i(\psi_J^i, \psi_J^{i*})} \right. \\ & \left. + \frac{(G(\psi_{J_1}, \psi_{J_1}), \psi_J^{i*})(G(\psi_J^i, \psi_{J_1}^*), \psi_{J_1})}{-\beta_J^i(\psi_J^i, \psi_J^{i*})} \right]. \end{aligned}$$

By (4.4.22) and (4.4.26),

$$\begin{aligned} \psi_{J_1} &= \begin{cases} u \sin(j_1 \alpha_1 x_1 + k_1 \alpha_2 x_2) \cos \pi x_3, \\ v \sin(j_1 \alpha_1 x_1 + k_1 \alpha_2 x_2) \cos \pi x_3, \\ \cos(j_1 \alpha_1 x_1 + k_1 \alpha_2 x_2) \sin \pi x_3, \\ T \cos(j_1 \alpha_1 x_1 + k_1 \alpha_2 x_2) \sin \pi x_3, \end{cases} \\ \psi_{J_1}^* &= \begin{cases} u^* \sin(j_1 \alpha_1 x_1 + k_1 \alpha_2 x_2) \cos \pi x_3, \\ v^* \sin(j_1 \alpha_1 x_1 + k_1 \alpha_2 x_2) \cos \pi x_3, \\ \cos(j_1 \alpha_1 x_1 + k_1 \alpha_2 x_2) \sin \pi x_3, \\ T^* \cos(j_1 \alpha_1 x_1 + k_1 \alpha_2 x_2) \sin \pi x_3, \end{cases} \end{aligned}$$

where

$$\begin{aligned} u &= -\frac{\pi}{\alpha_{j_1 k_1}^2} (j_1 \alpha_1 + k_1 \alpha_2 \omega A_1), & v &= -\frac{\pi}{\alpha_{j_1 k_1}^2} (k_1 \alpha_2 - j_1 \alpha_1 \omega A_1), & T &= A_2, \\ u^* &= -\frac{\pi}{\alpha_{j_1 k_1}^2} (j_1 \alpha_1 - k_1 \alpha_2 \omega A_1), & v^* &= -\frac{\pi}{\alpha_{j_1 k_1}^2} (k_1 \alpha_2 + j_1 \alpha_1 \omega A_1), & T^* &= \text{Pr } RA, \\ A_1 &= A_1(\beta_{J_1}^1), & A_2 &= A_2(\beta_{J_1}^1). \end{aligned}$$

Direct calculation shows that

$$G(\psi_{J_1}, \psi_{J_1}) = \frac{\pi}{2} P(u \sin 2\phi, v \sin 2\phi, 0, 0) - \frac{\pi}{2} (0, 0, 0, T \sin 2\pi x_3), \quad (4.4.62)$$

$$G(\psi_{J_1}, \psi_{J_1}^*) = \frac{\pi}{2} P(u^* \sin 2\phi, v^* \sin 2\phi, 0, 0) - \frac{\pi}{2} (0, 0, 0, T^* \sin 2\pi x_3), \quad (4.4.63)$$

where $\phi = j_1 \alpha_1 x_1 + k_1 \alpha_2 x_2$, and $P : L^2(\Omega)^4 \rightarrow \tilde{H}$ is the Leray projection. Also, we know that

$$G(\psi_{002}^1, \psi_{J_1}^*) = 0, \quad G(\psi_{2(j_1, k_1, 0)}, \psi_{J_1}^*) = 0. \quad (4.4.64)$$

It follows from (4.4.62) to (4.4.64) that

$$\begin{aligned} (G(\psi_{J_1}, \psi_{J_1}), \psi_J^{i*}) &= 0 & \forall \psi_J^{i*} \neq \psi_{2(j_1, k_1, 0)}^*, \psi_{002}^{1*}, \\ (G(\psi_{J_1}, \psi_{J_1}^*), \psi_J^i) &= 0 & \forall \psi_J^i \neq \psi_{2(j_1, k_1, 0)}, \psi_{002}^1, \\ (G(\psi_J^i, \psi_{J_1}^*), \psi_{J_1}) &= 0 & \text{for } \psi_J^i = \psi_{002}^1 \text{ or } \psi_{2(j_1, k_1, 0)}. \end{aligned}$$

Thus, we derive

$$\begin{aligned} \delta &= \frac{(G(\psi_{J_1}, \psi_{J_1}), \psi_{2(j_1, k_1, 0)}^*)(G(\psi_{J_1}, \psi_{J_1}^*), \psi_{2(j_1, k_1, 0)})}{-\beta_{2(j_1, k_1, 0)}(\psi_{2(j_1, k_1, 0)}, \psi_{2(j_1, k_1, 0)}^*)} \\ &\quad + \frac{(G(\psi_{J_1}, \psi_{J_1}), \psi_{002}^{1*})(G(\psi_{J_1}, \psi_{J_1}^*), \psi_{002}^1)}{-\beta_{002}^1(\psi_{002}^1, \psi_{002}^{1*})}. \end{aligned}$$

Note that $-\beta_{002}^1 = 4\pi^2$, $-\beta_{2(j_1, k_1, 0)} = 4\Pr \alpha_{j_1 k_1}^2$, and

$$\begin{aligned} \psi_{002}^1 &= \psi_{002}^{1*} = (0, 0, 0, \sin 2\pi x_3), \dots, \\ \psi_{2(j_1, k_1, 0)} &= \psi_{2(j_1, k_1, 0)}^* = (k_1 \alpha_2 \sin 2\phi, -j_1 \alpha_1 \sin 2\phi, 0, 0). \end{aligned}$$

By the calculation we get

$$\begin{aligned} \delta &= \frac{L_1 L_2 \pi^2}{8} \left[\Pr R A_2^2(\beta_{J_1}^1) - \frac{\pi^4 \omega^2 A_1^2(\beta_{J_1}^1)}{\Pr a_{j_1 k_1}^4} \right] \\ &= \frac{L_1 L_2 \pi^2}{8} \left[\frac{\Pr R}{(\alpha_{j_1 k_1}^2 + \pi^2 + \beta_{J_1}^1)^2} - \frac{\pi^4 \omega^2}{\Pr \alpha_{j_1 k_1}^4 (\Pr \alpha_{j_1 k_1}^2 + \Pr \pi^2 + \beta_{J_1}^1)^2} \right]. \quad (4.4.65) \end{aligned}$$

We know that

$$\beta_{J_1}^1(R_{c_1}) = 0, \quad R_{c_1} = \frac{(\alpha_{j_1 k_1}^2 + \pi^2)^3}{\alpha_{j_1 k_1}^2} + \frac{\pi^2 \omega^2}{\Pr^2 \alpha_{j_1 k_1}^2}.$$

Hence, by (4.4.65) at $R = R_{c_1}$ we have

$$\delta(R_{c_1}) = \frac{L_1 L_2 \pi^2}{8(\alpha_{j_1 k_1}^2 + \pi^2)^2 \alpha_{j_1 k_1}^2} \left[\Pr (\alpha_{j_1 k_1}^2 + \pi^2)^3 + \frac{\pi^2 \omega^2}{\Pr} - \frac{\pi^4 \omega^2}{\Pr^3 \alpha_{j_1 k_1}^2} \right]. \quad (4.4.66)$$

Thus, it follows from (4.4.66) that

$$\begin{aligned}\delta(R_{c_1}) > 0 &\iff P_r^4(\alpha_{j_1 k_1}^2 + \pi^2)^3 \alpha_{j_1 k_1}^2 > \pi^2(\pi^2 - \text{Pr}^2 \alpha_{j_1 k_1}^2) \omega^2, \\ \delta(R_{c_1}) < 0 &\iff P_r^4(\alpha_{j_1 k_1}^2 + \pi^2)^3 \alpha_{j_1 k_1}^2 < \pi^2(\pi^2 - \text{Pr}^2 \alpha_{j_1 k_1}^2) \omega^2.\end{aligned}\quad (4.4.67)$$

In addition, at $R = R_{c_1}$ we see that

$$\begin{aligned}(\psi_{J_1}, \psi_{J_1}^*) &= \pi^4 \left[\frac{1}{\alpha_{j_1 k_1}^2} - \frac{\omega^2 A_1}{\alpha_{j_1 k_1}^2} + 1 + \text{Pr} R_{c_1} A_2^2 \right] \\ &= \frac{\pi^2}{\alpha_{j_1 k_1}^2 (\alpha_{j_1 k_1}^2 + \pi^2)^2} \left[(\alpha_{j_1 k_1}^2 + \pi^2)^2 (\alpha_{j_1 k_1}^2 + 1) \right. \\ &\quad \left. + \frac{\pi^2 \omega^2}{\text{Pr}} + \text{Pr} (\alpha_{j_1 k_1}^2 + \pi^2)^3 - \frac{\omega^2}{\text{Pr}^2} \right].\end{aligned}$$

It is clear that for $\text{Pr} > \frac{1}{\pi^2}$,

$$(\psi_{J_1}, \psi_{J_1}^*) > 0. \quad (4.4.68)$$

In addition, by the spectral theorem, Theorem 3.4 in Ma and Wang (2005b), $(\psi_{J_1}, \psi_{J_1}^*) \neq 0$, which is continuous on Pr , ω , and $\alpha_{j_1 k_1}$. Therefore, (4.4.68) holds true provided $\beta_{J_1}^1$ being real and simple near $R = R_{c_1}$.

Hence, by Theorems 2.3.1 and 2.5.3, from (4.4.61), (4.4.67), and (4.4.68) we derive the conclusions of Theorems 4.4.9 and 4.4.10 in \tilde{H} for the condition (4.4.5).

Under the condition (4.4.4), the reduced equations of (4.4.1)–(4.4.3) on the center manifold in H are given by

$$\begin{aligned}\frac{dx}{dt} &= \beta_{J_1}^1(R)x + \frac{1}{(\psi_{J_1}, \psi_{J_1}^*)}(G(\psi, \psi), \psi_{J_1}^*), \\ \frac{dy}{dt} &= \beta_{J_1}^1(R)y + \frac{1}{(\tilde{\psi}_{J_1}, \tilde{\psi}_{J_1}^*)}(G(\psi, \psi), \tilde{\psi}_{J_1}^*),\end{aligned}\quad (4.4.69)$$

where $\psi = x\psi_{J_1} + y\tilde{\psi}_{J_1} + \Phi \in H$. It is easy to verify that

$$\begin{aligned}(G(x\psi_{J_1} + y\tilde{\psi}_{J_1}, x\psi_{J_1} + y\tilde{\psi}_{J_1}), \psi_{J_1}^*) &= 0, \\ (G(x\psi_{J_1} + y\tilde{\psi}_{J_1}, x\psi_{J_1} + y\tilde{\psi}_{J_1}), \tilde{\psi}_{J_1}^*) &= 0.\end{aligned}$$

Hence, (4.4.69) can be rewritten as

$$\begin{aligned}\frac{dx}{dt} &= \beta_{J_1}^1 x - \frac{x}{(\psi_{J_1}, \psi_{J_1}^*)} \left[(G(\psi_{J_1}, \psi_{J_1}^*), \Phi) + (G(\Phi, \psi_{J_1}^*), \psi_{J_1}) \right], \\ &\quad - \frac{y}{(\psi_{J_1}, \psi_{J_1}^*)} \left[(G(\tilde{\psi}_{J_1}, \psi_{J_1}^*), \Phi) + (G(\Phi, \psi_{J_1}^*), \tilde{\psi}_{J_1}) \right] + o(3), \\ \frac{dy}{dt} &= \beta_{J_1}^1 y - \frac{x}{(\tilde{\psi}_{J_1}, \tilde{\psi}_{J_1}^*)} \left[(G(\psi_{J_1}, \tilde{\psi}_{J_1}^*), \Phi) + (G(\Phi, \tilde{\psi}_{J_1}^*), \psi_{J_1}) \right] \\ &\quad - \frac{y}{(\tilde{\psi}_{J_1}, \tilde{\psi}_{J_1}^*)} \left[(G(\tilde{\psi}_{J_1}, \tilde{\psi}_{J_1}^*), \Phi) + (G(\Phi, \tilde{\psi}_{J_1}^*), \tilde{\psi}_{J_1}) \right] + o(3).\end{aligned}\quad (4.4.70)$$

Based on the approximation formula for center manifold functions in Theorem A.1.1, we have

$$L_R \Phi = -x^2 G(\psi_{J_1}, \psi_{J_1}) - xyG(\psi_{J_1}, \tilde{\psi}_{J_1}) - xyG(\tilde{\psi}_{J_1}, \psi_{J_1}) - y^2 G(\tilde{\psi}_{J_1}, \tilde{\psi}_{J_1}) + o(2).$$

Let $\Phi = \sum \varphi_J(x, y) \psi_J^i$. Then, by $L_R \Phi = \sum \varphi_J \beta_J^i \psi_J^i$, we get

$$\begin{aligned} \varphi_J &= \frac{1}{-\beta_J^i(\psi_J^i, \psi_J^{i*})} \left[x^2(G(\psi_{J_1}, \psi_{J_1}), \psi_J^{i*}) + y^2(G(\tilde{\psi}_{J_1}, \tilde{\psi}_{J_1}), \psi_J^{i*}) \right. \\ &\quad \left. + xy(G(\psi_{J_1}, \tilde{\psi}_{J_1}) + G(\tilde{\psi}_{J_1}, \psi_{J_1}), \psi_J^{i*}) \right] + o(2). \end{aligned}$$

By a direct calculation, we obtain

$$\begin{aligned} \Phi &= \varphi_1(x, y) \psi_{2(j_1, k_1, 0)} + \varphi_2(x, y) \tilde{\psi}_{2(j_1, k_1, 0)} + \varphi_3(x, y) \psi_{002}^1 + o(2), \quad (4.4.71) \\ \varphi_1(x, y) &= (x^2 - y^2) \frac{(G(\psi_{J_1}, \psi_{J_1}), \psi_{2(j_1, k_1, 0)}^*)}{-\beta_{2(j_1, k_1, 0)}(\psi_{2(j_1, k_1, 0)}, \psi_{2(j_1, k_1, 0)}^*)}, \\ \varphi_2(x, y) &= 2xy \frac{(G(\psi_{J_1}, \tilde{\psi}_{J_1}), \tilde{\psi}_{2(j_1, k_1, 0)}^*)}{-\beta_{2(j_1, k_1, 0)}(\tilde{\psi}_{2(j_1, k_1, 0)}, \tilde{\psi}_{2(j_1, k_1, 0)}^*)}, \\ \varphi_3(x, y) &= (x^2 + y^2) \frac{(G(\psi_{J_1}, \psi_{J_1}), \psi_{002}^{1*})}{-\beta_{002}^1(\psi_{002}^1, \psi_{002}^{1*})}. \end{aligned}$$

Likewise, we obtain that

$$\begin{aligned} (G(\Phi, \psi_{J_1}^*), \psi_{J_1}) &= (G(\Phi, \psi_{J_1}^*), \tilde{\psi}_{J_1}) = o(2), \\ (G(\Phi, \tilde{\psi}_{J_1}^*), \psi_{J_1}) &= (G(\Phi, \tilde{\psi}_{J_1}^*), \tilde{\psi}_{J_1}) = o(2). \end{aligned}$$

Thus, from (4.4.70) and (4.4.71) we derive

$$\begin{aligned} \frac{dx}{dt} &= \beta_{J_1}^1(R)x - \frac{\delta x(x^2 + y^2)}{(\psi_{J_1}, \psi_{J_1}^*)} + o(3), \\ \frac{dy}{dt} &= \beta_{J_1}^1(R)y - \frac{\delta y(x^2 + y^2)}{(\psi_{J_1}, \psi_{J_1}^*)} + o(3), \end{aligned} \quad (4.4.72)$$

where δ is as in (4.4.65). Therefore, the conclusions of Theorems 4.4.9 and 4.4.10 in H for the condition (4.4.4) follow from (4.4.72), (4.4.67), and the S^1 attractor bifurcation theorem, Theorem 2.2.4. The proof of the theorems is complete. \square

4.4.5 Transition from First Complex Eigenvalues

We now consider the transition of (4.4.1)–(4.4.3) with the condition (4.4.5) from the first complex eigenvalues of (4.4.7). In this section, we assume that the number R_{c_2} defined by (4.4.38) is the first critical Rayleigh number, and the first eigenvalues of (4.4.7) near R_{c_2} are complex simple, i.e., the multiplicity $m = 2$ in \tilde{H} . By Lemma 4.4.7, for almost all $(L_1, L_2) \in \mathbb{R}_+^2$, the first eigenvalues of (4.4.7) are simple in \tilde{H} .

To determine the transition type, we need to introduce the following parameter

$$\begin{aligned} b_2 = & \pi^3 \omega^2 (R_1 B_1 - I_1 B_2) C_1 + \pi^3 \omega^2 (I_1 B_1 + R_1 B_2) C_2 \\ & + \text{Pr} R (R_2 B_1 - I_2 B_2) C_3 + \text{Pr} R (I_2 B_1 + R_2 B_2) C_4, \end{aligned} \quad (4.4.73)$$

where

$$\begin{aligned} R_1 &= \text{Re } A_1(\beta_{J_2}^1(R_{c_2})) = \frac{\text{Pr}(\alpha^2 + \pi^2)}{\text{Pr}^2(\alpha^2 + \pi^2)^2 + \rho^2}, \\ R_2 &= \text{Re } A_2(\beta_{J_2}^1(R_{c_2})) = \frac{\alpha^2 + \pi^2}{(\alpha^2 + \pi^2)^2 + \rho^2}, \\ I_1 &= \text{Im } A_1(\beta_{J_2}^1(R_{c_2})) = \frac{-\rho}{\text{Pr}^2(\alpha^2 + \pi^2)^2 + \rho^2}, \\ I_2 &= \text{Im } A_2(\beta_{J_2}^1(R_{c_2})) = \frac{-\rho}{(\alpha^2 + \pi^2)^2 + \rho^2}, \\ B_1 &= \pi^2 \omega^2 I_1^2 - \text{Pr} R \alpha^2 I_2^2, \\ B_2 &= \pi^2 \omega^2 R_1 I_1 - \text{Pr} R \alpha^2 R_2 I_2, \\ C_1 &= \frac{3R_1}{2\text{Pr} \alpha^4} - \frac{\rho^2 R_1}{2\text{Pr} \alpha^4 (4\text{Pr}^2 \alpha^4 + \rho^2)} + \frac{\rho I_1}{\alpha^2 (4\text{Pr}^2 \alpha^4 + \rho^2)}, \\ C_2 &= -\frac{I_1}{2\text{Pr} \alpha^4} + \frac{\rho^2 I_1}{2\text{Pr} \alpha^4 (4\text{Pr}^2 \alpha^4 + \rho^2)} + \frac{\rho R_1}{\alpha^2 (4\text{Pr}^2 \alpha^4 + \rho^2)}, \\ C_3 &= -\frac{3R_2}{2\pi} + \frac{\rho^2 R_2}{2\pi (4\pi^4 + \rho^2)} - \frac{\rho \pi I_2}{4\pi^4 + \rho^2}, \\ C_4 &= \frac{I_2}{2\pi} - \frac{\rho^2 I_2}{2\pi (4\pi^4 + \rho^2)} - \frac{\rho \pi R_2}{4\pi^4 + \rho^2}, \\ \rho^2 &= \frac{1}{\alpha^2 + \pi^2} \left[\frac{1 - \text{Pr}}{1 + \text{Pr}} \pi^2 \omega^2 - \text{Pr}^2 (\alpha^2 + \pi^2)^3 \right], \\ R &= R_{c_2} = \frac{2(\text{Pr} + 1)(\alpha^2 + \pi^2)^3}{\alpha^2} + \frac{2\pi^2 \omega^2}{(\text{Pr} + 1)\alpha^2}, \end{aligned}$$

and $\alpha^2 = \alpha_{j_2 k_2}^2$ with (j_2, k_2) satisfying (4.4.38).

Then, we have the following transition theorem, which provides a criterion for the super-critical and sub-critical Hopf bifurcations by the sign of the number b_2 in (4.4.73).

Theorem 4.4.12 *Let b the number given by (4.4.73). Then, for the problems (4.4.1)–(4.4.3) with (4.4.5) we have the following assertions:*

- (1) *If $b_2 < 0$, then the problem has a continuous (Type-I) transition at $R = R_{c_2}$, and bifurcates to only one periodic solution on $R > R_{c_2}$ which is an attractor.*
- (2) *If $b_2 > 0$, then this problem has a jump (Type-II) transition at $R = R_{c_2}$, and bifurcates to a periodic solution on $R < R_{c_2}$, which is a repeller. Moreover, there exists a singular separation of periodic orbits at some R^* with $0 < R^* < R_{c_2}$.*

Proof. Let $J = (j_2, k_2, 1)$, $\alpha^2 = \alpha_{j_2 k_2}^2$, and $\beta_J = \beta_{J_2 k_2 l} = \sigma + i\rho$ be the first eigenvalue of (4.4.7) near $R = R_{c_2}$ in \tilde{H}_1 . By (4.4.24), the first eigenvectors $\psi_J = \psi_J^1 + i\psi_J^2$ are given by

$$\begin{aligned}\psi_J^1 &= (u_1 \sin \phi \cos \pi x_3, v_1 \sin \phi \cos \pi x_3, \cos \phi \sin \pi x_3, T_1 \cos \phi \sin \pi x_3), \\ \psi_J^2 &= (u_2 \sin \phi \cos \pi x_3, v_2 \sin \phi \cos \pi x_3, 0, T_2 \cos \phi \sin \pi x_3),\end{aligned}$$

where $\phi = j_2 \alpha_1 x_1 + k_2 \alpha_2 x_2$, and

$$\begin{aligned}(u_1, v_1, T_1) &= \left(-\frac{\pi j_2 \alpha_1}{\alpha^2} - \frac{\pi k_2 \alpha_2}{\alpha^2} \omega \operatorname{Re} A_1(\beta_J), -\frac{\pi k_2 \alpha_2}{\alpha^2} + \frac{\pi j_2 \alpha_1}{\alpha^2} \omega \operatorname{Re} A_1(\beta_J), R_e A_2(\beta_J) \right), \\ (u_2, v_2, T_2) &= \left(-\frac{\pi k_2 \alpha_2}{\alpha^2} \omega \operatorname{Im} A_1(\beta_J), \frac{\pi j_2 \alpha_1}{\alpha^2} \omega \operatorname{Im} A_1(\beta_J), \operatorname{Im} A_2(\beta_J) \right).\end{aligned}$$

By (4.4.29), the first conjugate eigenvectors $\psi_J^* = \psi_J^{1*} + i\psi_J^{2*}$ are given by

$$\begin{aligned}\psi_J^{1*} &= (u_1^* \sin \phi \cos \pi x_3, v_1^* \sin \phi \cos \pi x_3, \cos \phi \sin \pi x_3, T_1^* \cos \phi \sin \pi x_3), \\ \psi_J^{2*} &= (u_2^* \sin \phi \cos \pi x_3, v_2^* \sin \phi \cos \pi x_3, 0, T_2^* \cos \phi \sin \pi x_3),\end{aligned}$$

where

$$\begin{aligned}u_1^* &= -\frac{\pi j_2 \alpha_1}{\alpha^2} + \frac{\pi k_2 \alpha_2}{\alpha^2} \omega \operatorname{Re} A_1(\bar{\beta}_J), & v_1^* &= -\frac{\pi k_2 \alpha_2}{\alpha^2} - \frac{\pi j_2 \alpha_1}{\alpha^2} \omega \operatorname{Re} A_1(\bar{\beta}_J), \\ T_1^* &= \operatorname{Pr} R \operatorname{Re} A_2(\bar{\beta}_J), & u_2^* &= \frac{\pi k_2 \alpha_2}{\alpha^2} \omega \operatorname{Im} A_1(\bar{\beta}_J), \\ v_2^* &= -\frac{\pi j_2 \alpha_1}{\alpha^2} \omega \operatorname{Im} A_1(\bar{\beta}_J), & T_2^* &= \operatorname{Pr} R I_m A_2(\bar{\beta}_J).\end{aligned}$$

Note that

$$\beta_J(\psi_J, \psi_J^*) = (L_R \psi_J, \psi_J^*) = (\psi_J, L_R^* \psi_J^*) = \bar{\beta}_J(\psi_J, \psi_J^*).$$

Hence

$$(\psi_J^1, \psi_J^{1*}) = (\psi_J^2, \psi_J^{2*}), \quad (\psi_J^1, \psi_J^{2*}) = -(\psi_J^2, \psi_J^{1*}). \quad (4.4.74)$$

It is clear that for any real number α , the functions $\Phi_J^* = \Phi_J^{1*} + i\Phi_J^{2*}$, with

$$\Phi_J^{1*} = \psi_J^1 + \alpha\psi_J^{2*}, \quad \Phi_J^{2*} = -\alpha\psi_J^1 + \psi_J^{2*}$$

are also the first conjugate eigenvectors of L_R^* . By (4.4.74), they satisfy

$$\begin{aligned} (\psi_J^1, \Phi_J^{1*}) &= (\psi_J^2, \Phi_J^{2*}) \neq 0, \\ (\psi_J^1, \Phi_J^{2*}) &= (\psi_J^2, \Phi_J^{1*}) = 0, \\ (\psi_{J_k}^i, \Phi_J^*) &= 0, \quad \forall \psi_{J_k}^i \neq \Phi_J^{1*}, \Phi_J^{2*}, \end{aligned} \quad (4.4.75)$$

provided

$$\alpha = \frac{(\psi_J^1, \psi_J^{2*})}{(\psi_J^1, \psi_J^{1*})} = -\frac{(\psi_J^2, \psi_J^{1*})}{(\psi_J^2, \psi_J^{2*})}.$$

The two eigenvectors Φ_J^{1*} and Φ_J^{2*} satisfying (4.4.75) can be expressed as

$$\begin{aligned} \Phi_J^{1*} &= \frac{1}{(\psi_J^1, \psi_J^{1*})} [(\psi_J^1, \psi_J^{1*})\psi_J^{1*} + (\psi_J^1, \psi_J^{2*})\psi_J^{2*}], \\ \Phi_J^{2*} &= \frac{1}{(\psi_J^2, \psi_J^{1*})} [(\psi_J^2, \psi_J^{1*})\psi_J^{1*} + (\psi_J^2, \psi_J^{2*})\psi_J^{2*}] \end{aligned} \quad (4.4.76)$$

By (4.4.75), the reduced equations of (4.4.1)–(4.4.3) read

$$\begin{aligned} \frac{dx}{dt} &= \sigma x + \rho y + \frac{1}{(\psi_J^1, \Phi_J^{1*})} (G(\psi, \psi), \Phi_J^{1*}), \\ \frac{dy}{dt} &= -\rho x + \sigma y + \frac{1}{(\psi_J^2, \Phi_J^{2*})} (G(\psi, \psi), \Phi_J^{2*}), \end{aligned} \quad (4.4.77)$$

where $\psi \in \tilde{H}_1$ is written as

$$\psi = x\psi_J^1 + y\psi_J^2 + \Phi(x, y), \quad (4.4.78)$$

and Φ is the center manifold function.

Inserting (4.4.78) into (4.4.77) we obtain

$$\begin{aligned} \frac{dx}{dt} &= \sigma x + \rho y - \frac{1}{(\psi_J^1, \Phi_J^{1*})} [x(G(\psi_J^1, \Phi_J^{1*}), \Phi) \\ &\quad + y(G(\psi_J^2, \Phi_J^{1*}), \Phi) + x(G(\Phi, \Phi_J^{1*}), \psi_J^1) \\ &\quad + y(G(\Phi, \Phi_J^{1*}), \psi_J^2)] + o(3), \\ \frac{dy}{dt} &= -\rho x + \sigma y - \frac{1}{(\psi_J^2, \Phi_J^{2*})} [x(G(\psi_J^1, \Phi_J^{2*}), \Phi) \\ &\quad + y(G(\psi_J^2, \Phi_J^{2*}), \Phi) + x(G(\Phi, \Phi_J^{2*}), \psi_J^1) \\ &\quad + y(G(\Phi, \Phi_J^{2*}), \psi_J^2)] + o(3) \end{aligned} \quad (4.4.79)$$

As before we have

$$\begin{aligned} G(\psi_J^1, \psi_J^i) &= \frac{\pi}{2} P(u_i \sin 2\phi, v_i \sin 2\phi, 0, 0) + \frac{\pi}{2} (0, 0, 0, T_i \sin 2\pi x_3), \\ G(\psi_J^1, \psi_J^{i*}) &= \frac{\pi}{2} P(u_i^* \sin 2\phi, v_i^* \sin 2\phi, 0, 0) + \frac{\pi}{2} (0, 0, 0, T_i \sin 2\pi x_3), \\ G(\psi_J^2, \psi_J^i) &= G(\psi_J^2, \psi_J^{i*}) = 0, \end{aligned} \quad (4.4.80)$$

for $i = 1, 2$. By (A.1.19) and (4.4.80) with ρ replaced by $-\rho$ in (A.1.19), the center manifold function Φ satisfies

$$\begin{aligned} (L_R^2 + 4\rho^2)L_R\Phi &= -(L_R^2 + 4\rho^2)(x^2G_{11} + xyG_{12}) \\ &\quad + 2\rho^2((x^2 - y^2)G_{11} + 2xyG_{12}) - \rho L_R((y^2 - x^2)G_{12} + 2xyG_{11}), \end{aligned} \quad (4.4.81)$$

where $G_{ij} = G(\psi_J^i, \psi_J^j)$. Based on (4.4.81) we see that

$$\Phi = \Phi_1 + \Phi_2 + \Phi_3,$$

and

$$\begin{aligned} L_R\Phi_1 &= -(x^2G_{11} + xyG_{12}), \\ (L_R^2 + 4\rho^2)L_R\Phi_2 &= 2\rho^2((x^2 - y^2)G_{11} + 2xyG_{12}), \\ (L_R^2 + 4\rho^2)\Phi_3 &= -\rho((y^2 - x^2)G_{12} + 2xyG_{11}). \end{aligned} \quad (4.4.82)$$

By (4.4.80), G_{1i} ($i = 1, 2$) are linear combinations of the eigenvectors $\psi_k = \psi_{2(j_2k_20)}$ and ψ_{002}^1 of L_R . Hence, we have

$$\Phi_j = \varphi_{j1}\psi_k + \varphi_{j2}\psi_{002}^1, \quad j = 1, 2, 3.$$

From (4.4.80) and (4.4.82) we derive that

$$\begin{aligned} \Phi_1 &= -\frac{\pi^2\omega}{8\Pr\alpha^4}(R_1x^2 + I_1xy)\psi_k + \frac{1}{8\pi}(R_2x^2 + I_2xy)\psi_{002}^1, \\ \Phi_2 &= \frac{\pi^2\omega\rho^2}{16\Pr\alpha^4(4\Pr^2\alpha^4 + \rho^2)}(R_1(x^2 - y^2) + 2I_1xy)\psi_k \\ &\quad - \frac{\rho^2}{16\pi(4\pi^2 + \rho^2)}(R_2(x^2 - y^2) + 2I_2xy)\psi_{002}^1, \\ \Phi_3 &= \frac{\pi^2\omega\rho}{8\alpha^2(4\Pr^2\alpha^4 + \rho^2)}(I_1(y^2 - x^2) + 2R_1xy)\psi_k \\ &\quad - \frac{\pi\rho}{8(4\pi^2 + \rho^2)}(I_2(y^2 - x^2) + 2R_2xy)\psi_{002}^1, \end{aligned}$$

where R_j and I_j ($j = 1, 2$) are as in (4.4.73). Thus, we get the center manifold function Φ as follows:

$$\Phi = (E_1x^2 + E_2xy + E_3y^2)\psi_k + (D_1x^2D_2xy + D_3y^2)\psi_{002}^1, \quad (4.4.83)$$

where

$$\begin{aligned} E_1 &= -\frac{\pi^2 \omega R_1}{8\text{Pr } \alpha^4} + \frac{\pi^2 \omega \rho^2 R_1}{16\text{Pr } \alpha^4(4\text{Pr}^2 \alpha^4 + \rho^2)} - \frac{\pi^2 \omega \rho I_1}{8\alpha^2(4\text{Pr}^2 \alpha^4 + \rho^2)}, \\ E_2 &= -\frac{\pi^2 \omega I_1}{8\text{Pr } \alpha^4} + \frac{\pi^2 \omega \rho^2 I_1}{8\text{Pr } \alpha^4(4\text{Pr}^2 \alpha^4 + \rho^2)} + \frac{\pi^2 \omega \rho R_1}{4\alpha^2(4\text{Pr}^2 \alpha^4 + \rho^2)}, \\ E_3 &= -\frac{\pi^2 \omega \rho^2 R_1}{16\text{Pr } \alpha^4(4\text{Pr}^2 \alpha^4 + \rho^2)} + \frac{\pi^2 \omega \rho I_1}{8\alpha^2(4\text{Pr}^2 \alpha^4 + \rho^2)}, \\ D_1 &= \frac{R_2}{8\pi} - \frac{\rho^2 R_2}{16\pi(4\pi^4 + \rho^2)} + \frac{\rho\pi I_2}{4(4\pi^4 + \rho^2)}, \\ D_2 &= \frac{I_2}{8\pi} - \frac{\rho^2 I_2}{8\pi(4\pi^4 + \rho^2)} - \frac{\rho\pi R_2}{4(4\pi^4 + \rho^2)}, \\ D_3 &= \frac{\rho^2 R_2}{16\pi(4\pi^4 + \rho^2)} - \frac{\rho\pi I_2}{8(4\pi^4 + \rho^2)}. \end{aligned}$$

It follows from (4.4.80) and (4.4.83) that

$$(G(\psi_J^2, \psi_J^{i*}), \Phi) = 0, \quad (G(\Phi, \psi_J^{i*}), \psi_J^j) = 0.$$

for $i, j = 1, 2$. By (4.4.74) and (4.4.76), (4.4.79) can be rewritten as

$$\begin{aligned} \frac{dx}{dt} &= \sigma x + \rho y - \frac{x}{D^2} [(\psi_J^2, \psi_J^{2*})(G(\psi_J^1, \psi_J^{1*}), \Phi) \\ &\quad + (\psi_J^1, \psi_J^{2*})(G(\psi_J^1, \psi_J^{2*}), \Phi)] + o(3), \\ \frac{dy}{dt} &= -\rho x + \sigma y - \frac{x}{D^2} [(\psi_J^2, \psi_J^{1*})(G(\psi_J^1, \psi_J^{1*}), \Phi) \\ &\quad + (\psi_J^2, \psi_J^{2*})(G(\psi_J^1, \psi_J^{2*}), \Phi)] + o(3). \end{aligned} \quad (4.4.84)$$

where $D^2 = (\psi_J^1, \psi_J^{1*})^2 + (\psi_J^1, \psi_J^{2*})^2$.

Inserting (4.4.83) into (4.4.84) we find

$$\begin{aligned} \frac{dx}{dt} &= \sigma x + \rho y - \frac{\pi^5}{\alpha^2 D_2} [a_1 x^3 + a_2 x^2 y + a_3 x y^2] + o(3), \\ \frac{dy}{dt} &= -\rho x + \sigma y - \frac{\pi^5}{\alpha^2 D_2} [b_1 x^3 + b_2 x^2 y + b_3 x y^2] + o(3), \end{aligned} \quad (4.4.85)$$

where

$$\begin{aligned} a_1 &= \pi\omega(R_1 B_1 - I_1 B_2)E_1 + \text{Pr } R(R_2 B_1 - I_2 B_2)D_1, \\ a_3 &= \pi\omega(R_1 B_1 - I_1 B_2)E_3 + \text{Pr } R(R_2 B_1 - I_2 B_2)D_3, \\ b_2 &= -\pi\omega(I_1 B_1 + R_1 B_2)E_2 - \text{Pr } R(I_2 B_1 + R_2 B_2)D_2, \end{aligned}$$

and B_1, B_2 are as in (4.4.73).

Thus, the number b in Theorem 2.3.7 reads

$$\begin{aligned}
 b &= -(3a_1 + a_3 + b_2) \\
 &= -\pi\omega(R_1B_1 - I_1B_2)(3E_1 + E_3) \\
 &\quad -\text{Pr } R(R_2B_1 - I_2B_2)(3D_1 + D_3) + \pi\omega(I_1B_1 + R_1B_2)E_2 \\
 &\quad +\text{Pr } R(I_2B_1 + R_2B_2)D_2 \\
 &= \frac{1}{4}[\pi^3\omega^2(R_1B_1 - I_1B_2)C_1 + \text{Pr } R(R_2B_1 - I_2B_2)C_2 \\
 &\quad +\pi^3\omega^2(I_1B_1 + R_1B_2)C_3 + \text{Pr } R(I_2B_1 + R_2B_2)C_4] \\
 &= \frac{1}{4}b_2,
 \end{aligned}$$

where b_2 is as in (4.4.73).

By Theorem 2.3.7, if $b < 0$, the transition of (4.4.85) at $R = R_{c_2}$ is continuous, i.e., the transition is an attractor bifurcation which is a super-critical Hopf bifurcation, and if $b > 0$, the transition is jump, i.e., the transition is a sub-critical Hopf bifurcation. Based on Theorem 2.5.7, as the transition is jump, there exists a singular separation at some $R^* < R_{c_2}$. Thus the theorem is proved. \square

4.4.6 Physical Remarks

We start with a few remarks on the physical parameters appeared in the model. First, the periods $L_1\pi$ and $L_2\pi$ in (4.4.3) generally represent the horizontal size of the containers in physical experiments, and in the atmospheric and oceanic circulations, they characterize the horizontal motion scale of circulation cells.

Second, the nondimensional number ω is proportional to the rotating angular velocity, which is given by

$$\omega = \Omega L^2/\kappa,$$

where Ω is the rotating angular velocity, L is the length scale, and κ is the thermal diffusivity. In the mid-latitude atmospheric and oceanic circulations, ω is the reciprocal of the Rossby number Ro :

$$\omega = \frac{1}{R_0}.$$

The transition theorems obtained in the previous sections give precise criteria to distinguish the Type-I and Type-II transitions. Hereafter we shall discuss them in detail.

Transition Types of Steady-State Solutions Let $\alpha_0 = \min\{\alpha_1, \alpha_2\}$, $\alpha_i = L_i^{-1}$ ($i = 1, 2$). By Lemma 4.4.2, if ω^2/Pr^2 satisfies that

$$\lambda \stackrel{\text{def}}{=} \frac{\omega^2}{\Pr^2} \begin{cases} \geq 0 & \text{for } \Pr^2 \geq 1, \\ < \frac{(\pi^2 + \alpha_0^2)^3}{\pi^2} & \text{for } \Pr^2 < 1, \end{cases} \quad (4.4.86)$$

then R_{c_1} is the first critical Rayleigh number. Consider the number

$$\begin{aligned} & \Pr^4(\alpha_{j_1 k_1}^2 + \pi^2)^3 \alpha_{j_1 k_1}^2 - \pi^2(\pi^2 - \Pr^2 \alpha_{j_1 k_1}^2) \omega^2 \\ &= \Pr^2[\Pr^2 \alpha_{j_1 k_1}^2 (\alpha_{j_1 k_1}^2 + \pi^2)^3 - \pi^2(\pi^2 - \Pr^2 \alpha_{j_1 k_1}^2) \lambda]. \end{aligned}$$

It is clear that

$$\text{Condition (4.4.54) holds} \iff \Pr^2 > \frac{\pi^4 \lambda}{\alpha_{j_1 k_1}^2 (\pi^2 \lambda + (\alpha_{j_1 k_1}^2 + \pi^2)^3)}, \quad (4.4.87)$$

$$\text{Condition (4.4.55) holds} \iff \Pr^2 < \frac{\pi^4 \lambda}{\alpha_{j_1 k_1}^2 (\pi^2 \lambda + (\alpha_{j_1 k_1}^2 + \pi^2)^3)}. \quad (4.4.88)$$

We know that the function

$$f(\lambda) = \frac{\pi^4 \lambda}{\alpha_{j_1 k_1}^2 (\pi^2 \lambda + (\alpha_{j_1 k_1}^2 + \pi^2)^3)}$$

is exactly increasing on λ . Therefore, (4.4.87) and (4.4.88) show that if $\pi^2 > \Pr^2 \alpha_{j_1 k_1}^2$, small values of $\lambda = \omega^2/\Pr^2$ lead to a Type-I (continuous) transition at $R = R_{c_1}$, and big value of λ give rise to a Type-II (jump) transition.

Transition Types of Periodic Solutions By (4.4.37), when R_{c_2} is the first critical Rayleigh number, the number λ defined by (4.4.86) satisfied that

$$\lambda > \frac{(1 + \Pr)(\alpha^2 + \pi^2)^3}{(1 - \Pr)\pi^2} \quad \text{for } 0 < \Pr < 1,$$

where $\alpha^2 = \alpha_{j_2 k_2}^2$ as in (4.4.38). Let

$$\gamma^2 = \alpha_{j_2 k_2}^2 + \pi^2.$$

In the following, we show that both Type-I and Type-II transitions to the periodic solutions, dictated by the sign of b_2 , can appear in different physical parameter ranges.

First, we assume that \Pr ($0 < \Pr < 1$) is given, and

$$\omega^2 \gg \gamma^6 \gg 1, \quad (4.4.89)$$

Then, we have

$$\begin{aligned}
R = R_{c_2} &\cong \frac{2\pi^2\omega^2}{(\Pr + 1)\alpha^2}, & \rho^2 &\cong \frac{1}{\gamma^2} \frac{1 - \Pr}{1 + \Pr} \pi^2 \omega^2, \\
R_1 &\cong \frac{\Pr \gamma^2}{\rho^2}, & R_2 &\cong \frac{\gamma^2}{\rho^2}, \\
I_1 &\cong -\frac{1}{\rho}, & I_2 &\cong -\frac{1}{\rho}, \\
C_1 &\cong \frac{R_1}{\Pr \alpha^4} - \frac{1}{\alpha^2 \rho^2} = \frac{\pi^2}{\alpha^4} \frac{1}{\rho^2}, & C_2 &\cong \frac{\Pr(3\alpha^2 + \pi^2)}{\alpha^2 \rho^3}, \\
C_3 &\cong -\frac{\alpha^2}{\pi \rho^2}, & C_4 &\cong -\frac{\pi(\gamma^2 + 4\pi^2)}{\rho^3}, \\
B_1 &\cong \frac{(1 - \Pr) \pi^2 \omega^2}{\rho^2}, & B_2 &\cong \frac{(1 - \Pr) \Pr \gamma^2 \pi^2 \omega^2}{\rho^3}.
\end{aligned} \tag{4.4.90}$$

By (4.4.89) and (4.4.90), the number b_2 defined by (4.4.73) is approximated by

$$\begin{aligned}
b_2 &\cong \frac{\Pr(1 - \Pr) \pi^3 \omega^4}{\alpha^4 \rho^6} \left[2\pi^2 + \pi^4 \alpha^2 + \right. \\
&\quad \left. + \frac{2\alpha^4 \pi^2 + 10\alpha^2 \pi^4}{\Pr + 1} - 3\alpha^4 \pi^2 - 2\alpha^6 - 2\alpha^4 \pi^2 \right] + o(\rho^{-7}) \omega^4.
\end{aligned}$$

By (4.4.89) we see that $\rho^{-1} \cong 0$. Hence, as $\alpha^2 \leq \pi^2$ we deduce that $b_2 > 0$. Thus, by Theorem 4.4.12, we claim that when ω^2 is much large than $\gamma^6 = (\alpha^2 + \pi^2)^3$ and $\alpha^2 \geq \frac{\pi}{2}$, the transition of periodic solutions at $R = R_{c_2}$ is of Type-II.

Next, we assume that $\rho > 0$ is small enough, i.e.

$$1 \gg \rho > 0. \tag{4.4.91}$$

In this case, the number b_2 can be approximated by

$$\begin{aligned}
b_2 &= (\pi^4 \omega^2 (3R_1^2 - I_1^2) + \Pr^2 R \alpha^4 (I_2^2 - 3R_2^2)) B_1 \\
&\quad - 4(\pi^4 \omega^2 R_1 I_1 - \Pr^2 R \alpha^4 R_2 I_2) B_2.
\end{aligned} \tag{4.4.92}$$

By (4.4.91) we see that

$$\begin{aligned}
\gamma^6 &\cong \frac{1 - \Pr}{\Pr^2(1 + \Pr)} \pi^2 \omega^2, & R = R_{c_2} &\cong \frac{2\pi^2 \omega^2}{\alpha^2 \Pr^2(1 + \Pr)}, \\
R_1 &\cong \frac{1}{\Pr \gamma^2}, & R_2 &\cong \frac{1}{\gamma^2}, \\
I_1 &\cong -\frac{\rho}{\Pr^2 \gamma^4}, & I_2 &\cong -\frac{\rho}{\gamma^4}, \\
B_1 &\cong \frac{\pi^2 \omega^2 I_2^2}{\Pr} \left(\frac{1}{\Pr^3} - \frac{2}{1 + \Pr} \right), & B_2 &\cong \frac{\pi^2 \omega^2 R_2 I_2}{\Pr} \left(\frac{1}{\Pr^2} - \frac{2}{1 + \Pr} \right).
\end{aligned} \tag{4.4.93}$$

From (4.4.92) and (4.4.93) we derive

$$\begin{aligned} b_2 \cong & \frac{\pi^4 \omega^4 I_2^2 R_2^2}{\text{Pr}} [3(\frac{\pi^2}{\text{Pr}^2} - \frac{2}{\alpha^2(1 + \text{Pr})})(\frac{1}{P_r^3} - \frac{2}{1 + \text{Pr}}) \\ & - 4(\frac{\pi^2}{P_r^3} - \frac{2}{\alpha^2(1 + \text{Pr})})(\frac{1}{\text{Pr}^2} - \frac{2}{1 + \text{Pr}})]. \end{aligned}$$

It is clear that

$$b_2 < 0 \quad \text{if } \text{Pr} < \delta_0 \text{ for some } \frac{3}{5} < \delta_0 < 1.$$

Thus, by Theorem 4.4.12, under condition (4.4.91) if $\text{Pr} < \delta_0$, the transition of Periodic solutions at $R = R_{c_2}$ is of Type-II.

In summary, we have shown that (1) both continuous and jump transitions do occur in different flow regimes and (2) the jump transitions lead to hysteresis phenomena and a singularity (steady state or periodic orbit) separation, giving rise to more complex (possibly chaotic) flow regimes. The transitions are precisely quantified by specific computable parameters.

4.5 Convection Scale Theory

The convection scale law presented in this section is observed recently by the authors and has been used in Ma and Wang (2011b, 2010c,b). This is associated with a phenomenon appearing in convection problems. It is known that thermal convection appears when the Rayleigh number R crosses the critical Rayleigh number R_c . In (4.1.1), the Rayleigh number is given by

$$R = \frac{g\alpha T(T_0 - T_1)}{\kappa\nu} h^3 \quad (\text{h the height of the fluid}). \quad (4.5.1)$$

For the free boundary condition, by (4.1.38), the critical Rayleigh number is

$$R_c = \min_{\alpha^2} \frac{(\alpha^2 + \pi^2)^3}{\alpha^2} = \frac{(\alpha_c^2 + \pi^2)^3}{\alpha_c^2} = \frac{27\pi^4}{4}, \quad (4.5.2)$$

where $\alpha_c^2 = \frac{\pi^2}{2}$, and for the Dirichlet (rigid) boundary condition, by Chandrasekhar (1981), R_c is

$$R_c = 1700. \quad (4.5.3)$$

According to (4.5.1)–(4.5.3), as the horizontal scale L of the fluid is much larger than its height h , i.e., $L \gg h$, the thermal convection appears at the temperature difference

$$T_0 - T_1 = \begin{cases} \frac{27\kappa\nu\pi^4}{4g\alpha_T h^3} & \text{for the free boundary condition,} \\ \frac{1700\kappa\nu}{ga_T h^3} & \text{for the Dirichlet boundary condition.} \end{cases} \quad (4.5.4)$$

From (4.5.4) we see that the critical temperature difference satisfies

$$\Delta T_c = T_0 - T_1 \rightarrow \infty \quad \text{if } h \rightarrow 0, \quad (4.5.5)$$

$$\Delta T_c = T_0 - T_1 \rightarrow 0 \quad \text{if } h \rightarrow \infty. \quad (4.5.6)$$

Physically, (4.5.5) amounts to saying that for a given fluid there is a minimal size $h_0 > 0$ such that when $h < h_0$, no convection occurs for any $\Delta T = T_0 - T_1 > 0$. This is in agreement with the fact that in the real world, when h is small, it is hard to maintain a high temperature gradient to generate the vertical convection.

However, (4.5.6) has certain physical discrepancies with the energy conservation of the system. In fact, as h increases, it needs more energy to overcome the friction force to drive fluid to move. With ΔT representing this driving force, (4.5.6) shows that when h increases, ΔT decreases.

Here we present a method to resolve this issue. Note that the critical value $\alpha^2 = \pi^2/L^2$ defines the horizontal convection scale L (h scales to $h = 1$): $L^2 = \pi^2/\alpha^2$. Hence, at the critical value $\alpha_c^2 = \frac{\pi^2}{2}$ in (4.5.2), the convective cell size is

$$L_c^2 = 2. \quad (4.5.7)$$

When h is not large, this value (4.5.7) is a good approximation, we need to modify the model to accommodate the case when h is large.

Instead of (4.5.7), we consider the convection scale L_c as an increasing function of h :

$$L_c^2 = \psi(h), \quad \text{such that } \psi(0) = 2, \psi'(h) > 0. \quad (4.5.8)$$

Thus, the critical Rayleigh number R_c reaches its minimal at

$$\alpha_c^2 = \frac{\pi^2}{\psi(h)}, \quad (4.5.9)$$

rather than at $\alpha_c^2 = \pi^2/2$.

The convection scale law of (4.5.8) and (4.5.9) should be reflected by the hydrodynamical equations. In other words, we need to revise the Boussinesq equations leading to new critical values given by (4.5.8) and (4.5.9). The method we propose here is to add a term:

$$F = -(\sigma_0 u_1, \sigma_0 u_2, \sigma_1 u_3) \quad (4.5.10)$$

in the right-hand side of the first equations of (4.1.2), where σ_i depends only on h satisfying

$$\sigma_i(0) = 0, \quad \sigma'_i(h) \geq 0, \quad i = 0, 1. \quad (4.5.11)$$

Physically, (4.5.10) and (4.5.11) stand for the added resistance generated by pressure, called the damping terms.

In nondimensional form, the revised Boussinesq equations read

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla)u &= \text{Pr} [\Delta u - \nabla p - f + RTk], \\ \frac{\partial T}{\partial t} + (u \cdot \nabla)T &= \Delta T + u_3, \\ \text{div}u &= 0, \end{aligned} \quad (4.5.12)$$

where $f = (\delta_0 u_1, \delta_0 u_2, \delta_1 u_3)$, and

$$\delta_i(h) = \frac{h^2}{\nu} \sigma_i(h), \quad (i = 0, 1). \quad (4.5.13)$$

For the revised equation (4.5.12) with the free boundary conditions (4.1.34), the critical Rayleigh number is

$$R_c = \min_{\alpha^2} \left[(\alpha^2 + \pi^2)\delta_1 + \frac{(\alpha^2 + \pi^2)^3 + \pi^2(\alpha^2 + \pi^2)\delta_0}{\alpha^2} \right]. \quad (4.5.14)$$

Let $g(x) = (x + \pi^2)\delta_1 + ((x + \pi^2)^3 + \pi^2(x + \pi^2)\delta_0)/x$. Then, $x_c = \alpha_c^2$ satisfies $g'(x) = 0$. Thus we get

$$\delta_1 \alpha_c^4 - (\alpha_c^2 + \pi^2)^3 - \pi^2(\alpha_c^2 + \pi^2)\delta_0 + 3\alpha_c^2(\alpha_c^2 + \pi^2)^2 + \pi^2\alpha_c^2\delta_0 = 0. \quad (4.5.15)$$

We assume that $\delta_1 > \delta_0$ for $h \neq 0$. By $\alpha_c^2 = \pi^2/L_c^2$, using (4.5.10), (4.5.11), and (4.5.15), we derive the convection scale law of (4.5.8) and (4.5.9). In particular, when h is large, $\delta_1 \gg 1$ and $\delta_1 \gg \delta_0$, we derive from (4.5.14) and (4.5.15) that

$$\alpha_c^4 \cong \frac{\pi^4(\pi^2 + \delta_0)}{\delta_1}, \quad L_c^2 \cong \delta_1^{1/2}/(\pi^2 + \delta_0)^{1/2}, \quad R_c \cong \pi^2\delta_1 + \pi^2\delta_0 L_c^2, \quad (4.5.16)$$

These formulas are very useful in studying large-scale convection motion for the height of fluid $h \geq 10m$. It can fairly solve the contradiction caused by (4.5.6). Actually, by (4.5.1), (4.5.13), and (4.5.16), we obtain the critical temperature difference

$$\Delta T_c = \frac{\pi^2 \kappa \sigma_1(h)}{g \alpha_T h} \quad \text{for } h \gg 1m.$$

It suggests that $\sigma_1(h) \cong c \cdot h^2$, which results in that the critical temperature gradient is independent of h .

In summary, by (4.5.13), we propose to take

$$\delta_0 = C_0 h^4/\nu, \quad \delta_1 = C_1 h^4/\nu, \quad (4.5.17)$$

where C_0 and C_1 are constants.

For the rotating convection problem, the revised rotating Boussinesq equations are expressed as

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla)u &= \text{Pr} [\Delta u - \nabla p - f + RTk] - \omega k \times u, \\ \frac{\partial T}{\partial t} + (u \cdot \nabla)T &= \Delta T + u_3, \\ \text{div}u &= 0. \end{aligned} \quad (4.5.18)$$

For the boundary condition (4.4.2)–(4.4.3), similar to (4.4.21), the eigenvalues β of linearized equations of (4.5.18) satisfies

$$\begin{aligned} &\alpha^2(\text{Pr} \gamma^2 + \text{Pr} \delta_0 + \beta)(\text{Pr} \gamma^2 + \text{Pr} \delta_1 + \beta)(\gamma^2 + \beta) \\ &+ (\text{Pr} \gamma^2 + \text{Pr} \delta_0 + \beta)^2(\gamma^2 + \beta)\pi^2 + \omega^2(\gamma^2 + \beta)\pi^2 \\ &- \alpha^2\text{Pr} R(\text{Pr} \gamma^2 + \text{Pr} \delta_0 + \beta) = 0, \end{aligned} \quad (4.5.19)$$

where $\gamma^2 = \alpha^2 + \pi^2$, $\alpha^2 = \pi^2(j^2L_1^{-2} + k^2L_2^{-2})$.

Like (4.4.33), let $\beta = 0$ in (4.5.19), we get the first critical Rayleigh number as follows

$$R_{c_1} = \min_{\alpha^2} \left[(\alpha^2 + \pi^2)\delta_1 + \frac{(\alpha^2 + \pi^2)^3 + \pi^2(\alpha^2 + \pi^2)\delta_0}{\alpha^2} + \frac{\pi^2\omega^2(\alpha^2 + \pi^2)}{\alpha^2\text{Pr}^2(\alpha^2 + \pi^2 + \delta_0)} \right]. \quad (4.5.20)$$

When h is large, $\delta_1 \gg 1$, and $\delta_1 \gg \delta_0$, from (4.5.20) we obtain the following scale formulas of large-scale convection motion.

$$\alpha_c^4 \cong \frac{\pi^4}{\delta_1} \left(\pi^2 + \delta_0 + \frac{\omega^2}{\text{Pr}^2(\pi^2 + \delta_0)} \right), \quad (4.5.21)$$

$$\delta_1 \cong (\pi^2 + \delta_0)L_c^4 + \frac{\omega^2}{\text{Pr}^2(\pi^2 + \delta_0)}L_c^4, \quad (4.5.22)$$

$$R_{c_1} \cong \pi^2\delta_1 + \frac{\pi^6 + \pi^4\delta_0}{\alpha_c^2} + \frac{\pi^4\omega^2}{\alpha_c^2\text{Pr}^2(\pi^2 + \delta_0)}. \quad (4.5.23)$$

Let $\beta = i\rho(\rho \neq 0)$ in (4.5.19), then we get the second critical Rayleigh number R_{c_2} given by

$$\begin{aligned} R_{c_2} &= \min_{\alpha^2} \frac{1}{\alpha^2\text{Pr} [(\text{Pr} + 1)\gamma^4 + \text{Pr}(\alpha^2\delta_1 + \pi^2\delta_0)]} \\ &\times [\alpha^2\text{Pr}^2(\alpha^2\delta_1 + (\gamma^2 + \pi^2)\delta_0) \times B + 2\text{Pr} \pi^2\omega^2\gamma^4 \\ &+ \alpha^2\text{Pr} \gamma^4(2\text{Pr}^2r_0^2r_1^2 + (2\text{Pr} + 1)(r_1^2 + \gamma_0^2)r^2) \\ &+ 2\pi^2\text{Pr} r^4(\text{Pr}^2r_0^4 + (2\text{Pr} + 1)\gamma_0^2r^2)], \end{aligned} \quad (4.5.24)$$

where

$$r_0^2 = \alpha^2 + \pi^2 + \delta_0,$$

$$r_1^2 = \alpha^2 + \pi^2 + \delta_1,$$

$$B = \alpha^2 \text{Pr} (\text{Pr} r_0^2 r_1^2 + r_1^2 r^2 + r_0^2 r^2) + \pi^2 \text{Pr} (\text{Pr} r_0^4 + 2r_0^2 r^2) + \pi^2 \omega^2.$$

As we shall see later, these two sets of formulas (4.5.16) and (4.5.21)–(4.5.23) provide correct zonal and meridional circulation scales.



Chapter 5

Geophysical Fluid Dynamics and Climate Dynamics

Our Earth's atmosphere and oceans are rotating geophysical fluids that are two important components of the planet's climate system. The atmosphere and the oceans are extremely rich in their organization and complexity, and many phenomena that they exhibit, involving a broad range of temporal and spatial scales (Charney, 1948), cannot be reproduced in the laboratory. An understanding of the complex scientific issues of geophysical fluid dynamics requires the combined efforts of scientists in many fields.

The behavior of the atmosphere, the ocean, or the coupled atmosphere and ocean can be viewed as an initial and boundary value problem (Bjerknes, 1904; Rossby, 1926; Phillips, 1956). The ideas of dynamical systems theory were pioneered in Lorenz (1963a,b), Stommel (1961), Veronis (1963, 1966), among others, whose authors explored the bifurcation structure of low-order models of atmospheric and oceanic flows. The formulation of the primitive equations of atmosphere and ocean as an infinite-dimensional dissipative dynamical system was initiated in Lions et al. (1992a,b, 1993).

According to von Neumann (1960), the motion of the atmosphere can be divided into three categories, depending on the time scale of the phenomenon under consideration, or equivalently, on how far into the future one is attempting to make predictions of atmospheric conditions. It is convenient, then, to consider motions corresponding to short-, medium-, and long-term behavior of the atmosphere. Short-term weather forecasting, for example, clearly requires the efficient solution of an initial and boundary value problem of the atmospheric and oceanic systems.

Thanks to the pioneering work (Lorenz, 1963a), the medium- and long-term behavior of the atmosphere is often referred to as climate dynamics, and it is governed by the dynamical behavior of infinite-dimensional climate dynamical systems. Great advances in numerical and observational studies have enabled researchers in climate dynamics to set as one of their primary goals to document, through careful theoretical and numerical studies, the presence of low-frequency climate variability, to verify the robustness of this variability's characteristics under changes in the model's parameters, and to explain its physical mechanisms. Obtaining a thorough understanding of this variability is a challenging problem with important practical

implications for geophysical efforts to quantify predictability, analyze error growth in dynamical models, and develop efficient forecasting methods.

In mathematical terms, this primary goal is best described by dynamical transitions associated with typical sources of atmospheric and oceanic variability, including, for example, the meridional overturning oceanic circulation and the tropical atmosphere–ocean modes associated with the El Niño Southern Oscillations (ENSO).

Based on the authors’ recent work (Ma and Wang, 2010b,c, 2011b), the main objective of this chapter is to initiate a study of dynamical transitions and the stability of large-scale atmospheric and oceanic circulations, focusing on a few typical sources of climate variability. Such variability, independently and interactively, may play a significant role in past and future climate change.

Again, we reiterate that the philosophy of dynamical transition theory of geophysical flows is completely different from that of classical bifurcation studies. A full classification of transition states and their dynamical behavior, including, in particular, that of the transients, is crucial in climate studies, as indicated by existing numerical and observational data.

5.1 Modeling and General Characteristics of Geophysical Flows

The hydrodynamical equations governing the large-scale atmospheric and oceanic flows are the Navier–Stokes equations with the Coriolis force generated by the earth’s rotation, coupled with the first law of thermodynamics and with diffusion equation for the humidity, and salinity. There are many excellent books in this area; see among others (Pedlosky, 1987; Lorenz, 1967; Ghil and Childress, 1987; Dijkstra, 2000; Salby, 1996).

The general system of equations was basically the equations used by L. F. Richardson in his pioneering work (Richardson, 1922). However, they are in general too complicated to conduct theoretical analysis. As practiced by the earlier workers such as J. Charney, and from the lessons learned by the failure of Richardson’s pioneering work, one tries to be satisfied with simplified models approximating the actual motions to a greater or lesser degree instead of attempting to deal with the atmosphere in all its complexity. By starting with models incorporating only what are thought to be the most important of atmospheric influences, and by gradually bringing in others, one is able to proceed inductively and thereby to avoid the pitfalls inevitably encountered when a great many poorly understood factors are introduced all at once. The simplifications are usually done by taking into consideration of some main features of the large-scale atmosphere.

Hereafter we briefly mention a few general features of the large scale atmosphere and the ocean.

FIRST, the most important feature of the large scale atmospheric circulation is the smallness of the aspect ratio between the vertical and horizontal scales, leading to hydrostatic equation replacing the vertical momentum equation. The resulting

system of equation is called the primitive equations (PEs); see among others (Lions et al., 1992a; Salby, 1996).

More precisely, the PEs are obtained from the general equations of hydrodynamics and thermodynamics of the compressible atmosphere, by approximating the momentum equation in the vertical direction with the hydrostatic equation:

$$\frac{\partial p}{\partial z} = -\rho g. \quad (5.1.1)$$

This hydrostatic equation is based on the ratio between the vertical and horizontal scale being small. Here ρ is the density, g the gravitational constant, and $z = r - a$ height above the sea level, r the radial distance, and a the mean radius of the earth.

Equation (5.1.1) expresses the fact that p is a decreasing function along the vertical so that one can use p instead of z as the vertical variable. Motivated by this hydrostatic approximation, we can introduce a coordinate transformation from (θ, φ, z) to (θ, φ, p) .

SECOND, another important feature of the large scale geophysical flows is that the earth rotates much faster than the wind velocity or the velocity of the seawater. The simplification taking the fast rotation into consideration leads to the classical quasi-geostrophic equations (Charney, 1947), and many balanced equations.

THIRD, as discussed in Sect. 4.5, for large scale atmospheric and oceanic circulations, the critical convection scales of the classical theory are often too small. By carefully examining the Bénard convection, we show that in order to have a correct scaling for the large scale circulations, we need to add to the basic equations an anisotropic turbulent friction term $(\delta_0 u_1, \delta_0 u_2, \delta_1 u_3)$, where u_1 and u_2 stand for the horizontal velocity components, u_3 stands for the vertical velocity component, and

$$\delta_0 = C_0 h^4 / \nu, \quad \delta_1 = C_1 h^4 / \nu, \quad (5.1.2)$$

where C_0 and C_1 are constants, and h is the vertical length scale. We can think about this added friction term as due to the turbulent averaging process, although the mathematical derivation of this scaling law is based on the analysis of the dynamic transitions of convection problems.

FOURTH, the large scale atmospheric circulation can essentially be regarded as a thermal convection process caused by the temperature gradient between the earth's surface and the tropopause. There are two types of convective cells: the meridional circulation and the zonal circulation. The meridional circulation is characterized by the polar, the Ferrel, and the Hadley cells, as schematically illustrated in Fig. 5.1. These cells and their interactions are major players in global heat transport.

The large-scale zonal circulation consists of six convective cells over the equator, as shown in Fig. 5.2. The overall atmospheric motion is known as the zonal overturning circulation, and also called the Walker circulation (Salby, 1996).

One remarkable feature of the global atmospheric circulation is that the equatorial Walker circulation divides the whole earth into three invariant regions of atmospheric flow: the northern hemisphere, the southern hemisphere, and the equatorial zone. Also, note that although the large-scale structure of the zonal overturning circulation

Fig. 5.1 Three large-scale cells in the meridional direction: the Polar cell, the Ferrel cell, and the Hadley cell

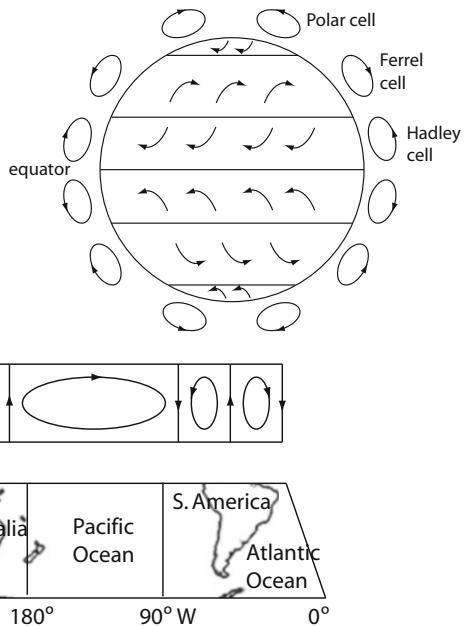


Fig. 5.2 The equatorial Walker circulation

varies from year to year, as seen, e.g., in the El Niño-Southern Oscillation phenomenon over the equatorial Pacific, its basic structure remains fairly stable. Hence to study the formation and dynamics of large scale atmospheric circulations, it is appropriate to consider the following Atmospheric Zone Hypothesis:

Circulation Zone Hypothesis of Large Scale Atmosphere *The atmospheric circulation has three invariant regions: the northern hemisphere domain ($0 < \theta \leq \frac{\pi}{2}$), the southern hemisphere domain ($-\frac{\pi}{2} \leq \theta < 0$), and the equatorial zone ($\theta = 0$). Namely, the large-scale circulations in these invariant regions can act alone with less influence on the others. In particular, the velocity field $u = (u_\varphi, u_\theta, u_z)$ of the atmospheric circulation has a vanished latitudinal component in a narrow equatorial zone, i.e., $u_\theta = 0$, for $-\varepsilon < \theta < \varepsilon$, where $\varepsilon > 0$ is a small number.*

5.2 El Niño-Southern Oscillation (ENSO)

ENSO is the known strongest interannual climate variability associated with strong atmosphere-ocean coupling, which has significant impacts on the global climate. ENSO has global impacts on temperature, precipitation, tropical cyclones, human

health, and even conflict. ENSO is in fact a phenomenon associated with interannual oscillation between warm events (El Niño phase) and cold events (La Niña phase) in the equatorial eastern Pacific sea-surface temperature (SST) anomaly. It is convenient and understandable to employ simplified coupled dynamical models to investigate some essential behaviors of ENSO dynamics; see among many others (Schopf and Suarez, 1987; Battisti and Hirst, 1989; Jin, 1996; Zebiak and Cane, 1987; Philander and Fedorov, 2003; Jin et al., 1996; Neelin et al., 1998; Neelin, 1990a,b; Ghil, 2000; Schneider et al., 2003; Sardeshmukh et al., 2000; Samelson, 2008).

An interesting current debate is whether ENSO is best modeled as a stochastic or chaotic system—linear and noise-forced, or nonlinear oscillatory and unstable system (Philander and Fedorov, 2003). It is obvious that a careful fundamental level examination of the problem is crucial. The main objective of this section is an attempt to address this fundamental question.

By establishing a rigorous mathematical theory on the formation of the Walker circulation over the tropics, we present in this section a mechanism of the ENSO, first derived in Ma and Wang (2010c, 2011b). Namely, ENSO is a self-organizing and self-excitation system, with two highly coupled processes:

- the oscillation between the two metastable warm (El Niño phase) and cold (La Niña phase) events, and
- the spatiotemporal oscillation of the sea surface temperature (SST) field.

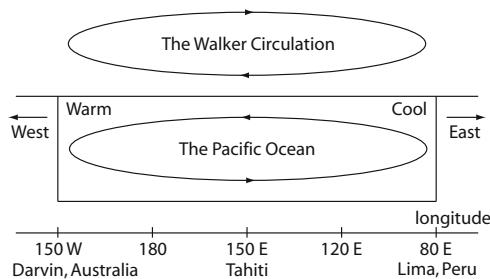
The symbiotic interplay between these two processes gives rise to the climate variability associated with the ENSO, leads to both the random and deterministic features of the ENSO, and defines a new natural *feedback* mechanism, which drives the sporadic oscillation of the ENSO. On one hand, the oscillation of metastable states has direct influence on the SST, leading to (1) the intensification or weakening of the oceanic upwelling, and (2) the variation of the atmospheric Rayleigh number. Consequently, forcing the system to adjust to a different metastable phase, and to intensify/weaken the corresponding event. On the other hand, the oscillation of the SST has a direct influence on the atmospheric Rayleigh number, leading to a different El Niño, La Niña, and normal conditions.

5.2.1 Walker Circulation and ENSO

Upwelling and zonal circulation over the tropics contains six convective cells, as shown in Fig. 5.2. Usually the Walker circulation is referred to the Pacific cell over the equator, responsible for creating ocean upwelling off the coasts of Peru and Ecuador. It was discovered by Jacob Bjerknes in 1969 and was named after the English physicist Gilbert Walker, an early-twentieth-century director of British observatories in India, who discovered an indisputable link between periodic pressure variations in the Indian Ocean and the Pacific, which he termed the “Southern Oscillation.”

The Walker cell is of such importance that its behavior is a key factor giving rise to the El Niño (more precisely, ENSO) phenomenon. When the convective activity weakens or reverses, an El Niño phase takes place, causing the ocean surface to be warmer than average, reducing or terminating the upwelling of cold water. A particularly strong Walker circulation causes a La Niña event, resulting in cooler SST due to stronger upwelling; see Fig. 5.3.

Fig. 5.3 A schematic diagram of the Walker circulation at the equatorial Pacific Ocean, a main factor of the El Niño and the la Niña phenomena



The atmospheric motion associated with the Walker circulation affects the loops on either side. Under normal years, the weather behaves as expected. But every few years, the winters become unusually warm or unusually cold, or the frequency of hurricanes increases or decreases. This entirely ocean-based cell is seen at the lower surface as easterly trade winds which drive the seawater and air warmed by sun moving westward. The western side of the equatorial Pacific is characterized by warm, wet low pressure weather, and the eastern side is characterized by cool, dry high pressure weather. The ocean is about 60 cm higher in the western Pacific as the result of this circulation. The air is returned at the upper surface to the east, and it now becomes much cooler and drier. An El Niño phase is characterized by a breakdown of this water and air cycle, resulting in relatively warm water and moist air in the eastern Pacific. Meanwhile, a La Niña phase is characterized by the strengthening of this cycle, resulting in much cooler water and drier air than the normal phase. During an El Niño or La Niña event, which occurs about every 3–6 years, the weather sets in for an indeterminate period.

It is then clear that the ENSO is a coupled atmospheric-ocean phenomena. The main objective of this section is to address the mechanism of atmospheric Walker circulations over the tropics from the dynamic transition point of view.

5.2.2 Equatorial Circulation Equations

The atmospheric motion equations over the tropics are the Boussinesq equations restricted on $\theta = 0$, where the meridional velocity component u_θ is set to zero, and

the effect of the turbulent friction is taken into considering using the scaling law introduced in Sect. 4.5:

$$\begin{aligned} \frac{\partial u_\varphi}{\partial t} &= -(u \cdot \nabla) u_\varphi - \frac{u_\varphi u_z}{r_0} + \nu \left(\Delta u_\varphi + \frac{2}{r_0^2} \frac{\partial u_z}{\partial \varphi} - \frac{2u_\varphi}{r_0^2} \right) \\ &\quad - \sigma_0 u_\varphi - 2\Omega u_z - \frac{1}{\rho_0 r_0} \frac{\partial p}{\partial \varphi}, \\ \frac{\partial u_z}{\partial t} &= -(u \cdot \nabla) u_z + \frac{u_\varphi^2}{r_0} + \nu \left(\Delta u_z - \frac{2}{r_0^2} \frac{\partial u_\varphi}{\partial \varphi} - \frac{2u_z}{r_0^2} \right) \\ &\quad - \sigma_1 u_z + 2\Omega u_\varphi - \frac{1}{\rho_0} \frac{\partial p}{\partial z} - (1 - \alpha_T(T - T_0))g, \\ \frac{\partial T}{\partial t} &= -(u \cdot \nabla) T + \kappa \Delta T + Q, \\ \frac{1}{r_0} \frac{\partial u_\varphi}{\partial \varphi} + \frac{\partial u_z}{\partial z} &= 0. \end{aligned} \tag{5.2.1}$$

Here $\sigma_i = C_i h^2$ ($i = 0, 1$) represent the turbulent friction, r_0 is the radius of the earth, the spatial domain is taken as $M = S_{r_0}^1 \times (r_0, r_0+h)$ with $S_{r_0}^1$ being the one-dimensional circle with radius r_0 , and

$$\begin{aligned} (u \cdot \nabla) &= \frac{u_\varphi}{r_0} \frac{\partial}{\partial \varphi} + u_z \frac{\partial}{\partial z}, & \Delta &= \frac{1}{r_0^2} \frac{\partial^2}{\partial \varphi^2} + \frac{\partial^2}{\partial z^2}, \\ (x_1, x_2) &= (r_0 \varphi, z), & (u_1, u_2) &= (u_\varphi, u_z). \end{aligned}$$

The temperature T_1 at the tropopause $z = r_0 + h$ is assumed to be a constant. We take T_0 as the average on the lower surface $z = r_0$. Also, let

$$\begin{aligned} x &= hx', & t &= h^2 t' / \kappa, & u &= \kappa u' / h, \\ T &= (T_0 - T_1)T' + T_0 - (T - T_0)x'_2, \\ p &= \kappa \nu \rho_0 p' / h^2 - g \rho_0 (hx'_2 + \alpha_T(T_0 - T_1)h(x'_2)^2 / 2). \end{aligned}$$

Also, we define the Rayleigh number, the Prandtl number and the scaling laws by

$$R = \frac{\alpha_T g (T_0 - T_1) h^3}{\kappa \nu}, \quad \text{Pr} = \frac{\nu}{\kappa}, \quad \delta_i = \frac{C_i h^4}{\nu} \quad (i = 0, 1). \tag{5.2.2}$$

Omitting the primes, the nondimensional form of (5.2.1) reads

$$\begin{aligned}
\frac{\partial u_1}{\partial t} &= \text{Pr} \left[\Delta u_1 + \frac{2}{r_0} \frac{\partial u_2}{\partial x_1} - \frac{2}{r_0} u_1 - \delta_0 u_1 - \frac{\partial p}{\partial x_1} \right] \\
&\quad - \omega u_2 - (u \cdot \nabla) u_1 - \frac{1}{r_0} u_1 u_2, \\
\frac{\partial u_2}{\partial t} &= \text{Pr} \left[\Delta u_2 - \frac{2}{r_0} \frac{\partial u_1}{\partial x_1} - \frac{2}{r_0} u_2 - \delta_1 u_2 + RT - \frac{\partial p}{\partial x_2} \right] \\
&\quad + \omega u_1 - (u \cdot \nabla) u_2 + \frac{1}{r_0} u_1^2, \\
\frac{\partial T}{\partial t} &= \Delta T + u_2 - (u \cdot \nabla) T + Q, \\
\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} &= 0,
\end{aligned} \tag{5.2.3}$$

where $(x_1, x_2) \in M = (0, 2\pi r_0) \times (r_0, r_0 + 1)$, δ_0 and δ_1 are as in (5.2.2), $(u \cdot \nabla)$ and Δ are usual differential operators, and

$$\omega = \frac{2\Omega h^2}{\kappa}. \tag{5.2.4}$$

The problem is supplemented with the natural periodic boundary condition in the x_1 -direction, and the free-slip boundary condition on the top and bottom boundary:

$$(u, T)(x_1 + 2\pi r_0, x_2, t) = (u, T)(x_1, x_2, t), \tag{5.2.5}$$

$$\begin{cases} u_2 = 0, \frac{\partial u_1}{\partial x_2} = 0, T = \varphi(x_1) & \text{at } x_2 = r_0, \\ u_2 = 0, \frac{\partial u_1}{\partial x_2} = 0, T = 0 & \text{at } x_2 = r_0 + 1. \end{cases} \tag{5.2.6}$$

Here $\varphi(x_1)$ is the temperature deviation from the average T_0 on the equatorial surface and is periodic:

$$\int_0^{2\pi r_0} \varphi(x_1) dx_1 = 0 \quad \text{and} \quad \varphi(x_1) = \varphi(x_1 + 2\pi r_0).$$

The deviation $\varphi(x_1)$ is mainly caused by a difference in the specific heat capacities between the sea water and land.

5.2.3 Walker Circulation Under Idealized Conditions

In the idealized case, the temperature deviation φ vanishes. In this case, the study of transition of (5.2.3) is of special importance to understand the longitudinal circulation. Here we address the dynamic bifurcation of (5.2.3), the Walker cell struc-

ture of bifurcated solutions, and the convection scale under the following idealized conditions:

$$\varphi(x_1) = 0 \quad \text{for any } 0 \leq x_1 \leq 2\pi r_0, \quad Q = 0 \quad \text{in (5.2.3).} \quad (5.2.7)$$

For the problem (5.2.3) with (5.2.5) and (5.2.6), let

$$\begin{aligned} H &= \{(u, T) \in L^2(M)^3 \mid \operatorname{div} u = 0, (u, T) \text{ satisfies (5.2.5)}\}, \\ H_1 &= \{(u, T) \in H^2(M)^3 \cap H \mid (u, T) \text{ satisfies (5.2.5) and (5.2.6)}\}. \end{aligned}$$

Then, define the operators $L_R = A + B_R$ and $G : H_1 \rightarrow H$ by

$$\begin{aligned} A\psi &= P \left(\operatorname{Pr} \left(\Delta u_1 + \frac{2}{r_0} \frac{\partial u_2}{\partial x_1} - \delta'_0 u_1 \right), \operatorname{Pr} \left(\Delta u_2 - \frac{2}{r_0} \frac{\partial u_1}{\partial x_1} - \delta'_1 u_2 \right), \Delta T \right), \\ B_R\psi &= P(-\omega u_2, \omega u_1 + \operatorname{Pr} RT, u_2), \\ G(\psi) &= -P \left((u \cdot \nabla) u_1 + \frac{u_1 u_2}{r_0}, (u \cdot \nabla) u_2 - \frac{u_1^2}{r_0}, (u \cdot \nabla) T \right), \end{aligned} \quad (5.2.8)$$

where $\psi = (u, T) \in H_1$, $P : L^2(M)^3 \rightarrow H$ is the Leray Projection, and

$$\delta'_i = \frac{2}{r_0^2} + \delta_i \quad (i = 0, 1).$$

Then the problem (5.2.3) with (5.2.5) and (5.2.6) is equivalent to

$$\frac{d\psi}{dt} = L_R\psi + G(\psi). \quad (5.2.9)$$

Consider the eigenvalue problem $L_R\psi = \beta(R)\psi$, which is equivalent to

$$\begin{aligned} \Delta u_1 + \frac{2}{r_0} \frac{\partial u_2}{\partial x_1} - \delta'_0 u_1 - \operatorname{Pr}^{-1} \omega u_1 - \frac{\partial p}{\partial x_1} &= \operatorname{Pr}^{-1} \beta u_1, \\ \Delta u_2 - \frac{2}{r_0} \frac{\partial u_1}{\partial x_1} - \delta'_1 u_2 + \operatorname{Pr}^{-1} \omega u_2 + RT - \frac{\partial p}{\partial x_2} &= \operatorname{Pr}^{-1} \beta u_2, \\ \Delta T + u_2 &= \beta T, \\ \operatorname{div} u &= 0, \end{aligned} \quad (5.2.10)$$

with the boundary conditions (5.2.5) and (5.2.6). We shall see later that these equations (5.2.10) are symmetric, which implies, in particular, that all eigenvalues $\beta_j(R)$ are real, and there exists a number R_c , called the first critical Rayleigh number, such that

$$\beta_i(R) \begin{cases} < 0 & \text{if } R < R_c, \\ = 0 & \text{if } R = R_c, \\ > 0 & \text{if } R > R_c \end{cases} \quad \text{for } i = 1, 2, \quad (5.2.11)$$

$$\beta_j(R_c) < 0 \quad \text{for } j > 2. \quad (5.2.12)$$

The following theorem provides a theoretical basis to understand the equatorial Walker circulation.

Theorem 5.2.1 Under the idealized condition (5.2.7), the problem (5.2.3) with (5.2.5) and (5.2.6) undergoes a Type-I transition at the critical Rayleigh number $R = R_c$. More precisely, the following statements hold true:

- (1) When the Rayleigh number $R \leq R_c$, the equilibrium solution $(u, T) = 0$ is globally stable in the phase space H .
- (2) When $R_c < R < R_c + \delta$ for some $\delta > 0$, this problem bifurcates from $((u, T), R) = (0, R_c)$ to an attractor $\mathcal{A}_R = S^1$, consisting entirely of steady-state solutions, which attracts $H \setminus \Gamma$, where Γ is the stable manifold of $(u, T) = 0$ with codimension two.
- (3) For each steady-state solution $\psi_R = (u_R, T_R) \in \mathcal{A}_R$, u_R is topologically equivalent to the structure as shown in Fig. 5.4.
- (4) For any initial value $\psi_0 = (u_0, T_0) \in H \setminus (\Gamma \cup E)$, there exists a time $t_0 \geq 0$ such that for any $t > t_0$ the velocity field $u(t, \psi_0)$ is topologically equivalent to the structure as shown in Fig. 5.5 either (a) or (b), where $\psi = (u(t, \psi_0), T(t, \psi_0))$ is the solution of the problem with $\psi(0) = \psi_0$, and

$$E = \{(u, T) \in H \mid \int_{r_0}^{r_0+1} u_1 dx_2 = 0\}.$$

Fig. 5.4 The cell structure of the steady-state solutions in the bifurcated attractor \mathcal{A}_R

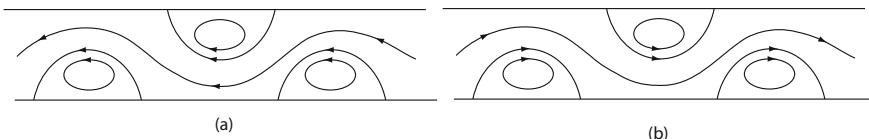
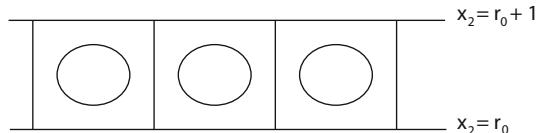


Fig. 5.5 The Walker cell structure with the cells separated by a cross channel flow: (a) a westbound cross channel flow, and (b) an eastbound cross channel flow. This cross-channel flow pattern has the same topological structure as the Walker circulation over the tropics and the Branstator–Kushnir waves in the atmospheric dynamics (Branstator, 1987; Kushnir, 1987)

A few remarks are now in order:

FIRST, in a more realized case where $T = T_0 + \varphi(x_1)$ at $x_2 = r_0$, the temperature deviation $\varphi(x_1)$ is small in comparison with the average temperature gradient $T_0 - T_1$, and $Q \neq 0$ is small as well. Therefore, the more realistic case can be considered as a perturbation of the idealized case.

SECOND, mathematically, the conclusion that $\mathcal{A}_R = S^1$ consisting of steady-state solutions is derived from the invariance of (5.2.3) under a translation $x_1 \rightarrow x_1 + \alpha$. It implies that for any two solutions $\psi_1, \psi_2 \in \mathcal{A}_R$, they are the same up to a phase angle α in the longitude, namely

$$\psi_2(x_1, x_2) = \psi_1(x_1 + \alpha, x_2) \quad \text{for some } \alpha \in [0, 2\pi r_0].$$

For the idealized case, this conclusion is natural as the equator is treated as homogeneous.

In addition, Assertion (2) amounts to saying that as the initial value ψ_0 varies, i.e., an external perturbation is applied, the rolls of the Walker circulation will be translated by a phase angle α . This behavior is termed as the translation oscillation, which can be used to explain the ENSO phenomenon for the idealized circumstance.

THIRD, when a deviation $\varphi(x_1)$ is present and $Q \neq 0$ in (5.2.3), the translation symmetry is broken, and consequently, the bifurcated attractor \mathcal{A}_R consists of isolated equilibrium solutions instead of a circle S^1 . Thus, the mechanism of the ENSO will be explained as state exchanges between these equilibrium solutions in \mathcal{A}_R , which are metastable. We shall address this point in detail later.

FOURTH, by the structural stability theorems in Ma and Wang (2005d), we see from Assertion (4) that the roll structures illustrated in Fig. 5.5a, b are structurally stable. Hence, for a deviation perturbation $\varphi(x_1)$, these structures are not destroyed, providing a realistic characterization of the Walker circulation.

FIFTH, in Assertion (3), these rolls in the bifurcated solutions are closed. However, under some external perturbation, a cross channel flow appears which separates the rolls apart, and globally transports heat between them.

SIXTH, using the classical methods as in Ma and Wang (2007a, 2005b), Chandrasekhar (1981), Drazin and Reid (1981), we can show that the critical Rayleigh number for the revised equation (5.2.10) with (5.2.5) and (5.2.6) is given by

$$R_c = \min_{\alpha^2} \left[(\alpha^2 + \pi^2)\delta'_1 + \frac{(\alpha^2 + \pi^2)^3 + \pi^2(\alpha^2 + \pi^2)\delta'_0}{\alpha^2} \right]. \quad (5.2.13)$$

Let $g(x) = (x + \pi^2)\delta'_1 + ((x + \pi^2)^3 + \pi^2(x + \pi^2)\delta'_0)/x$. Then, $x_c = \alpha_c^2$ satisfies $g'(x) = 0$. Thus we get

$$\delta'_1 \alpha_c^4 - (\alpha_c^2 + \pi^2)^3 - \pi^2(\alpha_c^2 + \pi^2)\delta'_0 + 3\alpha_c^2(\alpha_c^2 + \pi^2)^2 + \pi^2\alpha_c^2\delta'_0 = 0. \quad (5.2.14)$$

We assume that $\delta'_1 > \delta'_0$ for $h \neq 0$. By $\alpha_c^2 = \pi^2/L_c^2$, when h is large, $\delta'_1 \gg 1$ and $\delta'_1 \gg \delta'_0$. Here

$$\delta'_i = \delta_i + \frac{2}{r_0^2}.$$

We infer from (5.2.13) and (5.2.14) that

$$\alpha_c^4 \cong \frac{\pi^4(\pi^2 + \delta'_0)}{\delta'_1}, \quad L_c^2 \cong \frac{(\delta'_1)^{1/2}}{(\pi^2 + \delta'_0)^{1/2}}, \quad R_c \cong \pi^2 \delta'_1 + \pi^2 \delta'_0 L_c^2, \quad (5.2.15)$$

which are useful in studying large-scale convections, as shown in the next section.

Proof of Theorem 5.2.1. We proceed in the following four steps.

STEP 1. We show that equations (5.2.3) have an equivalent form as the classical 2D Bénard problem.

Since the velocity field u defined on $M = S_{r_0}^1 \times (r_0, r_0 + 1)$ is divergence-free, there exists a stream function φ such that

$$u_1 = \frac{\partial \varphi}{\partial x_2}, \quad u_2 = -\frac{\partial \varphi}{\partial x_1},$$

satisfying the given boundary conditions. Therefore, the following two vector fields

$$(-\omega u_2, \omega u_1) = \omega \nabla \varphi, \quad \left(\frac{2}{r_0} \frac{\partial u_2}{\partial x_1}, -\frac{2}{r_0} \frac{\partial u_1}{\partial x_1} \right) = -\frac{2}{r_0} \nabla \frac{\partial \varphi}{\partial x_1}, \quad (5.2.16)$$

are gradient fields, which can be balanced by ∇p in (5.2.3). Hence, (5.2.3) are equivalent to

$$\begin{aligned} \frac{\partial u_1}{\partial t} + (u \cdot \nabla) u_1 + \frac{u_1 u_2}{r_0} &= \text{Pr} \left[\Delta u_1 - \delta'_0 u_1 - \frac{\partial p}{\partial x_1} \right], \\ \frac{\partial u_2}{\partial t} + (u \cdot \nabla) u_2 - \frac{u_1^2}{r_0} &= \text{Pr} \left[\Delta u_2 - \delta'_1 u_2 + RT - \frac{\partial p}{\partial x_2} \right], \\ \frac{\partial T}{\partial t} + (u \cdot \nabla) T &= \Delta T + u_2, \\ \text{div } u &= 0. \end{aligned} \quad (5.2.17)$$

Therefore, equation (5.2.9) is also an abstract form of (5.2.17). Thus the eigenvalue problem (5.2.10) is symmetric. It is clear that the operator G defined in (5.2.8) is orthogonal. Hence Theorem 4.1.1 for the Rayleigh–Bénard convection proved in Ma and Wang (2004b) is also valid for the problem (5.2.3) with (5.2.5), (5.2.6). Thus, this transition of the problem is Type-I, and Assertion (1) is proved.

STEP 2. PROOF OF ASSERTION (2). We only need to verify that the attractor \mathcal{A}_R is homeomorphic to a circle, and consisting of singular points of (5.2.9). This proof is similar to that of Theorem 4.2.12.

We consider the first eigenvectors of (5.2.10) at $R = R_c$. By (5.2.16), equations (5.2.10) at $R = R_c$ are equivalent to the form

$$\begin{aligned}\Delta u_1 - \delta'_0 u_1 - \frac{\partial p}{\partial x_1} &= 0, \\ \Delta u_2 - \delta'_1 u_2 + R_c T - \frac{\partial p}{\partial x_2} &= 0, \\ \Delta T + u_2 &= 0, \\ \operatorname{div} u &= 0,\end{aligned}\tag{5.2.18}$$

with the boundary conditions (5.2.5) and (5.2.6)

Similar to the eigenvalue problem (4.2.57), we see that the multiplicity of R_c in (5.2.18) is $m = 2$, and the corresponding eigenvectors are given by

$$\psi_1 = (u_1, u_2, T) = \left(-\sin \frac{kx_1}{r_0} H'(x_2), \frac{k}{r_0} \cos \frac{kx_1}{r_0} H(x_2), \cos \frac{kx_1}{r_0} \Phi(x_2) \right), \tag{5.2.19}$$

$$\tilde{\psi}_1 = (\tilde{u}_1, \tilde{u}_2, \tilde{T}) = \left(\cos \frac{kx_1}{r_0} H'(x_2), \frac{k}{r_0} \sin \frac{kx_1}{r_0} H(x_2), \sin \frac{kx_1}{r_0} \Phi(x_2) \right), \tag{5.2.20}$$

where the functions $H(x_2)$ and $\Phi(x_2)$ satisfy that

$$\begin{cases} (D^4 - (\alpha^2 + \alpha_0^2)D^2 + \alpha^2 \alpha_1^2)H = \alpha R_c \Phi, \\ -(D^2 - \alpha^2)\Phi = \alpha H, \end{cases} \tag{5.2.21}$$

$$H = H'' = 0, \quad \Phi = 0 \quad \text{at } x_2 = r_0, r_0 + 1, \tag{5.2.22}$$

where $D = d/dx_2$, $\alpha = k/r_0$, $\alpha_i^2 = \alpha^2 + \delta'_i$ ($i = 0, 1$).

Since (5.2.21) with (5.2.22) are symmetric, the number defined by (4.2.71) and (4.2.72) is given by

$$\gamma = \frac{1}{\|\psi_1\|^2} (G(\Psi, \psi_1) + G(\psi_1, \Psi), \psi_1), \tag{5.2.23}$$

$$-L_{R_c} \Psi = G(\psi_1, \psi_1), \tag{5.2.24}$$

L_R is defined by (5.2.8), and $G(\tilde{\psi}, \psi)$ defined by

$$G(\tilde{\psi}, \psi) = P \left((\tilde{u} \cdot \nabla) u_1 + \frac{\tilde{u}_1 u_2}{r_0}, (\tilde{u} \cdot \nabla) u_2 + \frac{\tilde{u}_1 u_1}{r_0}, (\tilde{u} \cdot \nabla) T \right),$$

for any $\tilde{\psi} = (\tilde{u}, \tilde{T}), \psi = (u, T) \in H_1$. We derive from (5.2.23) and (5.2.24) that

$$\gamma = -\frac{1}{\|\psi_1\|^2} (G(\psi_1, \psi_1), \Psi) = -\frac{1}{\|\psi_1\|^2} (-L_{R_c} \Psi, \Psi).$$

Due to (5.2.16), $-L_R$ is a symmetric sectorial operator. Thus, we have

$$\gamma = -\frac{1}{\|\psi_1\|^2} ((-L_{R_c})^{1/2} \Psi, (-L_{R_c})^{1/2} \Psi) < 0.$$

Then, similar to Theorem 4.2.12, Assertion (2) follows.

STEP 3. PROOF OF ASSERTION (3). By Step 1 and the attractor bifurcation theorem, each steady-state solution $\varphi \in \mathcal{A}_R$ can be expressed as

$$\varphi = \frac{|\beta_1(R)|}{\gamma^{1/2}}(x\psi_1 + y\tilde{\psi}_1) + o(\beta_1(R)), \quad (5.2.25)$$

where $\psi_1, \tilde{\psi}_1$ are given by (5.2.19) and (5.2.20), $x^2 + y^2 = 1$, and $\beta_1(R)$ is the first eigenvalue of (5.2.10) satisfying (5.2.11).

Let $\varphi = (e, T)$. Then, by (5.2.25) the velocity field e is written as

$$e = \left(r(R) \cos \frac{k}{r_0} (x_1 + \theta) H'(x_2), \frac{k}{r_0} r(R) \sin \frac{k}{r_0} (x_1 + \theta) H(x_2) \right) + o(r(R)),$$

where $r(R) = |\beta_1(R)/\alpha|^{1/2}$, $(x, y) = (\cos \theta, \sin \theta)$.

By (5.2.21) and (5.2.22), $H(x_2) = \sin j\pi(x_2 - r_0)$, $j \in \mathbb{Z}$. But, for the first eigenvectors ψ_1 and $\tilde{\psi}_1$, we have $H(x_2) = \sin \pi(x_2 - r_0)$. Thus, e is expressed as

$$e = e_0 + o(r(R)), \quad (5.2.26)$$

$$e_0 = \left(\pi r(R) \cos \frac{k}{r_0} (x_1 + \theta) \cos \pi(x_2 - r_0), \frac{k}{r_0} r(R) \sin \frac{k}{r_0} (x_1 + \theta) \sin \pi(x_2 - r_0) \right).$$

It is easy to check that e_0 is regular, therefore e is also regular as $R_c < R < R_c + \delta$ for some $\delta > 0$. Obviously, the vector field e_0 is topologically equivalent to the structure as shown in Fig. 5.4.

By the structural stability theorem, Theorem 2.12 in Ma and Wang (2005d), the vector field e_0 is not structurally stable in H_1 , because the boundary saddle points on $x_2 = r_0$ are connected to saddle points on $x_2 = r_0 + 1$, a different connected component. However, e_0 is structurally stable in the space

$$\tilde{H} = \{(u, T) \in H \mid \int_0^{2\pi r_0} \int_{r_0}^{r_0+1} u dx = 0\}.$$

To see this, we know that if $u \in \tilde{H}$, then $u = (u_1, u_2) \neq \text{constant}$, and u_1 has the Fourier expansion

$$u_1 = \sum_{k,j} (x_{kj} \cos \frac{kx_1}{r_0} + y_{kj} \sin \frac{kx_1}{r_0}) \cos j\pi(x_2 - r_0).$$

It follows that

$$\int_{r_0}^{r_0+1} u_1 dx_2 = 0 \quad \forall u = (u_1, u_2) \in \tilde{H}. \quad (5.2.27)$$

According to the Connection Lemma, Lemma 2.3.1 in Ma and Wang (2005d), we derive from (5.2.27) that a vector field $u \in \tilde{H}$ is structurally stable if and only if

u is regular, and all interior saddle points of u are self-connected. Therefore, e_0 is structurally stable in \tilde{H} .

If we can prove that the vector field e given by (5.2.26) is in \tilde{H} , then, as $R_c < R < R_c + \delta$, e is topologically equivalent to e_0 . Hence, to prove Assertion (3), it suffices to verify that $e \in \tilde{H}$.

Obviously, \tilde{H} is an invariant space for the operator $L_R + G$ defined by (5.2.8), and the orthogonal complementary \tilde{H}^\perp of \tilde{H} in H is

$$\tilde{H}^\perp = \{(u, T) \in H \mid u = (c, 0), T = 0, c \in \mathbb{R}^1\}.$$

It is ready to prove that all steady-state solutions of (5.2.9) are in \tilde{H} . Thus, Assertion (3) is proved.

STEP 4. PROOF OF ASSERTION (4). For any initial value $\psi_0 = (u_0, T_0) \in H \setminus \tilde{H}$, ψ_0 can be written as

$$\psi_0 = c_0 + \Phi_0, \Phi_0 \in \tilde{H}, \quad c_0 = (\alpha_0, 0, 0) \in \tilde{H}^\perp, \quad \alpha_0 \neq 0.$$

Since $G(\psi) \in \tilde{H}$, $\forall \psi \in H$, the solution $\psi(t, \psi_0)$ of (5.2.9) with $\psi(0, \psi_0) = \psi_0$ must be in the form

$$\psi(t, \psi_0) = C_0 e^{-\delta'_0 t} + \Phi(t, \psi_0), \quad \Phi \in \tilde{H}.$$

In addition, we know that for each $\psi_0 \in H/\Gamma$ there is a steady-state solution $\varphi = (e, T) \in \mathcal{A}_R$ such that

$$\lim_{t \rightarrow \infty} \|\psi(t, \psi_0) - \varphi\|_{C^r} = 0 \quad \text{for any } r \geq 1.$$

Then, for any $\psi_0 \in H \setminus (\tilde{H} \cup \Gamma)$, it follows that there exists a time $t_0 > 0$ such that for any $t > t_0$, the velocity field $u(t, \psi_0)$ of $\psi(t_0, \psi_0)$ is topologically equivalent to the following vector field

$$\tilde{u} = e + c_0 e^{-\delta'_0 t} = e + (\alpha_0 e^{-\delta'_0 t}, 0, 0),$$

where e is as in (5.2.26).

The vector field e has the roll structure as shown in Fig. 5.4, and $c_0 e^{-\delta'_0 t}$ is a parallel channel flow. Hence, it is easy to show that $\tilde{u} = e + c_0 e^{-\delta'_0 t}$ is topologically equivalent to the structure as shown in Fig. 5.5a as $\alpha_0 < 0$, and to the structure as shown in Fig. 5.5b as $\alpha_0 > 0$, for $t > t_0$ large. Assertion (4) is proved. \square

5.2.4 Walker Circulation Under Natural Conditions

Physical Parameters and Effects of the Turbulent Friction Terms It is known that air properties vary with the temperature and pressure. Table 5.1 lists some common properties of air: density, kinematic viscosity, thermal diffusivity, expansion coefficient, and the Prandtl number for temperatures between -100 and 100°C .

Table 5.1 Some physical parameters

Temperature <i>t</i> (°C)	Density <i>ρ</i> (kg/m ³)	Kinematic viscosity <i>ν</i> (m ² /s) × 10 ⁻⁶	Thermal diffusivity <i>κ</i> (m ² /s) × 10 ⁻⁶	Expansion coefficient <i>α</i> (1/k) × 10 ⁻³	Prandtl number Pr
-100	1.980	5.95	8.4	5.82	0.74
-50	1.534	9.55	13.17	4.51	0.725
0	1.293	13.30	18.60	3.67	0.715
20	1.205	15.11	21.19	3.40	0.713
40	1.127	16.97	23.87	3.20	0.711
60	1.067	18.90	26.66	3.00	0.709
80	1.000	20.94	29.58	2.83	0.708
100	0.946	23.06	32.8	2.68	0.703

The average temperature over the equator is about 20–30 °C. Based on the data in Table 5.1, we take a set of physical parameters of air for the Walker circulation as follows:

$$\nu = 1.6 \times 10^{-5} \text{ m}^2/\text{s}, \quad \kappa = 2.25 \times 10^{-5} \text{ m}^2/\text{s}, \quad \alpha_T = 3.3 \times 10^{-3} / \text{°C}, \quad \text{Pr} = 0.71.$$

The height *h* of the troposphere is taken to be *h* = 8 × 10³ m. Thus, the Rayleigh number for the equatorial atmosphere is

$$R = \frac{g\alpha T(T_0 - T_1)}{\kappa\nu} h^3 = 3.6 \times 10^{19} (T_0 - T_1) / \text{°C}. \quad (5.2.28)$$

In the classical theory, we have the critical Rayleigh number and the convection scale:

$$R_c = \min_{\alpha^2} \frac{(\alpha^2 + \pi^2)^3}{\alpha^2} = \frac{27}{4} \pi^4 = 675, \quad L_c^2 = \pi^2 / \alpha_c^2 = 2. \quad (5.2.29)$$

Namely, the diameter of a convective roll is *d* = *L_c* × *h* = 20 km. However, based on atmospheric observations, there are six cells over the tropics, and it shows that the real convective scale is about

$$L = \frac{a\pi}{3} = 6600 \text{ km}. \quad (5.2.30)$$

As a comparison, the values (5.2.29) from the classical theory are too small to match the realistic data given in (5.2.28) and (5.2.30).

We use (4.5.16) to discuss this problem. With the values in (5.2.28) and (5.2.30), since *r₀* is large, we take

$$\delta'_1 \simeq \delta_1 = 2.7 \times 10^{20}, \quad \delta'_0 \simeq \delta_0 = 3.5 \times 10^8. \quad (5.2.31)$$

In comparison with (5.2.2), for air, the constants C_0 and C_1 take the following values

$$C_0 = 1.37 \times 10^{-12} \text{ m}^{-2} \cdot \text{s}^{-1}, \quad C_1 = 1.05 \text{ m}^{-2} \cdot \text{s}^{-1}.$$

Then, the convective scale L_c in (4.5.16) takes the value

$$L_c = [\delta_1/\delta_0]^{1/4} h = 7500 \text{ km}, \quad (5.2.32)$$

which is slightly bigger than the realistic value $L \approx 6300 \text{ km}$. And, the critical Rayleigh number in (4.5.16) takes

$$R_c \cong \pi^2 \delta_1 + \pi^2 \delta_0 L_c^2 \cong 2.7 \times 10^{21}.$$

Thus, the critical temperature gradient is

$$\Delta T_c \cong 60^\circ\text{C}, \quad (5.2.33)$$

which is realistic, although it is slightly smaller than the realistic temperature gradient over the equator.

However, with the classical theory with no friction terms, the critical values would be $L_c \cong 20 \text{ km}$ and $\Delta T_c \cong 10^{-17}^\circ\text{C}$, which are certainly unrealistic.

The relation between L_c and the wave number k in the first eigenvectors (5.2.19) and (5.2.20) is

$$k = \text{the closest even integer to } \frac{2\pi r_0}{L_c},$$

and in this case, we have $k = 6$.

Transition in Natural Conditions We now return to the natural conditions:

$$\varphi(x_1) \not\equiv 0, \quad Q \neq 0.$$

In this case, equation (5.2.3) admits a steady-state solution

$$\tilde{\psi} = (V, J), \quad (V = (V_1, V_2)). \quad (5.2.34)$$

Consider the deviation from this basic state:

$$u \rightarrow u + V, \quad T \rightarrow T + J.$$

Then (5.2.3) becomes

$$\begin{aligned}
\frac{\partial u_1}{\partial t} + (u \cdot \nabla) u_1 + \frac{u_1 u_2}{r_0} &= \text{Pr} [\Delta u_1 + \frac{2}{r_0} \frac{\partial u_2}{\partial x_1} - \delta'_0 u_1 - \frac{\partial p}{\partial x_1}] \\
&\quad - \omega u_2 - (V \cdot \nabla) u_1 - (u \cdot \nabla) V_1 - \frac{V_1}{r_0} u_2 - \frac{V_2}{r_0} u_1, \\
\frac{\partial u_2}{\partial t} + (u \cdot \nabla) u_2 - \frac{u_1^2}{r_0} &= \text{Pr} [\Delta u_2 - \frac{2}{r_0} \frac{\partial u_1}{\partial x_1} - \delta'_1 u_2 + RT - \frac{\partial p}{\partial x_2}] \\
&\quad + \omega u_1 - (V \cdot \nabla) u_2 - (u \cdot \nabla) V_2 + \frac{2V_1}{r_0} u_1, \\
\frac{\partial T}{\partial t} + (u \cdot \nabla) T &= \Delta T + u_2 - (u \cdot \nabla) J - (V \cdot \nabla) T, \\
\text{div } u &= 0.
\end{aligned} \tag{5.2.35}$$

The boundary conditions are the free-free boundary conditions given by

$$\begin{aligned}
(u, T) &\text{ is periodic in } x_1\text{-direction,} \\
T = 0, u_2 = 0, \frac{\partial u_1}{\partial x_2} &= 0 \text{ at } x_2 = r_0, r_0 + 1.
\end{aligned} \tag{5.2.36}$$

Let the operators $L_R = A + B_R$ and $G : H_1 \rightarrow H$ be as in (5.2.8), and $\mathcal{L}^\varepsilon : H_1 \rightarrow H$ be defined by

$$\begin{aligned}
\mathcal{L}^\varepsilon \psi &= P(-(V \cdot \nabla) u_1 - (u \cdot \nabla) V_1 - \frac{V_1}{r_0} u_2 - \frac{V_2}{r_0} u_1, \\
&\quad -(V \cdot \nabla) u_2 - (u \cdot \nabla) V_2 + \frac{2V_1}{r_0} u_1, -(u \cdot \nabla) J - (V \cdot \nabla) T).
\end{aligned}$$

Then, the problem (5.2.35) and (5.2.36) is equivalent to the abstract form

$$\frac{d\psi}{dt} = L_R \psi + \mathcal{L}^\varepsilon \psi + G(\psi). \tag{5.2.37}$$

Consider the eigenvalue problem

$$L_R \psi + \mathcal{L}^\varepsilon \psi = \beta^\varepsilon(R) \psi. \tag{5.2.38}$$

It is known that with $\Delta T = T_0 - T_1 \cong 100^\circ\text{C}$, $|\varphi(x_1)|$ and Q are small. Hence, the steady-state solution (V, J) is also small:

$$\|(V, J)\|_{L^2} = \varepsilon \ll 1.$$

Thus, (5.2.35) is a perturbed equation of (5.2.3).

Since perturbation terms involving (V, J) are not invariant under the zonal translation (in the x_1 -direction), for general small functions $\varphi(x_1) \neq 0$, the first eigenvalues of (5.2.35) are (real or complex) simple, and by the perturbation theorems in Ma and Wang (2005b), all eigenvalues of linearized equation of (5.2.35) satisfy the following principle of exchange of stability (PES):

$$\operatorname{Re}\beta_i^\varepsilon(R) \begin{cases} < 0 & \text{for } R < R_c^\varepsilon, \\ = 0 & \text{for } R = R_c^\varepsilon, \\ > 0 & \text{for } R > R_c^\varepsilon, \end{cases} \quad \text{for any } 1 \leq i \leq m,$$

$$\operatorname{Re}\beta_j^\varepsilon(R_c^\varepsilon) < 0 \quad \text{for any } j \geq m+1,$$

where $m = 1$ as $\beta_1^\varepsilon(R)$ is real, $m = 2$ as $\beta_1^\varepsilon(R)$ is complex near R_c^ε , and R_c^ε is the critical Rayleigh number of perturbed system (5.2.35).

When $\beta_1^\varepsilon(R)$ near $R = R_c^\varepsilon$ is real, by Theorem 2.6.3, we derive immediately the following transition theorem.

Theorem 5.2.2 *Let $\beta_1^\varepsilon(R)$ near $R = R_c^\varepsilon$ be a simple real eigenvalue. Then the system (5.2.35) has a transition at $R = R_c^\varepsilon$, which is either mixed (Type-III) or continuous (Type-I), depending on the temperature deviation $\varphi(x_1)$. Moreover, we have the following assertions:*

- (1) *If the transition is Type-I, then as $R_c^\varepsilon < R < R_c^\varepsilon + \delta$ for some $\delta > 0$, the system bifurcates at R_c^ε to exactly two steady-state solutions ψ_1 and ψ_2 in H , which are attractors. In particular, space H can be decomposed into two open sets U_1, U_2 :*

$$H = \overline{U}_1 + \overline{U}_2, \quad U_1 \cap U_2 = \emptyset, \quad \psi = 0 \in \partial U_1 \cap \partial U_2,$$

such that $\psi_i \in U_i$ ($i = 1, 2$), and ψ_i attracts U_i .

- (2) *If the transition is Type-III, then there is a saddle-node bifurcation at $R = R^*$ with $R^* < R_c^\varepsilon$ such that the following statements hold true:*

- (a) *if $R^* < R < R_c^\varepsilon + \delta$ with $R \neq R_c^\varepsilon$, the system has two steady-state solutions ψ_R^+ and ψ_R^- which are attractors, as shown in Fig. 5.6, such that*

$$\psi_R^+ = 0 \quad \text{for } R^* < R < R_c^\varepsilon.$$

- (b) *There is an open set $U \subset H$ with $0 \in U$ which can be decomposed into two disjoint open sets $\overline{U} = \overline{U}_+^R + \overline{U}_-^R$ with $\psi_R^\pm \in U_\pm^R$ and ψ_R^\pm attracts U_\pm^R .*

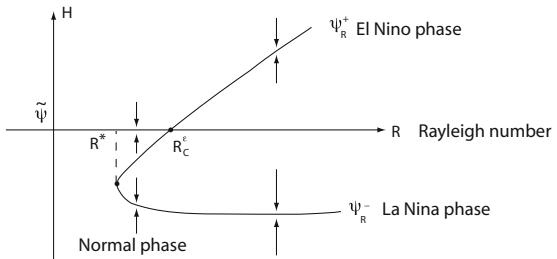
- (3) *For any initial value $\psi_0 = (u_0, T_0) \in U_-^R$ for $R > R^*$ or $\psi_0 = (u_0, T_0) \in U_+^R$ for $R > R_c^\varepsilon$, there exists a time $t_0 \geq 0$ such that for any $t > t_0$ the velocity field $u(t, \psi_0)$ is topologically equivalent to the structure as shown in Fig. 5.5 either (a) or (b), where $\psi = (u(t, \psi_0), T(t, \psi_0))$ is the solutions of the problem with $\psi(0) = \psi_0$.*

When $\beta_1^\varepsilon(R)$ is a complex eigenvalue, by Theorem 2.6.7, we obtain the following transition theorem of periodic solutions.

Theorem 5.2.3 *Let $\beta_1^\varepsilon(R)$ be complex near $R = R_c^\varepsilon$. Then the system (5.2.35) bifurcates from $(\psi, R) = (0, R_c^\varepsilon)$ to a periodic solution $\psi_R(t)$ on $R_c^\varepsilon < R$, which is an attractor, and $\psi_R(t)$ can be expressed as*

$$\psi_R(t) = A_R(\cos \rho t \psi_1 + \sin \rho t \tilde{\psi}_1) + o(|A_R|, \varepsilon), \quad (5.2.39)$$

Fig. 5.6 The transition diagram for the equatorial motion equations (5.2.3), R^* is the saddle-node bifurcation point, and R_c^ε is the first critical Rayleigh number



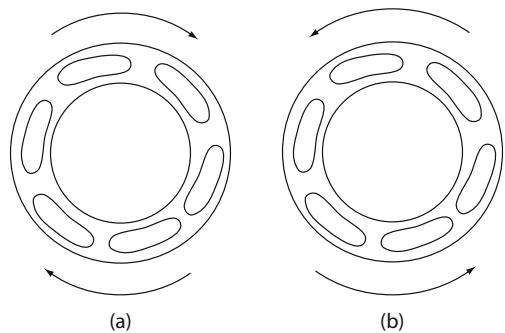
where $A_R = \alpha(R - R_c^\varepsilon)$, α and ρ are constants depending on $\varphi(x_1)$, and $\psi_1, \tilde{\psi}_1$ are first eigenfunctions of linearized equations of (5.2.3).

A few remarks are now in order.

FIRST, Theorems 5.2.1, 5.2.2, and 5.2.3 provide the possible dynamical behaviors for the zonal atmospheric circulation over the tropics. Theorem 5.2.1 describes the translation oscillation and does not represent a realistic explanation to the ENSO.

SECOND, the periodic solution (5.2.39) characterizes a roll pattern translating with a constant velocity over the equator, eastward or westward, as shown in Fig. 5.7a, b. The time-periodic oscillation obtained in Theorem 5.2.3 does not represent the typical oscillation in the ENSO phenomena, as the Walker circulation does not obey this zonal translational oscillation.

Fig. 5.7 Time-periodic translation of the Walker circulation pattern; (a) an eastward translation, (b) a westward translation



THIRD, Theorem 5.2.2 characterizes the oscillation between metastable states. The oscillation between the metastable states in Theorem 5.2.2 can be either Type-I or Type-III, depending on the number

$$b = (G(\Psi_1), \Psi_1^*),$$

where Ψ_1 is the first eigenvector of the linearized equation of (5.2.35) at $R = R_c^\varepsilon$, Ψ_1^* is the corresponding first eigenvector of the adjoint linearized problem, and G represents the nonlinear terms in (5.2.35).

Namely, if $b \neq 0$, the transition is Type-III and if $b = 0$, the transition is Type-I. From mathematical viewpoint, for almost all functions φ and $Q \neq 0$, $b \neq 0$. Hence,

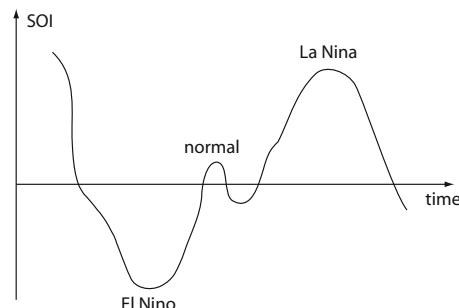
the case where $b \neq 0$, consequently the Type-III transition in Theorem 5.2.2, is generic. In other words, the Type-III transition derived in this theorem provides a correct oscillation mechanism of the ENSO between two metastable El Niño and La Niña events, and we shall explore this point of view in the next section in detail.

5.2.5 ENSO: Metastable Oscillation Theory

It is well known that the Walker cell at the equatorial Pacific ocean is closely related to the ENSO phenomenon. Its behavior is the key to the understanding of ENSO.

Southern Oscillation Phenomenon The Southern Oscillation Index (SOI) gives a simple measure of the strength and phase of the Southern Oscillation, and indicates the state of the Walker circulation. The SOI is defined by the pressure difference between Tahiti and Darwin. When the Walker circulation enters its El Niño phase, the SOI is strongly negative; when it is in the La Niña phase, the SOI is strongly positive; and in the normal state the SOI is small; see Fig. 5.8.

Fig. 5.8 SOI indicates the state of the Walker circulation



Also, the observational data of SOI as shown in Fig. 5.9 provides a more clear picture of both the spatiotemporal and random characteristics of ENSO.

El Niño and La Niña States The nonhomogeneous temperature distribution shows that the locations of the cells over the tropics are relatively fixed. This suggests that the homogeneous case as described by Theorem 5.2.1 is not a valid description of the realistic situation. In addition, the translation oscillation formation, suggested in Theorem 5.2.3, does not describe the ENSO either.

Hence the nature theory for the atmospheric circulation over the tropics associated with the Walker circulation and the ENSO is given by Theorem 5.2.2. It is known that the steady-state solutions of (5.2.3) can be written as

$$\tilde{\psi} + \psi_R^\pm, \quad (5.2.40)$$

where $\tilde{\psi} = (V, J)$ is the steady-state solution given by (5.2.34), and ψ_R^\pm are the stable equilibria of (5.2.35), derived in Theorem 5.2.2. In fact, the ENSO phenomenon

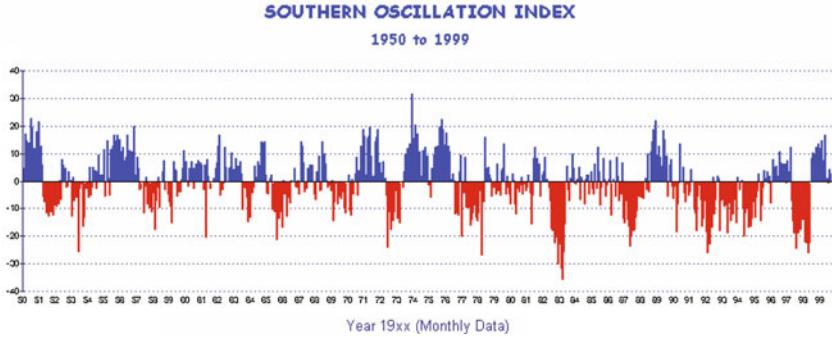


Fig. 5.9 SOI diagram, adopted from Web: BOM-Monitoring Climate-Climate variability and El Niño

can be explained as the transition between the two metastable steady-state solutions $\tilde{\psi} + \psi_R^\pm$.

Theorem 5.2.2 amounts to saying that for a mixed transition, there are two critical Rayleigh numbers R^* and R_c^ε , with $R^* < R_c^\varepsilon$, where $R^* \cong R_c \cong 2.7 \times 10^{21}$. Equation (5.2.35) has a saddle-node bifurcation at $R = R^*$, and has the mixed transition at $R = R_c^\varepsilon$, as shown in Fig. 5.6. When the Rayleigh number R is less than R^* , i.e., $R < R^*$, the system (5.2.3) has only one stable steady-state solution $\tilde{\psi} = (V, J)$ as in (5.2.34) which attracts H . When $R^* < R < R_c^\varepsilon$, this system has two stable equilibria

$$\tilde{\psi} \in U_+^R, \quad \text{and} \quad \tilde{\psi} + \psi_R^- \in U_-^R, \quad \text{for } R^* < R < R_c^\varepsilon, \quad (5.2.41)$$

where U_+^R , $U_-^R \subset H$ are basins of attraction of $\tilde{\psi}$ and $\tilde{\psi} + \psi_R^-$. And when $R_c^\varepsilon < R$, this system has also two stable equilibria

$$\tilde{\psi} + \psi_R^+ \in U_-^R \quad \text{and} \quad \tilde{\psi} + \psi_R^+ \in U_+^R \quad \text{for } R_c^\varepsilon < R, \quad (5.2.42)$$

with U_-^R and U_+^R as their basins of attraction.

Since the problem (5.2.3) with a natural boundary condition is a perturbation for the idealized condition, $\tilde{\psi} = (V, J)$ is small, and the velocity field V is almost zero. Hence the equilibrium

$$\Psi_R^+ = \begin{cases} \tilde{\psi} & \text{for } R < R_c^\varepsilon, \\ \tilde{\psi} + \psi_R^+ & \text{for } R > R_c^\varepsilon, \end{cases}$$

represents the El Niño phase, and the equilibrium

$$\Psi_R^- = \tilde{\psi} + \psi_R^-$$

represents the normal phase for $R^* < R < R_c^\varepsilon$, and the La Niña phase for $R_c^\varepsilon < R$. It is clear that both Ψ_R^\pm are metastable, with U_\pm^R as their basins of attractions respectively.

Oscillation Mechanism of ENSO The above theoretical studies suggest that ENSO is the *interplay* between two oscillation processes. The first is the oscillation between the metastable warm event (El Niño phase, represented by Ψ_R^+) and cold event (La Niña phase, represented by Ψ_R^-). The second oscillation is the oscillation of the Rayleigh number caused essentially by the some spatiotemporal oscillation of the sea surface temperature (SST) field.

Here we present a brief schematic description of the interplay between these two oscillation processes of the ENSO, based on the saddle-node transition diagram in Fig. 5.6, rigorously proved in Theorem 5.2.2.

We start with three physical conclusions:

- We observe that as R decreases, the normal and the La Niña phase Ψ_R^- weakens and its basin of attraction shrinks (to zero as R approaches to R^*).
- As R increases in the interval $R_c^\varepsilon < R$, the strength of the El Niño phase increases. As R increases in (R^*, R_c^ε) , the basin of attraction of the El Niño phase shrinks. In particular, near R_c^ε , the El Niño phase is close to the intersection of the two basins of attraction of the El Niño and La Niña phase; consequently, forcing the transition from the El Niño phase to the La Niña phase.
- Also, we see that as the El Niño event Ψ_R^+ intensifies, the SST increases, and as the normal and La Niña event Ψ_R^- intensifies, the SST decreases.

From the above three physical conclusions, we obtain the following new mechanism of the ENSO oscillation process:

- When $R^* < R < R_c^\varepsilon$, Ψ_R^- represents the normal condition, and the corresponding upwelling near Peru leads to the decreasing of the SST, and leads to R approaching R^* . On the other hand, near R^* , the basin of attraction of the normal condition Ψ_R^- shrinks, and due to the uncertainty of the initial data, the system can undergo a dynamic transition near R^* toward to the El Niño phase Ψ_R^+ .
- For the El Niño phase $\Psi_R^+ = \tilde{\psi}$ near R^* , however, the velocity V is small, and SST will increase due to the solar heating, which cannot be transported away by weak oceanic currents without upwelling. Hence the corresponding Rayleigh number R increases, and the El Niño phase intensifies. When the Rayleigh number approaches to the critical value R_c^ε , with high probability, the system undergoes a metastable transition from the El Niño phase to either the normal phase or the La Niña phase depending on the Rayleigh number R and the strength of the phase Ψ_R^- .
- With delaying effect, strong El Niño and La Niña occur at the Rayleigh number R larger than the critical number R_c^ε , and the system will repeat the process described in the above two items.

In summary, this new mechanism of ENSO as an interplay of the two processes leads to both the random and deterministic features of the ENSO, and defines a new natural *feedback* mechanism, which drives the sporadic oscillation of the ENSO. The randomness is closely related to the uncertainty/fluctuations of the initial data between the narrow basins of attractions of the corresponding metastable events, and

the deterministic feature is represented by a deterministic coupled atmospheric and oceanic model predicting the basins of attraction and the SST.

5.3 Thermohaline Ocean Circulation

Oceanic circulation is one of the key sources of internal climate variability. One important source of such variability is the thermohaline circulation (THC). Physically speaking, the buoyancy fluxes at the ocean surface give rise to gradients in temperature and salinity, which produce, in return, density gradients. These gradients are, overall, sharper in the vertical than in the horizontal and are associated therefore with an overturning or THC.

THC is the global density-driven circulation of the oceans, which is so named because it involves both heat, namely “thermo,” and salt, namely “haline.” The two attributes, temperature and salinity, together determine the density of seawater, and the gradient in density between the water masses in the oceans causes the water to flow.

THC is also called the great ocean conveyer, the ocean conveyer belt, or the global conveyer belt. The great ocean conveyer produces the greatest oceanic current on the planet. It works in a fashion similar to a conveyer belt transporting enormous volume of cold, salty water from the North Atlantic to the North Pacific, and bringing warmer, fresher water in return. Figure 5.10 gives a much simplified schematic of the great ocean conveyer and Fig. 5.11 gives a diagram of oceanic currents associated with THC.

In oceanography, the Conveyer is usually described as follows. Under and near the polar region sea ice in the North Atlantic, the warm and salty water that has been transported northward from the tropical regions is cooled to form frigid water in vast quantities, leading to denser seawater. Note that, unlike fresh water, salty water does not have a density maximum at 4°C , and gets denser as it cools all the way to its freezing point of approximatively -1.8°C . When this seawater freezes, the salt is separated (see ice contains almost no salt), increasing the salinity of the remaining, unfrozen water. This salinity makes the water denser again. The dense water then sinks into the deep basin of the sea to form the North Atlantic Deep Water (NADW), and drives today's oceanic THC.

The sinking, cold, and salty water in the North Atlantic flows very slowly and southward into the deep abyssal plains of the Atlantic. The deep water moves then through the South Atlantic around South Africa where it is split into two routes: one into the Indian Ocean and one past Australia into the Pacific. As it continues on its submarine migration, the current mixes with warmer fresh water, and slowly becomes warmer and fresher. Finally, in the North Pacific, the warmer and fresher water upwells, while a shallow-water counter-current is generated. This counter-current moves southward and westward, through the Indian Ocean, still heading west around southern Africa, then crosses through the South Atlantic, still on the surface (though it extends a kilometer and a half deep). It then moves up along

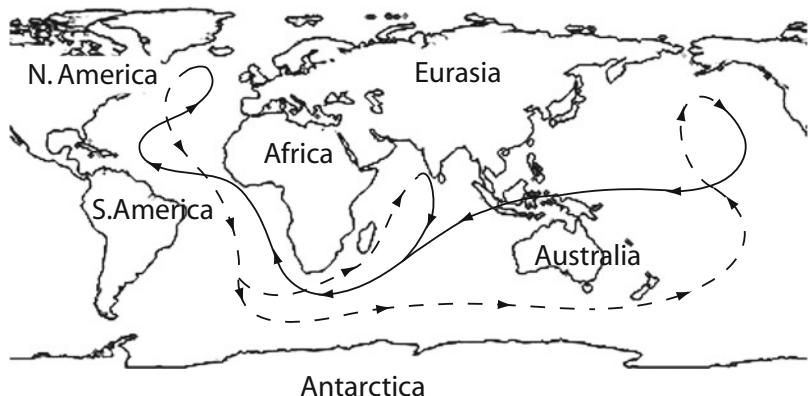
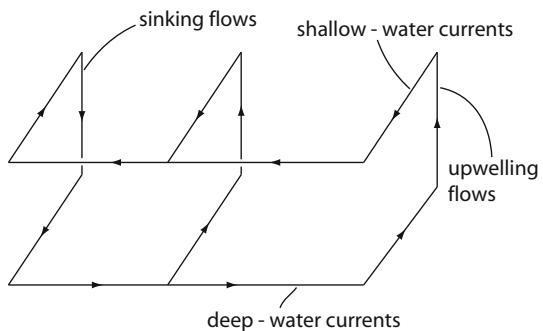


Fig. 5.10 The diagram of the thermohaline circulation. The dotted line represents deep-water currents, while the solid line represents shallow-water currents

Fig. 5.11 A schematic diagram of oceanic currents of the global thermohaline circulation



the east coast of the North America, and crosses the coast of Scandinavia. When this warmer, less salty water reaches high northern latitudes, it chills, and naturally becomes North Atlantic Deep Water, completing a circuit.

THC varies on timescales of decades or longer, and there have been extensive observational, physical, and numerical studies. We refer the interested readers to Dijkstra and Ghil (2005), Dijkstra (2000) for an extensive review of the topics; see also among others (Stommel, 1961; Rooth, 1982; Held and Suarez, 1974; Ghil, 1976; Quon and Ghil, 1992, 1995; Thual and McWilliams, 1992; Tziperman, 1997; Tziperman et al., 1994; Dijkstra and Molemaker, 1997, 1999; Dijkstra and Neelin, 1999, 2000).

The main objective of this section is to study dynamic stability and transitions in large scale oceanic circulations associated with THC. The main results presented here are first obtained by Ma and Wang (2010b).

5.3.1 Boussinesq Equations

For simplicity, we regard the region of the Pacific, the Indian Ocean, and the Atlantic, where the ocean conveyer belt occupies, as a rectangular domain: $\Omega = (0, l_1) \times (0, l_2) \times (0, h)$, where l_1 stands for the length of this conveyer belt, l_2 for the width, and h for the deep of the ocean.

The basic equations governing the thermohaline circulations are the Boussinesq type of hydrodynamic equations coupled with diffusion equations for salinity and temperature. As we see from Fig. 5.11, the dominated THC is in the zonal direction, leading to the approximation of the Coriolis force in the flat geometry by $(-2\Omega u_3, 0, 2\Omega u_1)$, which is balanced by a gradient force $\nabla\phi$. Hence, we can ignore the rotational effect of the earth, and the motion and states of the large scale ocean are governed by the following Boussinesq equations; see among others (Pedlosky, 1987; Ghil and Childress, 1987; Stern, 1960; Malkus and Veronis, 1958; Lions et al., 1992b):

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla)u &= \nu \Delta u - \frac{1}{\rho_0} (\nabla p + \rho g k), \\ \frac{\partial T}{\partial t} + (u \cdot \nabla)T &= \kappa_T \Delta T, \\ \frac{\partial S}{\partial t} + (u \cdot \nabla)S &= \kappa_S \Delta S, \\ \operatorname{div} u &= 0, \end{aligned} \tag{5.3.1}$$

where $u = (u_1, u_2, u_3)$ is the velocity field, T is the temperature function, S is the salinity, $k = (0, 0, 1)$, $\kappa_T > 0$ is the thermal diffusivity, $\kappa_S > 0$ is the salt diffusivity, $\rho_0 > 0$ is the fluid density at the lower surface $x_3 = 0$ for $(x_1, x_2, x_3) \in \Omega$, and ρ is the fluid density given by the equation of state

$$\rho = \rho_0 [1 - \alpha_T(T - T_0) + \alpha_S(S - S_0)]. \tag{5.3.2}$$

Here α_T and α_S are assumed to be positive constants. Moreover, the lower boundary ($x_3 = 0$) is maintained at a constant temperature T_0 and a constant salinity S_0 , while the upper boundary ($x_3 = h$) is maintained at a constant temperature T_1 and a constant salinity S_1 . The case where $T_0 < T_1$ and $S_0 > S_1$ often leads to thermal and solute instabilities, while each of the following three cases may give rises of instability:

$$T_0 > T_1, \quad S_0 < S_1, \tag{5.3.3}$$

$$T_0 > T_1, \quad S_0 > S_1, \tag{5.3.4}$$

$$T_0 < T_1, \quad S_0 < S_1. \tag{5.3.5}$$

The conditions (5.3.3)–(5.3.5) are satisfied respectively over some different regions of the oceans. In particular, in the high-latitude ocean regions the condition (5.3.3) or (5.3.4) is satisfied, and in the tropical ocean regions, the case (5.3.5) occurs. It is these properties (5.3.3)–(5.3.5) that give rise to the global thermoha-

line circulation, and the theoretic results derived in this section also support the viewpoint.

The trivial steady-state solution of (5.3.1)–(5.3.2) is given by

$$(u^0, T^0, S^0) = (0, T_0 - (T_0 - T_1)x_3/h, S_0 - (S_0 - S_1)x_3/h),$$

$$p^0 = p_0 - g\rho_0[x_3 + \frac{\alpha_T}{2}(T_0 - T_1)x_3^2/h - \frac{\alpha_S}{2}(S_0 - S_1)x_3^2/h],$$

where p_0 is a constant. To make the equations nondimensional, we consider the perturbation of the solution from the trivial steady-state

$$u'' = u - u^0, \quad T'' = T - T^0, \quad S'' = S - S^0, \quad p'' = p - p^0.$$

Then we set

$$x = hx', \quad t = h^2t'/\kappa_T,$$

$$u'' = \kappa_T u'/h, \quad T'' = (T_0 - T_1)T', \quad S'' = |S_0 - S_1|S', \quad p'' = \rho_0\nu\kappa_T p'/h^2.$$

Omitting the primes, the equations (5.3.1) can be written as

$$\begin{aligned} \frac{\partial u}{\partial t} &= \text{Pr}(\Delta u - \nabla p) + \text{Pr}(RT - \text{sign}(S_0 - S_1)\tilde{R}S)k - (u \cdot \nabla)u \\ \frac{\partial T}{\partial t} &= \Delta T + u_3 - (u \cdot \nabla)T, \\ \frac{\partial S}{\partial t} &= \text{Le} \Delta S + \text{sign}(S_0 - S_1)u_3 - (u \cdot \nabla)S, \\ \text{div}u &= 0, \end{aligned} \tag{5.3.6}$$

for $x = (x_1, x_2, x_3)$ in the nondimensional domain $\Omega = (0, L_1) \times (0, L_2) \times (0, 1)$, where $L_i = l_i/h$ ($i = 1, 2$), and the nondimensional parameters are the thermal Rayleigh number R , the saline Rayleigh number \tilde{R} , the Prandtl number Pr , and the Lewis number Le :

$$R = \frac{\alpha_T g(T_0 - T_1)h^3}{\kappa_T \nu}, \quad \tilde{R} = \frac{\alpha_S g(S_0 - S_1)h^3}{\kappa_T \nu}, \quad \text{Pr} = \frac{\nu}{\kappa_T}, \quad \text{Le} = \frac{\kappa_S}{\kappa_T}. \tag{5.3.7}$$

We consider the free boundary condition:

$$\begin{aligned} u_1 &= 0, \quad \frac{\partial u_2}{\partial x_1} = \frac{\partial u_3}{\partial x_1} = \frac{\partial T}{\partial x_1} = \frac{\partial S}{\partial x_1} = 0 \quad \text{at } x_1 = 0, L_1, \\ u_2 &= 0, \quad \frac{\partial u_1}{\partial x_2} = \frac{\partial u_3}{\partial x_2} = \frac{\partial T}{\partial x_2} = \frac{\partial S}{\partial x_2} = 0 \quad \text{at } x_2 = 0, L_2, \\ u_3 &= T = S = 0, \quad \frac{\partial u_1}{\partial x_3} = \frac{\partial u_2}{\partial x_3} = 0 \quad \text{at } x_3 = 0, 1. \end{aligned} \tag{5.3.8}$$

The initial value conditions are given by

$$(u, T, S) = (\tilde{u}, \tilde{T}, \tilde{S}) \quad \text{at} \quad t = 0. \quad (5.3.9)$$

For the problems (5.3.6)–(5.3.9), we set the spaces

$$\begin{aligned} H &= \{(u, T, S) \in L^2(\Omega)^5 \mid \operatorname{div} u = 0, u \cdot n|_{\partial\Omega} = 0\}, \\ H_1 &= \{(u, T, S) \in H^2(\Omega)^5 \cap H \mid (u, T, S) \text{ satisfies (5.3.8)}\}. \end{aligned}$$

Let $L_\lambda = A + B_\lambda : H_1 \rightarrow H$ and $G : H_1 \rightarrow H$ be defined by

$$\begin{aligned} A\psi &= P(\operatorname{Pr} \Delta u, \Delta T, \operatorname{Le} \Delta S), \\ B_x \psi &= P(\operatorname{Pr} (RT - \tilde{R}S \operatorname{sign}(S_0 - S_1))k, u_3, \operatorname{sign}(S_0 - S_1)u_3), \\ G(\psi) &= P((u \cdot \nabla)u, (u \cdot \nabla)T, (u \cdot \nabla)S), \end{aligned}$$

for $\psi = (u, T, S) \in H_1$, $\lambda = (R, \tilde{R})$. Here $P : L^2(\Omega, \mathbb{R}^5) \rightarrow H$ is the Leray projection. Then the problems (5.3.6)–(5.3.9) can be written as

$$\frac{d\psi}{dt} = L_\lambda \psi + G(\psi), \quad \psi(0) = \psi_0. \quad (5.3.10)$$

5.3.2 Linear Analysis

Linear Eigenvalue Problem To understand the dynamic transitions of the problem, we need to study the following eigenvalue problem for the linearized equations of (5.3.6)–(5.3.8):

$$\begin{aligned} \operatorname{Pr}(\Delta u - \nabla p) + \operatorname{Pr}(RT - \tilde{R}S \operatorname{sign}(S_0 - S_1))k &= \beta u, \\ \Delta T + u_3 &= \beta T, \\ \operatorname{Le} \Delta S + \operatorname{sign}(S_0 - S_1)u_3 &= \beta S, \\ \operatorname{div} u &= 0, \end{aligned} \quad (5.3.11)$$

supplemented with (5.3.8).

We proceed with the separation of variables. By the boundary condition (5.3.8), $\psi = (u, T, S)$ can be expressed in the following form

$$\begin{aligned} u_1 &= u_{jk}(x_3) \sin j\alpha_1 \pi x_1 \cos k\alpha_2 \pi x_2, \\ u_2 &= v_{jk}(x_3) \cos j\alpha_1 \pi x_1 \sin k\alpha_2 \pi x_2, \\ u_3 &= w_{jk}(x_3) \cos j\alpha_1 \pi x_1 \cos k\alpha_2 \pi x_2, \\ T &= T_{jk}(x_3) \cos j\alpha_1 \pi x_1 \cos k\alpha_2 \pi x_2, \\ S &= S_{jk}(x_3) \cos j\alpha_1 \pi x_1 \cos k\alpha_2 \pi x_2, \\ p &= p_{jk}(x_3) \cos j\alpha_1 \pi x_1 \cos k\alpha_2 \pi x_2, \end{aligned} \quad (5.3.12)$$

for integers j and k , where $\alpha_i = L_i^{-1}$ ($i = 1, 2$). Plugging (5.3.12) into (5.3.11), we obtain the following sets of ordinary differential equation systems:

$$\begin{aligned} \text{Pr } D_{jk} u_{jk} + \text{Pr } j\alpha_1 \pi p_{jk} &= \beta u_{jk}, \\ \text{Pr } D_{jk} v_{jk} + \text{Pr } k\alpha_2 \pi p_{jk} &= \beta v_{jk}, \\ \text{Pr } D_{jk} w_{jk} - \text{Pr } D p_{jk} + \text{Pr } R T_{jk} - \text{Pr } \tilde{R} \text{sign}(S_0 - S_1) S_{jk} &= \beta W_{jk}, \\ D_{jk} T + w_{jk} &= \beta T_{jk}, \\ \text{Le } D_{jk} S_{jk} + \text{sign}(S_0 - S_1) w_{jk} &= \beta S_{jk} \\ j\alpha_1 \pi u_{jk} + k\alpha_2 \pi v_{jk} + D w_{jk} &= 0, \\ D u_{jk} = D v_{jk} = w_{jk} = T_{jk} = S_{jk} &= 0 \quad \text{at } x_3 = 0, 1, \end{aligned} \tag{5.3.13}$$

where

$$D = \frac{d}{dx_3}, \quad D_{jk} = \frac{d^2}{dx_3^2} - \alpha_{jk}^2, \quad \alpha_{jk}^2 = \pi^2(j^2\alpha_1^2 + k^2\alpha_2^2).$$

If $w_{jk} \neq 0$, equations (5.3.13) can be reduced to a single equation

$$\begin{aligned} &[(D_{jk} - \beta)(\text{Le } D_{jk} - \beta)(\text{Pr } D_{jk} - \beta)D_{jk} \\ &+ \text{Pr } R\alpha_{jk}^2(\text{Le } D_{jk} - \beta) - \text{Pr } \tilde{R}\alpha_{jk}^2(D_{jk} - \beta)]w_{jk} = 0, \\ &w_{jk} = D^2 w_{jk} = D^4 w_{jk} = D^6 w_{jk} = 0 \quad \text{at } x_3 = 0, 1, \end{aligned} \tag{5.3.14}$$

It is clear that the solutions of (5.3.14) are sine functions

$$w_{jk} = \sin l\pi x_3 \quad \text{for } (j, k, l) \in \mathbb{Z}^2 \times \mathbb{N}. \tag{5.3.15}$$

Substituting (5.3.15) into (5.3.14), we see that the corresponding eigenvalues β of Problem (5.3.11) satisfy the cubic equation

$$\begin{aligned} &\gamma_{jkl}^2(\gamma_{jkl}^2 + \beta)(\text{Le } \gamma_{jkl}^2 + \beta)(\text{Pr } \gamma_{jkl}^2 + \beta) \\ &- \text{Pr } R\alpha_{jk}^2(\text{Le } \gamma_{jkl}^2 + \beta) + \text{Pr } \tilde{R}\alpha_{jk}^2(\gamma_{jkl}^2 + \beta) = 0, \end{aligned} \tag{5.3.16}$$

for any $(j, k, l) \in \mathbb{Z}^2 \times \mathbb{N}$, where $\gamma_{jkl}^2 = \alpha_{jk}^2 + l^2\pi^2$.

Moreover, to determine $u_{jk}(x_3)$, $v_{jk}(x_3)$, $T_{jk}(x_3)$, and $S_{jk}(x_3)$, we derive from (5.3.13) that

$$(\text{Pr } D_{jk} - \beta)u_{jk} = -\frac{j\alpha_1 \pi}{\alpha_{jk}^2}(\text{Pr } D_{jk} - \beta)D w_{jk}, \tag{5.3.17}$$

$$(\text{Pr } D_{jk} - \beta)v_{jk} = -\frac{k\alpha_2 \pi}{\alpha_{jk}^2}(\text{Pr } D_{jk} - \beta)D w_{jk}, \tag{5.3.18}$$

$$(D_{jk} - \beta)T_{jk} = -w_{jk}, \tag{5.3.19}$$

$$(\text{Le } D_{jk} - \beta)S_{jk} = -\text{sign}(S_0 - S_1)w_{jk} \tag{5.3.20}$$

With the above calculations, all eigenvalues and eigenvectors of (5.3.11) can be derived and are given in the following three groups:

1. For $(j, k, l) = (j, k, 0)$, we have

$$\begin{aligned}\beta_{jk0} &= -\text{Pr } \alpha_{jk}^2 = -\text{Pr } (j^2 \alpha_1^2 + k^2 \alpha_2^2), \\ \psi_{jk0} &= (k \alpha_2 \sin j \alpha_1 \pi x_1 \cos k \alpha_2 \pi x_2, -j \alpha_1 \cos j \alpha_1 \pi x_1 \sin k \alpha_2 \pi x_2, 0, 0, 0).\end{aligned}$$

2. For $(j, k, l) = (0, 0, l)$ with $l \neq 0$, we have

$$\begin{aligned}\beta_{00l}^1 &= -l^2 \pi^2, & \beta_{00l}^2 &= -\text{Le } l^2 \pi^2, \\ \psi_{00l}^1 &= (0, 0, 0, \sin l \pi x_3, 0), & \psi_{00l}^2 &= (0, 0, 0, 0, \sin l \pi x_3).\end{aligned}$$

3. For general (j, k, l) with $j^2 + k^2 \neq 0$ and $l \neq 0$, the solutions β of (5.3.16) are eigenvalues of (5.3.11). Let β_{jkl}^i ($1 \leq i \leq 3$) be three zeros of (5.3.16), with

$$\text{Re} \beta_{jkl}^1 \geq \text{Re} \beta_{jkl}^2 \geq \text{Re} \beta_{jkl}^3.$$

Then, by (5.3.12), (5.3.15), and (5.3.17)–(5.3.20), the eigenvector ψ_{jkl}^i corresponding to β_{jkl}^i can be written as

$$\psi_{jkl}^i = \begin{cases} u_{jkl}^i \sin j \alpha_1 \pi x_1 \cos k \alpha_2 \pi x_2 \cos l \pi x_3 \\ v_{jkl}^i \cos j \alpha_1 \pi x_1 \sin k \alpha_2 \pi x_2 \cos l \pi x_3 \\ w_{jkl}^i \cos j \alpha_1 \pi x_1 \cos k \alpha_2 \pi x_2 \sin l \pi x_3 \\ T_{jkl}^i \cos j \alpha_1 \pi x_1 \cos k \alpha_2 \pi x_2 \sin l \pi x_3 \\ S_{jkl}^i \cos j \alpha_1 \pi x_1 \cos k \alpha_2 \pi x_2 \sin l \pi x_3, \end{cases} \quad (5.3.21)$$

where

$$\begin{aligned}u_{jkl}^i &= -\frac{j \alpha_1 l \pi^2}{\alpha_{jk}^2} = -\frac{j \alpha_1 l}{j^2 \alpha_1^2 + k^2 \alpha_2^2}, & v_{jkl}^i &= -\frac{k \alpha_2 l \pi^2}{\alpha_{jk}^2} = -\frac{k \alpha_2 l}{j^2 \alpha_1^2 + k^2 \alpha_2^2}, \\ w_{jkl}^i &= 1, & T_{jkl}^i &= \frac{1}{\gamma_{jkl}^2 + \beta_{jkl}^2}, & S_{jkl}^i &= \frac{\text{sign}(S_0 - S_1)}{\text{Le } \gamma_{jkl}^2 + \beta_{jkl}^2}.\end{aligned}$$

The adjoint equations of (5.3.11) are given by

$$\begin{aligned}\text{Pr } (\Delta u^* - \nabla p^*) + (T^* + \text{sign}(S_0 - S_1) S^*) k &= \bar{\beta} u^*, \\ \Delta T^* + \text{Pr } R u_3^* &= \bar{\beta} T^*, \\ \text{Le } \Delta S^* - \text{Pr } \tilde{R} \text{sign}(S_0 - S_1) u_3^* &= \bar{\beta} S^* \\ \text{div } u^* &= 0.\end{aligned} \quad (5.3.22)$$

Thus, the conjugate eigenvector ψ_{jkl}^{i*} of (5.3.22) corresponding to β_{jkl}^i is as follows

$$\psi_{jkl}^{i*} = \begin{cases} u_{jkl}^{i*} \sin j\alpha_1 \pi x_1 \cos k\alpha_2 \pi x_2 \cos l\pi x_3 \\ v_{jkl}^{i*} \cos j\alpha_1 \pi x_1 \sin k\alpha_2 \pi x_2 \cos l\pi x_3 \\ w_{jkl}^{i*} \cos j\alpha_1 \pi x_1 \sin k\alpha_2 \pi x_2 \sin l\pi x_3 \\ T_{jkl}^{i*} \cos j\alpha_1 \pi x_1 \sin k\alpha_2 \pi x_2 \sin l\pi x_3 \\ S_{jkl}^{i*} \cos j\alpha_1 \pi x_1 \sin k\alpha_2 \pi x_3 \sin l\pi x_3, \end{cases} \quad (5.3.23)$$

where

$$u_{jkl}^{i*} = u_{jkl}^i = -\frac{j\alpha_1 l\pi^2}{\alpha_{jk}^2}, \quad v_{jkl}^{i*} = v_{jkl}^i = -\frac{k\alpha_2 l\pi^2}{\alpha_{jk}^2}, \quad w_{jkl}^{i*} = 1,$$

$$T_{jkl}^{i*} = \frac{\text{Pr } R}{\gamma_{jkl}^2 + \bar{\beta}_{jkl}^i}, \quad S_{jkl}^{i*} = \frac{-\text{sign}(S_0 - S_1)\text{Pr } \tilde{R}}{\text{Re}\tilde{\gamma}_{jkl}^2 + \bar{\beta}_{jkl}^i}.$$

Thus, all eigenvectors of (5.3.11) consist of ψ_{jk0} , ψ_{00l}^1 , ψ_{00l}^2 , and ψ_{jkl}^i ($i = 1, 2$). All conjugate eigenvectors of (5.3.22) consist of $\psi_{jk0}^* = \psi_{jk0}$, $\psi_{00l}^{1*} = \psi_{00l}^1$, $\psi_{00l}^{2*} = \psi_{00l}^2$, and ψ_{jkl}^{i*} as in (5.3.23).

Principle of Exchange of Stability We need to study the solutions β of (5.3.16), which is equivalent to the following form

$$\beta^3 + (\text{Pr} + \text{Le} + 1)\gamma_{jkl}^2\beta^2 + [(\text{Pr} + \text{Le} + \text{Pr Le})\gamma_{jkl}^4 - \text{Pr } \alpha_{jk}^2 \gamma_{jkl}^{-2} (R - \tilde{R})]\beta + \text{Pr Le } \gamma_{jkl}^6 - \text{Pr } \alpha_{jk}^2 (\text{Le } R - \tilde{R}) = 0. \quad (5.3.24)$$

First, we discuss the critical-crossing of the real eigenvalues. To this end we need to introduce a nondimensional parameter, called the R-Rayleigh number, defined by

$$\sigma = R - \text{Le}^{-1} \tilde{R} = \frac{gh^3}{\kappa_T \nu} (\alpha_T(T_0 - T_1) - \alpha_S \text{Le}^{-1} (S_0 - S_1)). \quad (5.3.25)$$

It is clear that $\beta = 0$ is a zero of (5.3.24) if and only if

$$\text{Le } \gamma_{jkl}^6 - \alpha_{jk}^2 (\text{Le } R - \tilde{R}) = 0.$$

In this case, we have

$$\sigma = \frac{\gamma_{jkl}^6}{\alpha_{jk}^2} \geq \sigma_c, \quad (5.3.26)$$

where σ_c , called the *critical R-Rayleigh number*, is defined by

$$\sigma_c = \min_{\substack{(j,k) \in \mathbb{Z}^2, j,k \geq 0, \\ j^2+k^2 \neq 0, l \geq 1}} \frac{\pi^4 (j^2 L_1^{-2} + k^2 L_2^{-2} + 1)^3}{j^2 L_1^{-2} + k^2 L_2^{-2}} = \frac{\pi^4 (j_1^2 L_1^{-2} + k_1^2 L_2^{-2} + 1)^3}{j_1^2 L_1^{-2} + k_1^2 L_2^{-2}}, \quad (5.3.27)$$

for some integer pair (j_1, k_1) such that $j_1 \geq 0, k_1 \geq 0, j_1^2 + k_1^2 \neq 0$. Here L_1 and L_2 are the nondimensional length scales in the zonal and meridional directions respectively (with the vertical length scaled to 1).

Next, we consider the critical-crossing of the complex eigenvalues. For this case, we introduce another nondimensional parameter, called the C-Rayleigh number, defined by

$$\eta = R - \frac{\text{Pr} + \text{Le}}{\text{Pr} + 1} \tilde{R} = \frac{gh^3}{\kappa_T v} \left[\alpha_T(T_0 - T_1) - \frac{(\text{Pr} + \text{Le})\alpha_S}{\text{Pr} + 1}(S_0 - S_1) \right]. \quad (5.3.28)$$

Let $i\rho_0$ ($\rho_0 \neq 0$) be a zero of (5.3.24). Then have

$$\begin{aligned} \rho_0^2 &= (\text{Pr} + \text{Le} + \text{Pr Le}) \gamma_{jkl}^4 - \text{Pr} \alpha_{jk}^2 \gamma_{jkl}^{-2} (R - \tilde{R}), \\ \rho_0^2 &= \frac{\text{Pr Le} \gamma_{jkl}^6 - \text{Pr} \alpha_{jk}^2 (\text{Le} R - \tilde{R})}{\gamma_{jkl}^2 (\text{Pr} + \text{Le} + 1)}. \end{aligned}$$

Therefore, equation (5.3.24) has a pair of purely imaginary solutions $\pm i\rho_0$ if and only if the following condition holds true

$$\begin{aligned} &(\text{Pr} + \text{Le} + 1)(\text{Pr} + \text{Le} + \text{Pr Le}) \gamma_{jkl}^6 - \text{Pr} \alpha_{jk}^2 (\text{Pr} + \text{Le} + 1)(R - \tilde{R}) \\ &= \text{Pr Le} \gamma_{jkl}^6 - \text{Pr} \alpha_{jk}^2 (\text{Le} R - \tilde{R}) > 0. \end{aligned} \quad (5.3.29)$$

It follows from (5.3.29) that

$$\eta = R - \frac{\text{Pr} + \text{Le}}{\text{Pr} + 1} \tilde{R} = \frac{(\text{Pr} + \text{Le})(1 + \text{Le})}{\text{Pr}} \frac{\gamma_{jkl}^6}{\alpha_{jk}^2}.$$

Hence we define the critical C-Rayleigh number by

$$\eta_c = \min_{(j,k,l) \in I} \frac{(\text{Pr} + \text{Le})(1 + \text{Le})}{\text{Pr}} \frac{\gamma_{jkl}^6}{\alpha_{jk}^2} = \frac{(\text{Pr} + \text{Le})(1 + \text{Le})}{\text{Pr}} \frac{\gamma_{j_1 k_1 1}^6}{\alpha_{j_1 k_1}^2}, \quad (5.3.30)$$

for the same integer pair (j_1, k_1) as in (5.3.27).

Definition 5.3.1 Let (j_1, k_1) satisfy (5.3.27). We call σ_c (resp. η_c) the first critical Rayleigh number if for all eigenvalues β of (5.3.11) we have

$$\text{Re}\beta(\sigma_c) \leq 0 \quad (\text{resp. } \text{Re}\beta(\eta_c) \leq 0).$$

To determine the first critical Rayleigh number for σ_c and η_c , we introduce another nondimensional parameter K :

$$K = \text{Le}^2 \left(1 + \frac{1}{\text{Pr}} \right) \sigma_c - (1 - \text{Le}) \tilde{R}. \quad (5.3.31)$$

Theorem 5.3.2 Let $\text{Le} \neq 1$, and (j_1, k_1) satisfy (5.3.27), and $\gamma^2 = \gamma_{j_1 k_1 1}^2$, $\alpha^2 = \alpha_{j_1 k_1}^2$.

(1) If $K > 0$ where K is defined by (5.3.31), then the number σ_c is the first critical Rayleigh number, and

$$\beta_{j_1 k_1 1}^1 = \begin{cases} < 0 & \text{if } \sigma < \sigma_c, \\ = 0 & \text{if } \sigma = \sigma_c, \\ > 0 & \text{if } \sigma > \sigma_c, \end{cases} \quad (5.3.32)$$

$$\operatorname{Re}\beta_{j_1 k_1 l}^r(\sigma_c) < 0 \quad \text{for all } (j, k, l) \text{ not satisfying (5.3.27).} \quad (5.3.33)$$

(2) If $K < 0$, then the number η_c is the first critical Rayleigh number, and

$$\operatorname{Re}\beta_{j_1 k_1 1}^1 = \operatorname{Re}\beta_{j_1 k_1 1}^2 = \begin{cases} < 0 & \text{if } \eta < \eta_c, \\ = 0 & \text{if } \eta = \eta_c, \\ > 0 & \text{if } \eta > \eta_c, \end{cases} \quad (5.3.34)$$

$$\operatorname{Re}\beta_{j_1 k_1 l}^r(\eta_c) < 0 \quad \text{for } (j, k, l) \text{ not satisfying (5.3.27).} \quad (5.3.35)$$

To prove Theorem 5.3.2, we begin with the following two lemmas. The first lemma ensures that the R-Rayleigh number σ defined by (5.3.25) is the parameter required to describe the critical-crossing for the real eigenvalues of (5.3.11).

Lemma 5.3.3 Assume that

$$\tilde{R} \neq \frac{-\text{Le}^2(1 + \Pr) \gamma_{j_1 k_1 1}^6}{(1 - \text{Le}) \Pr \alpha_{j_1 k_1}^2},$$

and for the R-Rayleigh number σ near σ_c , all real eigenvalues of (5.3.11) are real:

$$\beta_1 \geq \beta_2 \geq \cdots \geq \beta_m \geq \beta_{m+1} \geq \cdots. \quad (5.3.36)$$

Then $\beta_1 = \beta_{j_1 k_1 1}^1$, and with σ_c defined by (5.3.27),

$$\beta_{j_1 k_1 1}(\sigma) = \begin{cases} < 0 & \text{if } \sigma < \sigma_c, \\ = 0 & \text{if } \sigma = \sigma_c, \\ > 0 & \text{if } \sigma > \sigma_c. \end{cases} \quad (5.3.37)$$

Proof. Let $\alpha = \alpha_{j_1 k_1}$, $\gamma = \gamma_{j_1 k_1 1}$, $R_0 - \text{Le}^{-1} \tilde{R}_0 = \gamma^6 / \alpha^2$, where (j_1, k_1) as in (5.3.27). We shall show that

$$(\Pr + \text{Le} + \Pr \text{Le}) \gamma^4 - \Pr \alpha^2 \gamma^{-2} (R_0 - \tilde{R}_0) > 0. \quad (5.3.38)$$

Assume that (5.3.38) is not true, we consider the case

$$(\Pr + \text{Le} + \Pr \text{Le}) \gamma^4 - \Pr \alpha^2 \gamma^{-2} (R_0 - \tilde{R}_0) < 0. \quad (5.3.39)$$

Note that the solution $\beta(\sigma) = \beta_{j_1 k_1 l}^1(\sigma)$ of (5.3.24) with $(j, k, l) = (j_1, k_1, 1)$ is continuous on σ , and $\beta(\sigma_c) = 0$. Hence

$$\beta(\sigma) \rightarrow 0 \quad \text{as } \sigma \rightarrow \sigma_c = R_0 - \text{Le}^{-1} \tilde{R}. \quad (5.3.40)$$

Thus, near $\sigma = \sigma_c$ the equation (5.3.24) can be written as

$$\beta(\sigma) = \frac{-b_0(\sigma)}{b_1(\sigma)} + o(\beta(\sigma)), \quad (5.3.41)$$

where

$$\begin{aligned} b_0(\sigma) &= \text{Pr Le} \alpha^2 \left(\frac{\gamma^6}{\alpha^2} - \sigma \right), \\ b_1(\sigma) &= (\text{Pr} + \text{Le} + \text{Pr Le}) \gamma^4 - \text{Pr} \alpha^2 \gamma^{-2} (R - \tilde{R}). \end{aligned} \quad (5.3.42)$$

It follows from (5.3.39)–(5.3.41) that for $\sigma = R - \text{Le}^{-1} \tilde{R}$ near $\sigma_c = R_0 - \text{Le}^{-1} \tilde{R}_0$,

$$\beta(\sigma) \begin{cases} > 0 & \text{if } \sigma < \sigma_c, \\ = 0 & \text{if } \sigma = \sigma_c, \\ < 0 & \text{if } \sigma > \sigma_c. \end{cases} \quad (5.3.43)$$

We write the equation (5.3.24) in the following form

$$\beta^3 + b_2 \beta^2 + b_1(\sigma) \beta + b_0(\sigma) = 0 \quad (5.3.44)$$

where $b_0(\sigma), b_1(\sigma)$ are as in (5.3.42), and

$$b_2 = (\text{Pr} + \text{Le} + 1) \gamma^2 > 0.$$

Meanwhile, it is easy to see that

$$b_0(\sigma) \rightarrow +\infty, b_1(\sigma) \rightarrow +\infty, \text{ as } \sigma \rightarrow -\infty.$$

It implies that when σ is sufficiently small, the real solutions of (5.3.44) must be negative. Thus, by (5.3.43) there exists a number $\sigma_0 < \sigma_c$ such that the solution $\beta(\sigma)$ of (5.3.44) vanishes at $\sigma = \sigma_0$ which is a contradiction to that $b_0(\sigma_0) \neq 0$. Thus, we derive $b_1(\sigma_c) \geq 0$. By the assumption in the lemma, $b_1(\sigma_c) \neq 0$. Hence, (5.3.38) holds true, i.e., $b_1(\sigma_c) > 0$. By (5.3.41), we can obtain (5.3.37).

In the following, we shall prove that $\beta_{j_1 k_1 l}^1 = \beta_l$ as in (5.3.36). We only need to consider the real eigenvalues β satisfying (5.3.24). Let $\beta_m = \beta_{jkl}(\sigma_c)$ be a solution of (5.3.24) at $\sigma = \sigma_c$. We consider the coefficients of (5.3.24) at $\sigma = \sigma_c$. Thanks to (5.3.27),

$$(\text{Pr} + \text{Le} + 1) \gamma_{jkl}^2 > 0,$$

$$\text{Pr Le} \gamma_{jkl}^6 - \text{Pr} \alpha_{jk}^2 (\text{Le} R_0 - \tilde{R}) = \text{Pr} \alpha_{jk}^2 \text{Le} \left[\frac{\gamma_{jkl}^6}{\alpha_{jk}^2} - \sigma_c \right] \geq 0.$$

Thanks to (5.3.38), we have

$$\frac{(\Pr + \Le + \Pr \Le) \gamma^6}{\Pr \alpha^2} > R_0 - \tilde{R}_0.$$

Thus, we obtain

$$\begin{aligned} & (\Pr + \Le + \Pr \Le) \gamma_{jkl}^4 - \Pr \alpha_{jk}^2 \gamma_{jkl}^{-2} (R_0 - \tilde{R}_0) \\ &= \Pr \alpha_{jk}^2 \gamma_{jkl}^{-2} \left[\frac{\Pr + \Le + \Pr \Le}{\Pr} \frac{\gamma_{jkl}^6}{\alpha_{jk}^2} - (R_0 - \tilde{R}_0) \right] \\ &> \alpha_{jk}^2 \gamma_{jkl}^{-2} (\Pr + \Le + \Pr \Le) \left(\frac{\gamma_{jkl}^6}{\alpha_{jk}^2} - \frac{\gamma^6}{\alpha^2} \right) \\ &> 0 \quad (\text{by (5.3.27)}). \end{aligned}$$

Hence, the coefficients of (5.3.24) at $\sigma = \sigma_c$ are nonnegative and strictly positive provided that $\gamma_{jkl}^6/\alpha_{jkl}^2 \neq \sigma_c$. It follows that

$$\beta_m = \beta_{jkl}(\sigma_c) < 0 \quad \forall \sigma_c \neq \gamma_{jkl}^6/\alpha_{jkl}^2.$$

Thus, we derive that $\beta_{j_1 k_1 1} = \beta_1$ for σ near σ_c . The proof is complete. \square

The following lemma shows that the C-Rayleigh number η defined by (5.3.28) characterizes the critical-crossing at η_c for the complex eigenvalues of (5.3.11).

Lemma 5.3.4 *Let (j_1, k_1) satisfy (5.3.30), and the condition (5.3.29) holds true for $(j, k, l) = (j_1, k_1, 1)$. Then the pair of complex eigenvalues $\beta_{j_1 k_1 1}^1(\eta) = \bar{\beta}_{j_1 k_1 1}^2(\eta)$ are critical-crossing at $\eta = \eta_c$:*

$$Re \beta_{j_1 k_1 1}^1(\eta) \begin{cases} < 0 & \text{if } \eta < \eta_c, \\ = 0 & \text{if } \eta = \eta_c, \\ > 0 & \text{if } \eta > \eta_c. \end{cases}$$

Proof. Near $\eta = \eta_c$ the solution $\beta_{j_1 k_1 1}^1(\eta)$ of (5.3.44) takes the form

$$\begin{aligned} \beta_{j_1 k_1 1}^1(\eta) &= \lambda(\eta) + i\rho(\eta), \\ \lambda(\eta) &\rightarrow 0, \quad \rho(\eta) \rightarrow \rho_0 \quad \text{as } \eta \rightarrow \eta_c. \end{aligned} \tag{5.3.45}$$

Inserting (5.3.45) into (5.3.44) we get

$$\begin{aligned} (-3\rho^2 + b_1)\lambda + b_0 - b_2\rho^2 + o(\lambda) &= 0, \\ -\rho^3 + \rho b_1 + 2\rho b_2 \lambda + o(\lambda) &= 0. \end{aligned} \tag{5.3.46}$$

Since $\rho_0 \neq 0$, we derive from (5.3.46) that

$$\lambda(\eta) + o(\lambda) = \frac{b_0 - b_2 \rho^2}{3\rho^2 - b_1} = \frac{b_0 - b_1 b_2 - 2b_2^2 \lambda}{2b_1 + 6b_2 \lambda} + o(\lambda),$$

which yields

$$(1 + \frac{b_2^2}{b_1})\lambda(\eta) + o(\lambda) = \frac{b_0 - b_1 b_2}{2b_1 + 6b_2 \lambda}. \quad (5.3.47)$$

Note that

$$\rho_0^2 = b_1(\eta_c) > 0, \quad \lambda(\eta_c) = 0.$$

We derive from (5.3.47) that

$$\operatorname{Re}\beta_{j_1 k_1 1}^1(\eta) = \lambda(\eta) \begin{cases} < 0 & \text{if } b_0 < b_1 b_2, \\ = 0 & \text{if } b_0 = b_1 b_2, \\ > 0 & \text{if } b_0 > b_1 b_2, \end{cases} \quad (5.3.48)$$

for η near η_c . It is easy to check that

$$b_0 \begin{cases} < b_1 b_2 & \text{if } \eta < \eta_c, \\ = b_1 b_2 & \text{if } \eta = \eta_c, \\ > b_1 b_2 & \text{if } \eta > \eta_c. \end{cases}$$

Thus, the lemma follows from (5.3.48). \square

Proof of Theorem 5.3.2. Let $\sigma = R - \operatorname{Le}^{-1} \tilde{R}$ be at the critical state

$$R - \operatorname{Le}^{-1} \tilde{R} = \sigma_c = \gamma^6 / \alpha^2. \quad (5.3.49)$$

Assume that

$$\eta = R - \frac{\operatorname{Pr} + \operatorname{Le}}{\operatorname{Pr} + 1} \tilde{R} > \eta_c = \frac{(\operatorname{Pr} + \operatorname{Le})(1 + \operatorname{Le})}{\operatorname{Pr}} \frac{\gamma^6}{\alpha^2}. \quad (5.3.50)$$

Then, we deduce from (5.3.49) and (5.3.50) that

$$\tilde{R} \begin{cases} > \frac{\operatorname{Le}^2(\operatorname{Pr} + 1)}{(1 - \operatorname{Le})\operatorname{Pr}} \frac{\gamma^6}{\alpha^2} & \text{if } \operatorname{Le} < 1, \\ < \frac{\operatorname{Le}^2(\operatorname{Pr} + 1)}{(1 - \operatorname{Le})\operatorname{Pr}} \frac{\gamma^6}{\alpha^2} & \text{if } \operatorname{Le} > 1, \end{cases}$$

which implies that $K < 0$.

In addition, let $\eta = \eta_c$, namely

$$R - \frac{\operatorname{Pr} + \operatorname{Le}}{\operatorname{Pr} + 1} \tilde{R} = \frac{(\operatorname{Pr} + \operatorname{Le})(1 + \operatorname{Le})}{\operatorname{Pr}} \frac{\gamma^6}{\alpha^2}. \quad (5.3.51)$$

Then noticing that $K < 0$, we can infer from (5.3.51) that

$$\rho_0^2 = (\text{Pr} + \text{Le} + \text{Pr Le})\gamma^4 - \text{Pr} \alpha^2 \gamma^{-2}(R - \tilde{R}) > 0. \quad (5.3.52)$$

Thus, by (5.3.52) and Lemma 5.3.4, the conditions (5.3.49) and (5.3.50) imply that η_c is the first critical Rayleigh number, and (5.3.34) and (5.3.35) hold true.

Likewise, let

$$\eta = R - \frac{\text{Pr} + \text{Le}}{\text{Pr} + 1} \tilde{R} < \frac{(\text{Pr} + \text{Le})(1 + \text{Le})}{\text{Pr}} \frac{\gamma^6}{\alpha^2}. \quad (5.3.53)$$

Then, it is clear that (5.3.49) and (5.3.53) imply that σ_c is the first critical Rayleigh number, and

$$\tilde{R} \begin{cases} < \frac{\text{Le}^2(\text{Pr} + 1)}{(1 - \text{Le})\text{Pr}} \frac{\gamma^6}{\alpha^2} & \text{if } \text{Le} < 1, \\ > \frac{\text{Le}^2(\text{Pr} + 1)}{(1 - \text{Le})\text{Pr}} \frac{\gamma^6}{\alpha^2} & \text{if } \text{Le} > 1, \end{cases}$$

which implies $K > 0$. By Lemma 5.3.3 the conclusions (5.3.32) and (5.3.33) are valid. Thus, the theorem is proved. \square

5.3.3 Nonlinear Dynamic Transitions

Transitions and Onset of the First Real Eigenvalues By Theorem 5.3.2, we know that under the conditions $K > 0$ with K defined by (5.3.31), the Boussinesq equations (5.3.6) with (5.3.8) always undergo a dynamic transition at $\sigma = \sigma_c$.

In this section, we study the transition from a real simple eigenvalue. We know that generically, the first eigenvalues of (5.3.11) are simple. Hence we always assume that the first real eigenvalue of (5.3.11) near $\sigma = \sigma_c$ is simple.

The following main theorem shows that the transition is either a Type-I (continuous) transition, or a Type-II (jump) transition, depending on the sign of the following parameter:

$$b_1 = \sigma_c - \frac{1 - \text{Le}^2}{\text{Le}^3} \tilde{R}. \quad (5.3.54)$$

The Type-II transition leads also to the existence of metastable stables, saddle-node bifurcations, and the hysteresis associated with it. In this case, as the R-Rayleigh number goes beyond the critical value σ_c , the physical reality is represented by local attractors away from the basic equilibrium state.

Theorem 5.3.5 Assume that $\text{Le} \neq 1$, $K > 0$, and $b_1 > 0$, where K is defined by (5.3.31), and b_1 is defined by (5.3.54). Then the problem (5.3.6) with (5.3.8) undergoes a Type-I (continuous) transition at the critical R-Rayleigh number σ_c , and the following assertions hold true:

- (1) If the R-Rayleigh number σ crosses σ_c , the problem bifurcates to two steady-state solutions $\psi_i^\sigma = (u_i^\sigma, T_i^\sigma, S_i^\sigma)$, $i = 1, 2$.
- (2) There is an open set $U \subset H$ with $\psi = 0 \in U$ such that $\overline{U} = \overline{U}_1 + \overline{U}_2$, $U_1 \cap U_2 = \emptyset$, $\psi = 0 \in \partial U_1 \cap \partial U_2$, and $\psi_i^\sigma \in U_i$ attracts U_i ($i = 1, 2$).
- (3) ψ_i^σ can be expressed as

$$\psi_i^\sigma = (-1)^i a \sqrt{\beta(\sigma)} \psi_{j_1 k_1 1}^1 + o(\beta^{1/2}(\sigma)) \quad \text{for } i = 1, 2,$$

where $a > 0$ is a constant, $\beta(\sigma) = \beta_{j_1 k_1 1}^1(\sigma)$ is the first eigenvalue satisfying (5.3.37), and $\psi_{j_1 k_1 1}^1$ is the first eigenvector given by (5.3.21).

- (4) If $\psi_0 \in U_i$ ($i = 1, 2$), there exists a $t_0 > 0$ such that when $t > t_0$, the velocity component of $\psi(t, \psi_0)$ is topologically equivalent to the structure as shown in Fig. 5.12, with either the same or the reversed orientation. Here $\psi(t, \psi_0)$ is the solution of (5.3.6) with (5.3.8) with initial data ψ_0 .

Theorem 5.3.6 Assume that $\text{Le} \neq 1$, $K > 0$, and $b_1 < 0$. Then the problems (5.3.6)–(5.3.8) undergo a Type-II (jump) transition at σ_c , and the following assertions hold true:

- (1) The transition of this problem at σ_c is a subcritical bifurcation, i.e., there are steady-state solutions bifurcated on $\sigma < \sigma_c$, which are repellers, and no steady-state solutions bifurcated on $\sigma > \sigma_c$.
- (2) There is a saddle-node bifurcation of steady-state solutions from (ψ_1^*, σ^*) and (ψ_2^*, σ^*) with $\sigma^* < \sigma_c$, and there are four branches of steady-state solutions ψ_σ^i ($1 \leq i \leq 4$) in which ψ_σ^1 and ψ_σ^2 are bifurcated from (ψ_1^*, σ^*) , and $\psi_\sigma^3, \psi_\sigma^4$ bifurcated from (ψ_2^*, σ^*) on $\sigma > \sigma^*$, such that

$$\lim_{\sigma \rightarrow \sigma_c^-} \psi_\sigma^1 = \lim_{\sigma \rightarrow \sigma_c^-} \psi_\sigma^2 = 0, \quad \psi_{\sigma_c}^3, \psi_{\sigma_c}^4 \neq 0,$$

and ψ_σ^3 and ψ_σ^4 are attractors for $\sigma^* < \sigma < \sigma_c + \varepsilon$ with some $\varepsilon > 0$.

Proof of Theorems 5.3.5 and 5.3.6. Let $J_1 = (j_1, k_1, 1)$, $\psi_{J_1} = \psi_{j_1 k_1 1}^1$. The reduced equation of (5.3.6)–(5.3.8) in H reads

$$\frac{dx}{dt} = \beta_{J_1}^1(\sigma)x + \frac{1}{(\psi_{J_1}, \psi_{J_1}^*)}(G(\psi, \psi), \psi_{J_1}^*), \quad (5.3.55)$$

where $\psi \in H$ is written as $\psi = x\psi_{J_1} + \Phi$, Φ is the center manifold function, and for $\psi_i = (u_i, T_i, S_i) \in H$ ($i = 1, 2, 3$),

$$G(\psi_1, \psi_2) = -P((u_1 \cdot \nabla)u_2, (u_1 \cdot \nabla)T_2, (u_1 \cdot \nabla)S_2),$$

$$(G(\psi_1, \psi_2), \psi_3) = - \int_{\Omega} \left[\sum_{i,j=1}^3 u_{1i} \frac{\partial u_{2j}}{\partial x_i} u_{3j} + \sum_{i,j=1}^3 \left(u_{1i} \frac{\partial T_2}{\partial x_i} T_3 + u_{1i} \frac{\partial S_2}{\partial x_i} S_3 \right) \right] dx.$$

By Theorem A.1.1, the center manifold function Φ satisfies that

$$-L_\lambda \Phi = x^2 P_2 G(\psi_{J_1}, \psi_{J_1}) + \text{high order terms}.$$

Hence it is routine to calculate that

$$\begin{aligned} \Phi = & x^2 \left[\frac{(G(\psi_{J_1}, \psi_{J_1}), \psi_{2J}^*)}{-\beta_{2J}(\psi_{2J}, \psi_{2J}^*)} \psi_{2J} + \frac{(G(\psi_{J_1}, \psi_{J_1}), \psi_{002}^{1*})}{-\beta_{002}^1(\psi_{002}^1, \psi_{002}^{1*})} \psi_{002}^1 \right. \\ & \left. + \frac{(G(\psi_{J_1}, \psi_{J_1}), \psi_{002}^{2*})}{-\beta_{002}^2(\psi_{002}^2, \psi_{002}^{2*})} \psi_{002}^2 \right] + o(2), \end{aligned} \quad (5.3.56)$$

where

$$\begin{aligned} -\beta_{2J} &= 4\Pr \alpha^2, \quad -\beta_{002}^1 = 4\pi^2, \quad -\beta_{002}^2 = 4\Le \pi^2, \\ \psi_{2J} = \psi_{2J}^* &= (k_1 \alpha_2 \sin 2j_1 \alpha_1 \pi x_1 \cos 2k_1 \alpha_2 \pi x_2, \\ &\quad -j_1 \alpha_1 \cos 2j_1 \alpha_1 \pi x_1 \sin 2k_1 \alpha_2 \pi x_2, 0, 0, 0), \\ \psi_{002}^1 = \psi_{002}^{1*} &= (0, 0, 0, \sin 2\pi x_3, 0), \\ \psi_{002}^2 = \psi_{002}^{2*} &= (0, 0, 0, 0, \sin 2\pi x_3). \end{aligned}$$

Inserting $\psi = x\psi_{J_1} + \Phi$ into (5.3.55), we have

$$\frac{dx}{dt} = \beta_{J_1}^1(\sigma)x - \delta(\sigma)x^3 + o(3), \quad (5.3.57)$$

where

$$\begin{aligned} \delta(\sigma_c) &= \sum_{(J, i) \neq (J_1, 1)} \frac{(G(\psi_{J_1}, \psi_{J_1}), \psi_J^{i*}) \left[(G(\psi_{J_1}, \psi_{J_1}^*), \psi_J^i) + (G(\psi_J^i, \psi_{J_1}^*), \psi_{J_1}) \right]}{-\beta_J^i(\psi_{J_1}, \psi_{J_1}^*)(\psi_J^i, \psi_J^{i*})} \\ &= \frac{1}{(\psi_{J_1}, \psi_{J_1}^*)} \left[\frac{(G(\psi_{J_1}, \psi_{J_1}), \psi_{2J}^*)(G(\psi_{J_1}, \psi_{J_1}^*), \psi_{2J})}{-\beta_{2J}(\psi_{2J}, \psi_{2J}^*)} \right. \\ &\quad + \frac{(G(\psi_{J_1}, \psi_{J_1}), \psi_{002}^{1*})(G(\psi_{J_1}, \psi_{J_1}^*), \psi_{002}^1)}{-\beta_{002}^1(\psi_{002}^1, \psi_{002}^{1*})} \\ &\quad \left. + \frac{(G(\psi_{J_1}, \psi_{J_1}), \psi_{002}^{2*})(G(\psi_{J_1}, \psi_{J_1}^*), \psi_{002}^2)}{-\beta_{002}^2(\psi_{002}^2, \psi_{002}^{2*})} \right] \\ &= \frac{\pi^2}{8(\psi_{J_1}, \psi_{J_1}^*)} \left[\frac{(k_1 \alpha_2 u_{J_1} - J_1 \alpha_1 v_{J_1})(k_1 \alpha_2 u_{J_1}^* - J_1 \alpha_1 v_{J_1}^*)}{\Pr \alpha^4} + \frac{T_{J_1} T_{J_1}^*}{\pi^4} + \frac{S_{J_1} S_{J_1}^*}{\pi^4 \Le} \right]. \end{aligned}$$

Here

$$\begin{aligned} u_{J_1} = u_{J_1}^* &= -\frac{j_1 \alpha_1 \pi^2}{\alpha^2}, \quad v_{J_1} = v_{J_1}^* = -\frac{k_1 \alpha_2 \pi^2}{\alpha^2}, \quad T_{J_1} = \frac{1}{\gamma^2 + \beta_{J_1}^1}, \\ T_{J_1}^* &= \frac{\Pr R}{\gamma^2 + \beta_{J_1}^1}, \quad S_{J_1} = \frac{\text{sign}(S_0 - S_1)}{\Le \gamma^2 + \beta_{J_1}^1}, \quad S_{J_1}^* = \frac{\Pr \widetilde{R} \text{sign}(S_0 - S_1)}{\Le \gamma^2 + \beta_{J_1}^1}. \end{aligned}$$

By $\beta_{J_1}(\sigma_c) = 0$, direct calculation yields that at $\sigma = \sigma_c$,

$$\delta(\sigma_c) = \frac{\Pr(\text{Le}^3 R - \tilde{R})}{8\pi^2\gamma^4\text{Le}^3(\psi_{J_1}, \psi_{J_1}^*)} = \frac{\text{Le}^3 R - \tilde{R}}{8\pi^2\text{Le}^3 \left(\frac{\sigma_c}{\Pr} + R - \frac{\tilde{R}}{\text{Le}^2} \right)}.$$

Putting $R = \text{Le}^{-1}\tilde{R} + \sigma_c$ into $\delta(\sigma_c)$, we derive that

$$\delta = \delta(\sigma_c) = \frac{1}{8\pi^2\text{Le}^3} \frac{\text{Le}^3\sigma_c - (1 - \text{Le}^2)\tilde{R}}{(1 + \frac{1}{\Pr})\sigma_c - \frac{1-\text{Le}}{\text{Le}^2}\tilde{R}}.$$

Hence using $K > 0$, the sign of δ is the same as the sign of b_1 defined by (5.3.54), and then the theorems follow from Theorems 2.3.1 and 2.5.3, and from (5.3.57). \square

Transition to Spatiotemporal Oscillatory Patterns If $K < 0$, then, by Theorem 5.3.2, the problems (5.3.6)–(5.3.8) undergo a dynamic transition to the periodic solutions at the critical Rayleigh number η_c given by (5.3.30). It is easy to see that as $L_1/L_2 \gg 1$, η_c is given by

$$\eta_c = \frac{(\Pr + \text{Le})(1 + \text{Le})}{\Pr} \frac{\gamma_{j_101}^6}{\alpha_{j_10}^2}, \quad (5.3.58)$$

for some $(j_1, k_1) = (j_1, 0)$ with $j_1 > 0$. Here, we always assume the condition (5.3.58) hold true.

We now introduce a parameter for determining the transition type:

$$\begin{aligned} b_2 &= R_c(R_2B_1 - I_2B_2)C_3 + R_c(I_2B_1 + R_2B_2)C_4 \\ &\quad + \tilde{R}(R_3B_1 - I_3B_2)C_5 + \tilde{R}(I_3B_1 + R_3B_2)C_6, \end{aligned} \quad (5.3.59)$$

where R_2, I_2, C_3, C_4 are as in (4.4.73), $\alpha^2 = \alpha_{j_10}^2 = j_1^2\alpha_1^2\pi^2$, and

$$\begin{aligned} R_3 &= \text{Re}A_3 = \frac{\text{Le}(\alpha^2 + \pi^2)}{\text{Le}^2(\alpha^2 + \pi^2)^2 + \rho^2}, & I_3 &= \text{Im}A_3 = \frac{-\rho}{\text{Le}^2(\alpha^2 + \pi^2)^2 + \rho^2}, \\ B_1 &= \tilde{R}I_3^2 - R_cI_2^2, & B_2 &= \tilde{R}I_3R_3 - RI_2R_2, \\ C_5 &= \frac{3R_3}{2\pi\text{Le}} - \frac{\rho^2R_3}{2\pi\text{Le}(4\pi^4\text{Le}^2 + \rho^2)} - \frac{\rho I_3}{4\pi^4\text{Le}^2 + \rho^2}, \\ C_6 &= -\frac{I_3}{2\pi\text{Le}} + \frac{\rho^2I_3}{2\pi\text{Le}(4\pi^4\text{Le}^2 + \rho^2)} - \frac{\rho R_3}{4\pi^4\text{Le}^2 + \rho^2}, \\ \rho^2 &= \frac{\Pr(1 - \text{Le})\tilde{R}}{\Pr + 1} \frac{\alpha^2}{(\alpha^2 + \pi^2)} - \text{Le}^2(\alpha^2 + \pi^2)^2, \\ R_c &= \frac{(\Pr + \text{Le})(1 + \text{Le})}{\Pr} \frac{(\alpha^2 + \pi^2)^3}{\alpha^2} + \frac{\Pr + \text{Le}}{\Pr + 1}\tilde{R}. \end{aligned}$$

The following theorem characterizes the spatiotemporal transition.

Theorem 5.3.7 Let b_2 be given by (5.3.59), and $K < 0$ with K defined by (5.3.31). Then for problem (5.3.6)–(5.3.8), we have the following assertions:

- (1) If $b_2 < 0$, then the problem has a Type-I transition at $\eta = \eta_c$, and bifurcates to a periodic solution on $\eta > \eta_c$ which is an attractor, the periodic solution can be expressed as

$$\begin{aligned}\psi &= x(t)\operatorname{Re}\psi_{j_101} + y(t)\operatorname{Im}\psi_{j_101} + o(|x| + |y|), \\ (x(t), y(t)) &= \left[\frac{2\pi\operatorname{Re}\beta(\eta)|\rho|}{|b_2|} \right]^{1/2} (\sin \rho t, \cos \rho t),\end{aligned}$$

where $\beta(\eta) = \lambda(\eta) + i\rho(\eta)$ is the first eigenvalue with $\lambda(\eta_c) = \operatorname{Re}\beta(\eta_c) = 0$, $\rho(\eta_c) = \rho$ as given in (5.3.59).

- (2) If $b_2 > 0$, then the transition at $\eta = \eta_c$ is of Type-II, and there is a singular separation of periodic solutions at some (Γ^*, η^*) for $\eta^* < \eta_c$, where Γ^* is a periodic solution. In particular, there is a branch Γ_η of periodic solutions separated from Γ^* which are repellers, such that $\Gamma_\eta \rightarrow 0$ as $\eta \rightarrow \eta_c$.

Remark 5.3.8 The following example shows that both cases with $b_2 < 0$ and $b_2 > 0$ may appear in different physical regimes.

Let $\rho^2 \cong 0$. In this case we have

$$R_c \cong \frac{\Pr + \Le}{\Pr(1 - \Le)} \frac{\gamma^6}{\alpha^2}, \quad \widetilde{R} \cong \frac{(\Pr + 1)\Le^2}{\Pr(1 - \Le)} \frac{\gamma^6}{\alpha^2} \begin{cases} > 0 & \text{for } 1 > \Le, \\ < 0 & \text{for } 1 < \Le. \end{cases}$$

Thus, the number b_2 defined by (5.3.59) is given by

$$\begin{aligned}b_2 &\cong -\gamma^{-12}(R_c - \Le^{-4}\widetilde{R})(R_c - \Le^{-3}\widetilde{R}) \\ &= \frac{-1}{\Pr \alpha^2(1 - \Le)\Le^3} (\Le(\Pr + \Le) - \Pr - 1)(\Le^2(\Pr + \Le) - \Pr - 1).\end{aligned}$$

It is clear that for $\Le \approx 1$ and for \Pr small,

$$b_2 \begin{cases} < 0 & \text{if } \Le < 1, \\ > 0 & \text{if } \Le > 1. \end{cases}$$

Proof of Theorem 5.3.7. Let $J = (j_1, 0, 1)$ and $\beta(\eta) = \beta_J^1(\eta) = \lambda(\eta) + i\rho(\eta)$ be the first eigenvalue of (5.3.45) near $\eta = \eta_c$ with

$$\lambda(\eta_c) = 0, \quad \rho(\eta_c) = \rho > 0. \quad (5.3.60)$$

By (5.3.60), the eigenvectors $\psi_J = \psi_J^1 + i\psi_J^2$ corresponding to $\beta(\eta)$ are given by

$$\begin{aligned}\psi_J^1 &= \left(-\frac{1}{j_1\alpha_1} \sin \phi \cos \pi x_3, 0, \cos \phi \sin \pi x_3,\right. \\ &\quad \left.R\text{e}A_2(\beta) \cos \phi \sin \pi x_3, R\text{e}A_3(\beta) \cos \phi \sin \pi x_3,\right) \\ \psi_J^2 &= (0, 0, 0, \text{Im}A_2(\beta) \cos \phi \sin \pi x_3, \text{Im}A_3 \cos \phi \sin \pi x_3),\end{aligned}$$

where $\phi = j_1\alpha_1\pi x_1$. By (5.3.23) the conjugate eigenvectors $\psi_J^* = \psi_J^{1*} + i\psi_J^{2*}$ are given by

$$\begin{aligned}\psi_J^{1*} &= \left(-\frac{1}{j_1\alpha_1} \sin \phi \cos \pi x_3, 0, \cos \phi_1 \sin \pi x_3,\right. \\ &\quad \left.\text{Pr } R\text{Re}A_2(\bar{\beta}) \cos \phi \sin \pi x_3, -\text{Pr } R\text{Re}A_3(\bar{\beta}) \cos \phi \sin \pi x_3,\right) \\ \psi_J^{2*} &= (0, 0, 0, \text{Pr } R\text{Im}A_2(\bar{\beta}) \cos \phi \sin \pi x_3, -\text{Pr } R\text{Im}A_3(\bar{\beta}) \cos \phi \sin \pi x_3)\end{aligned}$$

and

$$A_2(\beta) = \frac{1}{\alpha^2 + \pi^2 + \beta}, \quad A_3(\beta) = \frac{1}{\text{Le}(\alpha^2 + \pi^2) + \beta}. \quad (5.3.61)$$

The conjugate eigenvectors Φ_J^{1*} and Φ_J^{2*} , satisfying

$$(\psi_J^1, \Phi_J^{1*}) = (\psi_J^2, \Phi_J^{2*}) \neq 0, \quad (\psi_J^1, \Phi_J^{2*}) = (\psi_J^2, \Phi_J^{1*}) = 0,$$

are given by

$$\Phi_J^{1*} = \psi_J^{1*} + C\psi_J^{2*}, \quad \Phi_J^{2*} = -C\psi_J^{1*} + \psi_J^{2*}, \quad (5.3.62)$$

where

$$C = \frac{(\psi_J^1, \psi_J^{2*})}{(\psi_J^1, \psi_J^{1*})} = -\frac{(\psi_J^2, \psi_J^{1*})}{(\psi_J^2, \psi_J^{2*})} = \frac{B_2}{B_1}. \quad (5.3.63)$$

The reduced equations of (5.3.6)–(5.3.8) read

$$\begin{aligned}\frac{dx}{dt} &= \lambda x + \rho y + \frac{1}{(\psi_J^1, \Phi_J^{1*})}(G(\psi, \psi), \Phi_J^{1*}), \\ \frac{dy}{dt} &= -\rho x + \lambda y + \frac{1}{(\psi_J^2, \Phi_J^{2*})}(G(\psi, \psi), \Phi_J^{2*}),\end{aligned} \quad (5.3.64)$$

where $\psi = x\psi_J^1 + y\psi_J^2 + \Phi$, and Φ is the center manifold function.

Note that for any gradient field $\nabla\varphi$, the Leray projection $P(\nabla\varphi) = 0$. Therefore, by (5.3.62)–(5.3.63) we have

$$\begin{aligned}G(\psi_J^1, \psi_J^1) &= \frac{\pi}{2}(0, 0, 0, \text{Re}A_2 \sin 2\pi x_3, \text{Re}A_3 \sin 2\pi x_3), \\ G(\psi_J^1, \psi_J^2) &= \frac{\pi}{2}(0, 0, 0, \text{Im}A_2 \sin 2\pi x_3, \text{Im}A_3 \sin 2\pi x_3), \\ G(\psi_J^1, \Phi_J^{1*}) &= \frac{\pi \text{Pr}}{2B_1}(0, 0, 0, R(B_1\text{Re}A_2 - B_2\text{Im}A_2) \sin 2\pi x_3, \\ &\quad - \tilde{R}(B_1\text{Re}A_3 - B_2\text{Im}A_3) \sin 2\pi x_3),\end{aligned}$$

$$\begin{aligned} G(\psi_J^1, \Phi_J^{2*}) &= \frac{\pi \Pr}{2B_1}(0, 0, 0, -R(B_2 \operatorname{Re} A_2 + B_1 \operatorname{Im} A_2) \sin 2\pi x_3, \\ &\quad R(B_2 \operatorname{Re} A_3 + B_1 \operatorname{Im} A_3) \sin 2\pi x_3) \\ G(\psi_J^2, \psi) &= 0, \quad \forall \psi \in H. \end{aligned}$$

Hence, by (A.1.15), the center manifold function Φ is expressed as

$$\Phi = \Phi_1 + \Phi_2 + \Phi_3 + o(2), \quad (5.3.65)$$

where $G_{ij} = G(\psi_J^i, \psi_J^j)$, and

$$\begin{aligned} \Phi_1 &= -\frac{\psi_{002}^1}{\beta_{002}^1 \|\psi_{002}^1\|^2} [x^2(G_{11}, \psi_{002}^1) + xy(G_{12}, \psi_{002}^1)] \\ &\quad - \frac{\psi_{002}^2}{\beta_{002}^2 \|\psi_{002}^2\|^2} [x^2(G_{11}, \psi_{002}^2) + xy(G_{12}, \psi_{002}^2)], \\ \Phi_2 &= \frac{2\rho^2 \psi_{002}^1}{\beta_{002}^1 (\beta_{002}^{12} + 4\rho^2) \|\psi_{002}^1\|^2} [(x^2 - y^2)(G_{11}, \psi_{002}^1) + 2xy(G_{12}, \psi_{002}^1)] \\ &\quad + \frac{2\rho^2 \psi_{002}^2}{\beta_{002}^2 ((\beta_{002}^2)^2 + 4\rho^2) \|\psi_{002}^2\|^2} [(x^2 - y^2)(G_{11}, \psi_{002}^2) + 2xy(G_{12}, \psi_{002}^2)], \\ \Phi_3 &= \frac{\rho \psi_{002}^1}{((\beta_{002}^1)^2 + 4\rho^2) \|\psi_{002}^1\|^2} [2xy(G_{11}, \psi_{002}^1) + (y^2 - x^2)(G_{12}, \psi_{002}^1)] \\ &\quad + \frac{\rho \psi_{002}^2}{((\beta_{002}^2)^2 + 4\rho^2) \|\psi_{002}^2\|^2} [2xy(G_{11}, \psi_{002}^2) + (y^2 - x^2)(G_{12}, \psi_{002}^2)]. \end{aligned}$$

It is clear that

$$(G(\Phi, \Phi_J^{i*}), \psi_J^j) = o(2) \quad \forall i, j = 1, 2.$$

Then, inserting $\psi = x\psi_J^1 + y\psi_J^2 + \Phi$ into (5.3.64), one gets

$$\begin{aligned} \frac{dx}{dt} &= \lambda x + \rho y - \frac{x(G(\psi_J^1, \Phi_J^{1*}), \Phi)}{(\psi_J^1, \Phi_J^{1*})} + o(3), \\ \frac{dy}{dt} &= -\rho x + \lambda y - \frac{x(G(\psi_J^1, \Phi_J^{2*}), \Phi)}{(\psi_J^2, \Phi_J^{2*})} + o(3). \end{aligned} \quad (5.3.66)$$

From (5.3.65) and (5.3.66) it follows that

$$\begin{aligned} \frac{dx}{dt} &= \lambda x + \rho y + \frac{\Pr \pi x[a_1 x^2 + a_2 xy + a_3 y^2]}{(\psi_J^1, \psi_J^{1*})^2 + (\psi_J^1, \psi_J^{2*})^2} + o(3), \\ \frac{dy}{dt} &= -\rho x + \lambda y + \frac{\Pr \pi x[b_1 x^2 + b_2 xy + b_3 y^2]}{(\psi_J^1, \psi_J^{1*})^2 + (\psi_J^1, \psi_J^{2*})^2} + o(3), \end{aligned} \quad (5.3.67)$$

where

$$\begin{aligned} a_1 &= R(-R_2B_1 + I_2B_2)D_1 + \tilde{R}(R_3B_1 - I_3B_2)F_1, \\ a_3 &= R(-R_2B_1 + I_2B_2)D_3 + \tilde{R}(R_3B_1 - I_3B_2)F_3, \\ b_2 &= R(I_2B_1 + R_2B_2)D_2 - \tilde{R}(I_3B_1 + R_3B_2)F_2, \end{aligned}$$

where B_1, B_2 are as in (5.3.63), and

$$\begin{aligned} D_1 &= \frac{R_2}{8\pi} - \frac{\rho^2 R_2}{16\pi(4\pi^4 + \rho^2)} + \frac{\rho\pi I_2}{4(4\pi^4 + \rho^2)}, \\ D_2 &= \frac{I_2}{8\pi} - \frac{\rho^2 I_2}{8\pi(4\pi^4 + \rho^2)} - \frac{\rho\pi R_2}{4(4\pi^4 + \rho^2)}, \\ D_3 &= \frac{\rho^2 R_2}{16\pi(4\pi^4 + \rho^2)} - \frac{\rho\pi I_2}{8(4\pi^4 + \rho^2)}, \\ F_1 &= \frac{R_3}{8\pi Le} - \frac{\rho^2 R_3}{16\pi Le(4\pi^4 Le^2 + \rho^2)} - \frac{\rho I_3}{8(4\pi^4 Le^2 + \rho^2)}, \\ F_2 &= \frac{I_3}{8\pi Le} - \frac{\rho^2 I_3}{8\pi Le(4\pi^4 Le^2 + \rho^2)} + \frac{\rho R_3}{4(4\pi^4 Le^2 + \rho^2)}, \\ F_3 &= \frac{\rho^2 R_3}{16\pi Le(4\pi^4 Le^2 + \rho^2)} + \frac{\rho I_3}{8(4\pi^4 Le^2 + \rho^2)}. \end{aligned}$$

Then we obtain

$$\begin{aligned} b &= 3a_1 + a_3 + b_2 \\ &= \frac{1}{4}[R(R_2B_1 - I_2B_2)C_3 + R(I_2B_1 + R_2B_2)C_4 \\ &\quad + \tilde{R}(R_3B_1 - I_3B_2)C_5 + \tilde{R}(I_3B_1 + R_3B_2)C_6] \\ &= \frac{1}{4}b_2, \end{aligned} \tag{5.3.68}$$

where b_2 is the number defined by (5.3.59).

By (5.3.67)–(5.3.68), this theorem follows from Theorems 2.3.7 and 2.5.7; the expression of bifurcated periodic solution in Assertion (1) can be derived by using the formula (2.3.22). The proof is complete. \square

5.3.4 Convections Scales and Dynamic Transition

Revised BEs and Critical Parameters Consider the equations (5.3.1) with a damping term δu . The revised nondimensional equations of (5.3.1) are given by

$$\begin{aligned}
\frac{\partial u}{\partial t} &= \text{Pr} \left[\Delta u - \delta u + RTk - \text{sign}(S_0 - S_1) \tilde{R} Sk - \nabla p \right] - (u \cdot \nabla) u, \\
\frac{\partial T}{\partial t} &= \Delta T + u_3 - (u \cdot \nabla) T, \\
\frac{\partial S}{\partial t} &= \text{Le} \Delta S + \text{sign}(S_0 - S_1) u_3 - (u \cdot \nabla) S, \\
\text{div} u &= 0,
\end{aligned} \tag{5.3.69}$$

supplemented with the boundary condition (5.3.8), where $\delta u = (\delta_0 u_1, \delta_0 u_2, \delta_1 u_3)$.

As in (5.3.16), the eigenvalues of the revised problem satisfy the following equation

$$\begin{aligned}
&\alpha_{jk}^2 (\gamma_{jk}^2 + \beta) (\text{Le} \gamma_{jk}^2 + \beta) (\text{Pr} \gamma_{jk}^2 + \text{Pr} \delta_1 + \beta) \\
&+ \pi^2 (\gamma_{jk}^2 + \beta) (\text{Le} \gamma_{jk}^2 + \beta) (\text{Pr} \gamma_{jk}^2 + \text{Pr} \delta_0 + \beta) \\
&- \text{Pr} R \alpha_{jk}^2 (\text{Le} \gamma_{jk}^2 + \beta) + \text{Pr} \tilde{R} \alpha_{jk}^2 (\gamma_{jk}^2 + \beta) = 0,
\end{aligned} \tag{5.3.70}$$

which is equivalent to

$$\begin{aligned}
&\gamma^2 \beta^3 + [\gamma^4 (1 + \text{Pr} + \text{Le}) + \text{Pr} (\alpha^2 \delta_1 + \pi^2 \delta_0)] \beta^2 \\
&+ [(\text{Pr} + \text{Le} + \text{Pr} \text{Le}) \gamma^6 + \text{Pr} (1 + \text{Le}) \gamma^2 (\alpha^2 \delta_1 + \pi^2 \delta_0) - \text{Pr} \alpha^2 (R - \tilde{R})] \beta \\
&+ \gamma^8 \text{Pr} \text{Le} + \gamma^4 \text{Pr} \text{Le} (\alpha^2 \delta_1 + \pi^2 \delta_0) - \text{Pr} \gamma^2 \alpha^2 (\text{Le} R - \tilde{R}) = 0.
\end{aligned} \tag{5.3.71}$$

Let $\beta = 0$, then we derive from (5.3.71) the critical Rayleigh number for the first real eigenvalues as

$$R_{c1} = \text{Le}^{-1} \tilde{R} + \min_{\alpha^2} \left[\gamma^2 \delta_1 + \frac{\gamma^6 + \pi^2 \gamma^2 \delta_0}{\alpha^2} \right], \tag{5.3.72}$$

where $\gamma^2 = \alpha^2 + \pi^2$. Let $\beta = i\rho_0$ ($\rho_0 \neq 0$) in (5.3.71), then we get the critical Rayleigh number for the first complex eigenvalues as follows

$$R_{c2} = \min_{\alpha^2} (\Gamma_1 + \Gamma_2 + \Gamma_3), \tag{5.3.73}$$

where

$$\begin{aligned}
\Gamma_1 &= \frac{(\text{Pr} + \text{Le}) \gamma^4 + \text{Pr} (\alpha^2 \delta_1 + \pi^2 \delta_0) \tilde{R}}{(\text{Pr} + 1) \gamma^4 + \text{Pr} (\alpha^2 \delta_1 + \pi^2 \delta_0)}, \\
\Gamma_2 &= \frac{(\text{Pr} + 1)(\text{Pr} + \text{Le})(1 + \text{Le}) \gamma^{10} + \text{Pr} (1 + \text{Le})(1 + 2\text{Pr} + \text{Le}) \gamma^6 (\alpha^2 \delta_1 + \pi^2 \delta_0)}{\text{Pr} \alpha^2 ((\text{Pr} + 1) \gamma^4 + \text{Pr} (\alpha^2 \delta_1 + \pi^2 \delta_0))}, \\
\Gamma_3 &= \frac{\text{Pr}^2 (1 + \text{Le}) \gamma^2 (\alpha^2 \delta_1 + \pi^2 \delta_0)^2}{\text{Pr} \alpha^2 ((\text{Pr} + 1) \gamma^4 + \text{Pr} (\alpha^2 \delta_1 + \pi^2 \delta_0))}.
\end{aligned}$$

It is ready to check that the first eigenvectors of the revised problem are the same as that in (5.3.21) and (5.3.23).

For (5.3.69), we need to modify the transition theorems, Theorems 5.3.5, 5.3.6, and 5.3.7, i.e., to modify the numbers b_1 and b_2 given by (5.3.54) and (5.3.59) associated with the new parameters (5.3.72) and (5.3.73). Thus, we can make a comparison between the new and old theories.

Physical Parameters in Oceanography In oceanography, main physical parameters are listed as

$$\begin{aligned} \text{Pr} &= 8, & \nu &= 1.1 \times 10^{-6} \text{ m}^2 \text{s}^{-1}, & \alpha_T &= 2.1 \times 10^{-4} \text{ K}^{-1}, \\ \text{Le} &= 10^{-2}, & \kappa_T &= 1.4 \times 10^{-7} \text{ m}^2 \text{s}^{-1}, & h &= 4 \times 10^3 \text{ m}. \end{aligned} \quad (5.3.74)$$

The haline contraction coefficient α_S satisfies

$$\alpha_S = \frac{1}{\rho} \frac{d\rho}{dS} \times 10^{-3}, \quad \rho = S\rho_S + (1 - S)\rho_W,$$

where S is the concentration of salt with unit $0/_{00}$ (psu), ρ_S and ρ_W are the densities of salt and water given by

$$\rho_W = 10^3 \text{ kg/m}^3, \quad \rho_S = 1.95 \times 10^3 \text{ kgm}^{-3}.$$

At the sea water density $\rho = 1.01 \sim 1.05 \times 10^3 \text{ kgm}^{-3}$, the haline concentration coefficient α_S is about

$$\alpha_S = \frac{1}{\rho} (\rho_S - \rho_W) \times 10^{-3} \cong 0.92 \times 10^{-3} (\text{psu})^{-1}.$$

Thus, for the oceanic motion we obtain the thermal and saline Rayleigh numbers as follows

$$R = \frac{g\alpha_T(T_0 - T_1)}{\kappa_T\nu} h^3 = 0.86 \times 10^{21} (T_0 - T_1) [\text{ }^\circ\text{C}^{-1}], \quad (5.3.75)$$

$$\tilde{R} = \frac{g\alpha_S(S_0 - S_1)}{\kappa_T\nu} h^3 = 3.75 \times 10^{21} (S_0 - S_1) (\text{psu})^{-1}. \quad (5.3.76)$$

For the thermohaline circulation, we know that its scale is tens of thousands kilometer, namely

$$L_c = O(10^4) \quad (h = 4 \text{ km as unit}). \quad (5.3.77)$$

For the damping coefficients δ_0 and δ_1 , by (4.5.17), we have

$$\delta_i = C_i h^4 / \nu = 2.33 \times 10^{20} C_i \text{ m}^2 \cdot \text{s} \quad (i = 0, 1).$$

Phenomenologically, the constants C_i depend on the density ρ , or equivalently on the pressure p , and we assume that C_i is proportional to ρ . Note that the ratio of densities of water and air is about 10^3 , i.e.

$$\rho_W / \rho_a \cong 10^3.$$

Therefore, the constant C_1 of water is about 10^3 times that of air. Thus, by (4.5.17), for water we take

$$C_1 \approx 10^3 (\text{m}^2 \cdot \text{s})^{-1}. \quad (5.3.78)$$

However, the constant C_0 of sea water is not larger than that of air. Thus, we take

$$C_0 \cong \frac{1}{2} \times 10^{-12}. \quad (5.3.79)$$

Then, we obtain the damping coefficients as

$$\delta_0 = 1.17 \times 10^8, \quad \delta_1 = 2.33 \times 10^{23}. \quad (5.3.80)$$

We remark that (5.3.78) and (5.3.79) are theoretically estimated values. However, they do provide a fairly reasonable explanation to the critical Rayleigh numbers and the convective scales in both the atmospheric and oceanic circulations.

Critical Rayleigh Numbers Due to (5.3.80), the critical Rayleigh numbers (5.3.72) and (5.3.73) can be approximatively expressed as

$$\begin{aligned} R_{c_1} - \text{Le}^{-1} \tilde{R} &\cong \min_{\alpha^2} \left[(\pi^2 + \alpha^2) \delta_1 + \frac{\pi^4 \delta_0}{\alpha^2} \right], \\ R_{c_2} - \tilde{R} &\cong (1 + \text{Le}) \min_{\alpha^2} \left[(\pi^2 + \alpha^2) \delta_1 + \frac{\pi^4 \delta_0}{\alpha^2} \right]. \end{aligned}$$

Thus, they have the same α_c^2 given by

$$\alpha_c^2 = \pi^2 \left[\frac{\delta_0}{\delta_1} \right]^{1/2} = 2.24 \times 10^{-7}. \quad (5.3.81)$$

The convective scale is

$$L_c = \frac{\pi}{\alpha_c} = \left[\frac{\delta_1}{\delta_0} \right]^{1/4} = 0.67 \times 10^4. \quad (5.3.82)$$

The critical Rayleigh numbers for real and complex eigenvalues are respectively given by

$$\sigma_c = (\pi^2 + \alpha_c^2) \delta_1 + \frac{\pi^4 \delta_0}{\alpha_c^2} = 2.33 \times 10^{24}, \quad (5.3.83)$$

$$\eta_c = (1 + \text{Le}) \left[(\pi^2 + \alpha_c^2) \delta_1 + \frac{\pi^4 \delta_0}{\alpha_c^2} \right] = (1 + \text{Le}) \sigma_c. \quad (5.3.84)$$

Here, as in (5.3.25) and (5.3.26), we define the R-Rayleigh number σ and C-Rayleigh number η as follows

$$\sigma = R - \text{Le}^{-1} \tilde{R}, \quad \eta = R - \tilde{R}. \quad (5.3.85)$$

It is seen that the theoretical value (5.3.82) agrees with the realistic length scale given in (5.3.77).

Dynamic Transition of Revised System From (5.3.83) to (5.3.85) we can see that if

$$\sigma = R - \text{Le}^{-1}\tilde{R} = \sigma_c, \quad \text{and} \quad \eta = R - \tilde{R} < \eta_c = (1 + \text{Le})\sigma_c,$$

then $\tilde{R} < \text{Le}^2\sigma_c/(1 - \text{Le})$. In this case σ_c is the first critical Rayleigh number.

In addition, if

$$R - \text{Le}^{-1}\tilde{R} = \sigma_c, \quad R - \tilde{R} > (1 + \text{Le})\sigma_c,$$

then $\tilde{R} > \text{Le}^2\sigma_c/(1 - \text{Le})$, and η_c is the first critical Rayleigh number. Thus, by Theorem 5.3.2, we obtain the following physical conclusions:

Physical Conclusion 5.3.9 *For the thermohaline circulation, we have the following assertions:*

1. If

$$\tilde{R} < \text{Le}^2\sigma_c = 2.35 \times 10^{20},$$

the number σ_c given by (5.3.83) is the first critical Rayleigh number.

2. If

$$\tilde{R} > \frac{\text{Le}^2}{1 - \text{Le}}\sigma_c = 2.35 \times 10^{20},$$

the number η_c given by (5.3.84) is the first critical Rayleigh number.

Note that the first eigenvectors of the linearized equations of (5.3.69) are the same as that of (5.3.11). For the transition of (5.3.69) from real eigenvalues, the parameter b_1 should be as in (5.3.54); namely

$$b_1 = \sigma_c - \frac{1 - \text{Le}^2}{\text{Le}^3}\tilde{R}.$$

Here $R = \sigma_c + \text{Le}^{-1}\tilde{R}$, $\text{Le} = 10^{-2}$, $\sigma_c = 2.33 \times 10^{24}$, $\text{Pr} = 8$, and $\alpha_c^2 = 2.24 \times 10^{-7}$. Thus we have

$$b_1 \cong 10^6(2.33 \times 10^{18} - \tilde{R}), \quad (5.3.86)$$

where \tilde{R} is given by (5.3.76).

By Physical Conclusion 5.3.9, as $\tilde{R} < 2.35 \times 10^{20}$, the transition of (5.3.69) is at $\sigma = \sigma_c$. The following physical conclusion is a revised version of Theorems 5.3.5 and 5.3.6.

Physical Conclusion 5.3.10 *Let $\tilde{R} < 2.35 \times 10^{20}$, and b_1 be as in (5.3.86). Then the problem (5.3.69) with (5.3.8) undergoes a dynamic transition from $\sigma = \sigma_c$, and the following assertions hold true:*

- (1) If $\tilde{R} < 2.33 \times 10^{18}$, the transition is continuous, and the problem bifurcates to two steady-state solutions ψ_1^σ and ψ_2^σ for $\sigma > \sigma_c$, σ as in (5.3.85), ψ_1^σ and ψ_2^σ have the expressions as in Assertion (3) of Theorem 5.3.5, with U_1 and U_2 as their basin of attractions in H respectively.
- (2) If $2.33 \times 10^{18} < \tilde{R} < 2.35 \times 10^{20}$, the transition is jump, and there are two saddle-node bifurcations from (ψ_1^*, σ^*) and (ψ_2^*, σ^*) with $\sigma^* < \sigma_c$.
- (3) When the transition is Type-I, if the initial value $\tilde{\psi} \in U_i$, then there is a time t_0 such that as $t > t_0$, the velocity component u , in the solution $\psi(t, \psi_0)$ with initial value ψ_0 , is topologically equivalent to the structure as shown in Fig. 5.12.

For the transition to periodic solutions, the revised parameters are as follows

$$\begin{aligned}\alpha^2 &= \alpha_c^2 = 2.24 \times 10^{-7} \ll 1, \\ \gamma^2 &= \alpha^2 + \pi^2 \cong \pi^2, \\ R &= R_{c_2} = \tilde{R} + \eta_c \quad (\eta_c \text{ as in (5.3.84)}).\end{aligned}$$

From (5.3.71), we can get the imaginary part $\beta = i\rho$ as

$$\rho^2 = (\Pr + \Le + \Pr \Le) \gamma^4 + \Pr (1 + \Le) (\alpha^2 \delta_1 + \pi^2 \delta_0) + \frac{\Pr \alpha^2 \eta_c}{\gamma^2} \cong 7.3 \times 10^{16}.$$

Thus, we derive that

$$\begin{aligned}R_2 &\cong \frac{\pi^2}{\rho^2}, & R_3 &\cong \frac{\Le \pi^2}{\rho^2}, & I_2 &\cong I_3 \cong -\frac{1}{\rho}, \\ C_3 &\cong -\frac{2\pi}{\rho^2}, & C_4 &\cong \frac{\pi(\pi^2 - 2)}{\rho^3}, & C_5 &\cong \frac{\pi + 1}{\rho^2}, \\ C_6 &\cong -\frac{\Le \pi^2(2\pi + 1)}{\rho^3}, & B_1 &\cong \frac{1}{\rho^2}(\tilde{R} - R_{c_2}), & B_2 &\cong \frac{\pi^2(R_{c_2} - \Le \tilde{R})}{\rho^3}.\end{aligned}$$

Then the parameter b_2 in (5.3.59) reads

$$\begin{aligned}b_2 &\cong \frac{\pi}{\rho^6} [\Le \pi(2\pi + 1)\tilde{R}^2 + (\pi^2 - 2)R_{c_2}^2 \\ &\quad - (2\pi^2 - \pi - 2 + \Le \pi(\pi + 2))\tilde{R}R_{c_2}] \\ &\cong \frac{\pi}{\rho^6} [-(1 - \Le)\pi(\pi - 1)\tilde{R}^2 + (\pi^2 - 2)\eta_c^2 \\ &\quad + (\pi - 2 - \Le \pi^2 - 2\Le \pi)\tilde{R}\eta_c] \\ &\cong \frac{\pi}{\rho^6} [-\pi(\pi - 1)\tilde{R}^2 + (\pi^2 - 2)\eta_c^2 + \tilde{R}\eta_c].\end{aligned}\tag{5.3.87}$$

Then Theorem 5.3.7 is rewritten as

Physical Conclusion 5.3.11 Let $\tilde{R} > 2.35 \times 10^{20}$, and b_2 be the parameter as in (5.3.87). Then for the problem (5.3.69) with (5.3.8), Assertions (1) and (2) of Theorem 5.3.7 hold true.

The above Physical Conclusions 5.3.9–5.3.11 provide the possible dynamical behaviors for the great ocean conveyer, depending on the saline Rayleigh number \tilde{R} .

We note that the temperature and salinity gradients $T_0 - T_1$ and $S_0 - S_1$ between the oceanic bottom and upper surfaces are different from the observed data. In fact, the observed values should be as follows

$$\begin{aligned}\Delta T &= T_0 - T_1 + T(x_1, x_2, 0) - T(x_1, x_2, 1), \\ \Delta S &= S_0 - S_1 + S(x_1, x_2, 0) - S(x_1, x_2, 1),\end{aligned}$$

where $T(x)$ and $S(x)$ are the transition solutions as described by Assertion (3) of Theorem 5.3.5. Due to the heat resources coming from the earth's crust, the bottom temperature T_0 of the ocean retains essentially a constant, and the bottom salinity S_0 is also a constant which equals to the average value

$$S_0 = \frac{1}{|V|} \int_V S(x) dx, \quad |V| \text{ the volume of the ocean.}$$

It is easy to see that on the upper surface, the temperature and salinity vary in different regions. In the equator, T is about $20\text{--}40^\circ\text{C}$ and near the Poles, T is 2 to -1.8°C . Here, T_1 takes an average on the upper surface. Since the elevated rate of evaporation in the tropical areas and the freezing of polar sea ice, which leave the salt behind in the remaining seawater, the densities of tropical and polar water are quite high. Hence, as an average, we determine that

$$T_0 - T_1 \leq 0, \quad S_0 - S_1 < 0. \quad (5.3.88)$$

By Physical Conclusion 5.3.9 and (5.3.88), the saline Rayleigh number $\tilde{R} < 0$. Therefore the transition of (5.3.69) with (5.3.8) is from real eigenvalues. In addition, by $\tilde{R} < 0$, we see that the number in (5.3.86) satisfies that $b_1 > 0$. Thus, by Physical Conclusion 5.3.10, the dynamical behavior of the THC is a Type-I transition to a pair of stable equilibrium states provided

$$\sigma = R - Le^{-1}\tilde{R} > \sigma_c,$$

which, by (5.3.75), (5.3.76), and (5.3.83), is equivalent to

$$0.86 \times 10^{21}(T_0 - T_1) + 3.75 \times 10^{23}(S_1 - S_0) > 2.33 \times 10^{24}. \quad (5.3.89)$$

For the large scale ocean circulation, $T_0 - T_1 < 100^\circ\text{C}$. Hence, (5.3.89) shows that if $S_1 - S_0 = O(10)$, then the great ocean conveyer is driven by the doubly diffusive convection.

In Assertion (1) of Physical Conclusion 5.3.10, the steady-state solutions describing the great Conveyer are approximatively expressed by

$$v^\pm = \begin{cases} \pm C\beta^{1/2}(\sigma)L_1 \sin \frac{\pi x_1}{L_1} \cos \pi x_3, \\ \mp C\beta^{1/2}(\sigma) \cos \frac{\pi x_1}{L_1} \sin \pi x_3, \end{cases} \quad (5.3.90)$$

$$T^\pm = T_0 + (T_1 - T_0)x_3 \mp \frac{C\beta^{1/2}(\sigma)}{\alpha_c^2 + \pi^2} \cos \frac{\pi x_1}{L_1} \sin \pi x_3, \quad (5.3.91)$$

$$S^\pm = S_0 + (S_1 - S_0)x_3 \mp \frac{\text{sign}(S_0 - S_1)C\beta^{1/2}}{\text{Le}(\alpha_c^2 + \pi^2)} \cos \frac{\pi x_1}{L_1} \sin \pi x_3, \quad (5.3.92)$$

where $C > 0$ is a constant, $\beta(\sigma)$ is the first real eigenvalue, and by Theorem 5.3.5, $\beta(\sigma)$ can be expressed as

$$\beta(\sigma) = k(\sigma - \sigma_c) + o(|\sigma - \sigma_c|) \quad (5.3.93)$$

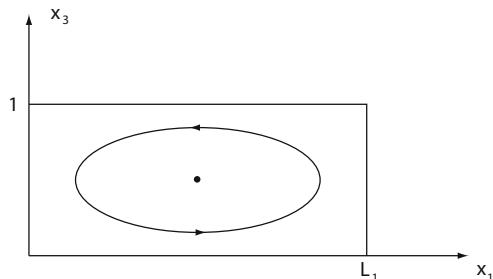
for some constant $k > 0$.

In oceanic dynamics, the Rayleigh number $\sigma = R - \text{Le}^{-1}\widetilde{R}$ is a main driving force for the great ocean conveyer, (5.3.90) and (5.3.93) show that the velocity v of the oceanic circulation is proportional to $\sqrt{\sigma - \sigma_c}$.

The velocity field v^+ given by (5.3.90) is topologically equivalent to the structure as shown in Fig. 5.12, which is consistent with the real oceanic flow structure.

Assertion (3) in the Physical Conclusion 5.3.10 shows also that the theoretical results are in agreement with the real thermohaline ocean circulation.

Fig. 5.12 The flow structure described by v^+ in (5.3.90)



At $x_3 = 1/2$, the temperature T^+ and salinity S^+ given by (5.3.91) and (5.3.92) are as follows

$$T_{1/2}^+(x_1) = \frac{1}{2}(T_0 + T_1) - \frac{C\beta^{1/2}}{\alpha_c^2 + \pi^2} \cos \frac{\pi x_1}{L_1},$$

$$S_{1/2}^+(x_1) = \frac{1}{2}(S_0 + S_1) + \frac{C\beta^{1/2}}{\text{Le}(\alpha_c^2 + \pi^2)} \cos \frac{\pi x_1}{L_1}.$$

The distributions $T_{1/2}^+(x_1)$ and $S_{1/2}^+(x_1)$ are illustrated in Figs. 5.13 and 5.14 respectively. In particular, if $x_1 = 0$ stands for the North Pacific, $x_1 = L_1$ for the North Pacific, and $x_3 = 1/2$ for the deep basin of the ocean, then the profiles of $T_{1/2}^+(x_1)$

and $S_{1/2}^+(x_1)$ confirm with observations, wherein the North Atlantic Deep Water is cold and salty water, and the North Atlantic Deep Water is warmer and fresher water.

Fig. 5.13 The temperature profile in deep water of the ocean

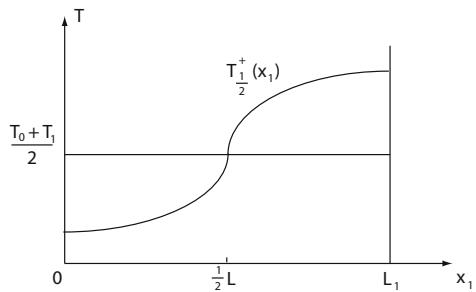
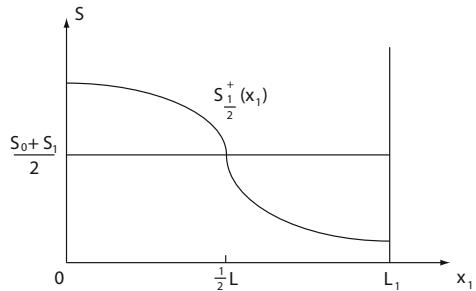


Fig. 5.14 The salinity profile in deep water of the ocean



Physical Conclusions 5.3.9 and 5.3.10 show that although the THC is mainly due to a continuous transition to multiple equilibria, analysis for other possible dynamical behaviors is still interesting.

We can see that when \tilde{R} satisfies

$$\frac{Le^3}{1 - Le^2}\sigma_c < \tilde{R} < \frac{Le^2}{1 - Le}\sigma_c, \quad (5.3.94)$$

the number b_1 in (5.3.86) is less than zero, i.e., $b_1 < 0$, then due to Physical Conclusions 5.3.9 and 5.3.10, the transition is a jump transition, leading to a saddle-node oscillation.

When $\tilde{R} > \frac{Le^2}{1 - Le}\sigma_c = 2.35 \times 10^{20}$, the problem undergoes a dynamic transition to a periodic solution, as $\eta > \eta_c$. This transition can be either Type-I, or Type-II, described as follows:

First, if $\tilde{R} > 2.35 \times 10^{24}$, the number b_2 in (5.3.87) is less than zero, and by Physical Conclusion 5.3.11, this system undergoes a Type-I transition to a stable periodic solution as $\eta = R - \tilde{R} > \eta_c$, its velocity field is written as

$$u(x, t) = \begin{cases} -\left(\frac{2\pi\lambda(\eta)}{|b_2|}\right)^{1/2} L_1 \sin \rho_0 t \sin \frac{\pi x_1}{L_1} \cos \pi x_3, \\ \left(\frac{2\pi\lambda(\eta)}{|b_2|}\right)^{1/2} \sin \rho_0 t \cos \frac{\pi x_1}{L_1} \sin \pi x_3, \end{cases} \quad (5.3.95)$$

where $\lambda(\eta) = \text{Re}\beta_1(\eta)$, $\eta = R - \tilde{R}$, and ρ_0 is as in (5.3.87).

Second, if $2.35 \times 10^{20} < \tilde{R} < 2.35 \times 10^{24}$, the number $b_2 > 0$. Hence the transition is jump to a stable periodic solution at $\eta = \eta_c$, leading to an oscillation between a time-periodic solution and the trivial equilibrium near η_c .

Finally, some remarks for the oceanic thermohaline circulation are in order.

First, the condition (5.3.88) provides a more realistic scenario, where \tilde{R} satisfies that $\tilde{R} < \frac{\text{Le}^3}{1-\text{Le}^2}\sigma_c$. In this case the circulation is a continuous transition to multiple stable equilibria, with flow structure as illustrated in Fig. 5.12, and with the temperature and salinity profiles for deep water of the ocean as shown in Figs. 5.13–5.14. This is consistent with observations.

Second, it is known that the velocity takes κ/h as its unit. Thus, the maximal value of v_1 in (5.3.90) is that

$$v_{\max} = \delta^{1/2} \beta^{1/2} \kappa h^{-1} L_1,$$

where $\delta = \delta(\sigma_c)$ is as in (5.3.57), and $\beta = \beta(\sigma)$ is as in (5.3.93).

The ratio between the vertical and horizontal velocities is given by

$$\frac{v_3}{v_1} = \frac{1}{L_1} = 1.56 \times 10^{-4}.$$

As a contrast, the oceanic circulation takes about 1600 years for its journey, i.e., the real velocity is also very small.

Third, the condition (5.3.94) would lead to a saddle-node bifurcation to metastable states. However, these states have not been observed in oceanography.

Fourth, if $\tilde{R} > 2.35 \times 10^{23}$, then the circulation is time-periodic, and we see from (5.3.95) that the period is

$$\tau = \frac{2\pi}{\rho_0} \cdot \frac{h^2}{\kappa} = 1.1 \times 10^6 \text{ s},$$

where ρ_0 is as in (5.3.87), and $\rho_0 \approx 6.5 \times 10^8$. This period is about 4 months. It is not realistic.

Fifth, when $2.35 \times 10^{20} < \tilde{R} < 2.35 \times 10^{24}$, the oceanic system undergoes a time-periodic oscillation phenomenon. However, this behavior has not been observed in a realistic oceanic regime.

5.4 Large Scale Meridional Atmospheric Circulation

5.4.1 Polar, Ferrel, and Hadley Cells

There are three primary large scale meridional atmospheric circulation cells: the Polar, the Ferrel, and the Hadley cells. Note that the Hadley cell may contain several cells within the equatorial zone which may shift, merge, and decouple in a complicated process over time. However, the Hadley cell is referred as the unique large-scale meridional cell.

The Polar cell is a relatively simple system. Warm air rises at lower latitudes and moves poleward through the upper troposphere at both the north and south poles. When the air reaches the polar areas, it has cooled considerably, and descends as a cold, dry high pressure area, moving away from the pole along the surface but twisting west-ward as a result of the Coriolis effect to generate the Polar easterlies.

The Ferrel cell is located between the Polar cell and the Hadley cell, and comes about as a result of the eddy circulations of the mid-latitudes. At its southern extent, it overrides the Hadley cell, and at its northern extent, it overrides the Polar cell. Just as the trade wind can be found beneath the Hadley cell, the Westerlies can be found below the Ferrel cell.

The Hadley cell provides an explanation for the trade winds and matches fairly with observations. It is a closed circulation loop, which begins at the equator with warm, moist air raised in equatorial low pressure areas to the tropopause and carried poleward. At about 30°N/S latitude, it descends in a cooler high pressure area. Part of the descending air travels towards the equator along the surface, closing the loop of the Hadley cell and creating the trade winds.

The meridional circulation is very different from the longitudinal circulation. In the equatorial motion equation (5.2.3), we see that the Coriolis force ($-\omega u_2, \omega u_1$) can be balanced by a potential term, playing the same role as the pressure gradient (5.2.16) so that the Coriolis effect has little influence for the equatorial circulation. However, in the latitudinal regions, the rotation effect of the earth is crucial:

- The Coriolis force deflects the poleward moving air to the east and the equator ward moving air to the west, which complicates the atmospheric motion, and
- The rotating effect reduces the convective scale, leading to an increase in the cells, i.e., there are twice as many loops in the latitudinal circulation as in the equatorial circulation.

Also the surface temperature gradient/deviation in the meridional direction is an order one term, and cannot be treated as a small perturbation to the average temperature gradient $\Delta T = T_0 - T_1$. This large surface temperature gradient leads to a strong climate variance, and to substantial difficulties for mathematical treatments.

The main objective of this section is to study dynamic transitions and pattern formations associated with polar, the Ferrel, and the Hadley cells.

5.4.2 β -plane Assumption

Consider the Boussinesq equations on a hemispheric region with the β -plane assumption (Pedlosky, 1987; Salby, 1996). Namely, let

$$(x_1, x_2, x_3) = \left(r_0 \varphi \cos \theta_0, r_0 \left(\frac{\pi}{2} - \theta \right), z \right),$$

for some $0 < \theta_0 < \frac{\pi}{2}$, and the northern hemispheric region D becomes a rectangle:

$$D = (0, 2\pi r_0) \times (0, \pi r_0 / 2) \times (r_0, r_0 + h).$$

The velocity field $(u_\varphi, u_\theta, u_r)$ becomes

$$u_1 = u_\varphi, u_2 = -u_\theta, u_3 = u_z.$$

The Coriolis force takes the following form:

$$2\Omega k \times u = (-2\Omega \sin \theta_0 u_2, 2\Omega \sin \theta_0 u_1, 0).$$

With the above considerations, the Boussinesq model takes the following form (Pedlosky, 1987; Salby, 1996):

$$\begin{aligned} \frac{\partial u_1}{\partial t} + (u \cdot \nabla) u_1 &= \nu \Delta u_1 - \sigma_0 u_1 - f u_2 - \frac{1}{\rho_0} \frac{\partial p}{\partial x_1}, \\ \frac{\partial u_2}{\partial t} + (u \cdot \nabla) u_2 &= \nu \Delta u_2 - \sigma_0 u_2 + f u_1 - \frac{1}{\rho_0} \frac{\partial p}{\partial x_2}, \\ \frac{\partial u_3}{\partial t} + (u \cdot \nabla) u_3 &= \nu \Delta u_3 - \sigma_1 u_3 - (1 - a(T - T_0))g - \frac{1}{\rho_0} \frac{\partial p}{\partial x_3}, \\ \frac{\partial T}{\partial t} + (u \cdot \nabla) T &= \kappa \Delta T, \\ \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} &= 0, \end{aligned} \quad (5.4.1)$$

where $f = 2\Omega \sin \theta_0$, and $\Delta, (u \cdot \nabla)$ are the usual operators in \mathbb{R}^3 .

Let T_1 be the temperature at the upper boundary $x_3 = r_0 + h$, and T_0 be the average temperature at the lower boundary $x_3 = r_0$, where T_1 and T_0 are constants. Take the standard nondimensional form as used in (5.2.3), the equations (5.4.1) for the deviation of the motionless state with linear temperature profile become

$$\begin{aligned} \frac{\partial u}{\partial t} &= \text{Pr} (\Delta u + RTk - \delta u - \nabla p) + \omega k \times u - (u \cdot \nabla) u, \\ \frac{\partial T}{\partial t} &= \Delta T + u_3 - (u \cdot \nabla) T, \\ \text{div} u &= 0, \end{aligned} \quad (5.4.2)$$

where $k = (0, 0, 1)$, $\delta u = (\delta_0 u_1, \delta_0 u_2, \delta_1 u_3)$, and

$$\omega = 2\Omega \sin \theta_0 h^2 / \kappa, \quad \delta_i = \frac{h^2}{\nu} \sigma_i, \quad (i = 0, 1).$$

The boundary conditions in the (x_1, x_2) -plane are periodic

$$(u, T)(x_1, x_2) = (u, T)(x_1 + 2\pi j r_0, x_2 + k L), \text{ for } j, k \in \mathbb{Z}, L = 2L_0 = \pi r_0. \quad (5.4.3)$$

At the pole $(x_1, x_2) = 0$, no horizontal flows appear, and we have to impose an additional condition:

$$u_1(0, 0, x_3) = u_2(0, 0, x_3) = 0. \quad (5.4.4)$$

For convenience, at the top and bottom boundary, we only consider the free-free boundary condition

$$\begin{aligned} u_3 &= 0, & \frac{\partial u_1}{\partial x_3} = \frac{\partial u_2}{\partial x_3} &= 0 & \text{at } x_3 = r_0, r_0 + 1, \\ T &= 0 & & & \text{at } x_3 = r_0 + 1, \\ T &= \varphi(x_1, x_2) & & & \text{at } x_3 = r_0, \end{aligned} \quad (5.4.5)$$

where $\varphi(x_1, x_2)$ is the temperature deviation satisfying that

$$\int_0^{2\pi r_0} \int_0^{\frac{\pi}{2} r_0} \varphi(x_1, x_2) dx_2 dx_1 = 0.$$

Namely, the real temperature distribution on the surface of the earth is $T = T_0 + \varphi(x_1, x_2)$. Different from that on the equator, here $\varphi(x_1, x_2)$ is not small in comparison with $\Delta T = T_0 - T_1$.

5.4.3 Meridional Circulation Under Idealized Condition

Model and Physical Parameters Consider the idealized case that the temperature distribution on the surface $x_3 = r_0$ is a constant, i.e.

$$\varphi(x_1, x_2) = 0 \quad \forall (x_1, x_2) \in (0, 2\pi r_0) \times (0, \frac{\pi}{2} r_0).$$

Based on Table 5.1, at $T_0 \cong 0^\circ\text{C}$, the main parameters of air are taken as follows

$$\begin{aligned} \nu &= 1.3 \times 10^{-5} \text{ m}^2/\text{s}, & \kappa &= 1.8 \times 10^{-5} \text{ m}^2/\text{s}, \\ a &= 3.7 \times 10^{-3} / \text{k}, & \text{Pr} &= 0.72, \\ \Omega &= 0.73 \times 10^{-4}, & h &= 8 \times 10^3 \text{ m}, \\ \sin \theta_0 &= 1/2, & \omega^2 &= 4\Omega^2 h^4 / \kappa^2 = 6.325 \times 10^{16}. \end{aligned} \quad (5.4.6)$$

Then the Rayleigh number is expressed as

$$R = \frac{g \cdot \alpha(T_0 - T_1)}{\kappa\nu} h^3 = 7.9 \times 10^{19}(T_0 - T_1). \quad (5.4.7)$$

We take δ_0 and δ_1 as in (5.2.31). Then by (4.5.21)–(4.5.23), we infer from (5.4.6) to (5.4.7) the following theoretical values of the critical length scale L_c of the circulation cells, the critical Rayleigh number R_c , and the critical temperature gradient ΔT_c :

$$L_c \cong \left[\frac{\delta_1}{\delta_0 + \frac{\omega^2}{Pr^2 \delta_0}} \right]^{1/4} \cdot h = 5000 \text{ km}, \quad (5.4.8)$$

$$R_c \cong \pi^2 \delta_1 + \frac{\pi^4 \omega^2}{\alpha_c^2 Pr^2 \delta_0} \cong 2.7 \times 10^{21}, \quad (5.4.9)$$

$$\Delta T_c = T_0 - T_1 \cong 34^\circ C. \quad (5.4.10)$$

It is known that the scale of the Polar, Ferrel, and Hadley cells is

$$L_0 = \frac{\pi}{6} r_0 \cong 3300 \text{ km}. \quad (5.4.11)$$

The average surface temperature $T_0 = 0^\circ C$ – $10^\circ C$, and the tropopause temperature $T_1 \cong -70^\circ C$. Hence, the real temperature gradient is about

$$\Delta T_c = 70^\circ C \sim 80^\circ C. \quad (5.4.12)$$

Linearized Eigenvalue Problem Consider the eigenvalue problem

$$\begin{aligned} & Pr(\Delta u + RTk - \delta u - \nabla p) + \omega k \times u = \beta u, \\ & \Delta T + u_3 = \beta T, \\ & \operatorname{div} u = 0, \end{aligned} \quad (5.4.13)$$

with boundary conditions (5.4.1)–(5.4.3).

The eigenvalues of (5.4.12) can be expressed in the form of (4.4.9) and (4.4.10), and the functions $(u_{jk}, v_{jk}, W_{jk}, T_{jk}, p_{jk})$ satisfy the ordinary differential equations

$$Pr(D_{jk}u_{jk} - \delta_0 u_{jk} + j\alpha_1 p_{jk}) - \omega v_{jk} = \beta u_{jk}, \quad (5.4.14)$$

$$Pr(D_{jk}v_{jk} - \delta_0 v_{jk} + k\alpha_2 p_{jk}) + \omega u_{jk} = \beta v_{jk}, \quad (5.4.15)$$

$$Pr(D_{jk}W_{jk} - \delta_1 w_{jk} - Dp_{jk} + RT_{jk}) = \beta W_{jk}, \quad (5.4.16)$$

$$D_{jk}T_{jk} + W_{jk} = \beta T_{jk}, \quad (5.4.17)$$

$$j\alpha_1 u_{jk} + k\alpha_2 v_{jk} + DW_{jk} = 0, \quad (5.4.18)$$

with the boundary condition

$$Du_{jk} = Dv_{jk} = W_{jk} = T_{jk} = 0, \quad \text{at } x_3 = 0, 1, \quad (5.4.19)$$

where D_{jk} , D , α_{jk}^2 are the same as in (4.4.14).

From (5.4.14), (5.4.15), and (5.4.18) we obtain

$$\begin{aligned} -(\Pr D_{jk} - \Pr \delta_0 - \beta)DW_{jk} + \Pr \alpha_{jk}^2 p_{jk} &= \omega(j\alpha_1 v_{jk} - k\alpha_2 u_{jk}), \\ -(\Pr D_{jk} - \Pr \delta_0 - \beta)(j\alpha_1 v_{jk} - k\alpha_2 u_{jk}) &= -\omega DW_{jk}. \end{aligned}$$

It follows that

$$(\Pr D_{jk} - \Pr \delta_0 - \beta)^2 DW_{jk} - \Pr \alpha_{jk}^2 (\Pr D_{jk} - \Pr \delta_0 - \beta) p_{jk} + \omega^2 DW_{jk} = 0. \quad (5.4.20)$$

Then, from (5.4.16), (5.4.17), and (5.4.20) we infer that W_{jk} satisfies the ordinary differential equation

$$\begin{aligned} &[\alpha_{jk}^2 (\Pr D_{jk} - \Pr \delta_0 - \beta) (\Pr D_{jk} - \Pr \delta_1 - \beta) (D_{jk} - \beta) \\ &-(\Pr D_{jk} - \Pr \delta_0 - \beta)^2 (D_{jk} - \beta) D^2 - \omega^2 (D_{jk} - \beta) D^2 \\ &-\alpha_{jk}^2 \Pr R (\Pr D_{jk} + \Pr \delta_0 - \beta)] W_{jk} = 0. \end{aligned} \quad (5.4.21)$$

From (5.4.19) and (5.4.14)–(5.4.18) we deduce that

$$W_{jk} = D^2 W_{jk} = D^4 W_{jk} = D^6 W_{jk} = 0. \quad (5.4.22)$$

It is clear that the solutions of (5.4.21)–(5.4.22) are sine functions $W_{jk} = \sin l\pi x_3 (l \in \mathbb{Z})$. Inserting $W_{jk} = \sin \pi x_3$ into (5.4.21) we obtain the equation (4.5.19) that the first eigenvalues β of (5.4.13) satisfy.

Similar to (4.4.18)–(4.4.20), u_{jk} , v_{jk} and T_{jk} satisfy

$$\begin{aligned} (\Pr D_{jk} - \Pr \delta_0 - \beta) u_{jk} &= -\frac{j\alpha_1 \pi}{\alpha_{jk}^2} (\Pr D_{jk} - \Pr \delta_0 - \beta + \frac{k\alpha_2}{j\alpha_1} \omega) DW_{jk}, \\ (\Pr D_{jk} - \Pr \delta_0 - \beta) v_{jk} &= -\frac{k\alpha_2 \pi}{\alpha_{jk}^2} (\Pr D_{jk} - \Pr \delta_0 - \beta - \frac{j\alpha_1}{k\alpha_2} \omega) DW_{jk}, \end{aligned} \quad (5.4.23)$$

$$(D_{jk} - \beta) T = -W_{jk}.$$

Let $W_{jk} = \sin \pi x_3$. Then, from (5.4.23) we get the eigenvectors of (5.4.13) corresponding to eigenvalue β as follows

$$\psi_{jk}^\beta = \begin{cases} u_{jk}^\beta \sin 2\pi(j\alpha_1 x_1 + k\alpha_2 x_2) \cos \pi x_3, \\ v_{jk}^\beta \sin 2\pi(j\alpha_1 x_1 + k\alpha_2 x_2) \cos \pi x_3, \\ \cos 2\pi(j\alpha_1 x_1 + k\alpha_2 x_2) \sin \pi x_3, \\ T_{jk}^\beta \cos 2\pi(j\alpha_1 x_1 + k\alpha_2 x_2) \sin \pi x_3, \end{cases} \quad (5.4.24)$$

where $\alpha_1 = (2\pi r_0)^{-1}$, $\alpha_2 = L^{-1}$, and

$$(u_{jk}^\beta, v_{jk}^\beta, T_{jk}^\beta) = \left(-\frac{\pi^2}{\alpha_{jk}^2} (j\alpha_1 - k\alpha_2 \omega A_1(\beta)), -\frac{\pi^2}{\alpha_{jk}^2} (k\alpha_2 + j\alpha_1 \omega A_1(\beta)), A_2(\beta) \right),$$

$$A_1(\beta) = \frac{1}{\Pr r_{jk}^2 + \Pr \delta_0 + \beta}, \quad A_2(\beta) = \frac{1}{r_{jk}^2 + \beta}, (r_{jk}^2 = \alpha_{jk}^2 + \pi^2).$$

Physically, (5.14) requires that $\alpha_{jk}^2 = k^2 \pi^2 / L_0^2$, i.e., $(j, k) = (0, k)$. In each period there are two cells, and the size of every cell in the meridional direction is about $\pi r_0 / 6$. Thus, when the length scale L in (5.4.3) takes $\pi r_0 / 3$, for the first eigenvectors, the integer pair (j, k) takes $(j, k) = (0, 1)$ in (5.4.22). Then, we obtain the first eigenvalue of (5.4.13) at R_c as

$$\psi_1 = \begin{cases} L\omega \Pr^{-1}(r^2 + \delta_0)^{-1} \sin \frac{2\pi x_2}{L} \cos \pi x_3, \\ -L \sin \frac{2\pi x_2}{L} \cos \pi x_3, \\ \cos \frac{2\pi x_3}{L} \sin \pi x_3, \\ r^{-2} \cos \frac{2\pi x_3}{L} \sin \pi x_3, \end{cases} \quad (5.4.25)$$

where $L = 2L_0 = \frac{\pi}{3}r_0$.

Dynamic Transitions It is ready to check that Theorems 4.4.9 and 4.4.10 are still valid for (5.4.2)–(5.4.5), only replacing b_1 by the following parameter

$$b = \Pr^4(r^2 + \delta_0)^2 R_{c_1} - \pi^4 \omega^2 \gamma^4 / \alpha^4, \quad (5.4.26)$$

where $\alpha^2 = \pi^2 L^{-2}$, $\gamma^2 = \alpha^2 + \pi^2$, and R_{c_1} is as in (4.5.20).

Theorem 5.4.1 Let R_{c_1} and R_{c_2} be given by (4.5.20) and (4.5.24), and b be the parameter as in (5.4.26). If $R_{c_1} < R_{c_2}$, then the following assertions hold true:

- 1 When $b > 0$ the system (5.4.2)–(5.4.5) undergoes a continuous dynamic transition at $R = R_{c_1}$, and bifurcates to exactly two steady-state solutions ψ_R^+ and ψ_R^- for $R > R_{c_1}$, which are attractors and can be expressed as

$$\psi_R^\pm = \pm C \beta_1^{1/2}(R) \psi_1 + o(|\beta_1|^{1/2}), \quad (5.4.27)$$

where ψ_1 is as in (5.4.25), $C > 0$ is a constant, and $\beta_1(R)$ is the first eigenvalue of (5.4.13).

- 2 When $b < 0$, the system undergoes a jump transition, and has exactly two saddle-node bifurcation points (ψ_i^*, R^*) and (ψ_i^*, R^*) with $R^* < R_{c_1}$ and $\psi_i^* \neq 0$ ($i = 1, 2$).

Also, Theorem 4.4.12 holds true for the revised problem (5.4.2)–(5.4.5). The parameter b_2 is still given by (4.4.73), with the following modifications:

$$R_1 = \frac{\Pr \gamma^2 + \Pr \delta_0}{\Pr^2(\gamma^2 + \delta_0)^2 + a^2}, \quad I_1 = \frac{-a}{\Pr^2(\gamma^2 + \delta_0)^2 + a^2},$$

$$\begin{aligned} a^2 = & \frac{1}{\gamma^2} [\pi^2 \omega^2 - \alpha^2 \text{Pr} R_{c_2} + \alpha^2 \text{Pr}^2 (r^2 + \delta_0)(r^2 + \delta_1) + \alpha^2 \text{Pr} \gamma^2 (\gamma^2 + \delta_1) \\ & + \alpha^2 \text{Pr} r^2 (r^2 + \delta_0) + \pi^2 \text{Pr}^2 (r^2 + \delta_0)^2 + 2\pi^2 \text{Pr} r^2 (r^2 + \delta_0)], \end{aligned}$$

where $r^2 = \alpha^2 + \pi^2$, and α^2 is as in (4.5.24).

5.4.4 Physical Implications

With an idealized model, we derive some physical conclusions. First, the predicated critical length scale and the critical temperature gradient of the large scale meridional atmospheric circulation is in gross agreement with observations. The slight differences of these values are due essentially to highly idealization of the model.

Second, set typical parameters in the revised problem (5.4.2)–(5.4.5) as:

$$\delta_0 = 3.5 \times 10^8, \quad \delta_1 = 2.7 \times 10^{20}, \quad \text{Pr} = 0.72.$$

By (4.5.21), (4.5.20), (4.5.24), and (5.4.26), we obtain that

$$\begin{aligned} \alpha_c^2 &\cong \pi^2 \delta_0^{1/2} / \delta_1^{1/2} = 1.16 \times 10^{-5}, \quad R_{c_1} \cong \pi^2 \delta_1 = 2.7 \times 10^{21}, \\ R_{c_2} &\cong \text{Pr}^2 \alpha_c^2 \delta_0 \delta_1 = 5.5 \times 10^{23}, \quad b \cong \text{Pr}^4 \delta_0^2 R_{c_1} - \pi^8 \omega^2 / \alpha_c^4 = 8.26 \times 10^{37}. \end{aligned}$$

Hence Theorem 5.4.1 indicates that the system (5.4.2)–(5.4.5) undergoes a continuous dynamic transition at $R = R_{c_1}$, and the polar, the Ferrel, and the Hadley cells are essentially depicted by two local attractors given by (5.4.27). In particular, the leading order approximation of the velocity field $u = (u_1, u_2, u_3)$ governing the roll structure of the meridional circulation is expressed by

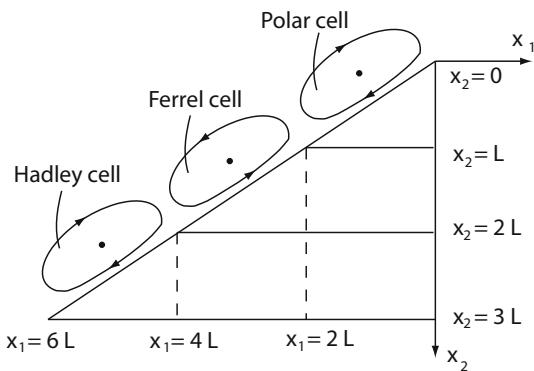
$$u \approx \left(\frac{L\omega}{\text{Pr}(\gamma^2 + \delta_0)} \sin \frac{2\pi x_2}{L} \cos \pi x_3, -L \sin \frac{2\pi x_2}{L} \cos \pi x_3, \cos \frac{2\pi x_3}{L} \sin \pi x_3 \right).$$

The easterlies and westerlies are characterized by u_1 , and the ratio between the zonal and meridional components is

$$\frac{u_1}{u_2} = -\frac{\omega}{\text{Pr}(\delta_0 + r^2)} \cong -2.$$

Hence along a straight line $x_2 = \frac{1}{2}x_1 + C$ (C is a constant), three closed loops of u in a hemisphere are formed in interval $kL < x_2 < (k+1)L$ for $k = 0, 1, 2$ and $L = \frac{\pi}{6}r_0$. They stand for the Polar cell for $0 < x_2 < L$, the Ferrel cell for $L < x_2 < 2L$, and the Hadley cell for $2L < x_2 < 3L$; see Fig. 5.15. We remark here that type of transitions derived in this section is independent of the added turbulent friction terms, although these friction terms are crucial for deriving proper critical convections of the large scale circulation.

Fig. 5.15 The Polar, Ferrel, and Hadley cells



Finally, the study here can be used for the atmospheric circulation on other planets than the earth. In particular, it is easy to derive from the theorems in this section that if the radius r_0 of the planet is sufficiently large, then the system can undergo a dynamic transition, leading to spatiotemporal oscillations.



Chapter 6

Dynamical Transitions in Chemistry and Biology

Chemical reaction systems and biological models were among the first dissipative systems described by the Belgium school; see, e.g., Prigogine and Lefever (1968), Glansdorff and Prigogine (1971). Among many other features, dissipative systems in nature demonstrate self-organized and self-assembled structures; see, for example, Nicolis and Prigogine (1977), Kapral and Showalter (1995), Pismen (2006), Desai and Kapral (2009), Cross and Hohenberg (1993), Swinney et al. (1990), Murray (2002).

This chapter studies dynamic transitions of a few typical nonequilibrium chemical and biological models. The main focus is on Belousov–Zhabotinsky models (Oregonator and Brusselator), the chemotactic model, and a population model. As in the previous chapters, a crucial component of our study is based on the general principle, Principle 1, following the Philosophy 2. Namely we classify dynamic transitions as continuous, catastrophic, or random, determine their patterns by searching for a complete set of transition states, whereby the chemical and biological implications of our rigorous analysis become transparent.

In all these models, linear analysis provides the onset of the dynamic transition, and nonlinear analysis offers precise information on the types of transitions and the phase diagrams in terms of typical biological and chemical parameters as well as the spatial geometry of the domain. As we know, catastrophic (jump) and random (mixed) transitions are far-from-equilibrium states. Hence the study here provides a precise characterization of the conditions and mechanisms of such far-from-equilibrium states. This has always been a difficult subject due to its complexity and global nature.

Pattern formation is an important topic in the study of nonequilibrium phase transitions. In this chapter, we focus mainly on dynamic transitions and touch only briefly on pattern formation. However, as mentioned earlier, our approach provides a new avenue for studying the formation and mechanisms of important patterns appearing in nature.

6.1 Modeling

6.1.1 Dynamical Equations of Chemical Reactions

Fick's Law of Diffusion If u stands for the density or temperature, then by the Fick law or the Fourier law, the flux of density or temperature is $J = -k\nabla u$, where k is the diffusion (or conduction) coefficient. The rate of change of density or temperature is then given by

$$\frac{\partial u}{\partial t} = -\operatorname{div} J = k\Delta u. \quad (6.1.1)$$

Thus Δu represents the diffusion. The heterogeneous distribution of density and temperature in $\Omega \subset \mathbb{R}^3$ implies that there exists a potential energy in some generalized sense, and the Laplace operator ∇^2 plays the role of reducing potential energy by diffusion.

Reaction Process We begin with an example. Consider the chemical reaction of synthetic ammonia as follows



where H_2 is the hydrogen, N_2 is the nitrogen, NH_3 is the ammonia, and k_1 is the synthetic coefficient, k_2 is the decomposition coefficient. For simplicity, we denote H_2 by X_1 , N_2 by X_2 and NH_3 by X_3 . Then the reaction equation (6.1.2) can be expressed in the abstract forms



Let u_i be the concentrations of X_i ($1 \leq i \leq 3$). For the synthetic reaction (6.1.3), the change rate of u_1 is proportional to the probability of collision of one X_2 and three X_1 , i.e.,

$$\frac{\delta_1 u_1(x)}{\delta t} = CP(3X_1 \cap X_2, x), \quad x \in \Omega \subset \mathbb{R}^3,$$

where $C \in \mathbb{R}^1$ is a constant, $P(3X_1 \cap X_2, x)$ represents the probability density of one X_1 and three X_2 at $x \in \Omega$. It is known that

$$P(3X_1 \cap X_2, x) = P^3(X_1, x) \cdot P(X_2, x),$$

$P(X_i, x)$ is the probability density of X_i at x . It is clear that

$$P(X_i, x) = Cu_i(x), \quad C > 0 \text{ a constant.}$$

Hence we derive that

$$\frac{\delta_1 u_1(x)}{\delta t} = Cu_1^3 u_2. \quad (6.1.5)$$

Since each synthetic reaction in (6.1.3) loses three X_1 , the proportional constant C in (6.1.5) is $C = -3k_1$. Thus, we get

$$\frac{\delta_1 u_1(x)}{\delta t} = -3k_1 u_1^3 u_2. \quad (6.1.6)$$

Equations (6.1.5) and (6.1.6) are also called the laws of mass action.

For the decomposition reaction (6.1.4), the change rate of u_1 is

$$\frac{\delta_2 u_1(x)}{\delta t} = 3k_2 u_3^2. \quad (6.1.7)$$

Taking the diffusion effect into account, see (6.1.1), the total rate of change of u_1 in (6.1.3) and (6.1.4) is given by

$$\frac{\partial u_1}{\partial t} = \mu_1 \Delta u_1 + \frac{\delta_1 u_1}{\delta t} + \frac{\delta_2 u_1}{\delta t}.$$

Therefore, it follows from (6.1.6) and (6.1.7) that

$$\frac{\partial u_1}{\partial t} = \mu_1 \Delta u_1 - 3k_1 u_1^3 u_2 + 3k_2 u_3^2. \quad (6.1.8)$$

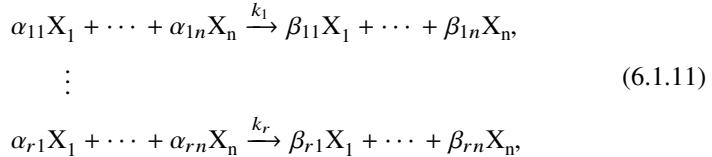
In the same fashion, we can derive the following rate equations of u_2 and u_3 :

$$\frac{\partial u_2}{\partial t} = \mu_2 \Delta u_2 - k_1 u_1^3 u_2 + k_2 u_3^2, \quad (6.1.9)$$

$$\frac{\partial u_3}{\partial t} = \mu_3 \Delta u_3 + 2k_1 u_1^3 u_3 - 2k_2 u_3^2. \quad (6.1.10)$$

Equations (6.1.8)–(6.1.10) describe the dynamical properties of the synthetic reaction and decomposition of ammonia.

General Case Let X_1, \dots, X_n represent n chemical reaction species, and there are r reactions:



where $\alpha_{ij}, \beta_{ij} \geq 0$ are the stoichiometric factors, and k_i ($1 \leq i \leq r$) are the reaction constants. Let u_j represent the concentration of X_j ($1 \leq j \leq n$). Then we derive from (6.1.11) the following chemical reaction-diffusion model:

$$\begin{aligned}\frac{\partial u_1}{\partial t} &= \mu_1 \Delta u_1 + \sum_{i=1}^r (\beta_{i1} - \alpha_{i1}) k_i u_1^{\alpha_{i1}} \cdots u_n^{\alpha_{in}}, \\ &\vdots \\ \frac{\partial u_n}{\partial t} &= \mu_n \Delta u_n + \sum_{i=1}^r (\beta_{in} - \alpha_{in}) k_n u_1^{\alpha_{i1}} \cdots u_n^{\alpha_{in}},\end{aligned}\tag{6.1.12}$$

where μ_j ($1 \leq j \leq n$) are the diffusive coefficients.

Equations (6.1.12) are the model governing the chemical reaction (6.1.11), which are supplemented with the Neumann boundary condition if there is no exchange of materials on the boundary, or the Dirichlet boundary condition if the system permits the flux of X_j ($1 \leq j \leq n$) to pass through the boundary $\partial\Omega$.

When the reaction system (6.1.11) is stirred in Ω , then (6.1.12) reduces to a system of ordinary differential equations as

$$\frac{du_j}{dt} = \sum_{i=1}^r (\beta_{ij} - \alpha_{ij}) k_i u_1^{\alpha_{i1}} \cdots u_n^{\alpha_{in}} \quad \text{for } 1 \leq j \leq n.\tag{6.1.13}$$

We have to impose a natural condition for (6.1.12) and (6.1.13) as follows

$$u \geq 0.\tag{6.1.14}$$

6.1.2 Population Models of Biological Species

Models with No Environmental Effect We consider the population problem of m species. Let X_1, \dots, X_m represent m species which coexist in a spatial domain $\Omega \subset \mathbb{R}^2$, and u_i the population density of X_i ($1 \leq i \leq m$). When we ignore the effect of the environment, the unknown functions u_i satisfy the following ordinary differential equations

$$\frac{du_i}{dt} = r_i u_i + \sum_{j=1}^m a_{ij} u_i u_j + g_i(u) \quad \forall 1 \leq i \leq m,\tag{6.1.15}$$

where $r_i > 0$ is the natural growth rate of X_i , the constant matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mm} \end{pmatrix}\tag{6.1.16}$$

is called the interaction matrix of species, and

$$g(u) = o(|u|^2).\tag{6.1.17}$$

Environmental Effect The environment effect to the population density is given by $\mu\Delta u$, and (6.1.15) should be revised as

$$\frac{\partial u_i}{\partial t} = \mu_i \Delta u_i + \sum_{j=1}^m b_{ij}(x)u_j + r_i u_i + \sum_{j=1}^n a_{ij}u_i u_j + g_i(u), \quad (6.1.18)$$

where $\mu_i > 0$ are the environment constants, and the function matrix

$$B(x) = \begin{pmatrix} b_{11}(x) & \cdots & b_{1m}(x) \\ \vdots & & \vdots \\ b_{m1}(x) & \cdots & b_{mm}(x) \end{pmatrix} \quad (6.1.19)$$

is called the environment factor.

As usual, (6.1.18) are supplemented with either the Dirichlet or the Neumann boundary condition. The Dirichlet boundary condition implies that the environment conditions near the boundary $\partial\Omega$ are not suitable for the species to exist. For example, near the desert, the North and South Pole, many animals cannot live. The Neumann boundary condition describes the species living in water area.

For (6.1.15) and (6.1.18), the nonnegative condition (6.1.14) is also necessary. Hence, the environment factor (6.1.19) is required to satisfy that the first eigenfunctions of the equations

$$\begin{aligned} \mu \cdot \Delta e + B(x)e + re &= \lambda e, \\ e|_{\partial\Omega} = 0 \quad (\text{or } \frac{\partial e}{\partial n}|_{\partial\Omega} = 0), \end{aligned} \quad (6.1.20)$$

are nonnegative in Ω , i.e.,

$$e_i(x) \geq 0, \quad \forall x \in \Omega, \quad e = (e_1, \dots, e_m). \quad (6.1.21)$$

The eigenfunctions $e_i(x)$ essentially determine the distribution of population density u_i . This is why we need the term Bu in (6.1.18), which characterizes the interaction between the species X_i ($1 \leq i \leq m$) and the environment.

The biological significance of the interaction matrix (6.1.16) is as follows:

$a_{ij} \cdot a_{ji} < 0$	the predator-prey relation between X_i and X_j ;
$a_{ij} > 0, a_{ji} > 0$	the mutually beneficial relation between X_i and X_j ;
$a_{ij} < 0, a_{ji} < 0$	the competition relation between X_i and X_j ;
$a_{ij} = a_{ji} = 0$	X_i and X_j are independent of each other;
$a_{ij} = 0, a_{ji} < 0$	X_j is independent of X_i , but X_i is harmful for X_j ; and
$a_{ij} = 0, a_{ji} > 0$	X_j is independent of X_i , but X_i is beneficial to X_j ,

where $i \neq j$. Based on the biological rules, a species with a large population density $u_i \gg 1$ is harmful for itself. Therefore we assume that

$$a_{ii} \geq 0 \quad \text{and} \quad g_i(u)u_i < 0 \quad \forall 1 \leq i \leq m. \quad (6.1.22)$$

In general, the environment factor B is a diagonal matrix. However, we here do not exclude the possibility that $b_{ij} \neq 0$ as $i \neq j$. The conditions (6.1.14), (6.1.21), and (6.1.22) are basic for the coexistence models (6.1.18) of m species.

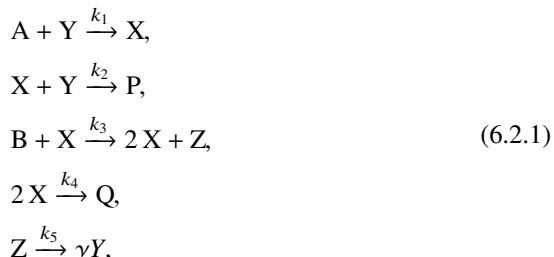
6.2 Belousov–Zhabotinsky Chemical Reactions: Oregonator

In 1950s, in his experiments, B. P. Belousov discovered a spatial-temporal oscillation phenomenon in the concentrations of intermediaries when citric acid was oxidized by acid bromate in the presence of a cerium ion catalyst (Belousov, 1959). It is also observed by Zhabotinski (1964) that organic acids and metal ions could be used as well in the reaction, leading to spatial-temporal oscillations. It has been considered nowadays that all of the chemical reactions give rise to oscillations, and the actions of catalyst are termed as the Belousov–Zhabotinsky (BZ) reactions. BZ reactions are now a classical example of nonequilibrium thermodynamics.

Brusselator was introduced by Prigogine and Lefever (1968). Some numeric and theoretical analyses were made by Glansdorff and Prigogine (1971), and Nicolis and Prigogine (1977), and a linear stability analysis was given by Reichl (1998). The existence of periodic solutions for the Oregonator was first proved by Hastings and Murray (1975). In this section, Theorem 6.2.1 was proved in Ma and Wang (2005b, 2007b), and the other results are from Ma and Wang (2011c,d).

6.2.1 Field-Körös–Noyes Equations

The Model In this section, we study the dynamic phase transitions associated with the spatial-temporal oscillations of the BZ reactions, given by Field et al. (1972). This BZ reaction consists of the following five irreversible steps:



where γ is a stoichiometric factor, P and Q are products which do not join the reaction again, and $X = \text{HBrO}_2$, $Y = \text{Br}^-$, $Z = \text{Ce}^{4+}$, $A = B = \text{BrO}_3^-$. It is the temporal oscillation of the curium ion ratio $\text{Ce(IV)}/\text{Ce(III)}$ which, with a suitable indicator, is displayed by a color change, when the reagent is stirred; see among

others (Field and Noyes, 1974), where a system of ordinary differential equations for this reaction was derived.

Here we consider the general cases. Let u_1, u_2, u_3, a, b represent the concentrations of X, Y, Z, A , and B , with u_i treated as variable functions, and with a and b treated as constants. Then the equations describing the chemical reaction process (6.2.1) are given by

$$\begin{aligned}\frac{\partial u_1}{\partial t} &= \sigma_1 \Delta u_1 + k_1 a u_2 - k_2 u_1 u_2 + k_3 b u_1 - 2k_4 u_1^2, \\ \frac{\partial u_2}{\partial t} &= \sigma_2 \Delta u_2 - k_1 a u_2 - k_2 u_1 u_2 + \gamma k_5 u_3, \\ \frac{\partial u_3}{\partial t} &= \sigma_3 \Delta u_3 + k_3 b u_1 - k_5 u_3,\end{aligned}\quad (6.2.2)$$

where σ_i ($i = 1, 2, 3$) are the diffusivities of u_i , and k_j ($1 \leq j \leq 5$) are the reaction coefficients as in (6.2.1). The model (6.2.2) is also called the Oregonator.

The coefficients σ_i and k_j are functions of the temperature T . In fact, $k_j = k_j^0 e^{-E_j/RT}$, E_j is the activation energy, and R is the Boltzmann constant.

The dimensions of the relevant quantities are given by:

$$\begin{aligned}k_j : M^{-1} t^{-1} \text{ for } 1 \leq j \leq 4, \quad k_5 : t^{-1}, \\ a, b, u_i : M \text{ for } 1 \leq i \leq 3, \quad \sigma_i : l^2 t^{-1} \text{ for } 1 \leq i \leq 3\end{aligned}$$

where t is the time, M is the mole density, and l is the length. Then we introduce the following nondimensional variables:

$$\begin{aligned}u_1 &= \frac{k_1 a}{k_2} u'_1, & u_2 &= \frac{k_3 b}{k_2} u'_2, & u_3 &= \frac{k_1 k_3}{k_2 k_5} a b u'_3, \\ t &= (k_1 k_3 a b)^{-1/2} t', & x &= L x', & \alpha &= \left(\frac{k_3 b}{k_1 a} \right)^{1/2}, \\ \beta &= \frac{2k_1 k_4 a}{k_2 k_3 b}, & \delta &= k_5 (k_1 k_3 a b)^{1/2}, & \mu_i &= \frac{\sigma_i}{l^2 (k_1 k_2 a b)^{1/2}}.\end{aligned}$$

Omitting the primes, we obtain the following nondimensional form of (6.2.2):

$$\begin{aligned}\frac{\partial u_1}{\partial t} &= \mu_1 \Delta u_1 + \alpha(u_1 + u_2 - u_1 u_2 - \beta u_1^2), \\ \frac{\partial u_2}{\partial t} &= \mu_2 \Delta u_2 + \frac{1}{\alpha}(\gamma u_3 - u_2 - u_1 u_2), \\ \frac{\partial u_3}{\partial t} &= \mu_3 \Delta u_3 + \delta(u_1 - u_3),\end{aligned}\quad (6.2.3)$$

where the unknown functions are $u_i \geq 0$ ($1 \leq i \leq 3$), and the parameters are positive constants:

$$\mu_1, \mu_2, \mu_3, \alpha, \beta, \gamma, \delta > 0.$$

Hence the control parameter is given by

$$\lambda = (\mu_1, \mu_2, \mu_3, \alpha, \gamma, \delta) \in \mathbb{R}_+^6.$$

We consider the case where the spatial domain is a bounded domain $\Omega \subset \mathbb{R}^n$. In the case where there is an exchange of materials on the boundary $\partial\Omega$ to maintain the level of concentrations of X, Y, Z , then the equations (6.2.3) are supplemented with the Dirichlet boundary condition

$$u = (u_1, u_2, u_3) = 0 \quad \text{on } \partial\Omega. \quad (6.2.4)$$

If there is no exchange of materials on the boundary, the Neumann boundary condition is used:

$$\frac{\partial u}{\partial n} = 0 \quad \text{on } \partial\Omega. \quad (6.2.5)$$

Invariant Regions and Mathematical Set-Up It is known in Smoller (1983) that the following region

$$D = \{(u_1, u_2, u_3) \in L^2(\Omega)^3 \mid 0 < u_i < a_i, 1 \leq i \leq 3\} \quad (6.2.6)$$

is invariant for (6.2.3)–(6.2.4), where a_i ($1 \leq i \leq 3$) satisfy

$$a_1 > \max\{1, \beta^{-1}\}, \quad a_2 > \gamma a_3, \quad a_3 > a_1.$$

As the solution $u = (u_1, u_2, u_3)$ of (6.2.3) represents concentrations of chemical materials, only nonnegative functions $u_i \geq 0$ ($1 \leq i \leq 3$) are chemically realistic.

We define the following function spaces:

$$H = L^2(\Omega)^3,$$

$$H_1 = \begin{cases} H^2(\Omega)^3 \cap H_0^1(\Omega)^3 & \text{for boundary condition (6.2.4),} \\ \left\{ u \in H^2(\Omega)^3 \mid \frac{\partial u}{\partial n} = 0 \text{ on } \partial\Omega \right\} & \text{for boundary condition (6.2.5).} \end{cases}$$

Define the operators $L_\lambda = A_\lambda + B_\lambda$ and $G_\lambda : H_1 \rightarrow H$ by

$$A_\lambda u = (\mu_1 \Delta u_1, \mu_2 \Delta u_2, \mu_3 \Delta u_3),$$

$$B_\lambda u = \left(\alpha u_1 + \alpha u_2, -\frac{1}{\alpha} u_2 + \frac{\gamma}{\alpha} u_3, \delta u_1 - \delta u_3 \right),$$

$$G(u, \lambda) = \left(-\alpha(u_1 u_2 + \beta u_1^2), -\frac{1}{\alpha} u_1 u_2, 0 \right).$$

Thus the Field–Körös–Noyes equations (6.2.3), with either (6.2.4) or (6.2.5), take the following operator form:

$$\frac{du}{dt} = L_\lambda u + G(u, \lambda), \quad u(0) = \varphi. \quad (6.2.7)$$

6.2.2 Transition Under the Dirichlet Boundary Condition

We consider the dynamic transitions of (6.2.3)–(6.2.4). Let ρ_k and e_k be the k th eigenvalue and eigenvector of the Laplacian $-\Delta$ with the Dirichlet boundary condition:

$$-\Delta e_k = \rho_k e_k, \quad e_k|_{\partial\Omega} = 0, \quad \text{for } k = 1, 2, \dots, \quad (6.2.8)$$

and the first eigenvector is simple and positive

$$e_1(x) > 0 \quad \forall x \in \Omega. \quad (6.2.9)$$

Then eigenvalues of L_λ are associated with eigenvalues of $E_k(\lambda)u = \beta_k(\lambda)u$, where

$$E_k(\lambda) = \begin{pmatrix} -(\mu_1\rho_k - \alpha) & \alpha & 0 \\ 0 & -(\mu_2\rho_k + \alpha^{-1}) & \alpha^{-1}\gamma \\ \delta & 0 & -(\mu_3\rho_k + \delta) \end{pmatrix}. \quad (6.2.10)$$

Let $N(\lambda)$ be a function of $\lambda = (\mu_1, \mu_2, \mu_3, \alpha, \gamma, \delta) \in \mathbb{R}_+^6$ defined by

$$N(\lambda) = \alpha \det E_1(\lambda) = \alpha\gamma\delta - (\mu_1\rho_1 - \alpha)(\mu_2\alpha\rho_1 + 1)(\mu_3\rho_1 + \delta).$$

Let $\lambda_0 = (\mu_1^0, \mu_2^0, \mu_3^0, \alpha^0, \gamma^0, \delta^0) \in \mathbb{R}_+^6$ be a solution of

$$N(\lambda_0) = 0. \quad (6.2.11)$$

Then the set $\Gamma \subset \mathbb{R}_+^6$ defined by

$$\Gamma = \{\lambda_0 \in \mathbb{R}_+^6 \mid \lambda_0 \text{ satisfies (6.2.11)}\}$$

is a 5D manifold, and separates \mathbb{R}_+^6 into two connected components U_1 and U_2 such that

$$N(\lambda) \begin{cases} < 0 & \text{if } \lambda \in U_1, \\ > 0 & \text{if } \lambda \in U_2. \end{cases}$$

Then we have the following dynamic transition theorem for (6.2.3)–(6.2.4), which characterizes the dynamic phase transition to a steady-state solution for the Belousov–Zhabotinsky reaction (6.2.1).

Theorem 6.2.1 *For any $\lambda_0 = (\mu_1^0, \mu_2^0, \mu_3^0, \alpha^0, \gamma^0, \delta^0) \in \Gamma$, the system (6.2.3)–(6.2.4) undergoes a Type-I dynamic transition at $(u, \lambda) = (0, \lambda_0)$, and the following assertions hold true:*

1. If $\lambda \in \mathbb{R}_+^6$ with $\lambda \neq \lambda_0$ and $N(\lambda) < N(\lambda_0) = 0$, then $u = 0$ is a stable equilibrium.
2. If $N(\lambda) > N(\lambda_0) = 0$, then the problem (6.2.3)–(6.2.4) bifurcates from $(u, \lambda) = (0, \lambda_0)$ to a unique steady-state solution $u_\lambda \in H_1$ with $u_\lambda > 0$ in Ω , given by

$$\begin{aligned} u_\lambda(x) &= \beta_1(\lambda)K_1 K_2^{-1} v_0(x) + o(|\beta_1(\lambda)|), \\ v_0(x) &= \left(\alpha^0(\delta^0 + \mu_3^0 \rho_1)e_1, \frac{\alpha^0 \gamma^0 \delta^0}{1 + \mu_2^0 \alpha^0 \rho_1} e_1, \alpha^0 \delta^0 e_1 \right), \end{aligned} \quad (6.2.12)$$

where $\beta_1(\lambda)$ is the positive eigenvalue of (6.2.10), and $K_1, K_2 > 0$ are defined by

$$\begin{aligned} K_1 &= \left[\alpha^0 \delta^0 (1 + \mu_2^0 \alpha^0 \rho_1) (\mu_3^0 \rho_1 + \delta^0) + \frac{(\alpha^0)^3 (\delta^0)^2 \gamma^0}{1 + \mu_2^0 \alpha^0 \rho_1} + \frac{(\alpha^0)^2 \gamma^0 (\delta^0)^2}{\delta^0 + \mu_3^0 \rho_1} \right] \cdot \|e_1\|^2, \\ K_2 &= \delta^0 (\alpha^0)^2 (\mu_3^0 \rho_1 + \delta^0) \left[\beta (\mu_3^0 \rho_1 + \delta^0) (\mu_2^0 \alpha^0 \rho_1 + 1) \right. \\ &\quad \left. + \alpha^0 \delta^0 \gamma^0 + \frac{\alpha^0 \gamma^0 \delta^0}{1 + \mu_2^0 \alpha^0 \rho_1} \right] \cdot \int_{\Omega} e_1^3 dx. \end{aligned}$$

In addition, u_λ is an attractor. Namely, there exists an open set $O \subset H$, such that $0 \in O$ and u_λ attracts $O \cap D$, where D is the invariant set given by (6.2.6), i.e., for any $\varphi \in O \cap D$, the solution $u(t, \varphi)$ of (6.2.7) satisfies that

$$\lim_{t \rightarrow \infty} \|u(t, \varphi) - u_\lambda\| = 0.$$

Proof. We use Theorem 2.3.1 to prove this theorem, and proceed in two steps.

STEP 1. VERIFICATION OF PES. It is well known that the operators L_λ and G in (6.2.7) satisfy (2.1.2) and (2.1.3). We now verify the PES: (1.2.4) and (1.2.5).

All eigenvalues $\beta_{jk}(\lambda)$ of L_λ ($1 \leq j \leq 3, k = 1, 2, \dots$) satisfy that

$$\det(\beta_{jk} I - E_k(\lambda)) = 0. \quad (6.2.13)$$

Observe that

$$\text{grad } N(\lambda) = \left(\frac{\partial N}{\partial \mu_1}, \frac{\partial N}{\partial \mu_2}, \frac{\partial N}{\partial \mu_3}, \frac{\partial N}{\partial \alpha}, \frac{\partial N}{\partial \gamma}, \frac{\partial N}{\partial \delta} \right) \neq 0 \quad \forall \lambda \in \Gamma,$$

and Γ can be expressed as zeroes of the following equation

$$\gamma = \frac{(\mu_1 \rho_1 - \alpha)(\mu_2 \alpha \rho_1 + 1)(\mu_3 \rho_1 + \delta)}{\alpha \delta},$$

where $\alpha, \delta, \mu_1, \mu_2, \mu_3 > 0$. Hence, it follows that $\Gamma \subset \mathbb{R}_+^6$ is a 5-dimensional manifold, and divides \mathbb{R}_+^6 into two disjoint open sets:

$$\mathbb{R}_+^6 = \overline{U}_1 + \overline{U}_2, \quad U_1 \cap U_2 = \emptyset, \quad \partial U_1 \cap \partial U_2 = \Gamma,$$

such that

$$\det E_1(\lambda) = \beta_{11}(\lambda) \beta_{21}(\lambda) \beta_{31}(\lambda) \begin{cases} < 0 & \forall \lambda \in U_1, \\ > 0 & \forall \lambda \in U_2. \end{cases} \quad (6.2.14)$$

On the other hand, the trace of $E_1(\lambda)$ is given by

$$\sum_{j=1}^3 \operatorname{Re}\beta_{j1}(\lambda) = -[(\mu_1\rho_1 - \alpha) + (\mu_2\rho_1 + \alpha^{-1}) + (\mu_3\rho_1 + \delta)]. \quad (6.2.15)$$

By (6.2.14) we deduce that

$$\mu_1\rho_1 - \alpha > 0, \quad \forall \lambda = (\mu_1, \mu_2, \mu_3, \alpha, \gamma, \delta) \in \overline{U}_1. \quad (6.2.16)$$

It follows from (6.2.14)–(6.2.16) that

$$\sum_{j=1}^3 \operatorname{Re}\beta_{j1}(\lambda) < 0, \quad \forall \lambda \in \overline{U}_1. \quad (6.2.17)$$

Then, by (6.2.14) and (6.2.17), we obtain that

$$\begin{aligned} \beta_{11}(\lambda) &= 0, \quad \operatorname{Re}\beta_{21}(\lambda) < 0, \quad \operatorname{Re}\beta_{31}(\lambda) < 0 && \text{for } \lambda \in \Gamma, \\ \beta_{11}(\lambda) &< 0, \quad \operatorname{Re}\beta_{21}(\lambda) < 0, \quad \operatorname{Re}\beta_{313}(\lambda) < 0 && \text{for } \lambda \in U_1, \\ \beta_{11}(\lambda) &> 0 && \text{for } \lambda \in U_2. \end{aligned} \quad (6.2.18)$$

Therefore (1.2.4) follows from (6.2.18). Now, we need to prove that

$$\operatorname{Re}\beta_{jk}(\lambda_0) < 0 \quad \forall \lambda_0 \in \Gamma, \quad 1 \leq j \leq 3, \quad k \geq 2. \quad (6.2.19)$$

By (6.2.16) we deduce that for any $k \geq 2$,

$$\begin{aligned} \det E_k(\lambda_0) &= \frac{1}{\alpha^0} [\alpha^0 \gamma^0 \delta^0 - (\mu_1^0 \rho_k - \alpha^0)(\mu_2^0 \rho_k \alpha^0 + 1)(\mu_3^0 \rho_k + \delta^0)] \\ &< \frac{1}{\alpha^0} [\alpha^0 \gamma^0 \delta^0 - (\mu_1^0 \rho_1 - \alpha^0)(\mu_2^0 \rho_1 \alpha^0 + 1)(\mu_3^0 \rho_1 + \delta^0)]. \end{aligned}$$

Namely,

$$\det E_k(\lambda_0) < \det E_1(\lambda_0) = 0 \quad \forall k \geq 2, \quad \lambda_0 \in \Gamma. \quad (6.2.20)$$

In addition, it is easy to see that

$$E_k(\lambda_0) = E_1(\tilde{\lambda}), \quad \tilde{\lambda} = (\mu_1^0 \rho_k \rho_1^{-1}, \mu_2^0 \rho_k \rho_1^{-1}, \mu_3^0 \rho_k \rho_1^{-1}, \alpha^0, \gamma^0, \delta^0). \quad (6.2.21)$$

Hence, we have

$$\beta_{jk}(\lambda_0) = \beta_{j1}(\tilde{\lambda}), \quad \forall k \geq 1, \quad j = 1, 2, 3. \quad (6.2.22)$$

Thus, from (6.2.20), (6.2.21), and (6.2.14) we find $\tilde{\lambda}_1 \in U_1$. Then, (6.2.19) follows from (6.2.18) and (6.2.22).

STEP 2. By Theorem 2.3.1, we now verify

$$\frac{(G(v_0, \lambda_0), v_0^*)}{(v_0, v_0^*)} < 0, \quad (6.2.23)$$

where v_0 and v_0^* are the eigenvectors of L_{λ_0} and $L_{\lambda_0}^*$ corresponding to $\beta_{11}(\lambda_0) = 0$. Let

$$v_0 = (x_1 e_1, x_2 e_1, x_3 e_1), \quad v_0^* = (x_1^* e_1, x_2^* e_1, x_3^* e_1),$$

where e_1 satisfies (6.2.9). Then, (x_1, x_2, x_3) is the eigenvector of the matrix $E_1(\lambda_0)$:

$$\begin{pmatrix} -(\mu_1^0 \rho_1 - \alpha^0) & \alpha^0 & 0 \\ 0 & -(\mu_2^0 \rho_1 + \frac{1}{\alpha^0}) & \gamma^0 / \alpha^0 \\ \delta^0 & 0 & -(\mu_3^0 \rho_1 + \delta^0) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0,$$

and (x_1, x_2, x_3) is expressed as

$$(x_1, x_2, x_3) = \left(\alpha^0 (\delta^0 + \mu_3^0 \rho_1), \frac{\alpha^0 \gamma^0 \delta^0}{1 + \mu_2^0 \alpha^0 \rho_1}, \alpha^0 \delta^0 \right).$$

Hence, the eigenvector $v_0(x)$ is as given in (6.2.12).

The conjugate eigenvector (x_1^*, x_2^*, x_3^*) satisfies

$$\begin{pmatrix} -(\mu_1^0 \rho_1 - \alpha^0) & 0 & \delta^0 \\ \alpha^0 & -\left(\mu_2^0 \rho_1 + \frac{1}{\alpha^0}\right) & 0 \\ 0 & \gamma^0 / \alpha^0 & -(\mu_3^0 \rho_1 + \delta^0) \end{pmatrix} \begin{pmatrix} x_1^* \\ x_2^* \\ x_3^* \end{pmatrix} = 0.$$

Hence we have

$$v_0^*(x) = \left(\delta^0 (1 + \mu_2^0 \alpha^0 \rho_1) e_1, \delta^0 (\alpha^0)^2 e_1, \frac{\alpha^0 \gamma^0 \delta^0}{\delta^0 + \mu_3^0 \rho_1} e_1 \right).$$

Then, inserting v_0 and v_0^* into (6.2.23), we deduce that

$$\begin{aligned} K_2 &= -(G(v_0, \lambda_0), v_0^*) = \int_{\Omega} [\alpha^0 (x_1 x_2 + \beta x_1^2) x_1^* + x_1 x_2 x_2^* / \alpha^0] e_1^3 dx, \\ K_1 &= (v_0, v_0^*) = (x_1 x_1^* + x_2 x_2^* + x_3 x_3^*) \int_{\Omega} e_1^2 dx. \end{aligned}$$

Then with simple substitutions, the theorem follows. \square

6.2.3 Transitions Under the Neumann Boundary Condition

Basic States We now consider the Neumann boundary condition (6.2.5). Now the system (6.2.3) admits two constant steady states:

$$\begin{aligned} U_0 &= (0, 0, 0), & U_1 &= (u_1^0, u_2^0, u_3^0), \\ u_1^0 = u_3^0 &= \sigma, & u_2^0 &= \frac{\gamma\sigma}{1 + \sigma} = \frac{1}{2}(1 + \gamma - \beta\sigma), \\ \sigma &= \frac{1}{2\beta} \left[(1 - \gamma - \beta) + \sqrt{(1 - \gamma - \beta)^2 + 4\beta(1 + \gamma)} \right]. \end{aligned} \quad (6.2.24)$$

It is easy to check that the steady-state solution $U_0 = 0$ is always unstable, because the linearized equations of (6.2.3) have three real eigenvalues λ_1, λ_2 , and λ_3 satisfying

$$\lambda_1\lambda_2\lambda_3 = \det \begin{pmatrix} \alpha & \alpha & 0 \\ 0 & -\frac{1}{\alpha} & \frac{\gamma}{\alpha} \\ \delta & 0 & -\delta \end{pmatrix} = \delta(\gamma + 1) > 0,$$

and there is at least one eigenvalue with positive real part. Therefore we only need to consider the transition of (6.2.3) with (6.2.5) at the steady-state solution U_1 .

Consider the deviation:

$$u_i = u'_i + u_i^0 \quad (1 \leq i \leq 3). \quad (6.2.25)$$

Omitting the primes, the problem (6.2.3) with (6.2.5) becomes

$$\begin{aligned} \frac{\partial u_1}{\partial t} &= \mu_1 \Delta u_1 + \alpha[(1 - u_2^0 - 2\beta u_1^0)u_1 + (1 - u_1^0)u_2 - u_1 u_2 - \beta u_1^2], \\ \frac{\partial u_2}{\partial t} &= \mu_2 \Delta u_2 + \frac{1}{\alpha}[-u_2^0 u_1 - (1 + u_1^0)u_2 + \gamma u_3 - u_1 u_2], \\ \frac{\partial u_3}{\partial t} &= \mu_3 \Delta u_3 + \delta(u_1 - u_3), \\ \frac{\partial u}{\partial n}|_{\partial\Omega} &= 0. \end{aligned} \quad (6.2.26)$$

The invariant region for the Neumann boundary condition case is the same as for the Dirichlet boundary condition. The following lemma is well known (Smoller, 1983; Temam, 1997).

Lemma 6.2.2 *The region D given by (6.2.6) is also invariant for the problem (6.2.3) with (6.2.5). In particular this problem possesses a global attractor in D .*

Linearized Problem In view of (6.2.24), the eigenvalue equations of (6.2.26) read

$$\begin{aligned} \mu_1 \Delta u_1 - \alpha \left[\frac{1}{2}(\gamma + 3\beta\sigma - 1)u_1 + (\sigma - 1)u_2 \right] &= \lambda u_1, \\ \mu_2 \Delta u_2 - \frac{1}{\alpha} \left[\frac{1}{2}(1 + \gamma - \beta\sigma)u_1 + (\sigma + 1)u_2 - \gamma u_3 \right] &= \lambda u_2, \\ \mu_3 \Delta u_3 + \delta(u_1 - u_3) &= \lambda u_3, \\ \frac{\partial u}{\partial n}|_{\partial\Omega} &= 0. \end{aligned} \quad (6.2.27)$$

For brevity, we still use $\rho_k \geq 0$ and e_k to denote the k th eigenvalue and eigenvector of the Laplace operator Δ with the Neumann boundary condition:

$$\Delta e_k = -\rho_k e_k, \quad \frac{\partial e_k}{\partial n}|_{\partial\Omega} = 0. \quad (6.2.28)$$

Let M_k be the following matrix:

$$M_k = \begin{pmatrix} -\mu_1\rho_k - \frac{\alpha}{2}(\gamma + 3\beta\sigma - 1) & -\alpha(\sigma - 1) & 0 \\ -\frac{1}{2\alpha}(1 + \gamma - \beta\sigma) & -\mu_2\rho_k - \frac{1}{\alpha}(\sigma + 1) & \frac{\gamma}{\alpha} \\ \delta & 0 & -\mu_3\rho_k - \delta \end{pmatrix}. \quad (6.2.29)$$

Thus, all eigenvalues $\lambda = \beta_{kj}$ of (6.2.27) satisfy

$$M_k x_{kj} = \beta_{kj} x_{kj}, \quad 1 \leq j \leq 3, \quad k = 1, 2, \dots,$$

where $x_{kj} \in \mathbb{R}^3$ is the eigenvector of M_k corresponding to β_{kj} . Hence, the eigenvector u_{kj} of (6.2.27) corresponding to β_{kj} is

$$u_{kj}(x) = x_{kj} e_k(x), \quad (6.2.30)$$

where $e_k(x)$ is as in (6.2.28). In particular, $\rho_1 = 0$ and e_1 is a constant, and

$$M_1 = \begin{pmatrix} -\frac{\alpha}{2}(\gamma + 3\beta\sigma - 1) & -\alpha(\sigma - 1) & 0 \\ -\frac{1}{2\alpha}(1 + \gamma - \beta\sigma) & -\frac{1}{\alpha}(\sigma + 1) & \frac{\gamma}{\alpha} \\ \delta & 0 & -\delta \end{pmatrix}. \quad (6.2.31)$$

The eigenvalues $\lambda = \beta_{1j}$ ($1 \leq j \leq 3$) of M_1 satisfy

$$\begin{aligned} \lambda^3 + A\lambda^2 + B\lambda + C &= 0, \\ A &= \delta + \left[\frac{\alpha}{2}(3\beta\sigma + \gamma - 1) + \frac{1}{\alpha}(\sigma + 1) \right], \\ B &= 2\beta\sigma^2 + \gamma - (1 - \beta)\sigma + \delta \left[\frac{\alpha}{2}(3\beta\sigma + \gamma - 1) + \frac{1}{\alpha}(\sigma + 1) \right], \\ C &= \delta\sigma(2\beta\sigma + \beta + \gamma - 1). \end{aligned} \quad (6.2.32)$$

It is known that all solutions of (6.2.32) have negative real parts if and only if

$$A > 0, \quad C > 0, \quad AB - C > 0. \quad (6.2.33)$$

Direct calculation shows that the two parameters A and C in (6.2.32) are positive:

$$A > 0, \quad C > 0; \quad (6.2.34)$$

see also Hastings and Murray (1975). Note that

$$\beta_{11}\beta_{12}\beta_{13} = -C < 0.$$

Hence all real eigenvalues of (6.2.31) do not change their signs, and at least one of these real eigenvalues is negative.

In addition, we derive from $AB - C = 0$ a critical number

$$\begin{aligned}\delta_0 &= \frac{c + b - a^2 + \sqrt{(c + b - a^2)^2 + 4a^2b}}{2a}, \\ a &= \frac{\alpha}{2}(3\beta\sigma + \gamma - 1) + \frac{1}{\alpha}(\sigma + 1), \\ b &= (1 - \beta)\sigma - 2\beta\sigma^2 - \gamma, \\ c &= \sigma(2\beta\sigma + \beta + \gamma - 1),\end{aligned}\quad (6.2.35)$$

such that $\delta_0 > 0$ if and only if

$$b = (1 - \beta)\sigma - 2\beta\sigma^2 - \gamma > 0, \quad (6.2.36)$$

and under the condition (6.2.36)

$$AB - C \begin{cases} > 0 & \text{if } \delta > \delta_0, \\ = 0 & \text{if } \delta = \delta_0, \\ < 0 & \text{if } \delta < \delta_0. \end{cases} \quad (6.2.37)$$

Now, we need to check the other eigenvalues β_{kj} with $k \geq 2$. By (6.2.29), $\lambda_k = \beta_{kj}$ ($k \geq 2$) satisfy

$$\lambda_k^3 + A_k \lambda_k^2 + B_k \lambda_k + C_k = 0, \quad (6.2.38)$$

where

$$\begin{aligned}A_k &= A + (\mu_1 + \mu_2 + \mu_3)\rho_k, \\ B_k &= B + (\mu_1 + \mu_2)\rho_k + a\mu_3\rho_k + \left(\frac{\sigma+1}{\alpha}\mu_1 + \frac{\alpha}{2}(3\beta\sigma + \gamma - 1)\mu_2\right)\rho_k \\ &\quad + (\mu_1\mu_2 + \mu_1\mu_3 + \mu_2\mu_3)\rho_k^2, \\ C_k &= C + \left[\left(\frac{\sigma+1}{\alpha}\mu_1 + \frac{\alpha}{2}(3\beta\sigma + \gamma - 1)\mu_2\right)\delta - b\mu_3\right]\rho_k \\ &\quad + \left[\mu_1\mu_2\delta + \left(\frac{1+\sigma}{\alpha}\mu_1 + \frac{\alpha}{2}(3\beta\sigma + \gamma - 1)\mu_2\right)\mu_3\right]\rho_k^2 + \mu_1\mu_2\mu_3\rho_k^3,\end{aligned}$$

where a, b are as in (6.2.35), and A, B, C are as in (6.2.32).

By (6.2.24) it is easy to check that

$$3\beta\sigma + \gamma - 1 > 0.$$

With the condition (6.2.36) we introduce another critical number

$$\delta_1 = \max_{\rho_k \neq 0} \left[\frac{\mu_3\rho_k b - C}{\left(\frac{\sigma+1}{\alpha}\mu_1 + \frac{\alpha}{2}(3\beta\sigma + \gamma - 1)\mu_2\right)\rho_k + \mu_1\mu_2\rho_k^2} - \mu_3\rho_k \right], \quad (6.2.39)$$

where b is as in (6.2.35).

Then the following lemma establishes the principle of exchange of stability for the eigenvalues β_{kj} of (6.2.27), as functions of δ .

Lemma 6.2.3 *Let δ_0 and δ_1 be the numbers given by (6.2.35) and (6.2.39), and b given by (6.2.35). When $b < 0$, all eigenvalues of (6.2.27) always have negative real part, and when $b > 0$ the following assertions hold true:*

- 1 *Let $\delta_0 < \delta_1$ and $k_0 \geq 2$ the integer that δ_1 in (6.2.39) reaches its maximum at ρ_{k_0} . Then $\beta_{k_0 l}$ is a real eigenvalue of (6.2.27), and*

$$\beta_{k_1}(\delta) \begin{cases} < 0 & \text{if } \delta > \delta_1, \\ = 0 & \text{if } \delta = \delta_1, \\ > 0 & \text{if } \delta < \delta_1 \end{cases} \quad \text{for } \rho_k = \rho_{k_0},$$

$$\operatorname{Re}\beta_{ij}(\delta_1) < 0 \quad \forall (i, j) \neq (k, 1) \text{ with } \rho_k = \rho_{k_0}.$$

- 2 *Let $\delta_0 > \delta_1$. Then $\beta_{11}(\delta) = \bar{\beta}_{12}(\delta)$ are a pair of complex eigenvalues of (6.2.27) near $\delta = \delta_0$, and*

$$\operatorname{Re}\beta_{11} = \operatorname{Re}\beta_{12} \begin{cases} < 0 & \text{if } \delta > \delta_0, \\ = 0 & \text{if } \delta = \delta_0, \\ > 0 & \text{if } \delta < \delta_0, \end{cases}$$

$$\operatorname{Re}\beta_{kj}(\delta_0) < 0 \quad \forall (k, j) \neq (1, 1), (1, 2).$$

Proof. In (6.2.32) we see that $A = a + \delta$, $B = a\delta - b$. By the direct calculation, we can see that

$$A_k > 0, \quad A_k B_k - C_k > 0, \quad \forall k \geq 2. \quad (6.2.40)$$

As $b < 0$, by (6.2.34)–(6.2.36), (6.2.38), and (6.2.40), for all physically sound parameters $\mu_1, \mu_2, \mu_3, \alpha, \beta, \delta > 0$, the following relations hold true

$$A_i > 0, \quad C_i > 0, \quad A_i B_i - C_i > 0, \quad \forall i \geq 1.$$

Hence, it follows that all eigenvalues β_{kj} ($k \geq 1, 1 \leq j \leq 3$) of (6.2.31) have negative real parts.

As $b > 0$, and $\delta_0 < \delta_1$, we infer from (6.2.34) and (6.2.37) that

$$\operatorname{Re}\beta_{1j}(\delta_1) < 0, \quad \forall 1 \leq j \leq 3. \quad (6.2.41)$$

In addition, it is clear that

$$C_{k_1} \begin{cases} > 0 & \text{if } \delta > \delta_1, \\ = 0 & \text{if } \delta = \delta_1, \\ < 0 & \text{if } \delta < \delta_1, \end{cases} \quad (6.2.42)$$

$$C_k > 0 \quad \text{at } \delta = \delta_1 \text{ for all } k \neq k_1.$$

Thus Assertion (1) follows from (6.2.40) to (6.2.42).

As $\delta_0 > \delta_1$, by (6.2.42) we know that $C_k > 0$ at $\delta = \delta_0$ for all $k \geq 2$. Since the real eigenvalues of β_{1j} ($1 \leq j \leq 3$) are negative, the condition (6.2.37) implies that there are a pair of complex eigenvalues $\beta_{11} = \bar{\beta}_{12}$ cross the imaginary axis at $\delta = \delta_0$. Then Assertions (2) follows. The lemma is proved. \square

Spatiotemporal Oscillations By Lemma 6.2.3, if $\delta_1 < \delta_0$, the problem (6.2.26) undergoes a dynamic transition to a periodic solution from $\delta = \delta_0$. To determine the types of transition, we introduce a parameter as follows:

$$\begin{aligned} b_1 = & \frac{\rho^2}{D^2 E} \left[\frac{3}{D_0} ((2D - 6 - \alpha\gamma\rho^2 D_3)F_3 - (2\gamma D_3 D_4 + \alpha D_5 D - 6)F_1) \right. \\ & - \frac{1}{D_0} \left((\alpha\gamma\rho^2 D_3 - 2D_6)F_1 + (2\gamma D_3 D_4) + \alpha D_5 D_6)F_3 \right) \\ & + \frac{1}{D_0} (2\gamma\rho D_3 + \alpha\rho D_6 + \alpha\gamma\rho D_3 D_5 - 2\rho^{-1} D_4 D_6)F_2 \\ & + \frac{\alpha}{2D^2 E\rho} (2\gamma D_3 D_8 + \alpha D_6 D_7)(2\rho^{-1} D_6 D_8 - \alpha\gamma\rho D_3 D_7) \\ & + \frac{\alpha\rho^2}{2D^2 E} (2\gamma D_3 + \alpha D_6)(2\gamma D_3 D_8 + \alpha D_6 D_7) \\ & \left. - \frac{\alpha\rho^2}{2D^2 E} (2D_6 - \alpha\gamma\rho^2 D_3)(2\rho^{-2} D_6 D_8 - \alpha\gamma D_3 D_7) \right], \end{aligned} \quad (6.2.43)$$

where

$$\begin{aligned} D^2 &= \frac{\gamma^2 \delta_0^2 \rho^2 (\sigma - 1)^2 + \rho^2 ((\delta_0^2 + \rho^2)^2 + \gamma \delta_0^2 (\sigma - 1))^2}{N^2 (\sigma - 1)^2 (\delta^2 + \rho^2)^4}, & D_0 &= \frac{\gamma \delta_0}{\alpha a^2} + \frac{a + 2\delta_0}{\alpha(\delta - 1)}, \\ D_1 &= \frac{\beta}{\alpha} + \frac{(\alpha A - 2)(\beta\sigma + 2\beta + \gamma - 1)}{2\alpha(\sigma - 1)^2}, & D_2 &= \frac{2 - \alpha A}{a^2(\sigma - 1)^2}, \\ D_4 &= A + \alpha\beta\sigma^2 + \alpha - 3\alpha\beta\sigma - \alpha\beta - \alpha\gamma, & D_3 &= \delta_0(\sigma - 1), \\ D_5 &= A + \alpha - \alpha\beta\sigma - 2\alpha\beta - \alpha\gamma, & D_6 &= (\delta_0^2 + \rho^2)^2 + \gamma \delta_0^2 (\sigma - 1), \\ D_7 &= 1 - \beta\sigma - \beta - \gamma, & D_8 &= \beta\sigma^2 + \beta + 1 - 3\beta\sigma - \gamma, \\ F_1 &= \frac{D_1}{A} - \frac{2\rho^2 D_1}{A(A^2 + 4\rho^2)} - \frac{\rho D_2}{A^2 + 4\rho^2}, & \rho &= \sqrt{B} = \sqrt{\frac{C}{A}}, \\ F_2 &= \frac{D_2}{A} - \frac{4\rho^2 D_2}{A(A^2 + 4\rho^2)} + \frac{2\rho D_1}{A^2 + 4\rho^2}, & E &= \alpha^3 (\sigma - 1)^3 (\delta_0^2 + \rho^2)^2, \\ F_3 &= \frac{2\rho D_1}{A(A^2 + 4\rho^2)} + \frac{\rho D_2}{A^2 + 4\rho^2}. \end{aligned}$$

Here A, B, C are as in (6.2.32), and a is as in (6.2.35).

Then we have the following dynamic transition theorem.

Theorem 6.2.4 *Let $\delta_1 < \delta_0$, and b_1 is the number given by (6.2.43).*

- 1 *If $b_1 < 0$, the problem (6.2.26) undergoes a Type-I dynamic transition at $\delta = \delta_0$, and bifurcates to a periodic solution on $\delta < \delta_0$ which is an attractor.*

2 If $b_1 > 0$, the problem (6.2.26) undergoes a Type-II dynamic transition at $\delta = \delta_0$, and bifurcates on $\delta > \delta_0$ to a periodic solution, which is a repeller.

Proof. By Lemma 6.2.3, at $\delta = \delta_0$ there is a pair of imaginary eigenvalues $\beta_{11} = \bar{\beta}_{12} = -i\rho$ of (6.2.27). Let $z = \xi + i\eta$ and $z^* = \xi^* + i\eta^*$ be the eigenvectors and conjugate eigenvectors of (6.2.27) corresponding to $-i\rho$, i.e., z and z^* satisfy that

$$(M_1 + i\rho)z = 0, \quad (M_1^* - i\rho)z^* = 0, \quad (6.2.44)$$

where M_1 is the matrix as in (6.2.31), and M_1^* the transpose of M . Because $\pm i\rho$ are solutions of (6.2.32), and $AB = C$ at $\delta = \delta_0$, we deduce that

$$\rho^2 = B = C/A. \quad (6.2.45)$$

For $z = (z_1, z_2, z_3)$, from the first equation of (6.2.44) we get

$$\delta_0 z_1 = (\delta_0 - i\rho)z_3, \quad \alpha(\sigma - 1)z_2 = \left(-\frac{\alpha}{2}(\gamma + 3\beta\sigma - 1) + i\rho\right)z_1. \quad (6.2.46)$$

Thus, we derive from (6.2.46) the eigenvectors $z = \xi + i\eta$ as follows:

$$\begin{aligned} \xi &= (\xi_1, \xi_2, \xi_3) = \left(1, -\frac{\gamma+3\beta\sigma-1}{2(\sigma-1)}, \frac{\delta_0^2}{\delta_0^2+\rho^2}\right), \\ \eta &= (\eta_1, \eta_2, \eta_3) = \left(0, \frac{\rho}{\alpha(\sigma-1)}, \frac{\rho\delta_0}{\delta_0^2+\rho^2}\right). \end{aligned} \quad (6.2.47)$$

In the same fashion, from the second equation of (6.2.44) we derive the conjugate eigenvectors $z^* = \xi^* + i\eta^*$ as

$$\begin{aligned} \xi^* &= (\xi_1^*, \xi_2^*, \xi_3^*) = \left(-\frac{\sigma+1}{\alpha^2(\sigma-1)}, 1, \frac{\gamma\delta_0}{\alpha(\delta_0^2+\rho^2)}\right), \\ \eta^* &= (\eta_1^*, \eta_2^*, \eta_3^*) = \left(-\frac{\rho}{\alpha(\sigma-1)}, 0, -\frac{\gamma\rho}{\alpha(\delta_0^2+\rho^2)}\right). \end{aligned} \quad (6.2.48)$$

As in (4.4.74), we have

$$(\xi, \xi^*) = (\eta, \eta^*) = -\frac{\gamma\rho^2 D_3}{H_1}, \quad (\xi, \eta^*) = -(\eta, \xi^*) = -\frac{\rho D_6}{H_1}, \quad (6.2.49)$$

where $H_1 = \alpha(\sigma - 1)(\delta_0^2 + \rho^2)^2$, D_3 and D_6 are as in (6.2.43). It is known that functions $\Psi_1^* + i\Psi_2^*$ given by

$$\begin{aligned} \Psi_1^* &= \frac{1}{(\xi, \xi^*)}[(\xi, \xi^*)\xi^* + (\xi, \eta^*)\eta^*], \\ \Psi_2^* &= \frac{1}{(\eta, \eta^*)}[(\eta, \xi^*)\xi^* + (\eta, \eta^*)\eta^*], \end{aligned} \quad (6.2.50)$$

also satisfy the second equation of (6.2.44) with

$$(\xi, \Psi_1^*) = (\eta, \Psi_2^*) \neq 0, \quad (\xi, \Psi_2^*) = (\eta, \Psi_1^*) = 0. \quad (6.2.51)$$

Let $u \in H$ be a solution of (6.2.26) expressed as

$$u = x\xi + y\eta + \Phi(x, y), \quad x, y \in \mathbb{R}^1.$$

where $\Phi(x, y)$ is the center manifold function of (6.2.26) at $\delta = \delta_0$. By (6.2.51), the reduced equations of (6.2.26) on the center manifold are given by

$$\begin{aligned} \frac{dx}{dt} &= -\rho y + \frac{1}{(\xi, \Psi_1^*)}(G(x\xi + y\eta + \Phi), \Psi_1^*), \\ \frac{dy}{dt} &= \rho x + \frac{1}{(\eta, \Psi_2^*)}(G(x\xi + y\eta + \Phi), \Psi_2^*), \end{aligned} \quad (6.2.52)$$

where $G(u) = G(u, u)$ is the bilinear operator defined by

$$G(u, v) = (-\alpha u_1 v_2 - \alpha \beta u_1 v_1, -\frac{1}{\alpha} u_1 v_2, 0) \quad (6.2.53)$$

for $u = (u_1, u_2, u_3), v = (v_1, v_2, v_3) \in H_1$.

We are now in a position to find the center manifold function $\Phi(x, y)$. To this end, we need to determine the third eigenvalue β_{13} and eigenvector ζ of M_1 at $\delta = \delta_0$. We know that

$$\beta_{13} \cdot (i\rho)(-i\rho) = -\rho^2 \beta_{13} = \det M_1 = C.$$

By (6.2.45) we obtain

$$\beta_{13} = -A = -(\delta_0 + a),$$

and a is the number as in (6.2.35). Then, from the equation

$$(M_1 - \beta_{13})\zeta = 0,$$

we derive the eigenvector

$$\zeta = (\zeta_1, \zeta_2, \zeta_3) = \left(1, \frac{A - \frac{\alpha}{2}(\gamma + 3\beta\sigma + 1)}{\alpha(\sigma - 1)}, -\frac{\delta_0}{a}\right). \quad (6.2.54)$$

In the same token, from $(M_1^* - \beta_{13})\zeta^* = 0$, we obtain the conjugate eigenvector:

$$\zeta^* = (\zeta_1^*, \zeta_2^*, \zeta_3^*) = \left(\frac{A - \frac{1}{\alpha}(\sigma + 1)}{\alpha(\sigma - 1)}, 1, -\frac{\gamma}{\alpha a}\right). \quad (6.2.55)$$

On the other hand, from (6.2.47) and (6.2.53) it follows that

$$\begin{aligned} G_{11} &= G(\xi, \xi) = (-\alpha(\xi_2 + \beta), -\frac{\xi_2}{\alpha}, 0), & G_{22} &= G(\eta, \eta) = 0, \\ G_{12} &= G(\xi, \eta) = (-\alpha\eta_2, -\frac{\eta_2}{\alpha}, 0), & G_{21} &= G(\eta, \xi) = 0. \end{aligned} \quad (6.2.56)$$

Direct calculation shows that

$$(\zeta, \zeta^*) = D_0, \quad (G_{11}, \zeta^*) = D_1, \quad (G_{12}, \zeta^*) = D_2,$$

and D_0, D_1, D_2 are as in (6.2.43).

By (A.1.19), the center manifold function Φ satisfies

$$\Phi = \Phi_1 + \Phi_2 + \Phi_3 + o(2), \quad (6.2.57)$$

where

$$\begin{aligned} \mathcal{L}\Phi_1 &= -x^2 P_2 G_{11} - xy P_2 G_{12}, \\ (\mathcal{L}^2 + 4\rho^2)\mathcal{L}\Phi_2 &= 2\rho^2(x^2 - y^2)P_2 G_{11} + 4\rho^2 xy P_2 G_{12}, \\ (\mathcal{L}^2 + 4\rho^2)\Phi_3 &= \rho(y^2 - x^2)P_2 G_{12} + 2\rho xy P_2 G_{11}, \end{aligned}$$

$P_2 : H \rightarrow E_2$ is the canonical projection, E_2 =the orthogonal complement of $E_1 = \text{span}\{\xi, \eta\}$, and \mathcal{L} is the linearized operator of (6.2.26). By the spectral theorem, Theorem 3.4 in Ma and Wang (2005b), and (6.2.30), it is clear that

$$P_2 G_{11} = (G_{11}, \zeta^*)\zeta = D_1\zeta, \quad P_2 G_{12} = (G_{12}, \zeta^*)\zeta = D_2\zeta. \quad (6.2.58)$$

Hence $\Phi_1, \Phi_2, \Phi_3 \in \text{span}\{\zeta\}$. It implies that

$$\mathcal{L}\Phi_j = M_1\Phi_j = -A\Phi_j. \quad (6.2.59)$$

We derive from (6.2.57) to (6.2.59) the center manifold function as follows:

$$\begin{aligned} \Phi &= \frac{\zeta}{D_0} \left[\left(\frac{D_1}{A} - \frac{2\rho^2 D_1}{A(A^2 + 4\rho^2)} - \frac{\rho D_2}{A^2 + 4\rho^2} \right) x^2 \right. \\ &\quad + \left(\frac{D_2}{A} - \frac{4\rho^2 D_2}{A(A^2 + 4\rho^2)} + \frac{2\rho D_1}{A^2 + 4\rho^2} \right) xy \\ &\quad \left. + \left(\frac{2\rho^2 D_1}{A(A^2 + 4\rho^2)} + \frac{\rho D_2}{A^2 + 4\rho^2} \right) y^2 \right] + o(2) \\ &= \frac{1}{D_0} (F_1 x^2 + F_2 xy + F_3 y^2) \zeta + o(2), \end{aligned} \quad (6.2.60)$$

where F_1, F_2, F_3 are as in (6.2.43). The remaining part of the proof is to use Theorem 2.3.7 with routine substitutions, and we omit the details; see also Ma and Wang (2011c). \square

Theorem 6.2.4 describes the temporal oscillation for the Belousov–Zhabotinsky reactions of (6.2.1) in a non-stirred condition. If the reagent is stirred, the equations (6.2.26) are reduced to a system of ordinary differential equations as follows:

$$\begin{aligned}\frac{du_1}{dt} &= -\alpha \left(\frac{1}{2}(\gamma + 3\beta\sigma - 1)u_1 + (\sigma - 1)u_2 - u_1u_2 - \beta u_1^2 \right), \\ \frac{du_2}{dt} &= -\frac{1}{\alpha} \left(\frac{1}{2}(1 + \gamma - \beta\sigma)u_1 + (\sigma + 1)u_2 - \gamma u_3 - u_1u_2 \right), \\ \frac{du_3}{dt} &= \delta(u_1 - u_3).\end{aligned}\quad (6.2.61)$$

In this case, only the transition to periodic solutions can take place.

Theorem 6.2.5 *Let $\delta_0 > 0$ be the number given by (6.2.35), and b be as in (6.2.43). Then the system (6.2.61) undergoes a dynamic transition to periodic solutions at $\delta = \delta_0$. Furthermore, the system bifurcates to an unstable periodic orbit on $\delta > \delta_0$ as $b > 0$, and to a stable periodic orbit on $\delta < \delta_0$ as $b < 0$. In addition, the bifurcated periodic solution can be expressed in the following form*

$$u_\delta = [-b^{-1} \operatorname{Re} \beta_{11}(\delta)]^{1/2} (\xi \cos \rho t + \eta \sin \rho t) + o(|\operatorname{Re} \beta_{11}|^{1/2}), \quad (6.2.62)$$

where ξ, η are as in (6.2.55) and (6.2.57), and $\beta_{11}(\delta)$ is the first complex eigenvalue as described in Lemma 6.2.3.

Also, by Lemma 6.2.2, for all physically sound parameters, each of the two systems (6.2.26) and (6.2.61) has a global attractor in the invariant region D in (6.2.6). Hence, when $b > 0$, the bifurcated periodic solution u_δ is a repeller, and its stable manifold divides D into two disjoint open sets D_1 and D_2 , i.e., $\overline{D} = \overline{D}_1 + \overline{D}_2$, $D_1 \cap D_2 = \emptyset$, such that the equilibrium $U_1 = (u_1^0, u_2^0, u_3^0)$ in (6.2.24) attracts D_1 , and there is another attractor $\mathcal{A}_2 \subset D_2$ which attracts D_2 .

Transition to Steady-State Solutions Due to Lemma 6.2.3, as $\delta_1 > \delta_0$, the transition of (6.2.26) occurs at $\delta = \delta_1$, which is from real eigenvalues. Let δ_1 achieve its maximum at ρ_{k_0} ($k_0 \geq 2$). Assume that $\beta_{kol}(\delta)$ is simple near δ_1 .

First, we consider the case where

$$\int_{\Omega} e_{k_0}^3 dx \neq 0. \quad (6.2.63)$$

In general, the condition (6.2.63) holds true for the case where $\Omega \neq (0, L) \times D$, and $D \subset \mathbb{R}^{n-1}$ ($1 \leq n \leq 3$) is a bounded open set.

Let

$$b_0 = [\alpha\beta(\sigma - 1)(\mu_2\rho_{k_0} + \frac{1}{\alpha}(\sigma + 1)) - (\alpha\mu_2\rho_{k_0} + 2\sigma)(\mu_1\rho_{k_0} + \frac{\alpha}{2}(\gamma + 3\beta\sigma - 1))].$$

Theorem 6.2.6 *Let the above number $b_0 \neq 0$, and $\delta_1 > \delta_0$. Under the condition (6.2.63), the transition of (6.2.26) at $\delta = \delta_1$ is mixed (Type-III), and the system bifurcates from $(u, \delta) = (0, \delta_1)$ to a unique steady-state solution u^δ such that u^δ is a saddle on $\delta > \delta_1$, and an attractor on $\delta < \delta_1$. The solution u_δ can be expressed as*

$$u^\delta = C\beta_{k_0 l}(\delta)\xi e_{k_0} + o(|\beta_{k_0}|), \quad (6.2.64)$$

where $\xi = (\xi_1, \xi_2, \xi_3)$ and constant C are given by

$$\xi_1 = \alpha(\sigma - 1)(\mu_3\rho_{k_0} + \delta_1),$$

$$\xi_2 = -(\mu_1\rho_{k_0} + \frac{\alpha}{2}(\gamma + 3\beta\sigma - 1))(\mu_3\rho_{k_0} + \delta_1),$$

$$\xi_3 = \alpha\delta_1(\sigma - 1),$$

$$C = \frac{[(\mu_3\rho_{k_0} + \delta_1)^2(\mu_1\rho_{k_0} + \mu_2\rho_{k_0} + \frac{1}{\alpha}(\sigma + 1) + \frac{\alpha}{2}(\gamma + 3\beta\sigma - 1)) - (\sigma - 1)\gamma\delta_1] \int_{\Omega} e_{k_0} dx}{b_0(\mu_3\rho_{k_0} + \delta_1)^3 \int_{\Omega} e_{k_0}^3 dx}.$$

Proof. We apply Theorem 2.3.2 to prove this theorem. Let ξ and $\xi^* \in \mathbb{R}^3$ be the eigenvectors of M_{k_0} and $M_{k_0}^*$ corresponding to $\beta_{k_0 l}(\delta_1) = 0$, i.e.

$$M_{k_0}\xi = 0, \quad M_{k_0}^*\xi^* = 0,$$

where M_{k_0} is the matrix (6.2.29) with $k = k_0$. Clearly, ξ is as in (6.2.64), and

$$\xi^* = (-(\mu_2\rho_{k_0} + \frac{1}{\alpha}(\sigma + 1))(\mu_3\rho_{k_0} + \delta_1), \alpha(\mu_3\rho_{k_0} + \delta_1)(\sigma - 1), \gamma(\sigma - 1)).$$

For the operator G in (6.2.7), we can derive that

$$\frac{1}{(\xi e_{k_0}, \xi^* e_{k_0})} (G(y\xi e_{k_0}), \xi^* e_{k_0}) = -\frac{1}{C}y^2 + o(2),$$

where C is as in (6.2.64). Therefore, the theorem follows from Theorem 2.3.2. \square

In the following, we assume that

$$\int_{\Omega} e_{k_0}^3 dx = 0.$$

Define

$$\begin{aligned} b_1 &= \frac{1}{(\xi, \xi^*)} \left[-(\mu_1\rho_{k_0} + \frac{\alpha}{2}(\gamma + 3\beta\sigma - 1))(\alpha\mu_2\rho_{k_0} + 2) \int_{\Omega} \phi_1 e_{k_0}^2 dx \right. \\ &\quad + 2\alpha^2\beta(\sigma - 1)(\mu_2\rho_{k_0} + \frac{\sigma + 1}{\alpha}) \int_{\Omega} \phi_1 e_{k_0}^2 dx \\ &\quad \left. + \alpha(\sigma - 1)(\alpha\mu_2\rho_{k_0} + 2) \int_{\Omega} \phi_2 e_{k_0}^2 dx \right], \end{aligned} \quad (6.2.65)$$

where

$$\begin{aligned} (\xi, \xi^*) &= \alpha(1 - \sigma) \left[(\mu_1\rho_{k_0} + \mu_2\rho_{k_0} + \frac{\sigma + 1}{\alpha} + \frac{\alpha}{2}(\gamma + 3\beta\sigma - 1)) \right. \\ &\quad \times (\mu_3\rho_{k_0} + \delta_1)^2 - \gamma\delta_1(\sigma - 1) \left. \right], \end{aligned}$$

the function $\phi = (\phi_1, \phi_2, \phi_3)$ satisfies

$$L\phi = -G(\xi)e_{k_0}^2, \quad (6.2.66)$$

and the operators L and G are defined by

$$\begin{aligned} L\phi &= \begin{cases} \mu_1\Delta\phi_1 - \alpha \left[\frac{1}{2}(\gamma + 3\beta\sigma - 1)\phi_1 + (\sigma - 1)\phi_2 \right], \\ \mu_2\Delta\phi_2 - \frac{1}{\alpha} \left[\frac{1}{2}(1 + \gamma - \beta\sigma)\mu_1 + (\sigma + 1)\phi_2 \right], \\ \mu_3\Delta\phi_3 + \delta(\phi_1 - \phi_3), \end{cases} \\ G(\xi) &= \begin{cases} \alpha^2(\sigma - 1)(\mu_3\rho_{k_0} + \delta_1)^2 \left[\mu_1\rho_{k_0} + \frac{\alpha}{2}(\gamma + 3\beta\sigma - 1) - \alpha\beta(\sigma - 1) \right], \\ (\sigma - 1)(\mu_3\rho_{k_0} + \delta_1)^2(\mu_1\rho_{k_0} + \frac{\alpha}{2}(\gamma + 3\beta\sigma - 1)), \\ 0. \end{cases} \end{aligned}$$

Theorem 6.2.7 Let $b_1 \neq 0$ be the parameter given by (6.2.65), and $\delta_0 < \delta_1$.

1. If $b_1 < 0$, then the system (6.2.26) undergoes a Type-I transition at $\delta = \delta_1$, and bifurcates to two steady-state solutions u_+^δ and u_-^δ on $\delta < \delta_1$ which are attractors, and no bifurcation on $\delta > \delta_1$.
2. If $b_1 > 0$, then the system undergoes a Type-II transition at $\delta = \delta_1$, and bifurcates from $(u, \delta) = (0, \delta_1)$ to two steady-state solutions u_+^δ and u_-^δ on $\delta > \delta_1$, which are saddles, and no bifurcation on $\delta < \delta_1$.
3. The bifurcated solutions u_\pm^δ can be expressed as

$$u_\pm^\delta = \pm \left[-\frac{\int_{\Omega} e_{k_0}^2 dx}{(\mu_2\rho_{k_0} + \delta_1)^2 b_1} \beta_{k_0 l}(\delta) \right]^{1/2} \xi e_{k_0} + o(|\beta_{k_0 l}|^{1/2}),$$

where $\xi \in \mathbb{R}^3$ is as in (6.2.64), and $\beta_{k_0 l}(\delta)$ is as in Lemma 6.2.3.

Proof. We use Theorem 2.3.1 to verify this theorem. Let

$$u = y\xi e_{k_0} + \Phi(y), \quad y \in \mathbb{R}^1,$$

and $\Phi(y)$ is the center manifold function. Denote by $\Phi = y^2\phi = o(2)$, then by Theorem A.1.1, ϕ satisfies

$$L\phi = -G(\xi e_{k_0}) = -G(\xi)e_{k_0}^2,$$

which is the equation (6.2.66).

We see that

$$\begin{aligned}
 (G(y\xi e_{k_0} + \Phi), \xi^* e_{k_0}) &= - \int_{\Omega} [\alpha(y\xi_1 e_{k_0} + \Phi_1)(y\xi_2 e_{k_0} + \Phi_2) \xi_1^* e_{k_0} \\
 &\quad + \alpha\beta(y\xi_1 e_{k_0} + \Phi_1)^2 \xi_1^* e_{k_0} + \frac{1}{\alpha}(y\xi_1 e_{k_0} + \Phi_1)(y\xi_2 e_{k_0} + \Phi_2) \xi_2^* e_{k_0}] dx \\
 &= y^3 [-(\alpha\xi_2\xi_1^* + 2\alpha\beta\xi_1\xi_1^* + \frac{1}{\alpha}\xi_2\xi_2^*) \int_{\Omega} \phi_1 e_{k_0}^2 dx \\
 &\quad - (\alpha\xi_1\xi_1^* + \frac{1}{\alpha}\xi_1\xi_2^*) \int_{\Omega} \phi_2 e_{k_0}^2 dx] + o(3).
 \end{aligned}$$

Hence, we have

$$\frac{1}{(\xi e_{k_0}, \xi^* e_{k_0})} (G(y\xi e_{k_0} + \Phi), \xi^* e_{k_0}) = \frac{1}{\int_{\Omega} e_{k_0}^2 dx} (\mu_2 \rho_{k_0} + \delta_1)^2 b_1 y^3 + o(y^3).$$

Thus the theorem follows from Theorem 2.3.1. \square

6.2.4 Phase Transition in the Realistic Oregonator

We begin with chemical parameters. In the chemical reaction (6.2.1), the constants given in Field and Noyes (1974) are as follows; see also Hastings and Murray (1975):

$$\begin{aligned}
 k_1 &= 1.34 \text{M}^{-1} \text{S}^{-1}, & k_2 &= 1.6 \times 10^9 \text{M}^{-1} \text{S}^{-1}, \\
 k_3 &= 8 \times 10^3 \text{M}^{-1} \text{S}^{-1}, & k_4 &= 4 \times 10^7 \text{M}^{-1} \text{S}^{-1},
 \end{aligned}$$

where

$$a = b = 6 \times 10^{-2} \text{M}, \quad \gamma = O(1), \quad k_5 = O(1) \text{ S}^{-1}.$$

Here we take

$$\gamma = 1, \quad k_5 = 1 \text{ S}^{-1},$$

and $a = b$ as a control parameter. The nondimensional parameters α, β, δ , and μ_i ($1 \leq i \leq 3$) are given by

$$\begin{aligned}
 \alpha &= \left(\frac{k_3 b}{k_1 a} \right)^{1/2} = 7.727 \times 10, \\
 \beta &= \frac{2k_1 k_4 a}{k_2 k_3 b} = 8.375 \times 10^{-6}, \\
 \delta &= k_5 (k_1 k_3 a b)^{1/2} = 1.035 \times 10^2 a \text{M}^{-1}, \\
 \mu_i &= \frac{1}{(k_1 k_2 a b)^{1/2}} \frac{\sigma_i}{l^2} = 4.664 \times 10^{-6} \frac{\sigma_i}{l^2 a} \text{M}^{-1} \text{S}^{-1}.
 \end{aligned} \tag{6.2.67}$$

Dirichlet Boundary Condition Case When the concentrations of $X = HBrO_2$, $Y = Br^-$ and Ce^{4+} are maintained to be vanished at the boundary, by Theorem 6.2.1, the reaction system (6.2.1) undergoes a dynamic transition to an equilibrium $u_\lambda \neq 0$ in Ω at the critical values $\lambda_0 = (\mu_1^0, \mu_2^0, \mu_3^0, \alpha^0, \gamma^0, \delta^0) \in \Gamma$. Furthermore, we see that the state of the system (6.2.1) depends on the sign of the parameter $N(\lambda)$ given by

$$N(\lambda) = \alpha\gamma\delta - (\mu_1\rho_1 - \alpha)(\mu_2\rho_1\alpha + 1)(\mu_3\rho_1 + \delta). \quad (6.2.68)$$

As $N(\lambda) < 0$, this system lies at the trivial state $u = 0$, and as $N(\lambda) > 0$ it is at the state $u_\lambda \neq 0$ given by (6.2.12).

Let

$$\begin{aligned} \alpha^0 &= \frac{\mu_1\mu_2\rho_1^2 + \sqrt{\mu_1^2\mu_2^2\rho_1^4 + 4} - 2}{2\mu_2\rho_1}, \\ \delta^0 &= \frac{(\mu_1\rho_1 - \alpha)(\mu_2\rho_1\alpha + 1)}{\alpha(2 + \mu_2\rho_1\alpha) - \mu_1\mu_2\rho_1^2\alpha - \mu_1\rho_1}. \end{aligned}$$

It is ready to derive from (6.2.68) that

- If $\mu_1\rho_1 \leq \alpha$, then $N(\lambda) > 0$,
- If $\alpha^0 < \alpha < \mu_1\rho_1$, then

$$N(\lambda) \begin{cases} > 0 & \text{for } \delta > \delta^0, \\ < 0 & \text{for } \delta < \delta^0, \end{cases}$$

- If $0 < \alpha \leq \alpha_0$, then $N(\lambda) < 0$.

Then, in view of (6.2.67), we claim that when $\mu_1\rho_1 \leq 77.27$, the reaction system (6.2.1) is at a nontrivial stable state; when $\alpha^0 < 77.27 < \mu_1\rho_1$, at $\delta = \delta^0$ this system undergoes a dynamic transition from the trivial state $u = 0$ to the equilibrium u_λ of (6.2.12); and when $77.27 \leq \alpha^0$, the system lies at $u = 0$. The first eigenvalue ρ_1 of (6.2.8) can be expressed as

$$\rho_1 = \frac{C}{L^2},$$

where $C > 0$ is a constant, and L is the length scale (or radius) of Ω . Then from (1) to (3) above we derive the following conclusion for the Belousov–Zhabotinsky reactions.

Let Ω_0 be a given domain, and Ω be the extension or contraction of Ω_0 from x_0 in the scale L ($0 < L < \infty$) defined by

$$\Omega(L, x_0) = \{(x - x_0)L \mid x, x_0 \in \Omega_0, x_0 \text{ is fixed}\}. \quad (6.2.69)$$

Physical Conclusion 6.2.8 Let $\Omega(L, x_0)$ be the domain defined by (6.2.69). Then, for the reaction system (6.2.1) with vanishing boundary condition (6.2.4), there exist two critical scales L_0 and L_1 ($L_0 < L_1$), depending on Ω_0 and x_0 , such that (a) as $L_1 < L$ the system is at a chemical state far from $u = 0$; (b) as $L_0 < L < L_1$ this

system bifurcates from $(u, \delta) = (0, \delta^0)$ to the stable equilibrium u_λ given by (6.2.12) on $\delta > \delta^0$; and (c) as $L < L_0$ the system will tend to the trivial state $u = 0$.

Stirred Case In view of (6.2.67), the numbers in (6.2.35) are as follows:

$$\sigma = 7 \times 10^2, \quad a = 9.74, \quad b = 690.79, \quad c = 8.21. \quad (6.2.70)$$

Hence

$$\delta_0 = 71.67. \quad (6.2.71)$$

Now, we need to compute the parameter b_1 in (6.2.43). At the critical value $\delta_0 = 71.67$, the numbers in (6.2.32) are

$$A = a + \delta_0 = 81.41, \quad B = a\delta_0 - b = 7.27, \quad C = \delta_0 c = 588.41. \quad (6.2.72)$$

Then we have

$$\begin{aligned} \rho &= \sqrt{B} \cong 2.7, & E &\cong 4 \times 10^{21}, & D_0 &\cong 1.3 \times 10^{-2}, \\ F_1 &\cong 8.3 \times 10^{-9}, & F_2 &\cong -2.6 \times 10^{-7}, & F_3 &\cong -8.4 \times 10^{-10}, \\ D_3 &\cong 5 \times 10^4, & D_4 &\cong 4 \times 10^2, & D_5 &\cong 80, \\ D_6 &\cong 3 \times 10^7, & D_7 &\cong -6 \times 10^{-3}, & D_8 &\cong 4. \end{aligned} \quad (6.2.73)$$

Thus, by (6.2.67) and (6.2.70)–(6.2.73), the number b_1 in (6.2.43) is given by

$$\begin{aligned} b_1 &\cong \frac{\rho^2}{D^2 E D_0} \times [(\alpha\rho D_6 - 2\rho^{-1} D_4 D_6)F_2 - \alpha D_5 D_6(3F_1 - F_3)] \\ &\quad + \frac{\rho^2}{D^2 E} \left[\frac{\alpha^2 D_6^2 D_7 D_8}{E D^2 \rho^2} + \frac{\alpha^3 \rho^2 D_6^2 D_7}{2 E D^2} - \frac{3.2 \times 10^7 D_6 D_8}{\rho^2 E D^2} \right] \\ &\cong -\frac{\rho^2}{E D_0 D^2} \times 4 \times 10^3 - \frac{3.2 D_6 D_8}{E^2 D^4} \times 10^7 \\ &\quad - \frac{6\alpha^2 D_6^2 D_8}{E^2 D^4} \times 10^{-3} - \frac{6\alpha^3 \rho^4 D_6^2}{2 E^2 D^4} \times 10^{-3}. \end{aligned} \quad (6.2.74)$$

Hence, in the stirred case, by Theorem 6.2.5 and (6.2.71)–(6.2.74), as $\delta < \delta_0 = 71.67$, the system (6.2.61) bifurcates from $(u, \delta) = (0, \delta_0)$ to a stable periodic solution, i.e., the reaction system (6.2.1) undergoes a temporal oscillation in the concentrations of $X = HBrO_2$, $Y = Br^-$, and $Z = Ce^{4+}$.

Non-stirred Case We consider the problem (6.2.26). In this case, there is another critical parameter δ_1 defined by (6.2.39). When $\delta_0 < \delta_1$, the system undergoes a transition to multiple equilibria at $\delta = \delta_1$, and when $\delta_0 > \delta_1$, the system has a transition to periodic solutions at $\delta = \delta_0$.

By (6.2.67) and (6.2.70), the number δ_1 is given by

$$\delta_1 = \max_{\rho_k} \left[\frac{690.8\mu_3}{9.1\mu_1 + 0.7\mu_2 + \mu_1\mu_2\rho_k} - \frac{588.4}{(9.1\mu_1 + 0.7\mu_2)\rho_k + \mu_1\mu_2\rho_k^2} - \mu_3\rho_k \right].$$

It is clear that if

$$\mu_3 \leq \mu_1, \mu_2, \quad (6.2.75)$$

then in view of (6.2.71) we have

$$\delta_1 \leq \frac{690.8\mu_3}{9.1\mu_1 + 0.7\mu_2} < \frac{690.8}{9.8} < \delta_0.$$

If (6.2.75) is not satisfied, we consider the domain $\Omega = \Omega(L, x_0)$ as defined by (6.2.69). In this case, the eigenvalues ρ_k ($k \geq 2$) of (6.2.28) can be expressed by

$$\rho_k = \frac{C_k}{L^2} \quad \text{with } C_k > 0 \quad \text{and } C_k \rightarrow \infty \text{ as } k \rightarrow \infty.$$

It follows that there exists an $L_0 > 0$ such that $\delta_1 < \delta_0$ for any $L < L_0$. Thus, by Theorem 6.2.4 and (6.2.74), we have the following conclusion.

Physical Conclusion 6.2.9 *Under the condition of (6.2.75) or $\Omega = \Omega(L, x_0)$ with $L < L_0$, the system (6.2.26) bifurcates from $(u, \delta) = (0, \delta_0)$ to a stable periodic solution on $\delta < \delta_0$, which implies that the reaction system (6.2.1) with the Neumann boundary condition undergoes a temporal oscillation for $\delta < \delta_0$. However, when $\delta > \delta_0$, this system is in the trivial state U_1 in (6.2.24).*

Phase Transition at $\delta = \delta_1$ By the above discussion, when $\mu_3 > \mu_i$ ($i = 1, 2$) and $L > L_0$ is large enough, the condition $\delta_0 < \delta_1$ may hold true. In the following, we assume that $L > L_0$ is sufficiently large, and

$$\mu_1 = \mu_2, \quad \mu_3 = z\mu_1 \quad (z > 1).$$

Then δ_1 becomes

$$\delta_1 \simeq \max_{x>0} \left[\frac{691z}{10+x} - \frac{589}{(10+x)x} - xz \right], \quad (x = \mu_1\rho_k).$$

It follows that $3 < x < 4$ for $z > 1$. We take $x = \mu_1\rho_{k_0} = 3.5$, then

$$\delta_1 = 47.7z - 12.5.$$

Thereby, we deduce that

$$\delta_1 > \delta_0 = 71.8 \quad \text{as } \frac{\mu_3}{\mu_1} = z > 1.8.$$

Now, we investigate the transition of (6.2.26) at δ_1 . For simplicity, we consider the case where

$$\Omega = (0, L) \subset \mathbb{R}^1.$$

In this case, the eigenvalues and eigenvectors of (6.2.28) are

$$\rho_k = (k - 1)^2 \pi^2 / L^2; \quad e_k = \cos(k - 1)\pi x / L.$$

It is clear that

$$\begin{aligned} e_{k_0}^2 &= (e_1 + e_j)/2 && \text{with } j = 2k_0 - 1, \\ \int_{\Omega} e_{k_0}^3 dx &= 0 && \text{by } k_0 \geq 2. \end{aligned}$$

We need to compute the number b_1 in (6.2.65). By (6.2.67), (6.2.70)–(6.2.71), and

$$\mu_3 \rho_{k_0} = z \mu_1 \rho_{k_0} = z \mu_2 \rho_{k_0} = 3.5z \quad (z > 1.8),$$

the vectors $G(\xi)$ and ϕ in (6.2.66) are given by

$$\begin{aligned} G(\xi) &\approx (1.6 \times 10^7 (\mu_3 \rho_{k_0} + \delta_1)^2, 3 \times 10^3 (\mu_3 \rho_{k_0} + \delta_1)^2, 0), \\ \phi &= \varphi_0 e_1 + \varphi_j e_j \quad (j = 2k_0 - 1), \\ M_1 \varphi_0 &= -\frac{1}{2} G(\xi), \\ M_j \varphi_j &= -\frac{1}{2} G(\xi), \end{aligned}$$

where

$$\begin{aligned} M_1 &= \begin{pmatrix} -0.7 & -5.4 \times 10^4 & 0 \\ -1.25 \times 10^{-2} & -12.5 & 1.25 \times 10^{-2} \\ \delta_1 & 0 & -\delta_1 \end{pmatrix}, \\ M_j &= \begin{pmatrix} -14.7 & -5.4 \times 10^4 & 0 \\ -1.25 \times 10^{-2} & -26.5 & 1.25 \times 10^{-2} \\ \delta_1 & 0 & -(\mu_3 \rho_j + \delta_1) \end{pmatrix}. \end{aligned}$$

The direct calculation shows that

$$\det M_1 = -8.75 \delta_1,$$

$$\det M_j \approx -1.5 \times 10^4 z + 4.9 \times 10^3,$$

$$M_1^{-1} = \frac{1}{\det M_1} \begin{pmatrix} 12.5 \delta_1 & -5.4 \times 10^4 \delta_1 & * \\ 0 & 0.7 \delta_1 & * \\ * & * & * \end{pmatrix},$$

$$M_j^{-1} = \frac{1}{\det M_j} \begin{pmatrix} 26.5(\mu_3 \rho_j + \delta_1) & -5.4 \times 10^4(\mu_3 \rho_j + \delta_1) & * \\ -1.25 \times 10^{-2} \mu_3 \rho_j & 14.7(\mu_3 \rho_j + \delta_1) & * \\ * & * & * \end{pmatrix}.$$

Then we get

$$\begin{aligned}\varphi_0 &= (\varphi_{01}, \varphi_{02}, \varphi_{03}) = -\frac{1}{2}M_1^{-1}G(\xi), & \varphi_{01} &= 2.5 \times 10^6(\mu_3\rho_{k_0} + \delta_1)^2, \\ \varphi_{02} &= 1.1 \times 10^2(\mu_3\rho_{k_0} + \delta_1)^2, & \varphi_j &= (\varphi_{j1}, \varphi_{j2}, \varphi_{j3}) = -\frac{1}{2}M_j^{-1}G(\xi), \\ \varphi_{j1} &= \frac{13.4 \times 10^3}{1.5z - 0.5}(\mu_3\rho_{k_0} + \delta_1)^2, & \varphi_{j2} &= \frac{2z + 27}{1.5z - 0.5}(\mu_3\rho_{k_0} + \delta_1)^2.\end{aligned}$$

Namely

$$\begin{aligned}\phi_1 &= \varphi_{01}e_1 + \varphi_{j1}e_j = (\mu_3\rho_{k_0} + \delta_1)^2 \left[2.5 \times 10^6 e_1 + \frac{13.4 \times 10^3}{1.5z - 0.5} e_j \right], \\ \phi_2 &= \varphi_{02}e_1 + \varphi_{j2}e_j = (\mu_3\rho_{k_0} + \delta_1)^2 \left[1.1 \times 10^2 e_1 + \frac{2z + 27}{1.5z - 0.5} e_j \right].\end{aligned}$$

Putting ϕ_1 and ϕ_2 into (6.2.65), we derive that

$$\begin{aligned}b_1 &= \frac{(\mu_3\rho_{k_0} + \delta_1)^2 L}{(\xi, \xi^*)} \left(5 \times 10^8 - \frac{5.7 \times 10^6}{1.5z - 0.5} - \frac{2z + 27}{1.5z - 0.5} \right), \\ (\xi, \xi^*) &= -699\alpha [16.7 \times (\delta_1 + 3.5z)^2 - 699\delta_1].\end{aligned}$$

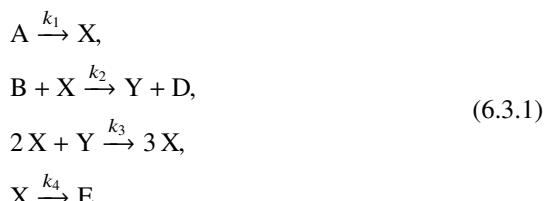
By $\delta_1 > \delta_0 = 71.8$ and $z > 1.8$ we see that $b_1 < 0$. Hence, by Theorem 6.2.7, the system (6.2.1) undergoes a dynamic transition at $\delta = \delta_1$ to the following steady-state solutions

$$u_{\pm} = U_1 \pm \left(-\frac{L}{2(\mu_2\rho_{k_0} + \delta_1)^2 b_1} \beta_{k_0 l}(\delta) \right)^{1/2} \xi \cos \frac{(k_0 - 1)\pi}{L} x + o(\beta_{k_0 l}^{1/2}),$$

which are attractors.

6.3 Belousov–Zhabotinsky Reactions: Brusselator

The main objective of this section is to study the dynamic phase transitions of the Belousov–Zhabotinsky reactions focusing on the Brusselator, first introduced by Prigogine and Lefever (1968). The Brusselator is one of the simplest models in nonlinear chemical systems. It has six components, four of which retain constants, and the other two permit their concentrations vary with time and space. The chemical reaction consists of four irreversible steps, given by



where A and B are constant components, D and E are products, and X and Y are the two components variable in time and space. Over the years, there have been extensive studies for the Brusselator and related chemical reaction problems; see among many others (Tzou et al., 2009; Guo and Han, 2009; Bao, 2002; Kaper and Kaper, 2002; Shi, 2009) and the references therein.

Dynamic phase transition of the Brusselator model is briefly addressed in this section without proofs, which we refer the interested readers to Ma and Wang (2011d) for details. We derive a complete characterization of the transition from the homogeneous state. There are two aspects of this characterization. First our analysis shows that both the transitions to multiple equilibria and to time-periodic solutions (spatiotemporal oscillations) can occur for the Brusselator model, and are precisely determined by the sign of an explicit nondimensional parameter $\delta_0 - \delta_1$ as defined by (6.3.12) and (6.3.13). Then in both transition cases, the dynamic behavior of the transition is classified, and the type of transitions is determined also by the signs of some nondimensional and computable parameters.

6.3.1 Prigogine-Lefever Model

Let u_1, u_2, a and b stand for the concentrations of X, Y, A and B . Then the reaction equations of (6.3.1) are given by

$$\begin{aligned}\frac{\partial u_1}{\partial t} &= \sigma_1 \Delta u_1 + k_1 a - (k_2 b + k_4) u_1 + k_3 u_1^2 u_2, \\ \frac{\partial u_2}{\partial t} &= \sigma_2 \Delta u_2 + k_2 b u_1 - k_3 u_1^2 u_2.\end{aligned}\tag{6.3.2}$$

To get the nondimensional form of (6.3.2), let

$$\begin{aligned}t &= k_4^{-1} t', & x &= l x', & u_i &= \left(\frac{k_4}{k_3}\right)^{1/2} u'_i, \\ a &= (k_4^3/k_1^2 k_3)^{1/2} \alpha, & b &= (k_4/k_2) \lambda, & \sigma_i &= l^2 k_4 \mu_i.\end{aligned}$$

Omitting the primes, the equations (6.3.2) become

$$\begin{aligned}\frac{\partial u_1}{\partial t} &= \mu_1 \Delta u_1 + \alpha - (\lambda + 1) u_1 + u_1^2 u_2, \\ \frac{\partial u_2}{\partial t} &= \mu_2 \Delta u_2 + \lambda u_1 - u_1^2 u_2,\end{aligned}\tag{6.3.3}$$

where $u_1, u_2 \geq 0$, $\Omega \subset \mathbb{R}^n$ ($1 \leq n \leq 3$) is a bounded domain, and

$$\mu_1, \mu_2, \alpha, \lambda > 0.$$

Equation (6.3.3) have a constant steady-state solution

$$u_0 = (\alpha, \lambda/\alpha).$$

Make the translation

$$u_1 = \alpha + v_1, \quad u_2 = \frac{\lambda}{\alpha} + v_2,$$

then the equations (6.3.3) are written as

$$\begin{aligned} \frac{\partial v_1}{\partial t} &= \mu_1 \Delta v_1 + (\lambda - 1)v_1 + \alpha^2 v_2 + \frac{2\lambda}{\alpha} v_1^2 + 2\alpha v_1 v_2 + v_1^2 v_2, \\ \frac{\partial v_2}{\partial t} &= \mu_2 \Delta v_2 - \lambda v_1 - \alpha^2 v_2 - \frac{2\lambda}{\alpha} v_1^2 - 2\alpha v_1 v_2 - v_1^2 v_2. \end{aligned} \quad (6.3.4)$$

There are two types of physically sound boundary conditions: the Dirichlet boundary condition

$$v = (v_1, v_2) = 0 \quad \text{on } \partial\Omega, \quad (6.3.5)$$

and the Neumann boundary condition

$$\frac{\partial v}{\partial n} = 0 \quad \text{on } \partial\Omega. \quad (6.3.6)$$

Define two function spaces

$$H = L^2(\Omega, \mathbb{R}^2), \quad H_1 = \begin{cases} H^2(\Omega, \mathbb{R}^2) \cap H_0^1(\Omega, \mathbb{R}^2) & \text{for (6.3.5),} \\ \{v \in H^2(\Omega, \mathbb{R}^2) \mid \frac{\partial v}{\partial n}|_{\partial\Omega} = 0\} & \text{for (6.3.6).} \end{cases}$$

Define the operators $L_\lambda = A + B_\lambda$ and $G : H_1 \rightarrow H$ by

$$\begin{aligned} Av &= (\mu_1 \Delta v_1, \mu_2 \Delta v_2), \\ B_\lambda v &= ((\lambda - 1)v_1 + \alpha^2 v_2, -\lambda v_1 - \alpha^2 v_2), \\ G(v, \lambda) &= \left(\frac{2\lambda}{\alpha} v_1^2 + 2\alpha v_1 v_2 + v_1^2 v_2, -\frac{2\lambda}{\alpha} v_1^2 - 2\alpha v_1 v_2 - v_1^2 v_2 \right). \end{aligned} \quad (6.3.7)$$

Thus, the equations (6.3.4) with (6.3.5) or with (6.3.6) can be written in the following abstract form

$$\frac{dv}{dt} = L_\lambda v + G(v, \lambda). \quad (6.3.8)$$

6.3.2 Linearized Problem

Consider the linearized eigenvalue problem of (6.3.4)

$$\begin{aligned} \mu_1 \Delta v_1 + (\lambda - 1)v_1 + \alpha^2 v_2 &= \beta v_1, \\ \mu_2 \Delta v_2 - \lambda v_1 - \alpha^2 v_2 &= \beta v_2, \end{aligned} \quad (6.3.9)$$

with the boundary condition (6.3.5) or (6.3.6).

Let ρ_k and e_k be the k th eigenvalue and eigenvector of the Laplacian with either the Dirichlet or the Neumann condition. All eigenvalues β_k^\pm and eigenvectors ϕ_k^\pm of (6.3.9) satisfy that

$$\phi_k^\pm = \xi_k^\pm e_k, \quad M_k \xi_k^\pm = \beta_k^\pm \xi_k^\pm, \quad (6.3.10)$$

where $\xi_k^\pm \in \mathbb{R}^2$ are the eigenvectors of M_k , and β_k^\pm are the eigenvalues of M_k :

$$\begin{aligned} \beta_k^\pm(\lambda) &= \frac{1}{2}[\lambda - (\mu_1\rho_k + \mu_2\rho_k + \alpha^2 + 1)] \\ &\pm \frac{1}{2}[(\lambda - \mu_1\rho_k - \mu_2\rho_k - \alpha^2 - 1)^2 \\ &+ 4(\lambda\mu_2\rho_k - (\mu_1\rho_k + 1)(\mu_2\rho_k + \alpha^2))]^{1/2}. \end{aligned} \quad (6.3.11)$$

It is clear that $\beta_k^-(\lambda) < \beta_k^+(\lambda) = 0$ if and only if

$$\lambda = \frac{1}{\mu_2\rho_k}(\mu_1\rho_k + 1)(\mu_2\rho_k + \alpha^2) < \mu_1\rho_k + \mu_2\rho_k + \alpha^2 + 1,$$

and $\beta_k^\pm(\lambda) = \pm\sigma_k(\lambda)i$ with $\sigma_k \neq 0$ if and only if

$$\lambda = \mu_1\rho_k + \mu_2\rho_k + \alpha^2 + 1 < \frac{1}{\mu_2\rho_k}(\mu_1\rho_k + 1)(\mu_2\rho_k + \alpha^2).$$

Thus we introduce two critical parameters

$$\lambda_0 = \min_{\rho_k} \frac{1}{\mu_2\rho_k}(\mu_1\rho_k + 1)(\mu_2\rho_k + \alpha^2), \quad (6.3.12)$$

$$\lambda_1 = \mu_1\rho_1 + \mu_2\rho_1 + \alpha^2 + 1. \quad (6.3.13)$$

Lemma 6.3.1 *Let λ_0 and λ_1 be the two numbers given by (6.3.12) and (6.3.13). Then we have the following assertions:*

- (1) *Let $\lambda_0 < \lambda_1$, and $k_0 \geq 1$ be the integer such that the minimum is achieved at ρ_{k_0} in the definition of λ_0 . Then $\beta_{k_0}^+(\lambda)$ is the first real eigenvalue of (6.3.9) near $\lambda = \lambda_0$ satisfying that*

$$\beta_k^+(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_0 \\ = 0 & \text{if } \lambda = \lambda_0 \\ > 0 & \text{if } \lambda > \lambda_0 \end{cases} \quad \forall k \in \mathbb{N} \text{ with } \rho_k = \rho_{k_0}, \quad (6.3.14)$$

$$\operatorname{Re}\beta_j^\pm(\lambda_0) < 0 \quad \forall \beta_j^\pm \neq \beta_{k_0}^+ \text{ with } \rho_k = \rho_{k_0}.$$

- (2) *Let $\lambda_1 < \lambda_0$. Then $\beta_1^+(\lambda) = \overline{\beta_1^-}(\lambda)$ are a pair of first complex eigenvalues of (6.3.9) near $\lambda = \lambda_1$, and*

$$Re\beta_1^+(\lambda) = Re\beta_1^-(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_1, \\ = 0 & \text{if } \lambda = \lambda_1, \\ > 0 & \text{if } \lambda > \lambda_1, \end{cases} \quad (6.3.15)$$

$$Re\beta_k^\pm(\lambda_1) < 0 \quad \forall k > 1.$$

6.3.3 Transition From Real Eigenvalues

Consider the case where $\lambda_0 < \lambda_1$. Based on Lemma 6.3.1, the system (6.3.8) always undergoes a dynamic transition at the real eigenvalue $\lambda = \lambda_0$. Hereafter, we always assume that the eigenvalue $\beta_{k_0}^+$ in (6.3.14) is simple. Let ρ_{k_0} be as in Lemma 6.3.1, and e_{k_0} be an eigenvector of the Laplacian with either the Dirichlet or the Neumann boundary condition, corresponding to ρ_{k_0} .

Theorem 6.3.2 *Let $\lambda_0 < \lambda_1$ and assume*

$$\int_{\Omega} e_{k_0}^3 dx \neq 0. \quad (6.3.16)$$

Then the system (6.3.4) with (6.3.5) or with (6.3.6) undergoes a Type-III (mixed) dynamic transition at $\lambda = \lambda_0$, and bifurcates on each side of $\lambda = \lambda_0$ to a unique branch v^λ of steady-state solutions, such that the following assertions hold true:

1. *On $\lambda < \lambda_0$, the bifurcated solution v^λ is a saddle, and the stable manifold Γ_λ^1 of v^λ separates the space H into two disjoint open sets U_1^λ and U_2^λ , such that $v = 0 \in U_1^\lambda$ is an attractor, and the orbits of (6.3.8) in U_2^λ are far from $v = 0$.*
2. *On $\lambda > \lambda_0$, the stable manifold Γ_λ^0 of $v = 0$ separates the neighborhood O of $u = 0$ into two disjoint open sets O_1^λ and O_2^λ , such that the transition is jump in O_1^λ , and is continuous in O_2^λ . The bifurcated solution $v^\lambda \in O_2^\lambda$ is an attractor such that for any $\varphi \in O_2^\lambda$,*

$$\lim_{t \rightarrow \infty} \|v(t, \varphi) - v^\lambda\| = 0,$$

where $v(t, \varphi)$ is the solution of (6.3.8) with $v(0, \varphi) = \varphi$.

3. *The bifurcated solution v^λ can be expressed as*

$$v^\lambda = C\beta_{k_0}^+(\lambda)\xi_{k_0}^+e_{k_0} + o(|\beta_{k_0}^\pm|),$$

$$\xi_{k_0}^+ = (-\mu_2\rho_{k_0}, \mu_1\rho_{k_0} + 1),$$

$$C = \frac{(\alpha\mu_2\rho_{k_0}(\mu_2\rho_{k_0} + \alpha^2) - \alpha^3(\mu_1\rho_{k_0} + 1)) \int_{\Omega} e_{k_0}^2 dx}{2\mu_2^3\rho_{k_0}^3(\mu_1\rho_{k_0} + 1) \int_{\Omega} e_{k_0}^3 dx}. \quad (6.3.17)$$

For the case

$$\int_{\Omega} e_{k_0}^3 dx = 0. \quad (6.3.18)$$

We introduce the following parameter

$$\begin{aligned} b_1 = & \left[\alpha \mu_2^2 \rho_{k_0}^2 (\mu_1 \rho_{k_0} + 1) \int_{\Omega} e_{k_0}^4 dx - 2\alpha^2 \mu_2 \rho_{k_0} \int_{\Omega} \psi_2 e_{k_0}^2 dx \right. \\ & \left. - 2(\mu_1 \rho_{k_0} + 1)(2\mu_2 \rho_{k_0} + \alpha^2) \int_{\Omega} \psi_1 e_{k_0}^2 dx \right] \\ & \times [\alpha^2(\mu_1 \rho_{k_0} + 1) - \mu_2 \rho_{k_0}(\mu_2 \rho_{k_0} + \alpha^2)]^{-1}, \end{aligned} \quad (6.3.19)$$

where $\psi = (\psi_1, \psi_2)$ satisfies

$$\begin{aligned} \mu_1 \Delta \psi_1 + (\lambda_0 - 1)\psi_1 + \alpha^2 \psi_2 &= -\frac{2\mu_2^2 \rho_{k_0}^2}{\alpha} (\mu_1 \rho_{k_0} + 1) e_{k_0}^2, \\ \mu_2 \Delta \psi_2 - \lambda_0 \psi_1 - \alpha^2 \psi_2 &= \frac{2\mu_2^2 \rho_{k_0}^2}{\alpha} (\mu_1 \rho_{k_0} + 1) e_{k_0}^2, \\ \psi|_{\partial\Omega} &= 0 \quad (\text{or } \frac{\partial \psi}{\partial n}|_{\partial\Omega} = 0). \end{aligned} \quad (6.3.20)$$

By the Fredholm Alternative Theorem, under the condition (6.3.18), the equation (6.3.20) has a unique solution.

Theorem 6.3.3 *Let $\lambda_0 < \lambda_1$, b_1 is the number given by (6.3.19) and assume (6.3.18).*

1. If $b_1 > 0$, (6.3.8) undergoes a Type-II transition at $\lambda = \lambda_0$. Moreover, (6.3.8) has no bifurcation on $\lambda > \lambda_0$, and bifurcates on $\lambda < \lambda_0$ to exactly two solutions v_+^λ and v_-^λ , which are saddles.

Moreover, the stable manifolds Γ_+^λ and Γ_-^λ of the two bifurcated solutions divide H into three disjoint open sets U_+^λ , U_0^λ , U_-^λ such that $v = 0 \in U_0^\lambda$ is an attractor, and the orbits in U_\pm^λ are far from $v = 0$.

2. If $b_1 < 0$, (6.3.8) undergoes a Type-I transition at $\lambda = \lambda_0$, has no bifurcation on $\lambda < \lambda_0$, and bifurcates on $\lambda > \lambda_0$ to exactly two solutions v_+^λ and v_-^λ , which are attractors.

In addition, there is a neighborhood $O \subset H$ of $v = 0$, such that the stable manifold Γ of $v = 0$ divides O into two disjoint open sets O_+^λ and O_-^λ such that $v_+^\lambda \in O_+^\lambda$, $v_-^\lambda \in O_-^\lambda$, and v_\pm^λ attracts O_\pm^λ ;

3. The bifurcated solutions v_\pm^λ can be expressed as

$$\begin{aligned} v_\pm^\lambda &= \pm C(\beta_{k_0}^+(\lambda))^{1/2} \xi_{k_0}^+ e_{k_0} + o(|\beta_{k_0}^+|^{1/2}), \\ \xi_{k_0}^+ &= (-\mu_2 \rho_{k_0}, \mu_1 \rho_{k_0} + 1), \\ C &= \left[\frac{-\alpha}{\mu_2 \rho_{k_0} b_1} \int_{\Omega} e_{k_0}^2 dx \right]^{1/2}, \end{aligned} \quad (6.3.21)$$

where b_1 is as in (6.3.19).

6.3.4 Transition from Complex Eigenvalues

For the case where $\lambda_1 < \lambda_0$, (6.3.8) undergoes a dynamic transition at $\lambda = \lambda_1$, leading to spatial-temporal oscillations. For the Neumann boundary condition, $\lambda_1 = \alpha^2 + 1$, and we have the following theorem.

Theorem 6.3.4 *Let $\lambda_1 < \lambda_0$. The problem (6.3.4) with (6.3.6) undergoes a Type-I dynamic transition at $\lambda = \lambda_1$, and bifurcates on $\lambda > \lambda_1$ to a periodic solution which is an attractor. Moreover, the bifurcated periodic solution $v^\lambda = (v_1^\lambda, v_2^\lambda)$ can be expressed as*

$$\begin{aligned} v_1^\lambda &= 2[C\beta_0(\lambda)]^{1/2}\alpha^2 \sin(\alpha t + \frac{\pi}{4}) + o(|\beta_0|^{1/2}), \\ v_2^\lambda &= [C\beta_0(\lambda)]^{1/2}\alpha((\alpha - 1)\cos \alpha t - (\alpha + 1)\sin \alpha t) + o(|\beta_0|^{1/2}), \end{aligned} \quad (6.3.22)$$

where $\beta_0 = \frac{1}{2}(\lambda - \lambda_1)$, and $C = (2\pi\alpha^2 + \frac{3}{2}\pi\alpha^4)^{-1}$.

For the Dirichlet boundary condition case, $\rho_1 > 0$ and $\lambda_1 = (\mu_1 + \mu_2)\rho_1 + \alpha^2 + 1$. We define the following parameter

$$\begin{aligned} b_1 &= \frac{2\pi\alpha^2[\int_{\Omega} e_1^3 dx]^2}{\sigma_0^2[\int_{\Omega} e_1^2 dx]^2}(\mu_1\rho_1 + 1)(\mu_2^2\rho_1^2 + 2\mu_2\rho_1(\mu_1\rho_1 + 1) - \sigma_0^2) \\ &\quad - \frac{\pi\alpha^2 \int_{\Omega} e_1^4 dx}{2 \int_{\Omega} e_1^2 dx}(2\mu_2\rho_1 + 3\alpha^2) \\ &\quad + \frac{2\pi\alpha^2}{\int_{\Omega} e_1^2 dx}[(3\mu_1\rho_1 + \mu_2\rho_1 + 3)A_1 + (\mu_1\mu_2\rho_1^2 + \mu_2\rho_1 + \delta_0)B_1] \\ &\quad - \frac{8\pi\alpha^2\sigma_0^2}{\int_{\Omega} e_1^2 dx}[(\mu_1\rho_1 + \mu_2\rho_1 + 1)A_2 - (\mu_1\mu_2\rho_1^2 + \mu_2\rho_1 - \sigma_0)B_2] \\ &\quad - \frac{4\pi\alpha^2}{\int_{\Omega} e_1^2 dx}[(\mu_1\mu_2\rho_1^2 + \mu_2\rho_1 - \sigma_0^2)A_3 + \sigma_0^2(\mu_1\rho_1 + \mu_2\rho_1 + 1)B_3], \end{aligned} \quad (6.3.23)$$

where

$$\begin{aligned} \sigma_0 &= [\alpha^2(\mu_1\rho_1 + 1) - \mu_2\rho_1(\mu_2\rho_1 + \alpha^2)]^{1/2}, \\ A_i &= (2\lambda_1 - \mu_2\rho_1 - \alpha^2)B_i + \alpha^2 C_i \quad \text{for } i = 1, 2, 3, \\ B_1 &= \sum_{k>1}^{\infty} \frac{\mu_2\rho_k[\int_{\Omega} e_1^2 e_k dx]^2}{\det M_k \int_{\Omega} e_k^2 dx}, \\ B_2 &= \sum_{k>1}^{\infty} \frac{[\int_{\Omega} e_1^2 e_k dx]^2}{\det M_k \det(M_k^2 + 4\sigma_0^2) \int_{\Omega} e_k^2 dx}[(\mu_2\rho_k + \alpha^2)D_k + \alpha^2 Q_k], \end{aligned}$$

$$\begin{aligned}
B_3 &= \sum_{k>1}^{\infty} \frac{[\int_{\Omega} e_1^2 e_k dx]^2}{\det(M_k^2 + 4\sigma_0^2) \int_{\Omega} e_k^2 dx} D_k, \\
C_1 &= - \sum_{k>1}^{\infty} \frac{(\mu_1 \rho_1 + 1)[\int_{\Omega} e_1^2 e_k dx]^2}{\det M_k \int_{\Omega} e_k^2 dx}, \\
C_2 &= \sum_{k>1}^{\infty} \frac{[\int_{\Omega} e_1^2 e_k dx]^2}{\det M_k \det(M_k^2 + 4\sigma_0^2) \int_{\Omega} e_k^2 dx} [(\mu_1 \rho_k + 1 - \lambda_1) Q_k - \lambda_1 D_k], \\
C_3 &= \sum_{k>1}^{\infty} \frac{[\int_{\Omega} e_1^2 e_k dx]^2}{\det(M_k^2 + 4\sigma_0^2) \int_{\Omega} e_k^2 dx} Q_k, \\
D_k &= (\mu_2 \rho_k + \alpha^2)^2 + 4\sigma_0^2 - \alpha^2(\mu_1 + \mu_2)\rho_k - \alpha^2(\alpha^2 + 1), \\
Q_k &= \lambda_1 \alpha^2 - \lambda_1(\mu_1 + \mu_2)(\rho_k - \rho_1) - (\mu_1 \rho_k + 1 - \lambda_1)^2 - 4\sigma_0^2.
\end{aligned}$$

Here M_k is the matrix defined by

$$M_k = \begin{pmatrix} -\mu_1 \rho_k + \lambda_1 - 1 & \alpha^2 \\ -\lambda_1 & -\mu_2 \rho_k - \alpha^2 \end{pmatrix}. \quad (6.3.24)$$

Theorem 6.3.5 Let b_1 be the parameter given by (6.3.23) and $\lambda_1 < \lambda_0$.

1. If $b_1 < 0$, then the problems (6.3.4)–(6.3.5) undergo a Type-I transition at $\lambda = \lambda_1$, and bifurcates to a stable periodic solution on $\lambda > \lambda_1$.
2. If $b_1 > 0$, then the problem undergoes a Type-II transition at $\lambda = \lambda_1$, and bifurcates to an unstable periodic solution on $\lambda < \lambda_1$.
3. The bifurcated periodic solution $v^\lambda = (v_1^\lambda, v_2^\lambda)$ can be expressed as

$$\begin{aligned}
v_1^\lambda &= 2\alpha^2[-\gamma(\lambda)/b_1]^{1/2} e_1 \sin(\sigma_0 t + \frac{\pi}{4}) + o(|\gamma|^{1/2}), \\
v_2^\lambda &= 2(\sigma_0^2 + (\mu_2 \rho_1 + \alpha^2)^2) \left[-\frac{\gamma(\lambda)}{b_1} \right]^{1/2} e_1 \cos(\sigma_0 t + \theta) + o(|\gamma|^{1/2}), \\
\theta &= \tan^{-1} \frac{\sigma_0 + \mu_2 \rho_1 + \alpha^2}{\sigma_0 - \mu_2 \rho_1 - \alpha^2},
\end{aligned} \quad (6.3.25)$$

where $\gamma = (\lambda - \lambda_1)/2$.

6.4 Bacterial Chemotaxis

Chemotaxis is an important phenomenon occurring in many biological individuals and involves mobility and aggregation of the species in two aspects: one is random walk and the other is the chemically directed movement. For example, in the slime mold *Dictyostelium discoideum*, the single-cell amoebae move towards regions of relatively high concentration of a chemical called cyclic-AAMP which is secreted by

the amoebae themselves. Many experiments demonstrate that under proper conditions a bacterial colony can form a rather regular pattern, which is relatively stable in certain time scale. For example, a series of experimental results on the patterns formed by the bacteria *Escherichia coli* (*E. coli*) and *Salmonella typhimurium* (*S. typhimurium*) were derived in Budrene and Berg (1991, 1995), where two types of experiments were conducted: one is in semi-solid medium and the other is in liquid medium. Both showed that when the bacteria are exposed to intermediates of TCA cycle, they can form various regular patterns, typically as ringlike and sunflowerlike formations. In all these experiments, the bacteria are known to secrete aspartate, a potent chemoattractant; also see Murray (2002); Brenner et al. (1998).

In their pioneering work (Keller and Segel, 1970), E. F. Keller and L. A. Segel proposed a model in 1970, called the Keller–Segel equations, to describe the chemotactic behavior of the slime mold amoebae. In their equations, the growth rate of amoeba cells was ignored, i.e., the model can only depict the chemotaxis process in a small timescale, as exhibited in the liquid medium experiments with Coli and Typhimurium by Budrene and Berg (1991, 1995). However, in the semi-solid medium experiments, the timescale of a pattern formation process is long enough to accommodate many generations of bacteria. Therefore, various revised models were presented by many authors, taking into consideration the effects of the stimulant (i.e., food source) and the growth rate of population; see among others (Murray, 2002) and the references therein. Also, there is a vast literature on the mathematical studies for the Keller–Segel model; see among others (Perthame et al., 2011; Guo and Hwang, 2010; Perthame and Dalibard, 2009; Nadin et al., 2008).

The main objective of this section is to study dynamic transition and pattern formation of the Keller–Segel model, and the main presentation is based on Ma and Wang (2011a).

6.4.1 Keller–Segel Models

The general form of the revised Keller–Segel models is given by

$$\begin{aligned}\frac{\partial u_1}{\partial t} &= k_1 \Delta u_1 - \nabla(\chi(u) \nabla u_2) + f(u), \\ \frac{\partial u_2}{\partial t} &= k_2 \Delta u_2 + g(u), \\ \frac{\partial u_3}{\partial t} &= k_3 \Delta u_3 + h(u) + q(x),\end{aligned}\tag{6.4.1}$$

where $u = (u_1, u_2, u_3)$, u_1 is the population density of biological individuals, u_2 is the chemoattractant concentration, u_3 is the stimulant concentration, $q(x)$ is the nutrient source, and χ, f, g, h are functions of u , which will be determined based on a few biological and mathematical rules as follows.

Basic Modeling Rules The chemotaxis equation (6.4.1) obeys a few biological and mathematical rules as follows:

- (R1) As stated in Sect. 1.1, all models describing natural phenomena obey basic physical laws. The first step to model a chemotaxis is to seek basic laws describing the underlying phenomena.
- (R2) It is clear that a medium consisting of only nutrients without organisms and chemoattractants can be observed in a realistic world. In particular, if some bacteria are inoculated in a nutrient medium, the cells will proliferate, and if inoculated in a nutrient-free medium the cells will vanish.
Equivalently, there exists a steady-state solution $U_0 = (0, 0, u_3^0)$ for (6.4.1) such that $u_3^0 \geq 0$ satisfy the equation

$$\begin{aligned} -k_3 \Delta u_3 &= q(x), & q \geq 0 \text{ in } \Omega, \\ \frac{\partial u_3}{\partial n} &= 0 \text{ on } \partial\Omega, \end{aligned}$$

and U_0 is stable if $q(x) = 0$, and unstable if $q \neq 0$.

- (R3) It is clear that with constant stimulant supply, the bacteria and the chemicals produced by themselves will exist permanently, but they cannot increase indefinitely even with abundant food supply.
Equivalently, when $q \neq 0$, the system possesses a nonnegative steady-state solution $u^* = (u_1^*, u_2^*, u_3^*)$ such that

$$u_1^*, u_2^* \leq M, \quad u^* \rightarrow 0 \text{ as } q \rightarrow 0, \quad \|u_3^*\| \rightarrow \infty \text{ as } \|q\| \rightarrow \infty.$$

where M is a constant independent of q .

- (R4) There are no stable steady states of the following types for the system:

$$(\tilde{u}_1, \tilde{u}_2, 0), \quad (\tilde{u}_1, 0, \tilde{u}_3), \quad (0, \tilde{u}_2, \tilde{u}_3), \quad (\tilde{u}_1, 0, 0), \quad (0, \tilde{u}_2, 0),$$

for $\tilde{u}_j \geq 0$ ($\neq 0$), $1 \leq j \leq 3$.

- (R5) The space of solutions of (6.4.1) is the set of nonnegative functions, denoted by H^+ , which is positively invariant, i.e., for any initial value $u_0 \in H^+$ the solution $u(t, u_0)$ of (6.4.1) with $u(0, u_0) = u_0$ satisfies

$$u(t, u_0) \in H^+, \quad \forall t > 0.$$

- (R6) For any initial data in H^+ , there exists a global in time weak solution in H^+ , and there exist a global attractor in H^+ .

- (R7) **SIMPLICITY PRINCIPLE:** A model usually takes its simplest form.

First Equation of (6.4.1) The first basic physical law for the chemotaxis is the conservation of bacterial population u_1 , which is written as

$$\frac{\partial u_1}{\partial t} = -\nabla J + f(u), \tag{6.4.2}$$

where J is the cell flux, and f the growth rate of u_1 .

The mobility of organisms is characterized by a random walk and a directed movement towards higher concentration of the chemoattractant:

$$J = J_d + J_c. \quad (6.4.3)$$

Physically, the random walk is similar to the Brown motion. Therefore, J_d follows the Fickian diffusion law:

$$J_d = -k_1 \nabla u_1, \quad (6.4.4)$$

where $k_1 > 0$ is the diffusion coefficient. It is reasonable to consider the chemotaxis as a motion of the attractants in the direction of a concentration gradient:

$$J_c = \chi(u) \nabla u_2, \quad (6.4.5)$$

where $\chi(u)$ is the chemotaxis response function, depending on u_1, u_2 , and the bacteria types. Based on the items (R5) and (R7), we take $\chi(u)$ in the form

$$\chi(u) = \chi u_1, \quad (6.4.6)$$

with a coefficient $\chi > 0$.

The growth rate f of u_1 is defined by

$$f = \text{birth rate } f_1 - \text{death rate } f_2.$$

By (R2) and the fact that f is independent of the chemoattractant u_2 , we assume

$$f_1 = \alpha(u_3)u_1, \quad f_2 = \alpha_1 u_1^n,$$

with a coefficient $\alpha_1 > 0$, an integer $n \geq 2$, and a function $\alpha_3(u_3) \geq 0$. Then (R6) suggests that $n \geq 3$, and by (R7), we take $n = 3$. According to (R3) and (R7), we employ

$$\alpha(u_3) = \frac{ku_3}{\alpha_0 + u_3},$$

where $k > 0$ is a constant. Thus we obtain that

$$f(u) = \alpha_1 u_1 \left(\frac{\alpha_2 u_3}{\alpha_0 + u_3} - u_1^2 \right), \quad (\alpha_1 \alpha_2 = k). \quad (6.4.7)$$

By (6.4.2)–(6.4.7), the first equation of (6.4.1) reads

$$\frac{\partial u_1}{\partial t} = k_1 \Delta u_1 - \chi \nabla(u_1 \nabla u_2) + \alpha_1 u_1 \left(\frac{\alpha_2 u_3}{\alpha_0 + u_3} - u_1^2 \right). \quad (6.4.8)$$

Remark 6.4.1 There are many discussions to find a precise expression for the chemotaxis function $\chi(u)$. Lapidus and Schiller (1976) suggest the following form

$$\chi(u) = \frac{\chi_1 u_1}{(\beta + u_2)^2}, \quad (6.4.9)$$

with two parameters $\chi_1, \beta > 0$, which is regarded as a best one in comparing with experimental data; see Murray (2002). In (6.4.6), for simplicity we use an average \bar{u}_2 to replace u_2 in (6.4.9), i.e., the parameter χ in (6.4.6) is considered as

$$\chi = \frac{\chi_1}{(\beta + \bar{u}_2)^2}, \quad (6.4.10)$$

where $\bar{u}_2 = \frac{1}{|\Omega|} \int_{\Omega} u_2^* dx$ for some steady-state u^* . \square

Remark 6.4.2 Mathematical theories show that for a parabolic system with the Dirichlet or Neumann boundary condition given by

$$\begin{aligned} \frac{\partial u_j}{\partial t} &= k_j \Delta u_j + F_j(x, u, \nabla u), \quad 1 \leq j \leq m, \\ u|_{\partial\Omega} &= 0 \quad \text{or} \quad \left. \frac{\partial u}{\partial n} \right|_{\partial\Omega} = 0, \\ u(0) &= u_0, \end{aligned} \quad (6.4.11)$$

if F_j are in the forms

$$\begin{aligned} F_j &= f_j(x, u, \nabla u)u_j + \sum_{i=1}^n g_i(x, u, \nabla u) \frac{\partial u_j}{\partial x_i} = h_j(x, u, \nabla u), \\ h_j(x, \xi, \eta) &\geq 0, \quad \forall (x, \xi, \eta) \in \Omega \times \mathbb{R}^n \times \mathbb{R}^{n^2}, \end{aligned} \quad (6.4.12)$$

then the nonnegative function space H^+ is an invariant set of (6.4.11). Namely under the condition (6.4.12), if $u_0 \in H^+$, then the solution $u(t, u_0)$ of (6.4.11) is nonnegative, i.e.

$$u_j(t, u_0) \geq 0 \quad \text{in } \Omega \quad \forall t > 0 \quad \text{for } 1 \leq j \leq m;$$

see Smoller (1983). Hence due to (R5) and (6.4.12), the expression of $\chi(u)$ must be taken in the form as follows

$$\chi(u) = \beta(u)u_1,$$

with some function $\beta(u)$.

The Second and Third Equations of (6.4.1) The chemoattractant u_2 and stimulant u_3 are two types of chemicals governed by the conservation law of mass and the Fickian law of diffusion:

$$\begin{aligned} \frac{\partial u_2}{\partial t} &= k_2 \Delta u_2 - \text{decay of } u_2 + \text{production of } u_2 \text{ by } u_1, \\ \frac{\partial u_3}{\partial t} &= k_3 \Delta u_3 - \text{consumption by } u_1 + \text{nutrient source}. \end{aligned}$$

Based on (R2)–(R5) and (R7), we take

$$\begin{aligned} \text{decay of } u_2 &= r_2 u_2, \\ \text{production of } u_2 \text{ by } u_1 &= r_1 u_1, \\ \text{consumption of } u_3 \text{ by } u_1 &= r_3 u_3 u_1, \end{aligned}$$

with coefficients $r_i > 0$ ($1 \leq i \leq 3$). Finally, the food supply of u_1 and u_3 are given by external sources:

$$\text{nutrient source} = q(x) \geq 0.$$

Explicit Formula of (6.4.1) We are now in position to write down a more explicit form of (6.4.1) for the Keller–Segel equations:

$$\begin{aligned} \frac{\partial u_1}{\partial t} &= k_1 \Delta u_1 - \chi \nabla(u_1 \nabla u_2) + \alpha_1 u_1 \left(\frac{\alpha_2 u_3}{\alpha_0 + u_3} - u_1^2 \right), \\ \frac{\partial u_2}{\partial t} &= k_2 \Delta u_2 + r_1 u_1 - r_2 u_2, \\ \frac{\partial u_3}{\partial t} &= k_3 \Delta u_3 - r_3 u_1 u_3 + q(x), \end{aligned} \tag{6.4.13}$$

which are supplemented with the Neumann boundary condition:

$$\frac{\partial u}{\partial n} = 0 \text{ on } \partial\Omega, \tag{6.4.14}$$

and with the initial condition:

$$u(0) = u_0. \tag{6.4.15}$$

We remark that the equations (6.4.13) are the simplest model satisfying the basic modeling rules (R1)–(R7). For the pattern formation and dynamic transition points of view, we shall see that there is no essential difference between the two terms of the simplest formula (6.4.6) and the Lapidus–Schiller formula (6.4.9) of $\chi(u)$.

Two-Component Systems For the case where the stimulant u_3 is ample, i.e., $q, u_3 = \infty$ in (6.4.13), the third equation of (6.4.13) can be ignored, and the birth rate in the first equation reads

$$\frac{\alpha_1 \alpha_2 u_1 u_3}{\alpha_0 + u_3} = \alpha_1 \alpha_2 u_1.$$

Then (6.4.13) are rewritten as

$$\begin{aligned} \frac{\partial u_1}{\partial t} &= k_1 \Delta u_1 - \chi \nabla(u_1 \nabla u_2) + \alpha_1 u_1 (\alpha_2 - u_1^2), \\ \frac{\partial u_2}{\partial t} &= k_2 \Delta u_2 + r_1 u_1 - r_2 u_2. \end{aligned} \tag{6.4.16}$$

This model depicts the chemotactic aggregating exhibited by the semi-solid medium experiments with sufficient nutrient supply as mentioned earlier.

In the liquid medium experiments with *E. coli* and *S. typhimurium*, the pattern formation process of the bacteria is in a timescale of minutes. It implies that the organisms neither birth nor die during the aggregating period, and the nutrient u_3 is also considered as sufficiently supplied. Thus, (6.4.16) become

$$\begin{aligned}\frac{\partial u_1}{\partial t} &= k_1 \Delta u_1 - \chi \nabla(u_1 \nabla u_2), \\ \frac{\partial u_2}{\partial t} &= k_2 \Delta u_2 + r_1 u_1 - r_2 u_2, \\ \int_{\Omega} u_1(x, t) dx &= \text{constant}.\end{aligned}\quad (6.4.17)$$

In fact, the third equation (conservation of bacterial population density) in (6.4.17) is compatible with the first equation. To see this, with the boundary condition (6.4.14) we deduce that

$$\frac{d}{dt} \int_{\Omega} u_1 dx = \int_{\partial\Omega} \left[k_1 \frac{\partial u_1}{\partial n} - \chi u_1 \frac{\partial u_2}{\partial n} \right] ds = 0.$$

Equations (6.4.17) are the original Keller–Segel model with an addition of the conservation of population density equation, which appears to be crucial to eliminate the existence of an infinite number of steady-state solutions.

Chemoattractant Balance Systems Oftentimes, the following form of the Keller–Segel equations is discussed in literature:

$$\begin{aligned}\frac{\partial u_1}{\partial t} &= k_1 \Delta u_1 - \chi \nabla(u_1 \nabla u_2) + \alpha_1 u_1 \left(\frac{\alpha_2 u_3}{\alpha_0 + u_3} - u_1^2 \right), \\ \frac{\partial u_3}{\partial t} &= k_3 \Delta u_3 - r_3 u_1 u_3 + q, \\ -k_2 \Delta u_2 + r_2 u_2 &= r_1 u_1.\end{aligned}\quad (6.4.18)$$

The biological significance of (6.4.18) is that the diffusion and degradation of the chemoattractant secreted by the bacteria themselves are essentially balanced by their production. The advantage of (6.4.18) lies in its mathematical simplicity, preserving the main characteristics of pattern formation and phase transition.

Nondimensional Form The main objective of this section is on the pattern formation and dynamic transition of the Keller–Segel models (6.4.13)–(6.4.15). For simplicity, the domain $\Omega = (0, l_1) \times (0, l_2)$ is a rectangle with $l_1 \neq l_2$. Also, let

$$(t, x) = (t'/r_2, \sqrt{k_2/r_2}x'), \quad (u_1, u_2, u_3) = (\sqrt{\alpha_2}u'_1, k_2u'_2/\chi, \alpha_0u'_3), \quad (6.4.19)$$

and introduce some nondimensional parameters

$$\begin{aligned}\lambda &= r_1 \sqrt{\alpha_2} \chi / r_2 k_2, \quad \alpha = \alpha_1 \alpha_2 / r_2, \quad \mu = k_1 / k_2, \\ r &= k_3 / k_2, \quad \delta = r_3 \sqrt{\alpha_2} / r_2, \quad \delta_0 = q / r_2 \alpha_0.\end{aligned}\quad (6.4.20)$$

Then the nondimensional form of (6.4.13)–(6.4.15) read (dropping the primes):

$$\begin{aligned}\frac{\partial u_1}{\partial t} &= \mu \Delta u_1 - \nabla(u_1 \nabla u_2) + \alpha u_1 \left(\frac{u_3}{1+u_3} - u_1^2 \right), \\ \frac{\partial u_2}{\partial t} &= \Delta u_2 - u_2 + \lambda u_1, \\ \frac{\partial u_3}{\partial t} &= r \Delta u_3 - \delta u_1 u_3 + \delta_0, \\ \frac{\partial u}{\partial n} \Big|_{\partial\Omega} &= 0, \\ u(0) &= u_0 \text{ in } \Omega.\end{aligned}\tag{6.4.21}$$

The nondimensional of Ω is written as

$$\Omega = (0, L_1) \times (0, L_2), \quad \text{with } L_1 \neq L_2.$$

The nondimensional form of the balanced chemoattractant system is given by

$$\begin{aligned}\frac{\partial u_1}{\partial t} &= \mu \Delta u_1 - \nabla(u_1 \nabla u_2) + \alpha u_1 \left(\frac{u_3}{1+u_3} - u_1^2 \right), \\ \frac{\partial u_3}{\partial t} &= r \Delta u_3 - \delta u_1 u_3 + \delta_0, \\ -\Delta u_2 + u_2 &= \lambda u_1, \\ \frac{\partial u}{\partial n} \Big|_{\partial\Omega} &= 0, \\ u(0) &= u_0.\end{aligned}\tag{6.4.22}$$

6.4.2 Dynamic Transitions for Rich Stimulant System

Rich Stimulant System We know that as nutrient u_3 is richly supplied, the Keller–Segel model (6.4.13) is reduced to a two-component system:

$$\begin{aligned}\frac{\partial u_1}{\partial t} &= \mu \Delta u_1 - \nabla(u_1 \nabla u_2) + \alpha u_1(1 - u_1^2), \\ \frac{\partial u_2}{\partial t} &= \Delta u_2 - u_2 + \lambda u_1, \\ \frac{\partial(u_1, u_2)}{\partial n} \Big|_{\partial\Omega} &= 0, \\ u(0) &= u_0.\end{aligned}\tag{6.4.23}$$

It is easy to see that $u^* = (1, \lambda)$ is a steady state of (6.4.23). Consider the deviation from u^* :

$$u = u^* + u'.$$

Suppressing the primes, the system (6.4.23) is then transformed into

$$\begin{aligned} \frac{\partial u_1}{\partial t} &= \mu \Delta u_1 - 2\alpha u_1 - \Delta u_2 - \nabla(u_1 \nabla u_2) - 3\alpha u_1^2 - \alpha u_1^3, \\ \frac{\partial u_2}{\partial t} &= \Delta u_2 - u_2 + \lambda u_1, \\ \left. \frac{\partial(u_1, u_2)}{\partial n} \right|_{\partial\Omega} &= 0, \\ u(0) &= u_0. \end{aligned} \tag{6.4.24}$$

Dynamic Transition and Pattern Formation for the Diffusion and Degradation Balanced Case We start with an important case where the diffusion and degradation of the chemoattractant secreted by the bacteria themselves are almost balanced by their production. In this case, the second equation in (6.4.24) is given by

$$0 = \Delta u_2 - u_2 + \lambda u_1.$$

With the Newman boundary condition for u_2 , we have $u_2 = [-\Delta + 1]^{-1}u_1$ and the functional form of the resulting equation is given by

$$\frac{\partial u_1}{\partial t} = \mathcal{L}_\lambda u_1 + G(u_1, \lambda), \tag{6.4.25}$$

where the operators $\mathcal{L}_\lambda : H_1 \rightarrow H$ and $G : H_1 \times \mathbb{R} \rightarrow \mathbb{R}$ are defined by

$$\begin{aligned} \mathcal{L}_\lambda u_1 &= \mu \Delta u_1 - 2\alpha u_1 - \lambda \Delta [-\Delta + I]^{-1} u_1, \\ G(u_1, \lambda) &= -\lambda \nabla(u_1 \nabla [-\Delta + I]^{-1} u_1) - 3\alpha u_1^2 - \alpha u_1^3. \end{aligned} \tag{6.4.26}$$

Here the two Hilbert spaces H and H_1 are defined by

$$H = L^2(\Omega), \quad H_1 = \left\{ u_1 \in H^2(\Omega) \mid \frac{\partial u_1}{\partial n} = 0 \text{ on } \Omega \right\}.$$

To study the dynamic transition of this problem, we need to consider the linearized eigenvalue problem of (6.4.25):

$$\mathcal{L}_\lambda e = \beta(\lambda)e. \tag{6.4.27}$$

Let ρ_k and e_k be the eigenvalues and eigenfunctions of $-\Delta$ with the Neumann boundary condition given by

$$e_k = \cos \frac{k_1 \pi x_1}{L_1} \cos \frac{k_2 \pi x_2}{L_2}, \quad \rho_k = \pi^2 \left(\frac{k_1^2}{L_1^2} + \frac{k_2^2}{L_2^2} \right), \tag{6.4.28}$$

for any $k = (k_1, k_2) \in \mathbb{N}_+^2$. Here \mathbb{N}_+ is the set of all nonnegative integers. In particular, $e_0 = 1$ and $\rho_0 = 0$.

Obviously, the functions in (6.4.28) are also eigenvectors of (6.4.27), and the corresponding eigenvalues β_k are

$$\beta_k(\lambda) = -\mu\rho_k - 2\alpha + \frac{\lambda\rho_k}{1 + \rho_k}. \quad (6.4.29)$$

Define a critical parameter by

$$\lambda_c = \min_{\rho_k} \frac{(\rho_k + 1)(\mu\rho_k + 2\alpha)}{\rho_k}. \quad (6.4.30)$$

Let

$$\mathcal{S} = \left\{ K = (K_1, K_2) \in \mathbb{N}_+^2 \text{ achieves the minimum in (6.4.30)} \right\}.$$

It follows from (6.4.29) and (6.4.30) that

$$\beta_K(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_c \\ = 0 & \text{if } \lambda = \lambda_c \\ > 0 & \text{if } \lambda > \lambda_c \end{cases} \quad \forall K = (K_1, K_2) \in \mathcal{S}, \quad (6.4.31)$$

$$\beta_k(\lambda_c) < 0 \quad \forall k \in \mathbb{Z}^2 \text{ with } k \notin \mathcal{S}. \quad (6.4.32)$$

Notice that for any $K = (K_1, K_2) \in \mathcal{S}$, $K \neq 0$ and

$$\lambda_c = \frac{(\rho_K + 1)(\mu\rho_K + 2\alpha)}{\rho_K}. \quad (6.4.33)$$

We note that for properly choosing spatial geometry, we have

$$\rho_K = \pi^2 \left(\frac{K_1^2}{L_1^2} + \frac{K_2^2}{L_2^2} \right) = \sqrt{\frac{2\alpha}{\mu}} \quad \forall K = (K_1, K_2) \in \mathcal{S}, \quad (6.4.34)$$

$$\lambda_c = 2\alpha + \mu + 2\sqrt{2\alpha\mu}. \quad (6.4.35)$$

Conditions (6.4.31) and (6.4.32) give rise to a dynamic transition of (6.4.25) from $(u, \lambda) = (0, \lambda_c)$. For simplicity, we denote

$$K_1 \triangleq (K_1, 0), \quad K_2 \triangleq (0, K_2),$$

and $K = (K_1, K_2) \in \mathcal{S}$. Also, we introduce a parameter as

$$\begin{aligned} b = & -3\mu\rho_K + \left[12 - \frac{24 - 15\text{sign}(K_1 K_2)}{4 - 2\text{sign}(K_1 K_2)} \right] \alpha \\ & - \frac{(2\mu\rho_K + \alpha)(2\mu\rho_K^2 + 28\alpha\rho_K + 4\alpha - \mu\rho_K)}{[1 + \text{sign}(K_1 K_2)] \cdot [(\mu\rho_{2K} + 2\alpha)(1 + \rho_{2K}) - \lambda_c \rho_{2K}]} \end{aligned}$$

$$\begin{aligned}
& - \frac{2(2\mu\rho_K\rho_{K_1} + 4\alpha\rho_{K_1} - 3\alpha\rho_K)}{(1 + \text{sign}K_1)\rho_K^2[(\mu\rho_{2K_1} + 2\alpha)(1 + \rho_{2K_1}) - \lambda_c\rho_{2K_1}]} \\
& \quad \times [(\mu\rho_K + 2\alpha)(2\rho_{K_1}^2 - 6\rho_{K_1}\rho_{K_2} - \rho_K) + 6\alpha\rho_K(4\rho_{K_1} + 1)] \\
& - \frac{2(2\mu\rho_K\rho_{K_2} + 4\alpha\rho_{K_2} - 3\alpha\rho_K)}{(1 + \text{sign}K_2)\rho_K^2[(\mu\rho_{2K_2} + 2\alpha)(1 + \rho_{2K_2}) - \lambda_c\rho_{2K_2}]} \\
& \quad \times [(\mu\rho_K + 2\alpha)(2\rho_{K_2}^2 - 6\rho_{K_1}\rho_{K_2} - \rho_K) + 6\alpha\rho_K(4\rho_{K_2} + 1)]. \quad (6.4.36)
\end{aligned}$$

The following is the main dynamic transition theorem, providing a precise criterion for the transition type and the pattern formation mechanism of the system.

Theorem 6.4.3 *Let b be the parameter defined by (6.4.36). Assume that the eigenvalue β_k satisfying (6.4.31) is simple. Then, for the system (6.4.25) we have the following assertions:*

- (1) *The system always undergoes a dynamic transition at $(u, \lambda) = (0, \lambda_c)$. Namely, the basic state $u = 0$ is asymptotically stable for $\lambda < \lambda_c$, and is unstable for $\lambda > \lambda_c$.*
- (2) *For the case where $b < 0$, this transition is continuous (Type-I). In particular, the system bifurcates from $(0, \lambda_c)$ to two steady-state solutions on $\lambda > \lambda_c$, which can be expressed as*

$$u_1^\pm(x, \lambda) = \pm \frac{1}{2} \sqrt{\frac{\beta_K(\lambda)}{2|b|}} \cos \frac{K_1 \pi x_1}{L_1} \cos \frac{K_2 \pi x_2}{L_2} + o(\beta_K^{1/2}), \quad (6.4.37)$$

and $u_1^\pm(x, \lambda)$ are attractors.

- (3) *For the case $b > 0$, this transition is jump (Type-II), and the system has two saddle-node bifurcation solutions at some $\lambda^*(0 < \lambda^* < \lambda_c)$ such that there are two branches v_1^λ and v_2^λ of steady states bifurcated from (v^*, λ^*) , and there are two other branches v_3^λ and v_4^λ bifurcated from (u^*, λ^*) . In addition, v_1^λ and v_3^λ are saddles, v_2^λ and v_4^λ are attractors, with $v_1^\lambda, v_3^\lambda \rightarrow 0$ as $\lambda \rightarrow \lambda_c$.*

Two remarks are now in order.

Remark 6.4.4 From the pattern formation point of view, for the Type-I transition, the patterns described by the transition solutions given in (6.4.37) are either lamella or rectangular:

lamella pattern	for $K_1 K_2 = 0$,
rectangular pattern	for $K_1 K_2 \neq 0$.

In the case where $b > 0$, the system undergoes a more drastic change. As $\lambda^* < \lambda < \lambda_c$, the homogeneous state, the new patterns v_2^λ and v_4^λ are metastable. For $\lambda > \lambda_c$, the system undergoes transitions to more complex patterns away from the basic homogeneous state form.

Remark 6.4.5 If we take the growth term $f(u)$ as $f = \alpha u_1(1 - u_1)$ instead of $f = \alpha u_1(1 - u_1^2)$ in (6.4.23), (6.4.24), and (6.4.25), then Theorem 6.4.3 still holds true except the assertion on the existence of the two saddle-node bifurcation solutions, and the parameter should be replaced by

$$\begin{aligned}
b = & -\mu\rho_K + \alpha - \frac{(2\mu\rho_K + \alpha)(2\mu\lambda_K^2 + 10\alpha\rho_K + \alpha - \mu\rho_K)}{2(1 + \text{sign}K_1 K_2)[(\mu\rho_{2K} + \alpha)(1 + \rho_{2K}) - \lambda_c \rho_{2K}]} \\
& - \frac{(2\mu\rho_K \rho_{K_1} + 2\alpha\rho_{K_1} - \alpha\rho_K)}{(1 + \text{sign}K_1)\lambda_K^2[(\mu\rho_{2K_1} + \alpha)(1 + \rho_{2K_1}) - \lambda_c \rho_{2K_1}]} \\
& \times [(\mu\rho_K + \alpha)(2\lambda_{K_1}^2 - 6\rho_{K_1}\rho_{K_2} - \rho_K) + 2\alpha\rho_K(4\rho_{K_1} + 1)] \\
& - \frac{(2\mu\rho_K \rho_{K_2} + 2\alpha\rho_{K_2} - \alpha\rho_K)}{(1 + \text{sign}K_2)\lambda_K^2[(\mu\rho_{2K_2} + \alpha)(1 + \rho_{2K_1}) - \lambda_c \rho_{2K_2}]} \\
& \times [(\mu\rho_K + \alpha)(2\lambda_{K_2}^2 - 6\rho_{K_2}\rho_{K_1} - \rho_K) + 2\alpha\rho_K(4\rho_{K_2} + 1)].
\end{aligned}$$

Pattern Formation and Dynamic Transition for the General Case Consider the general case (6.4.23). In this case, the unknown variable becomes $u = (u_1, u_2)$, and the basic function spaces are then defined by

$$H = L^2(\Omega, \mathbb{R}^2), \quad H_1 = \{u \in H^2(\Omega, \mathbb{R}^2) \mid \frac{\partial u}{\partial n} = 0 \text{ on } \Omega\}.$$

Let $L_\lambda : H_1 \rightarrow H$ and $G : H_1 \rightarrow H$ be defined by

$$\begin{aligned}
L_\lambda u &= \begin{pmatrix} \mu\Delta - 2\alpha & -\Delta \\ \lambda & \Delta - 1 \end{pmatrix} u, \\
G(u) &= \begin{pmatrix} -\nabla(u_1 \nabla u_2) - 3\alpha u_1^2 - \alpha u_1^3 \\ 0 \end{pmatrix}.
\end{aligned} \tag{6.4.38}$$

The linearized eigenvalue problem of (6.4.24) is $L_\lambda \varphi = \beta \varphi$, where $L_\lambda : H_1 \rightarrow H$ is defined by (6.4.38). Let B_k^λ be the matrices given by

$$B_k^\lambda = \begin{pmatrix} -(\mu\rho_k + 2\alpha) & \rho_k \\ \lambda & -(\rho_k + 1) \end{pmatrix}, \tag{6.4.39}$$

where ρ_k are the eigenvalues as in (6.4.28). It is easy to see that all eigenvectors φ_k and eigenvalues β_k of L_λ can be expressed as follows

$$\varphi_k = \begin{pmatrix} \xi_{k1} e_k \\ \xi_{k2} e_k \end{pmatrix}, \quad B_k^\lambda \begin{pmatrix} \xi_{k1} \\ \xi_{k2} \end{pmatrix} = \beta_k \begin{pmatrix} \xi_{k1} \\ \xi_{k2} \end{pmatrix}, \tag{6.4.40}$$

where e_k are as in (6.4.28), and β_k are also the eigenvalues of B_k^λ . By (6.4.39), β_k can be expressed by

$$\begin{aligned}
\beta_k^\pm(\lambda) &= \frac{1}{2} \left[-B \pm \sqrt{B^2 - 4((\rho_k + 1)(\mu\rho_k + 2\alpha) - \lambda\rho_k)} \right], \\
B &= (\mu + 1)\rho_k + 2\alpha + 1.
\end{aligned} \tag{6.4.41}$$

Let λ_c be the parameter as defined by (6.4.30). It follows from (6.4.41) and (6.4.30) that

$$\beta_K^+(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_c, \\ = 0 & \text{if } \lambda = \lambda_c, \\ > 0 & \text{if } \lambda > \lambda_c, \end{cases} \quad (6.4.42)$$

$$\begin{cases} \operatorname{Re}\beta_k^-(\lambda_c) < 0 \ \forall k \in \mathbb{Z}^2, \\ \operatorname{Re}\beta_k^+(\lambda_c) < 0 \ \forall k \in \mathbb{Z}^2 \text{ with } \rho_k \neq \rho_K, \end{cases} \quad (6.4.43)$$

with $K = (K_1, K_2)$ as in (6.4.33).

Then we have the following dynamic transition theorem.

Theorem 6.4.6 *Let b be the parameter defined by (6.4.36). Assume that the eigenvalue β_K^+ satisfying (6.4.42) is simple. Then the assertions of Theorem 6.4.3 hold true for (6.4.24), with the expression (6.4.37) replaced by*

$$u_\lambda^\pm = \pm \sqrt{a\beta_K^+(\lambda)} \left(\frac{\rho_K + 1}{\lambda_c} \right) \cos \frac{K_1 \pi x_1}{L_1} \cos \frac{K_2 \pi x_2}{L_2} + o(|\beta_K^+|^{1/2}),$$

$$a = \frac{8(\mu\rho_K + \rho_K + 2\alpha + 1)}{(\rho_K + 1)^3 |b|}.$$

Proof of Theorem 6.4.3. Assertion (1) follows directly from the general dynamic transition theorem in Chap. 2. To prove Assertions (2) and (3), we need to reduce (6.4.25) to the center manifold near $\lambda = \lambda_c$. We note that although the underlying system is now quasilinear in this general case, the center manifold reduction holds true as well; see Liu et al. (2011) for details.

To this end, let $u = xe_K + \Phi$, where $\Phi(x)$ the center manifold function of (6.4.25). Since $L_\lambda : H_1 \rightarrow H$ is symmetric, the reduced equation is given by

$$\frac{dx}{dt} = \beta_K(\lambda)x + \frac{1}{(e_K, e_K)}(G(xe_K + \Phi, \lambda), e_K), \quad (6.4.44)$$

where $G : H_1 \rightarrow H$ is defined by (6.4.26), and

$$(e_K, e_K) = \int_{\Omega} e_K^2 dx = \frac{2 - \operatorname{sign}(K_1 K_2)}{4} |\Omega|. \quad (6.4.45)$$

It is known that the center manifold function satisfies that $\Phi(x) = O(x^2)$. A direct computation shows that

$$\begin{aligned} (G(xe_K + \Phi, \lambda_c), e_K) = & -\alpha x^3 \int_{\Omega} e_K^4 dx - 6\alpha x \int_{\Omega} e_K^2 \Phi dx \\ & + \lambda_c x \int_{\Omega} [e_K \nabla e_k \cdot \nabla(-\Delta + I)^{-1} \Phi \\ & + \Phi \nabla e_k \cdot \nabla(-\Delta + I)^{-1} e_K] dx + o(x^3). \end{aligned} \quad (6.4.46)$$

It is clear that

$$(-\Delta + I)^{-1} e_K = \frac{1}{\rho_K + 1} e_K, \quad \Delta e_K = -\rho_K e_K.$$

We infer from (6.4.46) that

$$\begin{aligned} (G(xe_K + \Phi, \lambda_c), e_K) &= -\alpha x^3 \int_{\Omega} e_K^4 dx - 6\alpha x \int_{\Omega} e_K^2 \Phi dx \\ &\quad + \lambda_c x \int_{\Omega} \left[\frac{1}{\rho_K + 1} |\nabla e_K|^2 \Phi - |\nabla e_K|^2 (-\Delta_I)^{-1} \Phi \right. \\ &\quad \left. + \rho_K e_K^2 (-\Delta + I)^{-1} \Phi \right] dx + o(x^3). \end{aligned} \quad (6.4.47)$$

By (A.1.14), Φ satisfies the equation

$$\begin{aligned} -L_{\lambda_c} \Phi &= G_2(xe_K, \lambda_c) + o(x^2) \\ &= x^2 \left[\left(\frac{\rho_K \lambda_c}{\rho_K + 1} - 3\alpha \right) e_K^2 - \frac{\lambda_c}{\rho_K + 1} |\nabla e_K|^2 \right] + o(x^2). \end{aligned} \quad (6.4.48)$$

In view of (6.4.28), we find

$$\begin{aligned} e_K^2 &= \frac{1}{4} [e_0 + e_{2K_1} + e_{2K_2} + e_{2K}], \\ |\nabla e_K|^2 &= \frac{1}{4} [\rho_K e_0 + (\rho_{K_2} - \rho_{K_1}) e_{2K_1} + (\rho_{K_1} - \rho_{K_2}) e_{2K_2} - \rho_K e_{2K}]. \end{aligned} \quad (6.4.49)$$

Thus, (6.4.48) is written as

$$\begin{aligned} -L_{\lambda_c} \Phi &= \frac{x^2}{4} \left[-3\alpha e_0 + \left(\frac{2\rho_{K_1} \lambda_c}{\rho_K + 1} - 3\alpha \right) e_{2K_1} + \left(\frac{2\rho_{K_2} \lambda_c}{\rho_K + 1} - 3\alpha \right) e_{2K_2} \right. \\ &\quad \left. + \left(\frac{2\rho_K \lambda_c}{\rho_K + 1} e_{2K} - 3\alpha \right) e_{2K} \right] + o(x^2). \end{aligned} \quad (6.4.50)$$

Let

$$\Phi = \Phi_0 e_0 + \Phi_{2K_1} e_{2K_1} + \Phi_{2K_2} e_{2K_2} + \Phi_{2K} e_{2K}. \quad (6.4.51)$$

Note that

$$-L_{\lambda_c} e_{2K} = \frac{1}{1 + \rho_{2K}} [(1 + \rho_{2K})(\mu\rho_K + 2\alpha) - \lambda_c \rho_{2K}] e_{2K}. \quad (6.4.52)$$

Then by (6.4.33) and (6.4.50)–(6.4.52) we obtain that

$$\begin{aligned} \Phi_0 &= -\frac{3}{8}, \\ \Phi_{2K_1} &= \frac{(1 + \rho_{2K_1})(2\mu\rho_K \rho_{K_1} + 4\alpha \rho_{K_1} - 3\alpha \rho_K)}{4\rho_K [(1 + \rho_{2K_1})(\mu\rho_{2K_1} + 2\alpha) - \rho_{2K_1} \lambda_c]}, \\ \Phi_{2K_2} &= \frac{(1 + \rho_{2K_2})(2\mu\rho_K \rho_{K_2} + 4\alpha \rho_{K_2} - 3\alpha \rho_K)}{4\rho_K [(1 + \rho_{2K_2})(\mu\rho_{2K_2} + 2\alpha) - \rho_{2K_2} \lambda_c]}, \\ \Phi_{2K} &= \frac{(1 + \rho_{2K})(2\mu\rho_K + \alpha)}{4[(1 + \rho_{2K})(\mu\rho_{2K} + 2\alpha) - \rho_{2K} \lambda_c]}. \end{aligned} \quad (6.4.53)$$

Inserting (6.4.51) and (6.4.28) into (6.4.47) we get

$$\begin{aligned}
(G(xe_K + \Phi, \lambda_c), e_K) &= -\alpha x^3 \int_{\Omega} e_K^4 dx \\
&\quad - \frac{6\alpha x(2 - \text{sign}(K_1 K_2))}{4} \int_{\Omega} [\Phi_0 e_0^2 + \Phi_{2K_1} e_{2K_1}^2 + \Phi_{2K_2} e_{2K_2}^2 + \Phi_{2K} e_{2K}^2] dx \\
&\quad + \frac{\lambda_c x(2 - \text{sign}(K_1 K_2))}{4(\rho_K + 1)} \int_{\Omega} [\rho_K \Phi_0 e_0^2 + (\rho_{K_2} - \rho_{K_1}) \Phi_{2K_1} e_{2K_1}^2 \\
&\quad + (\rho_{K_1} - \rho_{K_2}) \Phi_{2K_2} e_{2K_2}^2 - \rho_K \Phi_{2K} e_{2K}^2] dx \\
&\quad - \frac{\lambda_c x(2 - \text{sign}(K_1 K_2))}{4} \int_{\Omega} \left[\rho_K \Phi_0 e_0^2 + \frac{\rho_{K_2} - \rho_{K_1}}{1 + \rho_{2K_1}} \Phi_{2K_1} e_{2K_1}^2 \right. \\
&\quad \left. + \frac{\rho_{K_1} - \rho_{K_2}}{1 + \rho_{2K_2}} \Phi_{2K_2} e_{2K_2}^2 - \frac{\rho_K}{1 + \rho_{2K}} \Phi_{2K} e_{2K}^2 \right] dx \\
&\quad + \frac{\lambda_c \rho_K x(2 - \text{sign}(K_1 K_2))}{4} \int_{\Omega} \left[\Phi_0 e_0^2 + \frac{\Phi_{2K_1}}{1 + \rho_{2K_1}} e_{2K_1}^2 + \frac{\Phi_{2K_2}}{1 + \rho_{2K_2}} e_{2K_2}^2 \right. \\
&\quad \left. + \frac{\Phi_{2K}}{1 + \rho_{2K}} e_{2K}^2 \right] dx + o(x^3) \\
&= -\alpha x^3 \int_{\Omega} e_K^4 dx + \frac{|\Omega| x(2 - \text{sign}(K_1 K_2))}{4} \\
&\quad \times \left[(\mu \rho_K - 4\alpha) \Phi_0 + \frac{1}{1 + \text{sign} K_1} \left(\frac{\lambda_c(\rho_{K_2} - \rho_{K_1})}{1 + \rho_K} + \frac{2\lambda_c \rho_{K_1}}{1 + \rho_{2K_1}} - 6\alpha \right) \Phi_{2K_1} \right. \\
&\quad + \frac{1}{1 + \text{sign} K_2} \left(\frac{\lambda_c(\rho_{K_1} - \rho_{K_2})}{1 + \rho_K} + \frac{2\lambda_c \rho_{K_2}}{1 + \rho_{2K_1}} - 6\alpha \right) \Phi_{2K_2} \\
&\quad \left. + \frac{1}{2(1 + \text{sign}(K_1 K_2))} \left(-\frac{\lambda_c \rho_K}{1 + \rho_K} + \frac{2\lambda_c \rho_K}{1 + \rho_{2K}} - 6\alpha \right) \Phi_{2K} \right] dx + o(x^3). \quad (6.4.54)
\end{aligned}$$

Also, we note that

$$\int_{\Omega} e_K^4 = \int_0^{L_1} e_{K_1}^4 dx_1 \int_0^{L_2} e_{K_2}^4 dx_2 = \frac{24 - 15\text{sign}(K_1 K_2)}{64}.$$

Then, putting (6.4.53) into (6.4.54) we arrive at

$$(G(xe_K + \Phi, \lambda_c), e_K) = \frac{(2 - \text{sign}(K_1 K_2)) |\Omega| x^3}{32} b + o(x^3), \quad (6.4.55)$$

where b is the parameter given by (6.4.36).

By (6.4.44) and (6.4.55), we derive the reduced equation on the center manifold as follows:

$$\frac{dx}{dt} = \beta_K(\lambda)x + \frac{b}{8}x^3 + o(x^3). \quad (6.4.56)$$

Based on the dynamic transition theory developed in Chap. 2, we obtain Assertions (2) and (3), except that two saddle-node bifurcations occur at the same point $\lambda = \lambda^*$. To prove this conclusion, we note that if $u^*(x)$ is a steady-state solution of (6.4.25),

then

$$v^*(x) = u^*(x + \pi) = u^*(x - \pi)$$

is also a steady-state solution of (6.4.25). This is because the eigenvectors (6.4.28) form an orthogonal base of H_1 . Hence, two saddle-node bifurcations on $\lambda < \lambda_c$ imply that they must occur at the same point $\lambda = \lambda^*$. Thus the proof of the theorem is complete. \square

Proof of Theorem 6.4.6. Assertion (1) follows from (6.4.42) and (6.4.43). To prove Assertions (2) and (3), we need to get the reduced equation of (6.4.24) to the center manifold near $\lambda = \lambda_c$.

Let $u = x \cdot \varphi_K + \Phi$, where φ_K is the eigenvector of L_λ corresponding to β_K at $\lambda = \lambda_c$, and $\Phi(x)$ the center manifold function of (6.4.24). Then the reduced equation of (6.4.24) reads

$$\frac{dx}{dt} = \beta_K^+(\lambda)x + \frac{1}{(\varphi_K, \varphi_K^*)}(G(x \cdot \varphi_K + \Phi), \varphi_K^*), \quad (6.4.57)$$

Here φ_K^* is the conjugate eigenvector of φ_K .

By (6.4.40), φ_K is written as

$$\varphi_K = \begin{pmatrix} \xi_1 e_K \\ \xi_2 e_K \end{pmatrix}, \quad \begin{pmatrix} -(\mu\rho_K + 2\alpha) & \rho_K \\ \lambda_c & -(\rho_K + 1) \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = 0, \quad (6.4.58)$$

which implies that

$$(\xi_1, \xi_2) = (\rho_K + 1, \lambda_c). \quad (6.4.59)$$

In the same token, we derive that

$$\varphi_K^* = (\xi_1^* e_K, \xi_2^* e_K)^T = (\rho_K + 1, \rho_K)^T e_K. \quad (6.4.60)$$

By (6.4.38), the nonlinear operator G is

$$\begin{aligned} G(u_1, u_2) &= G_2(u_1, u_2) + G_3(u_1, u_2), \\ G_2(u_1, u_2) &= -(\nabla u_1 \nabla u_2 + u_1 \Delta u_2 + 3\alpha u_1^2) \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\ G_3(u_1, u_2) &= -\alpha u^3 \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \end{aligned}$$

It is known that the center manifold function $\Phi(x) = (\Phi_1(x), \Phi_2(x)) = O(x^2)$. Then, in view of (6.4.58) and (6.4.59), by direct computation we derive that

$$\begin{aligned} &(G(x\xi_1 e_K + \Phi_1, x\xi_2 e_K + \Phi_2), \varphi_K^*) \\ &= (xG_2(\xi_1 e_K, \Phi_2) + xG_2(\Phi_1, \xi_2 e_K) + x^3 G_3(\xi_1 e_K, \xi_2 e_K), \varphi_K^*) + o(x^3) \\ &= x\xi_1^* \int_{\Omega} [\xi_2 \Phi_1 |\nabla e_K|^2 - \frac{1}{2} \xi_1 \Delta \Phi_2 e_K^2 - 6\alpha \xi_1 \Phi_1 e_K^2] dx \\ &\quad - \alpha \xi_1^* \xi_1^3 x^3 \int_{\Omega} e_K^4 dx + o(x^3). \end{aligned} \quad (6.4.61)$$

Using the approximation formula for center manifold functions given in (A.11) in Ma and Wang (2010b), $\Phi = (\Phi_1, \Phi_2)$ satisfies

$$\begin{aligned} -L_{\lambda_c}\Phi &= -x^2 G_2(\xi_1 e_k, \xi_2 e_K) + o(x^2) \\ &= -x^2 (\xi_1 \xi_2 |\nabla e_K|^2 + (3\alpha \xi_1^2 - \xi_1 \xi_2 \rho_K) e_K^2) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + o(x^2). \end{aligned} \quad (6.4.62)$$

From (6.4.28) we see that

$$\begin{aligned} e_K^2 &= \frac{1}{4}(1 + e_{2K_1})(1 + e_{2K_2}) = \frac{1}{4}(e_0 + e_{2K_1} + e_{2K_2} + e_{2K}), \\ |\nabla e_K|^2 &= \frac{\rho_{K_1}}{4}(1 - e_{2K_1})(1 + e_{2K_2}) + \frac{\rho_{K_2}}{4}(1 + e_{2K_1})(1 - e_{2K_2}) \\ &= \frac{\rho_K}{4}e_0 + \frac{\rho_{K_2} - \rho_{K_1}}{4}e_{2K_1} + \frac{\rho_{K_1} - \rho_{K_2}}{4}e_{2K_2} - \frac{\rho_K}{4}e_{2K}. \end{aligned}$$

Thus, (6.4.62) is written as

$$\begin{aligned} L_{\lambda_c}\Phi &= -\frac{\xi_1 x^2}{4}(3\alpha \xi_1 e_0 + (3\alpha \xi_1 - 2\xi_2 \rho_{K_1})e_{2K_1} \\ &\quad + (3\alpha \xi_1 - 2\xi_2 \rho_{K_2})e_{2K_2} + (3\alpha \xi_1 - 2\xi_2 \rho_K)e_{2K}) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + o(x^3). \end{aligned} \quad (6.4.63)$$

Let

$$\begin{pmatrix} \Phi_1 \\ \Phi_2 \end{pmatrix} = \begin{pmatrix} \Phi_1^0 \\ \Phi_2^0 \end{pmatrix} e_0 + \begin{pmatrix} \Phi_1^{2K_1} \\ \Phi_2^{2K_1} \end{pmatrix} e_{2K_1} + \begin{pmatrix} \Phi_1^{2K_2} \\ \Phi_2^{2K_2} \end{pmatrix} e_{2K_2} + \begin{pmatrix} \Phi_1^{2K} \\ \Phi_2^{2K} \end{pmatrix} e_{2K}. \quad (6.4.64)$$

It is clear that

$$L_\lambda \begin{pmatrix} \Phi_1^k \\ \Phi_2^k \end{pmatrix} e_k = B_k^\lambda \begin{pmatrix} \Phi_1^k \\ \Phi_2^k \end{pmatrix} e_k,$$

where B_k^λ is the matrix given by (6.4.39). Then by (6.4.63) and (6.4.64) we have

$$\begin{pmatrix} \Phi_1^{2k} \\ \Phi_2^{2k} \end{pmatrix} = -\frac{(3\alpha \xi_1^2 - 2\xi_1 \xi_2 \rho_K)x^2}{4} B_{2k}^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad B_{2k} = B_{2k}^{\lambda_c} \quad \text{for } k = K, K_1, K_2.$$

Direct computation shows that

$$\begin{pmatrix} \Phi_1^0 \\ \Phi_2^0 \end{pmatrix} = \frac{3\xi_1^2 x^2}{8} \begin{pmatrix} 1 \\ \lambda_c \end{pmatrix}, \quad (6.4.65)$$

$$\begin{pmatrix} \Phi_1^{2K_1} \\ \Phi_2^{2K_1} \end{pmatrix} = \frac{\xi_1(3\alpha \xi_1 - 2\xi_2 \rho_{K_1})}{4 \det B_{2K_1}} \begin{pmatrix} 1 + \rho_{2K_1} \\ \lambda_c \end{pmatrix}, \quad (6.4.66)$$

$$\begin{pmatrix} \Phi_1^{2K_2} \\ \Phi_2^{2K_2} \end{pmatrix} = \frac{\xi_1(3\alpha \xi_1 - 2\xi_2 \rho_{K_2})}{4 \det B_{2K_2}} \begin{pmatrix} 1 + \rho_{2K_2} \\ \lambda_c \end{pmatrix}, \quad (6.4.67)$$

$$\begin{pmatrix} \Phi_1^{2K} \\ \Phi_2^{2K} \end{pmatrix} = \frac{\xi_1(3\alpha\xi_1 - 2\xi_2\rho_K)}{4\det B_{2K}} \begin{pmatrix} 1 + \rho_{2K} \\ \lambda_c \end{pmatrix}. \quad (6.4.68)$$

Inserting (6.4.65) into (6.4.61), by (6.4.59) and (6.4.60) we get

$$\begin{aligned} (G(x\varphi_K + \Phi), \varphi_K^*) &= \frac{(2 - \text{sign}K_1 K_2)(\rho_K + 1)|\Omega|}{8} \left[-\frac{8\alpha(\rho_K + 1)^3 x^3 \int_{\Omega} e_K^4 dx}{(2 - \text{sign}K_1 K_2)|\Omega|} \right. \\ &\quad + 2(\xi_2\rho_K - 6\alpha\xi_1)\Phi_1^0 x + \frac{2(\xi_2(\rho_{K_2} - \rho_{K_1}) - 6\alpha\xi_1)\Phi_1^{2K_1} x}{1 + \text{sign}K_1} \\ &\quad + \frac{2(\xi_2(\rho_{K_1} - \rho_{K_2}) - 6\alpha\xi_1)\Phi_2^{2K_2} x}{1 + \text{sign}K_2} \\ &\quad - \frac{2 - \text{sign}K_1 K_2}{2}(\xi_2\rho_K + 6\alpha\xi_1)\Phi_1^{2K} x \\ &\quad + \frac{(\rho_K + 1)\rho_{2K_1}}{1 + \text{sign}K_1}\Phi_2^{2K_1} x + \frac{(\rho_K + 1)\rho_{K_2}}{1 + \text{sign}K_2}\Phi_2^{2K_2} x \\ &\quad \left. + \frac{(\rho_K + 1)\rho_{2K}(2 - \text{sign}K_1 K_2)}{4}\Phi_2^{2K} x \right] + o(x^3). \end{aligned}$$

By definition, we have

$$\begin{aligned} \rho_{K_1} + \rho_{K_2} &= \rho_K, \quad \rho_{2K} = 4\rho_K \quad \forall K = (K_1, K_2), \\ (\varphi, \varphi^*) &= [(\rho_K + 1)^2 + \rho_K \lambda_c] \int_{\Omega} e_K^2 dx \\ &= \frac{2 - \text{sign}(K_1 K_2)}{4}(\rho_K + 1)(\mu\rho_K + \rho_K + 2\alpha + 1)|\Omega|. \end{aligned}$$

In view of (6.4.65)–(6.4.68), the reduced equation (6.4.57) is given by

$$\frac{dx}{dt} = \beta_K^+(\lambda)x + \frac{(\rho_K + 1)^3 b x^3}{8(\mu\rho_K + \rho_K + 2\alpha + 1)} + o(x^3), \quad (6.4.69)$$

where b is the parameter as in (6.4.36). Then the theorem follows readily from (6.4.69). \square

6.4.3 Transition of Three-Component Systems

The Model Hereafter we always assume that $\delta_0 \geq 0$ is a constant. Then (6.4.22) possesses a positive constant steady-state u^* given by

$$(u_1^*, u_2^*, u_3^*) \quad \text{with } u_1^* = \left(\frac{u_3^*}{1 + u_3^*} \right)^{1/2}, \quad u_2^* = \lambda u_1^*, \quad u_3^* u_1^* = \frac{\delta_0}{\delta}. \quad (6.4.70)$$

It is easy to see that u_3^* is the unique positive real root of the cubic equation

$$x^3 - \left(\frac{\delta_0}{\delta}\right)^2 x - \left(\frac{\delta_0}{\delta}\right)^2 = 0.$$

Consider the translation

$$(u_1, u_2, u_3) \rightarrow (u_1^* + u_1, u_2^* + u_2, u_3^* + u_3). \quad (6.4.71)$$

Then equations (6.4.22) are equivalent to

$$\begin{aligned} \frac{\partial u_1}{\partial t} &= \mu \Delta u_1 - 2\alpha u_1^{*2} u_1 - u_1^* \Delta u_2 + \frac{\alpha u_1^*}{(1+u_3^*)^2} u_3 + g(u), \\ \frac{\partial u_3}{\partial t} &= r \Delta u_3 - \delta u_1^* u_3 - \delta u_3^* u_1 - \delta u_1 u_3, \\ -\Delta u_2 + u_2 &= \lambda u_1, \\ \frac{\partial (u_1, u_2, u_3)}{\partial n} \Big|_{\partial \Omega} &= 0, \\ u(0) &= u_0, \end{aligned} \quad (6.4.72)$$

where $u = (u_1, u_3)$, $u_2 = \lambda[-\Delta + 1]^{-1}u_1$, and

$$\begin{aligned} g(u) &= -\nabla(u_1 \nabla u_2) - 3\alpha u_1^* u_1^2 - \alpha u_1^3 + \frac{\alpha(u_1 + u_1^*)(u_3 + u_3^*)}{1 + u_3^* + u_3} \\ &\quad - \frac{\alpha u_1^* u_3^*}{1 + u_3^*} - \frac{\alpha u_1^* u_3}{(1 + u_3^*)^2} - \frac{\alpha u_3^* u_1}{1 + u_3^*}. \end{aligned} \quad (6.4.73)$$

The Taylor expansion of g at $u = 0$ is expressed as

$$\begin{aligned} g(u) &= -\nabla(u_1 \nabla u_2) - 3\alpha u_1^* u_1^2 + \frac{\alpha u_1 u_3}{(1 + u_3^*)^2} - \frac{\alpha u_1^* u_3^2}{(1 + u_3^*)^3} \\ &\quad - \alpha u_1^3 - \frac{\alpha u_1 u_3^2}{(1 + u_3^*)^3} + \frac{\alpha u_1^* u_3^3}{(1 + u_3^*)^4} + o(3). \end{aligned}$$

Let

$$H = L^2(\Omega, \mathbb{R}^2), \quad H_1 = \left\{ u \in H^2(\Omega, \mathbb{R}^2) \mid \frac{\partial u}{\partial n} = 0 \text{ on } \partial \Omega \right\}.$$

Define the operators $L_\lambda : H_1 \rightarrow H$ and $G_\lambda : H_1 \rightarrow H$ by

$$\begin{aligned} L_\lambda u &= \begin{pmatrix} \mu \Delta - 2\alpha u_1^{*2} - \lambda u_1^* \Delta [-\Delta + I]^{-1} & \frac{\alpha u_1^*}{(1+u_3^*)^2} \\ -\delta u_3^* & r \Delta - \delta u_1^* \end{pmatrix} \begin{pmatrix} u_1 \\ u_3 \end{pmatrix}, \\ G(u, \lambda) &= \begin{pmatrix} g(u) \\ -\delta u_1 u_3 \end{pmatrix}, \end{aligned} \quad (6.4.74)$$

Then the problem (6.4.72) takes the following abstract form:

$$\frac{du}{dt} = L_\lambda u + G(u, \lambda), \quad u(0) = u_0. \quad (6.4.75)$$

It is known that the inverse mapping $[-\Delta + I]^{-1} : H \rightarrow H_1$ is a bounded linear operator. Therefore we have $L_\lambda : H_1 \rightarrow H$ is a sectorial operator, and $G_\lambda : H_\theta \rightarrow H$ is a C^∞ bounded operator for $\theta \geq 1/2$.

Theorems 6.4.3 and 6.4.6 show that a two-component system undergoes a dynamic transition only at real eigenvalues. However, the three-component system (6.4.21) can undergo dynamic transitions at both real and complex eigenvalues, leading to both steady-state and spatiotemporal oscillations.

Linearized Problem The eigenvalue equations of (6.4.22) at the steady-state (u_1^*, u_2^*, u_3^*) given by (6.4.70) in their abstract form are given by

$$L_\lambda \varphi = \beta \varphi, \quad (6.4.76)$$

where $L_\lambda : H_1 \rightarrow H$ as defined in (6.4.74). The explicit form of (6.4.76) is given by

$$\begin{pmatrix} \mu\Delta - 2\alpha u_1^{*2} - \lambda u_1^* \Delta [-\Delta + I]^{-1} & \frac{\alpha u_1^*}{(1+u_3^*)^2} \\ -\delta u_3^* & r\Delta - \delta u_1^* \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_3 \end{pmatrix} = \beta \begin{pmatrix} \psi_1 \\ \psi_3 \end{pmatrix}.$$

As before, let ρ_k and e_k be the eigenvalue and eigenvector of $-\Delta$ with Neumann boundary condition given by (6.4.28), and let

$$\psi_k = (\psi_1^k, \psi_3^k) = (\xi_{k1} e_k, \xi_{k3} e_k).$$

Then, it is easy to see that ψ_k is an eigenvector of (6.4.76) provided that $(\xi_{k1}, \xi_{k3}) \in \mathbb{R}^2$ is an eigenvector of the matrix A_k^λ :

$$A_k^\lambda \begin{pmatrix} \xi_{k1} \\ \xi_{k3} \end{pmatrix} = \beta_k \begin{pmatrix} \xi_{k1} \\ \xi_{k3} \end{pmatrix}, \quad A_k^\lambda = \begin{pmatrix} \frac{\lambda \rho_k u_1^*}{1+\rho_k} - \mu \rho_k - 2\alpha u_1^{*2} & \frac{\alpha u_1^*}{(1+u_3^*)^2} \\ -\delta u_3^* & -r \rho_k - \delta u_1^* \end{pmatrix}. \quad (6.4.77)$$

The eigenvalues β_k of A_k^λ , which are also eigenvalues of (6.4.76), are expressed by

$$\begin{aligned} \beta_k^\pm(\lambda) &= \frac{1}{2} \left[a \pm \sqrt{a^2 - 4 \det A_k^\lambda} \right], \\ a &= \text{tr} A_k^\lambda = \frac{\lambda \rho_k u_1^*}{1+\rho_k} - \mu \rho_k - 2\alpha u_1^{*2} - r \rho_k - \delta u_1^*. \end{aligned} \quad (6.4.78)$$

To derive the PES, we introduce two parameters as follows:

$$\Lambda_c = \min_{\rho_K} \frac{(\rho_K + 1)}{\rho_K u_1^*} [\mu \rho_K + 2\alpha u_1^{*2} + r \rho_K + \delta u_1^*], \quad (6.4.79)$$

$$\lambda_c = \min_{\rho_K} \frac{(\rho_K + 1)}{\rho_K u_1^*} \left[\mu \rho_K + 2\alpha u_1^{*2} + \frac{\alpha \delta}{(1 + u_3^*)^2(r\rho_K + \delta u_1^*)} \right]. \quad (6.4.80)$$

Let $K = (K_1, K_2)$ and $K^* = (K_1^*, K_2^*)$ be the integer pairs such that ρ_K and ρ_{K^*} satisfy (6.4.79) and (6.4.80) respectively.

Theorem 6.4.7 Let Λ_c and λ_c be the parameters defined by (6.4.79) and (6.4.80) respectively. Then we have the following assertions:

1 If $\Lambda_c < \lambda_c$, then the eigenvalues $\beta_K^\pm(\lambda)$ of (6.4.78) are a pair of conjugate complex numbers near $\lambda = \Lambda_c$, and all eigenvalues of (6.4.78) satisfy

$$\operatorname{Re} \beta_K^\pm(\lambda) \begin{cases} < 0 & \text{if } \lambda < \Lambda_c, \\ = 0 & \text{if } \lambda = \Lambda_c, \\ > 0 & \text{if } \lambda > \Lambda_c, \end{cases} \quad (6.4.81)$$

$$\operatorname{Re} \beta_k^\pm(\Lambda_c) < 0 \quad \forall k \in \mathbb{Z}^2 \text{ with } \rho_k \neq \rho_K \quad (6.4.82)$$

2 If $\lambda_c < \Lambda_c$, then the eigenvalue $\beta_{K^*}^+(\lambda)$ is real near $\lambda = \lambda_c$, and all of (6.4.78) satisfy

$$\beta_{K^*}^+(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_c, \\ = 0 & \text{if } \lambda = \lambda_c, \\ > 0 & \text{if } \lambda > \lambda_c, \end{cases} \quad (6.4.83)$$

$$\begin{cases} \operatorname{Re} \beta_k^+(\lambda_c) < 0 & \forall k \in \mathbb{Z}^2 \text{ with } \rho_k \neq \rho_{K^*}, \\ \operatorname{Re} \beta_k^-(\lambda_c) < 0 & \forall |k| \geq 0. \end{cases} \quad (6.4.84)$$

Proof. By (6.4.78) we see that $\beta_k^\pm(\lambda)$ are a pair of complex eigenvalues of (6.4.76) near some $\lambda = \lambda^*$, and satisfy

$$\operatorname{Re} \beta_k^\pm(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda^*, \\ = 0 & \text{if } \lambda = \lambda^*, \\ > 0 & \text{if } \lambda > \lambda^*, \end{cases}$$

if and only if

$$\operatorname{tr} A_k^{\lambda^*} = 0, \quad \det A_k^{\lambda^*} > 0.$$

Likewise, $\beta_k^+(\lambda)$ is real near $\lambda = \lambda^*$ and satisfies

$$\beta_k^+(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda^*, \\ = 0 & \text{if } \lambda = \lambda^*, \\ > 0 & \text{if } \lambda > \lambda^*, \end{cases}$$

if and only if

$$\operatorname{tr} A_k^{\lambda^*} < 0, \quad \det A_k^{\lambda^*} = 0$$

By the definition of λ_c and Λ_c , when $\Lambda_c < \lambda_c$ we have

$$\begin{aligned} \operatorname{tr} A_K^{\Lambda_c} &= 0, \\ \operatorname{tr} A_k^{\Lambda_c} &< 0 \quad \forall k \in \mathbb{Z}^2 \text{ with } \rho_k \neq \rho_K, \\ \det A_k^{\Lambda_c} &> 0 \quad \forall |k| \geq 0, \end{aligned} \tag{6.4.85}$$

and when $\lambda_c < \Lambda_c$,

$$\begin{aligned} \det A_{K^*}^{\lambda_c} &= 0, \\ \det A_k^{\lambda_c} &> 0 \quad \forall k \in \mathbb{Z}^2 \text{ with } \rho_k \neq \rho_K, \\ \operatorname{tr} A_k^{\lambda_c} &< 0 \quad \forall |k| \geq 0. \end{aligned} \tag{6.4.86}$$

It is known that the real parts of β_k^\pm are negative at λ if and only if

$$\det A_k^\lambda > 0, \quad \operatorname{tr} A_k^\lambda < 0.$$

Then Assertions (1) and (2) follow from (6.4.85) and (6.4.86) respectively. \square

Dynamic Transition Theorem for (6.4.22) Based on Theorem 6.4.7, we immediately get the following transition theorem for (6.4.72).

Theorem 6.4.8 *Let Λ_c and λ_c be given by (6.4.79) and (6.4.80) respectively. Then the following assertions hold true for (6.4.72):*

- (1) *When $\Lambda_c < \lambda_c$, the system undergoes a dynamic transition to periodic solutions at $(u, \lambda) = (0, \Lambda_c)$. In particular, if the eigenvalues β_K^\pm satisfying (6.4.81) are complex simple, then there is a parameter b_0 such that the dynamic transition is continuous (Type-I) as $b_0 < 0$, and is jump (Type-II) as $b_0 > 0$ with a singularity separation of periodic solutions at some $\lambda^* < \Lambda_c$.*
- (2) *When $\lambda_c < \Lambda_c$, the system undergoes a dynamic transition to steady states at $(u, \lambda) = (0, \lambda_c)$. If $\beta_{K^*}^+(\lambda)$ satisfying (6.4.83) is simple, then there exists a parameter b_1 such that the transition is continuous as $b_1 < 0$, and jump as $b_1 > 0$ with two saddle-node bifurcation points at some $\tilde{\lambda} < \lambda_c$ from $(u^+, \tilde{\lambda})$ and $(u^-, \tilde{\lambda})$.*

Remark 6.4.9 By applying the standard procedure used in the preceding sections, we can derive explicit formulas for the two parameters b_0 and b_1 in Theorem 6.4.8. However, due to their complexity, we omit the details.

Transition for the System (6.4.21) We are now in a position to discuss the transition of (6.4.21). With the translation (6.4.71), the system (6.4.21) is rewritten in the following form

$$\begin{aligned}
\frac{\partial u_1}{\partial t} &= \mu\Delta u_1 - 2\alpha u_1^{*2} u_1 - u_1^* \Delta u_2 + \frac{\alpha u_1^* u + 3}{(1+u_3^*)^2} + g(u), \\
\frac{\partial u_2}{\partial t} &= \Delta u_2 - u_2 + \lambda u_1, \\
\frac{\partial u_3}{\partial t} &= r\Delta u_3 - \delta u_1^* u_3 - \delta u_3^* u_1 - \delta u_1 u_3, \\
\frac{\partial u}{\partial n} \Big|_{\partial\Omega} &= 0, \\
u(0) &= u_0,
\end{aligned} \tag{6.4.87}$$

where $g(u)$ is as in (6.4.73). Here the notation u stands for three-component unknown: $u = (u_1, u_2, u_3)$.

Let

$$L_\lambda u = \begin{pmatrix} \mu\Delta - 2\alpha u_1^{*2} & -u_1^* \Delta & \frac{\alpha u_1^*}{(1+u_3^*)^2} \\ \lambda & \Delta - 1 & 0 \\ -\delta u_3^* & 0 & r\Delta - \delta u_1^* \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}.$$

Then, all eigenvalues $\beta_k^j(\lambda)$ and eigenvectors ψ_k^j of L_λ satisfy

$$D_k^\lambda \begin{pmatrix} \xi_{k1}^j \\ \xi_{k2}^j \\ \xi_{k3}^j \end{pmatrix} = \beta_k^j(\lambda) \begin{pmatrix} \xi_{k1}^j \\ \xi_{k2}^j \\ \xi_{k3}^j \end{pmatrix}, \quad 1 \leq j \leq 3, \quad k \in \mathbb{Z}^2,$$

with

$$\psi_k^j = (\xi_{k1}^j e_k, \xi_{k2}^j e_k, \xi_{k3}^j e_k),$$

and e_k as in (6.4.28), D_k^λ is a 3×3 matrix given by

$$D_k^\lambda = \begin{pmatrix} -(\mu\rho_k + 2\alpha u_1^{*2}) & u_1^* \rho_k & \frac{\alpha u_1^*}{(1+u_3^*)^2} \\ \lambda & -(\rho_k + 1) & 0 \\ -\delta u_3^* & 0 & -(r\rho_k + \delta u_1^*) \end{pmatrix}.$$

We introduce the following three parameters:

$$\begin{aligned}
A_k^\lambda &= -\text{tr} D_k^\lambda = \mu\rho_k + 2\alpha u_1^{*2} + \rho_k + 1 + r\rho_k + \delta u_1^*, \\
B_k^\lambda &= \det \begin{pmatrix} -(\mu\rho_k + 2\alpha u_1^{*2}) & u_1^* \rho_k & \frac{\alpha u_1^*}{(1+u_3^*)^2} \\ \lambda & -(\rho_k + 1) & 0 \\ -\delta u_3^* & 0 & -(r\rho_k + \delta u_1^*) \end{pmatrix} + \det \begin{pmatrix} -(\mu\rho_k + 2\alpha u_1^{*2}) & \frac{\alpha u_1^*}{(1+u_3^*)^2} \\ -\delta u_3^* & -(r\rho_k + \delta u_1^*) \end{pmatrix} \\
&\quad + (\rho_k + 1)(r\rho_k + \delta u_1^*), \\
C_k^\lambda &= -\det D_k^\lambda = (\mu\rho_k + 2\alpha u_1^{*2})(\rho_k + 1)(r\rho_k + \delta u_1^*) \\
&\quad - u_1^* \rho_k \lambda (r\rho_k + \delta u_1^*) + \frac{\alpha u_1^*}{(1+u_3^*)^2} \delta u_3^* (\rho_k + 1).
\end{aligned}$$

By the Routh–Hurwitz theorem, we know that all eigenvalues β_k^j of D_k^j have negative real parts if and only if

$$A_k^\lambda > 0, \quad A_k^\lambda B_k^\lambda - C_k^\lambda > 0, \quad C_k^\lambda > 0. \quad (6.4.88)$$

Let Λ^c and $K = (K_1, K_2)$ satisfy

$$\begin{aligned} A_K^{\Lambda_c} &> 0, \quad A_K^{\Lambda_c} B_K^{\Lambda_c} - C_K^{\Lambda_c} = 0, \quad C_K^{\Lambda_c} > 0, \\ A_k^{\Lambda_c} &> 0, \quad A_k^{\Lambda_c} B_k^{\Lambda_c} - C_k^{\Lambda_c} > 0, \quad C_k^{\Lambda_c} > 0 \quad \forall k \text{ with } \rho_k \neq \rho_K. \end{aligned} \quad (6.4.89)$$

Then Λ_c satisfies that

$$\begin{aligned} \Lambda_c = \inf_{\rho_k} \frac{1}{\rho_k u_1^*} [(\mu + r)\rho_k + 2\alpha u_1^{*2} + \delta u_1^*] \\ \times \left[(r + 1)\rho_k + \delta u_1^* + 1 + \frac{\alpha\delta}{(\mu\rho_k + 2\alpha u_1^{*2} + \rho_k + 1)(1 + u_3^*)^2} \right], \end{aligned} \quad (6.4.90)$$

and ρ_K satisfies (6.4.90). In particular, under the condition (6.4.89), there is a pair of complex eigenvalues $\beta_K^1(\lambda)$ and $\beta_K^2(\lambda)$ of D_K^λ , such that

$$\operatorname{Re} \beta_K^{1,2}(\lambda) \begin{cases} < 0 & \text{if } \lambda < \Lambda_c, \\ = 0 & \text{if } \lambda = \Lambda_c, \\ > 0 & \text{if } \lambda > \Lambda_c, \end{cases} \quad (6.4.91)$$

and the other eigenvalues $\beta_k^j(\lambda)$ of L_λ satisfy

$$\begin{aligned} \operatorname{Re} \beta_k^j(\Lambda_c) &< 0 \quad \forall k \text{ with } \rho_k \neq \rho_K \text{ and } 1 \leq j \leq 3, \\ \beta_K^3(\Lambda_c) &< 0. \end{aligned} \quad (6.4.92)$$

Let λ_c and $K^* = (K_1^*, K_2^*)$ satisfy

$$\begin{aligned} A_{K^*}^{\lambda_c} &> 0, \quad A_{K^*}^{\lambda_c} B_{K^*}^{\lambda_c} - C_{K^*}^{\lambda_c} > 0, \quad C_{K^*}^{\lambda_c} = 0, \\ A_k^{\lambda_c} &> 0, \quad A_k^{\lambda_c} B_k^{\lambda_c} - C_k^{\lambda_c} > 0, \quad C_k^{\lambda_c} > 0 \quad \forall k \text{ with } \rho_k \neq \rho_{K^*}. \end{aligned} \quad (6.4.93)$$

Then λ_c is given by

$$\lambda_c = \inf_{\rho_k} \frac{(\rho_k + 1)}{\rho_k u_1^*} \left[\mu\rho_k + 2\alpha u_1^{*2} + \frac{\alpha\delta}{(1 + u_3^*)^2(r\rho_k + \delta u_1^*)} \right], \quad (6.4.94)$$

and λ_c arrives its minimal at ρ_{K^*} . By the Routh–Hurwitz criterion (6.4.88), we deduce that with (6.4.93) there is a real eigenvalue $\beta_{K^*}^1(\lambda)$ of $D_{K^*}^{\lambda_c}$ satisfies that

$$\beta_{K^*}(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_c, \\ = 0 & \text{if } \lambda = \lambda_c, \\ > 0 & \text{if } \lambda > \lambda_c, \end{cases} \quad (6.4.95)$$

$$\begin{cases} \operatorname{Re}\beta_{K^*}^j(\lambda_c) < 0 & j = 2, 3, \\ \operatorname{Re}\beta_k^j(\lambda_c) < 0 & \forall k \in \mathbb{Z}^2 \text{ with } \rho_k \neq \rho_{K^*} \text{ and } 1 \leq j \leq 3. \end{cases} \quad (6.4.96)$$

It is clear that (6.4.91) and (6.4.92) hold true as $\Lambda_c < \lambda_c$, and (6.4.94)–(6.4.95) hold true as $\lambda_c < \Lambda_c$. Hence, we have the following transition theorem for (6.4.87).

Theorem 6.4.10 *Let Λ_c and λ_c be given by (6.4.90) and (6.4.94) respectively. Then, Assertions (1) and (2) of Theorem 6.4.8 hold true for the system (6.4.87).*

6.4.4 Biological Conclusions

Biological Significance of Transition Theorems Pattern formation is one of the main characteristics for bacteria chemotaxis, and is fully characterized by dynamic transitions. Theorems 6.4.3–6.4.10 tell us that the nondimensional parameter λ , given by

$$\lambda = \frac{\sqrt{\alpha_2} r_1 \chi}{r_2 k_2}, \quad (6.4.97)$$

plays a crucial role to determine the dynamic transition and pattern formation. Actually, the key factor in (6.4.97) is the product of the chemotactic coefficient χ and the production rate $r_1 : \chi r_1$, which depends on the type of bacteria. When λ is less than some critical value λ_c , the uniform distribution of biological individuals is a stable state. When λ exceeds λ_c , the bacteria cells aggregate to form more complex and stable patterns.

As seen in (6.4.33), (6.4.79), (6.4.80), and (6.4.90), under different biological conditions, the critical parameter λ_c takes different forms and values. But, a general formula for λ_c is given by

$$\lambda_c = a_0 + \inf_{\rho_k} \left(a_1 \rho_k + \frac{a_2}{\rho_k} + \frac{a_3}{b_1 \rho_k + b_0} + \frac{a_4}{\rho_k (b_1 \rho_k + b_0)} \right), \quad (6.4.98)$$

where ρ_k are taken as the eigenvalues of $-\Delta$ with the Neumann boundary condition. When Ω is a rectangular region, ρ_k are given by (6.4.28), and the coefficients a_j ($1 \leq j \leq 4$), $b_0, b_1 \geq 0$ depend on the parameters in (6.4.20), with

$$a_0, a_1, a_2, b_0, b_1 > 0, \quad a_3, a_4 \geq 0.$$

In particular, for the system with rich nutrient supplies, (6.4.98) becomes

$$\lambda_c = a_0 + \inf_{\rho_k} \left[a_1 \rho_k + \frac{a_2}{\rho_k} \right].$$

The eigenvalues ρ_k , depending on the geometry of Ω , satisfy

$$0 = \rho_0 < \rho_1 \leq \cdots \leq \rho_k \leq \cdots, \quad \lim_{k \rightarrow \infty} \rho_k = \infty, \quad \rho_1 \propto \frac{1}{L^2}, \quad (6.4.99)$$

where L is the length scale of Ω . We infer from (6.4.98) and (6.4.99) that

$$\lambda_c \rightarrow \infty \quad \text{as} \quad |\Omega| \rightarrow 0 \quad (L \rightarrow 0).$$

It implies that when the container Ω is small, the homogenous state is stable and there is no pattern formation of bacteria under any biological conditions.

Spatiotemporal Oscillation Theorems 6.4.8 and 6.4.10 show that there are two critical parameters λ_c and Λ_c , such that if $\lambda_c < \Lambda_c$, the patterns formed by biological organisms are steady, as exhibited by many experimental results, and if $\Lambda_c < \lambda_c$ a spatial-temporal oscillatory behavior takes place.

For the case with rich nutrient, $u_1^* = 1$, $u_3^* = \infty$. Also, λ_c in (6.4.80) is reduced to (6.4.30), and obviously we have

$$\lambda_c < \Lambda_c \quad \text{for both (6.4.79) and (6.4.90)},$$

and the dynamic transition and pattern formation are determined by Theorems 6.4.3 and 6.4.6. Hence there is no spatiotemporal oscillations for the rich nutrient case, and the time periodic oscillation of chemotaxis occurs only for the case where the nutrient is moderately supplied.

On the other hand, if $\mu, r \cong 0$, and

$$\delta^2 u_1^{*2} (1 + u_3^*)^2 < \alpha \delta_0,$$

then for Λ_c defined by (6.4.79) and (6.4.90), we have $\Lambda_c < \lambda_c$. In this case, a spatial-temporal oscillation pattern is expected for $\lambda > \Lambda_c$.

Transition Types The main theorems provide precise information on the transition types. In all cases, types are precisely determined by the sign of some nondimensional parameters; see b , b_0 , and b_1 respectively in the main theorems. Hence a global phase diagram can be obtained easily by setting the related parameter to be zero.

For example, when $\Omega = (0, L_1)$ is one-dimensional or when $K = (K_1, 0)$ (resp. $K = (0, K_2)$), the parameter b in (6.4.36) can be simplified into the following form

$$b = 2 \left[-3\mu\rho_K + 9\alpha - \frac{(2\mu\rho_K + \alpha)(2\mu\lambda_K^2 + 28\alpha\rho_K + 4\alpha - \mu\rho_K)}{(\mu\rho_{2K} + 2\alpha)(\rho_{2K} + 1) - \rho_{2K}\lambda_c} \right]. \quad (6.4.100)$$

For a nongrowth system, $\alpha = 0$, $K = (1, 0)$, $\lambda_c = \mu(\rho_K + 1)$. Then, (6.4.100) becomes

$$b = \frac{\mu}{3}(1 - 20\lambda_1), \quad \lambda_1 = \frac{\pi^2}{L_1^2}, \quad (6.4.101)$$

and $\lambda = \frac{ar_1x}{r_2k_2}$, with $a = \frac{1}{|\Omega|} \int_{\Omega} u_1 dx$. It follows from (6.4.101) that

$$b \begin{cases} < 0 & \text{if } L_1 < 2\sqrt{5}\pi, \\ > 0 & \text{if } L_1 > 2\sqrt{5}\pi. \end{cases} \quad (6.4.102)$$

By Theorems 6.4.3 and 6.4.6, the phase transition of (6.4.25) and (6.4.23) from $(u, \lambda) = (u^*, \lambda_c)$ is continuous if the length scale L_1 of Ω is less than $2\sqrt{5}\pi$, and jump if L_1 is bigger than $2\sqrt{5}\pi$.

In addition, when we take $\chi(u) = \chi_1 u_1 / (\beta + u_2)^2$ as the chemotaxis function, by Remark 6.4.9, the parameter b of (6.4.101) is replaced by

$$b_1 = \frac{\mu}{3} \left(1 - \frac{20\pi^2}{L_1^2} \right) + \frac{4\kappa\mu^2\pi^2}{L_1^2}, \quad \kappa = \frac{k_2}{\beta\chi + k_2\lambda_c}, \quad \lambda_c = \mu \left(\frac{\pi^2}{L_1^2} + 1 \right).$$

The above conclusion shows that for a nongrowth system, the parameter

$$\lambda = \frac{r_1\chi}{r_2k_2} a, \quad \text{with } a = \frac{1}{|\Omega|} \int_{\Omega} u_1 dx,$$

is proportional to the average density of the initial condition of u_1 —note that u_1 is conservative. Hence, the biological individual is in a homogenous state provided

$$\frac{1}{|\Omega|} \int_{\Omega} \varphi dx < \frac{r_2k_2}{r_1\chi} \mu \left(\frac{\pi^2}{L_1^2} + 1 \right), \quad \varphi = u_1(0),$$

and the bacteria will aggregate to form high density regions provided that

$$\frac{1}{|\Omega|} \int_{\Omega} \varphi dx > \frac{r_2k_2}{r_1\chi} \mu \left(\frac{\pi^2}{L_1^2} + 1 \right). \quad (6.4.103)$$

Moreover, under the condition (6.4.103), if the scale L_1 of Ω is smaller than some critical value L_c (in (6.4.102) $L_c = 2\sqrt{5}\pi$), i.e., $L_1 < L_c$, the continuous transition implies that there is only one high density region of bacteria forming, and if $L_1 > L_c$, then the jump transition expects a large number of high density regions to appear.

Pattern Formation As mentioned before, the pattern formation behavior is dictated by the dynamic transition of the system. We analyzed the formation of two type patterns—the lamella and the rectangles—and the method and ideas can be generalized to study the formation of more complex patterns.

For a growth system, the critical parameter λ_c takes its value at some eigenvalue ρ_K of $-\Delta$ for $K = (K_1, K_2)$, as shown by (6.4.33) and (6.4.80). From the pattern formation point of view, for the Type-I transition, the patterns described by the transition solutions in the main theorems are either lamella or rectangular:

lamella pattern	for $K_1 K_2 = 0$,
rectangular pattern	for $K_1 K_2 \neq 0$.

In the case where $b > 0$, the system undergoes a more drastic change. As $\lambda^* < \lambda < \lambda_c$, the homogeneous state, the new patterns v_2^λ and v_4^λ are metastable. For $\lambda > \lambda_c$, the system undergoes transitions to more complex patterns away from the basic homogeneous state form.



Chapter 7

Fundamental Principles of Statistical and Quantum Physics

This chapter is aimed for the first-principle approach to statistical physics and quantum physics, in the spirit of the guiding principles of physics: Principles 1.1.1 and 1.1.2.

Classically, statistical physics relates the microscopic properties of individual particles to the macroscopic properties of the system, using methods of probability theory and statistics. It is often called statistical mechanics, and is closely related to thermodynamics.

We view statistical physics as a branch of physics, dealing with any physical system with the following three fundamental characteristics:

1. the system is made up of large collection of micro-particles, which form the physical body of the system;
2. the micro-particles are always in motion; and
3. on the macroscopic level, there is no viewable motion or state of the particles.

In a nutshell, *statistical physics studies the collective macroscopic physical properties of the motion and states of the microscopic particles in the system.*

Based on the above characteristics, we view statistical physics in a broad sense so that it consists of thermodynamics, statistical mechanics, and composite systems. Composite systems include nonequilibrium dynamics and phase transitions of statistical physical systems.

One aim of this chapter is to introduce a new principle, the potential-descending principle (PDP). We show that PDP, together with the classical principle of equal *a priori* probabilities (PEP), leads to the first and second laws of thermodynamics, irreversibility, and the three classical statistics: the Maxwell–Boltzmann, the Fermi–Dirac, and the Bose–Einstein statistics. Consequently, PDP and PEP are the first principles of statistical physics.

Also, PDP leads to the standard model of statistical physics given in the form of (7.1.1) and (7.1.2). Therefore, the basic issue boils down to deriving precise form of thermodynamic potentials F , which are fully dictated by underlying symmetries of the physical system as presented in Sect. 7.6.

Second aim of the chapter is to present a new statistical theory of heat presented in Sect. 7.5. One motivation of the theory is the lack of physical carriers of heat in classical thermodynamics. Another motivation is the recent discovered photon cloud structure of electrons (Ma and Wang, 2015a). This leads to a natural conjugate relation between electrons and photons, reminiscent of the conjugate relation between temperature and entropy. The new theory of heat provides a natural connection between these two conjugate relations: at the equilibrium of absorption and radiation, the average energy level of the system maintains unchanged, and represents the temperature of the system; at the same time, the number (density) of photons in the sea of photons represents the entropy (entropy density) of the system. The theory contains four parts: (1) the photon number formula of entropy, (2) the energy level formula of temperature, (3) the temperature theorem, and (4) the thermal energy formula. In particular, the photon number entropy formula is equivalent to the Boltzmann entropy formula, and, however, possesses new physical meaning that the physical carrier of heat is the photons.

The third aim is on the foundations of quantum physics in Sect. 7.3. The main focuses are on basic quantum rules, new field theoretic interpretation of quantum wave functions, elementary particles, photon cloud structure, and energy levels of particles.

The study presented in this chapter is based on Ma and Wang (2017a,d, 2019, 2015a), Liu et al. (2019), Ma et al. (2018).

7.1 Potential-Descending Principle

7.1.1 Potential-Descending Principle

We start with the classical principle of equal probability, which states as follows; see among many others (Pathria and Beale, 2011):

Principle 7.1.1 (Principle of Equal Probability) *An isolated equilibrium thermodynamic system has an equal probability of being in any microstate that is consistent with its current macrostate.*

We now introduce the potential-descending principle first postulated by Ma and Wang (2017a).

Principle 7.1.2 (Potential-Descending Principle (Ma and Wang, 2017a)) *For each thermodynamic system, there are order parameters $u = (u_1, \dots, u_N)$, control parameters λ , and the thermodynamic potential functional $F(u; \lambda)$. For a nonequilibrium state $u(t; u_0)$ of the system with initial state $u(0, u_0) = u_0$, we have the following properties:*

1. the potential $F(u(t; u_0); \lambda)$ is decreasing:

$$\frac{d}{dt} F(u(t; u_0); \lambda) < 0 \quad \forall t > 0;$$

2. the order parameters $u(t; u_0)$ have a limit

$$\lim_{t \rightarrow \infty} u(t; u_0) = \bar{u};$$

3. there is an open and dense set O of initial data in the space of state functions, such that for any $u_0 \in O$, the corresponding \bar{u} is a minimum of F , which is called an equilibrium of the thermodynamic system:

$$\delta F(\bar{u}; \lambda) = 0.$$

7.1.2 Standard Model of Thermodynamic Systems

Based on this principle, the dynamic equations of a thermodynamic system in a nonequilibrium state take the form

$$\frac{du}{dt} = -A\delta_{\mathcal{L}}F(u) \quad \text{for isolated systems,} \quad (7.1.1)$$

$$\begin{cases} \frac{du}{dt} = -A\delta_{\mathcal{L}}F(u) + B(u), \\ \int A\delta_{\mathcal{L}}F(u) \cdot B(u) = 0 \end{cases} \quad \text{for coupled systems,} \quad (7.1.2)$$

where $\delta_{\mathcal{L}}$ is the derivative operator with \mathcal{L} -constraint for some operator \mathcal{L} as defined by (1.1.7) and (1.1.8), B represents coupling operators, and A is a symmetric and positive definite matrix of coefficients.

Consider a thermodynamic system involving a conserved order parameter u , representing the mole density of the particles in the system. Let thermodynamic potential be given by $F = F(u, \lambda)$. We have

$$G(u) = \int_{\Omega} u(x, t) dx = \text{constant.} \quad (7.1.3)$$

By the definition of the chemical potential μ , at the equilibrium

$$\mu = \frac{\delta}{\delta u} F(u, \lambda). \quad (7.1.4)$$

On the other hand, by the Lagrange multiplier theorem, the equation for minimum state of the functional F under the constraint (7.1.3) is given by

$$\frac{\delta}{\delta u} F(u, \lambda) = \beta \delta G(u), \quad (7.1.5)$$

where G is the constraint functional in (7.1.3), β is the Lagrange multiplier, the variation operator of G is 1:

$$\delta G(u) = 1.$$

Hence (7.1.5) becomes

$$\frac{\delta}{\delta u} F(u, \lambda) = \beta. \quad (7.1.6)$$

In view of (7.1.4) and (7.1.6), we have

$$\mu = \text{Lagrange multiplier } \beta. \quad (7.1.7)$$

Notice that the Lagrange multiplier β is always a constant, even in the case where $u = u(x)$ is a function of $x \in \mathbb{R}^3$. Hence by (7.1.7), μ is different from the thermodynamic quantities such as the temperature T , generalized displacement f . Consequently, the energy contribution μN is different from the thermal energy $Q = ST$ and the mechanical energy $W = fX$.

The Lagrange multiplier theorem suggests that the thermodynamic potential functional F needs to be modified as follows:

$$H = F(u, \lambda) - \mu G(u) \quad (7.1.8)$$

where $G(u)$ is the constraint functional in (7.1.3), and μ is the chemical potential, and the equilibrium equation for the system becomes

$$\frac{\delta}{\delta u} H(u, \lambda) = 0. \quad (7.1.9)$$

In summary, we have the following physical conclusions for chemical potentials:

Physical Conclusion 7.1.3 (Physical Property of Chemical Potential) *For a thermodynamic system with conserved number of particles, the chemical potential μ enjoys the following*

- (1) μ is the Lagrangian multiplier of F ;
- (2) μ is neither an order parameter nor a control parameter; and
- (3) the functional H given in (7.1.8) is the potential functional.

The key point here is that chemical potential cannot be regarded either as an order parameter or as a control parameter for the thermodynamic system. Consequently, the model based on the potential-descending principle is still given by (7.1.1) and (7.1.2).

The above physical properties of the chemical potential μ lead to an important consequence that the classical Cahn–Hilliard equation:

$$\frac{\partial u}{\partial t} = -\alpha \Delta^2 u + \Delta f'(u), \quad (7.1.10)$$

should be replaced by the modified Cahn–Hilliard equation given by

$$\frac{\partial u}{\partial t} = \alpha \Delta u + f'(u) - \mu. \quad (7.1.11)$$

Mathematically, the phase transition behaviors of (7.1.11) and (7.1.10) are identical.

We have, however, retained the study in Chap. 3 on the Cahn–Hilliard equation, based on two considerations. First, the Cahn–Hilliard model has its historical meaning and can be also used in other applied fields. Second, its phase transition behavior is the same as the modified model derived from (7.1.9). Note that the thermodynamic potential for a binary system is given by

$$F = \int_{\Omega} \left[\frac{\alpha}{2} |\nabla u|^2 + f(u) \right] dx. \quad (7.1.12)$$

7.1.3 Irreversibility

Natural phenomena show that the motion of all statistical physical system (collection of large number of particles) possesses irreversibility. Namely, when the system evolves from one initial state to another, the system will not spontaneously return to its initial state without external interference. This phenomena is called (temporal) irreversibility.

Classical thermodynamics attributes this irreversibility to entropy-ascending principle. Based on the discussions in the previous sections, the physical quantity depicting irreversibility is thermodynamical potential. The Potential-Descending Principle (PDP), Principle 7.1.2, offers a clear description of the irreversibility of thermodynamical systems.

More precisely, consider a thermodynamic system with potential $F(u, \lambda)$, order parameters u , control parameters λ and a nonequilibrium initial state u_0 . Then the PDP amounts to saying that the potential is decreasing:

$$\frac{d}{dt} F(u(t; u_0); \lambda) < 0 \quad \forall t > 0.$$

This shows that the state of the system $u(t; u_0)$ will never return to its initial state u_0 in the future time. This is exactly the (temporal) irreversibility.

From the PDP-based mathematical model:

$$\begin{aligned} \frac{du}{dt} &= -\delta F(u; \lambda), \\ u|_{t=0} &= u_0, \end{aligned} \quad (7.1.13)$$

we also have a clear understanding of irreversibility. For example, given a nonhomogeneous system, its thermodynamic potential is

$$F = \int_{\Omega} \left[\frac{\kappa}{2} |\nabla u|^2 + g(u) \right] dx,$$

and the corresponding equation (7.1.13) is given by

$$\frac{du}{dt} = \kappa \Delta u - f(u), \quad u|_{t=0} = u_0, \quad (7.1.14)$$

where $f(z) = g'(z)$. If we take time reversal transformation $t \rightarrow -t$, then (7.1.14) can be written as

$$\frac{du}{dt} = -\kappa \Delta u - f(u), \quad u|_{t=0} = u_0. \quad (7.1.15)$$

Basic theory of partial differential equations shows that for general initial data u_0 , (7.1.15) has no solution for $t > 0$. This is another mathematical description for irreversibility. This description is also suitable for all dissipative systems, including fluid motion. In other words, irreversibility is a fundamental feature of dissipative systems.

It is worth emphasizing here that for isolated simple gas system, the thermodynamic potential is

$$F = U_0 - ST - \mu_1 N - \mu_2 E, \quad (7.1.16)$$

where U_0 is the internal energy, which is a constant, N is the number of particles, μ_1 and μ_2 are Lagrangian multipliers, and E is the total energy. For this system, the entropy S is an order parameter, and temperature T is a control parameter.

By (7.1.16), we see that potential-descending is equivalent to entropy ascending:

$$\frac{d}{dt} F(S(t)) < 0 \iff \frac{d}{dt} S(t) > 0.$$

In summary,

the entropy S is a state function, which is the solution of the basic thermodynamic equations. Thermodynamic potential F is a higher level physical quantity than entropy, and consequently, is the correct physical quantity, rather than the entropy, for describing irreversibility for all thermodynamic systems. PDP is exactly the law for irreversibility.

7.1.4 Problems in Boltzmann Equation

(1) Boltzmann Equation

In statistical physics, a classical problem is how to establish a mathematical model that can faithfully describe irreversible processes. The Boltzmann equation is introduced mainly for this purpose. We start with a brief introduction on the derivation of the Boltzmann equation.

Consider a system of ideal gases that is in a nonequilibrium state, and assume its dynamical equation takes the form:

$$\frac{d\rho}{dt} = G(\rho), \quad (7.1.17)$$

where ρ stands for the probability density function of the gas molecules.

By the general guiding principle of physics, (7.1.17) should be established based on some fundamental laws and/or principles. Classical theories lack such principles and laws. Therefore the Boltzmann equation is based on fulfilling the following two goals:

1. *Entropy-ascending principle*: equation (7.1.17) should yield entropy-ascending property:

$$\frac{d}{dt} S(\rho(t)) > 0, \quad (7.1.18)$$

where $S(\rho)$ is the entropy functional in terms of ρ . Physically, we know that

$$S = - \int \rho \ln \rho dx dv + S_0, \quad (7.1.19)$$

where S_0 is a constant, and v is the velocity of molecules.

2. For the Maxwell distribution (7.4.13)

$$\rho_0 = A e^{-v^2/2mkT} \quad (7.1.20)$$

to be an equilibrium state of ideal gases requires that ρ_0 is a steady-state solution of (7.1.17):

$$G(\rho_0) = 0. \quad (7.1.21)$$

The requirement (2) above shows that the probability density function ρ must be a function using both the space time coordinate (t, x) and velocity v as independent variables:

$$\rho = \rho(x, t, v).$$

Consequently the total derivative on the left-hand side of (7.1.17) becomes

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial x} \frac{dx}{dt} + \frac{\partial \rho}{\partial v} \frac{dv}{dt}.$$

Since

$$\frac{dx}{dt} = v, \quad \frac{dv}{dt} = \frac{F}{m} \quad (\text{Newtonian Second Law}),$$

we derive that

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + v \cdot \nabla \rho + \frac{F}{m} \cdot \frac{\partial \rho}{\partial v}. \quad (7.1.22)$$

Next in order to obtain the expression G in the right-hand side of (7.1.17), one needs to use phenomenological approach to come up with a formulation for G satisfying the above two requirements. Boltzmann obtained the following formula

for G ; see Lifschitz and Pitajewski (1990):

$$G(\rho) = \int [\rho' \rho'_1 - \rho \rho_1] \omega' d\Gamma_1 d\Gamma' d\Gamma'_1, \quad (7.1.23)$$

which is used to explain the change due to collisions.

Then (7.1.17) can be written as

$$\frac{\partial \rho}{\partial t} + v \cdot \nabla \rho + \frac{F}{m} \cdot \frac{\partial \rho}{\partial v} = \int [\rho' \rho'_1 - \rho \rho_1] \omega' d\Gamma_1 d\Gamma' d\Gamma'_1, \quad (7.1.24)$$

which is the famous Boltzmann equation.

(2) Problems in Boltzmann Equation

As indicated by Reichl (1998, S11.A). Beyond the Boltzmann equation), computer experiments and a study of correlation functions have shown that the picture of transport phenomena given by Boltzmann is not completely correct. Based on PDP presented earlier, we are now able to demonstrate more precisely the mathematical and physical problems faced by the Boltzmann equation.

1. *Boltzmann equation is not a physical law.* The Boltzmann equation is not established on fundamental principles and laws. The physical meaning of (7.1.24) is simply

$$\text{the rate of change of } \rho = \text{rate of change due to collisions}. \quad (7.1.25)$$

The equality here is the outcome of pure imagination. In fact,

$$\text{collision} = \text{very close distance interaction between particles}. \quad (7.1.26)$$

On the other hand, the real physical law should be

$$\begin{aligned} & \text{the rate of change of } \rho \\ &= \text{the driving force induced by} \\ & \quad \text{the total interaction potential between particles}. \end{aligned} \quad (7.1.27)$$

Comparing (7.1.25)–(7.1.26) and (7.1.27), we see that there is a large deviation between the Boltzmann equation and the real physical law.

2. The Boltzmann equation uses the velocity field v as an independent variable, but v itself is a state function. This inevitably induces large deviation between the Boltzmann equation and the reality. There is a force field in the Boltzmann equation (7.1.24):

$$F = F(t, x, v), \quad (7.1.28)$$

where v is velocity field of particles. Here F should be a function of v , since the velocity field v is treated as an independent variable. This force field is the sum of the external force and the force generated by the total interaction potential of all particles in the system, which includes the force due to collision. Hence F is a new unknown function, and

*the Boltzmann equation (7.1.24) is incomplete,
with another unknown function F .*

3. In order to ensure that the Maxwell distribution ρ_0 given by (7.1.20) is a steady-state solution of the Boltzmann equation (7.1.24), one has to assume that $d\rho_0/dt = 0$, which implies that

$$\frac{F}{m} \cdot \frac{\partial \rho_0}{\partial v} = 0 \quad \iff \quad F \cdot v = 0. \quad (7.1.30)$$

This requires that

$$F = 0. \quad (7.1.31)$$

This assumption (7.1.31) is nonphysical, since under (7.1.31), all particles in the system, with zero force, will make uniform rectilinear motion. This is in contradiction with the reality and leads to the following inconsistency of the Boltzmann equation with the real phenomena:

*the Maxwell distribution (7.1.20) is a steady-state solution
of the Boltzmann equation (7.1.24) under
the nonphysical assumption (7.1.31).*

4. In deriving the H-Theorem (i.e., the entropy-ascending principle), the following must be assumed:

$$F = F(t, x) \text{ is independent of } v. \quad (7.1.33)$$

However, the Boltzmann equation (7.1.24) treats the velocity v as an independent variable. In this case, the force field F must be treated as a function of v as well. Consequently the above assumption (7.1.33) is rather arbitrary, and any model using velocity v as an independent variable is inconsistent with the assumption (7.1.33). In other words, the H-Theorem is not a natural consequence of the Boltzmann equation (7.1.24).

Remark 7.1.4 *In statistical physics, one often regards the Boltzmann equation as a model only suitable for dilute ideal gas. The misperception is that the mutual interaction between particles in a dilute gas system is negligible, leading to the assumption $F = 0$ in (7.1.31). However, the force field generated by the total interaction potential Φ of all particles:*

$$F = -\nabla\Phi \quad (7.1.34)$$

is not small, and is not simply measured by r^{-2} between particles. The motion of particles in any dilute gas system is chaotic, and the velocities of the particles are not small even when there is no external force present. Consequently, the interaction force field is not small and cannot simply be ignored. \square

5. Ignoring the nonphysical nature of assumption (7.1.31), the space of all steady-state solutions of the Boltzmann equation (7.1.24) have five dimensions. Namely, the general form of the steady-state solutions is

$$\bar{\rho} = e^{\alpha_0 + \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v^2},$$

where α_i ($i = 0, \dots, 4$) are constants, and $v = (v_1, v_2, v_3)$. This shows that each steady state is not stable, and this does not fit the reality.

6. Entropy-increasing principle shows that a gaseous system in the equilibrium has the maximum entropy, i.e., the Maxwell distribution (7.1.20) should be the maximum of (7.1.19). However, the maximum of (7.1.19) is given by

$$\rho_0 = e^{-1}.$$

Again, this is nonphysical.

(3) Conclusions

In a nutshell, the above discussions clearly demonstrate the following conclusions:

1. Laws of physics (equations) should not use state functions as independent variables, which are themselves governed by physical laws. The Boltzmann equation violates this simple physical rule, and therefore is not able to faithfully reflect the true nature of the underlying physical phenomena.
2. Irreversibility is a common characteristic of all dissipative systems, and is not an entropy property. The Potential-Descending Principle is the basic mathematical model/description of irreversibility.
3. The Boltzmann equation was aimed to develop a mathematical model for entropy-increasing principle. This is an incorrect starting point, since thermodynamic potential is the right physical quantity for the mathematical characterization of irreversibility, rather than entropy.
4. Let $F(u, \lambda)$ be the thermodynamic potential of a gaseous system with order parameters u and control parameters λ . Then PDP gives rise to the following dynamic law (7.1.1) and (7.1.2). This dynamic law offers a complete description of the irreversibility of the gaseous system.

7.2 First and Second Laws of Thermodynamics

The classical thermodynamics is built upon the three fundamental laws of thermodynamics, the first law, the second law, and the third law of thermodynamics. In this section, we demonstrate that

the PDP leads to both the first and second laws of thermodynamics.

7.2.1 Derivation of the First and Second Laws from PDP

The first law of thermodynamics is an energy conservation law, which states that the internal energy of a thermodynamical system consists of thermal energy, mechanical energy, interaction energy between particles, and so on. These energies can convert from one form to another, or transfer from one body to another. However, during the transfer process, the total energy is conserved.

For an equilibrium state \bar{u} of a thermodynamic system, the potential-descending principle implies that

$$\frac{\delta}{\delta u} F(\bar{u}; \lambda) = 0, \quad (7.2.1)$$

$$dF(\bar{u}, \lambda) = \frac{\partial F(\bar{u}, \lambda)}{\partial \lambda} d\lambda. \quad (7.2.2)$$

These are equilibrium equations of the system. Indeed, for the equilibrium state, it is then easy to see that

$$dF(\bar{u}, \lambda) = \frac{\delta}{\delta u} F(\bar{u}; \lambda) \delta u + \frac{\partial F}{\partial \lambda} d\lambda = \frac{\partial F(\bar{u}; \lambda)}{\partial \lambda} d\lambda,$$

which is the *first law of thermodynamics*.

To derive the second law, for a given nonequilibrium thermodynamic state $u(t)$, we have

$$dF(u(t), \lambda) = \frac{\delta}{\delta u} F(u(t); \lambda) du + \frac{\partial F}{\partial \lambda} d\lambda.$$

The potential-descending principle tells us that

$$\frac{dF}{dt} = \frac{\delta}{\delta u} F(u(t); \lambda) \frac{du}{dt} < 0.$$

As $dt > 0$, we then derive that

$$dF(u, \lambda) < \frac{\partial F}{\partial \lambda} d\lambda,$$

which is the *second law of thermodynamics*.

7.2.2 Thermodynamic Potentials

To have a better understanding of the first and second laws, we examine now the basic thermodynamic quantities and the thermodynamic potentials.

(1) Legendre Transformation

For the precise description of various thermodynamic quantities and the thermodynamical potentials, we need to recall the mathematical problem of the Legendre transformation, stated as follows:

Problem 7.2.1 *Given a function $f(x, y)$ of $x, y \in \mathbb{R}^1$, find a function $F(y, x)$ from $f(x, y)$ and a relation $y = \varphi(x)$ such that on the curve $y = \varphi(x)$, $F(y, x)$ satisfies*

$$\frac{\partial}{\partial y} F(y, x) = 0, \quad y = \varphi(x). \quad (7.2.3)$$

This is achieved mathematically by

$$F(y, x) \stackrel{\text{def}}{=} f(x, y) - xy, \quad (7.2.4)$$

$$x = \frac{\partial f(x, y)}{\partial y} \quad \text{which defines } y = \varphi(x). \quad (7.2.5)$$

In fact, with (7.2.4) and (7.2.5), we have

$$\frac{\partial}{\partial y} F(y, x) = \frac{\partial f(x, y)}{\partial y} - x = 0,$$

which verifies (7.2.3). In addition,

$$dF(y, x) = \frac{\partial F(y, x)}{\partial y} dy + \frac{\partial F(y, x)}{\partial x} dx = \frac{\partial F(y, x)}{\partial x} dx. \quad (7.2.6)$$

which is in the form of the second equation (7.2.2) of the first law.

In summary, originally, for $f(x, y)$, the variable x is the order parameter, which is a function of the control parameter y : $x = \psi(y)$. Now through the Legendre transformation, y becomes the order parameter for the potential F , and is itself a function of the control parameter x : $y = \varphi(x)$ determined by (7.2.5).

Remark 7.2.2 (1) The classical definition of the Legendre transformation of a function $L : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ is

$$L^*(q) = \inf_{p \in \mathbb{R}^1} \{L(p) - pq\}. \quad (7.2.7)$$

This is equivalent to saying that

$$L^*(q) = L(p) - pq \quad (7.2.8)$$

with $p = p(q)$ solving

$$q = L'(p). \quad (7.2.9)$$

(2) To ensure the solvability of this equation, near the origin, we need the convexity condition:

$$L''(0) \neq 0. \quad (7.2.10)$$

In (7.2.5), we need to assume

$$\frac{\partial^2 f(0,0)}{\partial y^2} \neq 0. \quad (7.2.11)$$

This is a very important condition for deriving the exact forms of thermodynamic potentials later in this book. \square

(2) Thermodynamic Quantities

We now make precise the various thermodynamic functions appearing in PDP, and the first law of thermodynamics.

With PDP, for a thermodynamic system, we treat the state functions $u = (u_1, \dots, u_N)$ and the thermodynamic potential F separately as two different levels of thermodynamic quantities with F being the functional of the state functions u : $F = F(u)$.

The state functions are the physical quantities u_1, \dots, u_N , which contribute to the internal energy of the system. There are two types of internal energy:

$$U = \text{thermal energy } Q + \text{mechanical work } W. \quad (7.2.12)$$

The thermal energy is the energy corresponding to the microscopic motion of system particles. Its strength is reflected by a *measurable* quantity T , the temperature, which is a state function of the system. It is reasonable then to assume that there is a corresponding conjugate state function, denoted by S and called entropy of the system, so that

$$\text{thermal energy } Q = TS. \quad (7.2.13)$$

The mechanical energy is due to the work done by the system. For example, the work, done through the change of volume V when the pressure is applied, is given by

$$dW = -PdV, \quad (7.2.14)$$

where P is the pressure, and dV is the change of the volume. Here when $dV > 0$, i.e., volume increases, the system does work to the external and the internal energy decreases. Conversely, if the volume decreases, i.e., $dV < 0$, then $dW > 0$ and the external does work to the system and the internal energy increases.

There are different sources of mechanical energy, such as elastic force, surface tension, magnetic force, and electric strength. All mechanical energies are of the form:

$$\text{mechanical energy } W = fX, \quad (7.2.15)$$

where f is the generalized force, and X is the generalized displacement.

By (7.2.13) and (7.2.15), the general form of the internal energy functional U in (7.2.12) is written as follows:

$$U = U(T, S, f, X). \quad (7.2.16)$$

State functions can be classified into two categories:

extensive variables :	entropy S ,	generalized displacement X ,
intensive variables :	temperature T ,	generalized force f .

All state functions form conjugate pairs, and typical examples of (f, X) include (the pressure P , the volume V), (applied magnetic field H , magnetization M), (applied electric field E , electric polarization P).

In thermodynamics, the response functions provide the needed bridge between theories and experiments. Basically, one can derive theoretical expressions of these functions, and experimentally measure them. Consequently the theoretical results can be experimentally verified through the related measurements.

The main response variables are the heat capacity C , the compression coefficient κ , the thermal expansion coefficient α , and the magnetic susceptibility ξ , defined as follows:

- _ the heat capacity in the constant pressure: $C_p = -T \frac{\partial^2 F}{\partial T^2}$,
- _ the compression coefficient $\kappa = -\frac{1}{V} \frac{\partial^2 F}{\partial p^2}$,
- _ the thermal expansion coefficient $\alpha = \frac{1}{V} \frac{\partial^2 F}{\partial T \partial p}$, and
- _ the magnetic susceptibility: $\partial^2 F / \partial M \partial H$, where M is the magnetization of the material and H is magnetic field strength.

(3) Thermodynamic Potentials

Consider a thermodynamic system enclosed in a three-dimensional domain $\Omega \subset \mathbb{R}^3$. The system makes contacts with the external environment through the following physical quantities:

entropy S , temperature T , generalized displacement X , generalized force f .

We can classify the types of thermodynamic systems based on the way the systems contact with the external environment.

1. *Internal Energy System* An internal energy system refers to the case where the system in the domain Ω is adiabatic (with thermal exchange with the external), and the domain Ω is fixed. For such a system, we have

$$dS = 0 \text{ (adiabatic)}, \quad dV = 0 \text{ (\Omega is fixed)}. \quad (7.2.17)$$

Consequently, the control parameters for the system is

$$\lambda = (\lambda_1, \lambda_2), \quad \lambda_1 = S, \quad \lambda_2 = V, \quad (7.2.18)$$

and the order parameters are

$$u = (u_1, u_2), \quad u_1 = T, \quad u_2 = P \quad (7.2.19)$$

In general, when considering the generalized force f_i and the generalized displacement X_i , the order and control parameters are

$$\lambda = (\lambda_1, \lambda_2, \lambda_3) = (S, V, X_i), \quad u = (u_1, u_2, u_3) = (T, P, f_i). \quad (7.2.20)$$

Hence the thermodynamic potential is the internal energy, written as

$$U = U(u, \lambda). \quad (7.2.21)$$

By PDP, the first law of thermodynamics is given by

$$\frac{\delta}{\delta u} U(u, \lambda) = 0, \quad (7.2.22)$$

$$dU = \frac{\delta U}{\delta S} dS + \frac{\delta U}{\delta V} dV = T dS - P dV. \quad (7.2.23)$$

We remark here that by the implicit function theorem, one should be able to solve the equilibrium equation (7.2.22) to derive that the order parameters are functions of the control parameters: $u = u(\lambda)$.

2. *Enthalpy System* An enthalpy system refers to the case where the system is adiabatic with constant pressure:

$$dS = 0, \quad dP = 0 \quad (\text{or } df = 0). \quad (7.2.24)$$

In this case, P and S are treated as constants, as control parameters, and their conjugates are order parameters:

$$\lambda = (S, P), \quad u = (T, V). \quad (7.2.25)$$

To derive the thermodynamic potential of the enthalpy system, we use the Legendre transformation.

With the above discussions at our disposal, for an enthalpy system, we obtain the following thermodynamic potential with $u = (T, X)$ and $\lambda = (S, f)$:

$$H = H(u, \lambda) = U - fX, \quad f = \frac{\partial U}{\partial X}. \quad (7.2.26)$$

The corresponding first law is written as

$$\frac{\delta}{\delta u} H(u, \lambda) = 0, \quad (7.2.27)$$

$$dH = \frac{\delta H}{\delta S} dS + \frac{\delta H}{\delta f} df. \quad (7.2.28)$$

3. *Helmholtz System* This system refers to the case where the temperature is maintained as a constant and Ω is fixed:

$$dT = 0, \quad dV = 0. \quad (7.2.29)$$

Hence the order and control parameters are given by

$$\lambda = (T, V), \quad u = (S, P). \quad (7.2.30)$$

As for the enthalpy case, we now have the thermodynamic potential for a Helmholtz system given by

$$F = F(u, \lambda) = U - ST, \quad T = \frac{\partial U}{\partial S}. \quad (7.2.31)$$

The corresponding first law for a Helmholtz system is written as

$$\frac{\delta}{\delta u} F(u, \lambda) = 0, \quad (7.2.32)$$

$$dF = \frac{\delta F}{\delta T} dT + \frac{\delta F}{\delta V} dV. \quad (7.2.33)$$

4. *Gibbs System* This corresponds to the case where the temperature and pressure are maintained as constants:

$$dT = 0, \quad dp = 0 \quad (\text{or } df = 0). \quad (7.2.34)$$

so that the order and control parameters are

$$\lambda = (T, P), \quad u = (S, V). \quad (7.2.35)$$

The thermodynamic potential for a Gibbs system is then written as

$$G = U - ST - fX, \\ T = \frac{\partial}{\partial S} U, \quad f = \frac{\partial}{\partial X} U. \quad (7.2.36)$$

The corresponding first law for a Gibbs system is written as

$$\frac{\delta}{\delta u} G(u, \lambda) = 0, \quad (7.2.37)$$

$$dG = \frac{\delta G}{\delta T} dT + \frac{\delta G}{\delta P} dP. \quad (7.2.38)$$

7.3 Foundations of Quantum Physics

We introduce in this section some recent developments in quantum physics, which serve as the quantum physics foundations for the statistical theory of heat that we introduce in Sect. 7.5. For a more detailed account of these recent developments, see Ma and Wang (2015a).

7.3.1 Principles of Quantum Mechanics

(1) History and Wave-Particle Duality

Quantum physics describes the behavior of basic constituents of matter in the universe. With the discovery of the electron by J. J. Thomson in 1897, which marks the beginning of elementary particle physics. The Maxwell theory of electromagnetism can be viewed as a wave theory for photons. The attempt by Max Planck (1900) to explain the blackbody radiation required him to postulate that the energy levels of light must be quantized. It was Albert Einstein (1905), however, who showed that light as an electromagnetic wave behaves also as a particle (the photon), with a discrete quantum of energy that was dependent on its frequency. In 1924, Louis de Broglie postulated the wave nature of electrons and suggested that all matter has wave properties.

It is the wave-particle duality of microscopic energy states that forms the cornerstone and the starting point of quantum mechanics. On one hand, a particle is characterized by its energy ε and momentum \mathbf{P} , and the wave is characterized by its frequency ω and wave vector \mathbf{k} . The connections between these two characterizations are achieved by Einstein and de Broglie with the following de Broglie relations:

$$\varepsilon = \hbar\omega, \quad \mathbf{P} = \hbar\mathbf{k}, \quad (7.3.1)$$

where \hbar is the Planck constant.

For a massive particle with mass m , and velocity \mathbf{v} , according to the Einstein special theory of relativity, de Broglie relations are given by

$$\varepsilon = \hbar\omega = \frac{mc^2}{\sqrt{1 - v^2/c^2}}, \quad \mathbf{P} = \hbar\mathbf{k} = \frac{mv}{\sqrt{1 - v^2/c^2}}. \quad (7.3.2)$$

(2) Schrödinger Equation and Quantum Rules

Following the de Broglie relations (7.3.1), Schrödinger (1926) realized that the wave property of a particle is a plane-wave, described by the following wave function:

$$\Psi(x, t) = Ae^{-i(\omega t - \mathbf{k} \cdot \mathbf{x})} = Ae^{-\frac{i}{\hbar}(\varepsilon t - \mathbf{P} \cdot \mathbf{x})}. \quad (7.3.3)$$

Differentiating this wave function leads to the following

$$i\hbar \frac{\partial \Psi}{\partial t} = \varepsilon \Psi, \quad -i\hbar \nabla \Psi = \mathbf{P} \Psi. \quad (7.3.4)$$

According to classical mechanics, the energy ε and momentum \mathbf{P} satisfy

$$\varepsilon = \frac{1}{2m} \mathbf{P}^2 + V, \quad V \text{ is the external potential energy.} \quad (7.3.5)$$

In view of (7.3.4), (7.3.5) leads to the following Schrödinger equation: for the wave function Ψ :

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \Psi + V \Psi. \quad (7.3.6)$$

The main components of quantum physics include quantum mechanics and quantum field theory, which are based on the following basic postulates.

Quantum Rule 7.3.1 (Born Rule) *For a single particle system described by a wave function Ψ , its modular square $|\Psi(x, t)|^2$ represents the probability density of the particle being observed at point $x \in \mathbb{R}^3$ and at time t . Hence, Ψ satisfies that*

$$\int_{\mathbb{R}^3} |\Psi|^2 dx = 1.$$

Quantum Rule 7.3.2 *Each observable physical quantity L corresponds to an Hermitian operator \hat{L} , and the values of the physical quantity L are given by eigenvalues λ of \hat{L} :*

$$\hat{L} \Psi_\lambda = \lambda \Psi_\lambda,$$

and the eigenfunction Ψ_λ is the state function in which the physical quantity L takes value λ . In particular, the Hermitian operators corresponding to the momentum \mathbf{P} and the energy E are given by

$$\begin{aligned} \text{momentum operator: } & \hat{\mathbf{P}} \Psi = -i\hbar \nabla \Psi, \\ \text{energy operator: } & \hat{E} \Psi = i\hbar \frac{\partial \Psi}{\partial t}. \end{aligned} \quad (7.3.7)$$

When the particle is placed in an electromagnetic field $A_\mu = (A_0, \mathbf{A})$, the above energy and momentum operators are replaced by

$$\begin{aligned} \text{momentum operator: } & \hat{\mathbf{P}} = -i\hbar \nabla + \frac{g}{c} \mathbf{A}, \quad g \text{ is the charge,} \\ \text{energy operator: } & \hat{E} = i\hbar \frac{\partial}{\partial t} - g A_0. \end{aligned} \quad (7.3.8)$$

Quantum Rule 7.3.3 *For a quantum system Ψ and a physical Hermitian operator \hat{L} , Ψ can be expanded as*

$$\Psi = \sum \alpha_k \Psi_k + \int \alpha_\lambda \Psi_\lambda d\lambda, \quad (7.3.9)$$

where Ψ_k and Ψ_λ are the eigenfunctions of \hat{L} corresponding to discrete and continuous eigenvalues respectively. In (7.3.9) for the coefficients α_k and α_λ , their modular square $|\alpha_k|^2$ and $|\alpha_\lambda|^2$ represent the probability of the system Ψ in the states Ψ_k and Ψ_λ . In addition, the following integral, denoted by

$$\langle \Psi | \hat{L} | \Psi \rangle = \int \Psi^\dagger (\hat{L} \Psi) dx, \quad (7.3.10)$$

represents the average value of physical quantity \hat{L} of system Ψ .

Quantum Rule 7.3.4 For a quantum system with observable physical quantities L_1, \dots, L_N , if they satisfy a relation

$$R(L_1, \dots, L_N) = 0,$$

then the quantum system Ψ satisfies the equation

$$R(\hat{L}_1, \dots, \hat{L}_N) \Psi = 0,$$

where \hat{L}_k are the Hermitian operators corresponding to L_k ($1 \leq k \leq N$), provided that $R(\hat{L}_1, \dots, \hat{L}_N)$ is a Hermitian.

(3) Two Fundamental Principles of Quantum Physics

The Heisenberg uncertainty relation and Pauli exclusion principle are two fundamental principles of quantum physics.

We recall that particles are classified into two types:

$$\begin{aligned} \text{fermions} &= \text{particles with spin } J = \frac{n}{2} \text{ for odd } n, \\ \text{bosons} &= \text{particles with spin } J = n \text{ for integer } n. \end{aligned}$$

Fermions and bosons display very different characteristics. The fermions do not like to live together with the same fermions, but bosons are sociable particles. This difference is characterized by the Pauli exclusion principle.

Principle 7.3.5 (Heisenberg Uncertainty Principle) In a quantum system, the position x and momentum p , the time t and energy E satisfy the uncertainty relations given by

$$\Delta x \Delta p \geq \frac{1}{2} \hbar, \quad \Delta t \Delta E \geq \frac{1}{2} \hbar, \quad (7.3.11)$$

where ΔA represents a measuring error of A value. Namely, (7.3.11) implies that x and p , t and E cannot be precisely observed at the same moment.

Principle 7.3.6 (Pauli Exclusion Principle) In a quantum system, there are no two or more fermions living in the same quantum states, i.e., possessing entirely the same quantum numbers.

(4) Summary

The Schrödinger equation (7.3.6) provides the motion dynamic law of the particle. Its solution Ψ provide all the needed information for the particle. In other words, with the Born rule, the physical information of the quantum system is encoded in the wave function Ψ , and can be extrapolated as the expectation (7.3.10) of the Hermitian operator \hat{L} , associated with an observable physical quantity L .

Also, the process in deriving the Schrödinger equation (7.3.6) offers a general principle for deriving other quantum dynamic laws; see among others (Ma and Wang, 2015a) and the references therein.

7.3.2 New Interpretation of Quantum Wave Functions

(1) Fields as Fundamental Physical Quantities of Nature

Recently we discovered a new **field** theoretical interpretation of quantum wave functions (Ma and Wang, 2017e). The field theoretical approach for modern physics goes at least back to Maxwell, who unified the electric and magnetic forces through the electromagnetic potential field $A_\mu = (A_0, \mathbf{A})$. Einstein's theory of general relativity captures the gravity with Riemannian metric $\{g_{\mu\nu}\}$ as the gravitational potential fields. The weak and strong interactions are then described by the gauge potential fields. In a nutshell, fields are used as the fundamental level of physical quantities to describe the nature.

The system (7.3.6) conserves the energy, and hence the wave function Ψ can be expressed as

$$\Psi = e^{-iEt/\hbar} \psi(x), \quad (7.3.12)$$

where E is the energy, and ψ is the time-independent wave function. Putting (7.3.12) into (7.3.6), we derive that

$$-\frac{\hbar^2}{2m} \Delta \psi + V(x)\psi = E\psi. \quad (7.3.13)$$

The solution of (7.3.13) can be expressed as

$$\psi = |\psi| e^{i\varphi}, \quad (7.3.14)$$

where φ is the phase of ψ .

Recall from (7.3.4) or (7.3.7) that

$$-i\hbar\nabla = \text{momentum operator}. \quad (7.3.15)$$

Hence

$$\text{momentum density of } \psi = \text{imaginary part of } \hbar\nabla\psi \cdot \psi^*. \quad (7.3.16)$$

Let the momentum density of ψ be expressed as

$$\text{momentum density of the particle} = m|\psi|^2 v,$$

where v is the velocity of the particle. Then we infer from (7.3.14) and (7.3.16) that

$$v = \frac{\hbar}{m} \nabla \varphi. \quad (7.3.17)$$

Namely the velocity is the gradient of the phase of the wave function (7.3.14). (7.3.17) is consistent with the classical physical conclusion that the current density J is given by

$$J = \frac{ie\hbar}{2m} (\psi \nabla \psi^* - \psi^* \nabla \psi). \quad (7.3.18)$$

(2) New Field Theory Interpretation of QM Wave Function

Notice that (7.3.17) is the velocity *field* of particles, rather than the velocity of a single particle. This leads to the new interpretation of the wave function ψ as follows:

1. Under the external potential field $V(x)$, the wave function ψ in (7.3.13) is the field function for the motion of all particles with the same mass in the same class determined by the external potential $V(x)$. In other words, it is not the wave function of a particular particle in the classical sense;
2. When a particle is observed at a particular point $x_0 \in \Omega$, then the motion of the particle is fully determined by the solution of the following motion equation with initial position at x_0 :

$$\frac{dx}{dt} = \frac{\hbar}{m} \nabla \varphi(x), \quad x(0) = x_0, \quad (7.3.19)$$

where φ is the phase of the wave function ψ in (7.3.14);

3. With ψ being the field function,

$$|\psi(x)|^2 = \text{distribution density of particles at } x; \quad (7.3.20)$$

4. The energy E in (7.3.13) represents the average energy level of the particles and can be written as

$$E = \int_{\Omega} \left[\frac{\hbar^2}{2m} |\nabla \psi|^2 + \frac{\hbar^2}{2m} |\nabla \varphi(x)|^2 + V(x) |\psi|^2 \right] dx, \quad (7.3.21)$$

where in integrand on the right-hand side, the first term represents the nonuniform distribution potential of particles, the second term is the average kinetic energy, and the third term is the potential energy of the external field. Here $\nabla |\psi|$ is the characteristic of quantum mechanics and there is no such term in classical mechanics.

(3) Physical Meaning of New Quantum Interpretation

This new field theoretical interpretation was first discovered by Ma and Wang (2019). It does not alter the basic theories of quantum mechanics, and instead offers new understanding of quantum mechanics. It also plays a fundamental role for the quantum theory of condensed matter physics studied in the next chapter.

The classical Born statistical interpretation of quantum mechanics amounts to saying that without constraints, the motion of a micro-particle is random and there is no definite trajectory of the motion. Also $|\psi(x)|^2$ stands for the probability density of the particle appearing at the particular point x . This leads to the classical Einstein–Bohr debates, and is the main reason for absurdities associated with the interpretation of quantum mechanics.

The new interpretation says that $\psi = |\psi|e^{i\varphi}$ is the common wave function for all particles in the same class determined by the external potential $V(x)$, $|\psi(x)|^2$ represents the distribution density of the particles, and $\frac{\hbar}{m}\nabla\varphi$ is the velocity field of the particles. The trajectories of the motion of the particles are given by (7.3.19). The observed particles are the particles in the same class described by the same wave function, rather than a specific particle in the classical quantum mechanics. Hence the initial position of a particle is determined by the distribution $|\psi|^2$, and its motion trajectory is fully determined by (7.3.19). This is an entirely different interpretation from the classical Bohr interpretation, and will lead to impact to quantum physics and condensed matter physics.

In summary, there are field functions ψ_1 , ψ_2 , $|\psi|$, and φ in the wave function

$$\psi = \psi_1 + i\psi_2 = |\psi|e^{i\varphi},$$

whose physical meanings are given as follows:

1. ψ_1 and ψ_2 are the conjugate functions of the following Hamilton system

$$\begin{aligned} \hbar \frac{\partial \psi_1}{\partial t} &= k \frac{\delta}{\delta \psi_2} H(\psi), \\ \hbar \frac{\partial \psi_2}{\partial t} &= -k \frac{\delta}{\delta \psi_1} H(\psi); \end{aligned}$$

2. $|\psi(x)|^2$ represents the distribution density of particles;
3. $\nabla|\psi|$ is the potential describing the nonuniform distribution of particles, and $\frac{\hbar^2}{2m}|\nabla|\psi||^2$ is the potential energy for the nonuniform distribution of particles; and
4. $\frac{\hbar}{m}\nabla\varphi$ is the velocity field of particles.

7.3.3 Introduction to Elementary Particles

(1) Weakton Model of Elementary Particles

The weakton model of elementary particles was first introduced by Ma and Wang (2015c,a). This theory proposes six elementary particles, which we call weaktons, and their anti-particles:

$$\begin{array}{ccccccc} w^*, & w_1, & w_2, & \nu_e, & \nu_\mu, & \nu_\tau, \\ \bar{w}^*, & \bar{w}_1, & \bar{w}_2, & \bar{\nu}_e, & \bar{\nu}_\mu, & \bar{\nu}_\tau, \end{array} \quad (7.3.22)$$

where ν_e, ν_μ, ν_τ are the three generation neutrinos, and w^*, w_1, w_2 are three new particles, which we call w -weaktons. These are all massless particles with spin $J = \frac{1}{2}$. Each of them carries a weak charge, and only w^* and \bar{w}^* carry a strong charge. Also, the neutrinos do not carry electric charge, w^* carries 2/3 electric charge, w_1 carries $-1/3$ electric charge, and w_2 carries $-2/3$ electric charge.

The weakton constituents of charged leptons and quarks are given by

$$\begin{aligned} e &= \nu_e w_1 w_2, & \mu &= \nu_\mu w_1 w_2, & \tau &= \nu_\tau w_1 w_2, \\ u &= w^* w_1 \bar{w}_1, & c &= w^* w_2 \bar{w}_2, & t &= w^* w_2 \bar{w}_1, \\ d &= w^* w_1 w_2, & s &= w^* w_1 w_2, & b &= w^* w_1 w_2, \end{aligned} \quad (7.3.23)$$

where c, t and d, s, b are distinguished by the spin arrangements.

The weakton constituents of the mediators and their dual mediators are given by

$$\begin{aligned} \gamma &= \cos \theta_w w_1 \bar{w}_1 - \sin \theta_w w_2 \bar{w}_2 (\uparrow\uparrow, \downarrow\downarrow) && \text{vector photon,} \\ \gamma_0 &= \cos \theta_w w_1 \bar{w}_1 - \sin \theta_w w_2 \bar{w}_2 (\uparrow\downarrow, \downarrow\uparrow) && \text{scalar photon,} \\ g^k &= w^* \bar{w}^* (\uparrow\uparrow, \downarrow\downarrow), && \text{vector gluons,} \\ g_0^k &= w^* \bar{w}^* (\uparrow\downarrow, \downarrow\uparrow) && \text{scalar gluons} \end{aligned} \quad (7.3.24)$$

The ν -mediator ν has spin-0 with the following weakton constituents:

$$\nu = \alpha_1 \nu_e \bar{\nu}_e + \alpha_2 \nu_\mu \bar{\nu}_\mu + \alpha_3 \nu_\tau \bar{\nu}_\tau (\downarrow\uparrow), \quad \sum_{l=1}^3 \alpha_l^2 = 1. \quad (7.3.25)$$

Each gluon carries two strong charges and two weak charges, and participates both the weak and strong interactions. Both photon and the ν mediator only carry respectively two weak charges, and participate the weak interaction, but not the strong interaction. All three mediators carry no electric charge.

(2) Mass Generation Mechanism

For a particle moving with velocity v , its mass m and energy E obey the Einstein relation

$$E = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (7.3.26)$$

Usually, we regard m as a static mass which is fixed, and energy is a function of velocity v .

Now, taking an opposite viewpoint, we regard energy E as fixed, mass m as a function of velocity v , and the relation (7.3.26) is rewritten as

$$m = \sqrt{1 - \frac{v^2}{c^2} \frac{E}{c^2}}. \quad (7.3.27)$$

Thus, (7.3.27) means that a particle with an intrinsic energy E has zero mass $m = 0$ if it moves at the speed of light $v = c$, and will possess nonzero mass if it moves with a velocity $v < c$.

All particles including photons can only travel at the speed sufficiently close to the speed of light. Based on this viewpoint, we can think that if a particle moving at the speed of light (approximately) is decelerated by an interaction force \vec{F} , obeying

$$\frac{d\vec{P}}{dt} = \sqrt{1 - \frac{v^2}{c^2}} \vec{F},$$

then this massless particle will generate mass at the instant. In particular, by this mass generation mechanism, several massless particles can yield a massive particle if they are bound in a small ball, and rotate at velocities less than the speed of light.

For the mass problem, we know that the mediators

$$\gamma, g^k, \nu \quad \text{and their dual particles,} \quad (7.3.28)$$

have no masses. To explain this, we note that these particles in (7.3.24) and (7.3.25) consist of pairs as

$$w_1 \bar{w}_1, \quad w_2 \bar{w}_2, \quad w^* \bar{w}^*, \quad v_l \bar{v}_l. \quad (7.3.29)$$

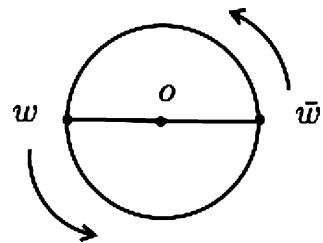
The weakton pairs in (7.3.29) are bound in a circle with radius R_0 as shown in Fig. 7.1. Since the interacting force on each weakton pair is in the direction of their connecting line, they rotate around the center 0 without resistance. As $\vec{F} = 0$ in the moving direction, by the relativistic motion law:

$$\frac{d}{dt} \vec{P} = \sqrt{1 - \frac{v^2}{c^2}} \vec{F}, \quad (7.3.30)$$

the massless weaktons rotate at the speed of light. Hence, the composite particles formed by the weakton pairs in (7.3.29) have no rest mass.

For the massive particles

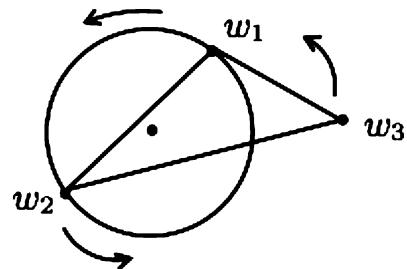
Fig. 7.1 The weakton pair rotate around the center O without resistance, forming a massless particle



$$e, \mu, \tau, u, d, s, c, t, b, \quad (7.3.31)$$

by (7.3.23), they are made up of weakton triplets with different electric charges. Hence the weakton triplets are arranged in an irregular triangle as shown in Fig. 7.2. Consequently, the weakton triplets rotate with nonzero interacting forces $\vec{F} \neq 0$ from the weak and electromagnetic interactions. By (7.3.30), the weaktons in the triplets at a speed less than the speed of light due to the resistance force. Thus, by the mass generating mechanism above, the weaktons become massive. Hence, the particles in (7.3.31) are massive.

Fig. 7.2 The weakton triplets, arranged in an irregular triangle, forming a massive particle



7.3.4 Photon Cloud Model of Electrons

(1) Photon Clouds of Electrons

The weakton constituents of an electron are $v_e w_1 w_2$. Therefore, an electron carries three weak charges, which are the source of the weak interaction. A photon has the weakton constituents as given by (7.3.24), and carries two weak charges.

The weak force formula between the naked electron and a photon γ is given by; see Ma and Wang (2015a):

$$F = -g_w(\rho_\gamma)g_w(\rho_e) \frac{d}{dr} \left[\frac{1}{r} e^{-kr} - \frac{B}{\rho} (1 + 2kr) e^{-2kr} \right]$$

$$= g_w(\rho_m)g_w(\rho_e)e^{-kr} \left[\frac{1}{r^2} + \frac{1}{rr_0} - \frac{4B}{\rho} \frac{r}{r_0^2} e^{-kr} \right], \quad (7.3.32)$$

where $k = 1/r_0 = 10^{16} \text{ cm}^{-1}$, $g_w(\rho_m)$ and $g_w(\rho_e)$ are the weak charges of mediators and the naked electron, expressed as

$$g_w(\rho_\gamma) = 2 \left(\frac{\rho_w}{\rho_\gamma} \right)^3 g_w, \quad g_w(\rho_e) = 3 \left(\frac{\rho_w}{\rho_e} \right)^3 g_w,$$

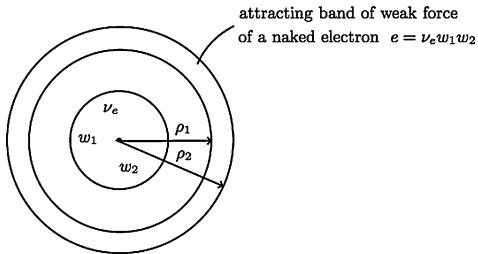
and B/ρ is a parameter determined by the naked electron and the photon.

By the weak force formula (7.3.32), there is an attracting shell region of weak interaction between naked electron and the photon

$$F < 0 \quad \text{for } \rho_1 < r < \rho_2, \quad (7.3.33)$$

as shown in Fig. 7.3. Since photons carry weak charges, they are attached to the

Fig. 7.3 Electron structure



electron in the attracting shell region (7.3.33), forming a cloud of photons. The irregular triangle distribution of the weakons ν_e, w_1, w_2 generate a small moment of force on the mediators. Meanwhile there also exist weak forces between them. Therefore the bosons will rotate at a speed less than the speed of light, and generate a small mass attached to the naked electron $\nu_e w_1 w_2$.

(2) Angular Momentum Rule

The Angular Momentum Rule 7.3.7 below was first discovered in Ma and Wang (2015c,a, 2016a). It ensures that the photons in the clouds of electrons can only be scalar photons $J = 0$, and consequently the photon cloud of an electron does not change the spin of the electron $J = 1/2$.

Angular Momentum Rule 7.3.7 *Only the fermions with spin $J = \frac{1}{2}$ and the bosons with $J = 0$ can rotate around a center with zero moment of force. The particles with $J \neq 0, \frac{1}{2}$ will move on a straight line unless there is a nonzero moment of force present.*

Also we remark that based on the mechanism of decay and scattering of particles, weakton exchanges may occur during the following γ - γ scattering process, leading to the transformation between scalar photons and vector photons:

$$\gamma + \gamma \longrightarrow \gamma + \gamma,$$

and the corresponding weakton constituent exchange is given by

$$w_1\bar{w}_1(\uparrow\uparrow) + w_1\bar{w}_1(\downarrow\downarrow) \rightleftharpoons w_1\bar{w}_1(\uparrow\downarrow) + w_1\bar{w}(\uparrow\downarrow).$$

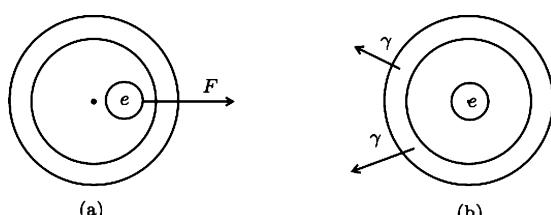
This observation shows that although the photons in the photon cloud of an electron can only be scalar photons, both scalar and vector photons are abundant in Nature.

(3) Photon Absorption and Radiation Mechanism of Electrons

The attracting shell region in Fig. 7.3 of an electron results in the ability for the electron to attract and emit photons. A macroscopic system is immersed in a sea of photons (mediators). When a photon enters the attracting shell region of an electron, it will be absorbed by the electron. An electron emits photons as its velocity changes, which is called the bremsstrahlung. Also, when the orbiting electron jumps from higher energy level to a lower energy level, it radiates photons. Hence electrons in the system are constantly in a state of absorbing and emitting photons, resulting changes on their energy levels. As we shall see in the next section, at the equilibrium of absorption and radiation, the average energy level of the system maintains unchanged, and represents the temperature of the system; at the same time, the number (density) of photons in the sea of photons represents the entropy (entropy density) of the system.

The reasons why bremsstrahlung can occur is unknown in classical theories. Based on the electron structure theory in Sect. 9.3.2, this phenomenon can be easily explained.

Fig. 7.4 (a) The naked electron is accelerated or decelerated in an electromagnetic field; and **(b)** the mediators (photons) fly away from the attracting shell region under a perturbation of moment of force



In fact, if an electron is situated in an electromagnetic field, then the electromagnetic field exerts a Coulomb force on the naked electron $v_e w_1 w_2$, but not on the attached neutral mediators. Thus, the naked electron changes its velocity, which

draws the mediator cloud to move as well, causing a perturbation to moment of force on the mediators. As the attracting weak force in the shell region (7.3.27) is small, under the perturbation, the centrifugal force makes some mediators in the cloud, such as photons, flying away from the attracting shell region, and further accelerated by the weak repelling force outside this shell region to the speed of light, as shown in Fig. 7.4.

7.3.5 Energy Levels of Particles

This section is based on Ma and Wang (2016b, 2015a). The mass m , energy E , and the momentum \vec{p} of a particle obey the Einstein energy-momentum relation; see among many others (Landau and Lifshitz, 1975; Ma and Wang, 2015a):

$$E^2 = m^2 c^4 + c^2 p^2. \quad (7.3.34)$$

There are different energy levels, which can undergo changes by (1) absorbing and/or emitting photons, and (2) exchanging the interior constituents. The weaktons are elementary particles, and all other particles are composite. For composite particles, the energy levels determined by their constituents are called intrinsic energy levels, which can change through exchanging constituents.

Energy levels of particles play an important role in statistical physics. There are many particles in Nature, and we are interested in energy levels of the following particles, which play crucial role in statistical physics:

$$\text{photons, electrons, and atoms.} \quad (7.3.35)$$

Hereafter we focus on the energy levels of these particles.

(1) Energy Levels of Photons

The weakton constituents of a photon are the following two weaktons, symmetrically bounded together by the weak force:

$$\gamma = w_i \bar{w}_i \quad i = 1, 2.$$

It suffices for us to consider the bounded states of one weakton. As the weakton $w_i (i = 1, 2)$ are massless, the wave function describing them is the two-component Weyl spinor:

$$\psi = (\psi_1, \psi_2), \quad (7.3.36)$$

and the corresponding wave equations are

$$(\vec{\sigma} \cdot \vec{D}) \frac{\partial \psi}{\partial t} = c(\vec{\sigma} \cdot \vec{D})^2 \psi - \frac{ig_w}{2\hbar} \{(\vec{\sigma} \cdot \vec{D}), A_0\} \psi, \quad (7.3.37)$$

where $\{A, B\} = AB + BA$ is the anti-commutator, $\vec{\sigma} = (\sigma^1, \sigma^2, \sigma^3)$ is the Pauli matrix operator, the operator \vec{D} is defined by

$$\vec{D} = \nabla + i \frac{g_w}{\hbar c} \vec{W}, \quad (7.3.38)$$

with $W_\mu = (W_0, \vec{W})$ being the weak interaction potential of weaktons.

The spectral equation of photons can be derived by setting

$$\psi = e^{-i\lambda t/\hbar} \varphi, \quad \varphi = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix},$$

where λ is the bounding energy. We infer from (7.3.37) that

$$-\hbar c(\vec{\sigma} \cdot \vec{D})^2 \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} + \frac{i g_w}{2\hbar} \{(\vec{\sigma} \cdot \vec{D}), W_0\} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} = i\lambda(\vec{\sigma} \cdot \vec{D}) \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}. \quad (7.3.39)$$

Since the weaktons are confined in the photon, we can set $\psi = 0$ outside of the photon. Consequently, we have the following boundary condition:

$$\varphi = 0 \quad \text{for } |x| = \rho_\gamma, \quad (7.3.40)$$

where ρ_γ is the radius of photons.

We can then derive the following conclusions for energy levels of photons using the above linear eigenvalue problems (7.3.39) and (7.3.40).

- There are finite number of negative eigenvalues for (7.3.39) and (7.3.40), representing the bounding energies of the weaktons:

$$\infty < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N < 0; \quad (7.3.41)$$

- There are finite number of energy levels for photons, given by

$$E_k = E_0 + \lambda_k \quad \text{for } 1 \leq k \leq N, \quad (7.3.42)$$

where E_0 is the intrinsic energy of the two weakton constituents of the photon. Hence the energy levels of a photon are finite:

$$0 < E_1 \leq E_2 \leq \dots \leq E_N; \quad (7.3.43)$$

- The frequencies of a photon are discrete:

$$\omega_k = E_k/\hbar, \quad \Delta\omega_k = \omega_{k+1} - \omega_k = (\lambda_{k+1} - \lambda_k)/\hbar; \quad (7.3.44)$$

- The number of energy levels of photons can be estimated as follows:

$$N = \left(\frac{B_w \rho_\gamma g_w^2}{\beta_1 \rho_w \hbar c} \right)^3 \simeq 10^{90}, \quad (7.3.45)$$

and the energy differences can be estimated as

$$\Delta E \simeq \frac{E_{\max} - E_{\min}}{N} = \frac{\lambda_N - \lambda_1}{N} \simeq 10^{-45} \text{ eV}, \quad (7.3.46)$$

which is small and unobservable.

(2) Energy Levels of Electrons

The electrons are massive with three weakton constituents: $\nu_e w_1 w_2$. As mentioned earlier, these three weaktons possess different electric charges, and are arranged in an irregular triangle as shown in Fig. 7.2, becoming the massive. Hence they are governed by three Dirac spinors:

$$\psi^j = (\psi_1^j, \dots, \psi_4^j), \quad j = 1, 2, 3.$$

It is then easy to derive the energy level equation for an electron as follows:

$$\begin{aligned} & -\frac{\hbar^2}{2m_j}(\nabla + i\frac{2g_w}{\hbar c}\mathbf{W})^2\varphi^j + 2(2g_w W_0 + \vec{\mu}_j \cdot \text{curl}\mathbf{W})\varphi^j = \lambda\varphi^j, \\ & \varphi^j = 0 \quad \text{for } j = 1, 2, 3, \quad |x| = \rho_e, \end{aligned} \quad (7.3.47)$$

where ρ_e is the radius of an electron, $W_\mu = (W_0, \mathbf{W})$ is the weak interaction potential, $\varphi^j = (\varphi_1^j, \varphi_2^j)$ are the eigenstates of the j -th weakton, and

$$\vec{\mu}_j = \frac{\hbar g_w}{2m_j} \vec{\sigma}$$

is the weak magnetic moment of the j -th weakton.

We derive from (7.3.47) the following conclusions:

1. The intrinsic energy levels of electrons are finite and discrete.
2. The number N of intrinsic energy levels of an electron can be approximately estimated as

$$N = \left[\frac{4}{\lambda_1} \frac{B_w \rho_e^2}{\rho_w} \frac{m_w c}{\hbar} \frac{g_w^2}{\hbar c} \right]^{3/2} \sim 10^{45},$$

where ρ_e is the radius of the electron, ρ_w is the radius of the weakton, B_w is the weak interaction parameter in (7.3.32), m_w is the mass of the constituent weaktons of the electron caused by nonzero interacting force from the weak and the electromagnetic interactions, and λ_1 is the first eigenvalue of $-\Delta$.

3. In view of the photon cloud structure that an electron consists of the naked electron and the shell-layer of its photons cloud, the total number N of energy levels of electrons is about

$$N = \text{number of intrinsic energy levels} \times \text{number of energy levels of photon} \simeq 10^{135}.$$

(3) Energy Levels of Atoms

Classical energy levels was developed based on the Bohr atomic models and the Schrödinger equations. The is made up of nucleus and the orbiting electrons, the nucleus is made up of protons p and neutrons n , which are made up of three quarks: $p = uud$, $n = udd$. In addition, the weakton constituents of upper and down quarks are $u = w^*w_1\bar{w}_1$ and $d = w^*w_1w_2$. Therefore, the energy levels of an atom is the sum of energy levels of the nucleus and the energy levels of the orbiting electrons, and the energy levels of the nucleus are determined by

$$\text{energy levels of a nucleus} = E_k^1 + \lambda_j^1,$$

where E_k^1 are the energy levels of nucleons, λ_j^1 are the negative eigenvalues of the spectral equation for the atom, representing the bounding energies bounding the nucleons. Therefore,

$$E_k^1 = E_l^2 + \lambda_j^2 + \text{energy levels of absorbed mediators of the nucleons}.$$

Here E_l^2 are the energy levels of quarks, λ_j^2 are the negative eigenvalues of the spectral equation of the nucleons, representing the bounding energy between quarks. Finally,

$$E_l^2 = E_0^3 + \lambda_j^3 + \text{energy levels of the mediators absorbed by quarks},$$

where E_0^3 is the intrinsic energy of the weaktons in the quark, and λ_j^3 is the negative eigenvalues of the spectral equation for the quark, representing the bounding energies bounding the weaktons inside the quarks.

(4) Physical Conclusions of Energy Levels of Particles

In summary, we have the following physical conclusions for energy levels of particles, which provide the particle physics foundation of statistical physics.

1. The energies of micro-particles are on their energy levels, and there are finite number of energy levels, which are discrete;
2. Particles can jump to different energy levels by (a) absorbing or emitting photons, and (b) exchanging their constituent particles.
3. The energy of a particle obeys the Einstein energy-momentum relation (7.3.34). When the energy level of a particle changes, its mass and momentum will undergo changes as well. For a fixed energy level, the mass and the momentum can undergo transformations between each other.
4. The number N of energy levels is large, and the gas between adjacent energy levels are small; they can be estimated roughly estimated as

$$\text{number of energy levels} \simeq \begin{cases} 10^{90} & \text{for photons,} \\ 10^{135} & \text{for electrons,} \\ 10^{300} & \text{for atoms,} \end{cases}$$

$$\Delta E_k = E_{k+1} - E_k \simeq 10^{-45} \text{ eV} \quad \text{for photons.}$$

7.4 Statistics of Equilibrium Systems

For a thermodynamic system, a main component of the statistical theory is to study the probability distribution of particles in different energy levels of the system in equilibrium. In this section, we aim to derive all three distributions: the Maxwell–Boltzmann distribution, the Fermi–Dirac distribution, and the Bose–Einstein distribution, based only on

- (1) the principle of equal *a priori* probabilities (PEP), and
- (2) the potential-descending principle (PDP), Principle 7.1.1.

The first-principle derivation was first given by Ma and Wang (2017a). Classically, these distributions are derived assuming that the maximum probability is achieved at the equilibrium. However there is no such principle of maximum probability in Nature.

In summary,

PDP leads to all three distributions: the Maxwell–Boltzmann distribution, the Fermi–Dirac distribution, and the Bose–Einstein distribution, and consequently, is the fundamental first principle of statistical physics.

7.4.1 Maxwell–Boltzmann Statistics for Classical Systems

Now consider an isolated classical thermodynamic system, the total energy E and the total number of particles N are constants. Each particle in the system must be situated on an energy level ε , and the total energy is finite. Since the system is at equilibrium, we have

$$a_n \stackrel{\text{def}}{=} [\text{number of particles on } \varepsilon_n] = [\text{probability of particles at } \varepsilon_n] \times N. \quad (7.4.1)$$

The PEP ensures the time-independence of a_n . We have then the following arrangement of number of particles on different energy levels:

$$\begin{array}{ccccccccc} \varepsilon_1 & < & \varepsilon_2 & < & \cdots & < & \varepsilon_{N_E}, \\ g_1 & & g_2 & & \cdots & & g_{N_E}, \\ a_1 & & a_2 & & \cdots & & a_{N_E}, \end{array} \quad (7.4.2)$$

where g_n represents the degeneracy factor (allowed quantum states) of the energy level ε_n . For an isolated system,

$$N = \sum_n a_n = \text{constant}, \quad E = \sum_n a_n \varepsilon_n = \text{constant}. \quad (7.4.3)$$

It is clear that the multiplicity function W for the distribution (7.4.2) is given by:

$$W = W(a_1, \dots, a_{N_E}). \quad (7.4.4)$$

The aim is then to find the relations between ε_n and a_n under constraint (7.4.3):

$$a_n = f(\varepsilon_n, T) \quad \text{for } 1 \leq n \leq N_E. \quad (7.4.5)$$

For an isolated thermodynamic system, the temperature T is a control parameter, and consequently, its thermodynamic potential functional is

$$F = E - ST, \quad (7.4.6)$$

where E is the total energy, and S is the entropy. Classically, the entropy is given by the famous Boltzmann formula:

$$S = k \ln W, \quad (7.4.7)$$

where $k = 1.381 \times 10^{-23}$ J/K is the Boltzmann constant.

By the potential-descending principle, Principle 7.1.1, the distribution $\{a_n\}$ at the thermodynamic equilibrium solves the following minimal potential variational equations of the potential functional (7.4.6):

$$\frac{\delta}{\delta a_n} \left[-kT \ln W + \alpha_0 \sum_n a_n + \beta_0 \sum_n a_n \varepsilon_n \right] = 0, \quad (7.4.8)$$

where α_0 and β_0 are the Lagrangian multipliers of constraints (7.4.3).

In view of the distribution of particles (7.4.2), the multiplicity function W is given by

$$W = \frac{N!}{\prod_n a_n!} \prod_n g_n^{a_n}, \quad (7.4.9)$$

By the Stirling formula

$$k! = k^k e^{-k} \sqrt{2\pi k},$$

and using $\ln(2\pi k) \ll k$, we have

$$\ln W = N \ln N - N - \sum_n a_n \ln a_n + \sum_n a_n + \sum_n a_n \ln g_n.$$

Since $\sum_n a_n = N$, we arrive at

$$\ln W = N \ln N - \sum_n a_n \ln \frac{a_n}{g_n},$$

which implies from (7.4.8) that

$$\ln \frac{a_n}{g_n} + \alpha + \beta \varepsilon_n = 0 \quad \text{for } 1 \leq n \leq N_E.$$

This gives rise to the famous Maxwell–Boltzmann distribution:

$$a_n = \frac{N}{Z} g_n e^{-\varepsilon_n/kT}. \quad (7.4.10)$$

Here Z is the partition function defined by

$$Z = \sum_n g_n e^{-\beta \varepsilon_n}, \quad (7.4.11)$$

Also, direct computation shows that the MB distribution is indeed the minimal point of the potential functional (7.4.6) under the constraints (7.4.3).

The partition function Z defined by (7.4.11) is another important thermodynamical quantity in statistical mechanics. In fact, once we know the detailed expression of the partition function, we can derive other related thermodynamical quantities as follows:

$$\begin{aligned} U &= -N \frac{\partial}{\partial \beta} \ln Z && \text{internal energy,} \\ S &= Nk \left(\ln Z - \beta \frac{\partial}{\partial \beta} \ln Z \right) && \text{entropy,} \\ f &= -\frac{N}{\beta} \frac{\partial}{\partial X} \ln Z && \text{generalized force,} \\ F &= -NkT \ln Z && \text{potential functional.} \end{aligned} \quad (7.4.12)$$

With the above MB distribution, it is easy to derive also the Maxwell velocity distribution of an ideal gas system consisting of particles with mass m :

$$f(v) = 4\pi \left(\frac{m}{2\pi kT} \right)^{3/2} v^2 e^{-\frac{mv^2}{2kT}}. \quad (7.4.13)$$

7.4.2 Quantum Systems

(1) Bose–Einstein Distribution

The Maxwell–Boltzmann statistics is for classical systems of particles. For systems where quantum behavior is prominent, quantum statistics is then ultimately needed. Quantum statistics consists of the Bose–Einstein (BE) statistics for systems of bosonic particles, and the Fermi–Dirac statistics for systems of fermionic particles.

The goal here is the same as the statistics for classical particle systems: to find the relations (7.4.5) between ε_n and a_n under constraint (7.4.3).

For a quantum system of bosonic particles, the multiplicity function associated with (7.4.2) is

$$W_{\text{BE}} = \prod_n \frac{g_n + a_n - 1)!}{a_n!(g_n - 1)!}. \quad (7.4.14)$$

As for the Maxwell–Boltzmann distribution, solving (7.4.8) leads to the following Bose–Einstein distribution:

$$a_n = \frac{g_n}{e^{(\varepsilon_n - \mu)/kT} - 1}, \quad (7.4.15)$$

where μ is the chemical potential.

For a quantum system of bosonic particles, the partition function is given by

$$Z = \prod_n \left(1 - e^{-\alpha - \beta \varepsilon_n}\right)^{-g_n}, \quad \alpha = -\frac{\mu}{kT}, \quad \beta = \frac{1}{kT}. \quad (7.4.16)$$

Also we can derive other related thermodynamical quantities, the total number of particles N , the total energy E , the generalized force f , the pressure p , the entropy S , the free energy F , and the Gibbs energy G , in terms of the partition function Z as follows:

$$\begin{aligned} N &= -\frac{\partial}{\partial \alpha} \ln Z, \\ E &= -\frac{\partial}{\partial \beta} \ln Z, \\ f &= -\frac{1}{\beta} \frac{\partial}{\partial X} \ln Z, \\ p &= \frac{1}{\beta} \frac{\partial}{\partial V} \ln Z, \\ S &= k \left[\ln Z - \left(\alpha \frac{\partial}{\partial \alpha} + \beta \frac{\partial}{\partial \beta} \right) \ln Z \right], \\ F &= E - ST = -kT \left(\ln Z - \alpha \frac{\partial}{\partial \alpha} \ln Z \right), \\ G &= -NkT\alpha = kT\alpha \frac{\partial}{\partial \alpha} \ln Z. \end{aligned} \quad (7.4.17)$$

(2) Fermi–Dirac Distribution

Fermions obey the Pauli exclusion principle. Hence for a quantum system of fermions, the multiplicity function is given by

$$W_{\text{FD}} = \prod_n \frac{g_n!}{a_n!(g_n - a_n)!}. \quad (7.4.18)$$

Then it is easy to find the following Fermi–Dirac distribution (7.4.5) between ε_n and a_n under constraint (7.4.3):

$$a_n = \frac{g_n}{e^{(\varepsilon_n - \mu)/kT} + 1}, \quad (7.4.19)$$

where μ is the chemical potential.

Also the partition function for a Fermi system is given by

$$Z = \prod_n \left(1 + e^{-\alpha - \beta \varepsilon_n}\right)^{g_n}, \quad \alpha = -\frac{\mu}{kT}, \quad \beta = \frac{1}{kT}. \quad (7.4.20)$$

The relations between the partition function and other thermodynamical quantities are given by (7.4.17) as well.

7.5 Statistical Theory of Heat

In this section, we present a new statistical theory of heat (Ma and Wang, 2017d). One motivation of the theory is the lack of physical carriers of heat in classical thermodynamics. Another motivation is the recently discovered photon cloud structure of electrons. This leads to a natural conjugate relation between electrons and photons, reminiscent of the conjugate relation between temperature and entropy. The new theory provides a natural connection between these two conjugate relations: at the equilibrium of absorption and radiation, the average energy level of the system maintains unchanged, and represents the temperature of the system; at the same time, the number (density) of photons in the sea of photons represents the entropy (entropy density) of the system. The theory contains four parts: (1) the photon number formula of entropy, (2) the energy level formula of temperature, (3) the temperature theorem, and (4) the thermal energy formula. In particular, the photon number entropy formula is equivalent to the Boltzmann entropy formula, and, however, possesses new physical meaning that the physical carrier of heat is the photons.

7.5.1 Energy Level Formula of Temperature

We have introduced the photon cloud structure of subatomic particles in Sect. 7.3.4. In this section, we use such structure of sub-atomic particles to reveal the nature of temperature and entropy.

Among basic constituents of matter, electrons, protons, and neutrons are fundamentally important. Both protons and neutrons are confined in the nucleons, and electrons are the only charged particles abundant inside the matter. The essence of thermal radiation is the radiation and absorption of photons. With the photon cloud structure of electrons, electrons, and photons form a pair of conjugate physical carri-

ers for absorption and emission associated with thermal radiation. On the other hand, thermal energy is the conjugate relation between temperature and entropy. Hence a correct statistical theory of heat must make a precise connection of the following correspondence:

$$\begin{array}{c} \text{conjugation between electrons and photons} \\ \Downarrow \\ \text{conjugation between temperature and entropy.} \end{array} \quad (7.5.1)$$

The main objective of this section is to derive the following temperature formula:

$$kT = \begin{cases} \sum_n \left(1 - \frac{a_n}{N}\right) \frac{a_n \varepsilon_n}{N(1 + \beta_n \ln \varepsilon_n)} & \text{for classical systems,} \\ \sum_n \left(1 + \frac{a_n}{g_n}\right) \frac{a_n \varepsilon_n}{N(1 + \beta_n \ln \varepsilon_n)} & \text{for Bose systems,} \\ \sum_n \left(1 - \frac{a_n}{g_n}\right) \frac{a_n \varepsilon_n}{N(1 + \beta_n \ln \varepsilon_n)} & \text{for Fermi systems.} \end{cases} \quad (7.5.2)$$

If we view

$$\bar{\varepsilon} = \frac{1}{N} \sum_n a_n \varepsilon_n$$

as the average energy level for the thermodynamical system, then the above formulas show that temperature is simply the weighted average energy level of the system.

Hereafter we derive these formulas using the basic distributions.

Classical Systems Consider a classical equilibrium thermodynamic system with energy levels of the particles given by

$$\varepsilon_1, \quad \varepsilon_2, \dots, \quad \varepsilon_{N_E}. \quad (7.5.3)$$

By the MB distribution (7.4.10), the total energy of the system is

$$E = \sum_n a_n \varepsilon_n = \frac{N}{Z} \sum_n g_n \varepsilon_n e^{-\varepsilon_n/kT}, \quad (7.5.4)$$

where N is the total number of particles, and $Z = \sum_n g_n e^{-\varepsilon_n/kT}$ is the partition function.

When we find the total energy E , we can view (7.5.4) as an equation defining an implicit function of the temperature T in terms of the energy levels in (7.5.3):

$$T = T(\varepsilon_1, \dots, \varepsilon_{N_E}).$$

Physically, it means that under the invariance of the total energy E , the distribution $\{a_n\}$ changes as $\{\varepsilon_n\}$ vary, leading to the change of the temperature T . Hence we can assume the following expression of T :

$$T = \sum_n \alpha_n T(\varepsilon_n), \quad (7.5.5)$$

where the coefficients α_n are to be determined.

Physically, it is natural to assume that

fluctuations on a specific energy level ε_n will only lead to fluctuations on the energy $a_n \varepsilon_n$ on the level ε_n in the total energy $E = \sum_m a_m \varepsilon_m$. (7.5.6)

Mathematically, by (7.5.6), the implicit function relation can be determined using the following variation:

$$0 = \frac{\partial E}{\partial \varepsilon_n} = N g_n \frac{\partial}{\partial \varepsilon_n} \left[\frac{\varepsilon_n}{Z} e^{-\varepsilon_n/kT} \right] \quad (7.5.7)$$

$$= N \left[g_n \left(\frac{1}{Z} - \frac{\varepsilon_n}{Z^2} \frac{\partial Z}{\partial \varepsilon_n} \right) e^{-\varepsilon_n/kT} + \frac{g_n \varepsilon_n}{Z} e^{-\varepsilon_n/kT} \left(-\frac{1}{kT} + \frac{\varepsilon_n}{kT^2} \frac{\partial T}{\partial \varepsilon_n} \right) \right],$$

$$\begin{aligned} \frac{\partial Z}{\partial \varepsilon_n} &= \left(-\frac{g_n}{kT} + \frac{g_n \varepsilon_n}{kT^2} \frac{\partial T}{\partial \varepsilon_n} \right) e^{-\varepsilon_n/kT} \\ &= -\frac{Z}{N} \frac{a_n}{kT} + \frac{a_n \varepsilon_n}{kT^2} \frac{Z}{N} \frac{\partial T}{\partial \varepsilon_n}. \end{aligned} \quad (7.5.8)$$

Consequently, we have

$$\begin{aligned} 0 &= a_n - \frac{a_n \varepsilon_n}{Z} \frac{\partial Z}{\partial \varepsilon_n} - \frac{a_n \varepsilon_n}{kT} + \frac{a_n \varepsilon_n^2}{kT^2} \frac{\partial T}{\partial \varepsilon_n} \\ &= a_n + \frac{a_n^2 \varepsilon_n}{N} \frac{1}{kT} - \frac{a_n^2 \varepsilon_n^2}{N kT^2} \frac{\partial T}{\partial \varepsilon_n} - \frac{a_n \varepsilon_n}{kT} + \frac{a_n \varepsilon_n^2}{kT^2} \frac{\partial T}{\partial \varepsilon_n}, \end{aligned}$$

which implies the following differential equation for $T(\varepsilon_n)$:

$$\frac{\partial T}{\partial \varepsilon_n} = \frac{T}{\varepsilon_n} - k \left(1 - \frac{a_n}{N} \right)^{-1} \frac{T^2}{\varepsilon_n^2}. \quad (7.5.9)$$

Let $x = \varepsilon$, then we infer from (7.5.9) that

$$T' = \frac{T}{x} - k \left(1 - \frac{a_n}{N} \right)^{-1} \frac{T^2}{x^2}. \quad (7.5.10)$$

Now let

$$y \stackrel{\text{def}}{=} \frac{T}{x}, \quad \text{then } T' = xy' + y.$$

Equation (7.5.10) becomes

$$xy' = -k \left(1 - \frac{a_n}{N} \right)^{-1} y^2. \quad (7.5.11)$$

Physically, we may assume that

$$\frac{a_n}{N} \simeq \text{constant.} \quad (7.5.12)$$

Then (7.5.11) is equivalent to

$$-\left(1 - \frac{a_n}{N}\right) \int \frac{dy}{ky^2} = \int \frac{dx}{x}. \quad (7.5.13)$$

Since $x = \varepsilon_n$ carries the dimension of energy, the solution of (7.5.13) is given by

$$\frac{(1 - a_n/N)}{ky} = \ln \frac{\varepsilon_n}{\varepsilon_0} + C_n,$$

where ε_0 is the unit of energy. Hence we obtain that

$$kT(\varepsilon) = \left(1 - \frac{a_n}{N}\right) \frac{\varepsilon_n}{C_n + \ln \varepsilon/\varepsilon_0}.$$

As discussed earlier, all $T(\varepsilon_m)$ should take the same form, we obtain that

$$kT = \sum \left(1 - \frac{a_n}{N}\right) \frac{\alpha_n \varepsilon_n}{C_n + \ln \varepsilon_n/\varepsilon_0}.$$

Let $\beta_n = 1/C_n$ and $\theta_n = \alpha_n/C_n$, then we have

$$kT = \sum_n \left(1 - \frac{a_n}{N}\right) \frac{\theta_n \varepsilon_n}{1 + \beta_n \ln \varepsilon/\varepsilon_0}. \quad (7.5.14)$$

We now try to determine the coefficients θ_n . For this purpose, we first define the translational derivative with respect to all energy levels:

$$\begin{aligned} T' &= \lim_{\Delta\varepsilon \rightarrow 0} \frac{1}{\Delta\varepsilon} \left[T(\varepsilon_1 + \Delta\varepsilon, \varepsilon_2 + \Delta\varepsilon, \dots) - T(\varepsilon_1, \varepsilon_2, \dots) \right] \\ &= \sum_n \theta_n \left(1 - \frac{a_n}{N}\right) \left(1 - \beta_n - \beta_n \ln \frac{\varepsilon_n}{\varepsilon_0}\right). \end{aligned} \quad (7.5.15)$$

Then we take the translation derivative on both sides of (7.5.4). By the physical assumption (7.5.6), we obtain that $\delta\varepsilon_n = \delta\varepsilon$ for all n and

$$0 = \delta E = \sum_n g_n \frac{\partial}{\partial \varepsilon_n} \left[\frac{\varepsilon_n}{Z} e^{-\varepsilon_n/kT} \right] \delta \varepsilon_n. \quad (7.5.16)$$

Also by (7.5.6), we have

$$g_n \frac{\partial}{\partial \varepsilon_n} \left[\frac{\varepsilon_n}{Z} e^{-\varepsilon_n/kT} \right] = g_n \left(\frac{1}{Z} - \frac{\varepsilon_n}{Z^2} \frac{\partial Z}{\partial \varepsilon} \right) e^{-\varepsilon_n/kT} + \frac{g_n \varepsilon_n}{Z} \left(-\frac{1}{kT} + \frac{\varepsilon_n}{kT^2} T' \right) e^{-\varepsilon_n/kT}, \quad (7.5.17)$$

where T' is as in (7.5.15), and by (7.5.6), we use the following approximation for the contribution of $\partial Z/\partial \varepsilon$ to the n -th energy level:

$$\frac{\partial Z}{\partial \varepsilon} = g_n \left(-\frac{1}{kT} + \frac{\varepsilon_n}{kT^2} T' \right) e^{-\varepsilon_n/kT}. \quad (7.5.18)$$

Hence by (7.5.16)-(7.5.18), we obtain that

$$kT = \sum_n \frac{a_n}{N} \left(1 - \frac{a_n}{N} \right) \varepsilon_n - \sum_n \frac{a_n}{N} \left(1 - \frac{a_n}{N} \right) \frac{T'}{T} \varepsilon_n^2. \quad (7.5.19)$$

On the other hand,

$$\begin{aligned} kT &= \sum_n \theta_n \left[1 - \frac{a_n}{N} \right] \varepsilon_n \times \frac{1}{1 + \beta_n \ln \varepsilon_n / \varepsilon_0} \\ &= \sum_n \left\{ \theta_n \left[1 - \frac{a_n}{N} \right] \varepsilon_n - \theta_n \left[1 - \frac{a_n}{N} \right] \beta_n \varepsilon_n \ln \frac{\varepsilon_n}{\varepsilon_0} \right\}. \end{aligned} \quad (7.5.20)$$

We deduce then from (7.5.19) and (7.5.20) that

$$\theta_n = \frac{a_n}{N}, \quad \beta_n = \frac{T' \varepsilon_n}{T \ln(\varepsilon_n / \varepsilon_0)}, \quad (7.5.21)$$

and consequently (7.5.2) for classical systems follows.

Quantum Systems For a quantum system, we first recall the Bose–Einstein statistics (7.4.15) or the Fermi–Dirac statistics (7.4.19):

$$a_n = \frac{g_n}{e^{(\varepsilon_n - \mu)/kT} \pm 1} \quad \left(\begin{array}{l} + \text{ for FD} \\ - \text{ for BE} \end{array} \right). \quad (7.5.22)$$

The total energy is written as

$$E = \sum_n \frac{g_n \varepsilon_n}{e^{(\varepsilon_n - \mu)/kT} \pm 1}. \quad (7.5.23)$$

As in the classical particle system case, with the assumption (7.5.6), by differentiating E with respect to ε_n , we obtain that

$$\frac{a_n}{g_n} \frac{\varepsilon_n (\varepsilon_n - \mu)}{kT^2} e^{\frac{\varepsilon_n - \mu}{kT}} T' = \frac{a_n}{g_n} \frac{\varepsilon_n}{kT} e^{\frac{\varepsilon_n - \mu}{kT}} - 1. \quad (7.5.24)$$

By (7.5.22), we have

$$e^{\frac{\varepsilon_n - \mu}{kT}} = \frac{g_n \pm a_n}{a_n} \quad \begin{cases} + \text{ for FD} \\ - \text{ for BE} \end{cases}. \quad (7.5.25)$$

We infer then from (7.5.24) and (7.5.25) that

$$T' = \frac{T}{\varepsilon - \mu} - \frac{g_n}{g_n \pm a_n} \frac{kT^2}{\varepsilon(\varepsilon - \mu)}. \quad (7.5.26)$$

Here $T' = \frac{\partial T}{\partial \varepsilon_n}$ and $\varepsilon = \varepsilon_n$. The solution of (7.5.26) is

$$kT(\varepsilon_n) = \left(1 \pm \frac{a_n}{g_n}\right) \frac{\varepsilon}{C_n + \ln \varepsilon} \quad \begin{cases} + \text{ for FD} \\ - \text{ for BE} \end{cases}. \quad (7.5.27)$$

Then as in the case for classical particle systems, we derive the following temperature formula:

$$kT = \sum_n \left(1 \pm \frac{a_n}{g_n}\right) \frac{a_n \varepsilon_n}{N(1 + \beta_n \ln \varepsilon_n)} \quad \begin{cases} + \text{ for FD} \\ - \text{ for BE} \end{cases}. \quad (7.5.28)$$

7.5.2 Physical Meaning of the Temperature Formula

Equation (7.5.2) enables us to have a better understanding on the essence of temperature. In short,

1. *the essence of temperature T is (weighted) average energy level,*
2. *the temperature T is a function of distributions $\{a_n\}$ and the energy levels $\{\varepsilon_n\}$, and*
3. *the parameters $\{\beta_n\}$ in the temperature formula reflect the property of the material.*

We now discuss some further physical implications of the temperature formula.

1. *Absolute Zero for Fermi Particle Systems* For a Fermi particle thermodynamic system,

$$T = 0 \text{ K} \iff \text{either } a_n = g_n \text{ or } a_n = 0. \quad (7.5.29)$$

Basic quantum mechanics shows that if the lower energy level is not fully occupied, then particles on the higher energy level are not stable, and will spontaneously jump to lower energy levels, unless there are always photons that excite the particles on the higher energy level. Consequently, (7.5.29) can be rewritten as

$$T = 0 \text{ K} \iff \begin{cases} a_n = g_n & \text{for } n = 1, \dots, m, \\ a_n = 0 & \text{for } n > m. \end{cases} \quad (7.5.30)$$

This is an exact solution of the temperature formula for Fermi particle systems.

Solid state systems at $T = 0\text{ K}$ are usually Fermi systems, since in a solid state system, atoms and molecules are fixed at lattice points, and the corresponding energy levels are determined by the orbiting electrons. Therefore such systems can be regarded as Fermi systems consisting of orbiting and free electrons.

2. *Absolute Zero for Bose Particle Systems* For a Bose particle system, we have

$$T = 0\text{ K} \iff \varepsilon_1 = 0, a_1 = N \text{ and } a_n = 0 \forall n > 1. \quad (7.5.31)$$

This corresponds exactly to the Bose–Einstein condensation. With temperature at absolute zero, states in a Bose particle system can only be in two forms: a gaseous state or a condensed state of a subsystem in an object.

3. *Classical Systems* For a classical particle system with temperature at absolute zero, we have

$$T = 0\text{ K} \iff a_1 = N \text{ and } a_n = 0 \forall n > 1. \quad (7.5.32)$$

Here we do not need to assume $\varepsilon_1 = 0$, which corresponds to superconductivity or condensation states of superfluids.

The above results derived from the temperature formula are in agreement with physical facts on $T = 0\text{ K}$.

4. *Existence of highest temperature.* Based on the theory of energy levels in Sect. 7.3.5, the number of energy levels of all particles are finite. Consequently, we infer from the temperature formula (7.5.2) the upper limit of T :

$$kT_{\max} < \varepsilon_{\max}, \quad \varepsilon_{\max} \stackrel{\text{def}}{=} \max_n \frac{\varepsilon_n}{1 + \beta_n \ln \varepsilon_n}. \quad (7.5.33)$$

We used the temperature formula for classical particle systems to derive (7.5.33), since at high temperature, the system can be regarded as a classical particle system.

7.5.3 Theory of Entropy

(1) Physical Meaning of Entropy

As the electrons in the system represent in a natural way all the particles in the system, the energy level theory of temperature amounts to saying that

$$\text{temperature } T = \text{average energy level of electrons}. \quad (7.5.34)$$

We now develop the theory of entropy based on the dual relation (7.5.1). In view of both (7.5.1) and (7.5.34), we deduce the following new description of entropy:

$$\text{entropy} = \text{certain sense of number of photons in the system}. \quad (7.5.35)$$

This equivalence (7.5.35) provides a starting point for the new theory of entropy, which we shall explore in this section.

(2) Physical Supports of Entropy as Number of Photons

1. The first law of thermodynamics amounts to saying that for a given thermodynamical system, the internal energy consists of thermal energy, mechanical energy, interaction energy, etc., which can transform among each other and from one system to another, maintaining the total internal energy invariant. In particular we have

$$SdT + TdS = 0, \quad (7.5.36)$$

which shows that in an isolated system, thermal fluctuation follows the rule that temperature increasing or decreasing corresponds to entropy decreasing or increasing.

At the same time, it is clear that

- *a particle absorbs photons if and only if its energy level increases and the number of photons between particles in the system decreases, and*
- *a particle emits photons if and only if its energy level decreases and the number of photons between particles in the system increases.*

$$(7.5.37)$$

It is clear that (7.5.36) and (7.5.37) are consistent. This shows clearly that the first law of thermodynamics offers a strong support for entropy being the number of photons in the system depicted in (7.5.35).

2. Long-range transfer is one important characteristic of thermal energy $Q = TS$. The temperature T is the average energy levels of particles, does not possess the long-range transfer feature, and can only be transferred through kinetic energies of particles. Therefore, the long-range transfer can only be achieved through the entropy S . On the other hand, it is clear that photons radiation is the only possible candidate. Hence (7.5.35) should be valid, and in other words, the characteristic of long-range transfer of thermal energy provides a physical support for (7.5.35).
3. First we can have the following law of entropy transfer.

Law of Entropy Transfer 7.5.1 *Assume the transfer of thermal and other forms energies is negligible. When two thermodynamic systems undergo thermal exchange, the entropy increasing for one system always leads to the entropy decreasing for the other system. Also, the entropy increases for the heat input system, and decreases for the heat output system.*

This law supports (7.5.35). Without particle exchange, thermal energy can only be transferred through either thermal radiation or transfer of kinetic energy of the system particles. With thermal radiation, energy levels of particles in heat *output* system decreases. It is clear then that with photon numbers and energy levels in equilibrium,

decreasing of energy levels leads to the absorption of more photons, reducing the number of photons. For heat input system, the kinetic energy and energy levels of particles increase. This increase of energy levels causes emission of more photons, for photon numbers and energy levels to returning to their original equilibrium.

This verifies the agreement between the law of entropy transfer and the entropy theory (7.5.35).

(3) Photon Number Formula of Entropy

For a given thermodynamic system, in view of (7.5.35), we characterize entropy as the number of photons in the photon clouds between system particles, or the photon density of the photon gas in the system. As discussed earlier, thermal radiation is simply photon radiation (γ radiation). Also, photons are Bosons and obey the Bose-Einstein distribution. In this case, since the total number of photons is not fixed, the chemical potential $\mu = 0$. Then the BE distribution is written as

$$a_n = \frac{g_n}{e^{(\varepsilon_n - \mu)/kT} - 1}. \quad (7.5.38)$$

Hence the total energy of the photon gas in the system is given by

$$E = \sum_n a_n \varepsilon_n. \quad (7.5.39)$$

The corresponding partition function Z is given by

$$Z_B = \prod_n [1 - e^{-\varepsilon_n/kT}]^{-g_n}, \quad \ln Z_B = \sum_n g_n \ln [1 - e^{-\varepsilon_n/kT}]^{-1}. \quad (7.5.40)$$

Consequently, by the entropy formula (7.4.17):

$$S = k \left[\ln Z_B - \beta \frac{\partial}{\partial \beta} \ln Z_B \right],$$

we obtain that

$$S = k \sum_n \left[g_n \ln \frac{e^{\varepsilon_n/kT}}{e^{\varepsilon_n/kT} - 1} + \frac{g_n}{e^{\varepsilon_n/kT} - 1} \cdot \frac{\varepsilon_n}{kT} \right]. \quad (7.5.41)$$

By the distribution (7.5.38), we have

$$\frac{e^{\varepsilon_n/kT}}{e^{\varepsilon_n/kT} - 1} = 1 + \frac{a_n}{g_n},$$

by which we infer from (7.5.41) that

$$S = k \sum_n \left[g_n \ln \left(1 + \frac{a_n}{g_n} \right) + \frac{a_n \varepsilon_n}{kT} \right]. \quad (7.5.42)$$

Since for any photon gaseous system, we always have $a_n \ll g_n$, which implies that

$$\ln \left(1 + \frac{a_n}{g_n} \right) \approx \frac{a_n}{g_n}.$$

Therefore, we derive from (7.5.42) the following photon number formula of entropy:

$$S = k N_0 \left[1 + \frac{1}{kT} \sum_n \frac{\varepsilon_n a_n}{N_0} \right], \quad (7.5.43)$$

where $N_0 = \sum_n a_n$ is the total number of photons, and $\sum_n \frac{\varepsilon_n}{kT} a_n$ represents the number of photons in the sense of average energy level. Notice that N_0 accounts only the photons between systems particles, not those in the clouds of electrons. Basically, thanks to the mechanism of photon radiation and absorption mechanism, at the equilibrium of absorption and radiation, the average energy level of the system maintains unchanged, and represents the temperature of the system; at the same time, the number (density) of photons in the sea of photons represents the entropy (entropy density) of the system.

(4) Law of Temperature

By the temperature formula and the entropy formula, we arrive immediately the following law of temperature.

Theorem 7.5.2 (Law of Temperature) *The following physical assertions hold true for temperature:*

1. *There are minimum and maximum values of temperature with $T_{\min} = 0$ and T_{\max} being given by (7.5.33);*
2. *When the number of photons in the system is zero, the temperature is at absolute zero; namely, the absence of photons in the system is the physical reason causing absolute zero temperature;*
3. *(Nernst Theorem) With temperature at absolute zero, the entropy of the system is zero;*
4. *With temperature at absolute zero, all particles fill all lowest energy levels.*

7.5.4 Essence of Heat

The theory of temperature and entropy developed in the previous sections provides a theoretical foundation for the theory of heat. We further explore in this section the consequence of the theory to reveal the nature of heat.

(1) Thermal Energy

First, in classical thermodynamics, thermal energy is defined as

$$\Delta Q = \Delta U - \Delta W, \quad (7.5.44)$$

where ΔQ represents the thermal energy absorbed by the system, ΔU is the change of internal energy, and ΔW is the work done by the system. At a thermal equilibrium, we have

$$dU = TdS - pdV. \quad (7.5.45)$$

Physically, this differential equation can be understood as follows. First, for a given thermodynamical system, the absorbed (released) heat dQ is given by

$$dQ = dU.$$

On the other hand,

$$dQ = TdS + SdT,$$

Therefore

$$dU = TdS + SdT. \quad (7.5.46)$$

Since the volume of the system can change, the change of system temperature corresponds to the work done by the system:

$$SdT = \text{work} (= -pdV). \quad (7.5.47)$$

Then substituting SdT in (7.5.46) by (7.5.47), we arrive at (7.5.45).

Now we use the statistical theory of heat presented in the previous two sections to explain the above thermodynamic process. By the entropy formula (7.5.43), the thermal energy can be written as

$$Q_0 = ST = E_0 + kN_0T, \quad (7.5.48)$$

where E_0 is the total energy of photons as given by (7.5.39), and N_0 is the number of photons. By (7.5.48), the absorbing thermal energy is given by

$$dQ = dQ_0 = dE_0 + kTdN_0 + kN_0dT,$$

by which we infer from (7.5.46) that

$$dU = dE_0 + kTdN_0 + kN_0dT. \quad (7.5.49)$$

With the same reason as (7.5.47), we obtain then that

$$dU = dE_0 + kTdN_0 - pdV. \quad (7.5.50)$$

Equation (7.5.50) is the differential equation for the new theory of heat expressed in (7.5.48), which is the equivalent form of the classical equation (7.5.45). Its physical

meaning is clear: When the system absorbs thermal energy dQ (photon energy), the increased internal energy consists of two parts: one is to do the work $-pdV$, and the other is to change the total energy E_0 and the number of photons N_0 in the system. By the entropy formula (7.5.43), the change of E_0 and N_0 amounts to the change of entropy with constant temperature; namely,

$$dE_0 + kT dN_0 = T dS.$$

Hence the new theory of heat we established here is consistent with the classical theory, and importantly, offers the new physical meaning.

(2) Balance Between Temperature and Entropy

The statistical theory of heat presented in the previous sections tells us that thermal energy is the product of temperature and entropy, representing the energy of the photons in the system. Entropy represents the number of photons, and the temperature is the average energy level of the matter particles in the system.

Also, temperature and entropy can transfer between each other. When the system matter particles absorb photons, the temperature increases, and the entropy decreases; while the particles radiate photons, the temperature decreases, and the entropy increases. We need now to examine conditions for radiation and absorption of photons.

1. *Absorption Condition* Based on the photon cloud structure of electrons, each electron possesses a layer which absorbs photons. The allowable photons in each such layer is finite, and if the layer is saturated with photons, no more photons can be absorbed. Based on the Bohr atomic theory, the orbiting energy level of an atom is given by

$$E_1 < E_2 < \dots < E_k < E_{\max}, \quad (7.5.51)$$

where E_k is the highest energy level, and E_{\max} is the escaping energy. For an electron at the energy level, it can absorb photons with energy level $E = E_{i+j} - E_i$ such that $E < E_{\max} - E_i$.

2. *Radiation Condition* There is no photon radiation for particles with uniform motion. Also, for an electron at energy level E_i with all lower energy levels in (7.5.51) filled, it will not radiate photons. The typical particle radiations include: atomic radiation, bremsstrahlung, Cherenkov radiation, and the radiation of electromagnetic polaritons. Basic electromagnetism shows that for bremsstrahlung, the emitting energy per unit time is given by

$$W = \frac{1}{6\pi^2\epsilon_0} \frac{e^2 a^2}{c^3}, \quad (7.5.52)$$

where ϵ_0 is the electric permittivity, e is the electric charge, a is the acceleration, and c the speed of light.

3. *Vibration Mechanism of Photon Absorption and Radiation* We can see then that a particle can only absorb and radiate photons while experiencing vibratory motion. By (7.5.52), the higher the frequency of the vibration of the particle, the larger the absorbing and radiating energy. Also, only the vibratory kinetic energy of the particle can be transferred to the energy levels of the particles, and the average kinetic energy of the macroscopic system does not affect much of the energy levels of the micro-particles in the system. The vibration of the particles in the system is caused by collisions between system particles, by collisions between the system particles and the photons (mediators), and by absorbing and radiating photons.
4. *Transformation and Balance Between Temperature and Entropy* For particles in high speed vibration and collision, the rate of photon emission and absorption increases, causing the energy and number density of photons to increase, and consequently leading to the increase of entropy density. Conversely, absorbing more photons by the system leads to the temperature increase. Hence it is clear that there is a natural connection between temperature and entropy.

(3) Zeroth Law of Thermodynamics

The zeroth law of thermodynamics states as follows:

Zeroth Law of Thermodynamics 7.5.3 *If two thermodynamic systems are each in thermal equilibrium with a third, then they are in thermal equilibrium with each other.*

This law is considered as the basis for temperature, and it is a common view that this law would lead to the following conclusion:

the temperatures are the same for all thermodynamic systems that are in equilibrium with each other. (7.5.53)

We examine this conclusion with the new theory of heat. First we recall the energy level temperature equation (7.5.2), restated here for convenience:

$$kT = \sum_n \left(1 - \frac{a_n}{N}\right) \frac{a_n \varepsilon_n}{N(1 + \beta_n \ln \varepsilon_n)}, \quad (7.5.54)$$

which shows clearly that the temperature depends on three ingredients: the distribution $\{a_n\}$, the energy levels $\{\varepsilon_n\}$, and the parameters $\{\beta_n\}$. The energy levels $\{\varepsilon_n\}$ and the parameters $\{\beta_n\}$ are system (object)-dependent. The distribution $\{a_n\}$ depends also on the total energy E_0 of photons and the total number N_0 of photons in the system. Hence with the same total photon energy and the same photon density, different systems (objects) induce different temperatures. Therefore the conclusion (7.5.53) is meaningless for different objects.

With the new theory of heat, what the zeroth law truly implies is that

the energies and number densities of photons are the same for all thermodynamic systems that are in equilibrium with each other. (7.5.55)

Therefore, with the new theory the photon density in the object can be viewed as a measure for its temperature.

(4) Caloric Theory of Heat

The current accepted theory of heat is also called the mechanical theory of heat, which related the heat with mechanical work. The theory was first introduced in 1798 by Benjamin Thompson, and further developed by such great scientists as Sadi Carnot, Rudolf Clausius, and James Clerk Maxwell.

The caloric theory of heat is an obsolete theory that heat is made up of a fluid called caloric that is massless and flows from hotter bodies to colder bodies. It was considered that caloric was a massless gas that exists in all matter, and is conserved. However the transfer between heat and mechanical work makes the caloric theory obsolete.

The statistical theory of heat is developed based on physical theories on fundamental interactions, the photon cloud model of electrons, the first law of thermodynamics, statistical theory of thermodynamics, radiation mechanism of photons, and energy level theory of micro-particles. The theory utilizes rigorous mathematics to reveal the physical essence of temperature, entropy and heat.

The statistical theory of heat established here revives the old caloric theory in the sense that

$$\text{photons are the caloric of heat.} \quad (7.5.56)$$

Photons possess all the required characteristics of caloric of heat: massless, ability to penetrate to matter, conserved in a certain sense, association of temperature with the quantity of photons in the system (object), and so on. One most important characteristic of the new theory is that it gives a natural explanation of the long-range heat transfer.

For example, for the heat phenomena associated with friction, the new theory indicates that the kinetic energy is transferred to system particles through friction, increasing the vibration kinetic energy of the particles. Then based on the radiation and vibration mechanism of photons, the high speed vibration of particles increases the energy levels of the particles and the absorbing and radiating frequencies of the surrounding photons; this leads to the cumulation of high density and high energy level photons. Hence the friction will increase the temperature of the object.

7.6 Thermodynamic Potentials

The potential-descending principle (PDP) is a fundamental principle of statistical physics and leads to the standard model (7.1.1) and (7.1.2). The standard model

offers a complete description of associated phase transitions and transformation of the system from nonequilibrium states to equilibrium states. With the standard model, an important issue boils down to find a better and more accurate account of the thermodynamic potentials.

In this section, we use a first-principle approach to systematically thermodynamic potentials. For condensates, the thermodynamic potentials and Hamiltonian energies are given in Sects. 8.3 and 8.4.

The study presented in this section was first given in Liu et al. (2019). Our study is based on (1) SO(3) symmetry of thermodynamical potentials, (2) theory of fundamental interaction of particles, (3) the statistical theory of heat, (4) quantum rules for condensates, and (5) the dynamical transition theory.

Of course, the study presented here relies on the rich previous work done by pioneers in the related fields; we refer the interested readers to, among many others, (Pathria and Beale, 2011; Reichl, 1998; Landau and Lifshitz, 1969; Lifshitz and Pitaevskii, 1980; Lifschitz and Pitajewski, 1990; Kadanoff, 2000; Fisher, 1998; Stanley, 1971; Chaikin and Lubensky, 2000; Kleman and Laverntovich, 2007; de Gennes, 1966; Ginzburg, 2004; Tinkham, 1996; Gor'kov, 1968; Pitaevskii and Stringari, 2016; Kosterlitz and Thouless, 1973) and the references therein for more details.

7.6.1 *SO(n) Symmetry and Basic Forms of Thermodynamic Potentials*

(1) *SO(n) Symmetry*

Symmetry plays a fundamental role in physics and is characterized by three ingredients:

spaces, transformation groups, and tensors,

which are applicable to different physical fields. The $SO(n)$ symmetry is also called rotational symmetry, and we shall see that $SO(3)$ symmetry plays a crucial role in determining the mathematical expression of thermodynamical potentials in statistical physics.

For $SO(n)$ symmetry, the space is the n-dimensional Euclidean space \mathbb{R}^n , and the transformation group is $SO(n)$ defined by

$$SO(n) = \{n \times n \text{ real matrices } A \mid AA^T = I, \det A = 1\}. \quad (7.6.1)$$

More precisely, let $x = (x_1, \dots, x_n)$ be the coordinates of a vector in \mathbb{R}^n under an orthogonal coordinate system. Consider the following orthogonal transformation

$$\tilde{x} = Ax \quad \text{or} \quad \tilde{x}_i = a_{ij} x_j \quad \text{for } A = (a_{ij}) \in SO(n), \quad (7.6.2)$$

where the summation convention is used for repeated indices. The corresponding tensors are Cartesian tensors with transformations given by (7.6.2).

For a thermodynamical system, there are three levels of variables: order parameters u (state variables), the control parameters λ , and the thermodynamic potential functional $F = F(u, \lambda)$, which are functionals of u and λ . The order parameters and control parameters are given by in their general form by

$$\begin{array}{ll} \text{order parameters} & u = (u_1, \dots, u_N), \\ \text{control parameters} & \lambda = (\lambda_1, \dots, \lambda_k). \end{array}$$

If $u = u(x)$ depends on the position $x \in \Omega$, the general expression of F is written as

$$F(u, \lambda) = \int_{\Omega} f(u, \dots, D^k u, \lambda) dx \quad (k \geq 1), \quad (7.6.3)$$

where $\Omega \subset \mathbb{R}^n$ ($n = 2, 3$) is the domain that the system occupies, $D^k u$ is the k -th derivative of u . As mentioned earlier, the potential functional (7.6.3) is $SO(n)$ invariant. Namely, under the coordinate transformation (7.6.2), the integrand $f(u, \dots, D^k u, \lambda)$ in (7.6.3) is invariant, and hence it is an $SO(n)$ invariant. This symmetry requirement leads to natural restrictions on the expression of f , which we explore hereafter.

1. An Example of Invariants Consider the following gradient square of a scalar function u :

$$f(u) = |\nabla u|^2 = \sum_{i=1}^n \left(\frac{\partial u}{\partial x_i} \right)^2. \quad (7.6.4)$$

Under the transformation (7.6.2), by

$$\tilde{\nabla} u^T = A \nabla u^T \quad (\text{or } \nabla u^T = A^T \tilde{\nabla} u^T),$$

where A is an orthogonal matrix as in (7.6.2). Then we have

$$f = \nabla u \cdot \nabla u^T = \tilde{\nabla} u A \cdot A^T \tilde{\nabla} u^T = \tilde{\nabla} u \cdot \tilde{\nabla} u^T, \quad (7.6.5)$$

where $A^T = A^{-1}$. Hence

$$f = |\tilde{\nabla} u|^2 = \sum_{i=1}^n \left(\frac{\partial u}{\partial \tilde{x}_i} \right)^2, \quad (7.6.6)$$

which shows that the expression (7.6.6) of f with coordinates \tilde{x} is the same as the expression (7.6.4) of f with coordinates x . This is the meaning of invariants.

2. Basic Forms of Invariants For a scalar field u , the following are invariants associated with the first-order derivatives of u :

$$\begin{aligned} |\nabla u|^2 &= \left(\frac{\partial u}{\partial x_1} \right)^2 + \dots + \left(\frac{\partial u}{\partial x_n} \right)^2, \\ \vec{a} \cdot \nabla u &= a_1 \frac{\partial u}{\partial x_1} + \dots + a_n \frac{\partial u}{\partial x_n} \end{aligned} \quad (7.6.7)$$

where \vec{a} is a given vector field.

For a vector field $\vec{u} = (u_1, \dots, u_n)$, the following are invariants associated with the first-order derivatives of u :

$$\begin{aligned} |\vec{u}|^2 &= u_1^2 + \dots + u_n^2, \\ \vec{a} \cdot \vec{u} &= a_1 u_1 + \dots + a_n u_n, \\ \operatorname{div} \vec{u} &= \frac{\partial u_1}{\partial x_1} + \dots + \frac{\partial u_n}{\partial x_n}, \\ |\nabla \vec{u}|^2 &= |\nabla u_1|^2 + \dots + |\nabla u_n|^2. \end{aligned} \quad (7.6.8)$$

3. *Basic Forms of Thermodynamic Potentials* Known physical facts are showing that thermodynamical potential functionals depend at most on derivatives of u up to the first-order; and this leads to the following hypothesis:

Hypothesis 7.6.1 *For a nonuniform thermodynamical system, the potential functional (7.6.3) depends only on u and its first-order derivatives with $k = 1$.*

With this assumption and the invariants (7.6.7)–(7.6.8) involving the first-order derivatives of u , we obtain the following general forms of thermodynamical potentials:

$$F = \int_{\Omega} \left[\frac{\alpha}{2} |\nabla u|^2 + \frac{\beta}{2} \vec{a} \cdot \nabla u + g(u, \lambda) \right] dx \quad \text{for scalar fields,} \quad (7.6.9)$$

$$F = \int_{\Omega} \left[\frac{\alpha}{2} |\nabla \vec{u}|^2 + \frac{\beta}{2} (\operatorname{div} \vec{u})^2 + g(\vec{u}, \lambda) \right] dx \quad \text{for vector fields.} \quad (7.6.10)$$

Here $\alpha > 0$ and β are coefficients, and $g(\vec{u}, \lambda)$ is an invariant with respect to \vec{u} . We see from (7.6.9) and (7.6.10) that $SO(n)$ invariance has already determined the main components of the potential, and what it remains is to determine $g(u, \lambda)$ and $g(\vec{u}, \lambda)$.

(2) $SO(3)$ Spinors

In quantum thermodynamical systems, we need to deal with quantum fields where the complex-valued wave functions are order parameters. In this case, the invariance is no longer described by tensors, rather by spinors. Now we introduce the concept of spinors, which are crucial for the mathematical foundation of the theory of condensates.

Let $\psi = (\psi_1, \dots, \psi_N)$ be a set of complex valued wave functions:

$$\psi : \mathbb{R}^3 \rightarrow \mathbb{C}^N. \quad (7.6.11)$$

Under an orthogonal coordinate transformation in \mathbb{R}^3 :

$$\tilde{x} = Ax \quad \text{for } A \in SO(3), \quad (7.6.12)$$

the wave functions ψ in (7.6.11) obey the spinor transformation defined as follows. For each orthogonal matrix A as in (7.6.12), there is a complex unitary matrix $U(A)$ such that

$$\tilde{\psi} = U(A)\psi, \quad U(A) \in SU(N). \quad (7.6.13)$$

Here $SU(N)$ is defined by

$$SU(N) = \{ N\text{-th order complex matrices } U \mid U^\dagger = U^{-1}, \det U = 1 \},$$

where $U^\dagger = (U^T)^*$ is the transpose of complex conjugate of U . We have then the following definition.

Definition 7.6.2 *For the wave functions ψ given by (7.6.11), if under the orthogonal transformation (7.6.12), there is a group homeomorphism*

$$U : SO(3) \rightarrow SU(N), \quad (7.6.14)$$

such that for each A in (7.6.12), ψ transform according to (7.6.13), then ψ is called an $SO(3)$ spinor. The homomorphism (7.6.14) is called an $SO(3)$ spinor representation of ψ .

In (7.6.14), the $SU(N)$ is required to ensure the normality of ψ :

$$|\tilde{\psi}|^2 = |\psi|^2.$$

(3) $SO(3)$ Spinor Representations

To derive the covariance of quantum field equations and the invariance of the Hamiltonian energy, we now determine the specific expression of the spinor representation $U(A)$ in (7.6.14) with respect to $A \in SO(3)$. We proceed as follows.

1. *Euler Representation of $SO(3)$* Consider an orthogonal matrix $A \in SO(3)$, and the corresponding transformation

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} = A \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (7.6.15)$$

For the Euler representation, we need to express A in terms of the three Euler angles (ϕ, θ, ψ) . The above transformation is the composition of the following rotations. First is the counterclockwise rotation with angle ϕ and with respect to the z -axis of the coordinate system (x, y, z) :

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = A_1(\phi) \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (7.6.16)$$

Then rotate the (x_1, y_1, z_1) system with respect to the x_1 -axis and with angle θ :

$$\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = A_2(\theta) \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}. \quad (7.6.17)$$

The final step is to rotate the (x_2, y_2, z_2) system around z_2 with angle ψ :

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} = \begin{pmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = A_1(\psi) \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}. \quad (7.6.18)$$

The combined transformation (7.6.16)–(7.6.18) leads to the equivalent form of the transformation (7.6.15):

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} = A_1(\psi)A_2(\theta)A_1(\phi) \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

where $A = A_1(\psi)A_2(\theta)A_1(\phi)$ is the matrix given in terms of (ϕ, θ, ψ) as

$$\begin{aligned} A &= \begin{pmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} f_{11}(\phi, \theta, \psi) & f_{12}(\phi, \theta, \psi) & \sin \psi \sin \theta \\ f_{21}(\phi, \theta, \psi) & f_{22}(\phi, \theta, \psi) & \cos \psi \sin \theta \\ \sin \theta \sin \phi & -\sin \theta \cos \phi & \cos \theta \end{pmatrix}, \end{aligned} \quad (7.6.19)$$

where

$$\begin{aligned} f_{11} &= \cos \phi \cos \psi - \cos \theta \sin \phi \sin \psi, \\ f_{12} &= \sin \phi \cos \psi + \cos \theta \cos \phi \sin \psi, \\ f_{21} &= -\cos \phi \sin \psi - \cos \theta \sin \phi \cos \psi, \\ f_{22} &= -\sin \phi \sin \psi + \cos \theta \cos \phi \cos \psi. \end{aligned}$$

The Euler angles (ϕ, θ, ψ) are independent, and determine all the matrices in $SO(3)$. Hence (7.6.19) is the Euler angle representation of $SO(3)$. Also, we have

$$0 \leq \phi \leq 2\pi, \quad 0 \leq \theta \leq \pi, \quad 0 \leq \psi \leq 2\pi.$$

2. $SO(3)$ Spinor Representation with $N = 2$

In this case, the spinor representation

$$U : SO(3) \rightarrow SU(2) \quad (7.6.20)$$

is a two-copy representation. Namely, for each $A \in SO(3)$ given in terms of the Euler angles (ϕ, θ, ψ) , the representation $U(A)$ in (7.6.20) has two-copies U_1, U_2 :

$$U_1 = \begin{pmatrix} e^{i(\psi+\phi)/2} \cos \frac{\theta}{2} & ie^{i(\psi-\phi)/2} \sin \frac{\theta}{2} \\ ie^{-i(\psi-\phi)/2} \sin \frac{\theta}{2} & e^{-i(\psi+\phi)/2} \cos \frac{\theta}{2} \end{pmatrix}, \quad (7.6.21)$$

$$U_2 = \begin{pmatrix} e^{-i(\psi+\phi)/2} \cos \frac{\theta}{2} & -ie^{i(\psi-\phi)/2} \sin \frac{\theta}{2} \\ -ie^{-i(\psi-\phi)/2} \sin \frac{\theta}{2} & e^{i(\psi+\phi)/2} \cos \frac{\theta}{2} \end{pmatrix}. \quad (7.6.22)$$

With these two copies, there are two types of spinors for the two-component wave function $\psi = (\psi_1, \psi_2)$: Under the coordinate transformation (7.6.12),

$$\begin{array}{lll} \text{Type-I spinor :} & \tilde{\psi} = U_1 \psi & U_1 \text{ is as (7.6.21),} \\ \text{Type-II spinor :} & \tilde{\psi} = U_2 \psi & U_2 \text{ is as (7.6.22).} \end{array}$$

3. *SO(3) Spinor Representation with $N = 3$* The spinor representation with $N = 3$ corresponds to the Bose–Einstein condensation problem with $J = 1$. For a Bose system with $J = 1$, there are three quantum states: $m = 1, 0, -1$, with m being the magnetic quantum number. The state of the system is described by a three-component wave function:

$$\psi = (\psi_1, \psi_0, \psi_{-1}),$$

where the subindices represent the magnetic quantum number $m = 1, 0, -1$. The corresponding $SO(3)$ spinor representation with $N = 3$

$$U : SO(3) \rightarrow SU(3) \quad (7.6.23)$$

is defined by

$$U(A) = M^\dagger A M \quad \forall A \in SO(3), \quad (7.6.24)$$

where M is given by

$$M = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 0 & 1 \\ -i & 0 & -i \\ 0 & \sqrt{2} & 0 \end{pmatrix}. \quad (7.6.25)$$

When $A \in SO(3)$ is given in terms of the Euler angles by (7.6.19):

$$A = A_1(\psi) A_2(\theta) A_1(\phi),$$

the corresponding $U(A)$ in (7.6.24) can be written as

$$U(A) = U_1(\psi) U_2(\theta) U_1(\phi), \quad (7.6.26)$$

where

$$\begin{aligned} U_1(\phi) &= M^\dagger A_1(\phi) M = \begin{pmatrix} e^{i\phi} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{-i\phi} \end{pmatrix}, \\ U_2(\theta) &= M^\dagger A_2(\phi) M = \frac{1}{2} \begin{pmatrix} 1 + \cos \theta & i\sqrt{2} \sin \theta & -1 + \cos \theta \\ i\sqrt{2} \sin \theta & 2 \cos \theta & i\sqrt{2} \sin \theta \\ -1 + \cos \theta & i\sqrt{2} \sin \theta & 1 + \cos \theta \end{pmatrix}. \end{aligned}$$

Consequently

$$\begin{aligned} U(A) &= U(\phi, \theta, \psi) \\ &= \begin{pmatrix} e^{i(\phi+\psi)}(1 + \cos \theta) & i\sqrt{2}e^{i\psi} \sin \theta & -e^{i(\psi-\phi)}(1 - \cos \theta) \\ i\sqrt{2} \sin \theta e^{i\phi} & 2 \cos \theta & i\sqrt{2}e^{-i\phi} \sin \theta \\ -e^{i(\phi-\psi)}(1 - \cos \theta) & i\sqrt{2}e^{-i\psi} \sin \theta & e^{-i(\phi+\psi)}(1 + \cos \theta) \end{pmatrix}. \end{aligned} \quad (7.6.27)$$

7.6.2 PVT Systems

The main objective of this section is to establish the thermodynamic potential of PVT (physical-vapor transport) systems. First we recall the van der Waals equation, which is the equation of state for gaseous PVT systems:

$$(bp + RT)\rho - a\rho^2 + ab\rho^3 - p = 0, \quad (7.6.28)$$

where a, b are van der Waals constants, $\rho = n/V$ is the mole density, R is the gas constant, T is the temperature, p is the pressure, and n is the mole number.

The relation between the thermodynamic potential F and the equation of state is given by

$$\delta F = 0 \Rightarrow \text{the equation of state.} \quad (7.6.29)$$

Hence the van der Waals equation (7.6.28) serves as an important reference for the following construction of the potential functional for PVT systems.

1. *Basic Form of the Thermodynamic Potential F* Since a PVT system is an isothermal-isopiestic process, and therefore a Gibbs system with

$$\text{order parameters} = (\rho, S), \quad \text{control parameters} \lambda = (T, p), \quad (7.6.30)$$

where ρ is the density, and S is the entropy density. By (7.1.8) and the general form of the Gibbs free energy, the potential functional for the PVT system is given by

$$F = \int_{\Omega} \left[\frac{\alpha}{2} |\nabla \rho|^2 + f(\rho, S, \lambda) - \mu \rho - ST - \frac{1}{\rho_0} p \rho \right] dx, \quad (7.6.31)$$

where $\alpha > 0$ is a constant, ρ_0 is the reference value of ρ , μ is the chemical potential as in (7.1.8) due to the conservation of the number of particles, and λ is as in (7.6.30).

By the statistical theory of heat presented in the previous section, entropy represents the number of photons, and by the vibratory mechanism of thermal radiation, the entropy density depends on the molar density ρ . Notice that $|\nabla \rho|^2$ represents the energy for maintaining the non-homogeneity of the particle distribution of the system, and that the nonhomogeneous distribution of entropy is due to nonhomogeneous distribution of particles. Therefore there is no $|\nabla S|^2$ term in the potential functionals for all thermodynamic systems, which is reflected by the term $|\nabla \rho|^2$.

2. *Physical Meaning of ρ^k* To derive an explicit form of $f(\rho, S, \lambda)$ in (7.6.31), we need to examine the physical meaning of ρ^k . In statistical physics, for $k \geq 2$,

$$\rho^k(x) = \text{probability density of collisions of } k \text{ particles at } x. \quad (7.6.32)$$

Also, ρ^2 , we have the following additional physical meaning:

$A\rho^2(x)$ represents the interaction potential energy density of particles at x , with $A > 0$ for repulsive interaction, and $A < 0$ for attractive interaction. (7.6.33)

Since the probability for more than $k \geq 3$ particles collide with each other at the same time and location is negligible, there is no ρ^k ($k \geq 3$) terms in $f(\rho, S, \lambda)$, excluding the Taylor expansion with respect to ρ .

In the free energy (7.6.31), f contains two parts: the energy corresponding to molar density ρ , denoted by $f_1(\rho, \lambda)$, and the energy corresponding to entropy density S , denoted by $f_2(S, \rho, \lambda)$:

$$\begin{aligned} f &= f_1(\rho, \lambda) + f_2(S, \rho, \lambda), \\ f_2(0, \rho, \lambda) &= 0. \end{aligned} \quad (7.6.34)$$

3. *Expression for $f_1(\rho, \lambda)$* By (7.6.33), f_1 contains a ρ^2 term, so that

$$f_1 = \frac{1}{2}A_1(\lambda)\rho^2 + g_1(\rho, \lambda). \quad (7.6.35)$$

Also as a particle system, the internal energy of a PVT system contains the following term

$$\beta \ln W,$$

where W is the number of states of particles. Hence g_1 is proportional to $\ln W$:

$$g_1 = a(\lambda) \ln W. \quad (7.6.36)$$

Basic thermodynamics shows that

$$W = (\rho_0 + \rho)! \simeq (\rho_0 + \rho)^{(\rho_0 + \rho)}. \quad (7.6.37)$$

where ρ_0 is the reference number density of particles. With the first-order terms of ρ included in the $\mu\rho$ in (7.6.31), we infer from (7.6.36) and (7.6.37) that

$$g_1 = A(\lambda)(1 + \rho/\rho_0) \ln(1 + \rho/\rho_0). \quad (7.6.38)$$

Then by (7.6.35) and (7.6.38), we derive the expression of f_1 as follows:

$$f_1 = \frac{1}{2}A_1\rho^2 + A(1 + \rho/\rho_0) \ln(1 + \rho/\rho_0), \quad (7.6.39)$$

where A and A_1 are system parameters depending on $\lambda = (T, p)$.

For van der Waals gas, we can use the van der Waals equation (7.6.28) and (7.6.29) to derive the following relation for the parameters A and A_1 in (7.6.39):

$$A = 2a/b^2, \quad A_1 = b^2 p + A_0 RT b, \quad (7.6.40)$$

where $\rho_0 = 1/b$, a , b are the van der Waals constants, and A_0 is a nondimensional parameter.

4. *Expression of $f_2(S, \rho, \lambda)$* By the condition (7.2.11) for the Legendre transformation, we have

$$\frac{\partial^2}{\partial S^2} f_2(0, \rho, \lambda) \neq 0. \quad (7.6.41)$$

Namely, there is an S^2 in the expression of f_2 . The statistical theory of heat indicates that the energy density of entropy S is zero when $T = 0$. Hence f_2 takes the form

$$f_2 = T \left(\frac{\beta}{2} S^2 + S g_2(\rho, \lambda) \right), \quad (7.6.42)$$

where β is a parameter with dimension $1/k\rho_0$, and g_2 is a to-be-determined function.

By the photon number entropy formula (7.5.43), S represents the photon number density in the system, and therefore the physical meaning similar to (7.6.33) for S holds true as well. Also, each photon carries two weak charges, and there is a weak interaction between photons. In particular, there is a radius $r_0 > 0$ such that

$$\text{interaction between photons} = \begin{cases} \text{attractive} & \text{for } r < r_0, \\ \text{repulsive} & \text{for } r > r_0. \end{cases} \quad (7.6.43)$$

By (7.6.33), we derive from the above properties of the weak interaction between photons that for f_2 , the term involving S^2 is given by ($\beta = -B$):

$$-\frac{1}{2} B TS(S - 2S_0), \quad (7.6.44)$$

where $B = B(\lambda) > 0$ is a parameter depending on λ , and $2S_0$ is the reference entropy density. By (7.6.43), the physical meaning of (7.6.44) is as follows:

the interaction between photons is attractive	for $S > 2S_0$,
the interaction between photons is repulsive	for $S < 2S_0$.

Also by (7.6.32), ρ^2 represents collision probability density of two particles. By the absorption and radiation mechanism of photons (Ma and Wang, 2017d), S is proportional to ρ^2 . Therefore in the internal energy, the coupling between S and ρ involves only the product of S and ρ^2 . Hence only the following appears as the coupling term in (7.6.42):

$$B_1 TS \rho^2$$

which represents thermal energy density due to the collision of particles. Finally by (7.6.44), the expression (7.6.42) for f_2 is

$$f_2 = BT \left(-\frac{1}{2}S^2 + S_0S \right) + B_1TS\rho^2, \quad (7.6.45)$$

where B, B_1 are system parameters depending on $\lambda = (T, p)$.

5. *Gibbs Free Energy for PVT Systems* By (7.6.39) and (7.6.45), we derive from (7.6.34) the expression for f :

$$\begin{aligned} f = & \frac{1}{2}A_1(T, p)\rho^2 + A(T, p)(1 + \rho/\rho_0)\ln(1 + \rho/\rho_0) \\ & + BT \left(-\frac{1}{2}S^2 + S_0S \right) + B_1TS\rho^2. \end{aligned} \quad (7.6.46)$$

In summary, we derive from (7.6.31) and (7.6.46) the general form of thermodynamic potentials for a general (gaseous, or liquid or solid) PVT system:

$$\begin{aligned} F = & \int_{\Omega} \left[\frac{\alpha}{2}|\nabla\rho|^2 + A(1 + \rho/\rho_0)\ln(1 + \rho/\rho_0) \right. \\ & + \frac{1}{2}A_1\rho^2 - \mu\rho - p\rho/\rho_0 \\ & \left. + BT \left(-\frac{1}{2}S^2 + S_0S \right) + B_1TS\rho^2 - ST \right] dx, \end{aligned} \quad (7.6.47)$$

where A, A_1, B, B_1 are system parameters depending on the control parameters T, p . These parameters are system dependent.

6. *Gibbs Free Energy for Gaseous PVT Systems* In this case, (7.6.47) can be written as

$$\begin{aligned} F = & \int_{\Omega} \left[\frac{\alpha}{2}|\nabla\rho|^2 + \frac{ART}{b}(1 + b\rho)\ln(1 + b\rho) \right. \\ & + \frac{1}{2}(b^2p + A_0bRT)\rho^2 - \mu\rho - bp\rho \\ & \left. + BT \left(-\frac{1}{2}S^2 + S_0S \right) + B_1TS\rho^2 - ST \right] dx, \end{aligned} \quad (7.6.48)$$

where A, A_0 are nondimensional parameters and b is the van der Waals constant.

7. *Relation Between Entropy Density S and (ρ, T, p)* By (7.6.47), we have

$$S = B_0\rho^2 + S_0 - \frac{1}{B} \quad (B_0 = B_1/B).$$

This shows that the entropy density (photon number density) S is proportional to ρ^2 , in agreement with physical facts.

7.6.3 N-Component Systems

An N -component system consists of N (≥ 2) different types of particles. The thermodynamical potential functional of an N -component system is either the Helmholtz free energy if the pressure is negligible, or the Gibbs free energy if the pressure is not negligible. We consider only the latter case.

The most important property of an N -component system is the phase separation phenomena as the temperature is lower than a critical value, leading to nonuniform distribution of different type of particles. At the phase separation, the total number density of particles can either be nonuniformly distributed in space, or uniformly distributed. We discuss both scenarios in turn below.

1. *Nonuniform Phase Separation Systems* In this case, the distribution of the total number density of the particles is nonuniformly distributed in space, so that the number densities of different types of particles are independent.

For an N -component system, let A_1, \dots, A_N ($N \geq 2$) be the different types of particles, and u_i be the number density of A_i particles. Then the order parameters of the system are entropy density S and $u = (u_1, \dots, u_N)$. The control parameters are $\lambda = T, p$, the length scale L , and number ratio x_i of A_i . Here the number ratio of A_i is defined by

$$x_i = \frac{1}{N_0} \int_{\Omega} u_i dx, \quad (7.6.49)$$

where N_0 is the total number of particles.

By the Fick Law of diffusion law and the Onsager reciprocal relations, the fluxes $J = (J_1, \dots, J_N)$ of the N -types of particles are given in terms of $(\nabla u_1, \dots, \nabla u_N)$ as follows:

$$J_i = -L_{ij} \nabla u_j, \quad (7.6.50)$$

where the summation convention of repeated indices are used, (L_{ij}) is a positive definite and symmetric matrix, called the Onsager coefficients. The potential energy induced by the particle flux J is $-\frac{1}{2} J \cdot \nabla u$:

$$\text{potential energy induced by } J = -\frac{1}{2} J \cdot \nabla u = \frac{1}{2} L_{ij} \nabla u_i \nabla u_j. \quad (7.6.51)$$

Hence the general form of the thermodynamic potential functional for an N -component system is

$$F = \int_{\Omega} \left[\frac{1}{2} L_{ij} \nabla u_i \nabla u_j + f(u, S, \lambda) - \mu_i u_i - ST - bp \sum_{i=1}^N u_i \right] dx, \quad (7.6.52)$$

where b is the van der Waals constant, and μ_i are the chemical potential of A_i . Since the number of particles is conserved, we have

$$\int_{\Omega} u_i dx = \text{constant} \quad \text{for } 1 \leq i \leq N. \quad (7.6.53)$$

We need to derive the expression of $f(S, u, \lambda)$. As for the PVT system, f consists of parts involving u and S :

$$\begin{aligned} f &= f_1(u, \lambda) + f_2(S, u, \lambda), \\ f_2(S, u, \lambda)|_{S=0} &= 0. \end{aligned} \quad (7.6.54)$$

By the Flory–Huggins theory (Reichl, 1998), f_1 takes the form

$$f_1 = AkTu_j \ln u_j + A_{ij}u_iu_j, \quad (7.6.55)$$

where (A_{ij}) is a symmetric matrix, depending on (T, p) and the type of particles, A_{ij} represent the interaction strength of particles A_i and A_j , $A_{ij} > 0$ stand for repulsive force, $A_{ij} < 0$ stand for attractive force, $A > 0$ is a nondimensional parameter, and k is the Boltzmann constant.

As for the PVT systems, f_2 can be expressed as

$$f_2 = -\frac{B_0bT}{2k}S^2 + b^2TSB_{ij}u_iu_j + \frac{B_0bT}{k}S_0S, \quad (7.6.56)$$

where B_{ij} , B_0 are nondimensional parameters, depending on (p, T) , (B_{ij}) are positive definite, S_0 is the physical unit of entropy, and $b^2TSB_{ij}u_iu_j$ represents the thermal energy due to the collision of particles.

Hence by (7.6.54)–(7.6.56), the potential functional (7.6.52) is written as

$$\begin{aligned} F = \int_{\Omega} \left[\frac{1}{2}L_{ij}\nabla u_i \nabla u_j + AkTu_i \ln u_i + A_{ij}u_iu_j \right. \\ \left. - \mu_iu_i - bp\rho - \frac{B_0bT}{2k}(S - 2S_0)S + b^2TSB_{ij}u_iu_j - ST \right] dx, \end{aligned} \quad (7.6.57)$$

where $\rho = \sum_j u_j$ is the total particle number density.

2. *Uniform Phase Separation Systems* In this case, when the phase separation occurs, the distributions of different types of parts are in general nonuniform, but the total number density of particles is uniformly distributed. In other words, the number densities of the N types of particles are not independent.

As in the previous case, consider an N -component system, with A_1, \dots, A_N ($N \geq 2$) being the types of particles, and u_i be the number density of A_i particles. By assumption, the total number density u_0 is a constant:

$$u_1 + \dots + u_N = u_0 \text{ (a constant).}$$

Hence there are only $N - 1$ independent variables among (u_1, \dots, u_N) . Let

$$\begin{aligned} \text{order parameters} &= \{S, u\}, \quad u = (u_1, \dots, u_{N-1}) \\ \text{control parameters} &= \{T, p, L, x_i\}, \end{aligned} \quad (7.6.58)$$

where x_i is the molar ratio for the i -th type particles, and u_N is given by

$$u_N = u_0 - \sum_{i=1}^{N-1} u_i. \quad (7.6.59)$$

With u_N in (7.6.57) given by (7.6.59), we deduce the potential functional for the uniform phase separation of an N -component system:

$$\begin{aligned} F = \int_{\Omega} & \left[l_{ij} \nabla u_i \nabla u_j + AkTu_i \ln u_i + a_{ij}u_i u_j \right. \\ & + AkT \left(1 - \sum_{i=1}^{N-1} u_i \right) \ln \left(1 - \sum_{i=1}^{N-1} u_i \right) - \mu_i u_i \\ & \left. - \frac{B_0 b T}{2k} (S - S_0) S + b^2 T S b_{ij} u_i u_j - bp \sum_{i=1}^{N-1} u_i - ST \right] dx, \end{aligned} \quad (7.6.60)$$

where some rearrangements are made: the constant term is ignored, the linear terms of u_i are combined with $\mu_i u_i$, and (l_{ij}) and (b_{ij}) are $(N-1)$ -th order positive definite and symmetric matrices.

Also we note that the parameters l_{ij} , a_{ij} , b_{ij} in (7.6.60) can be directly experimentally measured, or determined by physical laws. (l_{ij}) are the Onsager coefficients. The Onsager coefficients L_{ij} are for nonuniform phase separation systems, and (l_{ij}) are coefficients for uniform phase separation systems. The positive definiteness of these two matrices is the consequence of the well-posedness and dynamical stability of the systems.

7.6.4 Magnetic and Dielectric Systems

(1) Magnetic Systems

There are three types of magnetic systems: paramagnetic, ferromagnetic, and antiferromagnetic. A paramagnetic material can be magnetized under an external magnetic field, and a ferromagnetic material can form permanent magnetic field below the Curie temperature without the presence of the external magnetic field. An antiferromagnetic material is paramagnetic above the Neel temperature, and diamagnetic below the Neel temperature. Their equation of state is the Curie–Weiss law:

$$\chi = \frac{\partial M}{\partial H} = \frac{c}{T + T_0}, \quad T_0 \begin{cases} < 0 & \text{for ferromagnetic,} \\ = 0 & \text{for paramagnetic,} \\ > 0 & \text{for antiferromagnetic,} \end{cases} \quad (7.6.61)$$

where H is the external magnetic field, M is the magnetization (magnetic moment per unit volume), χ is the magnetic susceptibility, and T_0 is the fixed temperature (Weiss constant).

1. *Order Parameters* The physical meaning of the magnetization \mathbf{M} is as follows. In a ferromagnetic body, magnetic moment can be induced by the spin of an electron, or by the orbital rotation of an electron with respect to axis with direction n , as shown in Fig. 7.5:

Magnetic moments aligned in parallel induce a magnetization \mathbf{M} defined by

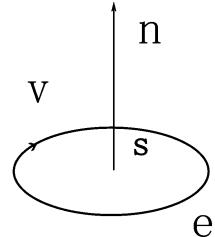
$$\mathbf{M} = \frac{1}{V} \sum m_i,$$

where $\sum m_i$ is the sum of all magnetic moments m_i in a region D , V is the volume of D , i.e., \mathbf{M} is the magnetic moment per unit volume.

As the coupling between \mathbf{M} and the number density ρ of particles in the system is weak, we can take the order parameters u and control parameters λ as follows:

$$u = (\mathbf{M}, S), \quad \lambda = (T, \mathbf{H}, L),$$

Fig. 7.5 An electron rotating around a direction n with velocity v induces a magnetic moment $m = evs$, where s is the area vector enclosed by the electron orbit



where \mathbf{H} is the applied magnetic field, and L is the length scale of the magnetic body.

2. *Thermodynamic Potential for Magnetic Systems* The general form of the Gibbs free energy of a magnetic system is given by

$$F = F_0 + \int_{\Omega} \left[\frac{\mu}{2} |\nabla \mathbf{M}|^2 + f(\mathbf{M}, S, \lambda) - ST - \mathbf{M} \cdot \mathbf{H} \right] dx, \quad (7.6.62)$$

where $\mu > 0$ is a coefficient. The density f in (7.6.62) can be decomposed into two parts:

$$f = f_1(M, \lambda) + f_2(S, M, \lambda). \quad (7.6.63)$$

Since the energy density of the magnetic field is a quadratic form of M , f_1 takes the form:

$$f_1 = \frac{1}{2} a_{ij} M_i M_j, \quad (7.6.64)$$

where (a_{ij}) is a second-order tensor, called magnetization matrix, reflecting the property of the material:

$$A = (a_{ij}) \text{ is a } 3 \times 3 \text{ symmetric matrix.} \quad (7.6.65)$$

There are three eigenvalues of this matrix: λ_k ($1 \leq k \leq 3$). By the equation of state (7.6.61), we derive the following relations between λ_k and the properties of

the magnetic materials:

$$\begin{aligned}\lambda_k < 0 \quad (1 \leq k \leq 3) &\quad \text{for ferromagnetic,} \\ \lambda_k = 0 \quad (1 \leq k \leq 3) &\quad \text{for paramagnetic,} \\ \lambda_k > 0 \quad (1 \leq k \leq 3) &\quad \text{for antiferromagnetic.}\end{aligned}\tag{7.6.66}$$

Also, $\lambda_k (1 \leq k \leq 3)$ characterizes the orientation of the materials:

$$\begin{aligned}\lambda_1 = \lambda_2 = \lambda_3 &\iff \text{isotropic,} \\ \text{otherwise} &\iff \text{anisotropic.}\end{aligned}\tag{7.6.67}$$

For f_2 in (7.6.63), notice that the entropy S coupling with M is not the total entropy for the system, S can take negative values. As for PVT systems, in view of (7.6.61), we derive the expression for f_2 as

$$f_2 = -\frac{T}{2\alpha}S^2 - \beta TS|\mathbf{M}|^2,\tag{7.6.68}$$

where $\alpha, \beta > 0$ are parameters. The second term in the right-hand side stands for loss of thermal energy caused by magnetization, indicating that thermal energy prevents magnetization.

By (7.6.64) and (7.6.68), the Gibbs free energy (7.6.62) for a magnetic system can be written as

$$F = F_0 + \int_{\Omega} \left[\frac{\mu}{2} |\nabla \mathbf{M}|^2 + \frac{a_{ij}}{2} M_i M_j - \mathbf{M} \cdot \mathbf{H} - \frac{T}{2\alpha} S^2 - \beta TS|\mathbf{M}|^2 - ST \right] dx.\tag{7.6.69}$$

In (7.6.69), there are six parameters in the symmetric matrix $A = (a_{ij})$. However, there are only three independent parameters, which are the three eigenvalues $\lambda_k (1 \leq k \leq 3)$ of A . Take the three eigenvectors (e_1, e_2, e_3) , as the basis, A is given by

$$A = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_3 \end{pmatrix}.$$

Then (7.6.69) is written as

$$F = F_0 + \int_{\Omega} \left[\frac{\mu}{2} |\nabla \mathbf{M}|^2 + \frac{1}{2} \lambda_k M_k^2 - M_k H_k - \frac{T}{2\alpha} S^2 - \beta TS|\mathbf{M}|^2 - ST \right] dx.\tag{7.6.70}$$

(2) Dielectric Systems

Dielectric systems can also be classified into three types: paraelectric, ferroelectric, and anti-ferroelectric. Paraelectricity is the ability of materials to become polarized under an applied electric field. Ferroelectricity is a characteristic of materials that

have a spontaneous electric polarization without an applied electric field. An antiferroelectric state is defined as one in which lines of ions in the crystal are spontaneously polarized, but with neighboring lines polarized in antiparallel directions.

In a dielectric body, for lattice made by charged atoms and molecules, the nonuniform distribution of positive and negative charges induce electric dipoles, leading to polarization phenomena. The polarization dipole moment of a single electric dipole is $p = qr$, where q is the charge of the electric dipole, and r is the distance vector of the pair of electric charges. When the temperature is lower than the Curie point, or under an applied electric field E , the electric dipoles in a dielectric body align in parallel in the same direction, leading to polarization. Such small region is called electric domain. Its electric polarization strength is defined by $P = (\sum p_i)/V$, which stands for electric dipole moment per unit volume. The order parameters u and control parameters λ are

$$u = (P, S), \quad \lambda = (T, E, L),$$

where E is the applied electric field.

As in (7.6.61), the equation of state for the dielectric system is

$$\chi = \frac{\partial P}{\partial E} = a + \frac{b}{T + T_0}, \quad (7.6.71)$$

where $a, b > 0$ are constants and

$$T_0 \begin{cases} < 0 & \text{for ferroelectric,} \\ = 0 & \text{for paraelectric,} \\ > 0 & \text{for anti-ferroelectric.} \end{cases} \quad (7.6.72)$$

As the thermodynamic potential for the magnetic systems (7.6.70), by (7.6.71) and (7.6.72), the thermodynamic potential functional of a dielectric system takes the form:

$$F = F_0 + \int_{\Omega} \left[\frac{\mu}{2} |\nabla P|^2 + \frac{\lambda_k p_k^2}{2} - (1 + \eta T) P \cdot E - \frac{TS^2}{2\alpha} - \beta TS|P|^2 - ST \right] dx. \quad (7.6.73)$$

where $\mu, \eta, \alpha, \beta > 0$ are coefficients, and

$$\begin{cases} \lambda_k < 0 (1 \leq k \leq 3) & \text{for ferroelectric,} \\ \lambda_k = 0 (1 \leq k \leq 3) & \text{for paraelectric,} \\ \lambda_k > 0 (1 \leq k \leq 3) & \text{for anti-ferroelectric,} \end{cases} \quad (7.6.74)$$

$$\begin{cases} \lambda_1 = \lambda_2 = \lambda_3 \Leftrightarrow \text{isotropic,} \\ \lambda_1 \neq \text{one of } \lambda_j \Leftrightarrow \text{anisotropic.} \end{cases} \quad (7.6.75)$$



Chapter 8

Quantum Mechanism of Condensates and High Tc Superconductivity

The main objective of this chapter is to introduce a new quantum mechanism of condensates and superconductivity based on the field theoretical interpretation of quantum mechanical wave functions presented in Sect. 7.3.2, and on recent developments in quantum physics and statistical physics. The quantum mechanism presented serves as a unified physical picture and formation mechanism for superconductivity, superfluidity, and Bose–Einstein condensation. The study presented in this chapter was first obtained by Ma and Wang (2019).

8.1 Quantum Mechanism of Condensation

8.1.1 *Quantum Mechanism of Condensates*

1. *Physical Picture of the Formation of Condensates* We start with the mechanism of the formation of condensates. The new interpretation of quantum mechanics in Sect. 7.3.2 shows that the wave function is no longer the state function of a specific particle, and is now considered as the field function governing the motion of particles. The motion of particles in the field of a wave function is similar to the motion of macroscopic objects in a gravitational field. This new viewpoint enables us to describe precisely the physical picture of condensates.

By the Pauli exclusion principle, under the condition of very weak interactions between particles, a large collection of Bose particles can undergo a motion governed by the same wave function field, as long as all the particles are in the same external potential field. Hence the physical picture of a condensate is as follows:

a condensate is a large collection of Bosons in the same wave function field ψ determined by the same external potential field V . (8.1.1)

We call ψ the condensation wave function or the condensation field.

2. Microscopic Quantum Mechanism Condensation occurs at low temperatures.

The main reason is that at higher temperature, particles possess relatively high kinetic energy, causing them to collide (scattering) with each other. Collisions are interactions. Hence under such collisions, each particle is situated at an independent interaction field with its own wave function, and consequently the condensation picture (8.1.1) does not happen.

However, at a relative low temperature, the kinetic energy of particles decreases since temperature is the average energy level of particles. Hence the possibility of collision decreases, leading to the formation of a common external potential field. Also, since particles with the same wave function field do not collide, the system energy decreases. By the minimum potential principle, the system has the tendency toward condensation of particles. This is the mechanism of the physical condensation depicted in (8.1.1).

3. Collisionless Since the large collection of particles share a common wave function field $\psi = |\psi|e^{i\varphi}$, their motion is dictated by the velocity field $\frac{\hbar}{m}\nabla\varphi$; see (7.3.19):

$$\begin{aligned} \frac{dx}{dt} &= \frac{\hbar}{m}\nabla\varphi(x), \\ x(0) &= x_0, \end{aligned} \tag{8.1.2}$$

For different particles, their initial positions are different so that they will not collide with each other. For example, consider two different particles *A* and *B* with initial positions at x_1 and x_2 , i.e.

$$x_A(0, x_1) = x_1, \quad x_B(0, x_2) = x_2,$$

where $x_A(t, x_1)$ and $x_B(t, x_2)$ are the trajectories of particles *A* and *B*, which are solutions of

$$\frac{dx}{dt} = \frac{\hbar}{m}\nabla\varphi(x)$$

with initial data x_1 and x_2 . It is then clear that if $x_1 \neq x_2$, then

$$x_A(t, x_1) \neq x_B(t, x_2), \quad \forall t > 0.$$

In other words, particles *A* and *B* can never collide with each other. The motion of particles in a condensate resembles the motion of planets around the nucleus of a galaxy.

4. Minimal Energy For N particles on a state described by one wave function field ψ_0 , by (7.3.21), their total energy is given by

$$NE_0 = N \int \left[\frac{\hbar^2}{2m}|\nabla|\psi_0||^2 + \frac{\hbar^2}{2m}|\nabla\varphi_0|^2 + V_0|\psi_0|^2 \right] dx.$$

On the other hand, for N particles such that each particle occupies its own state described by its own wave function ψ_i , their total energy is

$$\sum_{i=1}^N E_i = \sum_{i=1}^N \int \left[\frac{\hbar^2}{2m} |\nabla|\psi_i||^2 + \frac{\hbar^2}{2m} |\nabla\varphi_i|^2 + V_i |\psi_i|^2 \right] dx.$$

When condensation occurs, the particles collectively occupy the minimal energy state, so that

$$E_0 \leq E_i \quad (1 \leq i \leq N).$$

Hence the total energy of the system satisfies:

$$NE_0 < \sum_{i=1}^N E_i \quad (\text{if there is an } E_i \text{ such that } E_i > E_0).$$

Namely, when the condensate particles occupy the minimal energy levels, the total energy of the system is less than the total energy of the particles in the normal states. By the potential-descending principle (minimal potential principle), Principle 7.1.2, the system will be in a condensate state.

5. *Condensation Mechanism of Superconductivity and Superfluidity* In summary, we have derived the following condensation formation mechanism. This is the mechanism of superconductivity, superfluidity, and BEC.

Condensation Mechanism 8.1.1

1. *The condition for condensation of bosonic particles is that their mutual interaction is sufficiently weak to ensure that they are in a state governed by the same wave function field under the common bounding potential V ;*
2. *a large collection of bosonic particles occupy the same wave function ψ , called condensation wave function, that is determined by the wave equation under the same potential V ;*
3. *the condensation wave function ψ is the smallest allowed energy state of the wave equation; and*
4. *the motion of particles in a condensate system under the same wave function with velocity field $\frac{\hbar}{m}\nabla\varphi$ resembles the motion of stars in a galaxy, and can last permanently.*

8.1.2 Field Equations of Condensates

There are two cases of condensates: The first is the case where the system is at near the critical temperature T_c , the condensation is in its early stage, and the condensed particle density is small. The second is the case where away from the critical temperature, the system enters a deeper condensate state. The field equations for these two cases are different and will be addressed separately below.

1. *Condensation Field Equations Near T_c* At this stage, the field equations are the variational equations of the thermodynamic potential. Let $F(\psi, \Phi)$ be the thermodynamic potential, ψ be the wave function, and $\Phi = (S, \mathbf{H}, \mathbf{E})$, where S

is the entropy, \mathbf{H} is the electromagnetic field, and E is the electric field. Then the field equations are the equilibrium equations of the thermodynamic phase transition:

$$\frac{\delta}{\delta \psi^*} F(\psi, \Phi) = 0, \quad \frac{\delta}{\delta \Phi} F(\psi, \Phi) = 0. \quad (8.1.3)$$

2. *Fully Developed Condensation Field Equations* For the fully developed condensation away from critical temperature, the system obeys both the principle of Lagrangian dynamics and the principle of Hamiltonian dynamics, leading to two equivalent field equations.

Let H be the Hamiltonian energy functional. Then the Lagrangian dynamical equation is given by

$$i\hbar \frac{\partial \Psi}{\partial t} = \frac{\delta H(\Psi)}{\delta \Psi^*}. \quad (8.1.4)$$

Also, the Hamiltonian dynamic equations are

$$\begin{aligned} 2\hbar \frac{\partial \Psi_1}{\partial t} &= \frac{\delta H(\Psi)}{\delta \Psi_2}, \\ 2\hbar \frac{\partial \Psi_2}{\partial t} &= -\frac{\delta H(\Psi)}{\delta \Psi_1}, \end{aligned} \quad (8.1.5)$$

where $\Psi = \Psi_1 + i\Psi_2$.

Equations (8.1.4) and (8.1.5) are equivalent, and both will be used in describing different properties of quantum systems of condensates.

3. *Thermodynamic Potentials and Hamiltonian Energy Functionals of Condensates* are certain types of energies and are derived explicitly based on first principles and the underlying symmetries of the condensate systems in Sects. 8.3 and 8.4 hereafter. Consequently the field equations are then explicitly derived in Sect. 8.5 using (8.1.3)–(8.1.5).

8.1.3 Topological Structure Equations of Condensates

Let the wave function ψ of a condensate system be expressed as

$$\psi = \zeta e^{i\varphi}, \quad \zeta = |\psi|. \quad (8.1.6)$$

In Sect. 7.3.2, we have derived the physical meaning of ζ and φ :

$$\begin{aligned} \rho &= \zeta^2 \quad \text{distribution density of particles in the condensate,} \\ \frac{\hbar}{m} \nabla \varphi &\quad \text{the velocity field of the flow of particles.} \end{aligned} \quad (8.1.7)$$

Hence ζ and φ describe the structure of condensates in the physical space, and the equations of ζ and φ are called *topological structure equations* or *pattern formation equations*.

The pattern formation equations for ζ and φ can be derived from the field equations (8.1.3) and (8.1.4) as follows.

First, near the critical temperature, (8.1.3) is the steady-state equation of (8.1.6). Away from the criticality, (8.1.4) are evolution equations. The energy of the Hamiltonian system (8.1.5) is conserved, i.e., the number of particles is conserved. Hence the solution of (8.1.4) can be written as:

$$\Psi = e^{-i\lambda t/\hbar} \psi(x), \quad \psi \text{ is as (8.1.6)}, \quad (8.1.8)$$

where $\lambda \in \mathbb{R}$ represents the chemical potential. By (8.1.8), we infer from (8.1.4) that

$$\lambda\psi = \frac{\delta}{\delta\psi^*} H(\psi). \quad (8.1.9)$$

Hence we reduce (8.1.4) to the steady-state form given by (8.1.9).

1. Gaseous and Liquid Condensate Systems For gaseous and liquid systems, (8.1.3) and (8.1.9) can be written in a unified form as follows:

$$-\frac{\hbar^2}{2m}\Delta\psi + f(|\psi|)\psi = 0, \quad (8.1.10)$$

where $f(x)$ is a function of $x \in \mathbb{R}$. If ψ is a spinor, then $f(x)$ is a matrix-valued function of $x \in \mathbb{R}^3$, and ψ can be written as

$$\psi = \zeta e^{i\varphi}, \quad \zeta = (\zeta_+, \zeta_0, \zeta_-),$$

as the three types of particles in the system possess the same mass, and consequently share the same velocity field $-\frac{\hbar}{m}\nabla\varphi$.

In view of (8.1.6), we infer from (8.1.10) the following pattern formation equation for gaseous and liquid condensate systems:

$$\operatorname{div}(\zeta^2\nabla\varphi) = 0, \quad (8.1.11)$$

$$-\frac{\hbar^2}{2m}[\Delta\zeta - \zeta|\nabla\varphi|^2] + f(\zeta)\zeta = 0. \quad (8.1.12)$$

2. Superconducting Systems For superconducting systems, (8.1.3) and (8.1.9) can be written in the following unified form:

$$\frac{1}{2m_s} \left(i\hbar\nabla + \frac{e_s}{c} A \right) \psi + f(|\psi|)\psi = 0, \quad (8.1.13)$$

with superconducting currents:

$$J_s = -\frac{e_s^2}{m_s c^2} |\psi|^2 A - \frac{e_s \hbar}{2m_s c} i(\psi^* \nabla \psi - \psi \nabla \psi^*).$$

In view of (8.1.6), J_s becomes

$$J_s = -\frac{e_s^2}{m_s c^2} \zeta^2 A + \frac{e_s \hbar}{m_s c} \zeta^2 \nabla \varphi. \quad (8.1.14)$$

Also by (8.1.6), (8.1.13) is rewritten as

$$\operatorname{div} \left(\hbar \zeta^2 \nabla \varphi - \frac{e_s}{c} \zeta^2 A \right) = 0, \quad (8.1.15)$$

$$-\frac{\hbar^2}{2m_s} \left[\Delta \zeta - \zeta |\nabla \varphi|^2 \right] + \frac{e_s^2}{2m_s c^2} A^2 \zeta + f(\zeta) \zeta = 0. \quad (8.1.16)$$

In (8.1.15), we have used the Coulomb gauge

$$\operatorname{div} A = 0.$$

By (8.1.14) and (8.1.15), we obtain that $\operatorname{div} J_s = 0$.

The above two systems of Eqs: (8.1.11)–(8.1.12), and (8.1.15)–(8.1.16) are the pattern formation equations in the physical space for superfluidity, superconductivity, and BEC. These are important equations. On one hand, they dictate two important physical structures of condensate particles:

- *physical picture of the distribution function $\rho = \zeta^2$, and*
 - *topological structure of the flow determined by the velocity field $\frac{\hbar}{m} \nabla \varphi$.*
- (8.1.17)

On the other hand, the topological structure equations are fundamental equations for studying quantum phase transitions. In fact, quantum phase transition deals with the phenomena of structure changes in the physical space of pattern formation equation (8.1.17) of condensate systems at critical control parameters.

8.2 Theory for High-Temperature Superconductivity

8.2.1 Physical Mechanism of Superconductivity

In the previous section, we introduced Quantum Condensation Mechanism 8.1.1. For superconducting systems, this mechanism can be explicitly stated as follows:

- *there is a critical temperature $T_c > 0$ such that for $T < T_c$, a large number of free electrons in the system form bosonic electron-pairs; and*
- (8.2.1)

- *the interactions between the bosonic electron-pairs and other particles in the system are negligible, so that the electron-pairs share a common interaction potential Φ . This common potential Φ enables the electron-pairs to occupy a single wave function field ψ , forming the condensate, and Φ is the potential given by the interaction between the lattice of the system and the external field.*
- (8.2.2)

This is the common physical mechanism of superconductivity. The classical BCS theory (Bardeen et al., 1957) can be used to explain superconductivity in low temperatures. The Cooper pair theory provides the mechanism of electron-phonon interaction in low temperature superconductors, and a physical mechanism form the Bose electron-pair as required in (8.2.1). However, the BCS theory is applicable only for $T < 35$ K.

However, in 1986, Bednorz and Müller (1986) discovered superconductivity in a lanthanum-based cuprate perovskite material, which had a transition temperature of 35 K. Bednorz and Müller's work opened the door for high T_c superconductors. The current highest temperature of superconductors reaches $T_c = 130$, and $T_c = 160$ K under high pressure.

The discovery of high T_c superconductors shows the limitation of the BCS theory. In fact, in the BCS theory, the Cooper electron-pair is explained as

$$\text{the electron-phonon interaction,} \quad (8.2.3)$$

whose physical meaning (picture) is hard to understand. This interaction is reminiscent to the Yukawa interaction mechanism in quantum field theory. The electron-phonon interaction (8.2.3) means that by exchanging phonons, electrons induce an attractive interaction; the physical picture of such an interaction is not clear. Phonon is a hypothetical pseudo-particle. In fact, in quantum field theory, to compute the binding energy Δ of the Cooper pair, a quantized energy needs to be introduced as the binding energy, and is called phonon. In other words, phonon is essentially a hypothesis, written in the form (8.2.3), and is introduced for computing Δ . This hypothesis can be expressed in the following easier understandable fashion:

$$\text{In addition to the repulsive Coulomb potential between electrons,} \\ \text{the deformation of lattice induces an attractive interaction potential.} \quad (8.2.4)$$

In summary, both (8.2.3) and (8.2.4) are different expressions of the same hypothesis. For the BCS theory, regardless of the validity of this hypothesis, the key issue is if there is a good agreement between experiments and the computational results derived under the existence assumption of the binding energy Δ between electrons. Indeed, such agreement is only observed in low temperature cases, but not in high temperature cases. Hence it is worth pointing out that

the validity of a result does not ensure the correctness of the hypothesis introduced to derive the result, since reasons for the same result are usually not unique.

8.2.2 PID Interaction Potential of Electrons

The classical theory of electromagnetism shows that there is a repulsive interaction force between two electrons:

$$F = \frac{e^2}{r^2} \quad (e \text{ is the electric charge}),$$

preventing the paring of electrons, called screening of Coulomb repulsion. The formation of Cooper electron-pairs in the BCS theory makes the hypothesis that the lattice deformation induces the electron-phonon attractive interaction: when free electrons are near the Fermi surface,

$$\text{attractive force } F_1 > \text{Coulomb force } F.$$

Notice also that there exists Coulomb repulsive force between two Cooper pairs:

$$F = \frac{4e^2}{r^2}, \quad (8.2.5)$$

which will cause collision and scattering between Cooper-pairs, preventing condensation and superconducting behaviors of particles. In other words, the repulsive force (8.2.5) makes the mechanism (8.2.2) impossible.

In Ma and Wang (2015a), using the principle of interaction dynamics (PID) and the $SU(N)$ gauge theory for multi-particle systems, the authors obtained a layered electromagnetic interaction potential formula, which can be used here to derive conditions to ensure the validity of the mechanism (8.2.1) and (8.2.2) for superconductivity. For the sake of completion, we recapitulate this formula below.

Consider a particle A with electric charge $Q = ne_0$ with either $e_0 < 0$ or $e_0 > 0$. Let there be a test electron e at distance r from particle A . By the field equations (Ma and Wang, 2015a, (6.5.20) and (6.5.21)) for N particles, the $SU(N)$ ($N = n + 1$) electromagnetic potentials generated by particle A acting on the electron:

$$\{A_\mu^a \mid 0 \leq \mu \leq 3, \quad 1 \leq a \leq N^2 - 1\} \quad (8.2.6)$$

satisfy the following field equations, using the standard $SU(N)$ generators to ensure that $\text{Tr}(\tau_a \tau_b^+) = 2\delta_{ab}$:

$$\partial^\nu A_\nu^\mu - \frac{e_0}{\hbar c} \lambda_{bc}^a g^{ab} A_{a\mu}^b A_{\beta}^c - e_0 \bar{\Psi} \gamma_\mu \tau^a \Psi = [\partial_\mu - \frac{1}{2} k^2 x_\mu] \phi^a, \quad (8.2.7)$$

$$i\gamma^\mu \left(\partial_\mu + \frac{ie_0}{\hbar c} A_\mu^b \tau_b \right) \Psi - \frac{c}{\hbar} M \Psi = 0, \quad (8.2.8)$$

where M is the mass matrix of the particles, $\Psi = (\Psi_1, \dots, \Psi_N)^T$, Ψ_i ($1 \leq i \leq N$) represent i wave function (Dirac spinors), $(g^{\alpha\beta})$ is the Minkowski metric, λ_{bc}^a are the structure constants of $SU(N)$, ϕ^a are the dual potentials of $\{A_\mu^a\}$, γ^μ are the Dirac matrices, $\gamma_\mu = g_{\mu\alpha}\gamma^\alpha$, $\tau^a = \tau_a$, and

$$A_{\nu\mu}^a = \partial_\nu A_\mu^a - \partial_\mu A_\nu^a + \frac{e_0}{\hbar c} \lambda_{bc}^a A_\mu^b A_\nu^c.$$

We note here that the field equations (Ma and Wang, 2015a, (6.5.20) and (6.5.21)) for N particles are derived from first principles: principle of Lorentz invariance, principle of gauge invariance, Einstein's principle of general relativity, principle of representation invariance (PRI), principle of Lagrangian dynamics, and principle of interaction dynamics (PID).

We now derive from (8.2.7) and (8.2.8) the electron interaction potential. Thanks to the PRI applied to $SU(N)$, there is an $SU(N)$ representation tensor $\{\theta_a \mid 1 \leq a \leq N^2 - 1\}$ such that

$$A_\mu = \theta_a A_\mu^a \quad (\mu = 0, 1, 2, 3) \quad (8.2.9)$$

representing the total electromagnetic potential induced by the N electron charges. Namely,

$$\begin{aligned} A_0 & \quad \text{electric potential,} \\ A = (A_1, A_2, A_3) & \quad \text{magnetic potential.} \end{aligned}$$

Let the radius of particle be r_A , $\Phi(r) = A_0(r)$ be the electric potential acting on the test electron e . The additivity of electric potential shows that the $\mu = 0$ component of (8.2.7) is linear, taking the following form; see Ma and Wang (2015a) for details:

$$\begin{aligned} -\Delta\Phi &= e_0 Q_0 - \frac{1}{4} k^2 c\tau\phi, \\ -\Delta\phi + k^2\phi &= -e_0 \operatorname{div}\vec{Q}, \end{aligned} \quad (8.2.10)$$

where $c\tau$ represents the wave-length of the electron, k is the electric charge interaction parameter,

$$\phi = \theta_a \phi^a, \quad (Q_0, \vec{Q}) = Q_\mu = -\theta_a \bar{\Psi} \gamma_\mu \tau^a \Psi,$$

and θ_a is as in (8.2.9). Here

$$Q_0 = nB_0 \left(\frac{r_e}{r_A} \right)^3 \delta(r), \quad \operatorname{div}\vec{Q} = \frac{B_1}{r_A} \delta(r), \quad (8.2.11)$$

where r_A , r_e are radii the particle A and the electron respectively, $\delta(r)$ is the Dirac function, B_0 , B_1 are parameters.

Then with the same procedure and derivation in Ma and Wang 2015a, (the layered strong interaction formula (4.5.41)), we obtain from (8.2.10) to (8.2.11) the following

$$\Phi(r) = nB_0 \left(\frac{r_e}{r_A} \right)^3 e_0 \left[\frac{1}{r} - A_0(1 + kr)e^{-kr} \right], \quad (8.2.12)$$

where B_0 is as (8.2.11), k is as (8.2.10), and A_0 can be expressed explicitly as

$$A_0 = \frac{k^2 c \tau r_A^2 B_1}{4nr_e^3 B_0}.$$

Here A_0 depends on the types of particle A and the test electron, and is determined by experiments.

Formula (8.2.12) represents the electric potential induced by particle A , where

$$q_A = nB_0 \left(\frac{r_e}{r_A} \right)^3 e_0 \quad (8.2.13)$$

represents the effective electric charge of A , and B_0 depends on the particle type of A . The physical meaning of the effective charge is apparent in (8.2.14) and (8.2.15) below. For two effective charges q_1 and q_2 , the electric potential between them is

$$V_{12}(r) = q_1 q_2 \left[\frac{1}{r} - A_{12}(1 + kr)e^{-kr} \right], \quad (8.2.14)$$

where A_{12} is a parameter depending on the types of the two particles. The force between the two particles with effective charges q_1 and q_2 is then given by

$$F_{12} = -q_1 q_2 \frac{d}{dr} \left[\frac{1}{r} - A_{12}(1 + kr)e^{-kr} \right]. \quad (8.2.15)$$

8.2.3 Condition for Formation of Electron-Pairs

The field equations (8.2.7) and (8.2.8), together with the electron interaction potential and force formulas (8.2.14) and (8.2.15), form the theoretical basis of the new theory on electron-pairs in superconductivity in this section.

1. *Electric Charge Interaction Between Two Electrons* When both particles are electrons, $n = 1$, $B_0 = 1$, $r_A = r_e$. Then (8.2.12) becomes

$$\Phi_e = \frac{e_0}{r} - e_0 A_0(1 + kr)e^{-kr}, \quad (8.2.16)$$

which is the modification of the classical Coulomb potential. The first term in the right-hand side is the Coulomb potential, and the second term represents the effect of the dual field ϕ of A_μ , induced from PID.

By (8.2.16), the interaction force between two electrons is given by

$$F_e = -e_0 \nabla \Phi_e = \frac{e_0^2}{r^2} - k^2 A_0 e_0^2 r e^{-kr}. \quad (8.2.17)$$

There is a short range attractive force in the right-hand side of (8.2.17) ($A_0 > 0$):

$$f = -k^2 A_0 e_0^2 r e^{-kr} < 0.$$

It is this force f , together with magnetic moment of electrons, that induces a short range attractive force to counteract the repulsive Coulomb force.

2. *Spin Magnetic Moment Interaction of Electrons* Electrons possess spin magnetic moment, with numerical value given by

$$\mu_0 = -\frac{e_0 \hbar}{2mc}, \quad (8.2.18)$$

where $e_0 < 0$ is the electric charge of an electron, and m is the mass of electron. We can define the direction of the spin magnetic moment, as the north pole N , and the opposite direction as the south pole S . Then there is a spin magnetic moment interaction force

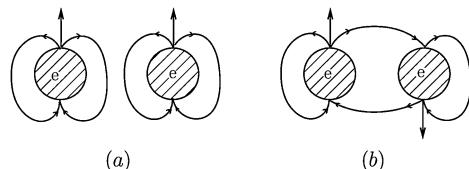
$$F_m = \pm \frac{\alpha \mu_0^2}{r^2}, \quad (8.2.19)$$

where α is the magnetic moment interaction coefficient. Physically, we know that

$$0 < \alpha < e_0^2 / \mu_0^2. \quad (8.2.20)$$

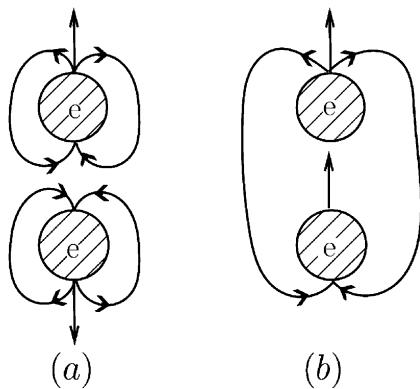
There are two electron orientation arrangements that can produce the magnetic moment force given by (8.2.19): one is the parallel arrangement, as shown in Fig. 8.1a, b; the other is collinear arrangement as shown in Fig. 8.2a, b. The plus sign in (8.2.19) represents repulsive force, corresponding to the cases of Figs. 8.1a and 8.2a. Hence only Figs. 8.1b and 8.2b may give rise to superconducting electron-pairs.

Fig. 8.1 (a) Parallel orientation leads to repulsive magnetic moment force; and **(b)** anti-parallel orientation leads to attractive magnetic moment force



3. *Stability of Orientation Arrangements of Electrons* For the two electron orientation arrangements that produce attractive spin magnetic moment force, the antiparallel electron-pair shown in Fig. 8.1b is stable, and the one in Fig. 8.2b is unstable. There are two main reasons behind the stability and instability of these two cases. First, since the collinear arrangement of an electron-pair possesses electric charge $2e_0$ and spin $J = 1$, it induces a magnetic moment with intensity: $\mu = -2e_0 \hbar / mc$. This causes magnetic moment interaction between electron-pairs. Such long arrangement is also the source of instability, and two reversed parallel

Fig. 8.2 (a) Anti-collinear orientation leads to repulsive magnetic moment force; and (b) collinear orientation leads to attractive magnetic moment force



electron-pairs as shown in Fig. 8.2b can break form two electron-pairs with the structure as shown in Fig. 8.1b. Second, by the Heisenberg uncertainty relation, for an electron-pair, the distance between the two electrons and the kinetic energy $p = mv$ of the electrons satisfy the following inequality:

$$v \geq \frac{\hbar}{2mr}.$$

This implies that the two electrons in an electron-pair must rotate around a mass center. Hence the rotating arrangement of electrons in Fig. 8.1b is stable, and the one in Fig. 8.2b is unstable. We arrive then at the following conclusion:

$$\text{only electron-pairs with total spin } J = 0 \text{ is stable.} \quad (8.2.21)$$

4. *Attracting Shell Region of Electron-Pairs* By (8.2.17) and (8.2.19), the interaction force formula of two anti-parallel electrons is given by

$$F = \frac{e_0^2}{r^2} \left(1 - \frac{\alpha \mu_0^2}{e_0^2} \right) - k^2 A_0 r e^{-kr} e_0^2. \quad (8.2.22)$$

The interval where F takes negative values defines the attraction shell region of two electrons:

$$\bar{r}_0 < r < \bar{r}_1 \quad (\bar{r}_0 > 0), \quad (8.2.23)$$

where \bar{r}_0 , \bar{r}_1 are two positive roots of $F(r) = 0$, satisfying

$$\frac{r^3}{e^{kr}} = \frac{1}{k^2 A_0} \left(1 - \frac{\alpha \mu_0^2}{e_0^2} \right). \quad (8.2.24)$$

Here the condition for (8.2.24) having two positive roots is that the maximum of r^3/e^{kr} is greater than the right-hand side of the above equation:

$$\left(\frac{3}{e}\right)^3 > \frac{k}{A_0} \left(1 - \frac{\alpha\mu_0^2}{e_0^2}\right),$$

where $e \approx 2.718$. By (8.2.18), the above condition can be rewritten as

$$A_0 > \left(\frac{e}{3}\right)^3 k \left(1 - \frac{\alpha\hbar^2}{4m^2c^2}\right), \quad (8.2.25)$$

which is the condition to ensure the existence of the attracting shell region (8.2.23).

5. *Attracting Potential of Electron-Pairs* The attracting interaction potential of magnetic moment for an electron-pair is

$$V_m = -\frac{\alpha\mu_0^2}{r} = -\frac{\alpha\hbar^2}{4m^2c^2 r} e_0^2.$$

By (8.2.16), the total interaction potential of an anti-parallel electron-pair is

$$V = \frac{e_0^2}{r} \left(1 - \frac{\alpha\hbar^2}{4m^2c^2}\right) - e_0^2 A_0 (1 + kr) e^{-kr}. \quad (8.2.26)$$

By the relation (8.2.22) between the potential V and the force F :

$$F = -\frac{d}{dr} V,$$

The first zero \bar{r}_0 of F in (8.2.23) is the minimum point V :

$$V(\bar{r}_0) = \min V(r). \quad (8.2.27)$$

The condition for an anti-parallel electron-pair to possess a shell region of attracting potential is $V(\bar{r}_0) < 0$, i.e.

$$A_0 > \frac{e^{k\bar{r}_0}}{\bar{r}_0(1+k\bar{r}_0)} \left(1 - \frac{\alpha\hbar^2}{4m^2c^2}\right). \quad (8.2.28)$$

This is the condition for an electron-pair to possess attracting interaction potential. It is clear that we can view \bar{r}_0 as the theoretical radius of electron-pairs. Let

$$A_0 = A/\bar{r}_0, \quad (8.2.29)$$

then (8.2.28) becomes

$$A > \frac{e^{k\bar{r}_0}}{1+k\bar{r}_0} \left(1 - \frac{\alpha\hbar^2}{4m^2c^2}\right). \quad (8.2.30)$$

Obviously if this condition holds true, then (8.2.25) holds true as well for $k\bar{r}_0 > 0$.

8.2.4 Binding Energy of Superconducting Electron-Pairs

The key ingredient of the BCS theory of superconductivity is to postulate the existence of the binding energy of electron-pairs for $T < T_c$:

$$\begin{aligned} \text{binding energy of electron-pair} &= 2\Delta(T) \quad \text{for } T < T_c, \\ \Delta(T_c) &= 0. \end{aligned} \quad (8.2.31)$$

The relation between $\Delta(T)$ and the energy spectrum $E(p)$ of the electron-pair is

$$E(p) = \sqrt{(E_p - E_F)^2 + \Delta(T)}. \quad (8.2.32)$$

where E_F is the energy of Fermi level, and $E_p = \frac{1}{2m}p^2$ is the kinetic energy of the particle. Also, Δ is determined by the following equation, called energy-gap equation:

$$N(0)V \int_0^\infty \frac{1}{(2\pi\hbar)^3 E(p)} \left[1 - \frac{2}{1 + \exp(E(p)/kT)} \right] dp = 1, \quad (8.2.33)$$

where $E(p)$ is as in (8.2.32), $N(0)$ is the Fermi energy density with temperature at absolute zero, and V is the intensity of attracting portion of potential of electrons.

The above three formulas (8.2.31)–(8.2.33) are the three main ingredients of the BCS theory, and the remaining components of the BCS theory can be derived from these three formulas. By $\Delta(T_c) = 0$, we derive from the energy-gap equation (8.2.33) the following formula for the critical temperature T_c :

$$kT_c = 1.14\hbar\omega_D \exp[-1/N(0)V], \quad (8.2.34)$$

where ω_D is the Debye frequency, and $N(0)$ and V are as in (8.2.33). The highest critical temperature formula (8.2.34). Hence for high Tc superconductivity, the BCS theory is invalid, and in particular for high Tc superconductivity, the electron-phonon interaction mechanism for electron-pairs is not convincing.

We now establish the new theory of binding energy of electrons, suitable for high Tc superconductivity, based on the new interaction theory of electrons established earlier.

1. *Binding Energy of Electron-Pairs* The previous section has shown that under condition (8.2.30), two electrons with reversed parallel orientation can form an electron-pair, with radius \bar{r}_0 satisfying (8.2.27).

In the BCS theory, the binding energy 2Δ of electron-pairs is determined by the energy-gap equation (8.2.33). The binding energy 2Δ can also be expressed using the potential in (8.2.26) as follows:

$$2\Delta = -V(\bar{r}_0) - m\bar{v}^2, \quad (8.2.35)$$

where $m\bar{v}^2$ stands for the total average kinetic energy of the two electrons. By the average energy level formula of temperature, we can view the average kinetic energy of the two electrons as a function of temperature:

$$\frac{1}{2}m\bar{v}^2 = \varepsilon(T). \quad (8.2.36)$$

Hence (8.2.35)–(8.2.36) can be written as

$$\Delta = \frac{e_0^2}{2\bar{r}_0} \left[A(1 + k\bar{r}_0)e^{-k\bar{r}_0} + \frac{\alpha\hbar^2}{4m^2c^2} - 1 \right] - \varepsilon(T), \quad (8.2.37)$$

where A is a parameter as in (8.2.29)–(8.2.30).

Formula (8.2.37) is the formula for the binding energy Δ of electron-pairs of superconductivity, in terms of average kinetic energy of electrons and temperature function $\varepsilon(T)$. At low temperature, Δ as well as $\varepsilon(T)$ are determined by (8.2.33), showing that the new theory is consistent with the BCS theory at low temperatures. Also, physically, by (8.2.35), the critical temperature T_c satisfies $\Delta(T_c) = 0$, which amounts to saying that

$$\varepsilon(T_c) = \frac{e_0^2}{2\bar{r}_0} \left[A(1 + k\bar{r}_0)e^{-k\bar{r}_0} + \frac{\alpha\hbar^2}{4m^2c^2} - 1 \right], \quad (8.2.38)$$

which is the new temperature formula, applicable for both low and high T_c cases.

2. *Verification of Mechanism (8.2.1) of Superconductivity* By (8.2.36), $\varepsilon(T)$ is an increasing function of T :

$$\frac{d}{dT}\varepsilon(T) > 0.$$

Hence we infer from (8.2.37) and (8.2.38) that

$$\begin{aligned} \Delta(T) &= 0 && \text{for } T \geq T_c, \\ \frac{d}{dT}\Delta(T) &< 0 && \text{for } T < T_c. \end{aligned} \quad (8.2.39)$$

This verifies the mechanism (8.2.1) of superconductivity. Namely,

$$\begin{array}{ll} \text{no superconducting electron-pairs emerge} & \text{for } T \geq T_c, \\ \text{superconducting electron-pairs emerge} & \text{for } T < T_c. \end{array}$$

By (8.2.21), we note that

$$\text{total spin of electron-pairs } J = 0.$$

3. *Verification of Mechanism (8.2.2) of Superconductivity* For $T < T_c$, electron-pairs with radius \bar{r}_0 emerge in a superconducting system. By (8.2.13), the effective charge of a superconducting electron-pair is

$$q_0 = 2B_0 \left(\frac{r_e}{\bar{r}_0} \right)^3 e_0.$$

The classical radius of an electron is

$$r_e = \frac{e_0^2}{mc^2} = 2.8 \times 10^{-13} \text{ cm},$$

and the radius of an electron-pair \bar{r}_0 has the same order of magnitude as the diameter of the particle lattice of the system: $\rho \sim 10^{-8} \text{ cm}$. Hence

$$q_0 \sim 10^{-13} B_0 e_0.$$

In other words, the effective electric charge of an electron-pair is negligibly small, almost electric neutral. Hence according to the interaction force formula of electric charge (8.2.15), the interaction between superconducting electron-pairs and particles in the system are extremely weak. Consequently this verifies the mechanism (8.2.2) of superconductivity.

8.2.5 Formulas for the Critical Temperature T_c

We now compute the critical temperature T_c from both the microscopic and macroscopic approaches. The macroscopic approach is based on the thermodynamic potential (free-energy) in the standard model of thermodynamics, and the microscopic approach is based on (8.2.37).

(1) Formula for T_c Based on the Standard Model

First we recall that the thermodynamic potential for superconductivity. Under no external magnetic field and homogeneous conditions, the potential functional (8.3.17) for superconductivity becomes

$$F = -g_0|\psi|^2 + \frac{1}{2}g_1|\psi|^4 - \frac{\alpha_0}{2}TS^2 - \alpha_1 TS|\psi|^2 - ST, \quad (8.2.40)$$

where $g_0 > 0$ is the binding potential of condensates, g_1 is the coupling constant of particles, α_0 and α_1 are entropy coupling constants. The entropy is determined by

$$\frac{\partial F}{\partial S} = 0 \Rightarrow S = -\frac{1}{\alpha_0} - \frac{\alpha_1}{\alpha_0}|\psi|^2. \quad (8.2.41)$$

On the other hand, we have

$$-\frac{\partial}{\partial \psi^*} F(\psi) = g_0\psi - g_1|\psi|^2\psi + \alpha_1 TS\psi. \quad (8.2.42)$$

In view of (8.2.41) for entropy S , we have

$$-\frac{\partial}{\partial \psi^*} F(\psi) = \beta_1 \psi - \left(g_1 + \frac{\alpha_1^2}{\alpha_0} T \right) |\psi|^2 \psi, \quad (8.2.43)$$

where β_1 represents the eigenvalue of the linearized operator of (8.2.43), given by

$$\beta_1(T) = g_0 - \frac{\alpha_1}{\alpha_0} T.$$

By Theorem 3.1.3, $\beta_1(T) = 0$ provides the critical temperature:

$$T_c = \alpha_0 g_0 / \alpha_1. \quad (8.2.44)$$

Consider T_c in (8.2.44). If $T < T_c$, by (8.2.43) and

$$\frac{\partial}{\partial \psi^*} F(\psi) = 0,$$

we deduce the solution for superconducting phase

$$|\psi|^2 = \frac{\beta_1(T)}{g_1 + \alpha_1^2 T / \alpha_0} = \frac{\alpha_1}{\alpha_0} \frac{(T_c - T)}{g_1 + \alpha_1^2 T / \alpha_0}. \quad (8.2.45)$$

Plugging (8.2.45) into (8.2.41), we obtain that

$$S = \begin{cases} -\frac{1}{\alpha_0} - \left(\frac{\alpha_1}{\alpha_0} \right)^2 \frac{(T_c - T)}{g_1 + \alpha_1^2 T / \alpha_0} + S_0(T) & \text{for } T < T_c, \\ S_0(T) & \text{for } T_c < T. \end{cases}$$

Therefore, we derive that

$$\Delta C_V = T_c \left[\frac{\partial S}{\partial T} \Big|_{T_c^-} - \frac{\partial S_0}{\partial T} \right] = \left(\frac{\alpha_1}{\alpha_0} \right)^2 \frac{T_c}{g_1 + \alpha_1^2 T_c / \alpha_0}, \quad (8.2.46)$$

where ΔC_V is the increment of heat capacity at T_c . By (8.2.46), we have

$$\frac{\alpha_0^2}{\alpha_1^2} = \frac{T_c}{g_1 \Delta C_V} (1 - \alpha_0 \Delta C_V), \quad (8.2.47)$$

where it is required that

$$\alpha_0 \Delta C_V < 1. \quad (8.2.48)$$

This is the condition for superconductivity to occur. By (8.2.44) and (8.2.47), we deduce that

$$T_c = \frac{g_0^2}{g_1 \Delta C_V} (1 - \alpha_0 \Delta C_V). \quad (8.2.49)$$

This is the formula for critical temperature T_c based on the standard model.

Remark 8.2.1 The formula (8.2.49) of the critical temperature T_c depends on the parameters g_0 , g_1 , α_0 and ΔC_V . Here ΔC_V is the increment of heat capacity at the critical temperature T_c , which is a theoretical value without taking into consideration of fluctuations and does not reflect the real measurement. \square

(2) Formula for T_c Based on Microscopic Theory

Formula (8.2.38) provides the theoretical basis for computing T_c . To obtain a formula for T_c , we need consider the expression for the function $\varepsilon(T)$. First, $\varepsilon(T)$ is defined by (8.2.36), and represents the average energy level of electron gas in the superconducting system. By the statistical theory of heat (Ma and Wang, 2017d), we can derive an expression for $\varepsilon(T)$ as follows:

$$\varepsilon(T) = \gamma k_B T, \quad k_B \text{ is the Boltzmann constant,} \quad (8.2.50)$$

where γ is called the energy ratio of the system, and can be expressed as

$$\gamma = \frac{N^2}{N_1 N_2 (1 + \theta)}. \quad (8.2.51)$$

Here N is the total number of particles in the system, N_1 , N_2 are respectively the numbers of free electrons and lattice particles such that $N_1 + N_2 = N$, and θ is

$$\theta = \frac{\text{average energy level of lattice particles}}{\text{average energy level of free electrons}}. \quad (8.2.52)$$

We shall prove (8.2.50)–(8.2.52) later.

It is clear that θ defined by (8.2.52) depends on the temperature, and consequently γ depends on the temperature as well. Let γ_0 be the temperature ratio of the electron gas at the critical temperature T_c , then by (8.2.50)–(8.2.51), the formula for T_c in (8.2.38) can be rewritten as

$$k_B T_c = \frac{E_0}{\gamma_0} = \frac{N_1 N_2 (1 + \theta_0)}{N^2} E_0. \quad (8.2.53)$$

Here E_0 is the intrinsic energy of electron-pairs, expressed as

$$E_0 = \frac{e_0^2}{2\bar{r}_0} \left[A(1 + k\bar{r}_0)e^{-kr_0} + \frac{\alpha\hbar^2}{4m^2 c^2} - 1 \right]. \quad (8.2.54)$$

Formulas (8.2.53) and (8.2.54) provide the needed formula for T_c based on the microscopic theory.

(3) Derivation of the Expression for $\varepsilon(T)$

First we recall the average energy level temperature formula in Ma and Wang (2017d):

$$k_B T = \frac{1}{N} \sum_n \frac{a_n \varepsilon_n}{1 + \beta_n \ln \varepsilon_n} \left(1 - \frac{a_n}{N} \right), \quad (8.2.55)$$

where ε_n is the n -th energy level, a_n is the number of particles on the energy level ε_n , $N = \sum_n a_n$ is the total number of particles. A superconducting system can be considered as a binary system of free electron gases and lattice-particles. Therefore, (8.2.55) can be written as

$$\frac{N_1 N_2}{N^2} \frac{\varepsilon_{\text{electron}}}{1 + \beta_1 \ln \varepsilon_{\text{electron}}} + \frac{N_1 N_2}{N^2} \frac{\varepsilon_{\text{lattice}}}{1 + \beta_2 \ln \varepsilon_{\text{lattice}}}, \quad (8.2.56)$$

where N_1, N_2 are respectively the numbers of free electrons and lattice-particles in the system, $\varepsilon_{\text{electron}}$ and $\varepsilon_{\text{lattice}}$ are the average energy levels of electrons and lattice-particles. We approximately take $\beta_1, \beta_2 = 0$, then the above expression reduces to

$$k_B T = \frac{N_1 N_2}{N^2} \varepsilon_{\text{electron}} + \frac{N_1 N_2}{N^2} \varepsilon_{\text{lattice}}. \quad (8.2.57)$$

Let $\theta = \varepsilon_{\text{lattice}} / \varepsilon_{\text{electron}}$, then we infer from (8.2.57) that

$$k_B T = \frac{N_1 N_2}{N^2} (1 + \theta) \varepsilon_{\text{electron}}, \quad (8.2.58)$$

where $\varepsilon_{\text{electron}} = \frac{1}{2} m \bar{v}^2$ is the average kinetic energy. Therefore, (8.2.58) implies (8.2.50).

(4) Theoretical Analysis of T_c

From (8.2.53), we can deduce the following property of high Tc superconducting materials:

the energy level for the lattice-particle is much larger than that of free electrons, and consequently $\theta \gg 1$.

We have derived basic analysis of the new microscopic theory of superconductivity, and further analysis needs to consider the effect and physical meaning of $\beta_1, \beta_2 \neq 0$ in (8.2.56).

Also, the formula (8.2.49) based on the standard model indicates that the critical temperature T_c is a decreasing function of the increment of heat capacity. We now demonstrate that the microscopic theory of T_c can reveal such property as well. Recall the heat capacity formula

$$C_V = \frac{\partial E}{\partial T}, \quad E \text{ is the internal energy.}$$

By (8.2.50), we have

$$N_1 k_B \gamma = N_1 \frac{\partial \varepsilon}{\partial T} = \text{heat capacity of electron gas.}$$

The phase transition is due to condensation of the electron gas, and is independent of the lattice. Hence the heat capacity for the lattice at T_c is continuous. Hence at T_c , γ_0 is given by

$$\gamma_0 \sim \Delta C_V + \delta \quad (\delta = \gamma(T_c^-)).$$

We infer then from (8.2.53) that

$$k_B T_c \sim \frac{E_0}{\Delta C_V + \delta}.$$

This provides the same property as the macroscopic formula (8.2.49) based on the standard model.

In summary, (8.2.49) is derived using the standard model of thermodynamics, and the microscopic formula (8.2.53) is derived using PID interaction theory of electrons and the statistical theory of heat. With different origins, these two formulas possess the same property.

8.3 Thermodynamic Potentials for Condensates

Condensates near or away from the critical temperature T_c behave differently, and the study of condensates can be classified into thermodynamic and quantum categories. Thermodynamic systems of condensates study transitions between normal and condensate states, which is thermal driven as studied in Chap. 3 in this book. This section derives the thermodynamic potential of condensates near the critical temperature.

Quantum system of condensates deal with the quantum behavior of fully developed condensates. The mathematical model of quantum condensate systems are the Hamiltonian systems to be developed in the next section, and the related phase transition is the quantum phase transition to be studied in the next chapter of this book.

8.3.1 Quantum Rules of Condensates

1. *Quantum Rules* When a thermodynamic system forms a condensate, the order parameter is a family of complex-valued wave functions $\psi : \Omega \rightarrow \mathbb{C}^N$ such that

$$|\psi|^2 \text{ represents the particle number density of the condensate.} \quad (8.3.1)$$

The wave functions ψ are governed by physical laws, and describes the collective behavior of the condensate particles, which possess mutual interactions. Hence for condensates these wave functions obey their own quantum rules as follows:

Quantum Rule 8.3.1 *For quantum physics of condensates, the particle interaction is attractive in long-range, and repulsive in short-range. Hence the wave functions ψ of condensates obey the following quantum rules:*

- a. *the (attractive) binding energy between particles is*

$$\text{binding energy density} = -g_0|\psi|^2, \quad (8.3.2)$$

where $g_0 > 0$ stands for interaction potential;

- b. *the repulsive interaction energy between particles is*

$$\text{repulsive energy density} = \frac{1}{2}g_1|\psi|^4, \quad (8.3.3)$$

where $g_1 > 0$ is the interaction constant.

Remark 8.3.2 Quantum Rule 8.3.1 is the consequence of the combination of physical facts and the repulsive-attractive characteristic of the three interactions: the electromagnetic, the weak, and the strong interactions; see Ma and Wang (2015a). It is the attractive potential (8.3.2) that ensures a large collection of particles to condensate, and it is also the short-range repulsive characteristic (8.3.3) that ensures the finite density of particles. In (8.3.2) and (8.3.3), the interaction constants g_0 and g_1 are macroscopic statistical average values, and g_0 depends on the external magnetic field. \square

In condensed matter physics, ψ represents the wave function for the collection of particles in the condensate. The differential operator $-i\hbar\nabla$ stands for the gradient potential, and $-\frac{\hbar^2}{2m}\nabla^2$ stands for the gradient potential energy. In addition, we have the following correspondences of differential operators and their physical meaning.

In the absence of external electromagnetic field, the following are the basic Hermitian differential operators:

$$\begin{aligned} \text{energy operator : } & \widehat{E} = i\hbar \frac{\partial}{\partial t}, \\ \text{gradient potential operator : } & \widehat{P} = -i\hbar\nabla, \\ \text{interaction potential operator : } & \widehat{V}\psi = V\psi, \\ \text{gradient energy operator : } & \widehat{K} = -\frac{\hbar^2}{2m}\nabla^2. \end{aligned} \quad (8.3.4)$$

The energies induced by the operators in (8.3.4) are

$$\text{potential energy} \quad V = \int_{\Omega} V|\psi|^2 dx, \quad (8.3.5)$$

$$\text{gradient energy} \quad K = \int_{\Omega} \hat{K} \psi \cdot \psi^\dagger dx = \int_{\Omega} \frac{\hbar^2}{2m} |\nabla \psi|^2 dx. \quad (8.3.6)$$

We remark that (8.3.2) in Quantum Rule 8.3.1 is the consequence of (8.3.5) and the potentials of the three interactions (Ma and Wang, 2015a).

In the presence of electromagnetic field, the Hermitian operators are given by

$$\begin{aligned} \text{energy operator :} \quad & \hat{E} = i\hbar \frac{\partial}{\partial t} + eA_0, \\ \text{gradient potential operator :} \quad & \hat{P} = -\left(i\hbar \nabla + \frac{e}{c} \mathbf{A}\right), \\ \text{gradient energy operator :} \quad & \hat{K} = -\frac{\hbar^2}{2m} \left(\nabla - i \frac{e}{\hbar c} \mathbf{A}\right)^2. \end{aligned} \quad (8.3.7)$$

The energy induced by the gradient energy operator \hat{K} is

$$K = \int_{\Omega} \frac{1}{2m} \left| \left(-i\hbar \nabla - \frac{e}{c} \mathbf{A} \right) \psi \right|^2 dx, \quad (8.3.8)$$

where (A_0, \mathbf{A}) is the four-dimensional electromagnetic potential, and K is the energy induced by the nonuniform distribution.

2. *Implications of Basic Principles* By the Pauli exclusion principle, only one Fermion is allowed in each quantum state. The particles in a condensate share the same quantum state; this shows that the particles in a condensate must be Bosons. By the Heisenberg uncertainty relation, for a particle in the system, the position x and momentum P , time t energy E cannot be determined simultaneously determined. Their deviations Δx , ΔP , Δt , and ΔE satisfy

$$\Delta x \Delta P \geq \frac{1}{2} \hbar, \quad \Delta t \Delta E \geq \frac{1}{2} \hbar. \quad (8.3.9)$$

These relations imply that there exist quantum fluctuations in a condensate system—spontaneous deviations of the position, momentum, and energy from the equilibrium can occur, and are due to the quantum uncertainty, rather than to thermo effect.

8.3.2 Ginzburg-Landau Free Energy of Superconductivity

The classical Ginzburg-Landau free energy is a phenomenological model. In this section, we derive the potential functionals for superconducting systems based on the following *first principles*:

1. Quantum Rule 8.3.1,
2. quantum kinetic energy formula (8.3.8),

3. theory of electromagnetism, and
4. the statistical theory of heat in Sect. 7.5.

First the order parameters u and the control parameters λ are

$$u = (\psi, \mathbf{A}, S), \quad \lambda = (T, \mathbf{H}_a, L),$$

where $\psi : \Omega \rightarrow \mathbb{C}$ is the wave function of superconducting electrons, \mathbf{A} is the magnetic potential, \mathbf{H}_a is the external magnetic field, and

$$|\psi|^2 \text{ represents superconducting electron density.} \quad (8.3.10)$$

The Gibbs free energy consists of

$$\begin{aligned} F = & \text{gradient energy} + \text{binding energy} \\ & + \text{repulsive interaction energy} + \text{magnetic energy} \\ & + \text{entropy energy} + \int_{\Omega} \left(-ST - \frac{1}{4\pi} \operatorname{curl} \mathbf{A} \cdot \mathbf{H}_a \right) dx. \end{aligned} \quad (8.3.11)$$

As in (8.3.8), the gradient energy of ψ under the electromagnetic field is given by

$$\text{gradient energy} = \int_{\Omega} \frac{1}{2m_s} \left| \left(-i\hbar \nabla - \frac{e_s}{c} \mathbf{A} \right) \psi \right|^2 dx. \quad (8.3.12)$$

By (8.3.2), the binding energy of superconducting electrons is

$$\text{binding energy} = \int_{\Omega} -g_0 |\psi|^2 dx. \quad (8.3.13)$$

By (8.3.3), the superconducting repulsive interaction energy is

$$\text{repulsive interaction energy} = \int_{\Omega} \frac{g_1}{2} |\psi|^4 dx. \quad (8.3.14)$$

The magnetic field energy is

$$\text{magnetic field energy} = \int_{\Omega} \frac{1}{8\pi} |\operatorname{curl} \mathbf{A}|^2 dx. \quad (8.3.15)$$

Also, as in (7.6.68) for magnetic and dielectric systems, the entropy energy of condensates has the form

$$\text{entropy energy} = \int_{\Omega} \left[-\frac{T}{2\alpha k} S^2 - \beta ST |\psi|^2 \right] dx, \quad (8.3.16)$$

where $\alpha, \beta > 0$ are parameters, and $-\beta ST |\psi|^2$ represents the loss of thermal energy ST due to condensation.

By (8.3.12)–(8.3.16), the Gibbs free energy (8.3.11) for superconductivity can be written as

$$F = \int_{\Omega} \left[\frac{1}{2m_s} \left| \left(-i\hbar\nabla - \frac{e_s}{c}\mathbf{A} \right) \psi \right|^2 - g_0 |\psi|^2 + \frac{g_1}{2} |\psi|^4 \right. \\ \left. + \frac{1}{8\pi} |\operatorname{curl}\mathbf{A}|^2 - \frac{T}{2k\alpha} S^2 - \beta TS |\psi|^2 - ST - \frac{1}{4\pi} \operatorname{curl}\mathbf{A} \cdot \mathbf{H}_a \right] dx. \quad (8.3.17)$$

Remark 8.3.3 It is worth mentioning that each term in the potential functional (8.3.17) is derived based on basic physical principles and physical laws, while the classical GL free energy is a phenomenological model. The parameters g_0 and g_1 in (8.3.17) are different from the corresponding coefficients in the classical GL free energy, which are Taylor expansion coefficients. Here g_0 and g_1 possess clear physical meaning, and are independent of the temperature.

8.3.3 Thermodynamic Potential for Liquid 4He

4He is a nonradioactive isotope of the element helium. Its nucleus is identical to an alpha particle, and consists of two protons and two neutrons, with two orbiting electrons. It has an integer spin and is a boson, and consequently 4He may condensate.

Kapitza (1938) discovered in 1938 that liquid 4He undergoes a transition from the normal state to superfluid state with vanishing viscosity at the critical temperature $T_c = 2.17K$. As a condensate system, liquid 4He is a binary system with density ρ decomposed into

$$\rho = \rho_n + \rho_s, \quad (8.3.18)$$

where ρ_n is the density of normal fluid, and ρ_s is the density of superfluid. As a superfluid, its state function is a complex-valued wave function $\psi : \Omega \rightarrow \mathbb{C}$ such that

$$\rho_s = |\psi|^2.$$

The order parameters u and the control parameters λ are

$$u = (\psi, \rho_n, S), \quad \lambda = (T, p).$$

4He can be viewed as a combined system of a PVT system and a condensate phase, and then its thermodynamic potential is their coupling. First, we have the following conservation of total number of particles:

$$\int_{\Omega} [\rho_n + |\psi|^2] dx = \text{constant}. \quad (8.3.19)$$

Also, the Gibbs free energy of the superfluid is given by

$$F = \rho_n \text{ part} + \psi \text{ part} + \text{coupling of } \rho_n \text{ and } \psi \quad (8.3.20)$$

$$+ \text{ part for entropy } S - \int_{\Omega} (\mu\rho + bp\rho_n + ST) dx,$$

where $\mu\rho$ corresponds to Lagrangian multiplier of the conservation law (8.3.19), b is the van der Waals constant, and ρ is as in (8.3.18).

In (8.3.20), the ρ_n part obeys the PVT system. Hence

$$\rho_n \text{ term} = \int_{\Omega} \left[\frac{\alpha}{2} |\nabla \rho_n|^2 + \frac{1}{2} (b^2 p + A_1 T) \rho_n^2 \right] dx, \quad (8.3.21)$$

By Quantum Rule 8.3.1 and (8.3.6), ψ part is written as

$$\psi \text{ part} = \int_{\Omega} \left[\frac{\hbar^2}{2m} |\nabla \psi|^2 - g_0 |\psi|^2 + \frac{g_1}{2} |\psi|^4 \right] dx. \quad (8.3.22)$$

By (8.3.18) and the log term of ρ for the PVT systems, the coupling part of ρ_n and ψ is

coupling term of ρ_n and ψ

$$= \int_{\Omega} \left[A_2 T (1 + b \rho_n + b |\psi|^2) \ln(1 + b \rho_n + b |\psi|^2) + g_2 b \rho_n |\psi|^2 \right] dx, \quad (8.3.23)$$

where the second term in the integrand represents the interaction energy between normal particles and superfluid particles, and

$$g_2 = \begin{cases} > 0 & \text{for attractive interaction,} \\ < 0 & \text{for repulsive interaction.} \end{cases}$$

Also, in view of the case for PVT systems and for condensates, the part involving the entropy is given by

$$\text{entropy part} = \int_{\Omega} \left[- \frac{T}{2\beta_0 k} S^2 + \beta_1 S T \rho_n^2 - \beta_2 S T |\psi|^2 \right] dx. \quad (8.3.24)$$

By (8.3.21)–(8.3.24), the Gibbs free energy of ${}^4\text{He}$ superfluid is written as

$$\begin{aligned} F = & \int_{\Omega} \left[\frac{\alpha}{2} |\nabla \rho_n|^2 + \frac{1}{2} (b^2 p + A_1 T) \rho_n^2 - b p \rho_n - \mu \rho \right. \\ & + \frac{\hbar^2}{2m} |\nabla \psi|^2 - g_0 |\psi|^2 + \frac{g_1}{2} |\psi|^4 + g_2 b \rho_n |\psi|^2 \\ & + A_2 T (1 + b \rho_n + b |\psi|^2) \ln(1 + b \rho_n + b |\psi|^2) \\ & \left. - \frac{T}{2\beta_0 k} S^2 + \beta_1 S T \rho_n^2 - \beta_2 S T |\psi|^2 - S T \right] dx, \end{aligned} \quad (8.3.25)$$

where m is the mass of ${}^4\text{He}$ atom, A_1 , A_2 , β_0 , β_1 , β_2 are parameters, μ is the chemical potential, b is the van der Waals constant, and $\rho = \rho_n + |\psi|^2$.

8.3.4 Liquid ^3He Superfluidity

^3He is an isotope of ^4He , and the superfluidity of liquid ^3He was found in 1971 by D. M. Lee, D. D. Osheroff, and R. C. Richardson, and its transition temperature is $T \approx 10^{-3} \text{ K}$ under $p = 1 \text{ atm}$ (10^5 Pa). ^3He is a fermion and obeys the Fermi–Dirac statistics, and its atoms have to be paired (Cooper pair) to form the superfluid phase, leading to the existence of multiple superfluid phases; see Fig. 8.3.

Note that there are four possible spin states for the Cooper pairs of ^3He atoms:

$$\uparrow\uparrow, \quad \downarrow\downarrow, \quad \uparrow\downarrow, \quad \downarrow\uparrow, \quad (8.3.26)$$

where \uparrow represents the $J = 1/2$ -spin, and \downarrow stands for the $J = -1/2$ -spin. Consequently, all Cooper pairs in (8.3.26) have integral spins $J = 1, -1, 0$:

$$J = 1 : |\uparrow\uparrow\rangle, \quad J = -1 : |\downarrow\downarrow\rangle, \quad J = 0 : |\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle. \quad (8.3.27)$$

For a ^3He superfluid system, the three spin states in (8.3.27) are described by three wave functions:

$$\psi_+ = |\uparrow\uparrow\rangle \text{ state}, \quad \psi_0 = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle) \text{ state}, \quad \psi_- = |\downarrow\downarrow\rangle \text{ state}. \quad (8.3.28)$$

Hence ψ_+ stands for $J = 1$ state, ψ_0 for spin $J = 0$ state, and ψ_- for $J = -1$ state.

Experimental evidences show that in the absence of external magnetic field, there are two superfluid states of ^3He , A , and B superfluid phases:

$$A \text{ superfluid phase} : |\uparrow\uparrow\rangle \text{ state} + |\downarrow\downarrow\rangle \text{ state},$$

$$B \text{ superfluid phase} : |\uparrow\uparrow\rangle \text{ state} + |\downarrow\downarrow\rangle \text{ state} + (|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle) \text{ state}.$$

With the presence of an external magnetic field, there are three ^3He superfluid phases: A_1 , A , B phases as defined above and

$$A_1 \text{ superfluid phase} : |\uparrow\uparrow\rangle \text{ state}.$$

Note that A_1 superfluid phase consists of only $J = 1$ Cooper pairs. A phase consists of both $J = 1, -1$ Cooper pairs, and the B phase consists of $J = 1, -1, 0$ Cooper pairs. The corresponding wave functions of these phases are

$$A_1 \text{ superfluid phase} : \psi_+ \neq 0, \quad \psi_- = 0, \quad \psi_0 = 0,$$

$$A \text{ superfluid phase} : \psi_+ \neq 0, \quad \psi_- \neq 0, \quad \psi_0 = 0, \quad (8.3.29)$$

$$B \text{ superfluid phase} : \psi_+ \neq 0, \quad \psi_- \neq 0, \quad \psi_0 \neq 0.$$

Figures 8.3 and 8.4 provide the experimental phase diagrams for liquid ^3He without or with external magnetic fields, and provide valuable guidance for deriving the thermodynamic functional for liquid ^3He superfluid systems.

Fig. 8.3 The coexistence curve of ${}^3\text{He}$ without an applied magnetic field

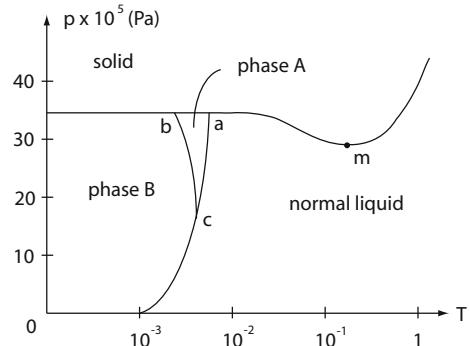
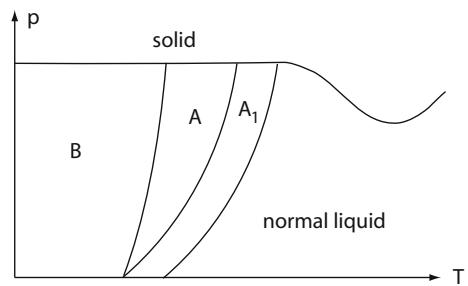


Fig. 8.4 PT -phase diagram of ${}^3\text{He}$ in a magnetic field



1. *Case Without External Magnetic Field ($\mathbf{H}_a = 0$)* The order parameters u and control parameters λ are

$$u = (\psi_+, \psi_0, \psi_-, \rho_n, S), \quad \lambda = (T, p),$$

where $\Psi = (\psi_+, \psi_0, \psi_-)$ is as in (8.3.28), and ρ_n is the normal fluid density. As in (8.3.19), for a liquid ${}^3\text{He}$ superfluid system, the number of particles is conserved:

$$\int_{\Omega} \left[\rho_n + |\Psi|^2 \right] dx = \text{constant}.$$

Hence the Gibbs free energy takes the following form

$$F = \int_{\Omega} \left[\frac{\hbar^2}{2m} |\nabla \Psi|^2 + \frac{\alpha}{2} |\nabla \rho_n|^2 + f(\Psi, \rho_n, S) - \mu \rho - bp \rho_n - ST \right] dx, \quad (8.3.30)$$

where μ is the chemical potential, $\rho = \rho_n + |\Psi|^2$ is the fluid density, and f consists of the following:

$$\begin{aligned} f &= f_1 + f_2 + f_3 + f_4 + f_5, \\ f_1 &= \text{interaction energy of } \rho_n, \\ f_2 &= \text{interaction energy density of } \Psi \text{ particles}, \end{aligned}$$

f_3 = coupling between ρ_n and Ψ ,

f_4 = entropy part,

f_5 = spin interaction energy density.

As in (8.3.21)–(8.3.24), for liquid ^3He superfluid system,

$$f_1 = \frac{1}{2}(b^2 p + A_1 T)\rho_n^2, \quad (8.3.31)$$

$$f_2 = -g_0|\psi_0|^2 - g_{\pm}(|\psi_+|^2 + |\psi_-|^2) + \frac{g_1}{2}|\Psi|^4, \quad (8.3.32)$$

$$f_3 = A_2 T(1 + b\rho) \ln(1 + b\rho) + g_2 b \rho_n |\Psi|^2, \quad (8.3.33)$$

$$f_4 = -\frac{T}{2\beta_0 k} S^2 + \beta_1 S T \rho_n^2 - \beta_2 S T |\Psi|^2, \quad (8.3.34)$$

where $\rho = \rho_n + |\Psi|^2$. For the spin interaction energy f_5 , by the theory of condensates in the next section, we have

$$f_5 = \frac{g_s}{2} |\Psi^\dagger \widehat{F} \Psi|^2. \quad (8.3.35)$$

Here g_s is the spin coupling constant, and \widehat{F} is the spin operator

$$\widehat{F} = (F_1, F_2, F_3), \quad (8.3.36)$$

with F_i ($1 \leq i \leq 3$) being Hermitian matrices:

$$F_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad F_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad F_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Hence $|\Psi^\dagger \widehat{F} \Psi|^2$ in (8.3.35) can be written as

$$|\Psi^\dagger \widehat{F} \Psi|^2 = |\Psi^\dagger F_1 \Psi|^2 + |\Psi^\dagger F_2 \Psi|^2 + |\Psi^\dagger F_3 \Psi|^2,$$

$$|\Psi^\dagger F_k \Psi|^2 = \left[(\psi_+^*, \psi_0^*, \psi_-^*) F_k \begin{pmatrix} \psi_+ \\ \psi_0 \\ \psi_- \end{pmatrix} \right]^2 \quad 1 \leq k \leq 3.$$

Hence by (8.3.31)–(8.3.35), the Gibbs free energy (8.3.30) for liquid ^3He superfluid system takes the form

$$\begin{aligned} F = & \int_{\Omega} \left[\frac{\hbar^2}{2m} |\nabla \Psi|^2 - g_0 |\psi_0|^2 - g_{\pm} (|\psi_+|^2 + |\psi_-|^2) + \frac{g_1}{2} |\Psi|^4 \right. \\ & + \frac{g_s}{2} |\Psi^\dagger \widehat{F} \Psi|^2 + g_2 b \rho_n |\Psi|^2 - \mu(\rho_n + |\Psi|^2) \\ & \left. + \frac{\alpha}{2} |\nabla \rho_n|^2 + \frac{1}{2} (b^2 p + A_1 T) \rho_n^2 - b p \rho_n \right] \end{aligned}$$

$$+ A_2 T (1 + b\rho_n + b|\Psi|^2) \ln(1 + b\rho_n + b|\Psi|^2) \\ - \frac{T}{2\beta_0 k} S^2 + \beta_1 ST \rho_n^2 - \beta_2 ST |\Psi|^2 - ST \Big] dx. \quad (8.3.37)$$

Remark 8.3.4 The coefficients g_0 and g_{\pm} represent the binding energies for $J = 0$ and $J = 1$ superfluid particles. Since there exists spin magnetic moment potential for $J = 1$ particles,

$$g_{\pm} \neq g_0. \quad (8.3.38)$$

It is this relation (8.3.38) that causes the differences between phase diagrams in Figs. 8.3 and 8.4. \square

2. *Case with External Magnetic Field ($\mathbf{H}_a \neq 0$)* With the presence of external magnetic field $\mathbf{H}_a \neq 0$, we can see from Fig. 8.4 that \mathbf{H}_a influences the potential for a ${}^3\text{He}$ superfluid system. Since the total electric charge of Cooper atom pairs is zero, the momentum operator of Ψ is still $-i\hbar\nabla$. However, under the magnetic field \mathbf{H}_a , there is an additional energy density due to the spin magnetic moment of the Cooper pairs:

$$\text{spin magnetic moment energy density} = -\boldsymbol{\mu}_0 \cdot \mathbf{H}(|\psi_+|^2 - |\psi_-|^2), \quad (8.3.39)$$

where direction of the external magnetic field is the same as the spin direction of ψ_+ , $\boldsymbol{\mu}_0$ is the magnetic moment of the Cooper atom pair, and \mathbf{H} is the induced magnetic field. Also there is a magnetic part in the free energy:

$$\int_{\Omega} \left[\frac{1}{8\pi} \mathbf{H}^2 - \frac{1}{4\pi} \mathbf{H} \cdot \mathbf{H}_a \right] dx, \quad (8.3.40)$$

where the first term is the induced magnetic field energy, and the second term is the factor due to Legendre transformation. Hence under $\mathbf{H}_a \neq 0$, the Gibbs free energy F_1 for ${}^3\text{He}$ superfluid system is the sum of (8.3.37), (8.3.39), and (8.3.40):

$$F_1 = F + \int_{\Omega} \left[\frac{1}{8\pi} \mathbf{H}^2 - \boldsymbol{\mu}_0 \cdot \mathbf{H}(|\psi_+|^2 - |\psi_-|^2) - \frac{1}{4\pi} \mathbf{H} \cdot \mathbf{H}_a \right] dx, \quad (8.3.41)$$

where F is as in (8.3.37).

8.3.5 Gibbs Free Energy for Gaseous Condensates (BEC)

Gaseous condensate is called Bose–Einstein condensate (BEC). It is the prediction of a matter form by Einstein in 1924 (Bose, 1924; Einstein, 1924, 1925) based on the Bose–Einstein (BE) distribution, which is given by

$$a_n = \frac{g_n}{e^{(\epsilon_n - \mu)/kT} - 1}.$$

Einstein hypothesized that at $T = 0\text{ K}$, $\mu = 0$, and all particles are on the energy level ε_1 :

$$a_1 = \infty, \quad a_n = 0 \text{ for } n > 1.$$

Hence with temperature at absolute zero, particles of the Bose gas are all on a quantum state with zero kinetic energy. This is BEC.

By the temperature formula (7.5.2) for a Bose gas system:

$$kT = \sum_n \left(1 + \frac{a_n}{g_n}\right) \frac{a_n \varepsilon_n}{N(1 + \beta_n \ln \varepsilon_n)}, \quad (8.3.42)$$

BEC can be viewed clearly as follows:

$$T = 0 \Leftrightarrow \varepsilon_1 = 0, \quad a_1 = N \quad \text{and} \quad a_n = 0 \quad \forall n > 1.$$

The first gaseous condensate was produced by Eric Cornell and Carl Wieman (Anderson et al., 1995) in 1995 in a gas of rubidium atoms cooled to $T = 1.7 \times 10^{-7}\text{ K}$. At the same time, Davis et al. (1995) developed a BEC using sodium. For their achievements Cornell, Wieman, and Ketterle received the 2001 Nobel Prize in Physics. Since then, BEC phenomena were also discovered in the following nine elements: Rb, Na, Li, ^1H , ^4He , ^{41}K , ^{52}Cr , ^{133}Cs , and ^{174}Yb .

In 1997, W. Ketterle's group discovered the first $J = 1$ spinor BEC, using optical trapping of condensates, opening a door for the study of spinor condensates.

BEC phenomena refers specifically to condensates of gaseous systems, and has quite different properties than solid condensates (superconductivity) and liquid condensates (superfluidity). The condensation of gaseous systems occurs at zeroth momentum level, corresponding to momentum condensation. BEC occurs for dilute gases at every extra-low temperatures. Systems with higher densities become liquid or solid states. Hence with temperature at absolute zero, a system can rarely be a gaseous state.

1. Thermodynamic Potential of Scalar BEC Systems In this case, the order parameters u and control parameters λ are

$$u = (\psi, \rho_n, S, \mathbf{H}), \quad \lambda = (p, T, \mathbf{H}_a).$$

The Gibbs free energy for scalar BEC is similar to (8.3.25) for ^4He superfluid systems. The ψ terms for BEC are different from (8.3.22), and are given by

$$\psi \text{ terms} = \int_{\Omega} \left[\frac{\hbar^2}{2m} |\nabla \psi|^2 + \frac{g_1}{2} |\psi|^4 \right] dx. \quad (8.3.43)$$

The reason is that due to the relative large distances between particles for dilute gases, the attracting interaction between particles as described by the g_0 term in (8.3.2) is negligible. However, the repulsive behavior due to the strong interaction is still present, i.e., $g_1 \neq 0$.

Also for a BEC system, the binding energy, which confines the atoms of the gases in a region is due to the applied magnetic field, called magnetic potential well.

As in (8.3.41), the Gibbs free energy for a BEC system contains an additional magnetic potential well energy

$$\text{magnetic potential well energy} \quad (8.3.44)$$

$$= \int_{\Omega} \left[\frac{1}{8\pi} \mathbf{H}^2 - \boldsymbol{\mu}_0 \cdot \mathbf{H} |\psi|^2 - \frac{1}{4\pi} \mathbf{H} \cdot \mathbf{H}_a \right] dx,$$

where $\boldsymbol{\mu}_0$ is the magnetization moment of atoms, representing the collective quantum effect of particles.

Besides (8.3.43) and (8.3.44), other terms of the free energy are similar to those in (8.3.21), (8.3.23), and (8.3.24) for ${}^4\text{He}$. Then we derive the Gibbs free energy for a BEC system as follows:

$$\begin{aligned} F = & \int_{\Omega} \left[\frac{\alpha}{2} |\nabla \rho_n|^2 + \frac{1}{2} (b^2 p + A_1 T) \rho_n^2 - b p \rho_n - \mu (\rho_n + |\psi|^2) \right. \\ & + \frac{\hbar^2}{2m} |\nabla \psi|^2 + \frac{g_1}{2} |\psi|^4 + g_2 b \rho_n |\psi|^2 \\ & + A_2 T (1 + b \rho_n + b |\psi|^2) \ln (1 + b \rho_n + b |\psi|^2) \\ & - \frac{1}{2\beta_0 k} T S^2 + \beta_1 T S \rho_n^2 - \beta_2 T S |\psi|^2 - S T \\ & \left. + \frac{1}{8\pi} \mathbf{H}^2 - \boldsymbol{\mu}_0 \cdot \mathbf{H} |\psi|^2 - \frac{1}{4\pi} \mathbf{H} \cdot \mathbf{H}_a \right] dx. \end{aligned} \quad (8.3.45)$$

2. *Thermodynamic Potential for $J = 1$ spinor BEC Systems* The binding energy for spinor BEC is induced by the applied electric field, and the order and control parameters u and λ for the system are

$$u = (\psi_+, \psi_0, \psi_-, \rho_n, S, \mathbf{E}), \quad \lambda = (p, T, \mathbf{E}_a),$$

where \mathbf{E}_a is the applied electric field, and \mathbf{E} is the induced electric field.

The Gibbs free energy for the system is similar to (8.3.41) for ${}^3\text{He}$. The differences are two-fold. First there is no attracting interaction energy between particles. Second, the electromagnetic binding energy is the following optical potential well energy (optical trapping energy), rather than the magnetic potential well energy (8.3.44):

$$\text{optical trapping energy} = \int_{\Omega} \left[\frac{1}{8\pi} \mathbf{E}^2 - \boldsymbol{\epsilon}_0 \cdot \mathbf{E} |\Psi|^2 - \frac{1}{4\pi} \mathbf{E} \cdot \mathbf{E}_a \right] dx, \quad (8.3.46)$$

where $\boldsymbol{\epsilon}_0$ is the polarization, representing the collective quantum effect of particles, and $\Psi = (\psi_+, \psi_0, \psi_-)$. Hence the Gibbs free energy for a spinor BEC system is

$$F = \int_{\Omega} \left[\frac{\alpha}{2} |\nabla \rho_n|^2 + \frac{1}{2} (b^2 p + A_1 T) \rho_n^2 - b p \rho_n - \mu (\rho_n + |\Psi|^2) \right.$$

$$\begin{aligned}
& + \frac{\hbar^2}{2m} |\nabla \Psi|^2 + \frac{g_1}{2} |\Psi|^4 + \frac{g_s}{2} |\Psi^\dagger \widehat{F} \Psi| + g_2 b \rho_n |\Psi|^2 \\
& + A_2 T (1 + b \rho_n + b |\Psi|^2) \ln(1 + b \rho_n + b |\Psi|^2) \\
& - \frac{T}{2\beta_0 k} S^2 + \beta_1 T S \rho_n^2 - \beta_2 T S |\Psi|^2 - ST \\
& + \frac{1}{8\pi} \mathbf{E}^2 - \boldsymbol{\varepsilon}_0 \cdot \mathbf{E} |\Psi|^2 - \frac{1}{4\pi} \mathbf{E} \cdot \mathbf{E}_a \Big] dx. \tag{8.3.47}
\end{aligned}$$

We remark that the optical potential well spinor BEC system bounds atoms with the electric potential. There are no constraints on the spin of the atoms so that the $N = 2J + 1 (J \neq 0)$ spin states are independent and Ψ is a spinor. In fact, this is also the consequence of the SO(3) spinor invariance of the functional (8.3.47).

3. *BEC Systems with Electromagnetic Potential Well* For a BEC system with spin $J \geq 1$ of atoms and with applied magnetic field \mathbf{H}_a and electric field \mathbf{E}_a as the binding energy, the order and control parameters u and λ are

$$u = (\Psi, \rho_n, \mathbf{H}, \mathbf{E}), \quad \lambda = (p, T, \mathbf{H}_a, S, \mathbf{E}_a),$$

where $\Psi = (\psi_J, \dots, \psi_0, \dots, \psi_{-J})$ has $2J + 1$ components. Then electromagnetic energy is

$$\begin{aligned}
\text{EM energy} &= \int_{\Omega} \left[\frac{1}{8\pi} \mathbf{H}^2 - \Psi^\dagger (\mathbf{H} \cdot \widehat{\mu}) \Psi - \frac{1}{4\pi} \mathbf{H} \cdot \mathbf{H}_a \right. \\
&\quad \left. + \frac{1}{8\pi} \mathbf{E}^2 - \boldsymbol{\varepsilon}_0 \cdot \mathbf{E} |\Psi|^2 - \frac{1}{4\pi} \mathbf{E} \cdot \mathbf{E}_a \right] dx, \tag{8.3.48}
\end{aligned}$$

where $\boldsymbol{\varepsilon}_0$ is the polarization, and $\widehat{\mu}$ is the atomic magnetic moment operator:

$$\widehat{\mu} = \boldsymbol{\mu}_0 + \mu_s \widehat{S}. \tag{8.3.49}$$

Here $\boldsymbol{\mu}_0$ is the orbital magnetic moment of the atom, μ_s the spin magnetic moment of the atom, and \widehat{S} is the spin operator; see the next section.

In view of (8.3.47) and (8.3.48), the Gibbs free energy for spin J BEC system with electromagnetic potential well is

$$\begin{aligned}
F &= \int_{\Omega} \left[\frac{\alpha}{2} |\nabla \rho_n|^2 + \frac{1}{2} (b^2 p + A_1 T) \rho_n^2 - b p \rho_n - \mu (\rho_n + |\Psi|^2) \right. \\
&\quad + \frac{\hbar^2}{2m} |\nabla \Psi|^2 + \frac{g_1}{2} |\Psi|^4 + \frac{g_s}{2} |\Psi^\dagger \widehat{S} \Psi|^2 + g_2 b \rho_n |\Psi|^2 \\
&\quad + A_2 T (1 + b \rho_n + b |\Psi|^2) \ln(1 + b \rho_n + b |\Psi|^2) \\
&\quad + \frac{1}{8\pi} \mathbf{H}^2 - \Psi^\dagger (\mathbf{H} \cdot \widehat{\mu}) \Psi - \frac{1}{4\pi} \mathbf{H} \cdot \mathbf{H}_a \\
&\quad \left. + \frac{1}{8\pi} \mathbf{E}^2 - \boldsymbol{\varepsilon}_0 \cdot \mathbf{E} |\Psi|^2 - \frac{1}{4\pi} \mathbf{E} \cdot \mathbf{E}_a \right] dx
\end{aligned}$$

$$-\frac{1}{2\beta_0 k} TS^2 + \beta_1 TS\rho_n^2 - \beta_2 TS|\Psi|^2 - ST \Big] dx. \quad (8.3.50)$$

Remark 8.3.5 The coupling term for the magnetic field $\Psi^\dagger(H \cdot \hat{\mu})\Psi$ in (8.3.50) breaks the symmetry of spinors for the system. When $\mathbf{H}_a = 0$, the symmetry is then restored. This is the magnetic screening phenomena, which can be more clearly viewed in the theory of phase transitions.

8.4 Quantum Systems of Condensates

8.4.1 Spin Operators of Spinors

As we mentioned earlier, the study of condensates can be classified into thermodynamic and quantum categories. Thermodynamic systems of condensates study transitions between normal and condensate states, and quantum system of condensates deal with the quantum behavior of fully developed condensates. For example, in the phase diagrams given in Figs. 8.3 and 8.4 for ${}^3\text{He}$ superfluid, when (T, p) are in the normal phase and A, A_1 phases, the superfluidity is of thermodynamic, and when (T, p) are region B, the system is a quantum system of condensates.

A quantum system of condensates is a conservative system with the Hamiltonian energy as its potential functional. The wave function Ψ is an N-component spinor, which depends on the spin J of the particles:

$$\Psi = (\psi_1, \dots, \psi_N), \quad N = 2J + 1, \quad (8.4.1)$$

where J is the spin of particles, which has $2J + 1$ quantum states, called magnetic quantum number m , and each component ψ_i in (8.4.1) represents the wave function of the i -th magnetic quantum state:

$$\begin{aligned} J = 0 : & \quad m = 0, \\ J = \frac{1}{2} : & \quad m = \frac{1}{2}, -\frac{1}{2}, \\ J = 1 : & \quad m = 1, 0, -1, \\ J = \frac{3}{2} : & \quad m = \frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}, \\ J = 2 : & \quad m = 2, 1, 0, -1, -2. \end{aligned} \quad (8.4.2)$$

Spin $J = 0$ corresponds to zeroth order spinors, and $J \neq 0$ represents J -spinors. For J -spinors, its Hamiltonian energy includes spin interaction energy. Hence we need to derive explicit expression for spin interaction energy.

1. *Spin Operators* For $J \neq 0$ spinors in (8.4.1), the spin operator \hat{J} is a vector operator:

$$\widehat{J} = (J_1, J_2, J_3),$$

where each component J_i is a Hermitian matrix of $N = 2J + 1$, and

$$\vec{J} = \int_{\Omega} \Psi^{\dagger} \widehat{J} \Psi dx \quad \text{represents total spin vector.} \quad (8.4.3)$$

The spin operator \widehat{J} enjoys the following properties:

- a. If x_3 is in the direction of the magnetic field, then the third component of \widehat{J} is

$$J_3 = \begin{pmatrix} J & 0 \\ \ddots & -J \\ 0 & -J \end{pmatrix} \quad (\text{tr} J_3 = 0), \quad (8.4.4)$$

where the entries on the diagonal of J_3 correspond to the magnetic quantum number in (8.4.2), as the total spin vector \vec{J} in (8.4.3) is parallel to the magnetic field \mathbf{H} .

- b. \widehat{J} is $SO(3)$ invariant. Namely, under

$$\tilde{x} = Ax, \quad A \in SO(3), \quad (8.4.5)$$

the spinor Ψ transforms as follows:

$$\tilde{\Psi} = U_A \Psi, \quad U_A \in SU(N) \text{ is the } N (= 2J + 1) \text{ spinor representation,}$$

then \widehat{J} satisfies

$$U_A^{\dagger} (a_{jk} J_j) U_A = J_k, \quad (8.4.6)$$

where $(a_{jk}) = A$ is the matrix in (8.4.5).

- 2. *Spin Interaction Energy* The spin of a particle induces magnetic moment, and in (4.4.1) we have

$$g_J \hbar \Psi^{\dagger} \widehat{J} \Psi \text{ represents spin magnetic moment,} \quad (8.4.7)$$

where g_J is the magnetic moment coefficient, depending on the types of particles, and corresponds to the Bohr magneton. Also the module square of magnetic moment is its energy density, and we have

$$\text{spin energy} = g_J^2 \hbar^2 \int_{\Omega} |\Psi^{\dagger} \widehat{J} \Psi|^2 dx. \quad (8.4.8)$$

The constant $g_s = 4g_J^2 \hbar^2$ is called spin coupling constant, as in (8.3.35). The expression (8.4.8) shows that once the spin operator \widehat{J} is given, the spin interaction energy is determined.

- 3. *Spin Operators for $J = \frac{1}{2}$* They are given by the Pauli matrices:

$$\widehat{\sigma} = (\sigma_1, \sigma_2, \sigma_3), \quad (8.4.9)$$

where σ_i are given by

$$\sigma_1 = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \frac{1}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Pauli matrices (8.4.9) satisfy the invariance of the SO(3) spinor representation in (7.6.21) and (7.6.22).

4. *Spin Operators for $J = 1$* They are given by (8.3.36).

8.4.2 Hamiltonian Energy of Superconductors

For a superconductor or a superfluid system, when $T \simeq 0$ K, the system is a quantum system of condensate. The order parameter for the system is the wave function. The temperature T , the pressure p , and entropy S are no longer the main parameters. The control parameters are the applied electromagnetic field and the geometric length scales of the system, and the potential functional is the Hamiltonian energy, whose expression plays a crucial role in condensed matter physics.

1. *Scalar Superconductors* In a condensate state, the order parameters of a scalar superconducting system is the scalar wave function $\psi : \Omega \rightarrow \mathbb{C}$ and the electromagnetic fields \mathbf{E} and \mathbf{A} :

$$\text{order parameters} = (\psi, \mathbf{E}, \mathbf{A}).$$

The control parameters are the applied electromagnetic fields \mathbf{E}_a and \mathbf{H}_a :

$$\lambda = (\mathbf{E}_a, \mathbf{H}_a).$$

The Hamiltonian energy \mathcal{H} of the system contains superconducting electron energy and the electromagnetic energy:

$$\mathcal{H} = \mathcal{H}_\psi + \mathcal{H}_{\text{EM}}. \quad (8.4.10)$$

By (8.3.12)–(8.3.14), the superconducting electron energy is

$$\mathcal{H}_\psi = \int_{\Omega} \left[\frac{1}{2m_s} \left| \left(-i\hbar\nabla - \frac{e_s}{c} \mathbf{A} \right) \psi \right|^2 - g_0 |\psi|^2 + \frac{g_1}{2} |\psi|^4 \right] dx. \quad (8.4.11)$$

The electromagnetic energy is

$$\begin{aligned} \mathcal{H}_{\text{EM}} = & \int_{\Omega} \left[\frac{1}{8\pi} \mathbf{E}^2 - \frac{1}{4\pi} \mathbf{E} \cdot \mathbf{E}_a - \boldsymbol{\epsilon} \cdot \mathbf{E} |\psi|^2 \right. \\ & \left. + \frac{1}{8\pi} \mathbf{H}^2 - \frac{1}{4\pi} \mathbf{H} \cdot \mathbf{H}_a - \boldsymbol{\mu} \cdot \mathbf{H} |\psi|^2 \right] dx, \end{aligned} \quad (8.4.12)$$

where (\mathbf{E}, \mathbf{H}) are the electromagnetic fields, $(\mathbf{E}_a, \mathbf{H}_a)$ are the applied electromagnetic fields, e_s is the charge of the Cooper pair, $\boldsymbol{\varepsilon}$ is the polarization, $\boldsymbol{\mu}$ is the induction magneton

$$\boldsymbol{\mu} = \frac{e_s \hbar}{m_s c} \mu_0 \hat{H}_a, \quad (8.4.13)$$

and \hat{H}_a is the unit vector in the direction of \mathbf{H}_a , and μ_0 is

$$\mu_0 = \frac{N^+ - N^-}{N}.$$

Here N^+, N^- represent the numbers of the Cooper pairs with spin $s = 1$ and $s = -1$, N is the total number of superconducting electrons. The control parameter λ is

$$\lambda = (\mathbf{E}_a, \mathbf{H}_a). \quad (8.4.14)$$

By (8.4.11)–(8.4.13), the Hamiltonian energy (8.4.10) is written as

$$\begin{aligned} \mathcal{H} = \int_{\Omega} \left[\frac{1}{2m_s} \left| \left(-i\hbar \nabla - \frac{e_s}{c} \mathbf{A} \right) \psi \right|^2 - g_0 |\psi|^2 + \frac{g_1}{2} |\psi|^4 \right. \\ \left. + \frac{1}{8\pi} \mathbf{E}^2 - \frac{1}{4\pi} \mathbf{E} \cdot \mathbf{E}_a - \boldsymbol{\varepsilon} \cdot \mathbf{E} |\psi|^2 \right. \\ \left. + \frac{1}{8\pi} \mathbf{H}^2 - \frac{1}{4\pi} \mathbf{H} \cdot \mathbf{H}_a - \boldsymbol{\mu} \cdot \mathbf{H} |\psi|^2 \right] dx. \end{aligned} \quad (8.4.15)$$

Remark 8.4.1 In (8.4.15), if the total spin of \mathbf{S} is in the reversed direction of \mathbf{H}_a , i.e., $\mu_0 = (N^+ - N^-)/N < 0$, then the superconducting magnetic moment can negate the applied field:

$$\boldsymbol{\mu} |\psi|^2 \text{ negates } \frac{1}{4\pi} \mathbf{H}_a.$$

This is the Meissner effect. In particular, if

$$\boldsymbol{\mu} |\psi|^2 = -\frac{1}{4\pi} \mathbf{H}_a,$$

then a superconductor becomes a perfect diamagnetic system. \square

2. *Spinor Superconductors* In general, the superconducting Cooper electron-pairs have spin zero. However, under magnetic field, nonzero spin Cooper electron-pairs may occur, leading to spinor superconducting electron-pairs, whose spin arrangements are given by

$$J = 1 : |\uparrow\uparrow\rangle, \quad J = -1 : |\downarrow\downarrow\rangle, \quad J = 0 : |\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle.$$

The corresponding wave functions are :

$$\psi_+ = |\uparrow\uparrow\rangle, \quad \psi_0 = \frac{1}{\sqrt{2}} (|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle), \quad \psi_- = |\downarrow\downarrow\rangle.$$

We then have a three-component spinor wave function:

$$\Psi = (\psi_+, \psi_0, \psi_-). \quad (8.4.16)$$

The Hamiltonian energy \mathcal{H} of the spinor consists of three parts:

$$\mathcal{H} = \mathcal{H}_\Psi + \mathcal{H}_{\text{spin}} + \mathcal{H}_{\text{EM}}.$$

They are given respectively as follows:

$$\begin{aligned} \mathcal{H}_\Psi &= \int_{\Omega} \left[\frac{1}{2m_s} \left| \left(-i\hbar\nabla - \frac{e_s}{c}\mathbf{A} \right) \Psi \right|^2 - g_0 |\psi_0|^2 \right. \\ &\quad \left. - g_{\pm} (|\psi_+|^2 + |\psi_-|^2) + \frac{g_1}{2} |\Psi|^4 \right] dx, \end{aligned} \quad (8.4.17)$$

$$\mathcal{H}_{\text{spin}} = \int_{\Omega} \frac{g_s}{2} |\Psi^\dagger \hat{F} \Psi|^2 dx, \quad (8.4.18)$$

$$\begin{aligned} \mathcal{H}_{\text{EM}} &= \int_{\Omega} \left[\frac{1}{8\pi} (\mathbf{E}^2 + \mathbf{H}^2) - \frac{1}{4\pi} \mathbf{E} \cdot \mathbf{E}_a - \boldsymbol{\epsilon} \cdot \mathbf{E} |\psi|^2 \right. \\ &\quad \left. - \frac{1}{4\pi} \mathbf{H} \cdot \mathbf{H}_a + \frac{e_s \hbar}{m_s c} \Psi^\dagger \mathbf{H} \cdot \hat{F} \Psi \right] dx. \end{aligned} \quad (8.4.19)$$

Here \hat{F} in (8.4.18) is the spin operator defined by (8.3.36).

By (8.4.17)–(8.4.19), the Hamiltonian energy of a spinor superconducting system is

$$\begin{aligned} \mathcal{H} &= \int_{\Omega} \left[\frac{1}{2m_s} \left| \left(-i\hbar\nabla - \frac{e_s}{c}\mathbf{A} \right) \Psi \right|^2 - g_0 |\psi_0|^2 \right. \\ &\quad \left. - g_{\pm} (|\psi_+|^2 + |\psi_-|^2) + \frac{g_1}{2} |\Psi|^4 + \frac{g_s}{2} |\Psi^\dagger \hat{F} \Psi|^2 \right. \\ &\quad \left. + \frac{1}{8\pi} (\mathbf{E}^2 + \mathbf{H}^2) - \frac{1}{4\pi} \mathbf{E} \cdot \mathbf{E}_a - \boldsymbol{\epsilon} \cdot \mathbf{E} |\psi|^2 \right. \\ &\quad \left. - \frac{1}{4\pi} \mathbf{H} \cdot \mathbf{H}_a + \frac{e_s \hbar}{m_s c} \Psi^\dagger \mathbf{H} \cdot \hat{F} \Psi \right] dx. \end{aligned} \quad (8.4.20)$$

8.4.3 Superfluid Systems

Superfluid systems include liquid ${}^4\text{He}$ and liquid ${}^3\text{He}$ systems. ${}^4\text{He}$ are made up of bosons and can condense without forming Cooper atoms. Also, the electric dipole moment of induced by ${}^4\text{He}$ atoms is small, indicating that the total spin of the ground state electrons is zero, and the total spin of the nuclei is also zero. Hence an applied electromagnetic field has negligible effect on ${}^4\text{He}$ superfluid.

Atoms of ^3He are Fermions, and need to form Cooper pairs to condensate. Also, the spin arrangements of the pairs may lead to magnetic moments strong enough as depicted in (8.4.17).

In summary, superfluid systems are less sensitive to the applied electric fields, since the two electrons of a helium atom are on an orbital ground state without electric polarization, very different from Alkali metals. In view of these general properties, we now introduce, respectively, the Hamiltonians of liquid ^4He superfluid systems, liquid ^3He without magnetic fields, and liquid ^3He superfluid systems with the presence of magnetic fields.

1. ^4He Superfluidity This is a scalar system with a single component complex-valued wave function $\psi : \Omega \rightarrow \mathbb{C}$, and without coupling with electromagnetic fields. Its Hamiltonian is as (8.4.11):

$$\mathcal{H} = \int_{\Omega} \left[\frac{\hbar^2}{2m} |\nabla \psi|^2 - g_0 |\psi|^2 + \frac{g_1}{2} |\psi|^4 \right] dx, \quad (8.4.21)$$

which is also called the Gross-Pitaevskii energy functional.

2. ^3He Superfluid Systems Without Magnetic Fields This is a three-component spinor system with spinor wave function

$$\Psi = (\psi_+, \psi_0, \psi_-). \quad (8.4.22)$$

Its Hamiltonian energy is the Ψ part of the Gibbs free energy (8.3.37) for ^3He superfluid system:

$$\begin{aligned} \mathcal{H} = \int_{\Omega} & \left[\frac{\hbar^2}{2m} |\nabla \Psi|^2 - g_0 |\psi_0|^2 - g_{\pm} (|\psi_+|^2 + |\psi_-|^2) \right. \\ & \left. + \frac{g_1}{2} |\Psi|^4 + \frac{g_s}{2} |\Psi^\dagger \hat{F} \Psi|^2 \right] dx, \end{aligned} \quad (8.4.23)$$

where the spin operator \hat{F} is given by (8.3.36), and the attracting interaction potentials g_0 and g_{\pm} satisfy

$$\begin{aligned} g_0 &\neq g_{\pm} && \text{phase A in Fig. 8.3,} \\ g_0 &= g_{\pm} && \text{phase B region in Fig. 8.3.} \end{aligned}$$

3. ^3He Superfluid Systems Coupling Magnetic Fields Since the atom of ^3He is electric neutral, the gradient operator is the usual ∇ , which is different from the superconducting systems. The wave function is as in (8.4.22), and the Hamiltonian energy contains the energy involving Ψ and the magnetic energy:

$$\mathcal{H} = \mathcal{H}_{\Psi} + \mathcal{H}_M, \quad (8.4.24)$$

where \mathcal{H}_{Ψ} is the same as (8.4.23), and \mathcal{H}_M is the magnetic field portion of the electromagnetic energy in (8.4.19):

$$\mathcal{H}_M = \int_{\Omega} \left[\frac{1}{8\pi} \mathbf{H}^2 - \frac{1}{4\pi} \mathbf{H} \cdot \mathbf{H}_a - \mu_0 \Psi^\dagger \mathbf{H} \cdot \widehat{\mathbf{F}} \Psi \right] dx. \quad (8.4.25)$$

Hence the Hamiltonian energy is

$$\begin{aligned} \mathcal{H} = & \int_{\Omega} \left[\frac{\hbar^2}{2m} |\nabla \Psi|^2 - g_0 |\psi_0|^2 - g_{\pm} (|\psi_+|^2 + |\psi_-|^2) \right. \\ & \left. + \frac{g_1}{2} |\Psi|^4 + \frac{g_s}{2} |\Psi^\dagger \widehat{\mathbf{F}} \Psi|^2 + \frac{1}{8\pi} \mathbf{H}^2 - \frac{1}{4\pi} \mathbf{H} \cdot \mathbf{H}_a - \mu_0 \Psi^\dagger \mathbf{H} \cdot \widehat{\mathbf{F}} \Psi \right] dx, \end{aligned} \quad (8.4.26)$$

where μ_0 is the magnetic moment of the ${}^3\text{He}$ Cooper atom-pairs, and the attracting interaction potentials g_0 and g_{\pm} enjoy

$$\begin{aligned} g_0 \neq g_{\pm} & \quad \text{phases A and } A_1 \text{ regions in Fig. 8.4,} \\ g_0 = g_{\pm} & \quad \text{phase B region in Fig. 8.4.} \end{aligned}$$

8.4.4 Gaseous BEC Systems

BEC is a condensate system, different from superconducting and superfluid systems. BEC occurs at the condensate state with zero kinetic energy, and superconductor and superfluid condensate occur at nonzero lowest energy levels.

There are three types of BEC systems: scalar systems, optical potential well spinor systems, and electromagnetic potential well spinor systems with $J \geq 1$.

- Gross-Pitaevskii Energy for Scalar Systems* For a scalar such system, the atoms of the dilute gases are confined by the magnetic potential well, leading to a condensate state. Its Hamiltonian energy is the Gross-Pitaevskii functional (Gross, 1961; Pitaevskii, 1961):

$$\mathcal{H} = \int_{\Omega} \left[\frac{\hbar^2}{2m} |\nabla \psi|^2 - V(x) |\psi|^2 + \frac{g_1}{2} |\psi|^4 \right] dx, \quad (8.4.27)$$

which has the same form as (8.4.21) for ${}^4\text{He}$, but with different binding potential. In a superfluid system, g_0 is the attracting potential between particles, but in (8.4.27), $V(x)$ is the binding potential induced by the external magnetic field. Also g_1 here takes the form

$$g_1 = \frac{4\pi\hbar^2 a}{m}, \quad (8.4.28)$$

where a is the length of the s -wave scattering.

- Ho-Ohmi-Machida Energy for Optical Potential Well Spinor Systems* Atoms of such a dilute gas system are binding together by an electric potential. The Hamiltonian energy for such a BEC system with $J = 1$ is derived by Ho (1998), Ohmi and Machida (1998) in the following form, which is SO(3) spinor invariant:

$$\mathcal{H} = \int_{\Omega} \left[\frac{\hbar^2}{2m} |\nabla \Psi|^2 - V|\Psi|^2 + \frac{g_1}{2} |\Psi|^4 + \frac{g_s}{2} |\Psi^\dagger \widehat{F} \Psi|^2 \right] dx, \quad (8.4.29)$$

where $\Psi = (\psi_+, \psi_0, \psi_-)$, \widehat{F} is given by (8.3.36), and

$$g_1 = \frac{4\pi\hbar^2}{m} \frac{a_0 + 2a_1}{3}, \quad g_s = \frac{4\pi\hbar^2}{m} \frac{a_1 - a_0}{3}. \quad (8.4.30)$$

Here a_0 , a_1 are the lengths of s-wave scattering for $J = 0$ and $J = 2$ respectively, and

$$\begin{aligned} g_s < 0 &\quad \text{for ferromagnetic,} \\ g_s > 0 &\quad \text{for anti-ferromagnetic.} \end{aligned} \quad (8.4.31)$$

8.5 Field Equations for Condensates

8.5.1 Field Equations of Superconductivity

The field equations governing condensate systems are (8.1.3)–(8.1.5), determined by the principle of Hamiltonian dynamics (PHD) and the principle of Lagrangian dynamics (PLD). For a scalar superconductor, the Hamiltonian energy functional is given by (8.4.15), and the field equations are written in the form

$$i\hbar \frac{\partial \psi}{\partial t} = \frac{1}{2m_s} \left(i\hbar \nabla + \frac{e_s}{c} \mathbf{A} \right)^2 \psi - V\psi + g|\psi|^2\psi - (\boldsymbol{\mu} \cdot \mathbf{H} + \boldsymbol{\epsilon} \cdot \mathbf{E})\psi, \quad (8.5.1)$$

together with the electromagnetic equilibrium equations

$$\frac{\delta}{\delta \mathbf{E}} \mathcal{H} = 0, \quad \text{and} \quad \frac{\delta}{\delta \mathbf{H}} \mathcal{H} = 0,$$

which are expressed as

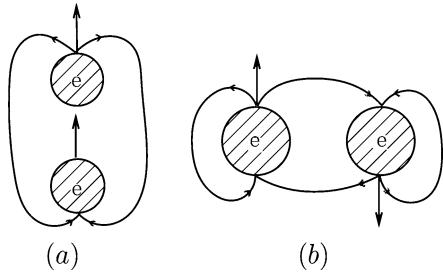
$$\begin{aligned} \mathbf{E} &= \mathbf{E}_a + 4\pi|\psi|^2\boldsymbol{\epsilon}, \\ \mathbf{H} &= \mathbf{H}_a + 4\pi|\psi|^2\boldsymbol{\mu}, \end{aligned} \quad (8.5.2)$$

Remark 8.5.1 For the magnetic field \mathbf{H} in (8.5.2) for a superconductor, we see that

$$\mathbf{H} = 0 \iff \mathbf{H}_a = -4\pi|\psi|^2\boldsymbol{\mu}.$$

This is the Meissner effect. It shows that the demagnetic effect is caused by the Cooper pairs with spin $s = -1$, i.e., their orientation is reverse to \mathbf{H}_a . The spins of Cooper pairs as shown in Fig. 8.5. \square

Fig. 8.5 Arrangements of Cooper pairs: (a) collinear orientation with spin $J = 1$; (b) anti-parallel orientation with spin $J = 0$; they lead to attractive magnetic moment force



8.5.2 Field Equations of Superfluidity and BEC

1. ^4He Superfluid and Scalar BEC Systems The Hamiltonian energy for ^4He superfluid and scalar BEC systems is the Gross-Pitaevskii functional (8.4.21) and (8.4.27), and the Lagrangian dynamic equation is given by

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi - V\psi + g|\psi|^2\psi. \quad (8.5.3)$$

2. ^3He Systems Without Applied Magnetic Fields and Spinor BEC Systems These are spinor systems with 3-component wave function

$$\Psi = (\psi_+, \psi_0, \psi_-), \quad (8.5.4)$$

where ψ_+, ψ_0, ψ_- represent the particles with spin $s = 1, 0, -1$ respectively. The energy functional is given by (8.4.23), which can also be written as

$$\mathcal{H} = \int_{\Omega} \left[\frac{\hbar^2}{2m} |\nabla \Psi|^2 - \Psi^\dagger V \Psi + \frac{g}{2} |\Psi|^4 + \frac{g_s}{2} |\Psi^\dagger \hat{F} \Psi| \right] dx, \quad (8.5.5)$$

where V is the matrix representing bounding potential, and $g, g_s > 0$ are coupling parameters. The Lagrangian field equation for (8.5.5) is

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \Delta \Psi - V\Psi + g|\Psi|^2\Psi + g_s (\Psi^\dagger \hat{F} \Psi) \cdot \hat{F} \Psi, \quad (8.5.6)$$

where V is the matrix as

$$\begin{pmatrix} V_+ & 0 \\ 0 & V_0 \\ 0 & V_- \end{pmatrix} \quad (8.5.7)$$

and V_+, V_0, V_- are the bounding potential acting on ψ_+, ψ_0, ψ_- respectively. The control parameter is $\lambda = (V_+, V_0, V_-)$.

3. ^3He Superfluid Systems with Applied Magnetic Fields When an external magnetic field \mathbf{H}_a is exerted, the energy functional of a ^3He superfluid system is given by (8.4.26), rewritten as

$$\begin{aligned}\mathcal{H} = \int_{\Omega} & \left[\frac{\hbar^2}{2m} |\nabla \Psi|^2 - \Psi^+ V \Psi + \frac{g}{2} |\Psi|^4 + \frac{g_s}{2} |\Psi^\dagger \hat{F} \Psi|^2 \right. \\ & \left. + \frac{1}{8\pi} \mathbf{H}^2 - \frac{1}{4\pi} \mathbf{H} \cdot \mathbf{H}_a - \mu_0 \Psi^\dagger (\mathbf{H} \cdot \hat{F}) \Psi \right] dx.\end{aligned}\quad (8.5.8)$$

The Lagrangian system (8.1.4) for the Hamiltonian (8.5.8) is in the form

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \delta \Psi - V \Psi + g |\Psi|^2 \Psi + g_s (\Psi^+ \hat{F} \Psi) \cdot \hat{F} \Psi - \mu_0 (\mathbf{H} \cdot \hat{F}) \Psi,\quad (8.5.9)$$

with the magnetic field equation $\frac{\delta}{\delta H} \mathcal{H} = 0$ as

$$\mathbf{H} = \mathbf{H}_a + 4\pi \mu_0 (\mathbf{H} \cdot \hat{F}) \Psi.\quad (8.5.10)$$

The control parameter of the system (8.5.9)–(8.5.10) is

$$\lambda = (V_0, V_1, V_2, \mathbf{H}_a).$$



Chapter 9

Topological Phase Transitions

As mentioned in the preface to this second edition, all phase transitions in Nature that we have encountered can be categorized into the following two types:

1. dynamical phase transitions, and
2. topological phase transitions (TPTs), also called the pattern formation transitions.

This chapter aims to develop a systematic theory of TPTs and explores a few typical examples, including

1. quantum phase transitions (QPTs),
2. galactic spiral structures,
3. electromagnetic eruptions on solar surface,
4. boundary-layer separation of fluid flows, and
5. interior separation of fluid flows.

9.1 Topological Theory of 2D Incompressible Flows

In this section, we recapitulate the geometric theory of incompressible flows developed by the authors to study the structure and its stability and transitions of incompressible fluid flows in the physical spaces. The complete account of this geometric theory is given in the authors' research monograph (Ma and Wang, 2005d). This geometric theory is then directly applied to study the TPTs associated with the quantum phase transitions and the boundary-layer and interior separations of both classical and geophysical fluid flows later in this chapter.

9.1.1 Basic Concepts

The state function describing the fluid motion is the velocity field u . The topological phase transition of a fluid system is defined as the change in its topological structure

of the velocity field $u(x, \lambda)$ at a critical parameter λ_c , where λ is the time or other physical control parameters.

Let v be a vector field defined on domain $\Omega \subset \mathbb{R}^2$. For each point $x_0 \in \Omega$, v possesses an orbit $x(t, x_0)$ passing through x_0 , which is a solution of the ordinary differential equation with x_0 as its initial value:

$$\frac{dx}{dt} = v(x), \quad x(0) = x_0.$$

The set of all orbits is called the flow of v . Each vector field has its own flow structure, called the topological structure of v . Therefore, we can introduce the notion of topological equivalence for two vector fields.

Definition 9.1.1 *Let v_1 and v_2 be two vector fields in Ω . We say that v_1 and v_2 are topologically equivalent if there exists a homeomorphism $\varphi : \Omega \rightarrow \Omega$ that takes the orbits of v_1 to the orbits of v_2 , preserving orientation.*

The main aim is to study the structure transitions of 2D incompressible fluid flows represented by velocity fields u , which are the solutions of the fluid dynamical equations. To this end, let $C^r(\Omega, \mathbb{R}^2)$ be the space of all r -th order continuously differentiable 2D vector fields on Ω , and let

$$\begin{aligned} D^r(\Omega, \mathbb{R}^2) &= \{v \in C^r(\Omega, \mathbb{R}^2) \mid \operatorname{div} v = 0, \quad v_n = 0 \text{ on } \partial\Omega\}, \\ B^r(\Omega, \mathbb{R}^2) &= \{v \in D^r(\Omega, \mathbb{R}^2) \mid \frac{\partial v_\tau}{\partial n} = 0 \text{ on } \partial\Omega\}, \\ B_0^r(\Omega, \mathbb{R}^2) &= \{v \in D^r(\Omega, \mathbb{R}^2) \mid v = 0 \text{ on } \partial\Omega\}, \end{aligned}$$

where $v_n = v \cdot n$, $v_\tau = v \cdot \tau$, and n , τ are the unit normal and tangent vectors on $\partial\Omega$. The vector fields in $B^r(\Omega, \mathbb{R}^2)$ satisfy the free-slip boundary condition, and vector fields in $B_0^r(\Omega, \mathbb{R}^2)$ satisfy the rigid boundary condition.

Definition 9.1.2 *Let X be either $B^r(\Omega, \mathbb{R}^2)$ or $B_0^r(\Omega, \mathbb{R}^2)$. A vector field $v \in X$ is called structurally stable in X , if there exists an open neighborhood $U \subset X$ of v such that for any $v_1 \in U$, v and v_1 are topologically equivalent.*

9.1.2 Structural Stability Theorems

In Ma and Wang (2005d), the authors established the geometric theory for 2D divergence-free vector fields, including the structural stability theorems, the boundary-layer and the interior separation theory. In this section, we recapitulate the two structural stability results respectively for $v \in B^r(\Omega, \mathbb{R}^2)$ and $v \in B_0^r(\Omega, \mathbb{R}^2)$, which lay the needed mathematical foundation for the topological phase transitions studied in this chapter.

1. *Structural Stability in $B^r(\Omega, \mathbb{R}^2)$* The vector fields $v \in B^r(\Omega, \mathbb{R}^2)$ satisfy the free-slip boundary condition, given by

$$v_n|_{\partial\Omega} = 0, \quad \left. \frac{\partial v_\tau}{\partial n} \right|_{\partial\Omega} = 0. \quad (9.1.1)$$

A point $p \in \Omega$ is called a nondegenerate zero point (or singular point) of v if $v(p) = 0$, and the Jacobian matrix $Dv(p)$ is nondegenerate. A vector field v is regular if all zero points of v are nondegenerate. For the vector fields with condition (9.1.1), we have the following structural stability theorem.

Theorem 9.1.3 (Ma and Wang 2005d, Theorem 2.1.2) *Let $\Omega \subset \mathbb{R}^2$ be a bounded domain. A vector field $v \in B^r(\Omega, \mathbb{R}^2)$ ($r \geq 1$) is structurally stable if and only if*

- a. v is regular;
- b. all interior saddles of v are self-connected; and
- c. each boundary saddle is connected to a boundary saddle on the same connected component of the boundary.

Moreover, all structurally stable vector fields in $B^r(\Omega, \mathbb{R}^2)$ form an open and dense set in $B^r(\Omega, \mathbb{R}^2)$.

2. *Structural Stability in $B_0^r(\Omega, \mathbb{R}^2)$* A vector field $v \in B_0^r(\Omega, \mathbb{R}^2)$ ($r \geq 1$) satisfies the rigid boundary condition, also called the Dirichlet boundary condition:

$$v|_{\partial\Omega} = 0. \quad (9.1.2)$$

With condition (9.1.2), all boundary points are singular in the usual sense. Hence we need to introduce the ∂ -regular and the ∂ -singular points for $p \in \partial\Omega$.

Definition 9.1.4 *Let $v \in B_0^r(\Omega, \mathbb{R}^2)$.*

1. *A point $p \in \partial\Omega$ is called a ∂ -regular point of v if*

$$\frac{\partial v_\tau(p)}{\partial n} \neq 0;$$

otherwise, $p \in \partial\Omega$ is called a ∂ -singular point;

- 2) *a ∂ -singular point p of v is called nondegenerate if*

$$\det \begin{pmatrix} \frac{\partial^2 v_\tau(p)}{\partial n \partial \tau} & \frac{\partial^2 v_\tau(p)}{\partial n^2} \\ \frac{\partial^2 v_n(p)}{\partial n \partial \tau} & \frac{\partial^2 v_n(p)}{\partial n^2} \end{pmatrix} \neq 0.$$

and a nondegenerate ∂ -singular point is also called ∂ -saddle of v ; and

2. *a vector field $v \in B_0^r(\Omega, \mathbb{R}^2)$ is said D-regular, if v is regular in the interior of Ω and all ∂ -singular points are nondegenerate.*

Let $v \in B_0^r(\Omega, \mathbb{R}^2)$ be D-regular, then the number of ∂ -saddles of v is finite, and there is only one orbit connected to a ∂ -saddle from the interior. In particular, no orbits are connected to a ∂ -singular point.

We have the following structural stability theorem for incompressible flows with the Dirichlet boundary condition (9.1.2).

Theorem 9.1.5 (Ma and Wang 2005d, Theorem 2.2.9) *Let $\Omega \subset \mathbb{R}^2$ be a bounded domain. Then a vector field $v \in B_0^r(\Omega, \mathbb{R}^2)$ ($r \geq 2$) is structurally stable if and only if*

1. *v is D-regular;*
2. *all interior saddles of v are self-connected; and*
3. *each ∂ -saddle of v is connected to a ∂ -saddle on the same connected component of the boundary.*

Moreover, all structurally stable vector fields in $B_0^r(\Omega, \mathbb{R}^2)$ form an open and dense set in $B_0^r(\Omega, \mathbb{R}^2)$.

9.1.3 Structural Bifurcations on Boundary

Let $t \in [0, T]$ be the time parameter, or another physical parameter, and u be a family of vector fields with t as parameter:

$$u : [0, T] \rightarrow X \quad \text{for } 0 < T < \infty, \quad (9.1.3)$$

where X is $B^r(\Omega, \mathbb{R}^2)$ or $B_0^r(\Omega, \mathbb{R}^2)$. Let $C^k([0, T], X)$ be the space of all one-parameter family of vector fields $u(t)$ as in (9.1.3), where $k \geq 0$ is the order of continuous derivatives of u with respect to t .

Definition 9.1.6 *Let $u \in C^0([0, T], X)$ be a one-parameter family of vector fields in X . We say that $u(x, t)$ has a structural bifurcation at t_0 ($0 < t_0 < T$), if for any $t^- < t_0$ and $t_0 < t^+$ with t^- and t^+ sufficiently close to t_0 the vector field $u(\cdot, t^-)$ is not topologically equivalent to $u(\cdot, t^+)$.*

We remark that the structural stability theorems, Theorems 9.1.3 and 9.1.5, ensure the rationality of Definition 9.1.6, i.e., the bifurcation point t_0 is isolated.

In the following we introduce the structural bifurcation theorems on the boundary for vector fields in $B^r(\Omega, \mathbb{R}^2)$ or $B_0^r(\Omega, \mathbb{R}^2)$.

1. *Structural Bifurcations for Free Boundary Condition* Let $u \in C^1([0, T], B^r(\Omega, \mathbb{R}^2))$. Take the first-order Taylor expression of $u(x, t)$ at $t_0 \in \partial\Omega$ as

$$\begin{aligned} u(x, t) &= u^0(x) + (t - t_0)u^1(x) + o(|t - t_0|), \\ u^0(x) &= u(x, t_0), \\ u^1(x) &= \frac{\partial u}{\partial t}(x, t_0). \end{aligned} \quad (9.1.4)$$

For the vector fields u^0 and u^1 in (9.1.4), we make the following assumption. Let $\bar{x} \in \partial\Omega$ satisfy

$$\begin{aligned}
 u^0(\bar{x}) &= 0, \text{ and } \bar{x} \text{ is isolated singular point,} \\
 u_\tau^1(\bar{x}) &\neq 0, \\
 \text{ind}(u^0, \bar{x}) &\neq -\frac{1}{2}.
 \end{aligned} \tag{9.1.5}$$

Here $\text{ind}(u^0, \bar{x})$ is the Poincaré index of u^0 at \bar{x} , defined by

$$\text{ind}(u^0, \bar{x}) = -\frac{n}{2} \quad (n = 0, 1, 2, \dots),$$

where n is the number of interior orbits of u^0 connected to \bar{x} .

Theorem 9.1.7 (Structural Bifurcation for Free Boundary Condition) *Let $u \in C^1([0, T], B^r(\Omega, \mathbb{R}^2))$ have the Taylor expression (9.1.4) at $t_0 > 0$, and for $\bar{x} \in \partial\Omega$ satisfy condition (9.1.5). Then $u(x, t)$ has a structural bifurcation at (\bar{x}, t_0) .*

2. Structural Bifurcation for Rigid Boundary Condition Let $u \in C^1([0, T], B_0^r(\Omega, \mathbb{R}^2))$ have the Taylor expression (9.1.4). For u^0 and u^1 in (9.1.4) we assume that

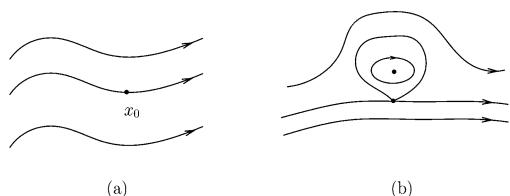
$$\begin{aligned}
 \frac{\partial u^0(\bar{x})}{\partial n} &= 0 \text{ and } \bar{x} \in \partial\Omega \text{ is an isolated singular point,} \\
 \frac{\partial u_\tau^1(\bar{x})}{\partial n} &\neq 0, \\
 \text{ind}\left(\frac{\partial u^0}{\partial n}, \bar{x}\right) &\neq -\frac{1}{2}.
 \end{aligned} \tag{9.1.6}$$

Theorem 9.1.8 (Structural Bifurcation for Rigid Boundary Condition) *Let $u \in C^1([0, T], B_0^r(\Omega, \mathbb{R}^2))$ have the Taylor expression (9.1.4) and satisfy condition (9.1.6). Then u has a structural bifurcation at (\bar{x}, t_0) .*

9.1.4 Interior Separation

Definition 9.1.9 We say that a family of 2D incompressible vector fields $u(x, \lambda)$ has an interior separation from (x_0, λ_0) with $x_0 \in \Omega$, if $u(x, \lambda)$ is locally topologically equivalent to the structure of Fig. 9.1a near x_0 for any $\lambda < \lambda_0$, and to the structure of Fig. 9.1b for $\lambda_0 < \lambda$.

Fig. 9.1 A schematic diagram of interior separation, (a) a parallel flow near x_0 , and (b) a vortex flow



Usually, the parameter λ is taken as the time t . Let $u(x, t)$ be the first-order Taylor expansion with respect to t at $t_0 > 0$:

$$\begin{aligned} u(x, t) &= u^0(x) + (t - t_0)u^1(x) + o(|t - t_0|), \\ u^0(x) &= u(x, t_0), \\ u^1(x) &= \left. \frac{\partial u}{\partial t} \right|_{t=t_0}. \end{aligned} \tag{9.1.7}$$

Let $x_0 \in \Omega$ be a degenerate singular point of u^0 , and the Jacobian $Du^0(x_0) \neq 0$. Then, there are two unit vectors e_1 and e_2 orthogonal to each other, satisfying that

$$\begin{aligned} Du^0(x_0)e_1 &= 0, \\ Du^0(x_0)e_2 &= \alpha e_1 \quad \text{with } \alpha \neq 0. \end{aligned} \tag{9.1.8}$$

Theorem 9.1.10 (Interior Separation (Ma and Wang, 2005d)) *Let $u(x, t)$ be a one-parameter family of 2D divergence-free vector fields, and have the Taylor expansion (9.1.7) at $t = t_0$. If $x_0 \in \Omega$ is an isolated singular point if $u^0(x)$ with $Du^0(x_0) \neq 0$, and satisfies that*

$$\begin{aligned} \text{ind}(u^0, x_0) &= 0, \\ u^1(x_0) \cdot e_2 &\neq 0, \end{aligned} \tag{9.1.9}$$

where e_2 is the vector as in (9.1.8), then $u(x, t)$ has an interior separation from (x_0, t_0) .

Theorem 9.1.10 holds also true for $u(x, \lambda)$, where λ is another parameter instead of the time t .

9.2 Quantum Phase Transitions

The aim of this section is to provide a precise mathematical description and theory for quantum phase transitions (QPTs) associated with the Bose–Einstein condensates, the superfluidity, and the superconductivity. The study presented here is based on Ma and Wang (2017e).

9.2.1 Definition and Examples

Classical QPTs

To the best knowledge of the authors, there is no clear definition for QPTs in classical statistical physics. Usually, a QPT is regarded as a transition with the following three basic characteristics:

1. the transition occurs with temperature at or near absolute zero;
2. the control parameters are nonthermal; and
3. the change of the physical states at a threshold is induced by quantum fluctuations, rather than thermal fluctuations.

Here the quantum fluctuations mean that they are caused by the Heisenberg uncertainty relation.

A popular description for a QPT is given in Fig. 9.2; see among others (Vojta, 2003). In this diagram, the horizontal axis represents the control parameter λ used to tune the system through the QPT, and the vertical axis is the temperature. In Fig. 9.2a, the solid line separates the region into two parts D and E, where D is the region of ordered phase, E is the region of disordered phase, and the QPT occurs between the phases. In Fig. 9.2b, the region E is separated into three parts A, B, C by the dashed lines, where A represents the domain where the states are controlled by quantum fluctuations, C is the domain by thermal fluctuations, and B is the quantum critical region where both types of fluctuations are important.

Fig. 9.2 A QPT

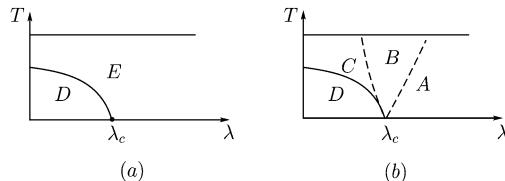


Figure 9.2 offers the physical definition of the classical QPT: There are two ways that result in the transition, the first one is from region C to D, which is the thermodynamical phase transition, and the second one is from A through B to D, which is called the QPT.

Hence, from Fig. 9.2 and the three basic characteristics of QPTs listed in the beginning of this section, we arrive at the following classical criterion for QPT:

$$\text{the QPT is caused by quantum fluctuations.} \quad (9.2.1)$$

Definition of QPTs

We note that the classical criterion of QPT is given by (9.2.1). However, it is hard to distinguish both thermal fluctuations and quantum fluctuations. In particular, we know that the reason causing the instability of a system is the state functions lose their stability, rather than the fluctuations, which only make the system deviate from the unstable states. Hence, the fluctuations are not the main characters of phase transitions, and cannot be used to define QPTs.

A precise definition is crucial to establish a QPT theory. Based on the condensation theory developed in Sect. 8.1, a better description for QPTs can be derived. First, a QPT should possess the following three physical characteristics:

Characteristics of QPT 9.2.1

1. A QPT is a transition between quantum states, and the state quantities of the quantum system describing the transition should be the wave functions ψ of the quantum states;
2. the control parameters are nonthermal; and
3. since the dynamical equations (9.3.15) of a QPT system are energy conserved, a QPT is a topological phase transition of a condensate system, rather than a dynamical phase transition. Consequently, the state functions describing a QPT are (ζ, φ) in the wave function $\psi = \zeta e^{i\varphi}$; see (9.3.14).

Thanks to the above three characteristics, we introduce the following definition of QPTs.

Definition 9.2.2 Let $\psi(\lambda) = \zeta e^{i\varphi}$ be a wave function of a QPT system, where λ be the control parameter. If there is a critical value λ_c , for any $\Delta\lambda > 0$ sufficiently small such that the topological structure of $(\zeta(\lambda_c - \Delta\lambda), \varphi(\lambda_c - \Delta\lambda))$ is different from that of $(\zeta(\lambda_c + \Delta\lambda), \varphi(\lambda_c + \Delta\lambda))$.

Remark 9.2.3 In the above definition, the topological structure of (ζ, φ) is described as follows; see the description at the end of Sect. 8.1:

- (1) the flow structure of velocity field $\zeta^2 \nabla \varphi$, and
- (2) the function structure of distribution density $\rho = \zeta^2$.

In quantum condensation systems, $\zeta^2 \nabla \varphi$ represents the superfluid flows or superconducting currents. \square

Examples

Based on the definition of QPTs, Definition 9.2.2, we present in this section a few examples of QPTs.

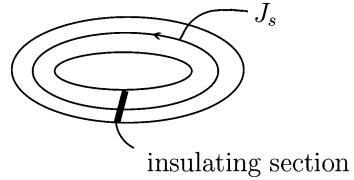
1. *Josephson Superconductor-Insulator Transition* In superconductivity, the Josephson tunnel effect is that a superconducting current remains even if a thin insulating layer is sandwiched in a circular superconductor. The quantum effect shows a QPT behavior.

Consider a circular superconductor with an insulating section in the conductor, as shown in Fig. 9.3. Let $\lambda > 0$ be the thickness of the insulating layer. It is easy to know that there exists critical thickness $\lambda_c > 0$, such that the superconducting current J_s has a transition at λ_c , i.e.

$$J_s \begin{cases} \neq 0 & \text{if } 0 < \lambda < \lambda_c, \\ = 0 & \text{if } \lambda_c < \lambda. \end{cases} \quad (9.2.2)$$

The phenomena (9.2.2) is the QPT, called the Josephson superconductor-insulator transition. If the wave function of the superconductivity is as $\psi = \zeta e^{i\varphi}$, then J_s is as in (9.3.22):

Fig. 9.3 An insulating section in the conductor



$$J_s = \frac{e_s \hbar}{m_s c^2} \zeta^2 \nabla \varphi(\lambda), \quad \nabla \varphi = a(\lambda) \sin(\varphi_2 - \varphi_1),$$

where φ_2 and φ_1 are the phases of ψ on the two sides of the insulating section, $a(\lambda)$ is phase factor with $a(\lambda) = 0$ for $\lambda > \lambda_c$. Hence, (9.2.2) can be rewritten as

$$\nabla \varphi(\lambda) \begin{cases} \neq 0 & \text{for } 0 < \lambda < \lambda_c, \\ = 0 & \text{for } \lambda_c < \lambda. \end{cases}$$

It is clear that this is a typical QPT as defined by Definition 9.2.2.

2. *Mott Superfluid-Insulator Transition* The Mott superfluid-insulator transition is the QPT of a gaseous BEC system, which undergoes a condensation transition as the temperature $T < T_c$. Similar to the liquid He, under an external potential field, the gaseous BEC system also exhibits the superfluidity behavior. In this case, the wave function is

$$\psi = \zeta e^{i\varphi},$$

superfluidity: $\frac{\hbar}{m} \zeta^2 \nabla \varphi \neq 0.$ (9.2.3)

The superfluidity in (9.2.3) occurs in the system is exerted some potential barrier, then it will yield a damping for the superfluidity. The damping can be described by a control parameter:

$$\lambda = V/J,$$

where V represents the barrier intensity, and J is the coefficient of tunnel breakdown. Then, the experiments show that there exists a critical value λ_c , such that

$$\text{the superfluidity } \nabla \varphi \begin{cases} \neq 0 & \text{if } 0 \leq \lambda < \lambda_c, \\ = 0 & \text{if } \lambda_c < \lambda. \end{cases} \quad (9.2.4)$$

The phenomena (9.2.4) is the Mott superfluid-insulator transition.

3. *Abrikosov Vortex-Meissner Phase Transition* In superconductivity, the Abrikosov vortex-Meissner phase transition is a typical QPT phenomena. Based on the Abrikosov theory, for a Type-II superconductor, there are two types of superconducting states:

$$\begin{array}{ll} \text{the Meissner phase} & \text{denoted by } \psi_M, \\ \text{the vortex phase} & \text{denoted by } \psi_V. \end{array} \quad (9.2.5)$$

The control parameter λ is the external magnetic field:

λ = the applied magnetism H_a .

Then there is a critical magnetic field λ_c , such that

$$\text{the superconducting wave function} = \begin{cases} \psi_M & \text{as } \lambda_c < \lambda < \lambda_N, \\ \psi_V & \text{as } 0 < \lambda < \lambda_c, \end{cases} \quad (9.2.6)$$

where ψ_M , ψ_V are as in (9.2.5), λ_N is the critical magnetic field for the normal state. The property (9.2.6) is called the Abrikosov vortex–Meissner phase transition.

4. *Liquid ^3He A–B Phase Transitions* In the absence of applied magnetic field, the superfluidity of liquid ^3He has two phases:

$$\begin{aligned} \text{A phase :} \quad \psi_A &= \psi_+ + \psi_-, \\ \text{B phase :} \quad \psi_B &= \psi_+ + \psi_- + \psi_0, \end{aligned} \quad (9.2.7)$$

where ψ_+ , ψ_- , ψ_0 represent the wave functions for the particles with spin $s = 1, -1, 0$ respectively.

When a magnetic field \mathbf{H} is applied, the system has three phases:

$$\text{A, B phases and } A_1 \text{ phase : } \psi_{A_1} = \psi_+. \quad (9.2.8)$$

For (9.2.7) and (9.2.8), the control parameters are

$$\lambda = \begin{cases} \text{pressure } p & \text{for } \mathbf{H} = 0, \\ \text{applied magnetic field } \mathbf{H} & \text{for } \mathbf{H} \neq 0. \end{cases}$$

Then there exist three types of QPTs for the ^3He superfluid as follows:

- (i) For $\mathbf{H} = 0$, there are two critical temperatures T_0 and T_1 . For each fixed T with $T_0 < T < T_1$, there exists a critical pressure p_c , such that the following QPT occurs as p crosses p_c and is near p_c :

$$\text{the superfluid wave function } \psi = \begin{cases} \psi_A & \text{for } p_c < p, \\ \psi_B & \text{for } p < p_c, \end{cases} \quad (9.2.9)$$

where ψ_A , ψ_B are as in (9.2.7).

- (ii) For $\mathbf{H} \neq 0$ with $H = |\mathbf{H}|$, there are three critical temperatures

$$T_0 < T_1 < T_2.$$

If $T_1 < T < T_2$, there is a critical magnetic field H_{c1} , such that for H near H_{c1} ,

$$\text{the superfluid wave function } \psi = \begin{cases} \psi_A & \text{if } H_{c1} < H, \\ \psi_{A_1} & \text{if } H < H_{c1}. \end{cases} \quad (9.2.10)$$

- (iii) For $\mathbf{H} \neq 0$ and for $T_0 < T < T_1$, there is an H_{c_2} such that for H near H_{c_2} ,

$$\text{the superfluid wave function } \psi = \begin{cases} \psi_B & \text{if } H_{c_2} < H, \\ \psi_A & \text{if } H < H_{c_2}, \end{cases} \quad (9.2.11)$$

where ψ_A is as in (9.2.8).

The three transitions (9.2.9), (9.2.10), and (9.2.11) characterize the QPTs for the liquid ^3He superfluid.

9.2.2 Topological Index of Quantum Condensates

The four groups of equations (8.5.1)–(8.5.2), (8.5.3), (8.5.6), and (8.5.9)–(8.5.10) constitute the basic models to study topological phase transitions for quantum condensation systems. They can be in the unified abstract form as

$$i\hbar \frac{\partial \psi}{\partial t} = \frac{\delta}{\delta \psi^*} \mathcal{H}(\psi, \lambda). \quad (9.2.12)$$

There are two types of transitions for (9.2.12) given as follows:

- (i) *Condensate Particle Number Conserved Systems* In this case, the equation (9.2.12) is equivalent to (9.3.17), i.e.

$$\frac{\delta}{\delta \psi^*} \mathcal{H}(\psi, \lambda) = \mu \psi, \quad (9.2.13)$$

where μ is the chemical potential.

- (ii) *Non-conserved Condensate Particle Number Systems* For this transition, the number of condensate particles is not conserved, and the model describing the transition is the steady-state equation of (9.2.12) given by

$$\frac{\delta}{\delta \psi^*} \mathcal{H}(\psi, \lambda) = 0. \quad (9.2.14)$$

QPTs for systems (9.2.13) and (9.2.14) are topological phase transitions (pattern formation) as defined by Definition 9.2.2. To establish the QPT theory for (9.2.13) and (9.2.14), we need to introduce the mathematical theory of the topological structure of (ζ, φ) for the solutions ψ of (9.2.13) and (9.2.14) written as

$$\psi(\lambda) = \zeta(\lambda) e^{i\varphi(\lambda)}. \quad (9.2.15)$$

- To justify Definition 9.2.2, it is necessary to ensure the topological structure given by $(\zeta(\lambda), \varphi(\lambda))$ in (9.2.15) is physically stable:

for a noncritical value λ , the topological structure defined by $(\zeta(\lambda), \varphi(\lambda))$ is the same (in equivalent sense) as that of $(\zeta(\lambda \pm \varepsilon), \varphi(\lambda \pm \varepsilon))$ for any $\varepsilon > 0$ sufficiently small.

- Both equations (9.2.13) and (9.2.14) are topological structure equations as given in Sect. 8.1.3. For gaseous and liquid condensation systems, they are given by (8.1.11) and (8.1.12), recalled here for convenience:

$$\operatorname{div}(\zeta^2 \nabla \varphi) = 0, \quad (9.2.17)$$

$$-\frac{\hbar^2}{2m} [\Delta \zeta - \zeta |\nabla \varphi|^2] + f(\zeta) \zeta = 0. \quad (9.2.18)$$

For superconducting systems, they are given by (8.1.15) and (8.1.16), also recalled here for convenience:

$$\operatorname{div} J = 0, \quad (9.2.19)$$

$$-\frac{\hbar^2}{2m_s} \left[\Delta \zeta - \zeta |\nabla \varphi|^2 \right] + \frac{e_s^2}{2m_s c^2} A^2 \zeta + f(\zeta) \zeta = 0. \quad (9.2.20)$$

Here $\zeta^2 = \rho$ represents the density of the condensed particles, $\rho \nabla \varphi$ is the superfluid current field, and J is the supercurrent

$$J = \begin{cases} \frac{\hbar}{m} \zeta^2 \nabla \varphi & \text{for BEC and liquid He,} \\ \frac{e_s}{m_s c} \zeta^2 \left(\hbar \nabla \varphi - \frac{e_s}{c} \mathbf{A} \right) & \text{for superconductivity} \end{cases} \quad (9.2.21)$$

Here we have used the Coulomb gauge

$$\operatorname{div} A = 0.$$

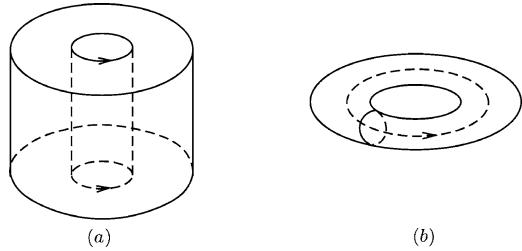
- The authors have developed a geometric theory for incompressible flows to study the structure and its stability and transitions of 2-D incompressible fluid flows in the physical spaces. The complete account of this theory is given in the authors' research monograph (Ma and Wang, 2005d). A brief recapitulation of the theory is given in Sect. 9.1.

The divergence-free condition of J in (9.2.17) and (9.2.19) shows the incompressibility of the supercurrents. For the 2D divergence-free vector fields, the structural classification and structural stability theorems, proved in Ma and Wang (2005d) and recalled in Sect. 9.1, show that a 2D divergence-free regular vector field consists of finite number of vortices, and they are structurally stable as defined in (9.2.16).

Based on the authors' geometric theory for 2D incompressible flows, we can define a topological index for 2D systems. For (ζ, φ) in (9.2.15) as follows

$$\operatorname{Ind}(\psi) = \operatorname{Ind}(\zeta, \varphi) = N, \quad (9.2.22)$$

Fig. 9.4 (a) Columnar vortex flows; **(b)** torus flows



where J is as in (9.2.21), and N is the vortex number of J . The index (9.2.22) represents the topological structure of (ζ, φ) satisfying the stability defined by (9.2.16).

4. For 3D systems, the vector fields J given by (9.2.21) are harmonic fields in essence. Hence, if $J \neq 0$, then in mathematics the harmonic field J consists of finite numbers of columnar vortex flows and torus flows as shown in Fig. 9.4a, b, which are stable in the sense that their columnar and torus structures remain invariant.

Hence, we can define the topological index for the 3D system by

$$\text{Ind}(\psi) = \text{Ind}(\zeta, \varphi) = (N_1, N_2), \quad (9.2.23)$$

where N_1, N_2 represent the numbers of columnar flows and torus flows of J . Such index possesses the stability as required by (9.2.16).

5. For the case $\nabla\varphi = 0$, the wave function

$$\psi = \text{real function}, \quad i.e. \text{Im}\psi = 0. \quad (9.2.24)$$

For the real function ψ , the domain $\Omega \subset \mathbb{R}^n$ ($1 \leq n \leq 3$) can be divided into three kinds of connected regions as

$$\begin{aligned} \Omega^+ &= \{x \in \Omega \mid \psi(x) > 0\}, \\ \Omega^- &= \{x \in \Omega \mid \psi(x) < 0\}, \\ \Omega^0 &= \{x \in \Omega \mid \psi(x) = 0\}. \end{aligned} \quad (9.2.25)$$

It is clear that the topological structure defined by the numbers of connected components of Ω^+ and Ω^- is mathematically stable if and only if the following two conditions hold true:

- (i) the numbers N^+ and N^- of connected components of Ω^+ and Ω^- are finite; and
- (ii) the dimension of each Ω^0 is less than the dimension n of the domain Ω :

$$\dim \Omega^0 < n.$$

For physical stability, it suffices that only (i) holds true, because in condensation systems, the normal state regions Ω^0 with $\dim \Omega^0 = n$ are usually caused by the

material properties, which possess stability under small perturbation of λ . Hence, ψ is called physically stable, if only condition (i) above holds true.

For physically stable field ψ , we define the index as

$$\text{Ind}(\psi) = (N^+, N^-), \quad (9.2.26)$$

where N^+ and N^- are the numbers of connected regions Ω^+ and Ω^- defined by (9.2.25).

Remark 9.2.4 In the case that $\nabla\varphi = 0$ and $\zeta \neq 0$, the system is in condensation state, but no particle superflows exist. The states of ψ in Ω^+ and Ω^- are different from phases each other, i.e.

$$\begin{aligned} \varphi^+ - \varphi^- &= (2m + 1)\pi, \\ \varphi_i^+ - \varphi_j^+ &= 2m\pi, \quad i \neq j, \quad (m = 0, \pm 1, \pm 2, \dots), \\ \varphi_k^- - \varphi_l^- &= 2m\pi, \quad k \neq l, \end{aligned} \quad (9.2.27)$$

where φ_i^+, φ_k^- represent the phases of φ in Ω_i^+ and Ω_k^- and Ω_i^\pm and $\Omega_j^\pm (i \neq j)$ are different connected domains. \square

Although the phase differences in (9.2.27) cannot be detected by experiments, the different indices of (9.2.26) represent different states of ψ , and under some external driving forces, the states will illustrate their own superflow structure.

9.2.3 Microscopic Mechanism of the Meissner Effect

The Meissner effect is one of two main properties: superconductivity and diamagnetism. When the temperature is below T_c , a superconductor not only excludes an external magnetic field to enter its body, but also removes the magnetism, which was originally present in its interior in the normal state. A superconducting state is called the Meissner phase if the superconductor lies in a completely diamagnetic state.

The Abrikosov vortex–Meissner phase transition is a QPT from the Meissner phase to vortex state in a Type-II superconductor when an applied magnetic field H_a increases to exceed to a critical field H_c . To establish the QPT theory for the Abrikosov transition, we have to consider the microscopic mechanism for the Meissner effect.

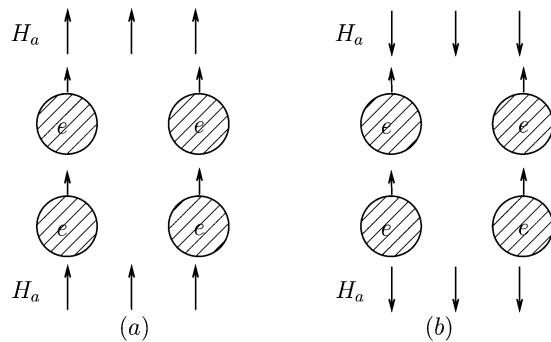
In Fig. 8.5a, b, we note that there are two arrangements for Cooper pairs, leading to different spins:

$$\begin{aligned} \text{spin } J = 0 &\quad \text{in Fig. 8.5a,} \\ \text{spin } J = 1 &\quad \text{in Fig. 8.5b.} \end{aligned}$$

For the case without an applied magnetic field, or for a moving Cooper pair, the spin $J = 0$ state is stable and the $J = 1$ state is unstable; see Ma and Wang (2019).

When an external magnetic field H_a is applied, the arrangement of the $J = 1$ Cooper pairs become stable since the electrons are held in their position by the

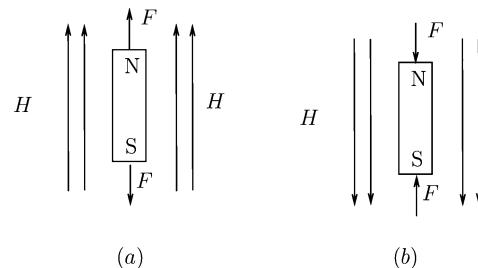
Fig. 9.5 (a) Cooper pairs parallel to \mathbf{H}_a ($s = 1$), and (b) Cooper pairs reversely parallel to \mathbf{H}_a ($s = -1$)



applied magnetic field. There are two orientations for the $J = 1$ Cooper pairs (i.e., $s = \pm 1$): parallel to the field \mathbf{H}_a ($s = 1$) and reversely parallel to \mathbf{H}_a ($s = -1$); see Fig. 9.5.

We know that for a magnetic body under a magnetic field \mathbf{H} , if the north-pole N points to the same direction as \mathbf{H} , then the force acting on the north-pole of this body points to the same direction as N , and on the force acting on the south pole S points to the direction of $-\mathbf{H}$; see Fig. 9.6a. Conversely, if the north pole is in the direction of $-\mathbf{H}$, then the forces acting on both the N and S poles are reversed as shown in Fig. 9.6b.

Fig. 9.6 Directions of the forces F acting on the two ends of the magnetic body under an applied magnetic field \mathbf{H}



Based on this principle of magnetic interaction as described in Fig. 9.6, it is easy to see that the Cooper pairs as shown in Fig. 9.5a are exerted a stretching force by \mathbf{H}_a , and these as shown in Fig. 9.5b are exerted a squeezing force. Consequently, the Cooper pairs parallel to \mathbf{H}_a are unstable as the two electrons can break easily due to the stretching force. However, the Cooper pairs reversely parallel to \mathbf{H}_a are stable.

In addition, each Cooper pair with spin $J = 1$ ($s = \pm 1$) induces a magnetic moment $\vec{\mu}_0$ given by

$$\boldsymbol{\mu}_0 = \frac{e_s \hbar}{m_s c} \mathbf{S},$$

where \mathbf{S} is the unit vector at the spin direction. Because the orientation of $\boldsymbol{\mu}$ is reversely parallel to that of \mathbf{H}_a , the total magnetic moment of all Cooper pairs with spin $s = -1$ can counteract the applied magnetic field \mathbf{H}_a . Hence, we derive the following microscopic mechanism for the Meissner effect.

Mechanism of Meissner Effect 9.2.5

1. Below the critical temperature T_c , an applied magnetic field \mathbf{H}_a induces spin $J = 1$ Cooper pairs;
2. Only the Cooper pairs with spin $s = -1$ (reversely parallel to \mathbf{H}_a) are stable, leading to their physical formation;
3. The total magnetic moment of all Cooper pairs with $s = -1$, together with the surface supercurrents, can cancel out the magnetism induced by the applied field \mathbf{H}_a in the superconductor, and resists the applied field \mathbf{H}_a from entering its body.

Remark 9.2.6 The current explanation of the Meissner effect is that the magnetic field \mathbf{M} induced by the surface supercurrents cancels out the applied field \mathbf{H}_a . However, if the applied field \mathbf{H}_a is homogeneous, it is impossible to be completely eliminated by the nonhomogeneous magnetic field induced by the surface supercurrents. This fact shows that the magnetic field

$$\mathbf{M} = -\frac{e_s \hbar}{m_s c} \frac{N^-}{N} |\psi|^2 \hat{\mathbf{H}}_a$$

plays an important role to counteract the applied field \mathbf{H}_a in the Meissner effect, where N^- , N , and $\hat{\mathbf{H}}_a$ are as in (8.4.13).

9.2.4 Abrikosov Vortex–Meissner Phase Transition

In Sect. 9.2.1, we briefly introduced the Abrikosov vortex–Meissner phase transition (A–M transition). Based on the mechanism of the Meissner effect and the superconducting field equations (8.5.1) and (8.5.2), we can derive the theory for this topological phase transition.

1. *Meissner State and Vortex State* In a Type-II superconductor, in general there are two superconducting states (or three states in some materials): the Meissner state and the vortex state (also called the mixed state). The Meissner state is that in the surface there is a penetrating magnetism, and in the interior the magnetism is zero:

$$\text{the Meissner state} = \text{the interior magnetism } \mathbf{H} = 0.$$

The vortex state is a mixed state, i.e., the normal and the superconducting states coexist. In this phase, the magnetism penetrates the sample to form some columnar vortex supercurrents with central column being in the normal state, as shown in Fig. 9.7, called the Abrikosov vortices.

2. *Phase Diagram for Critical Magnetic Fields* For general Type-II superconductors, there are two (or three for certain materials) critical magnetic fields, denoted by H_{c1} , H_{c2} , and H_{c3} . The phase diagram for Type-II superconductors is given in Fig. 9.8.

Fig. 9.7 Schematic diagram for the Abrikosov vortex: the central column region marked by N is in the normal state, the magnetism penetrates the region, and J_s is the supercurrents, reversely parallel to \mathbf{H}_a

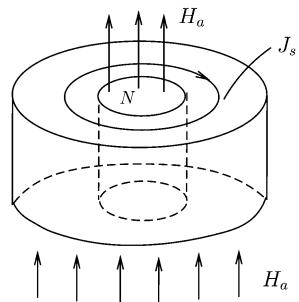
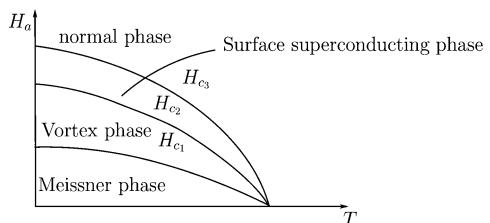


Fig. 9.8 Phase diagram for Type-II superconductors



For the A–M transition, a Type-II superconductor enters the vortex phase from the Meissner phase as the magnetic field H_a increases crossing the first critical magnetic field H_{c1} .

3. *Theory for A–M Transition* The A–M transition is an equilibrium state transition process. The model describing this system is the steady-state equations of (8.5.1) and (8.5.2) with $\mathbf{E} = 0$:

$$\frac{1}{2m_s} \left(i\hbar\nabla + \frac{e_s}{c} \mathbf{A} \right)^2 \psi - V\psi + g|\psi|^2\psi - \boldsymbol{\mu} \cdot \mathbf{H}\psi = 0, \quad (9.2.28)$$

$$\mathbf{H} = \mathbf{H}_a + 4\pi|\psi|^2\boldsymbol{\mu}. \quad (9.2.29)$$

By (8.4.13) and the mechanism of the Meissner effect, the induction magneton $\boldsymbol{\mu}$ is as

$$\boldsymbol{\mu} = -\frac{e_s\hbar}{m_s c} \frac{N^-}{N} \hat{\mathbf{H}}_a, \quad (9.2.30)$$

where N^- is the number of the Cooper pairs with spin $s = -1$, N is the number of Cooper pairs, and $\hat{\mathbf{H}}_a$ is the unit vector in the direction of \mathbf{H}_a , the bounding potential V is

$$V = g_0 - \alpha T \quad (V > 0). \quad (9.2.31)$$

Here g_0 is the particle interacting potential, α is the entropy coupling parameter, and

$$T_c = g_0/\alpha, \quad \text{is the critical temperature.} \quad (9.2.32)$$

For (9.2.31) and (9.2.32), see Sect. 8.2.

For a Type-II superconductor and for a fixed $T < T_c$, by (9.2.28) and (9.2.29), we have the following conclusions for the A–M transition:

if $0 \leq H_a < H_{c_1}$, then the system is in the Meissner phase with the interior magnetic field $\mathbf{H} = 0$ or $H_a = -4\pi|\psi|^2\mu$, and with $\psi = \sqrt{V/g}$;

if $H_{c_1} < H_a < H_{c_2}$, then the system is in the vortex phase with the interior magnetic field $\mathbf{H} \neq 0$ or $H_a > \frac{4\pi e_s \hbar}{m_s c} \frac{N^-}{N} |\psi|^2$, and ψ satisfies (9.2.28) with $\mathbf{A} \neq 0$.

By using the topological index (9.2.22), the transition from (9.2.33) to (9.2.34) can be simply stated as

$$\text{Ind}(\psi) = \begin{cases} 0 & \text{for } 0 \leq H_a < H_{c_1}, \\ (N, 0) \text{ with } N \neq 0 & \text{for } H_{c_1} < H_a < H_{c_2}, \end{cases} \quad (9.2.35)$$

where N is the number of Abrikosov vortices, H_{c_1} and H_{c_2} are as in Fig. 9.8. In particular, the number N can be computed from the equation (9.2.28)–(9.2.29), which are equivalently written as

$$\begin{aligned} \frac{1}{2m_s} \left(i\hbar\nabla + \frac{e_s}{c} \mathbf{A} \right)^2 \psi - \left(V - \frac{e_s \hbar}{m_s c} \frac{N^-}{N} H_a \right) \psi \\ + \left(g - \frac{4\pi e_s^2 \hbar^2}{m_s^2 c^2} \left(\frac{N^-}{N} \right)^2 \right) |\psi|^2 \psi = 0, \end{aligned} \quad (9.2.36)$$

$$\text{curl}^2 \mathbf{A} = \text{curl} \mathbf{H}_a - \frac{4\pi e_s^2}{m_s c^2} |\psi|^2 \mathbf{A} - \frac{2\pi e_s^2}{m_s c} i(\psi^* \nabla \psi - \psi \nabla \psi^*), \quad (9.2.37)$$

supplemented with a physical boundary condition.

4. *Critical Induction Magnetron μ_c* By the Abrikosov theory, H_{c_1} can be expressed as

$$H_{c_1} = \frac{H_c}{\sqrt{2}\kappa} \ln \kappa, \quad (9.2.38)$$

where $\kappa > \sqrt{2}$ is the Ginzburg–Landau parameter, and H_c is the critical H_a , which makes the coefficient of ψ in (9.2.36) to be zero:

$$\mu_c H_c = V, \quad V \text{ as in (9.2.31)}. \quad (9.2.39)$$

The μ_c in (9.2.39) is also the critical magneton which satisfies $H = 0$ of (9.2.29) at $H_a = H_{c_1}$, i.e.

$$H_{c_1} = 4\pi|\psi|^2 \mu_c.$$

By (9.2.33), $|\psi|^2 = V/g$. Then μ_c satisfies

$$H_{c_1} = \frac{4\pi V}{g} \mu_c. \quad (9.2.40)$$

From (9.2.38)–(9.2.40), we derive the critical induction magneton as

$$\mu_c^2 = \frac{g}{4\sqrt{2}\pi\kappa} \ln \kappa. \quad (9.2.41)$$

Equivalently,

$$\boldsymbol{\mu}_c = -\mu_c \hat{\mathbf{H}}_a.$$

In view of (9.2.30), we have

$$N_c^- = N \frac{m_s c}{e_s \hbar} \mu_c, \quad N = |\Omega| |\psi|^2, \quad (9.2.42)$$

where $|\Omega|$ is the volume of the superconductor, and N_c^- represents the maximal value of N^- in this system.

9.2.5 QPT Theorem for Scalar BEC Systems

In this section, we introduce a QPT theorem for the scalar BEC system established in Ma et al. (2016). The dynamic equation governing the systems is given by (8.5.3), which is recalled here for convenience:

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi - V\psi + g|\psi|^2\psi, \quad (9.2.43)$$

supplemented with either the Neumann boundary condition:

$$\left. \frac{\partial \psi}{\partial n} \right|_{\partial\Omega} = 0, \quad (9.2.44)$$

or the Dirichlet boundary condition:

$$\psi|_{\partial\Omega} = 0. \quad (9.2.45)$$

The physical significance of (9.2.44) is that $T < T_c$ on Ω , and (9.2.45) is that $T < T_c$ in the interior of Ω and $T \geq T_c$ on $\partial\Omega$.

In the BEC state, the particle number is conserved:

$$\int_{\Omega} |\psi|^2 = \text{constant}.$$

For this case, the solutions of (9.2.43) can be written as

$$\psi = \varphi(x) e^{-i\mu t/\hbar}, \quad (9.2.46)$$

where μ is the chemical potential. Inserting (9.2.46) into (9.2.43), we deduce that the equations

$$-\frac{\hbar^2}{2m}\Delta\varphi + g|\varphi|^2\varphi = (V + \mu)\varphi \quad (9.2.47)$$

with φ satisfying either the boundary conditions (9.2.44) or (9.2.45).

1. *Main Mathematical Theorem* Consider the eigenvalue problem for the Laplace operator

$$\begin{aligned} -\Delta e_k &= \lambda_k e_k, \\ e_k|_{\partial\Omega} &= 0 \quad \left(\text{or } \frac{\partial e_k}{\partial n}\Big|_{\partial\Omega} = 0\right). \end{aligned} \quad (9.2.48)$$

Mathematically, (9.2.48) possess infinite number of eigenvalues (counting multiplicities):

$$0 < \lambda_1 < \dots \leq \lambda_k \leq \dots \quad (\lambda_1 = 0 \text{ for (9.2.44)}), \quad (9.2.49)$$

where λ_1 is simple. Let e_k be the eigenfunction corresponding to λ_k . For $n = 1$ case, the topological index of e_k defined by (9.2.26) is given by

$$\text{Ind}(e_k) = \begin{cases} (N, N) & \text{if } k = 2N, \\ (N+1, N) & \text{if } k = 2N+1. \end{cases}$$

For $n \geq 2$, the expression of $\text{Ind}(e_k)$ is complex. Then we have the following theorem, which is an equivalent form of the theorem in Ma et al. (2016).

Theorem 9.2.7 *Let $\lambda = (V + \mu)$ be the control parameter. For the system (9.2.47), we have the following assertions:*

- a. *For each eigenvalue λ_k of (9.2.48), the equation (9.2.47) bifurcates from $\lambda_k \frac{\hbar^2}{2m}$ two nonzero solutions $\pm\varphi_k(\lambda)$ on the side of $\lambda > \lambda_k \frac{\hbar^2}{2m}$, and*

$$\text{Ind}(\varphi_k(\lambda)) = \text{Ind}(e_k), \quad \forall \lambda > \lambda_k.$$

- b. *If $\lambda_k \frac{\hbar^2}{2m} < \lambda \leq \lambda_{k+1} \frac{\hbar^2}{2m}$, then (9.2.47) possesses at least $2k$ solutions $\pm\varphi_j(\lambda)$ ($1 \leq j \leq k$), where $\pm\varphi_j(\lambda)$ are the solutions bifurcated from λ_j .*

2. *Physical Conclusions* Theorem 9.2.7 leads to the following physical conclusions:

- (i) Let φ be a solution of (9.2.47). Then the Hamiltonian energy E and the chemical potential μ for φ are

$$\begin{aligned} E &= \int_{\Omega} \left[\frac{\hbar^2}{2m_0} |\nabla\varphi|^2 - V|\varphi|^2 + \frac{g}{2} |\varphi|^4 \right] dx, \\ \mu &= \int_{\Omega} \left[\frac{\hbar^2}{2m_0} |\nabla\varphi|^2 - V|\varphi|^2 + g|\varphi|^4 \right] dx, \\ \mu - E &= \frac{g}{2} \int_{\Omega} |\varphi|^4 dx; \end{aligned}$$

- (ii) The control parameters for this system are (V, N, Ω) , where V is the bounding potential, N is the particle number, and Ω is the container. The parameter $\lambda = V + \mu$ can be viewed as a combined control parameter, which is determined by the parameters V, N , and Ω ;
- (iii) Assertion (1) of Theorem 9.2.7 amounts to saying that if $\lambda = V + \mu$ passes through $\hbar^2 \lambda_j / 2m$ from the left-hand side to the right-hand side, then the system (9.2.47) will add $2k$ quantum states, where k is the multiplicity of λ_j :

$$2k \text{ new quantum states for (9.2.47)} \text{ emerge as } \lambda \text{ crosses } \frac{\hbar^2 \lambda_k}{2m}; \quad (9.2.50)$$

- (iv) If $\lambda = V + \mu$ satisfies

$$\frac{\hbar^2}{2m} \lambda_k < \lambda < \frac{\hbar^2}{2m} \lambda_{k+1}, \quad (9.2.51)$$

- then the system (9.2.47) has at least $2k$ quantum states $\pm \varphi_j$ ($1 \leq j \leq k$), such that φ_j has the same index as the j -th eigenfunction e_j of (9.2.48);
- (v) If $\lambda = V + \mu$ is fixed as in (9.2.51), then we can tune the physical parameters: N, Ω , the electromagnetic fields \mathbf{E}, \mathbf{H} and the pressure p so that the system undergoes QPT between the $2k$ quantum states $\pm \varphi_j$ ($1 \leq j \leq k$).

9.3 Formation of Galactic Spiral Structure

The aim of this section is to derive the mechanism for the formation of the galactic spiral patterns, and the presented study is based on Ma and Wang (2017g).

9.3.1 Formation Mechanism of Spiral Galaxies

(1) Galactic Structures

A typical galaxy consists of two main parts: the disc and the halo. The halo is a ball shaped spatial region located in the center of the galaxy, and the disk is composed of stars and nebulae, rotating around the halo. A schematic diagram for a galactic section is illustrated in Fig. 9.9.

Fig. 9.9 The central ball shaped region is the halo of a galaxy, and the outer region is the disk

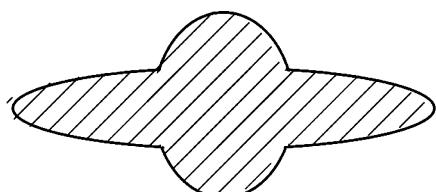


Fig. 9.10 An elliptical galaxy



(a)



(b)

Fig. 9.11 Typical spiral galaxies

There are three types of galactic structures: the spiral, the elliptical, and the irregular. Near 83% of galaxies are spiral, 13%–14% are elliptical, and the others are irregular. Figure 9.10 illustrates an elliptical galaxy and Fig. 9.11a, b shows the typical spiral galaxies.

The spiral structure of galaxies is an important astrophysical phenomenon, which has attracted much attention for a long time. Originally, it was thought that the arms of a spiral galaxy were material. However, this interpretation leads to the winding problem, i.e., the arms would become more and more tightly wound, because the matter near the center of the galaxy moves faster than the matter farther away from the center. Through many years, the arms cannot be distinguished. The density wave theory established by Lin and Shu (1964) proposed that the arms, being nonmaterial, are caused by the nonhomogeneous velocity of stars and nebulae, similar to a traffic jam in a highway. In a road with the traffic jam, the density of cars is larger than in the open roads. In spiral galaxies, stars and nebulae move through the density waves are compressed, and then go out of them.

The traffic jam version of the density wave theory has been generally accepted. However, the reasons to cause the nonhomogeneous velocity of stars and nebulae are still not very clear. The density wave theory provided an explanation, but not completely satisfied. Here we propose a different theory to explain the phenomena, based on three theories developed recently by Ma and Wang (2015a, 2017c):

1. the models for astrophysical fluid dynamics,
2. the gravitational radiations, and
3. the dynamic theory of phase transitions.

9.3.2 Spiral Pattern Formation

First we consider a thermal convection phenomenon in fluid motions, demonstrating the formation mechanism similar to the structure of spiral galaxies.

Let Ω be a two-dimensional ring domain:

$$\Omega = \{(r, \theta) \mid 0 \leq \theta \leq 2\pi, r_0 < r < r_1\}, \quad (9.3.1)$$

where (r, θ) is the polar coordinate system, r_0, r_1 are the inner and outer radii, as shown in Fig. 9.12a. Consider a fluid in Ω coupling heat, with a centripetally gravitational force. The dynamical model for the system is the standard Boussinesq equations

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla)u &= \nu \Delta u - \frac{1}{\rho} \nabla p - g \vec{k}(1 - \alpha T), \\ \frac{\partial T}{\partial t} + (u \cdot \nabla)T &= \kappa \Delta T, \\ \operatorname{div} u &= 0, \end{aligned} \quad (9.3.2)$$

where $u = (u_r, u_\theta)$ is the velocity field, T is the temperature, ν is the viscosity, κ is the diffusion coefficient, ρ is the density, p is the pressure, g is the gravitational constant, α is the expansion coefficient, and $\vec{k} = (1, 0)$ is the unit vector in the r -direction.

On the boundary $r = r_0$ and $r = r_1$, there is a thermal gradient:

$$T = T_0 \text{ at } r = r_0, \quad T = T_1 \text{ at } r = r_1 \quad \text{with } T_0 > T_1. \quad (9.3.3)$$

The boundary value problem (9.3.2)–(9.3.3) has a steady-state solution

$$(U_r, U_\theta) = (U(r), 0), \quad \mathcal{T} = T_0 - \beta(r - r_0), \quad p = P(r),$$

where $\beta = (T_0 - T_1)/(r_1 - r_0)$, and the basic solution $(U(r), 0)$ has the flow structure as shown in Fig. 9.12a. Take the translation

$$u = U(r) + \tilde{u}, \quad T = \mathcal{T} + \tilde{T}, \quad p = P(r) + \tilde{p}, \quad (9.3.4)$$

then equations (9.3.2) become

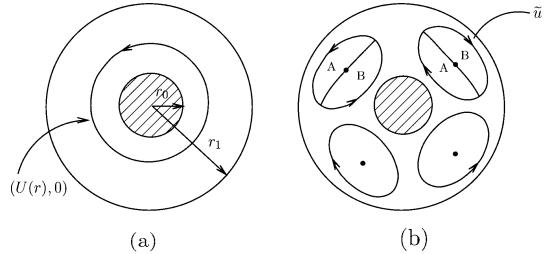
$$\begin{aligned} \frac{\partial \tilde{u}}{\partial t} + (\tilde{u} \cdot \nabla) \tilde{u} &= \nu \Delta \tilde{u} - (U \cdot \nabla) \tilde{u} - (\tilde{u} \cdot \nabla) U - \frac{1}{\rho} \nabla \tilde{p} - g \vec{k} \beta \tilde{T}, \\ \frac{\partial \tilde{T}}{\partial t} + (\tilde{u} \cdot \nabla) \tilde{T} &= \kappa \Delta \tilde{T} + \beta \tilde{u}_r - \frac{a}{r^2} \frac{\partial \tilde{T}}{\partial \theta}, \\ \operatorname{div} \tilde{u} &= 0, \end{aligned} \quad (9.3.5)$$

with a physical boundary condition.

Mathematically, we can prove that there exists a critical value $\beta_c > 0$, such that if

$$\beta = \frac{T_0 - T_1}{r_1 - r_0} > \beta_c, \quad (9.3.6)$$

Fig. 9.12 (a) The basic flow $(U(r), 0)$, and **(b)** the transition solution \tilde{u} , possessing the vortex structure



then equations (9.3.5) bifurcate to a nonzero solution (\tilde{u}, \tilde{T}) with \tilde{u} possessing the vortex structure as shown in Fig. 9.12b.

We now see that if the temperature difference $T_0 - T_1 > 0$ satisfies (9.3.6), then the fluid flow is given by (9.3.4) as the superposition of the basic flow $(U(r), 0)$ in Fig. 9.12a and the bifurcating flow \tilde{u} in Fig. 9.12b:

$$u = (U(r), 0) + \tilde{u}. \quad (9.3.7)$$

It is clear that the flows of \tilde{u} in regions B and A in Fig. 9.12b are oriented in reversed directions, while $(U(r), 0)$ flows only in one direction. Hence, the superposition (9.3.7) results in slower fluid motion in region B and faster fluid motion in regions A, and leads to the spiral pattern of the fluid flow.

The above observation will give rise to a new mechanism for the formation of the spiral galactic structure. However, we are facing the following two basic problems in galactic dynamics:

- (a) *the matter in a galaxy is a noncontinuous, discrete field, and therefore the classical fluid dynamical equations are not suitable for the galactic motion; and*
 - (b) *the temperature is very low in a galaxy, and obviously cannot play the role to derive a vortex motion.*
- (9.3.8)

These two problems in (9.3.8) are solved in the following section by introducing the (1) momentum fluid equations for galactic dynamics, and (2) the gravitational radiation which leads to the needed force for generating galactic vortex motion.

9.3.3 Momentum Fluid Model and Gravitational Radiation

(1) Momentum Fluid Dynamical Equations

In Ma and Wang (2014a, 2015a), the momentum density field $P(x, t)$ was used to replace the velocity field $u(x, t)$ as the state function for galactic objects, because the momentum density field P is the energy flux taking into consideration of the

mass, the heat, and all interaction energy fluxes, and consequently can be regarded as a continuous field. In this section, we introduce the momentum form of the astrophysical fluid dynamical model.

The physical law governing the energy flux is the Newtonian Second Law. Namely, for the momentum density P ,

$$\frac{dP}{dt} = \text{density of the driving force.} \quad (9.3.9)$$

This is the original form of the Newtonian Second Law.

Let ρ be the energy density. As the momentum is energy flux, we have

$$\frac{dx}{dt} = \frac{1}{\rho} P.$$

Hence, we have

$$\frac{dP}{dt} = \frac{\partial P}{\partial t} + \frac{\partial P}{\partial x_k} \frac{dx_k}{dt} = \frac{\partial P}{\partial t} + \frac{1}{\rho} (P \cdot \nabla) P. \quad (9.3.10)$$

Also,

$$\begin{aligned} v \Delta P + \mu \nabla(\operatorname{div} P) & \quad \text{friction,} \\ - \nabla p & \quad \text{pressure force,} \\ \rho \nabla \varphi & \quad \text{gravity,} \end{aligned} \quad (9.3.11)$$

where φ represents the gravitational potential. In view of (9.3.10) and (9.3.11), the equation (9.3.9) is written as

$$\frac{\partial P}{\partial t} + \frac{1}{\rho} (P \cdot \nabla) P = v \Delta P + \lambda \nabla(\operatorname{div} P) - \nabla p + \rho \nabla \varphi. \quad (9.3.12)$$

Also, by the conservation of energy,

$$\frac{\partial \rho}{\partial t} + \operatorname{div} P = 0. \quad (9.3.13)$$

Equations (9.3.12) and (9.3.13) are the momentum fluid model for astrophysical fluid dynamics, replacing classical fluid dynamical equations. Here, ρ and P stand for

ρ = total density of all forms of energy,

P = total flux of all form of energy.

Remark 9.3.1 As we take ρ = mass density, and P = particle flux (i.e., $P = \rho u$), then (9.3.12) and (9.3.13) are reduced to the classical Navier–Stokes equations.

For galactic rotation in the domain (9.3.1), where $r_0 > 0$ represents the halo radius and r_1 is the galactic radius, the differential operators in (9.3.12) and (9.3.13) in the polar coordinates are

$$\begin{aligned}
\Delta P &= \left(\tilde{\Delta}P_r - \frac{2}{r^2} \frac{\partial P_\theta}{\partial \theta} - \frac{P_r}{r^2}, \tilde{\Delta}P_\theta + \frac{2}{r^2} \frac{\partial P_r}{\partial \theta} - \frac{P_\theta}{r^2} \right), \\
\tilde{\Delta}f &= \frac{\partial^2 f}{\partial r^2} + \frac{1}{r} \frac{\partial f}{\partial r} + \frac{1}{r^2} \frac{\partial^2 f}{\partial \theta^2}, \\
(P \cdot \nabla)P &= \left(P_r \frac{\partial P_r}{\partial r} + \frac{P_\theta}{r} \frac{\partial P_r}{\partial \theta} - \frac{P_\theta^2}{r}, \right. \\
&\quad \left. P_r \frac{\partial P_\theta}{\partial r} + \frac{P_\theta}{r} \frac{\partial P_\theta}{\partial \theta} + \frac{P_r P_\theta}{r} \right), \\
\text{div}P &= \frac{\partial P_r}{\partial r} + \frac{1}{r} \frac{\partial P_\theta}{\partial \theta} + \frac{P_r}{r}, \\
\nabla &= \left(\frac{\partial}{\partial r}, \frac{1}{r} \frac{\partial}{\partial \theta} \right).
\end{aligned} \tag{9.3.14}$$

(2) Graviton Thermodynamics

In Ma and Wang (2017d), a statistical theory of heat was derived, and the crucial points of the theory can be stated as follows:

1. the temperature T is the average energy level for a thermodynamic system;
2. entropy is essentially the photon number in the gap of the system particles; and
3. temperature increasing and decreasing are caused by the system particles absorbing and radiating photons respectively. Hence, the photon density and energy levels can characterize the temperature of a system. This is the reason why the temperature diffusion obeys the same law as the particle diffusion does.

Also, we recall that the equation of state for a thermodynamical system

$$\rho_s = f(p, T), \quad \frac{\partial \rho_s}{\partial T} < 0 \tag{9.3.15}$$

provides a relation between the mass density ρ_s , the pressure p , and the temperature T . Based on the new theory of heat, T is a function of the photon energy density σ . Hence (9.3.15) is also equivalent to the form:

$$\rho_s = F(p, \sigma) \quad \text{with} \quad \frac{\partial \rho_s}{\partial \sigma} < 0. \tag{9.3.16}$$

In summary, the photons are the main resource of thermal energy. Hence, the classical thermal physics can also be called as the photon thermodynamics. However, the galactic dynamics cannot be established on the photon thermodynamics as stated in (9.3.16), rather on the graviton thermodynamics to be described as follows.

Now we recapitulate the gravitational radiation theory derived in Ma and Wang (2017c). First, we recall the PID gravitational field equations (Ma and Wang, 2014b, 2015a, 2017c):

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = -\frac{8\pi G}{c^4} T_{\mu\nu} - \nabla_\mu \Phi_\nu, \tag{9.3.17}$$

where there are three groups of state functions, whose physical meaning is given as follows:

1. $\{g_{\mu\nu}\}$ is the Riemannian metrics of four-dimensional space-time, representing the gravitational potential, depicting the curved space-time;
2. $\{\Phi_\nu\}$ is the dual gravitational potential, representing the gravitational field particle and carrying the field energy, which is similar to the electromagnetic interaction field particle: the photon;
3. $\{T_{\mu\nu}\}$ is the energy-momentum tensor of the visible matter field.

It is the dual gravitational potential $\{\Phi_\nu\}$ that plays the role of the dark matter and dark energy. In galactic dynamics, $\{\Phi_\nu\}$ plays the same role as the photon in thermodynamics. In fact, as a state function describing the gravitational field particle,

$$\Phi_\nu \text{ represents the graviton with spin } J = 1. \quad (9.3.18)$$

Physically, an interaction field particle must be a massless and electric neutral boson. Hence, as a graviton $\{\Phi_\nu\}$ has to satisfy the Klein-Gordon type of equations. In fact, by the Bianchi identity,

$$\nabla^\mu (R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R) = 0,$$

we derive from (9.3.17) that

$$\nabla^\mu \nabla_\mu \Phi_\nu = -\frac{8\pi G}{c^4} \nabla^\mu T_{\mu\nu}, \quad (9.3.19)$$

where $\nabla^\mu \nabla_\mu$ is the Klein-Gordon operator. Equation (9.3.19) validate the claim (9.3.18). In vacuum, $T_{\mu\nu} = 0$, and (9.3.19) become

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) \Phi_\nu = 0, \quad (9.3.20)$$

which are the gravitational radiation wave equations.

Remark 9.3.2 The gravitational radiation wave is different from the gravitational wave. The radiation wave is about the field particle $\{\Phi_\nu\}$ in (9.3.20), representing the propagation of gravitational field energy. The gravitational wave is about space metric $\{g_{\mu\nu}\}$, representing the propagation of space deformation. \square

We now examine at the analogy to electromagnetic field particle equations

$$\nabla^\mu \nabla_\mu A_\nu = \nabla_\nu (\text{div} A) + e J_\nu,$$

and to the wave equations of electromagnetic radiation equations

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) A_\nu = 0.$$

In view of (9.3.19) and (9.3.20) we can claim that the graviton $\{\Phi_\nu\}$ in gravitational interaction is the analog of the photon in electromagnetism.

Hence, similar to the photon thermodynamical state equations (9.3.15) and (9.3.16), we propose that there exists a scalar field \mathcal{T} , called the gravitational temperature, such that the energy density ρ in (9.3.12) and (9.3.13) satisfies the equation of state:

$$\rho = f(p, \mathcal{T}), \quad \frac{\partial \rho}{\partial \mathcal{T}} < 0. \quad (9.3.21)$$

In a nutshell,

the absorption and radiation of gravitons could generate gravitational temperature, representing the average energy level of massive matter, reminiscent of the photons yielding the temperature in thermodynamical systems. (9.3.22)

Based on this conclusion, the gravitational temperature \mathcal{T} also satisfies the diffusion equation as

$$\frac{\partial \mathcal{T}}{\partial t} + \frac{1}{\rho}(P \cdot \nabla)\mathcal{T} = \kappa \Delta \mathcal{T} + Q, \quad (9.3.23)$$

where Q represents the gravitational source.

9.3.4 Model for Galactic Dynamics

(1) Boussinesq Approximation

The dynamical equations governing galactic rotation are based on the momentum fluid equations (9.3.12)–(9.3.13) and the gravitational temperature equations (9.3.21) and (9.3.23). Here, we need to make the Boussinesq approximation.

First, the function φ in (9.3.12) is the Newton potential:

$$\nabla \varphi = -\frac{GM_r}{r^2} \hat{r},$$

where \hat{r} is the unit vector in the r -direction, G is the gravitational constant, M_r is the total mass in the ball of radius r . We take approximately $\nabla \varphi$ to be a constant vector

$$\nabla \varphi = -g \hat{r}, \quad g = \frac{M_{r_0} G}{r_0^2}, \quad (9.3.24)$$

where r_0 is the radius of the halo.

We take the standard Boussinesq approximation as follows: First, the equation of state (9.3.21) is taken in the form

$$\rho = \rho_0(1 - \alpha \mathcal{T}), \quad \rho_0 = \text{constant}, \quad (9.3.25)$$

and the momentum density P is divergence-free:

$$\operatorname{div} P = 0. \quad (9.3.26)$$

Second, by (9.3.24) and (9.3.25), the term $\rho\nabla\varphi$ in (9.3.12) is then given by

$$\rho\varphi = -\rho_0 g(1 - \alpha\mathcal{T})\hat{r}. \quad (9.3.27)$$

Also, we set $\rho = \rho_0$ in other terms of (9.3.12) and (9.3.23).

In addition, the gravitational source in (9.3.23) is zero,

$$Q = 0 \quad \text{in } \Omega, \quad (9.3.28)$$

where Ω is as in (9.3.1).

Thus, under the above Boussinesq approximation, equations (9.3.12)–(9.3.13) and (9.3.23) are written as

$$\begin{aligned} \frac{\partial P}{\partial t} + \frac{1}{\rho_0}(P \cdot \nabla)P &= \mu\Delta P - \nabla p - \rho_0 g(1 - \alpha\mathcal{T})\hat{r}, \\ \frac{\partial \mathcal{T}}{\partial t} + \frac{1}{\rho_0}(P \cdot \nabla)\mathcal{T} &= \kappa\Delta\mathcal{T}, \\ \operatorname{div} P &= 0, \end{aligned} \quad (9.3.29)$$

where g is as in (9.3.24), and the differential operators Δ , ∇ , div , and $(P \cdot \nabla)P$ are as in (9.3.14), and

$$(P \cdot \nabla)\mathcal{T} = \left(P_r \frac{\partial \mathcal{T}}{\partial r}, \frac{P_\theta}{r} \frac{\partial \mathcal{T}}{\partial \theta} \right).$$

Now we continue to simply the equations (9.3.29). Consider the ring domain Ω of (9.3.1). For the polar coordinate (r, θ) , we take the transformation

$$x_1 = r\theta, \quad x_2 = r - r_0. \quad (9.3.30)$$

Then the domain Ω becomes

$$\Omega = \{(x_1, x_2) \mid 0 \leq x_1 \leq 2\pi r_0, 0 < x_2 < r_1 - r_0\}. \quad (9.3.31)$$

Here, the domain of (9.3.31) looks as if Ω is a rectangular region. However, when (9.3.29) are supplemented with the periodic boundary condition in the x_1 -direction:

$$(P, \mathcal{T})(x_1 + 2\pi r_0, x_2) = (P, \mathcal{T})(x_1, x_2), \quad (9.3.32)$$

then Ω of (9.3.31) is still a ring domain.

Because the radius r_0 of the halo of a galaxy is very large, usually $r_0 \sim 10^4$ ly (1 ly = 9.46×10^{12} km), we can ignore all terms containing the factor $1/(r_0 + x_2)$ in (9.3.29) after the coordinate transformation (9.3.30). Thus, equations (9.3.29) are approximatively in the form

$$\begin{aligned} \frac{\partial P_1}{\partial t} + \frac{1}{\rho_0}(P \cdot \nabla)P_1 &= \mu\Delta P_1 - \frac{\partial p}{\partial x_1}, \\ \frac{\partial P_2}{\partial t} + \frac{1}{\rho_0}(P \cdot \nabla)P_2 &= \mu\Delta P_2 - \frac{\partial p}{\partial x_2} - g\rho_0(1 - \alpha\mathcal{T}), \\ \frac{\partial \mathcal{T}}{\partial t} + \frac{1}{\rho_0}(P \cdot \nabla)\mathcal{T} &= \kappa\Delta\mathcal{T}, \\ \operatorname{div}P &= 0, \end{aligned} \quad (9.3.33)$$

with the periodic boundary condition (9.3.32), where Δ , ∇ , div , and $(P \cdot \nabla)$ are the usual differential operators in the orthogonal coordinate system in \mathbb{R}^2 .

(2) Basic Steady-State Solution

The steady-state solution of (9.3.33) must be consistent with the physical reality in astronomy. In 1970, V. Rubin and J. W. K. Ford first observed that most stars in a spiral galaxy have the same average orbital velocity; see Rubin and Ford (1970). Namely, if $v(r)$ represents the velocity of stars with distance r from the galactic center, then $v(r)$ is almost independent of r , i.e.

$$v(r) \simeq v_0 \quad \text{a constant.} \quad (9.3.34)$$

A typical galactic rotation curve is as shown in Fig. 9.13. The Rubin rotational curve suggests that the momentum density \bar{P} in the steady-state solution $(\bar{P}, \bar{\mathcal{T}}, \bar{p})$ of (9.3.33) should be constant:

$$\bar{P} = (\bar{P}_1, \bar{P}_2) = (\rho_0 v_0, 0), \quad (9.3.35)$$

where v_0 is the constant velocity in the Rubin rotational curve.

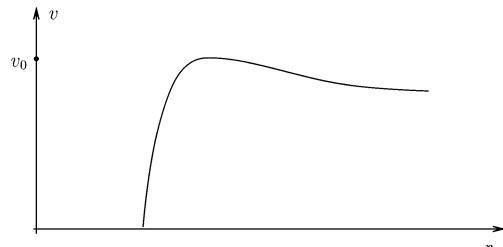
By the gravitational radiation theory, a large amount of gravitons are emitted from the halos of the galaxy, leading to a gravitational temperature gradient:

$$\bar{\mathcal{T}} = \mathcal{T}_0 - \beta x_2, \quad 0 < x_2 < r_1 - r_0, \quad (9.3.36)$$

where $\beta = (\mathcal{T}_0 - \mathcal{T}_1)/(r_1 - r_0)$ with $\mathcal{T}_0 > \mathcal{T}_1$.

Inserting $(\bar{P}, \bar{\mathcal{T}})$ into (9.3.33), we deduce that

Fig. 9.13 Rubin galactic rotational curve, where the horizontal axis represents the distance from the galactic center, and the vertical axis is the velocity of stars



$$\bar{p} = -\rho_0 g \int (1 - \alpha \bar{\mathcal{T}}) dx_2. \quad (9.3.37)$$

The functions $(\bar{P}, \bar{\mathcal{T}}, \bar{p})$ given by (9.3.35)–(9.3.36) are the basic steady state of (9.3.33) obeying the Rubin rotational curve.

(3) Standard form of Model

For the steady-state solution $(\bar{P}, \bar{\mathcal{T}}, \bar{p})$ in (9.3.35)–(9.3.36), we take the translation:

$$P = P' + \bar{P}, \quad \mathcal{T} = \mathcal{T}' + \bar{\mathcal{T}}, \quad p = p' + \bar{p}. \quad (9.3.38)$$

Then, after suppressing the primes, equations (9.3.33) become

$$\begin{aligned} \frac{\partial P_1}{\partial t} + \frac{1}{\rho_0} (P \cdot \nabla) P_1 &= \mu \Delta P_1 - v_0 \frac{\partial P_1}{\partial x_1} - \frac{\partial p}{\partial x_1}, \\ \frac{\partial P_2}{\partial t} + \frac{1}{\rho_0} (P \cdot \nabla) P_2 &= \mu \Delta P_2 - v_0 \frac{\partial P_2}{\partial x_1} - \frac{\partial p}{\partial x_2} + \alpha g \rho_0 \mathcal{T}, \\ \frac{\partial \mathcal{T}}{\partial t} + \frac{1}{\rho_0} (P \cdot \nabla) \mathcal{T} &= \kappa \Delta \mathcal{T}, \\ \operatorname{div} P &= 0, \end{aligned} \quad (9.3.39)$$

Let

$$\begin{aligned} x &= r_0 x', & t &= r_0^2 t' / \kappa, & P &= \kappa \rho_0 P' / r_0, \\ \mathcal{T} &= \beta r_0 \mathcal{T}' / \sqrt{\text{Ra}}, & p &= \rho_0 \kappa^2 r_0^2 p', \end{aligned} \quad (9.3.40)$$

where Ra is the Rayleigh number defined by

$$\text{Ra} = \frac{g \alpha \rho_0 \beta}{\kappa \mu} r_0^4, \quad \beta = \frac{T_0 - T_1}{r_1 - r_0}. \quad (9.3.41)$$

Then, suppressing the primes again, we deduce the following nondimensional form of (9.3.39):

$$\begin{aligned} \frac{1}{\text{Pr}} \left[\frac{\partial P}{\partial t} + (P \cdot \nabla) P \right] &= \Delta P - a \frac{\partial P}{\partial x_1} - \frac{1}{\text{Pr}} \nabla P + \sqrt{\text{Ra}} \mathcal{T} \vec{k}, \\ \frac{\partial \mathcal{T}}{\partial t} + (P \cdot \nabla) \mathcal{T} &= \Delta \mathcal{T} - a \text{Pr} \frac{\partial \mathcal{T}}{\partial x_1} + \sqrt{\text{Ra}} P_2, \\ \operatorname{div} P &= 0. \end{aligned} \quad (9.3.42)$$

The nondimensional domain is

$$\Omega = [0, 2\pi] \times (0, \ell).$$

Here Pr is the Prandtl number, a is the Rubin number, and ℓ is the ratio between the disk width and the halo radius. These are nondimensional parameters given by

$$\text{Pr} = \frac{\mu}{\kappa}, \quad a = \frac{r_0 v_0}{\mu}, \quad \ell = \frac{r_1 - r_0}{r_0}. \quad (9.3.43)$$

Equation (9.3.42) are supplemented with the physically sound boundary conditions:

$$(P, \mathcal{T}, p) \text{ are periodic in } x_1\text{-direction,} \\ P_n = 0, \quad \frac{\partial P_\tau}{\partial n} = 0, \quad \mathcal{T} = 0 \text{ at } x_2 = 0, \ell. \quad (9.3.44)$$

9.3.5 Theory for Galactic Spiral Structure

Based on the galactic dynamical model (9.3.42) and (9.3.44), we now establish the dynamical and topological phase transitions for the galactic spiral pattern formation.

(1) Dynamic Transitions

First, we introduce the dynamical transition theorems for the model.

1. *Critical Rayleigh Number* Consider the eigenvalue problem of the linearized equations of (9.3.42) with the boundary conditions (9.3.44) as follows

$$\begin{aligned} \Delta P - \frac{1}{\text{Pr}} \nabla P - a \frac{\partial P}{\partial x_1} + \sqrt{\text{Ra}} \mathcal{T} \vec{k} &= \beta(\lambda) P, \\ \Delta \mathcal{T} - a \text{Pr} \frac{\partial \mathcal{T}}{\partial x_1} + \sqrt{\text{Ra}} P_2 &= \beta(\lambda) \mathcal{T}, \\ \text{div} P &= 0, \\ (P, T) &\text{ satisfies BC (9.3.44),} \end{aligned} \quad (9.3.45)$$

where λ represents the control parameters:

$$\lambda = (\text{Ra}, a, \text{Pr}, \ell). \quad (9.3.46)$$

For the eigenvalue problem (9.3.45) we have the following theorem, which ensures that the dynamic transition for (9.3.42) with (9.3.44) does occur.

Theorem 9.3.3 *For any given $a, \text{Pr}, \ell > 0$ in (9.3.43), there exists a critical Rayleigh number $R_c > 0$ such that all eigenvalues β_k ($k = 1, 2, \dots$) of (9.3.45) satisfy*

$$Re\beta_i(Ra) \begin{cases} < 0 & \text{if } Ra < R_c, \\ = 0 & \text{if } Ra = R_c, \\ > 0 & \text{if } Ra > R_c \end{cases} \quad \text{for } 1 \leq i \leq m, \quad (9.3.47)$$

$$Re\beta_j|_{Ra=R_c} < 0 \quad \text{for } j > m. \quad (9.3.48)$$

The proof of Theorem 9.3.3 relies on the continuity of $\beta(\lambda)$ on the parameter λ in (9.3.46). In fact, the eigenvalue equations (9.3.45) are symmetric as $a = 0$, and the first eigenvalue β_1 satisfies

$$\beta_1(Ra) \begin{cases} < 0 & \text{if } Ra < R_c^\circ, \\ = 0 & \text{if } Ra = R_c^\circ, \\ > 0 & \text{if } Ra > R_c^\circ, \end{cases} \quad (9.3.49)$$

$\beta_1(Ra) \rightarrow +\infty$ as $Ra \rightarrow +\infty$,

for some $R_c^\circ > 0$. Theorem 9.3.3 then follows from (9.3.49).

2. *Transition Theorem* Thanks to Theorem 9.3.3, by the dynamic transition theorem, Theorem 2.1.3, we immediately derive the following result, which is crucial for the theory of the galactic spiral structure.

Theorem 9.3.4 *For any given $\tilde{\lambda} = (a, Pr, \ell)$, there is a critical Rayleigh number $R_c(\tilde{\lambda})$ satisfying (9.3.47) and (9.3.48), such that the equations (9.3.42) with (9.3.44) undergo a dynamic transition from the basic state $(P, T) = 0$ to a new stable state (\tilde{P}, \tilde{T}) at $Ra=R_c(\tilde{\lambda})$, and the following statements hold true:*

1. if $\beta_1(Ra)$ in (9.3.47) is a real number, then (\tilde{P}, \tilde{T}) is an equilibrium state; and
2. if $\beta_1(Ra)$ is an imaginary number, then (\tilde{P}, \tilde{T}) is a periodic solution.

In particular, the field \tilde{P} has the vortex structure as shown in Fig. 9.12b.

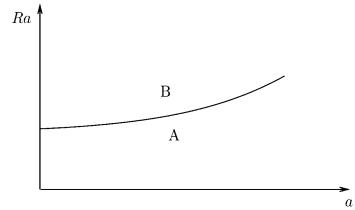
(2) Galactic Pattern Formation

Based on Theorems 9.3.3 and 9.3.4 above, we now establish a new topological phase transition theory for the galactic pattern formation.

1. *Critical Phase Diagram* There are four main parameters to influence the galactic pattern formation: Ra , a , ℓ and Pr given by (9.3.41) and (9.3.43). From the mathematical point of view, the Prandtl number Pr plays a less role in both dynamical and topological phase transitions. The other three parameters are very important for the formation of different galactic structures:

- Ra representing the G-temperature difference $T_0 - T_1$,
- a representing the average velocity of stars in the galaxy,
- ℓ representing the ratio $(r_1 - r_0)/r_0$.

Fig. 9.14 Galactic a -Ra phase diagram, where A is the elliptic galaxy region, and B is the spiral galaxy region



We can obtain the a -Ra phase diagram in Fig. 9.14 from Theorem 9.3.4 for given parameters ℓ and Pr , and the resulting critical curve $R_c = R_c(a)$ divides the domain of (a, Ra) into two regions A and B , whose physical meaning is

$$\begin{aligned} (a, \text{Ra}) \in A &\Rightarrow \text{the galaxy rotates with constant momentum } \bar{P}, \\ (a, \text{Ra}) \in B &\Rightarrow \text{the galaxy rotates with momentum } \bar{P} + \tilde{P}, \end{aligned} \quad (9.3.50)$$

where \bar{P} is as in (9.3.35), and \tilde{P} is the transition solution of (9.3.42) with (9.3.44), given in Theorem 9.3.4. Since \tilde{P} possesses vortex pattern as given in Fig. 9.15, this state $\bar{P} + \tilde{P}$ is spiral in its structure. Hence (9.3.50) can be restated as follows:

$$\begin{aligned} (a, \text{Ra}) \in A &\Rightarrow \text{the galaxy is elliptic,} \\ (a, \text{Ra}) \in B &\Rightarrow \text{the galaxy is spiral.} \end{aligned} \quad (9.3.51)$$

2. *Spiral Arm Pattern* By Theorem 9.3.4 and (9.3.50)–(9.3.51), galactic rotation undergoes two types of transitions: the dynamic phase transition of (9.3.42) with (9.3.44) from $(\bar{P}, \bar{\tau})$ to $(\bar{P} + \tilde{P}, \bar{\tau} + \tilde{\tau})$, and the topological phase transition from the elliptic pattern to the spiral pattern.

Here, we discuss the topological phase transition. In fact, direct numerical computation shows that the dynamic phase transition of (9.3.42) with (9.3.44) is a Hopf bifurcation for any $a > 0$, from the equilibrium $(\bar{P}, \bar{\tau})$ to a periodic solution $(\bar{P} + \tilde{P}, \bar{\tau} + \tilde{\tau})$. In particular, the flow field of \tilde{P} has convective vortex structure as given by the computer simulation in Fig. 9.15, and the vortices of \tilde{P} move around the center of the ring domain Ω with period τ :

$$\text{topological structure of } \tilde{P} = m \text{ right-handed vortices} \quad (9.3.52)$$

$$+ m \text{ left-handed vortices, } m \geq 1;$$

$$\text{motion state of } \tilde{P} = \text{the } 2m \text{ vortices form a zonal motion} \quad (9.3.53)$$

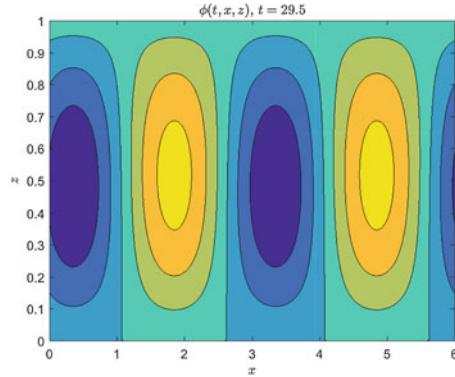
$$\text{in the } \bar{P} \text{ direction crossing the entire ring } \Omega.$$

Here the ring region Ω is defined by (9.3.1).

The above properties (9.3.52) and (9.3.53) of \tilde{P} depict the spiral pattern of galactic rotating motion as follows:

$$\text{state of } \bar{P} + \tilde{P} = m \text{ rotating arms.} \quad (9.3.54)$$

Fig. 9.15 Convective vortex structure



In the following, we explain the claim (9.3.54). Consider the superposition of the constant momentum field $\bar{P} = (\rho_0 v_0, 0)$ of (9.3.35) and the vortex field \tilde{P} as shown in Fig. 9.15. The local structure of $\bar{P} + \tilde{P}$ is schematically illustrated in Fig. 9.16a, in which each vortex has a D region and an E region where the horizontal component of the vortex velocity in D region is reversal to v_0 , and in E is the same as v_0 . By the superposition principle of velocity, we have

$$\begin{aligned} \text{in D : } & v_0 + \tilde{v} = \text{deceleration in } x_1\text{-direction,} \\ \text{in E : } & v_0 + \tilde{v} = \text{acceleration in } x_1\text{-direction,} \end{aligned}$$

where \tilde{v} is the rotating velocity of vortices. Then, it yields a band in Ω where the velocity of $v_0 + \tilde{v}$ in x_1 -direction is slower than v_0 , as shown in Fig. 9.16b. Then, 2m vortices generate m bands where the galactic matter will be jammed to form m spiral arms. Besides, the spiral arms also move in the v_0 direction due to (9.3.53).

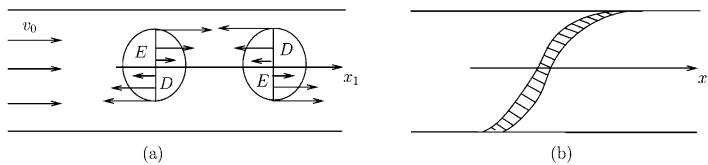


Fig. 9.16 (a) The superposition of v_0 and two vortex flows can yield a slower speed band as the shadow region in (b)

3. *Relation Between the Parameters ℓ and m* In mathematics, the vortex number $2m$ of \tilde{P} , corresponding to the arm number m , depends on the parameter $\ell = (r_1 - r_0)/r_0$, and is a decreasing function of ℓ :

$$\frac{dm}{d\ell} < 0.$$

In fact, as ℓ is very small, we have the asymptotical relation

$$m \sim \ell^{-1} = \frac{r_0}{r_1 - r_0}. \quad (9.3.55)$$

Hence, we infer from (9.3.55) the following conclusion:

$$\ell = \frac{r_1 - r_0}{r_0} \text{ is small} \implies \text{the galaxy is elliptical.} \quad (9.3.56)$$

The reason is that a small ℓ leads to a large arm number m , and it is hard to distinguish the arms from one to another.

4. *Summary* By (9.3.51) and (9.3.56), the parameters R_a , a , ℓ determine galactic structure as follows:

$$\begin{aligned} R_a < R_c(a) \text{ or } \ell \text{ is small} &\implies \text{the galaxy is elliptic,} \\ R_a > R_c(a) \text{ and } \ell \text{ is relatively large} &\implies \text{the galaxy is spiral.} \end{aligned}$$

9.4 Solar Surface Eruptions and Sunspots

This section is aimed to provide the theory for the formation of the solar surface eruptions and sunspots based on Ma and Wang (2017f). The key ingredient of the study is the anti-diffusive effect of heat, derived from the statistical theory of heat presented in Sect. 7.5. The anti-diffusive effect of heat states that due to the higher rate of photon absorption and emission of the particles with higher energy levels, the photon flux will move to the higher temperature regions from the lower temperature regions. This anti-diffusive effect of heat leads to a modified law of heat transfer, which includes a reversed heat flux counteracting the heat diffusion. It is this anti-diffusive effect of heat and thereby the modified law of heat transfer that lead to the temperature blow-up and consequently the formation of sunspots, solar eruptions, and solar prominences. This anti-diffusive effect of heat may be utilized to design a plasma instrument, directly converting solar energy into thermal energy. This may likely offer a new form of fuel much more efficient than the photovoltaic devices.

9.4.1 Sun's Surface Fluid Dynamics

(1) Astronomical Phenomenon on the Sun's Surface

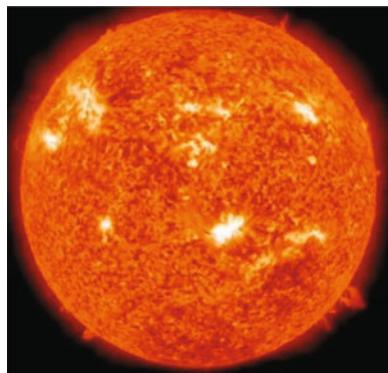
The Sun is a star which we are most familiar with. The solar mass is about 2×10^{30} kg, and the Sun's radius is 7×10^5 km. Its average density is $\rho = 1.4 \text{ g/cm}^3$ and, for comparison, we recall that the density of water is $\rho = 1 \text{ g/cm}^3$. The Sun consists mainly of hydrogen (94%) and helium (6%), and is entirely in a gaseous state. Hence it can be viewed a gaseous fluid ball.

Solar structure is divided into two parts: the interior and the atmosphere. The atmosphere is composed of three distinct parts: the photosphere, the chromosphere, and the corona. The photosphere is in the bottom layer, the chromosphere is in the

middle, and the corona is in the outer of the solar atmosphere. The most visible light comes from the photosphere, and its temperature is ranged from 4000 K to 7000 K. The chromospheric temperature is 1.5×10^4 K, and the corona has the highest temperature at about 2×10^6 K. But the corona density is very low, at 10^{-9} times the density of the earth's atmosphere. Based on astronomical observations, the following are known important phenomena for the solar atmosphere:

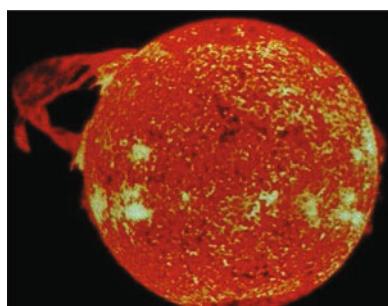
1. *Solar Flare* It is a very attracting event which occurs in Sun's atmosphere, and is a sudden flash in Sun's surface brightness. Solar flares affect all three layers: photosphere, chromosphere, and corona, and are often accompanied by a coronal mass ejection, or by an erupting prominence. When the plasma medium is heated to near $T = 5 \times 10^6$ K, flares are powered by the sudden release of huge energy, together with very strong electromagnetic radiations and high speed (near the speed of light) particle eruptions. The active period of solar flares follows the 11-year cycle, called the solar cycle. See Fig. 9.17 for solar flares.

Fig. 9.17 Solar flares



2. *Solar Prominence* A prominence is a large, bright gaseous feature extending outward from the Sun's surface, usually in a loop shape; see Fig. 9.18. Prominences occur often along with solar flares, are anchored to the Sun's surface in the photosphere, and extend outward into the Sun's corona, reaching as high as thousands of kilometers.

Fig. 9.18 Solar prominence



3. *Sunspots* They are temporary phenomena on the Sun's surface that appear as spots darker than the surrounding areas. The sunspot regions have lower surface temperature. Sunspots occur in an approximatively 11-year solar cycle, the same as that of the solar flares.

Usually, sunspots accompany secondary phenomena such as coronal loops, prominences, and solar flares. Most solar flares and coronal mass ejections originate in active regions around sunspots.

Astronomical observations show that solar flares, prominences, and sunspots are intimately related. The theory that we established here verifies clearly this claim, and explains their relations.

(2) Anti-diffusive Effect of Heat

In Ma and Wang (2017d), the authors developed a statistical theory of heat, consisting of the following main ingredients:

1. *Energy Level Temperature Formulas* These formulas are derived from the well-known Maxwell–Boltzmann, the Fermi–Dirac, and the Bose–Einstein distributions, and they show that the temperature is essentially the average energy level of system particles.
2. *Photon Number Entropy Formula* This formula shows that entropy is the number of photons in the gap between system particles, and the physical carrier of heat is the photons.
3. *Vibratory Mechanism of Photon Absorption and Radiation* A particle can only absorb and radiate photons while experiencing vibratory motion. The higher the frequency of the vibration of the particle, the larger the absorbing and radiating energy. The vibration or irregular motion of the particles in a system is caused by collisions between particles and by absorbing and radiating photons.
4. *Law of Heat Transfer* For particles in high speed vibration and irregular motion, the rate of photon emission and absorption increases, leading to the number density of photons to increase, and further causing the particle energy levels to elevate. Hence, the photon absorption and emission induce the concentration of the temperature, which we call the *anti-diffusive effect of heat*:

Due to the higher rate of photon absorption and emission of the particles with higher energy levels, the photon flux will move toward the higher temperature regions from the lower temperature regions. (9.4.1)

On the other hand, we know that temperature obeys the Fourier law, i.e., photons are dictated by the Fick law. Hence, the heat transfer follows the balance between the anti-diffusive effect (9.4.1) and the Fick diffusion law for the system photons, i.e., the law of heat transfer can be expressed as

$$\frac{dT}{dt} = \kappa\Delta T + Q + \text{the concentration rate of photons}, \quad (9.4.2)$$

where $\kappa\Delta T$ represents the diffusion term, Q is the heat resource.

Remark 9.4.1 In the classical heat conduction theory, the law of heat transfer is written as

$$\frac{dT}{dt} = \kappa\Delta T + Q, \quad (9.4.3)$$

which is different from (9.4.2). In fact, in the case where the temperature is not very high, or the system medium is not plasma, the third term in the right-hand side of (9.4.2) is very small and can be ignored. However, for the Sun's surface plasma fluid, this term will play a crucial role for the appearance of the solar flares. \square

Based on the anti-diffusive effect of heat (9.4.1), the concentration of photons is a reversed process of heat diffusion. Hence, by the Stefan-Boltzmann law, the third term on the right-hand side of (9.4.2) should be proportional to T^4 :

$$\text{the concentrating rate of photons} = \beta_0 T^4. \quad (9.4.4)$$

Equivalently the anti-diffusive effect of heat is expressed as

$$\left(\frac{dT}{dt} \right)_{ADE} = \beta_0 T^4,$$

where β_0 is the anti-diffusive effect coefficient.

In fact, it is the anti-diffusive effect of heat (9.4.1) that leads to the formation of sunspots, and it is the relation (9.4.4) that yields the solar flares and prominences.

(3) Sun's Surface Fluid Dynamical Equations

Sun's surface fluid is composed of gaseous plasma. The state functions describing solar flares and prominences are: the velocity field u of the plasma fluid, the temperature T , the electromagnetic fields \mathbf{E} and \mathbf{H} . Hence, the dynamical model governing Sun's surface fluid is the three groups of equations coupling the Navier-Stokes equations, the heat equation, and the Maxwell equations.

We start with the spatial domain given by

$$\Omega = S^2 \times (r_0, r_1),$$

where S^2 is the two-dimensional unit sphere, r_0 is the solar radius, and $r_1 = r_0 + h$ with the thickness of solar atmosphere h . We now describe the three set of equations, and their coupling.

1. The Navier-Stokes equations are written as

$$\rho \left[\frac{\partial u}{\partial t} + (u \cdot \nabla) u \right] = \mu \Delta u - \nabla p + f, \quad (9.4.5)$$

where ρ is the mass density, p is the pressure, and f is the force density. Because the fluid is plasma, each particle is charged. Hence, the force field f includes

$$f = \text{charge force} + \text{Lorentz force} + \text{thermal force}.$$

Based on classical electromagnetic theory,

$$\begin{aligned}\text{charge force} &= \rho_e \mathbf{E}, \\ \text{Lorentz force} &= J \times \mathbf{H},\end{aligned}$$

where ρ_e is the effective charge density in the plasma, $J = \rho_e u$ is the effective current density. By theory of thermodynamics, we have

$$\text{thermal force} = -g \mathbf{k} \rho (1 - \alpha T),$$

where g is the solar gravitational constant, α is the thermal expression coefficient, and \mathbf{k} is the radial unit vector.

Then equations (9.4.5) are expressed as

$$\rho \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right] = \mu \Delta \mathbf{u} - \nabla p + \rho_e \mathbf{E} + \rho_e \mathbf{u} \times \mathbf{H} - g \vec{\mathbf{k}} \rho (1 - \alpha T). \quad (9.4.6)$$

Also, (9.4.6) is complemented with the continuous equation

$$\frac{\partial \rho}{\partial t} + \text{div}(\rho \mathbf{u}) = 0. \quad (9.4.7)$$

2. In view of (9.4.2) and (9.4.4), the heat equation is given by

$$\frac{\partial T}{\partial t} + (\mathbf{u} \cdot \nabla) T = \kappa \Delta T + Q + \beta_0 T^4, \quad (9.4.8)$$

where κ is the heat conduction coefficient, and Q is the heat source excited by the solar electromagnetic fields, written as

$$Q = \beta_1 (\mathbf{E}^2 + \mathbf{H}^2).$$

Thus, the heat equation (9.4.8) becomes

$$\frac{\partial T}{\partial t} + (\mathbf{u} \cdot \nabla) T = \kappa \Delta T + \beta_0 T^4 + \beta_1 (\mathbf{E}^2 + \mathbf{H}^2). \quad (9.4.9)$$

3. The Maxwell equations read

$$\begin{aligned}\mu_0 \frac{\partial \mathbf{H}}{\partial t} &= -\operatorname{curl} \mathbf{E}, \\ \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} &= \operatorname{curl} \mathbf{H} - \mathbf{J}, \quad \mathbf{J} = \rho_e \mathbf{u}, \\ \operatorname{div} \mathbf{H} &= 0, \\ \operatorname{div} \mathbf{E} &= \rho_e,\end{aligned}\tag{9.4.10}$$

where μ_0 is the magnetic permeability, and ε_0 is the dielectric constant.

4. *Model for Sun's Surface Plasma Fluid* Combing the fluid dynamical equations (9.4.6)–(9.4.7), the heat equation (9.4.9), and the Maxwell equations (9.4.10), we derive the model governing Sun's surface plasma fluid as follows

$$\rho \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right] = \mu \Delta \mathbf{u} - \nabla p + \rho_e (\mathbf{E} + \mathbf{u} \times \mathbf{H}) - g \mathbf{k} \rho (1 - \alpha T),\tag{9.4.11}$$

$$\frac{\partial T}{\partial t} + (\mathbf{u} \cdot \nabla) T = \kappa \Delta T + \beta_0 T^4 + \beta_1 (\mathbf{E}^2 + \mathbf{H}^2),\tag{9.4.12}$$

$$\frac{\partial \mathbf{H}}{\partial t} = -\frac{1}{\mu_0} \operatorname{curl} \mathbf{E},\tag{9.4.13}$$

$$\frac{\partial \mathbf{E}}{\partial t} = \frac{1}{\varepsilon_0} \operatorname{curl} \mathbf{H} - \frac{1}{\varepsilon_0} \rho_e \mathbf{u},\tag{9.4.14}$$

$$\operatorname{div} \mathbf{H} = 0,\tag{9.4.15}$$

$$\operatorname{div} \mathbf{E} = \rho_e,\tag{9.4.16}$$

$$\frac{\partial \rho}{\partial t} = -\operatorname{div}(\rho \mathbf{u}).\tag{9.4.17}$$

The system (9.4.11)–(9.4.17) constitutes the basis to establish Sun's electromagnetic eruption theory.

9.4.2 Theory on Formation of Sunspots and Solar Eruptions

(1) Blow-Up Theorem

From the mathematical point of view, the phenomena of solar flares and prominences correspond to the blow-up of the solutions of the equations (9.4.11)–(9.4.17). Blow-up is a mathematical property that a solution $\Phi(x, t) = (u, T, \mathbf{H}, \mathbf{E})$ is called to blow-up at (x_0, t_0) if

$$\lim_{t \rightarrow t_0} |\Phi(x_0, t)| = \infty,\tag{9.4.18}$$

where $|\Phi|^2 = u^2 + T^2 + \mathbf{E}^2 + \mathbf{H}^2$.

Hence, the blow-up theorem for the system (9.4.11)–(9.4.17) introduced in the following is crucial for us to understand the Sun's electromagnetic eruptions. For simplicity, consider the case where ρ is a constant, and (9.4.17) becomes

$$\operatorname{div} u = 0. \quad (9.4.19)$$

Consider the initial and boundary conditions:

$$u_n|_{\partial\Omega} = 0, \quad \frac{\partial u_\tau}{\partial n}\Big|_{\partial\Omega} = 0, \quad \frac{\partial T}{\partial n}\Big|_{\partial\Omega} = 0, \quad (9.4.20)$$

$$\Phi = (u, T, \mathbf{H}, \mathbf{E})|_{t=0} = \Phi_0, \quad (9.4.21)$$

where n, τ are the unit normal and tangent vectors on $\partial\Omega$. Then we have the following theorem.

Theorem 9.4.2 (Blow-Up Theorem) *Let $\Phi(x, t)$ be a solution of (9.4.11)–(9.4.16) and (9.4.19), and satisfy the initial and boundary conditions (9.4.20) and (9.4.21). If the initial value Φ_0 is bounded, i.e.*

$$\sup_{\Omega} |\Phi_0(x)| < \infty,$$

then there exist $x_0 \in \Omega$ and $t_0 > 0$, such that the temperature T blows up at (x_0, t_0) , and consequently $\Phi(x, t)$ blows up (x_0, t_0) as well, i.e., (9.4.18) holds true.

Proof. Take the integration on both sides of (9.4.12):

$$\frac{d}{dt} \int_{\Omega} T dx = \int_{\Omega} [\kappa \Delta T - (u \cdot \nabla)T + \beta_0 T^4 + \beta_1 (\mathbf{E}^2 + \mathbf{H}^2)] dx. \quad (9.4.22)$$

In view of (9.4.19) and (9.4.20), by the Gauss formula, we have

$$\begin{aligned} \int_{\Omega} \Delta T dx &= \int_{\partial\Omega} \frac{\partial T}{\partial n} dx = 0, \\ \int_{\Omega} (u \cdot \nabla) T dx &= - \int_{\Omega} T \operatorname{div} u dx + \int_{\partial\Omega} T u_n dS = 0. \end{aligned}$$

Then (9.4.22) becomes

$$\frac{d}{dt} \int_{\Omega} T dx = \int_{\Omega} [\beta_0 T^4 + \beta_1 (\mathbf{E}^2 + \mathbf{H}^2)] dx. \quad (9.4.23)$$

Physically, $T > 0$, which can also be proved using the same method as in the paper (Foias et al., 1987). Hence, we have

$$\int_{\Omega} |T| dx = \int_{\Omega} T dx > 0. \quad (9.4.24)$$

In addition, by the anti-Hölder inequality (see Ma (2011)), for any $0 < p < 1$ and $q = p/(p-1)$ we have

$$\int_{\Omega} |fg| dx \geq \left[\int_{\Omega} |f|^p dx \right]^{1/p} \left[\int_{\Omega} |g|^q dx \right]^{1/q}. \quad (9.4.25)$$

Take $f = T^4$, $g = 1$, $p = 1/4$, and $q = -1/3$, and note (9.4.24), then (9.4.25) becomes

$$\int_{\Omega} T^4 dx \geq \frac{1}{|\Omega|^3} \left[\int_{\Omega} T dx \right]^4. \quad (9.4.26)$$

Recall the comparison theorem of differential equations: For two nonnegative functions f_1 and f_2 satisfying

$$f_1(t) \leq f_2(t), \quad \forall t \geq 0,$$

we consider the initial value problems

$$\begin{aligned} \frac{dx_1}{dt} &= f_1(t) \quad \text{with } x_1(0) = a, \\ \frac{dx_2}{dt} &= f_2(t) \quad \text{with } x_2(0) = a, \end{aligned}$$

where $a \geq 0$. Their solutions $x_1(t)$ and $x_2(t)$ satisfy

$$x_1(t) \leq x_2(t), \quad \forall t \geq 0.$$

Then, consider the equation

$$\frac{d}{dt} \int_{\Omega} T dx = \frac{\beta_0}{|\Omega|^3} \left[\int_{\Omega} T dx \right]^4. \quad (9.4.27)$$

Based on the comparison theorem, for the solution $\int_{\Omega} T dx$ of (9.4.23) and the solution $\int_{\Omega} T_1 dx$ of (9.4.27) with the initial value conditions

$$\int_{\Omega} T dx \Big|_{t=0} = a, \quad (9.4.28)$$

$$\int_{\Omega} T_1 dx \Big|_{t=0} = a, \quad (9.4.29)$$

we deduce from (9.4.26) that

$$\int_{\Omega} T_1 dx \leq \int_{\Omega} T dx, \quad \forall t \geq 0. \quad (9.4.30)$$

Denote $y = \int_{\Omega} T_1 dx$, then equation (9.4.27) with the initial value condition (9.4.29) is written as

$$\frac{dy}{dt} = k\beta_0 y^4, \quad y(0) = a, \quad (9.4.31)$$

where $k = 1/|\Omega|^3$. It is easy to see that the solution of (9.4.31) is

$$y = \frac{a}{[1 - 3ka^3\beta_0 t]^{1/3}}, \quad k = 1/|\Omega|^3. \quad (9.4.32)$$

Again, by (9.4.30) we have,

$$y \leq \int_{\Omega} T dx, \quad (9.4.33)$$

where T is the solution of (9.4.12). By (9.4.32), we see that

$$\lim_{t \rightarrow t_0} y(t) = \infty, \quad t_0 = \frac{1}{3ka^3\beta_0}.$$

Then we deduce from (9.4.33) that

$$\lim_{t \rightarrow t_0} \int_{\Omega} T dx = \infty, \quad (9.4.34)$$

which implies this theorem holds true.

(2) Sun's Surface Eruptions and Sunspots

The blow-up theorem, Theorem 9.4.2, provides the mathematical foundation for the solar eruption theory developed here. It shows that the eruptions are typically topological phase transitions at the blow-up point $t_0 = 1/3ka^3\beta_0$, at which the state functions $\Phi(x, t) = (u, T, \mathbf{H}, \mathbf{E})$ tend to infinite, depicting the huge and complicated explosions as high speed mass ejections, sudden flares of flashlight, very strong radiations, and large spurts of loop shaped magnetic energy.

Based on Theorem 9.4.2 and the anti-diffusive effect of heat (9.4.1), we discuss the sunspot and solar eruption problems in the following.

1. *Sunspots* We know that sunspots are the regions on Sun's surface, and possess the following two main characteristics:

- the temperature in sport is lower than the surrounding areas, and consequently the brightness is darker; and
- it often accompanies secondary events as the solar flares and prominences—the solar eruptions.

The above two characteristics can be well explained as follows by the anti-diffusive effect of heat (9.4.1) and the blow-up of the temperature:

- Due to thermal fluctuations, the temperature in the solar atmosphere is non-homogeneous, leading to elevated temperature in some local places. Based on the anti-diffusive effect, the regions with higher temperature absorb more photons from their surrounding areas, leading to the temperature decreasing, and consequently generating the sunspots;

- When the sunspots appear, the anti-diffusive effect (9.4.4) makes the temperature in the regions around sunspots rapidly increasing to generate the blow-up as described in Theorem 9.4.2, which causes solar eruptions.
2. *Solar Flares* In the proof of Theorem 9.4.2, we derived (9.4.34), which implies that there exists a point $x_0 \in \Omega$ such that

$$\lim_{t \rightarrow t_0} T(x_0, t) = \infty \quad \text{with } t_0 = \frac{1}{3ka^3\beta_0}. \quad (9.4.35)$$

It is the blow-up (9.4.35) that generates the solar flare. In addition, it also induces a chain of tragical variations in the velocity field u and the electromagnetic fields \mathbf{H} and \mathbf{E} , resulting in the coronal mass ejections, the strong radiations, and the solar prominences.

3. *Coronal Mass Ejections* Equation (9.4.11) dictates the behavior of mass ejections. When the temperature T blows up at (x_0, t_0) as given in (9.4.35), the maximal forces acting on the particles near x_0 are just ∇p , as $0 < \alpha T < 1$ in (9.4.11). Hence, in the neighborhood of (x_0, t_0) , (9.4.11) can be approximatively expressed as

$$\frac{du}{dt} = -\frac{1}{\rho} \nabla p. \quad (9.4.36)$$

By the gaseous equation of state:

$$p = R\rho T/m,$$

where R is the gas constant and m is the particle mass, the equation (9.4.36) is written as

$$\frac{du}{dt} = -\frac{R}{m} \nabla T. \quad (9.4.37)$$

By (9.4.35) we have

$$\lim_{t \rightarrow t_0} |\nabla T(x_0, t)| = \infty.$$

Therefore we deduce from (9.4.37) that

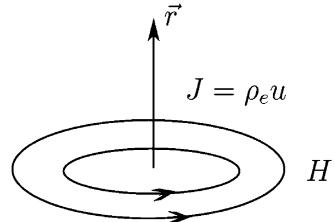
$$\lim_{t \rightarrow t_0} \left| \frac{du(x_0, t)}{dt} \right| = \infty. \quad (9.4.38)$$

The equality (9.4.38) represents the high speed gas explosion and particle ejections. The ejection direction is

$$\vec{r} = -\lim_{t \rightarrow t_0} \frac{\nabla T(x_0, t)}{|\nabla T(x_0, t)|}. \quad (9.4.39)$$

4. *Strong Radiations* The blow-up (9.4.38) of the velocity causes the formation of the high speed jet of charged particles, inducing the electromagnetic eruptions. By the Maxwell equations (9.4.13) and (9.4.14), we have

Fig. 9.19 Huge current jet J generates a violent magnetic loop, which is the solar prominence



$$\begin{aligned}\frac{\partial^2 \mathbf{E}}{\partial t^2} + \frac{1}{\varepsilon_0 \mu_0} \operatorname{curl}^2 \mathbf{E} &= -\frac{\rho_e}{\varepsilon_0} \frac{\partial \mathbf{u}}{\partial t}, \\ \frac{\partial^2 \mathbf{H}}{\partial t^2} + \frac{1}{\varepsilon_0 \mu_0} \operatorname{curl}^2 \mathbf{H} &= \frac{\rho_e}{\varepsilon_0} \mathbf{u}.\end{aligned}\quad (9.4.40)$$

In view of (9.4.38) and (9.4.39), the equations (9.4.40) generate very strong electromagnetic radiation in the \vec{r} direction.

5. *Solar Prominences* Consider the equation (9.4.14), which can be approximatively expressed as (i.e., the Ampère law):

$$\operatorname{curl} \mathbf{H} = \rho_e \mathbf{u}. \quad (9.4.41)$$

By (9.4.38) and (9.4.39), we have

$$\rho_e \mathbf{u} \cdot \vec{r} \gg 1 \quad \text{at } (x_0, t_0), \quad (9.4.42)$$

which represents the huge current jet in the \vec{r} direction.

Then, by the Ampère law (9.4.41), the strong current of (9.4.42) gives rise to violent magnetic loops, perpendicular to the direction \vec{r} in (9.4.39), as shown in Fig. 9.19. As the current orientation \vec{r} is not in the normal direction of solar surface, the erupting magnetic loops are the solar prominences that we observe.

(3) Period of Solar Eruptions

Astronomical observations show that sunspots and solar flares occur periodically in a 11-year cycle. The temperature solution (9.4.32) provides also the formula of solar eruption period, which is rewritten as

$$\int_{\Omega} T dx = \frac{a}{[1 - 3ka^3\beta_0 t]^{1/3}}, \quad (9.4.43)$$

where $k = 1/|\Omega|^3$, and a is the initial value

$$a = \int_{\Omega} T(x, 0) dx.$$

When a solar eruption ends, the temperature distribution in the solar chromosphere is almost homogenous. Let T_0 be the homogenous temperature, i.e., $T(x, 0) = T_0$. Then the initial value $a = T_0|\Omega|$, and (9.4.43) becomes

$$\int_{\Omega} T dx = \frac{|\Omega|T_0}{[1 - 3T_0^3\beta_0 t]^{1/3}}. \quad (9.4.44)$$

It is easy to deduce from (9.4.44) that the time t_p for the next eruption to occur satisfies

$$1 - 3T_0^3\beta_0 t_p = 0.$$

Hence the period t_p is

$$t_p = \frac{1}{3T_0^3\beta_0}, \quad (9.4.45)$$

where β_0 represents the heat effect coefficient, depending on the material, and T_0 is the initial average temperature.

The period formula (9.4.45) is also applicable to all stars, because it was found that other stars also have the eruption phenomena. For the Sun,

$$t_p = 11 \text{ years} = 3.469 \times 10^8 \text{ s},$$

and T_0 is the average temperature of the chromosphere:

$$T_0 = 1.5 \times 10^4 \text{ K}.$$

Therefore, the heat effect coefficient for solar gaseous plasma is derived from (9.4.45) as follows

$$\beta_0 = 2.85 \times 10^{-22} / (\text{K}^3 \cdot \text{s}). \quad (9.4.46)$$

Remark 9.4.3 The heat effect coefficients for non-plasma materials are much smaller than the value in (9.4.46). \square

(4) Non-Blow-Up Condition of Temperature

In this section, we demonstrate that under the Dirichlet boundary condition for the temperature T ,¹ the *non-blow-up condition* for temperature in domain Ω is

$$\kappa > \frac{16}{9}\beta_0|\Omega|^{1/15} \left[\int_{\Omega} T_0^5 dx \right]^{3/5}, \quad (9.4.47)$$

where T_0 is the initial temperature, κ is the heat conduction coefficient, and β_0 is the heat effect coefficient. Note that β_0 is negligible for non-plasma material. Therefore,

¹ The Dirichlet boundary condition for T amounts to saying that there is heat exchange of the system with outside. Hence there is no contradiction between the non-blow-up condition and the blow-up theorem 9.4.2, where the Neumann boundary condition for T is used.

in the usual cases we often encounter, both T_0 and β_0 are small, so that no temperature blow-up occurs.

To verify (9.4.47), we consider the heat conduction equation with anti-diffusive effect of heat:

$$\begin{aligned}\frac{\partial T}{\partial t} &= \kappa \Delta T + \beta_0 T^4, \quad x \in \Omega \subset \mathbb{R}^3, \\ T|_{\partial\Omega} &= 0.\end{aligned}\tag{9.4.48}$$

Multiplying both sides of (9.4.48) by T and taking integral, we have

$$\frac{1}{2} \frac{d}{dt} \int_{\Omega} T^2 dx = \int_{\Omega} [-\kappa |\nabla T|^2 + \beta_0 T^5] dx.\tag{9.4.49}$$

By the Sobolev inequality (see, e.g., p. 23 in Ma (2011)):

$$\left[\int_{\Omega} |\nabla T|^2 dx \right]^{1/2} \geq \frac{n(n-2)}{2(n-1)} \left[\int_{\Omega} T^{\frac{2n}{n-2}} dx \right]^{\frac{n-2}{2n}},$$

where n is the dimension of space.² For the study presented here, $n = 3$, and we have

$$\int_{\Omega} |\nabla T|^2 dx \geq \frac{9}{16} \left[\int_{\Omega} T^6 dx \right]^{1/3}.\tag{9.4.50}$$

In addition, by the Hölder inequality, we derive

$$\|T\|_{L^6} \geq \frac{1}{|\Omega|^{1/30}} \|T\|_{L^5}.$$

Hence

$$\left[\int_{\Omega} T^6 dx \right]^{1/3} \geq \frac{1}{|\Omega|^{1/15}} \left[\int_{\Omega} T^5 dx \right]^{2/5}.\tag{9.4.51}$$

It follows from (9.4.50) and (9.4.51) that

$$\int_{\Omega} |\nabla T|^2 dx \geq \frac{9}{16|\Omega|^{1/15}} \left[\int_{\Omega} T^5 dx \right]^{2/5}.$$

Then we deduce from (9.4.49) that

$$\frac{1}{2} \frac{d}{dt} \int_{\Omega} T^2 dx \leq -\frac{9\kappa}{16|\Omega|^{1/15}} \left[\int_{\Omega} T^5 dx \right]^{2/5} + \beta_0 \int_{\Omega} T^5 dx\tag{9.4.52}$$

² We remark here that by the above inequality and the Hölder inequality, we have

$$\|\nabla u\|_{L^p} \geq \frac{c}{|\Omega|^{(N^*-q)/N^*q}} \|u\|_{L^q} \quad \text{for } n > p, N^* = \frac{np}{n-p}.$$

$$\leq - \left[\frac{9\kappa}{16|\Omega|^{1/15}} - \beta_0 \left(\int_{\Omega} T^5 dx \right)^{3/5} \right] \left[\int_{\Omega} T^5 dx \right]^{\frac{2}{5}}.$$

It is clear that if

$$\kappa \geq \frac{16}{9} \beta_0 |\Omega|^{1/15} \left[\int_{\Omega} T^5 dx \right]^{3/5} + \varepsilon, \quad \forall t > 0, \quad (9.4.53)$$

where $\varepsilon > 0$ is arbitrarily small, then we have

$$\lim_{t \rightarrow \infty} \int_{\Omega} T^2 dx = 0. \quad (9.4.54)$$

In fact, by the inequality

$$\left[\int_{\Omega} T^5 dx \right]^{2/5} \geq C_0 \left[\int_{\Omega} T^2 dx \right],$$

and by the Gronwall inequality, we deduce from (9.4.52) that

$$\int_{\Omega} T^2 dx \leq e^{-C_0 \int_0^t \alpha(\tau) d\tau}, \quad (9.4.55)$$

where by (9.4.53)

$$\alpha = \frac{9\kappa}{16|\Omega|^{1/15}} - \beta_0 \left[\int_{\Omega} T^5 dx \right]^{3/5} \geq \frac{9\varepsilon}{16|\Omega|^{1/15}}.$$

Then, (9.4.55) implies that (9.4.54) holds true, and consequently the thermal system (9.4.48) has no blow-up. Physically, (9.4.53) means that (9.4.47) is the condition to ensure no temperature blow-up.

9.4.3 Plasma Fuel

The anti-diffusive effect of heat (9.4.1) may provide a new source of fuel. Basically, in a high temperature plasma system where free electrons are abundant, photons in the low temperature area experience the anti-diffusive effect and can move to the high temperature region, further raising the temperature in the high temperature region. The conversion of the thermal energy in the high-temperature region offers an alternative fuel source, which we call plasma fuel. The Sun is a natural source of photons for the low-temperature regions required for such a process to continue. This direct way of converting solar energy into a new form of fuel might be much more efficient than the photovoltaic devices.

In view of the non-blow-up condition (9.4.47), the following condition is required for the anti-diffusive effect of heat to dominate the normal heat conduction:

$$\kappa < \frac{16}{9} \beta_0 |\Omega|^{1/15} \left[\int_{\Omega} T_0^5 dx \right]^{3/5}, \quad (9.4.56)$$

where T_0 is the initial temperature, κ is the heat conduction coefficient, and β_0 is the anti-diffusive effect coefficient. Both κ and β_0 depend on the plasma material. Therefore, for a given plasma system, the larger the domain and the higher the initial temperature, the stronger the anti-diffusive effect. Consequently an effective plasma device for converting solar energy into a fuel source can be achieved by increasing the domain size and/or the initial temperature.

In conclusion, the anti-diffusive effect of heat (9.4.1) may lead to a new efficient way of converting solar energy directly to a form of fuel, much more efficient than the photovoltaic devices.

9.5 Dynamic Theory of Boundary-Layer Separations

Boundary-layer separation phenomenon is one of the most important processes in fluid flows, with a long history of studies which go back to the pioneering work of Prandtl (1904). Basically, in the boundary-layer, the shear flow can detach/separate from the boundary, generating vortices and leading to more complicated turbulent behavior. It is important to characterize the separation.

We present in this section a systematic dynamic theory for boundary-layer separations of fluid flows and its applications to large scale ocean circulations. The study here is based on Ma and Wang (2017h).

9.5.1 Phenomena and Problems on Boundary-Layer Separations

(1) Boundary-Layer Separation Phenomena

Boundary-layer separation is a universal phenomenon in fluid flows, and says that a shear flow near the boundary generates suddenly vortices from the boundary. More precisely, we say that a velocity field $u(x, t)$ has a boundary-layer separation near $\bar{x} \in \partial\Omega$ at $t_0 > 0$, if $u(x, t)$ is topological equivalent to the structure as shown in Fig. 9.20a for $t < t_0$, and to the structure as in Fig. 9.20b for $t > t_0$. Namely, if $t < t_0$, then $u(x, t)$ is topological equivalent to a parallel shear flow, and if $t > t_0$, $u(x, t)$ bifurcates to a vortex from $\bar{x} \in \partial\Omega$.

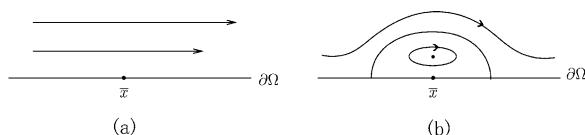


Fig. 9.20 (a) A parallel shear flow, and (b) a boundary vortex flow

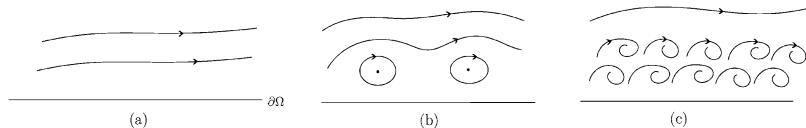


Fig. 9.21 (a) If the surface velocity $u_0 < u_c$, surface flow is parallel; (b) boundary-layer separation occurs for u_0 near u_c ; and (c) if u_0 exceeds u_c , surface turbulence appears

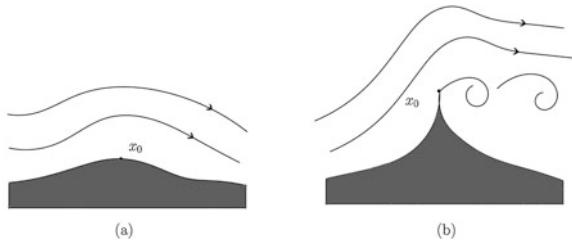


Fig. 9.22 (a) As the curvature $k(x_0) < k_c$, no vortices near x_0 , (b) As $k(x_0) > k_c$, the vortices appear

In the following, we give three remarkable physical phenomena associated with boundary-layer separations.

1. *Formation of Surface Turbulence* A surface flow is the fluid motion on a boundary surface, as the parallel shear flows. We know that if the velocity of a surface flow exceeds certain critical value, then the boundary-layer separation will lead to turbulence. During the transition from a parallel shear flow to a surface turbulence, boundary-layer separation must occur. Figure 9.21 provides a schematic diagram to illustrate the formation process of surface turbulence.
2. *Vortex Flows Near a Boundary Tip Point* At a tip point on the boundary, a fluid flow generates vortices, as shown in Fig. 9.22b; see also Oertel (2001, Section 4.1.4). Let $x_0 \in \partial\Omega$ be a tip point with curvature $k(x_0)$. For a given injection velocity u_0 , there is a critical curvature k_c such that if $k(x_0) < k_c$, there are no vortices forming near x_0 for the boundary-layer flow shown in Fig. 9.22a, and if $k(x_0) > k_c$ ($k(x_0) = \infty$ at a cusp point), the vortices appear near x_0 ; see Fig. 9.22b.
3. *Wind-Driven Atlantic Gyres* The oceanic gyres are typical large structure of large-scale ocean surface circulations. Figure 9.23 provides the map of the five major oceanic gyres, where the left one is the Indian ocean gyre, the middle two are the Pacific gyres, and the right two are the Atlantic gyres.
The Atlantic gyres consist of the north Atlantic gyre located in the northern Atlantic and the south Atlantic gyre in the southern Atlantic.
In the northern Atlantic, the trade winds blow westward in the tropics and the westerly winds blow eastward in mid-latitudes. This wind-driven ocean surface flows form a huge and clockwise gyre in the mid-latitude ocean basin, usually called the north Atlantic gyre. In the north of the north Atlantic gyre, there is the north Atlantic subpolar gyre. The mid-latitude gyre and subpolar gyre together are called the north Atlantic double gyres, which are present permanently; see Fig. 9.24a. Due to the influences of the Gulf Stream and seasonal winds, vortices

Fig. 9.23 A global map of major ocean gyres

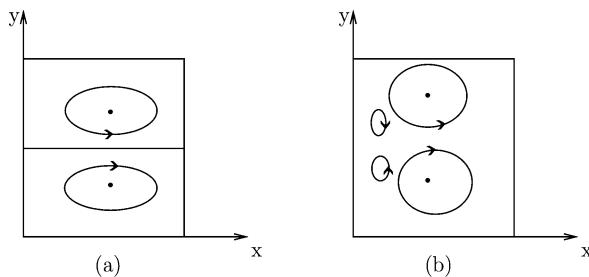
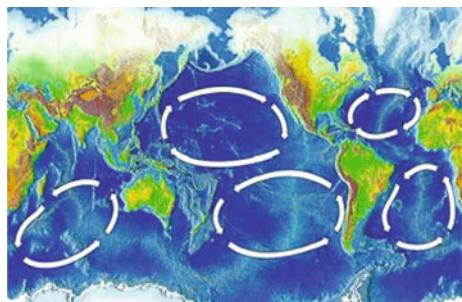


Fig. 9.24 (a) North Atlantic double gyres, which are present permanently; (b) vortices separated from the North Atlantic coast

are separated from the margins of the gyres, and are essentially the phenomena of boundary-layer separation; see Fig. 9.24b.

(2) Main Problems

Based on the physical phenomena of boundary-layer separations given above, we now state some basic problems that need to be addressed by the theory of boundary-layer separations.

The boundary-layer separation phenomenon is a typical example of topological phase transitions, and is governed by fluid dynamical equations. Consider the Navier-Stokes equations

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla) u &= \nu \Delta u - \frac{1}{\rho} \nabla p + f, \quad x \in \Omega \subset \mathbb{R}^2, \\ \operatorname{div} u &= 0, \\ u|_{\partial\Omega} = 0 \quad \left(\text{or } u_n|_{\partial\Omega} = 0, \quad \frac{\partial u_\tau}{\partial n} \Big|_{\partial\Omega} = 0 \right), \\ u(0) &= \varphi(x). \end{aligned} \tag{9.5.1}$$

The state function describing boundary-layer separations is the 2D velocity field $u(x, \lambda)$, where λ is the parameter controlling the topological structural transitions of u , and can be taken as the following physical quantities

$$\lambda = t, \nu, f, \varphi, k, \quad (9.5.2)$$

where t is the time, ν is the viscosity, f is the external force, φ is the initial velocity, and k is the curvature of $\partial\Omega$.

The main aspects of a boundary-layer separation theory should include the following:

1. establish conditions under which the separations occur and determine the critical thresholds;
2. (predicable problem) develop a theory to determine when, where and how the separation occurs, based on the observable data such as the initial velocity φ , the external force f , and the geometric condition of $\partial\Omega$, etc.;
3. for the surface turbulent phenomena as shown in Fig. 9.21, if we take the initial injection velocity u_0 as the control parameter, then we need to know the critical velocity u_c for the turbulence to occur; and
4. at the tip point vortices as shown in Fig. 9.22b, provide a relation between the critical curvature k_c , the surface frictional coefficient β of material, and other related physical parameters.

In the remaining part of this section, we develop a systematic boundary-layer theory, based on the geometric theory of 2D incompressible flows recapitulated in Sect. 9.1. The theory we establish solves the problems listed above associated with the main components of a boundary-layer theory.

9.5.2 Basic Boundary-Layer Theory

(1) Boundary-Layer Separation Theorem

Let $u(x, t)$ be the solution of the Navier–Stokes equations (9.5.1), and $\Gamma \subset \partial\Omega$ be an open part. For $\bar{x} \in \Gamma$, if there is a time $t_0 > 0$ such that the following conditions hold true:

$$\begin{cases} \bar{x} \in \Gamma \text{ is an isolated zero of } \frac{\partial u_\tau(\cdot, t_0)}{\partial n}, \\ \left. \frac{\partial u_\tau}{\partial n} \right|_{\Gamma} \neq 0 \forall t < t_0, \end{cases} \quad \text{for (9.1.2),}$$

$$\begin{cases} \bar{x} \in \Gamma \text{ is an isolated zero of } u_\tau(\cdot, t_0), \\ u_\tau|_{\Gamma} \neq 0 \forall t < t_0, \end{cases} \quad \text{for (9.1.1),}$$

$$\begin{cases} u_\tau(\bar{x}, t_0) = 0 \end{cases}$$

then by the homotopic invariance of the topological index, we have

$$\begin{aligned} \text{ind}\left(\frac{\partial u(\cdot, t_0)}{\partial n}, \bar{x}\right) &= 0 \quad \text{for (9.1.2),} \\ \text{ind}(u(\cdot, t_0), \bar{x}) &= 0 \quad \text{for (9.1.1).} \end{aligned} \quad (9.5.4)$$

We also assume that

$$\begin{aligned} \frac{\partial}{\partial t} \frac{\partial u_\tau(\bar{x}, t_0)}{\partial n} &\neq 0 \quad \text{for (9.1.2),} \\ \frac{\partial}{\partial t} u_\tau(\bar{x}, t_0) &\neq 0 \quad \text{for (9.1.1).} \end{aligned} \quad (9.5.5)$$

It is clear that conditions (9.5.3) replace conditions (9.5.4). Then by Theorems 9.1.7 and 9.1.8, we derive the following boundary-layer separation theorem for the solutions of the Navier–Stokes equation (9.5.1).

Theorem 9.5.1 (Boundary-Layer Separation) *Let $u(x, t)$ be the solution of the Navier–Stokes equation (9.5.1). If $u(x, t)$ satisfies (9.5.3) and (9.5.5), then u has a boundary-layer separation at (\bar{x}, t_0) .*

Remark 9.5.2 Let $X = B_0^r(\Omega, \mathbb{R}^2)$ or $B^r(\Omega, \mathbb{R}^2)$. Mathematically, conditions (9.5.5) are generic. Namely, there is an open and dense set $U \subset X \times L^2(\Omega, \mathbb{R}^2)$, such that for any $(\varphi, f) \in U$ the solution $u(x, t)$ of (9.5.1) satisfies (9.5.5). Hence conditions (9.5.5) are a physically sound condition. Moreover, in Theorem 9.5.1, the conditions (9.5.5) can also be replaced by

$$\frac{\partial u_\tau}{\partial n}(\bar{x}, t) \quad \text{or} \quad u_\tau(\bar{x}, t) \begin{cases} > 0 & \text{for } t < t_0 \quad (\text{or } t > t_0), \\ = 0 & \text{for } t = t_0, \\ < 0 & \text{for } t > t_0 \quad (\text{or } t < t_0). \end{cases}$$

(2) Separation Equations for the Rigid Boundary Condition

To verify the condition (9.5.3), it is very useful to transform the Navier–Stokes equations (9.5.1) into the separation equation introduced in this and next subsections.

1. *Separation Equation in (τ, n) -Representation* In Ma and Wang (2005d), we proved that the Navier–Stokes equations (9.5.1) lead to the following boundary-layer separation equation of (9.5.1) under the rigid boundary condition:

$$\frac{\partial u_\tau(x, t)}{\partial n} = \frac{\partial \varphi_\tau}{\partial n} + \int_0^t [\nu \nabla \times \Delta u + k \nu \Delta u \cdot \tau + \nabla \times f + k f_\tau] dt, \quad (9.5.6)$$

where n, τ are the unit normal and tangent vectors on $\partial\Omega$, k is the curvature of $\partial\Omega$ at $x \in \partial\Omega$, and

$$\nabla \times v = \frac{\partial v_\tau}{\partial n} - \frac{\partial v_n}{\partial \tau}. \quad (9.5.7)$$

2. *Separation Equation in (x_1, x_2) -Representation* Consider an orthogonal coordinate system (x_1, x_2) , the equivalent separation equation of the Navier–Stokes equations was derived in Wang et al. (2015b):

$$\frac{\partial u_\tau(x, t)}{\partial n} = -\nabla \times \varphi - \int_0^t [\nu \nabla \times \Delta u + \nabla \times f] dt, \quad x \in \partial\Omega, \quad (9.5.8)$$

where $\nabla \times v = \partial v_2 / \partial x_1 - \partial v_1 / \partial x_2$ for $v = (v_1, v_2)$.

Remark 9.5.3 Although the curvature k of $\partial\Omega$ doesn't appear in the separation equation (9.5.8), the curvature of $\partial\Omega$ is hidden in $\nabla \times \Delta u$ and $\nabla \times f$, i.e., in the curl of Δu and f in the orthogonal coordinate system.

(3) Separation Equation for the Free Boundary Condition

Consider the Navier–Stokes equations with the free boundary condition:

$$\frac{\partial u}{\partial t} + (u \cdot \nabla)u = \nu \Delta u - \frac{1}{\rho} \nabla p + f, \quad (9.5.9)$$

$$\operatorname{div} u = 0, \quad (9.5.10)$$

$$\begin{cases} u_n = 0, & \frac{\partial u_\tau}{\partial n} = 0 \quad \text{on } \partial\Omega, \\ u(x, 0) = \varphi(x). \end{cases} \quad (9.5.11)$$

To deduce the separation equation of (9.5.9)–(9.5.11), we recall that for any external $f \in L^2(\Omega, \mathbb{R}^2)$, as a vector field we know that f can be uniquely decomposed as

$$f = F + \nabla\phi, \quad \operatorname{div} F = 0, \quad \text{and } F_n = 0 \quad \text{on } \partial\Omega. \quad (9.5.12)$$

Also, $(u \cdot \nabla)u$ is decomposed as

$$\begin{aligned} (u \cdot \nabla)u &= g(u) + \nabla\Phi(u), \\ \operatorname{div} g(u) &= 0, \quad g_n|_{\partial\Omega} = 0. \end{aligned} \quad (9.5.13)$$

Then in the following we shall prove that the separation equation of (9.5.9)–(9.5.11) is written in the following form

$$u_\tau(x, t) = \varphi_\tau(x) + \int_0^t \left[\nu \left(\frac{\partial^2 u_\tau}{\partial \tau^2} + \frac{\partial^2 u_\tau}{\partial n^2} \right) - g_\tau(u) + F_\tau \right] dt, \quad (9.5.14)$$

where F and g are as in (9.5.12) and (9.5.13).

Remark 9.5.4 If the portion of the boundary $\Gamma \in \partial\Omega$ is flat:

$$\Gamma = \{(x_1, x_2) \in \partial\Omega \mid -\delta < x_1 < \delta, x_2 = 0\}, \quad (9.5.15)$$

and the velocity gradient $\partial u_1 / \partial x_1$ is small on Γ , then the equation (9.5.14) can be approximatively written as

$$\begin{aligned} u_1(x, t) &= \varphi_1 + \int_0^t [\nu \Delta u_1 + F_1 - A] dt, \quad x = (x_1, 0), \\ A &= u_1 \frac{\partial u_1}{\partial x_1} \Big|_{x=0}, \end{aligned} \tag{9.5.16}$$

where u_1 , φ_1 , and F_1 are the x_1 -components of u , φ , and F . \square

We are now in position to verify (9.5.14). By (9.5.10) and (9.5.11), we induce from the free boundary condition that

$$\Delta u \cdot n|_{\partial\Omega} = 0,$$

and consequently

$$\operatorname{div}(\Delta u) = 0.$$

Hence, by (9.5.12) and (9.5.13), equation (9.5.9) can be decomposed as

$$\begin{aligned} \frac{\partial u}{\partial t} &= \nu \Delta u - g(u) + F, \\ \frac{1}{\rho} p &= \phi - \Phi(u). \end{aligned} \tag{9.5.17}$$

Then, we derive from the first equation of (9.5.17) that

$$u_\tau = \varphi_\tau + \int_0^t [\nu \Delta u_\tau - g_\tau(u) + F_\tau] dt,$$

which is the equation (9.5.14).

We now verify (9.5.16). By (9.5.11), it is clear that

$$(u \cdot \nabla) u \cdot \tau = u_1 \frac{\partial u_1}{\partial x_1} \quad \text{on } \Gamma. \tag{9.5.18}$$

By (9.5.13), in the divergence-free part $g = (g_1, g_2)$ of $(u \cdot \nabla) u$, $g_1|_{x=0}$ represents the leading order of g_1 . Hence the Taylor expression of g_1 on Γ is given by

$$g_1(x_1) = g_1(0) + x_1 f(x_1), \quad f_1 = g'(0) + \frac{1}{2} g''(0) x_1 + o(x_1).$$

Since $\partial u_1 / \partial x_1$ is small on Γ , by (9.5.13) and (9.5.18) we have

$$g_1(0) \gg x_1 f(x_1) \quad \text{for } (x_1, 0) \in \Gamma.$$

Hence, we get that

$$g_1(0) = u_1 \frac{\partial u_1}{\partial x_1} \Big|_{x=0} \simeq g_\tau(x) \text{ on } \Gamma.$$

Replacing g_τ in (9.5.14) by $g_1(0)$, we obtain (9.5.16).

We note that the separation equation (9.5.16) is more useful than (9.5.14) in dealing with fluid boundary-layer separations.

9.5.3 Predictable Conditions and Critical Thresholds

(1) Predictable Condition

Predictable problem for boundary-layer separations of 2D fluid flows governed by Navier–Stokes equations is an important topic in both classical and geophysical fluid dynamics. It mainly concerns when, where, and how a boundary-layer separation will occur. In particular, we need to know the conditions for the separation to appear in terms of the initial values and the external forces that are observable. Based on the basic theory introduced in the last section, we now address this problem.

1. *Predictable Condition of Flows with Dirichlet Boundary Condition on a Flat Boundary* When the separation occurs on a flat portion Γ of $\partial\Omega$, a predictable condition was given in Luo et al. (2015b). For the sake of completeness, we introduce it in the following.

Let $\Gamma \subset \partial\Omega$ be as in (9.5.15), and $x_2 > 0$ be the interior of Ω . By

$$\varphi|_{\partial\Omega} = 0, \quad \operatorname{div}\varphi = 0,$$

near Γ , the initial value $\varphi = (\varphi_1, \varphi_2)$ can be expressed as

$$\begin{aligned} \varphi_1 &= x_2 \varphi_{11}(x_1) + x_2^2 \varphi_{12}(x_1) + x_2^3 \varphi_{13}(x_1) + o(x_2^3), \\ \varphi_2 &= x_2^2 \varphi_{21}(x_1) + o(x_2^3), \\ \varphi_{21} &= -\frac{1}{2} \varphi'_{11}. \end{aligned} \tag{9.5.19}$$

The first-order Taylor expression of the external force f on x_2 near Γ is given by

$$\begin{aligned} f_1 &= f_{11}(x_1) + x_2 f_{12}(x_1) + o(x_2), \\ f_2 &= f_{21}(x_1) + x_2 f_{22}(x_1) + o(x_2). \end{aligned} \tag{9.5.20}$$

If the following condition holds true

$$0 < \min_{\Gamma} \frac{-\varphi_{11}}{2\nu\varphi''_{11} + 6\nu\varphi_{13} + f_{12} - f'_{21}} \ll 1, \tag{9.5.21}$$

then there are $t_0 > 0$ and $x_0 \in \Gamma$ such that the solution u of (9.5.1) has a boundary layer separation at (x_0, t_0) , where ν is the viscosity, φ_{11} , φ_{13} , f_{12} , f_{21} are as in (9.5.19) and (9.5.20). In addition, t_0 and $x_0 = (x_1^0, 0)$ approximately satisfy

$$t_0 = g(x_0), \quad \text{and} \quad g(x_0) = \min_{\Gamma} g(x), \quad (9.5.22)$$

where

$$g = \frac{-\varphi_{11}}{2\nu\varphi_{11}'' + 6\nu\varphi_{13} + f_{12} - f_{21}'}.$$

The condition (9.5.21), expressed in terms of the initial value φ and the external force f , provides a criterion for a boundary-layer separation to occur at (x_0, t_0) , and the relations in (9.5.22) give the time t_0 and the position x_0 where the separation occurs.

The proof (9.5.21) is based on Theorem 9.5.1 by applying the separation equation (9.5.6). In fact, on Γ (9.5.6) can be written as

$$\frac{\partial u_1(x, t)}{\partial x_2} = \frac{\partial \varphi_1}{\partial x_2} + \int_0^t \left[\nu \frac{\partial \Delta u_1}{\partial x_2} - \nu \frac{\partial \Delta u_2}{\partial x_1} + \frac{\partial f_1}{\partial x_2} - \frac{\partial f_2}{\partial x_1} \right] dt. \quad (9.5.23)$$

Take the first-order expression of u on t as

$$u = \varphi + tu'(x, t),$$

and insert it in the right-hand side of (9.5.23). By (9.5.19) and (9.5.20), we obtain that

$$\frac{\partial u_1}{\partial x_2} = \varphi_{11} + t[\nu(\varphi_{11}'' + 6\varphi_{13} - 2\varphi_{21}') + f_{12} - f_{21}'] + o(t).$$

Note that $\varphi_{21} = -\frac{1}{2}\varphi_{11}''$. Then we have

$$\frac{\partial u_1}{\partial x_2} = \varphi_{11} + t[\nu(2\varphi_{11}'' + 6\varphi_{13}) + f_{12} - f_{21}'] + o(t). \quad (9.5.24)$$

Then, under the condition (9.5.21), we deduce that there are $x_0 \in \Gamma$ and $t_0 > 0$ sufficiently small such that

$$\begin{aligned} \frac{\partial u_\tau}{\partial n}(t, x_0) &\begin{cases} \neq 0 & \text{for } 0 \leq t < t_0, \\ = 0 & \text{for } t = t_0. \end{cases} \\ \frac{\partial}{\partial t} \frac{\partial u_\tau}{\partial n} \Big|_{(t_0, x_0)} &\neq 0. \end{aligned}$$

Conclusions of (9.5.21) and (9.5.22) follow then from Theorem 9.5.1.

2. *Predictable Condition for Flows with Dirichlet Boundary Condition on a Curved Boundary* On a curved boundary, a predictable condition for boundary-layer separations can be derived from the separation equations (9.5.6) and (9.5.8) respectively, also based on Theorem 9.5.1.

Let $\Gamma \subset \partial\Omega$ be an open and curved portion of the boundary. By (9.5.6), we can derive the following predictable condition:

$$0 < \min_{\Gamma} \frac{-\frac{\partial \varphi_\tau}{\partial n}}{\nu(\nabla \times \Delta \varphi + k \Delta \varphi \cdot \tau) + \nabla \times f + k f_\tau} \ll 1, \quad (9.5.25)$$

where $k = k(x)$ is the curvature of $\partial\Omega$ at $x \in \Gamma$, and $\nabla \times v$ is as in (9.5.7) for the vector field v .

The other predictable condition in the orthogonal coordinate (x_1, x_2) can be derived from (9.5.8) as follows (see Wang et al. (2015b)),

$$0 < \min_{\Gamma} \frac{\frac{\partial \varphi_2}{\partial x_1} - \frac{\partial \varphi_1}{\partial x_2}}{\nu \left(\frac{\partial \Delta \varphi_1}{\partial x_2} - \frac{\partial \Delta \varphi_2}{\partial x_1} \right) + \frac{\partial f_1}{\partial x_2} - \frac{\partial f_2}{\partial x_1}} \ll 1. \quad (9.5.26)$$

3. *Predictable Conditions of Free Boundary Condition* By the separation equations (9.5.14) and (9.5.16), based on Theorem 9.5.1, we can also derive predictable conditions for fluid flows with the free boundary condition.

If $\Gamma \subset \partial\Omega$ is a flat portion of the boundary, as given by (9.5.15), we have the following criterion to determine a boundary-layer separation to occur on Γ :

$$0 < \min_{\Gamma} \frac{-\varphi_1}{\nu \Delta \varphi_1 + F_1 - \varphi_1 \frac{\partial \varphi_1}{\partial x_1}|_{x=0}} \ll 1, \quad (9.5.27)$$

where φ_1 and F_1 are the x_1 -components of φ and F , and F is as in (9.5.12). If $\Gamma \subset \partial\Omega$ is curved, then by (9.5.14) we deduce the following predictable condition:

$$0 < \min_{\Gamma} \frac{-\varphi_\tau}{\nu \left(\frac{\partial^2 \varphi_\tau}{\partial \tau^2} + \frac{\partial^2 \varphi_\tau}{\partial n^2} \right) + F_\tau - g_\tau(\varphi)} \ll 1, \quad (9.5.28)$$

where g is as in (9.5.13).

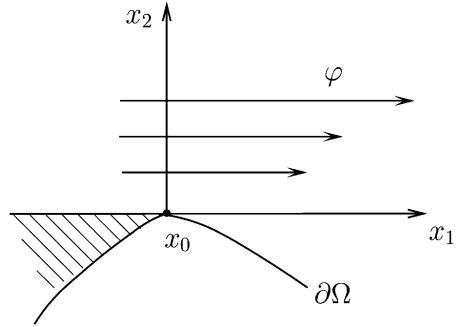
The five criteria (9.5.21) and (9.5.25)–(9.5.28) for boundary-layer separations are very useful in wide range of applications.

Remark 9.5.5 With conditions (9.5.21) and (9.5.25)–(9.5.28), the critical time t_0 might appear to be very small; see for example (9.5.22). However, this misperception is related to scaling. For a large scale fluid motion, the underlying dimensions are large, and the time t_0 is not small in real applications.

(2) Critical Curvature of Boundary Tip Point

To discuss vortices separated from a boundary tip point we need to use the separation equation (9.5.6). Let t_0 be the elapsed time between the instant when a vortex separated from the tip point disappears and the time when a new one forms. The elapsed time t_0 is called the relaxation time for the tip point separation.

Fig. 9.25 A schematic diagram of parallel gradient flow near a boundary tip point x_0 , the shadow part is a dead angle where fluid velocity $u = 0$



Let $x_0 \in \partial\Omega$ be a boundary tip point with curvature k . Take k as the control parameter. Then, equation (9.5.6) becomes

$$\begin{aligned} \frac{\partial u_\tau(k)}{\partial n} &= \frac{\partial \varphi_\tau(x_0)}{\partial n} + \left(\frac{\partial f_\tau(x_0)}{\partial n} - \frac{\partial f_n(x_0)}{\partial \tau} + k f_\tau(x_0) \right) t_0 \\ &\quad + \int_0^{t_0} \nu \left[\frac{\partial(\Delta u \cdot \tau)}{\partial n} - \frac{\partial(\Delta u \cdot n)}{\partial \tau} + k \Delta u \cdot \tau \right]_{x_0} dt. \end{aligned} \quad (9.5.29)$$

Here, the initial injection flow $\varphi(x)$ is taken parallel to the tangent vector τ at x_0 , i.e., φ represents a parallel gradient flow as

$$\varphi = (ax_2, 0) \quad \text{near } x_0 \in \partial\Omega, \quad (9.5.30)$$

where the coordinate (x_1, x_2) is taken so that the x_1 -axis is in the tangent τ -direction, x_2 -axis is in the normal n -direction, and x_0 is the origin of this coordinate system; see Fig. 9.25. By the definition of the relaxation time, it is known that t_0 is very small. Hence we have

$$u(x, t) \simeq \varphi \quad \forall 0 \leq t < t_0 \quad \text{near } x_0 \in \partial\Omega.$$

By (9.5.30), this implies that

$$\int_0^{t_0} \left[\frac{\partial(\Delta u \cdot \tau)}{\partial n} - \frac{\partial(\Delta u \cdot n)}{\partial \tau} + k \Delta u \cdot \tau \right]_{x_0} dt \simeq 0,$$

and (9.5.29) can be approximatively written as

$$\frac{\partial u_\tau(k)}{\partial n} = a + k f_\tau t_0 + \left(\frac{\partial f_\tau}{\partial n} - \frac{\partial f_n}{\partial \tau} \right) t_0. \quad (9.5.31)$$

Because f is the resistance force generated by the friction at x_0 , it is reversely parallel to φ , and we have

$$\frac{\partial f_n}{\partial \tau} = 0, \quad \frac{\partial f_\tau}{\partial n} \simeq 0.$$

Thus, (9.5.31) becomes

$$\frac{\partial u_\tau(k)}{\partial n} = a + k f_\tau t_0 \quad (a > 0, f_\tau < 0). \quad (9.5.32)$$

For the control parameter k (instead of t), it follows from (9.5.32) that

$$\begin{aligned} \frac{\partial u_\tau(k)}{\partial n} &\begin{cases} < 0 & \text{for } k < k_c, \\ = 0 & \text{for } k = k_c, \end{cases} \\ \frac{\partial}{\partial k} \frac{\partial u_\tau(k_c)}{\partial n} &= f_\tau t_0 \neq 0, \end{aligned} \quad (9.5.33)$$

where k_c is the critical curvature, given by

$$k_c = \frac{a}{|f_\tau|t_0}. \quad (9.5.34)$$

Namely, by the boundary-layer separation theorem, Theorem 9.5.1, under the conditions (9.5.33), the critical curvature k_c in (9.5.34) is the critical threshold where vortices begin to form.

Physically, $a = |\nabla \varphi(x_0)|$ is the strength of the injection velocity, which is proportional to the viscosity ν :

$$a = \gamma \nu \quad (\gamma \text{ is the coefficient}).$$

The frictional force f_τ is proportional to the smoothness of the material surface, denoted by κ :

$$f_\tau = \theta \kappa,$$

where θ is the proportional coefficient. Therefore, the critical curvature k_c in (9.5.34) is

$$k_c = \frac{\alpha \nu}{\kappa t_0}, \quad (9.5.35)$$

where $\alpha = \gamma/\theta$ is the proportional coefficient.

Formula (9.5.35) shows that the vortices are easier to form if the viscosity of fluid is relatively small, and the surface of the material is rougher (i.e., κ is relatively large). This conforms to the physical reality.

(3) Critical Velocity of Surface Turbulence

For a boundary flow, when the injection velocity u_0 is smaller than a critical threshold u_c , it is a parallel shear flow, and when $u_0 > u_c$, surface turbulence occurs. The threshold u_c is defined as

$$u_c = \text{the critical velocity of boundary-layer separation.} \quad (9.5.36)$$

Based on this definition, to obtain the critical threshold u_c of surface turbulence, we only need to consider the critical injection velocity u_0 at which the boundary-layer separation occurs. We use the separation equation (9.5.16) for the free boundary condition to study this problem.

Let $t_0 > 0$ be the relaxation time when the first vortex appears as the injection velocity u_0 arrives at u_c , so that $t_0 > 0$ is small. Hence, (9.5.16) can be approximatively written in the form

$$u_\tau = \varphi_\tau + \left[F_\tau + \nu \left(\frac{\partial^2 \varphi_\tau}{\partial \tau^2} + \frac{\partial^2 \varphi_\tau}{\partial n^2} \right) - \varphi_\tau \frac{\partial \varphi_\tau}{\partial \tau} \Big|_{x=0} \right] t_0, \quad (9.5.37)$$

where F_τ represents the tangent damping resistance, and $\Gamma \subset \partial\Omega$ is

$$\Gamma = \{(x_1, 0) \mid 0 < x_1 < L\}. \quad (9.5.38)$$

Let u_0 be the injection velocity. Take u_0 as the control parameter, and

$$\varphi_\tau = u_0 - \beta_1 x_1 + \beta_2 x_1^2 \quad (\beta_1 > \beta_2 L), \quad (9.5.39)$$

where $\beta_1, \beta_2 > 0$ are two small parameters, depending on the viscosity ν of the fluid and the surface physical property of Γ . By the physical law of the damping force,

$$F_\tau = -\gamma u_0^k \quad (k > 1, \gamma > 0). \quad (9.5.40)$$

Inserting (9.5.39) and (9.5.40), we deduce that

$$u_\tau = u_0 - \beta_1 x_1 + \beta_2 x_1^2 - (\gamma u_0^k - 2\beta_2 \nu - \beta_1 u_0) t_0.$$

Because $\beta_1 x_1$ and $\beta_2 x_1^2$ are relatively small with respect to u_0 , and ν is very small, u_τ can be expressed as

$$u_\tau = (1 + \beta_1 t_0) u_0 - \gamma u_0^k t_0.$$

Let $u_\tau = 0$. Then we obtain the following form of u_c :

$$u_c = \left(\frac{1}{\gamma t_0} + \frac{\beta_1}{\gamma} \right)^{1/(k-1)} \quad (k > 1), \quad (9.5.41)$$

where γ is the damping coefficient, depending on the surface property of Γ , and β_1 is as in (9.5.39) which is an increasing function of the viscosity ν . Let $\beta_1 = \beta_0 \nu$. Then (9.5.41) is rewritten as

$$u_c = \left[\frac{1}{\gamma} \left(\frac{1}{t_0} + \beta_0 \nu \right) \right]^{1/(k-1)}. \quad (9.5.42)$$

Remark 9.5.6 For the surface turbulent problem, the length L in (9.5.38) cannot be too large because the velocity decay formula (9.5.39) holds true only for L not large. In fact, the physical phenomena show that surface fluid turbulence can occur only for L being relative small.

9.5.4 Boundary-Layer Separation of Ocean Circulations

The atmospheric and oceanic flows exhibit recurrent large-scale patterns. The formation of these patterns are important topological phase transition problem. The main objective of this section is to study the mechanism for the formation of the subpolar gyre and the vortices separated from the western boundary of the Atlantic ocean.

(1) Vortices Separated from the Boundary

Let $\Gamma \subset \partial\Omega$ be a flat portion of the coast, denoted by

$$\Gamma = \{(x_1, 0) \mid -\delta < x_1 < \delta\} \quad (9.5.43)$$

and $x_2 > 0$ represents the sea area. Let $\varphi = (\varphi_1, \varphi_2)$ represent the oceanic flow, expressed as

$$\varphi_1 = \frac{x_2 u_0}{2\delta + x_1}, \quad \varphi_2 = \frac{x_2^2 u_0}{2(2\delta + x_1)^2} \quad (9.5.44)$$

where δ is as in (9.5.43), satisfying

$$\delta \gg 1, \quad u_0 = o(1). \quad (9.5.45)$$

Consider the wind-driven force $f = (f_1, f_2)$ as

$$f_1 = f_1(x_1), \quad f_2 = -\frac{1}{2}x_1^2 + 2\delta x_1 + 5\delta^2. \quad (9.5.46)$$

In view of (9.5.19) and (9.5.20), for (9.5.44) and (9.5.46) we see that

$$\begin{aligned} \varphi_{11} &= \frac{u_0}{2\delta + x_1}, & \varphi_{13} &= 0, \\ f_{12} &= 0, & f_{12} &= -\frac{1}{2}x_1^2 + 2\delta x_1 + 5\delta^2, \end{aligned}$$

and the function

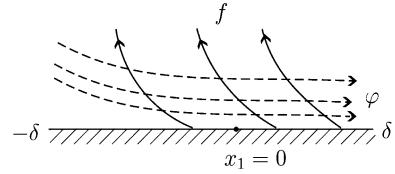
$$g(x_1) = \frac{-\varphi_{11}}{2\nu\varphi_{11}'' + 6\nu\varphi_{13} + f_{12} - f_{21}'} = \frac{(2\delta + x_1)^2 u_0}{(2\delta + x_1)^3 (2\delta - x_1) - 2\nu u_0}.$$

By (9.5.45) we have

$$0 < g(x_1) \ll 1 \quad \forall -\delta < x_1 < \delta.$$

It follows from (9.5.21) that under the wind-driven action of (9.5.46), the oceanic flow represented by (9.5.44) will generate a vortex from the boundary Γ . In addition, by (9.5.45) and $0 < \nu \ll 1$,

Fig. 9.26 Solid lines represent the wind, and the dashed lines stand for the oceanic flow



$$\min_{\Gamma} g(x_1) \simeq \frac{u_0}{(2\delta + x_1)(2\delta - x_1)} \Big|_{x_1=0} = \frac{u_0}{4\delta^2}.$$

Then we derive from (9.5.22) that the separation position x_1° and the time t_0 are given by

$$x_1^\circ = 0, \quad t_0 = \frac{u_0}{4\delta^2}.$$

Figure 9.26 schematically shows the oceanic flow diagram of the wind-driven force (9.5.46)

(2) Wind-Driven North Atlantic Gyres

In Sect. 9.5.1 we introduced the north Atlantic gyres, where double gyres (as shown in Fig. 9.24a) are permanent, and some small vortices (as shown in Fig. 9.24b) are seasonal. In this section we discuss the wind-driven vortices by applying the boundary-layer separation theory.

The dynamical equations governing the wind-driven north Atlantic circulations are the Navier–Stokes equations with free boundary condition, given by

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla) u &= \nu \Delta u - \beta y \vec{k} \times u - \frac{1}{\rho} \nabla p + f, \quad (x, y) \in \Omega, \\ \operatorname{div} u &= 0, \\ u_n|_{\partial\Omega} &= 0, \quad \frac{\partial u_\tau}{\partial n} \Big|_{\partial\Omega} = 0, \\ u(0) &= \varphi, \end{aligned} \tag{9.5.47}$$

where the domain Ω represents the north Atlantic region, approximatively in a rectangular region, as

$$\Omega = (0, L) \times (-L, L),$$

with coordinate (x, y) , and the x -axis points eastward, the y -axis is northward, $y = 0$ represents the mid-latitude. The term representing the Coriolis force in (9.5.47) is

$$-\beta y \vec{k} \times u = -\beta y(-u_2, u_1),$$

where βy is the first-order approximation parameter of the Coriolis force. The wind-driven force f can be expressed as

$$f = F + \mathcal{F}, \quad (9.5.48)$$

where F is the constant force driving the permanent double gyre as shown in Fig. 9.24a, and \mathcal{F} represents the force by seasonal winds. For convenience, \mathcal{F} is written in the divergence free part, by (9.5.14) and (9.5.16), which dictates the boundary-layer separation:

$$\mathcal{F} = \lambda(\mathcal{F}_1, \mathcal{F}_2), \text{ with } \mathcal{F}_2(0, y) > 0 \text{ for } y < L, \quad (9.5.49)$$

where λ is the strength, $\operatorname{div}\mathcal{F} = 0$, and $\mathcal{F} \cdot n|_{\partial\Omega} = 0$.

The initial field φ represents the north Atlantic double gyre, which satisfies the stationary equation of (9.5.47):

$$\begin{aligned} v\Delta\varphi - \beta y \vec{k} \times \varphi - (\varphi \cdot \nabla)\varphi - \frac{1}{\rho} \nabla p + F &= 0, \\ \operatorname{div}\varphi &= 0, \end{aligned} \quad (9.5.50)$$

with the free boundary condition, where F is as in (9.5.48).

To use the separation equation (9.5.16), we take

$$\Gamma = \{(0, y) \in \partial\Omega \mid 0 < y < L\}.$$

Since $u_n|_{\partial\Omega} = 0$, we have

$$(\vec{k} \times u) \cdot \tau|_{\partial\Omega} = 0.$$

Hence, in view of (9.5.47)–(9.5.50), the equation (9.5.16) can be written as

$$u_2(y, t) = \varphi_2(y) + \int_0^t \left[v\Delta u_2 - u_2 \frac{\partial u_2}{\partial y} \Big|_{y=0} + f_2 \right] dt. \quad (9.5.51)$$

Let u be denoted by

$$u = \varphi + v \quad \text{with} \quad v|_{t=0} = 0. \quad (9.5.52)$$

Putting u_2 of (9.5.52) in the right-hand side of (9.5.51), by (9.5.48)–(9.5.50), we obtain that

$$u_2(y, t) = \varphi_2(y) + \int_0^t \left[v\Delta u_2 - \varphi_2 \frac{\partial v_2}{\partial y} - v_2 \frac{\partial \varphi_2}{\partial y} - \lambda \mathcal{F}_2 \right] dt. \quad (9.5.53)$$

for $0 < y < L$.

Fixing $t_0 > 0$ small, then by $v(y, 0) = 0$ and (9.5.49), the equation (9.5.53) can be approximatively expressed in the form

$$u_2(y, t_0) = \varphi_2(y) + \lambda t_0 \mathcal{F}_2(y). \quad (9.5.54)$$

By (9.5.50), in the region of $0 < y < L$, the field φ describes the north Atlantic subpolar gyre, i.e., the northern counter-clockwise gyre in Fig. 9.24a. Hence, we have

$$\begin{aligned}\varphi_2(y) &< 0 \quad \text{for } 0 < y < L, \\ \varphi_2(0) &= \varphi_2(L) = 0.\end{aligned}\tag{9.5.55}$$

In addition, by (9.5.49) and $\mathcal{F} \cdot n|_{\partial\Omega} = 0$, for the rectangular Ω we have

$$\begin{aligned}\mathcal{F}_2(y) &> 0, \quad \text{for } 0 < y < L, \\ \mathcal{F}_2(L) &= 0.\end{aligned}\tag{9.5.56}$$

It is clear that $\varphi_2 \neq \mathcal{F}_2$. Hence it follows from (9.5.54) to (9.5.56) that there are $\lambda_c > 0$ and an isolated point $y_0 \in (0, L)$ such that

$$u_2(y_0, \lambda) \begin{cases} < 0 & \text{for } \lambda < \lambda_c, \\ = 0 & \text{for } \lambda = \lambda_c, \\ > 0 & \text{for } \lambda > \lambda_c. \end{cases}\tag{9.5.57}$$

By Theorem 9.5.1 or Remark 9.5.2, we infer from (9.5.57) that the solution u of (9.5.43) has a boundary-layer separation at $(0, y_0) \in \Gamma$ for $\lambda > \lambda_c$. Namely, we have proved the following physical conclusion.

Physical Conclusion 9.5.7 *If the mid-latitude seasonal wind strength λ exceeds certain threshold λ_c , vortices near the north Atlantic coast will form, as shown in Fig. 9.24b. Moreover, the scale (radius) of the vortices is an increasing function of $\lambda - \lambda_c$.*

(3) North Atlantic Subpolar Gyre

The North Atlantic double gyres are formed by the northern subpolar gyre and the southern subtropical gyre. The subtropical gyre is caused mainly by winds, and the subpolar gyre is generated by the Gulf stream along the western boundary and the regional topography; see Fig. 9.27 in which a schematically topographic diagram of the North Atlantic double gyre is illustrated. Here we shall discuss the

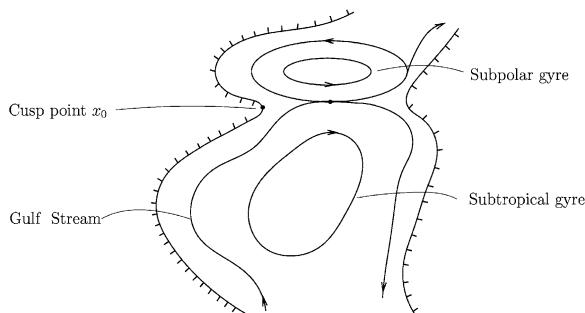
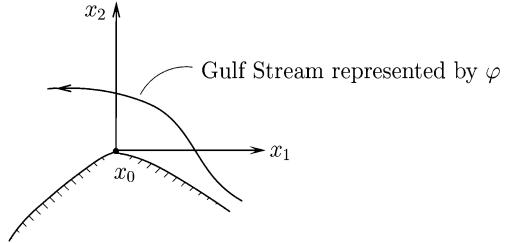


Fig. 9.27 A schematically topography diagram of the North Atlantic double gyre: the northern subpolar and the southern subtropical gyre

Fig. 9.28 Gulf stream

mathematical mechanism for the subpolar gyre formation by applying the separation equation (9.5.6); we shall see that the topographical influence of the North Atlantic, i.e., the curvature $k(x_0)$ of the tip point x_0 in Fig. 9.27, plays a very important role for the formation of the subpolar gyre.

In a neighborhood of the point x_0 , we take the coordinate (x_1, x_2) with x_0 as its origin, the x_1 -axis in the tangent direction pointing southward, and the x_2 -axis in the normal direction pointing eastward; see Fig. 9.28. Take $t_0 > 0$ small, then the separation equation (9.5.6) at x_0 (i.e., $x = 0$) can be approximatively written as

$$\begin{aligned} \frac{\partial u_\tau(k)}{\partial n} &= \frac{\partial \varphi_\tau(x_0)}{\partial n} + \nu t_0 \left[\frac{\partial(\Delta\varphi \cdot \tau)}{\partial n} - \frac{\partial(\Delta\varphi \cdot n)}{\partial \tau} + k \Delta\varphi \cdot \tau \right]_{x_0} \\ &\quad + \left(\frac{\partial f_\tau(x_0)}{\partial n} - \frac{\partial f_n(x_0)}{\partial \tau} + k f_\tau(x_0) \right) t_0, \end{aligned} \quad (9.5.58)$$

where k is the curvature of $\partial\Omega$ at x_0 ($x = 0$). Because φ represents the Gulf stream, and $\varphi|_{\partial\Omega} = 0$, we have

$$\begin{aligned} \frac{\partial \varphi_\tau(0)}{\partial n} &= \frac{\partial \varphi_1(0)}{\partial x_2} < 0 \quad (\text{see Fig. 9.27}), \\ \Delta\varphi \cdot \tau|_{x=0} &= \frac{\partial^2 \varphi_1(0)}{\partial x_2^2}, \end{aligned}$$

Due to the curvature $k \gg 1$, if

$$\frac{\partial^2 \varphi_1(0)}{\partial x_2^2} \neq 0, \quad f_1(0) \neq 0,$$

then we can ignore the following terms in (9.5.58)

$$\nu t_0 \left(\frac{\partial(\Delta\varphi \cdot \tau)}{\partial n} - \frac{\partial(\Delta\varphi \cdot n)}{\partial \tau} \right) \quad \text{and} \quad \frac{\partial f_\tau}{\partial n} - \frac{\partial f_n}{\partial \tau}.$$

Then (9.5.58) becomes

$$\frac{\partial u_\tau(k)}{\partial n} = - \left| \frac{\partial \varphi_1(0)}{\partial x_2} \right| + k t_0 \left[f_1(0) + \nu \frac{\partial^2 \varphi_1(0)}{\partial x_2^2} \right]. \quad (9.5.59)$$

Under the following condition

$$f_1(x_0) + \nu \frac{\partial^2 \varphi_1(0)}{\partial x_2^2} > 0, \quad (9.5.60)$$

by (9.5.59) we deduce that there is a $k_c > 0$ such that

$$\frac{\partial u_\tau(k)}{\partial n} \begin{cases} < 0 & \text{for } k < k_c, \\ = 0 & \text{for } k = k_c, \\ > 0 & \text{for } k > k_c. \end{cases} \quad (9.5.61)$$

By Theorem 9.5.1, it follows from (9.5.61) that if $k > k_c$, then there exists a counterclockwise gyre in the north of the subtropic gyre. Hence, we have the following results.

Physical Conclusion 9.5.8 *The condition for the subpolar gyre to appear is that the curvature k of $\partial\Omega$ at x_0 in Fig. 9.27 is sufficiently large, and the Gulf stream velocity field $\varphi = (\varphi_1, \varphi_2)$ and external forcing $f = (f_1, f_2)$ at x_0 satisfy the inequality (9.5.60).*

9.6 Interior Separations and Cyclone Formation Theory

Interior separation of fluid flows is a common phenomenon in fluid dynamics, especially in geophysical fluid dynamics, such as the formation of hurricanes, typhoons and tornados, and gyres of oceanic flows. In general, the interior separation refers to sudden appearance of a vortex from the interior of a fluid flow.

The aim of this section is to establish a systematic theory for interior separations of fluid flows, and to apply the theory to the formation of tornados and hurricanes. The study is based on Liu et al. (2017b).

9.6.1 Horizontal Heat-Driven Fluid Dynamical Equations

Hurricane, typhoon, and tornado are typical interior separation phenomena in geophysical fluid dynamics, which are caused by external wind-driven forces and by the nonhomogenous temperature distributions. Therefore, the crucial factors for the appearance of interior separations in the atmospheric and oceanic flows are the three ingredients: (1) the initial velocity field, (2) the external force, and (3) the temperature. Hence the dynamical fluid model for interior separations has to incorporate properly the heat effect.

The Boussinesq equations are the classical fluid dynamical model coupling the heat conduction equation. However, this model cannot be used as the dynamical equations to describe interior separation of fluid flows. To show this point, we recall the Boussinesq equations, given by

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla)u &= \nu \Delta u - \frac{1}{\rho} \nabla p - (1 - \alpha T)g \vec{k} + f, \\ \frac{\partial T}{\partial t} + (u \cdot \nabla)T &= \kappa \Delta T + Q, \\ \operatorname{div} u &= 0. \end{aligned} \quad (9.6.1)$$

We note that the driving force acting on the fluid by heat is the term in the first equation of (9.6.1) as follows

$$f_T = -\alpha T g \vec{k}, \quad (9.6.2)$$

where g is the acceleration due to gravity, α is the thermal expansion coefficient, T represents the temperature, and $\vec{k} = (0, 0, 1)$ is the unit vector orthogonal to the horizontal plane. By (9.6.2), it is easy to see that equations (9.6.1) mainly govern the thermal convective fluid motion in the vertical direction, rather than the horizontal motion such as the interior separation fluid flows.

In Yang and Liu (2016), a fluid dynamical model coupling heat was established, which can describe the plane fluid flows as the interior separation behaviors. We introduce this model in the following. Consider the Navier–Stokes equations

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla)u &= \nu \Delta u - \frac{1}{\rho} \nabla p + f, \quad x \in \Omega, \\ \operatorname{div}(\rho u) &= 0, \end{aligned} \quad (9.6.3)$$

and the coupling heat equation

$$\frac{\partial T}{\partial t} + (u \cdot \nabla)T = \kappa \Delta T, \quad x \in \Omega, \quad (9.6.4)$$

where $\Omega \subset \mathbb{R}^2$ is an open set, and $u = (u_1, u_2)$ is the velocity governing a horizontal fluid motion.

To make the temperature T into the Navier–Stokes equation (9.6.3), we introduce the state equations in thermodynamics. For gaseous systems, we have

$$pV = nRT, \quad (9.6.5)$$

where n is the molar number, and R is the gaseous constant. In addition, the gas density is given by

$$\rho = nm/V, \quad m \text{ is the molecular mass.} \quad (9.6.6)$$

Thus, (9.6.5) can be written as

$$p = \beta_0 \rho T, \quad \beta_0 = R/m \text{ is the special gas constant.} \quad (9.6.7)$$

For liquid systems, the equation of state is expressed as

$$V = \alpha_1 T - \alpha_2 p + V_0, \quad (9.6.8)$$

where $\alpha_1, \alpha_2 > 0$ are constants. By (9.6.6), (9.6.8) is rewritten as

$$p = \beta_1 T - \frac{\beta_2}{\rho} + \beta_3, \quad (9.6.9)$$

where $\beta_1, \beta_2, \beta_3$ are parameters.

Equations (9.6.3), (9.6.4), (9.6.7), and (9.6.9) are together to constitute the horizontal heat-driven fluid dynamical equations:

$$\frac{\partial u}{\partial t} + (u \cdot \nabla)u = \nu \Delta u - \frac{1}{\rho} \nabla p + f, \quad (9.6.10)$$

$$\frac{\partial T}{\partial t} + (u \cdot \nabla)T = \kappa \Delta T, \quad (9.6.11)$$

$$\operatorname{div}(\rho u) = 0, \quad (9.6.12)$$

$$p = \begin{cases} \beta_0 \rho T & \text{for gaseous systems,} \\ \beta_1 T - \beta_2 \rho^{-1} + \beta_3 & \text{for liquid systems.} \end{cases} \quad (9.6.13)$$

Remark 9.6.1 As $\rho \neq 0$, the orbit of u with $\operatorname{div} u = 0$ is the same as those of v with $\operatorname{div}(\rho v) = 0$. Since the proof of Theorem 9.1.10 relies only on the topological structure of orbits of divergence-free vector fields, Theorem 9.1.10 is also valid for the vector fields u satisfying (9.6.12).

9.6.2 Interior Separation Equations

Fluid separation equations are crucial mathematical tools to study vortex generation problem. For boundary-layer separations, separation equations for both the free and the rigid boundary conditions were established in Ma and Wang (2005d, 2017h); they play a very important role in both the boundary-layer separation theory and its applications. For interior separation, such separations are clearly needed as well.

To this end, we introduce two physically sound conditions:

1. The gradient of ρ is small:

$$\nabla \rho \simeq 0. \quad (9.6.14)$$

By (9.6.12), it implies that

$$\operatorname{div} u = -\frac{1}{\rho} \nabla \rho \cdot u \simeq 0. \quad (9.6.15)$$

Physically, for large scale motion, the relaxation time for the fluid mobility is much smaller than the relaxation time of compression. Hence for the large scale fluid, the change of density ρ is small, leading to the validity of assumption (9.6.14).

2. The relaxation time t_0 of an interior separation is very small:

$$0 < t_0 \ll 1, \quad (9.6.16)$$

where the relaxation time refers to the time elapsed from the initial time when the separation condition is formed to the time when the separation takes place. It is also called the separation relaxation time.

Based on the model (9.6.10)–(9.6.13) and the two physical conditions (9.6.14) and (9.6.16), we now derive the separation equation.

Let the initial values of u and T are given by

$$u(x, 0) = \varphi(x), \quad T(x, 0) = T_0(x). \quad (9.6.17)$$

Inserting (9.6.13) into (9.6.10), integrating both sides of the resulting equation with respect to t , and ignoring the terms of $\nabla\rho$ (by (9.6.14)), we deduce that

$$u(x, t) = \varphi(x) + \int_0^t [\nu\Delta u - (u \cdot \nabla)u - \alpha\nabla T + f]dt, \quad (9.6.18)$$

for $0 < t < t_0$, where φ is as in (9.6.17), α is given by

$$\alpha = \begin{cases} \beta_0 & \text{for a gaseous system,} \\ \beta_1/\rho & \text{for a liquid system,} \end{cases} \quad (9.6.19)$$

$\beta_0 = R/m$ is as in (9.6.7), and β_1 is as in (9.6.9). Take the first-order Taylor expansion on time t for the solution $(u(x, t), T(x, t))$ of (9.6.10)–(9.6.13):

$$(u(x, t), T(x, t)) = (\varphi(x) + t\tilde{u}(x, t), T_0(x) + t\tilde{T}(x, t)),$$

and insert it into the right-hand side of (9.6.18). Then we obtain that

$$\begin{aligned} u(x, t) &= \varphi(x) + [\nu\Delta\varphi - (\varphi \cdot \nabla)\varphi - \alpha\nabla T_0 + f]t \\ &\quad + \int_0^t [\nu\Delta\tilde{u} - \alpha\nabla\tilde{T} - (\varphi \cdot \nabla)\tilde{u} - (\tilde{u} \cdot \nabla)\varphi]tdt \\ &\quad + \int_0^t (\tilde{u}\nabla\tilde{u}) t^2 dt. \end{aligned} \quad (9.6.20)$$

By (9.6.15), we have

$$\operatorname{div}\varphi \simeq 0, \quad \operatorname{div}[\nu\Delta\varphi - (\varphi \cdot \nabla)\varphi - \alpha\nabla T_0 + f] \simeq 0.$$

Hence we deduce that

$$\frac{\partial\varphi_1}{\partial x_2} \frac{\partial\varphi_2}{\partial x_1} + \left(\frac{\partial\varphi_1}{\partial x_1}\right)^2 + \frac{\alpha}{2}\Delta T_0 - \frac{1}{2}\operatorname{div}f \simeq \nu\Delta\operatorname{div}\varphi \simeq 0, \quad (9.6.21)$$

where $\varphi = (\varphi_1, \varphi_2)$. It is clear that

$$\int_0^t [\nu\Delta\tilde{u} - \alpha\nabla\tilde{T} - (\varphi \cdot \nabla)\tilde{u} - (\tilde{u} \cdot \nabla)\varphi]tdt = o(t).$$

Hence, by (9.6.16), the first-order approximation of (9.6.20) on t ($0 < t < t_0$) and (9.6.21) leads to the following interior separation equations:

$$u(x, t) = \varphi + [\nu\Delta\varphi - (\varphi \cdot \nabla)\varphi - \alpha\nabla T_0 + f]t, \quad (9.6.22)$$

$$\frac{\partial\varphi_1}{\partial x_2} \frac{\partial\varphi_2}{\partial x_1} + \left(\frac{\partial\varphi_1}{\partial x_1} \right)^2 + \frac{\alpha}{2}\Delta T_0 - \frac{1}{2}\operatorname{div}f \simeq 0, \quad (9.6.23)$$

where $\varphi = (\varphi_1, \varphi_2)$ and T_0 are the initial functions as in (9.6.17).

Remark 9.6.2 The interior separation equations (9.6.22) and (9.6.23) include all physical information about the interior separations of the solution u for system (9.6.10)–(9.6.13) given by the initial state (φ, T_0) and the external force f .

9.6.3 U-Flow Theory

(1) Interior Separation of Fluid Dynamical Equations

Theorem 9.1.10 provides a sufficient condition for general 2D incompressible vector fields to undergo an interior separation. Now, we apply the separation equation (9.6.22) to derive the interior separation theorem for solutions of the horizontal heat-driven fluid dynamical equations (9.6.10)–(9.6.13). Let

$$\nu(x) = \nu\Delta\varphi - (\varphi \cdot \nabla)\varphi - \alpha\nabla T_0 + f. \quad (9.6.24)$$

Then $u(x, t)$ in (9.6.22) is written as

$$u(x, t) = \varphi(x) + tv(x). \quad (9.6.25)$$

For such $u(x, t)$, condition (9.1.9) becomes the following:

1. there is a point $x_0 \in V$, with $V \subset \Omega$ being a neighborhood of x_0 , such that

$$\begin{aligned} \varphi(x) + tv(x) &\neq 0, & \forall 0 \leq t < t_0, \\ \varphi(x_0) + t_0v(x_0) &= 0, \end{aligned} \quad (9.6.26)$$

2. the Jacobian of $u^0 = \varphi + t_0v$ at x_0 is nonzero:

$$Du^0(x_0) = D(\varphi + t_0v)|_{x_0} \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad (9.6.27)$$

3. for the unit vector e_2 as in (9.1.8), we have

$$v(x_0) \cdot e_2 \neq 0. \quad (9.6.28)$$

Namely, Conditions (9.6.26)–(9.6.28) for $u(x, t)$ in (9.6.25) are equivalent to (9.1.9). Therefore, by Theorem 9.1.10, we obtain immediately the following interior separation theorem for the solution of equations (9.6.10)–(9.6.13).

Theorem 9.6.3 *Let (φ, T_0) be the initial functions as in (9.6.17), f be the external force, and v be defined by (9.6.24). If (φ, T_0, f) satisfies (9.6.26)–(9.6.28), then under the conditions (9.6.14) and (9.6.16), the solution $u(x, t)$ of (9.6.10)–(9.6.13) has an interior separation from (x_0, t_0) .*

(2) U-flow Pattern for Interior Separation

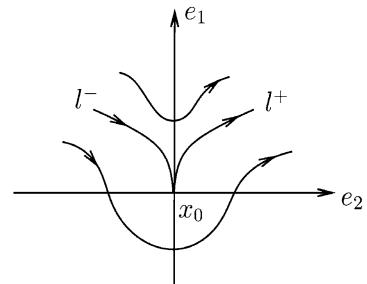
Theorem 9.6.3 shows that under conditions (9.6.26)–(9.6.28) an interior separation occurs. However, we need to know which expressions of (φ, T_0, f) satisfy (9.6.26)–(9.6.28). To solve this problem, we should investigate the topological structure of the two vector fields φ and v in (9.6.25).

For convenience, we rewrite (9.6.24) and (9.6.25) in the form

$$\begin{aligned} u(x, t) &= u^0(x) + (t - t_0)v(x), \\ u^0(x) &= \varphi(x) + t_0v(x), \\ v(x) &= \nu\Delta\varphi - (\varphi \cdot \nabla)\varphi - \alpha\nabla T_0 + f. \end{aligned} \quad (9.6.29)$$

In Ma and Wang (2005d), Ma and Wang proved that if $u(x, t)$ in the form of (9.6.29) satisfies (9.6.26) and (9.6.27), then $\text{ind}(u^0, x_0) = 0$ and $u^0(x)$ is topologically equivalent to the structure near $x_0 \in \Omega$, as shown in Fig. 9.29; see Ma and Wang (2005d, Section 5.6.1).

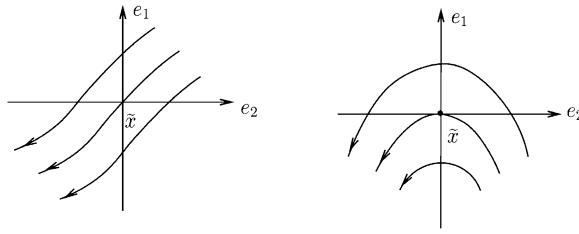
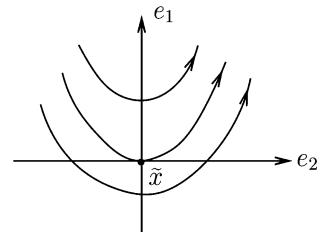
Fig. 9.29 The flow pattern of the vector field u^0 satisfying $u^0(x_0) = 0$, $\text{ind}(u^0, x_0) = 0$, and $Du^0(x_0) \neq 0$. Here the two orbits l^- and l^+ of u^0 connected to x_0 are tangent to the e_1 -axis at x_0 , e_1 , e_2 are as in (9.1.8)



Thus, it is clear that $u^0 = \varphi + t_0v$ is a superposition of φ and t_0v . It implies that if u^0 has the structure as in Fig. 9.29 and $\varphi \neq 0$ in a neighborhood of x_0 , then by the evolution procedure:

$$u(x, 0) = \varphi \xrightarrow{t \rightarrow t_0} u(x, t_0) = \varphi + t_0v,$$

Fig. 9.30 U-flow pattern of a vector field u with $u \cdot e_2 > 0$ at \tilde{x}



(a)

(b)

Fig. 9.31 (a) A flat flow pattern of u with $u \cdot e_2 < 0$ at \tilde{x} , and (b) U-flow pattern of u with $u \cdot e_2 < 0$ at \tilde{x}

we can see that φ and v have the reversed orientations, and their flow structures are either the U-flow pattern as shown in Fig. 9.30 or one of the two patterns as shown in Fig. 9.31a, b. In particular, if φ (or v) is in the pattern of Fig. 9.30 then v (or φ) must be in one of the patterns (a) and (b) in Fig. 9.31.

(3) Physical Conclusions of U-Flow Theory

Mathematically, the three situations that lead to the formation of interior separations are as follows:

- (i) φ is a U-flow and v is a flat flow;
- (ii) φ is a flat flow and v is a U-flow; and
- (iii) φ and v are U-flows.

However, in reality, case (i) appears more often. Hence, here we only consider the case:

$$\begin{aligned} \varphi &\text{ is a U-flow,} \\ v &\text{ is a flat flow, and} \\ \varphi &\text{ and } v \text{ have reversed orientations,} \end{aligned} \tag{9.6.30}$$

where φ and v are as in (9.6.29).

Based on (9.6.30) we now establish the U-flow theory of the interior separation for the solutions of equations (9.6.10)–(9.6.13). The main components of this theory are as follows:

1. The physical data (φ, T_0, f) are approximated by their first-order polynomials. In fact, we take the Taylor expansions near \tilde{x} :

$$\begin{aligned}\varphi &= \varphi(\tilde{x}) + D\varphi(\tilde{x})(x - \tilde{x}) + \text{h.o.t.,} \\ T_0 &= T(\tilde{x}) + \nabla T(\tilde{x})(x - \tilde{x}) + \text{h.o.t.,} \\ f &= f(\tilde{x}) + Df(\tilde{x})(x - \tilde{x}) + \text{h.o.t.,}\end{aligned}\quad (9.6.31)$$

where h.o.t. represents the higher-order terms. Because φ, T_0, f are *nearly independent*, by the second separation equation (9.6.23), we see that the coefficients of the n -th order terms for $n \geq 2$ are almost zero, and

$$\frac{\partial \varphi_1}{\partial x_2} \frac{\partial \varphi_2}{\partial x_1} \Big|_{\tilde{x}}, \quad \frac{\partial \varphi_1}{\partial x_1} \Big|_{\tilde{x}}, \quad \frac{\partial \varphi_2}{\partial x_2} \Big|_{\tilde{x}}, \quad \frac{\partial f_1}{\partial x_1} \Big|_{\tilde{x}}, \quad \frac{\partial f_2}{\partial x_2} \Big|_{\tilde{x}} \simeq 0. \quad (9.6.32)$$

Then we derive the first physical conclusion for the U-flow theory.

Physical Conclusion 9.6.4 *The physical data φ, T_0 and f are, in essence, the first-order polynomials as*

$$\begin{aligned}\varphi &= \varphi(\tilde{x}) + \begin{pmatrix} 0 & \frac{\partial \varphi_1(\tilde{x})}{\partial x_2} \\ \frac{\partial \varphi_2(\tilde{x})}{\partial x_1} & 0 \end{pmatrix} \begin{pmatrix} x_1 - \tilde{x}_1 \\ x_2 - \tilde{x}_2 \end{pmatrix} + \text{h.o.s.t.,} \\ T_0 &= T(\tilde{x}) + \frac{\partial T(\tilde{x})}{\partial x_1}(x_1 - \tilde{x}_1) + \frac{\partial T(\tilde{x})}{\partial x_2}(x_2 - \tilde{x}_2) + \text{h.o.s.t.,} \\ f &= f(\tilde{x}) + \begin{pmatrix} 0 & \frac{\partial f_1(\tilde{x})}{\partial x_2} \\ \frac{\partial f_2(\tilde{x})}{\partial x_1} & 0 \end{pmatrix} \begin{pmatrix} x_1 - \tilde{x}_1 \\ x_2 - \tilde{x}_2 \end{pmatrix} + \text{h.o.s.t.,}\end{aligned}\quad (9.6.33)$$

and by (9.6.32), $\frac{\partial \varphi_1}{\partial x_2} \frac{\partial \varphi_2}{\partial x_1} \Big|_{\tilde{x}} \simeq 0$, where $\tilde{x} = (\tilde{x}_1, \tilde{x}_2)$ is a point near the separation point x_0 , and the abbreviation h.o.s.t. refers to higher-order terms with small coefficients.

2. For the U-flow in Fig. 9.30, we take the coordinate (x_1, x_2) with \tilde{x} as its origin, the x_1 -axis in e_2 direction, and the x_2 -axis in e_1 direction. By (9.6.24), let φ be the U-flow as in Fig. 9.30. By (9.6.26) and (9.6.33), $\varphi = (\varphi_1, \varphi_2)$ should be in the form:

$$\begin{aligned}\varphi_1 &= a_0 + a_1 x_2 + a_2 x_2^2 + \text{h.o.s.t.,} \\ \varphi_2 &= b_0 x_1 + \text{h.o.s.t.,}\end{aligned}\quad (9.6.34)$$

where

$$a_0, a_2, b_0 > 0, \quad a_1 \simeq 0, \quad a_0 \text{ is relatively small.} \quad (9.6.35)$$

The second-order term $a_2 x_2^2$ in φ_1 is specifically retained to ensure the zero of the superposition is an isolated zero, and to ensure the formation of interior separation. We need the condition that

$$b_0 \text{ is of the same order as or larger than } a_0, \quad (9.6.36)$$

to ensure the flow φ in (9.6.34) is a U-flow.

Also by (9.6.33)–(9.6.35), T_0 and $f = (f_1, f_2)$ are as

$$\begin{aligned} T_0 &= \tau_0 + \tau_1 x_1 + \tau_2 x_2 + \text{h.o.s.t.,} \\ f_1 &= -f_1^0 + f_1^2 x_2 + \text{h.o.s.t.,} \\ f_2 &= -f_2^0 + f_2^1 x_1 + \text{h.o.s.t.,} \end{aligned} \quad (9.6.37)$$

and the vector field $v = (v_1, v_2)$ in (9.6.29) can be written as

$$\begin{aligned} v_1 &= -\alpha\tau_1 - f_1^0 + f_1^2 x_2 + \text{h.o.s.t.,} \\ v_2 &= -a_0 b_0 - \alpha\tau_2 - f_2^0 + f_2^1 x_1 + \text{h.o.s.t..} \end{aligned} \quad (9.6.38)$$

Because v is a flat flow as in Fig. 9.31a, and v and φ have reversed orientations, (9.6.38) is rewritten as

$$v = -(\alpha\tau_1 + f_1^0, \alpha\tau_2 + a_0 b_0 + f_2^0) + \text{h.o.s.t.,} \quad (9.6.39)$$

where

$$\alpha\tau_1 + f_1^0 > 0. \quad (9.6.40)$$

Based on (9.6.34)–(9.6.40), we get the second physical conclusion for the U-flow theory.

Physical Conclusion 9.6.5 *The first separation equation (9.6.22) for the U-flow φ and flat flow v in (9.6.30) is expressed as*

$$\begin{aligned} u_1 &= (a_0 + a_1 x_2 + a_2 x_2^2) - t(\alpha\tau_1 + f_1^0) + \text{h.o.s.t.,} \\ u_2 &= b_0 x_1 - t(\alpha\tau_2 + a_0 b_0 + f_2^0) + \text{h.o.s.t.,} \end{aligned} \quad (9.6.41)$$

and the parameters in (9.6.41) satisfy (9.6.35), (9.6.40), and

$$0 < 4a_0 a_2 - a_1^2 \ll 4a_2(\alpha\tau_1 + f_1^0). \quad (9.6.42)$$

In particular, the separation equation (9.6.41) satisfying (9.6.35), (9.6.40), and (9.6.42) obeys conditions (9.6.26)–(9.6.28); consequently an interior separation takes place from (x_0, t_0) with

$$\begin{aligned} x_0 &= \left(\frac{t_0(\alpha\tau_2 + a_0 b_0 + f_2^0)}{b_0}, -\frac{a_1}{2a_2} \right), \\ t_0 &= \frac{4a_0 a_2 - a_1^2}{4a_2(\alpha\tau_1 + f_1^0)}, \end{aligned} \quad (9.6.43)$$

where α is the thermal parameter as in (9.6.19).

Remark 9.6.6 The expression (9.6.41) is the unique form for the U-flow (9.6.30) to generate a vortex from an interior point $x_0 \in \Omega$. Namely, the initial physical data φ

and v in (9.6.30) must take the form as in (9.6.41) for the formation of an interior separation at x_0 . In particular, the parameters a_0 and b_0 represent the strength of the U-flow φ , (τ_1, τ_2) is the initial temperature gradient ∇T_0 at \tilde{x} , and (f_1^0, f_2^0) is the main part of the external force f . The six parameters $\{a_0, b_0, \tau_1, \tau_2, f_1^0, f_2^0\}$ determine the energy level and scale of the generated vortex.

In addition, for the cases (ii) and (iii) listed in the beginning of Sect. 3.2.2, the associated unique forms of the U-flow for the formation of interior separations are just slight variations of (9.6.41).

9.6.4 Tornado and Hurricane Formations

(1) Atmospheric Cyclone Dynamical Model

In nature, tornados and hurricanes (or typhoons) are the most important phenomena in the atmospheric sciences, which are topological phase transitions described by interior separations of the atmospheric flows.

In this section, we shall address the tornado and hurricane formation mechanism by using the U-flow theory established in the last section.

The main factors for the formation of tornados and hurricanes are

- the initial velocity field φ ,
- the initial temperature field T_0 , and
- the Coriolis force f , measuring the earth's rotational effect.

To derive the atmospheric cyclone dynamic model, we adopt the conventional β -plane assumption.

Let the earth's hemisphere be approximatively given as a rectangle:

$$\Omega = (0, 2\pi r_0) \times (0, \pi r_0/2), \quad (9.6.44)$$

where r_0 is the earth's radius. The Coriolis force is taken as

$$f = -\beta x_2 \vec{k} \times u = (\beta x_2 u_2, -\beta x_2 u_1), \quad (9.6.45)$$

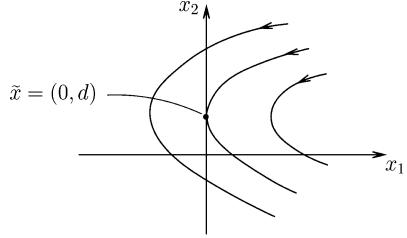
where $\beta = 2\omega/r_0$, and ω is the angular velocity of the earth's rotation. For the domain Ω , the coordinate system is (x_1, x_2) with x_1 -axis being the longitude coordinate and x_2 -axis for the latitude. Equations (9.6.44) and (9.6.45) are called the β -plane assumption.

With the above Coriolis force f , the system (9.6.1) serves as the basic atmospheric cyclone dynamical model for the initial stages of the formation of hurricanes.

With the Coriolis force (9.6.45), the interior separation equations (9.6.22) and (9.6.23) become

$$\begin{cases} u(x, t) = \varphi + tv(x), \\ v(x) = \nu \Delta \varphi - (\varphi \cdot \nabla) \varphi - \alpha \nabla T_0 - \beta x_2 \vec{k} \times \varphi, \end{cases} \quad (9.6.46)$$

Fig. 9.32 A U-flow in the northern hemisphere, $x_2 = 0$ represents the equator, and $\tilde{x} = (0, d)$ with $d > 0$ is the point as in (9.6.33)



$$\frac{\partial \varphi_1}{\partial x_2} \frac{\partial \varphi_2}{\partial x_1} + \left(\frac{\partial \varphi_1}{\partial x_1} \right)^2 + \frac{\alpha}{2} \Delta T_0 + \frac{\beta x_2}{2} \left(\frac{\partial \varphi_1}{\partial x_2} - \frac{\partial \varphi_2}{\partial x_1} \right) + \frac{\beta}{2} \varphi_1 \simeq 0. \quad (9.6.47)$$

(2) Atmospheric Cyclone Generations

We now discuss tornado and hurricane in the northern hemisphere based on the U-flow theory. We proceed in a few steps as follows.

1. *Initial Physical Data φ , T_0 and the Coriolis Force f* In the northern hemisphere, the atmospheric flows near the earth surface often appear in the form of the U-flows as shown in Fig. 9.32, in which $x_2 = 0$ represents the equator.

Corresponding to (9.6.34) and (9.6.37), as the initial velocity field $\varphi = (\varphi_1, \varphi_2)$, the U-flow in Fig. 9.32 should be rewritten in the following form

- the initial flow:

$$\begin{aligned} \varphi_1 &= -b_0(x_2 - d) + \text{h.o.s.t.,} \\ \varphi_2 &= -(a_0 + a_1 x_1 + a_2 x_1^2) + \text{h.o.s.t.,} \end{aligned} \quad (9.6.48)$$

- the initial temperature T_0 :

$$T_0 = \tau_1(\theta - x_2) + \text{h.o.s.t.} \quad (\theta > 0 \text{ is a constant}), \quad (9.6.49)$$

- the Coriolis force $f = (f_1, f_2)$:

$$\begin{aligned} f_1 &= \beta x_2 \varphi_2 = -a_0 \beta x_2 + \text{h.o.s.t.,} \\ f_2 &= -\beta x_2 \varphi_1 = b_0 \beta x_2 (x_2 - d) + \text{h.o.s.t.,} \end{aligned} \quad (9.6.50)$$

where β is as in (9.6.45), and a_i ($0 \leq i \leq 2$), b_0 and τ_1 satisfy (9.6.35) and (9.6.40) with $f_1^0 = 0$.

2. *Interior Separation Equation* By (9.6.46), it follows from (9.6.48) to (9.6.50) that the interior separation equation for the atmospheric cyclone is

$$\begin{aligned} u_1 &= -b_0(x_2 - d) - t[a_0 b_0 + a_0 \beta x_2] + \text{h.o.s.t.,} \\ u_2 &= -(a_0 + a_1 x_1 + a_2 x_1^2) + t[\alpha \tau_1 + b_0 \beta x_2 (x_2 - d)] + \text{h.o.s.t.,} \end{aligned} \quad (9.6.51)$$

3. *Conditions for the Formation of Tornados* Tornadoes form in higher latitudes; hence $d > 0$ is relatively large. Let $x = x_1$, $y = x_2 - d$ in (9.6.51). Then the separation point (x_0, y_0, t_0) satisfies the equation

$$\begin{aligned} -b_0 y_0 - t_0 a_0 (b_0 + \beta d + \beta y_0) &= 0, \\ a_0 + a_1 x_0 + a_2 x_0^2 - t_0 (\alpha \tau_1 + b_0 \beta d y_0 + b_0 \beta y_0^2) &= 0. \end{aligned} \quad (9.6.52)$$

The solution of (9.6.52) satisfying (9.6.16), (9.6.35), and (9.6.40) is

$$\begin{aligned} (x_0, y_0) &= \left(-\frac{a_1}{2a_2}, -\frac{a_0(b_0 + \beta d)}{b_0 + \beta a_0 t_0} t_0 \right), \\ t_0 &= \frac{4a_0 a_2 - a_1^2}{4a_2(\alpha \tau_1 + b_0 \beta d y_0 + b_0 \beta y_0^2)} \simeq \frac{4a_0 a_2 - a_1^2}{4\alpha a_2 \tau_1}. \end{aligned} \quad (9.6.53)$$

By $0 < t_0 \ll 1$ and $y_0 \simeq 0$, (9.6.53) implies that

$$0 < \frac{1}{\alpha \tau_1} \left(a_0 - \frac{a_1^2}{4a_2} \right) \ll 1, \quad \frac{a_0(b_0 + \beta d)}{b_0} \sim o(1). \quad (9.6.54)$$

These are exactly the conditions required for the formation of tornados.

4. *Conditions for the Formation of Hurricanes* A hurricane can form near the equator, i.e., $d \simeq 0$. Let $x = x_1$, $y = x_2$, and $d = 0$ in (9.6.51). Then the location and time (x_0, y_0, t_0) , where a hurricane forms, satisfies

$$\begin{aligned} -b_0 y_0 - t_0 a_0 (b_0 + \beta y_0) &= 0, \\ a_0 + a_1 x_0 + a_2 x_0^2 - \alpha \tau_1 t_0 &= 0, \end{aligned} \quad (9.6.55)$$

whose solution is

$$(x_0, y_0) \simeq \left(-\frac{a_1}{2a_2}, -a_0 t_0 \right), \quad t_0 = \frac{4a_0 a_2 - a_1^2}{4\alpha \tau_1 a_2}. \quad (9.6.56)$$

In summary, from the above discussions, we see that the initial velocity field (9.6.48) of a U-flow, the initial temperature distribution (9.6.49), and the Coriolis force (9.6.50) determine the necessary conditions for the formation of tornadoes and hurricanes.

Appendix A

A.1 Formulas for Center Manifold Functions

The stable and center manifold theorems are classical (Henry, 1981; Ma and Wang, 2005b). The main objective of this appendix is to give a brief account of the center manifold reduction, and to prove a few important formulas in Theorem A.1.1. Part of Theorem A.1.1 is first derived and proved in Ma and Wang (2005b, 2007b), and part of it is introduced for the first time in this appendix. These formulas play a crucial role in many applications addressed in this book.

Let X_1 and X be two Banach spaces, and $X_1 \subset X$ a dense and compact inclusion. Consider the following nonlinear evolution equation

$$\begin{aligned} \frac{du}{dt} &= L_\lambda u + G(u, \lambda), & \lambda \in \mathbb{R}^1, \\ u(0) &= u_0, \end{aligned} \tag{A.1.1}$$

where $L_\lambda = -A + B_\lambda : X_1 \rightarrow X$ are parameterized linear completely continuous fields depending continuously on $\lambda \in \mathbb{R}^1$, i.e.,

$$\begin{aligned} A : X_1 &\rightarrow X \quad \text{a linear homeomorphism,} \\ B_\lambda : X_1 &\rightarrow X \quad \text{parameterized linear compact operators.} \end{aligned} \tag{A.1.2}$$

Consider the case where $L_\lambda : X_1 \rightarrow X$ is a sectorial operator. Then one can define fractional power operators L_λ^α for $\alpha \in \mathbb{R}^1$ with domains $X_\alpha = D(L_\lambda^\alpha)$. Assume that the nonlinear mappings $G(\cdot, \lambda) : X_\theta \rightarrow X$ for some $\theta < 1$ are parameterized C^r ($r \geq 1$) bounded operators continuously depending on the parameter $\lambda \in \mathbb{R}^1$, such that

$$G(u, \lambda) = o(\|u\|_{X_\theta}) \quad \forall \lambda \in \mathbb{R}^1 \quad (\theta < 1). \tag{A.1.3}$$

In addition, let $G(u, \lambda)$ have the Taylor expansion at $u = 0$ as follows

$$G(u, \lambda) = \sum_{m=k}^r G_m(u, \lambda) + o(\|u\|_{X_1}^r) \quad \text{for some } 2 \leq k \leq r, \quad (\text{A.1.4})$$

where $u \in X_1$, $G_m : X_1 \times \cdots \times X_1 \rightarrow X$ is an m -multiple linear operator, and $G_m(u, \lambda) = G_m(u, \dots, u, \lambda)$.

Let $\{\beta_i(\lambda) \in \mathbb{C} \mid i = 1, 2, \dots\}$ be all eigenvalues of L_λ (counting multiplicities), and satisfy

$$\operatorname{Re} \beta_i(\lambda) \begin{cases} < 0 & \text{if } \lambda < \lambda_0, \\ = 0 & \text{if } \lambda = \lambda_0, \\ > 0 & \text{if } \lambda > \lambda_0, \end{cases} \quad \forall 1 \leq i \leq m, \quad (\text{A.1.5})$$

$$\operatorname{Re} \beta_j(\lambda_0) \neq 0 \quad \forall j \geq m+1, \quad (\text{A.1.6})$$

for some $\lambda_0 \in \mathbb{R}^1$. Let $\{e_1(\lambda), \dots, e_m(\lambda)\}$ and $\{e_1^*(\lambda), \dots, e_m^*(\lambda)\}$ be the eigenvectors of L_λ and the conjugate operator L_λ^* corresponding to the eigenvalues $\beta_i(\lambda)$ ($1 \leq i \leq m$) respectively. By the spectral theorem, Theorem 3.4 in Ma and Wang (2005b), the spaces X_1 and X can be decomposed into the direct sum

$$X_1 = E_1^\lambda \bigoplus E_2^\lambda, \quad X = E_1^\lambda \bigoplus \overline{E}_2^\lambda,$$

where

$$E_1^\lambda = \operatorname{span} \{e_1(\lambda), \dots, e_m(\lambda)\},$$

$$E_2^\lambda = \{u \in X_1 \mid (u, e_i^*(\lambda)) = 0 \quad \forall 1 \leq i \leq m\},$$

$$\overline{E}_2^\lambda = \text{closure of } E_2^\lambda \text{ in } X.$$

The linear operator L_λ can be decomposed into

$$L_\lambda = J_\lambda \bigoplus \mathcal{L}_\lambda, \quad (\text{A.1.7})$$

where $J_\lambda : E_1^\lambda \rightarrow E_1^\lambda$ is the Jordan matrix of L_λ at $\beta_i(\lambda)$ ($1 \leq i \leq m$), and $\mathcal{L}_\lambda = L_\lambda|_{E_2^\lambda} : E_2^\lambda \rightarrow \overline{E}_2^\lambda$ possesses eigenvalues $\beta_j(\lambda)$ ($j \geq m+1$). In this case the Eq. (A.1.1) can be written as

$$\frac{dx}{dt} = J_\lambda x + P_1 G(x+y, \lambda), \quad (\text{A.1.8})$$

$$\frac{dy}{dt} = \mathcal{L}_\lambda y + P_2 G(x+y, \lambda), \quad (\text{A.1.9})$$

where $P_i : X \rightarrow E_i^\lambda$ ($i = 1, 2$) are the canonical projections.

The existence of center manifold functions are classical; see among others (Henry, 1981; Ma and Wang, 2005b). As the center manifold function is implicitly defined,

one ingredient in the theory and its applications of the phase transition dynamics in this book is that an approximation of the center manifold function to certain order will lead to a complete understanding of the transitions of a dynamical system. We now present some approximation formulas of the center manifold functions.

First, we show how the center manifold functions are constructed. Let $\rho_\varepsilon : E_1^\lambda \rightarrow [0, 1]$ be a C^∞ cut-off function defined by

$$\rho_\varepsilon(x) = \begin{cases} 1 & \text{if } \|x\| < \varepsilon, \\ 0 & \text{if } \|x\| > 2\varepsilon, \end{cases}$$

for some $\varepsilon > 0$, and let

$$C^{0,1}(E_1^\lambda, E_2^\lambda(\theta)) = \{h : E_1^\lambda \rightarrow E_2^\lambda(\theta) \mid h(0) = 0, h \text{ is Lipschitz}\}.$$

As in Henry (1981), we need to find a function $h \in C^{0,1}(E_1^\lambda, E_2^\lambda(\theta))$ satisfying

$$h(\cdot) = \int_{-\infty}^0 e^{-\mathcal{L}_\lambda \tau} \rho_\varepsilon(x(\tau, \cdot)) G_2(x(\tau, \cdot) + h(x(\tau, \cdot)), \lambda) d\tau, \quad (\text{A.1.10})$$

where $x(t, x_0)$ is a solution of the ordinary differential equation

$$\frac{dx}{dt} = J_\lambda x + \rho_\varepsilon(x) G_1(x + h(x), \lambda), \quad x(0) = x_0, \quad (\text{A.1.11})$$

Then the function $y(t, h(x_0)) = h(x(t, x_0))$ is a solution of the equation

$$\frac{dy}{dt} = \mathcal{L}_\lambda y + \rho_\varepsilon(x(t, x_0)) G_2(x(t, x_0), y), \quad y(0) = h(x_0).$$

Thus, $x(t, x_0) + h(x(t, x_0))$ is a local solution of (A.1.8) and (A.1.9), and the manifold M_λ is locally invariant for (A.1.1).

Hence, the existence of the center manifold of (A.1.1) is referred to as the existence of the fixed point of the mapping $F : C^{0,1}(E_1^\lambda, E_2^\lambda(\theta)) \rightarrow C^{0,1}(E_1^\lambda, E_2^\lambda(\theta))$ defined by

$$F(h) = \int_{-\infty}^0 e^{-\mathcal{L}_\lambda \tau} \rho_\varepsilon(x(\tau)) G_2(x(\tau), h(x(\tau))) d\tau \quad (\text{A.1.12})$$

Using some well-known properties for a sectorial operator (Pazy, 1983; Henry, 1981), one can prove that under the condition (A.1.3) the mapping F defined by (A.1.12) is a contraction operator for $\varepsilon > 0$ small, therefore F has a fixed point in $C^{0,1}(E_1^\lambda, E_2^\lambda(\theta))$. For details of the proof, see Temam (1997) and Henry (1981).

The following theorem gives a first-order approximation formula of the center manifold function of (A.1.1) near $\lambda = \lambda_0$.

Theorem A.1.1 *Assume that the nonlinear operator $G(u, \lambda)$ has the Taylor expansion as (A.1.4) at $u = 0$, and the conditions (A.1.5) and (A.1.6) hold true. Then the center manifold function $\Phi : E_1 \rightarrow E_2$ of (A.1.1) near $\lambda = \lambda_0$ can be expressed as*

$$\Phi(x, \lambda) = \int_{-\infty}^0 e^{-\tau \mathcal{L}_\lambda} \rho_\varepsilon P_2 G_k(e^{\tau J_\lambda} x, \lambda) d\tau + o(\|x\|^k), \quad (\text{A.1.13})$$

where J_λ and \mathcal{L}_λ are the linear operators given by (A.1.7), $G_k(u, \lambda)$ is the lowest order k -multiple linear operator as in (A.1.4), and $x = \sum_{i=1}^m x_i e_i \in E_1$. In particular, we have the following assertions:

(1) If J_λ is diagonal near $\lambda = \lambda_0$, then (A.1.13) can be written as

$$-\mathcal{L}_\lambda \Phi(x, \lambda) = P_2 G_k(x, \lambda) + o(\|x\|^k) + O(|\beta| \|x\|^k), \quad (\text{A.1.14})$$

where $\beta(\lambda) = (\beta_1(\lambda), \dots, \underline{\beta_m(\lambda)})$ is as in (A.1.5).

(2) Let $m = 2$ and $\beta_1(\lambda) = \beta_2(\lambda) = \alpha(\lambda) + i\rho(\lambda)$ with $\rho(\lambda_0) \neq 0$. If $G_k(u, \lambda) = G_2(u, \lambda)$ is bilinear, then $\Phi(x, \lambda)$ can be expressed as

$$\begin{aligned} [(-\mathcal{L}_\lambda)^2 + 4\rho^2(\lambda)](-\mathcal{L}_\lambda)\Phi(x, \lambda) &= [(-\mathcal{L}_\lambda)^2 + 4\rho^2(\lambda)]P_2 G_2(x, \lambda) \\ &\quad - 2\rho^2(\lambda)P_2 G_2(x, \lambda) + 2\rho^2 P_2 G_2(x_1 e_2 - x_2 e_1) \\ &\quad + \rho(-\mathcal{L}_\lambda)P_2[G_2(x_1 e_1 + x_2 e_2, x_2 e_1 - x_1 e_2) \\ &\quad + G_2(x_2 e_1 - x_1 e_2, x_1 e_1 + x_2 e_2)] + o(k), \end{aligned} \quad (\text{A.1.15})$$

where we have used $o(k) = o(\|x\|^k) + O(|\operatorname{Re}\beta(\lambda)| \|x\|^k)$.

(3) Let $\beta_1(\lambda) = \dots = \beta_m(\lambda)$ have algebraic and geometric multiplicities $m \geq 2$ and $r = 1$ near $\lambda = \lambda_0$, i.e., J_λ has the Jordan form:

$$J_\lambda = \begin{pmatrix} \beta(\lambda) & \delta & \cdots & 0 & 0 \\ 0 & \beta(\lambda) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \beta(\lambda) & \delta \\ 0 & 0 & \cdots & 0 & \beta(\lambda) \end{pmatrix} \quad \text{for some } \delta \neq 0. \quad (\text{A.1.16})$$

Let

$$z = \sum_{j=1}^m \xi_j e_j \in E_1 \quad \text{with} \quad \xi_j = \sum_{r=0}^{m-j} \frac{\delta^r t^r x_{j+r}}{r!},$$

where $x = (x_1, \dots, x_m) \in \mathbb{R}^m$, $\delta \neq 0$ is as in J_λ , and $t \geq 0$. Then the k -linear term $G_k(z, \lambda)$ can be expressed as

$$G_k(z, \lambda) = F_1(x) + tF_2(x) + \cdots + t^{m-1}F_m(x),$$

and the center manifold function Φ can be expressed as

$$\Phi = \sum_{j=1}^m \Phi_j + o(k), \quad -\mathcal{L}_\lambda^j \Phi_j = (j-1)! P_2 F_j(x) \quad \forall 1 \leq j \leq m. \quad (\text{A.1.17})$$

Remark A.1.2 The formula (A.1.14) of center manifold functions can be equivalently expressed as

$$-\mathcal{L}_\lambda \Phi(x, \lambda) = \sum_{1 \leq j_1, \dots, j_k \leq m} g_{j_1 \dots j_k} x_{j_1} \cdots x_{j_k} + o(k), \quad (\text{A.1.18})$$

where $g_{j_1 \dots j_k} = P_2 G_k(e_{j_1}, \dots, e_{j_k}, \lambda)$. Formula (A.1.15) is equivalent to the form

$$\Phi = \Phi_1 + \Phi_2 + \Phi_3 + o(3), \quad (\text{A.1.19})$$

where

$$\begin{aligned} -\mathcal{L}_\lambda \Phi_1 &= x_1^2 G_{11} + x_2^2 G_{22} + x_1 x_2 (G_{12} + G_{21}), \\ [(-\mathcal{L}_\lambda)^2 + 4\rho^2](-\mathcal{L}_\lambda) \Phi_2 &= 2\rho^2 [(x_1^2 - x_2^2)(G_{22} - G_{11}) - 2x_1 x_2 (G_{12} + G_{21})], \\ [(-\mathcal{L}_\lambda)^2 + 4\rho^2] \Phi_3 &= \rho(-\mathcal{L}_\lambda) [(x_2^2 - x_1^2)(G_{12} + G_{21}) + 2x_1 x_2 (G_{11} - G_{22})], \end{aligned}$$

and $G_{ij} = P_2 G_2(e_i, e_j, \lambda)$, $1 \leq i, j \leq 2$.

Proof of Theorem A.1.1. STEP 1. By (A.1.10) and (A.1.11), the center manifold function $y = \Phi(x, \lambda)$ of (A.1.8) and (A.1.9) satisfies

$$\Phi(x, \lambda) = \int_{-\infty}^0 e^{-\tau \mathcal{L}_\lambda} \rho_\varepsilon P_2 G(z(\tau, x) + \Phi, \lambda) d\tau \quad (\text{A.1.20})$$

where $z(t, x)$ is a solution of the equations

$$\frac{dz}{dt} = J_\lambda z + \rho_\varepsilon(z) P_1 G(z + \Phi(z), \lambda), \quad z(0) = x. \quad (\text{A.1.21})$$

By (A.1.4) and $\Phi(x, \lambda) = o(\|x\|)$, the solution $z(t, x)$ of (A.1.21) is given by

$$z = e^{tJ_\lambda} x + o(\|x\|).$$

Inserting $z(t, x)$ in (A.1.20) we get (A.1.13).

As J_λ is diagonal, $z = \sum_{i=1}^m z_i e_i$ is expressed as

$$z_i(t, x) = x_i e^{\beta_i t} + o(\|x\|).$$

Then, from (A.1.20) we obtain

$$\begin{aligned} \Phi(x, \lambda) &= \int_{-\infty}^0 e^{-\tau \mathcal{L}_\lambda} \rho_\varepsilon P_2 G_k(x e^{\beta \tau}, \lambda) d\tau + o(\|x\|^k) \\ &= (-\mathcal{L}_\lambda)^{-1} \int_{-\infty}^0 e^{-\tau \mathcal{L}_\lambda} \rho_\varepsilon P_2 G_k d(-\tau \mathcal{L}_\lambda) + o(\|x\|^k) \\ &= (-\mathcal{L}_\lambda)^{-1} P_2 G_k(x, \lambda) + o(\|x\|^k) + O(|\beta| \|x\|^k). \end{aligned}$$

Hence, Assertion (1) is proved.

STEP 2. When the eigenvalues satisfying (A.1.5) are a pair of complex eigenvalues $\beta_1 = \bar{\beta}_2 = \alpha + i\rho$ and $G_k = G_2$ is bilinear, the solution $z(t, x)$ of (A.1.21) reads

$$\begin{aligned} z &= z_1 e_1 + z_2 e_2 + o(\|x\|), \\ z_1 &= e^{\alpha t} (x_1 \cos \rho t - x_2 \sin \rho t), \\ z_2 &= e^{\alpha t} (x_1 \sin \rho t + x_2 \cos \rho t), \quad (\alpha(\lambda_0) = 0), \end{aligned} \tag{A.1.22}$$

and $G_2(e^{\tau J_\lambda} x, \lambda)$ is given by

$$\begin{aligned} P_2 G_2(e^{\tau J_\lambda} x, \lambda) &= P_2 G_2(z_1 e_1 + z_2 e_2, z_1 e_1 + z_2 e_2, \lambda) \\ &= z_1^2 G_{11} + z_1 z_2 (G_{12} + G_{21}) + z_2^2 G_{22} \end{aligned}$$

where $G_{ij} = P_2 G_2(e_i, e_j, \lambda)$. It follows from (A.1.22) that

$$\begin{aligned} P_2 G_2(e^{t J_{\lambda_0}} x, \lambda_0) &= [\frac{1}{2}(x_1^2 + x_2^2) + \frac{1}{2}(x_1^2 - x_2^2) \cos 2\rho t - x_1 x_2 \sin 2\rho t] G_{11} \\ &\quad + [\frac{1}{2}(x_1^2 + x_2^2) - \frac{1}{2}(x_1^2 - x_2^2) \cos 2\rho t + x_1 x_2 \sin 2\rho t] G_{22} \\ &\quad + [\frac{1}{2}(x_1^2 - x_2^2) \sin 2\rho t + x_1 x_2 \cos 2\rho t] [G_{12} + G_{21}]. \end{aligned}$$

By the formulas

$$\begin{aligned} \int_{-\infty}^0 e^{-\tau \mathcal{L}_\lambda} \cos 2\rho \tau d\tau &= (-\mathcal{L}_\lambda)((-\mathcal{L}_\lambda)^2 + 4\rho^2)^{-1}, \\ \int_{-\infty}^0 e^{-\tau \mathcal{L}_\lambda} \sin 2\rho \tau d\tau &= -2\rho((-\mathcal{L}_\lambda)^2 + 4\rho^2)^{-1}, \end{aligned}$$

we deduce that

$$\begin{aligned} \int_{-\infty}^0 e^{-\tau \mathcal{L}_\lambda} \rho_e P_2 G_2(e^{\tau J_\lambda} x, \lambda) d\tau &= \int_{-\infty}^0 e^{-\tau \mathcal{L}_\lambda} P_2 G_2(e^{\tau J_{\lambda_0}} x, \lambda_0) d\tau + O(|\alpha| \|x\|^3) \\ &= (-\mathcal{L}_\lambda)^{-1} [P_2 G_2(x, \lambda) - 2\rho^2 ((-\mathcal{L}_\lambda)^2 + 4\rho^2)^{-1} P_2 G_2(x, \lambda)] \\ &\quad + 2\rho^2 ((-\mathcal{L}_\lambda)^2 + 4\rho^2)^{-1} (x_1^2 G_{22} + x_2^2 G_{11} - x_1 x_2 (G_{12} + G_{21})) \\ &\quad + \rho (-\mathcal{L}_\lambda) ((-\mathcal{L}_\lambda)^2 + 4\rho^2)^{-1} [(x_2^2 - x_1^2) (G_{12} + G_{21}) \\ &\quad + 2x_1 x_2 (G_{11} - G_{22})] + O(|\alpha| \|x\|^3). \end{aligned}$$

Thus, from (A.1.20) we derive (A.1.15), and Assertion (2) is proved.

STEP 3. When J_λ has the Jordan form as given by (A.1.16), the solution z of (A.1.21) can be written as

$$z = \sum_{j=1}^m z_j e_j + o(\|x\|), \quad z_j = \xi_j e^{\beta(\lambda)t}, \quad \xi_j = \sum_{r=0}^{m-j} \frac{\delta^r t^r x_{j+r}}{r!},$$

where $\beta = \beta_j$. Thus, (A.1.20) reads

$$\begin{aligned}\Phi(x, \lambda) &= \int_{-\infty}^0 e^{-\tau \mathcal{L}_\lambda} P_2 G(\sum_{j=1}^m \xi_j e^{\beta \tau} e_j, \lambda) d\tau + o(k) \\ &= \sum_{j=1}^m \left[\int_{-\infty}^0 \tau^{j-1} e^{-\tau \mathcal{L}_\lambda} d\tau \right] P_2 F_j(x) + o(k).\end{aligned}$$

Direct computation shows that

$$\int_{-\infty}^0 \tau^{j-1} e^{-\tau \mathcal{L}_\lambda} d\tau = -(j-1)! \mathcal{L}_\lambda^{-j},$$

and Assertion (3) follows. The proof is complete. \square

A.2 Dynamics of Gradient-Type Systems

Equilibrium phase transitions in statistical physics are governed by gradient-type flows, where the Gibbs free-energies are the corresponding functionals. We study in this appendix dynamical systems generated by gradient-type operators.

Let H_1 and H be two Hilbert spaces, $L : H_1 \rightarrow H$ be a sectorial operator, and $G : H_\theta \rightarrow H$ be a C^r ($r \geq 1$) mapping for some $0 < \theta < 1$. Let $Y \subset H_\alpha$ ($0 < \alpha < 1$) be a Banach space. Throughout this section, we always use $\|\cdot\|$ and (\cdot, \cdot) to denote the norm and inner product in H . Consider the following dynamical system

$$\frac{du}{dt} = Lu + G(u), \quad u(0) = u_0. \quad (\text{A.2.1})$$

Definition A.2.1 A mapping $L + G : H_1 \rightarrow H$ is called a gradient-type operator if there exist a C^1 function $F : Y \rightarrow \mathbb{R}^1$ and a constant $C > 0$ such that $DF(u) \in H$ for any $u \in H_\beta$ for some $\beta \leq 1$, and

$$-(DF(u), Lu + Gu) \geq C \|DF(u)\|^2 \quad \forall u \in H_1 \quad (\text{A.2.2})$$

$$DF(u_0) = 0 \quad \Leftrightarrow \quad Lu_0 + Gu_0 = 0. \quad (\text{A.2.3})$$

In this case, F is called the energy functional of $L + G$.

Obviously, a gradient operator $L + G = -DF$ is of the gradient-type structure. Conversely, if $L + G : H_1 \rightarrow H$ is a gradient-type operator, then in general $L + G$ is in the form

$$L + G = ADF + B : H_1 \rightarrow H \quad (\text{A.2.4})$$

where $A : H_1 \rightarrow H$ is a negative definite linear operator, and

$$\begin{aligned} (Au, u) &\leq -C\|u\|^2 && \text{for some } C > 0, \\ (Bu, DF(u)) &\leq 0 && \forall u \in H_1, \\ DF(u_0) = 0 &\Rightarrow Bu_0 = 0. \end{aligned} \quad (\text{A.2.5})$$

In this section, we always assume that $L + G : H_1 \rightarrow H$ is a gradient-type operator with an energy functional $F : Y \rightarrow \mathbb{R}^1$. Let $u(t, \varphi) \in H_1$ be a solution of (A.2.1). Then

$$\frac{d}{dt}F(u(t, \varphi)) \leq -C\|DF(u(t, \varphi))\|^2 \quad \forall t \geq 0, \quad (\text{A.2.6})$$

which implies that the function $t \rightarrow F(u(t, \varphi))$ is strictly decreasing except at critical points of DF :

$$F(u(t, \varphi)) < F(u(s, \varphi)) \quad \forall 0 \leq s < t \quad \text{if } DF(\varphi) \neq 0. \quad (\text{A.2.7})$$

The following theorem shows that all solutions of gradient-type equations converge to steady-state solutions, and is a direct consequence of (A.2.6) and (A.2.7).

Theorem A.2.2 *Assume that the energy functional $F : Y \rightarrow \mathbb{R}^1$ is weakly lower semicontinuous, i.e., for any $u_n \rightharpoonup v$ in Y ,*

$$\liminf_{n \rightarrow \infty} F(u_n) \geq F(v),$$

and $F(u) \rightarrow +\infty \Leftrightarrow \|u\|_Y \rightarrow \infty$, then for any $\varphi \in H$, there exists $u_0 \in H$ such that

$$\lim_{t \rightarrow \infty} u(t, \varphi) = u_0 \text{ in } H, \quad Lu_0 + G(u_0) = 0, \quad (\text{A.2.8})$$

where $u(t, \varphi) \in L^\infty(0, \infty; Y)$ is the weak solution of (A.2.1).

Now we consider a gradient operator

$$L + G = -DF, \quad (\text{A.2.9})$$

where the function $F : Y \rightarrow \mathbb{R}^1$, and $Y \subseteq H_{1/2}$ is a Banach space. We also assume that

$$F(u) \rightarrow +\infty \Leftrightarrow \|u\|_Y \rightarrow \infty, \quad (\text{A.2.10})$$

$$(Lu + Gu, u) \leq -C_1\|u\|_Y^2 + C_2, \quad (\text{A.2.11})$$

where $C_1, C_2 > 0$ are constants.

The following theorem can be derived using the existence theorem of global attractors in Ma et al. (2002).

Theorem A.2.3 *Under the assumptions (A.2.9)–(A.2.11), Eq. (A.2.1) generates an operator semigroup $S(t) : H \rightarrow H$, which has a global attractor \mathcal{A} in H attracting any bounded set in H -norm.*

The dynamical systems associated with gradient-type equations have some nice properties in their global attractors and asymptotical behaviors. In particular, for the systems generated by the Morse type operators, the attractors possess simpler topological structure. A gradient-type operator $L + G : H_1 \rightarrow H$ is called a Morse field if all singular points of $L + G$ are regular, i.e., the derivative operators of $L + G$ at their singular points are invertible

$$L + DG(u_0) : H_1 \rightarrow H.$$

It is known that a Morse field has a finite number of singular points, and the unstable manifolds at the singular points are homeomorphic to finite-dimensional open ball.

The following theorem can be derived directly from the classical invariant manifold theorems, and Lemma A.2.7 below.

Theorem A.2.4 *Assume that (A.2.1) possesses a global attractor \mathcal{A} in H . Then we have the following assertions:*

- (1) *The attractor \mathcal{A} consists of all singular points of $L + G$ and their unstable sets, and*

$$\dim \mathcal{A} = \text{maximal dimension of unstable sets.}$$

- (2) *\mathcal{A} contains a set S of singular points such that for all $\varphi \in H$ the solutions $u(t, \varphi)$ of (A.2.1) converge to S*

$$\lim_{t \rightarrow \infty} \text{dist}(u(t, \varphi), S) = 0, \quad \forall \varphi \in H.$$

- (3) *If $L + G$ is a Morse field, then \mathcal{A} consists of finite number of finite-dimensional closed balls and singular points.*

In order to understand the global dynamics of gradient flows, we need to study the local topological structure of singular points. In particular, we are interested in isolated singular points because they have nice properties, and they are generic.

Let G_C be a continuous group, which acts on H

$$g \circ u \in H, \quad (g \circ u = 0 \Leftrightarrow u = 0), \quad \forall g \in G_C, \quad u \in H.$$

We can see that if under the action of the group G_C the operator $L + G$ is invariant, i.e.

$$g \circ (Lu + Gu) = L(g \circ u) + G(g \circ u), \quad \forall g \in G_C, \quad u \in H,$$

then, for any singular point u_0 of $L + G$, the set

$$S = \{g \circ u_0 \mid g \in G_C\}$$

consists of singular points of $L + G$. Obviously, in this case $L + G$ has no isolated singular points in the usual sense. However, we can define the (generalized) isolated singular points as follows.

Definition A.2.5 Let G_C be a continuous group, and $L + G$ be invariant under the action of G_C . A singular point u_0 of $L + G$ is called isolated if there is a neighborhood U of $S = \{g \circ u_0 | g \in G_C\}$ such that there are no other singular points $\tilde{u} \notin S$ of $L + G$ in U .

Hereafter isolated singular points are always understood in the sense of Definition A.2.5.

Definition A.2.6 Let $u_0 \in H$ be a singular point of $L + G$. The sets W^s and $W^u \subset H$ are called the stable and unstable sets of u_0 respectively, if

$$\begin{aligned} W^s &= \{\varphi \in H | u(t, \varphi) \rightarrow u_0 \text{ in } H \text{ as } t \rightarrow \infty\}, \\ W^u &= \{\varphi \in H | u(t, \varphi) \rightarrow u_0 \text{ in } H \text{ as } t \rightarrow -\infty\}, \end{aligned}$$

where $u(t, \varphi)$ is the solution of (A.2.1).

For any two singular points u_1 and u_2 of (A.2.1), we have

$$W_1^u \cap W_2^s = \emptyset \quad \text{if } F(u_2) \geq F(u_1).$$

Namely, the unstable manifold of the singular point with lower energy does not intersect with the stable manifold of singular point with higher energy.

Lemma A.2.7 Let $u_0 \in H$ be an isolated singular point of (A.2.1), and $U \subset H$ be a neighborhood of u_0 . Then

$$\begin{aligned} U &= W_0^s + W_0^u + D, \quad D = \{\varphi \in U | \lim_{t \rightarrow \pm\infty} u(t, \varphi) \neq u_0\}, \\ W_0^s &= W^s \cap U, \quad W_0^u = W^u \cap U. \end{aligned} \tag{A.2.12}$$

Moreover, D is an open set, called the hyperbolic set of u_0 , and

$$\dim W^u < \infty, \quad \dim W^s = \infty, \quad W^u \cap W^s = \{u_0\}. \tag{A.2.13}$$

Proof. Since $G : H_\theta \rightarrow H$ ($0 < \theta < 1$) is a C^r ($r \geq 1$) mapping, $L + DG(u_0) : H_1 \rightarrow H$ is a linear completely continuous field and sectorial operator. Hence $L + DG(u_0)$ has an eigenvalue sequence $\{\lambda_k\} \subset \mathbb{C}$ (counting the multiplicity), such that

$$\begin{aligned} \operatorname{Re} \lambda_i &\geq 0 && \text{if } 1 \leq i \leq m, \\ \operatorname{Re} \lambda_j &< 0 && \text{if } j \geq m+1. \end{aligned}$$

The corresponding eigenspaces are

$$\begin{aligned} E_1 &= \left\{ \sum_{i=1}^m x_i e_i \mid (L + DG(u_0) - \lambda_j)^{k_j} e_j = 0, \quad 1 \leq j \leq m, \quad k_j \in \mathbb{N} \right\}, \\ E_2 &= \{u \in H \mid (u, e_i^*) = 0, \quad \forall 1 \leq i \leq m\}, \end{aligned}$$

where e_j^* are the conjugate eigenvectors of e_j ($1 \leq j \leq m$).

Equation (A.2.1) has an invariant manifold M^I and a stable manifold M^s at u_0 such that M^I and M^s are tangent to E_1 and E_2 respectively:

$$\begin{aligned} T_{u_0} M^I &= E_1, \quad T_{u_0} M^s = E_2, \\ \dim M^I &= \dim E_1 = m, \quad \dim M^s = \dim E_2 = \infty. \end{aligned}$$

It is clear that

$$W_0^u \subset M^I, \quad M^s \cap U \subset W_0^s.$$

Hence

$$\dim W_0^u < \infty, \quad \dim W_0^s = \infty.$$

By (A.2.7) we get

$$W_0^u \cap W_0^s = \{u_0\}.$$

Thus, we derive (A.2.13).

From (A.2.7) it follows that there are only three cases to occur:

- (i) $\lim_{t \rightarrow \infty} u(t, \varphi) = u_0, \quad \lim_{t \rightarrow -\infty} u(t, \varphi) \neq u_0;$
- (ii) $\lim_{t \rightarrow -\infty} u(t, \varphi) = u_0, \quad \lim_{t \rightarrow \infty} u(t, \varphi) \neq u_0;$ and
- (iii) $\lim_{t \rightarrow \pm\infty} u(t, \varphi) \neq u_0.$

Hence, we have

$$U = W_0^u + W_0^s + D.$$

By $\dim M^I < \infty$ and $W_0^u \subset M^I$, it is easy to see that W_0^u is relatively closed in U . We need to prove that W_0^s is also relatively closed in U .

By assumption, $G : H_\theta \rightarrow H$ is C^r ($r \geq 1$) for some $0 < \theta < 1$. It is known that (A.2.1) generates a local operator semigroup $S(t) : U \rightarrow H$ for $0 \leq t < \infty$. Noting that $S(t)\varphi \rightarrow u_0, \forall \varphi \in W_0^s$, and $S(t)$ is continuous for any $0 \leq t < \infty$, hence if $\varphi_n \in W_0^s$ and $\varphi_n \rightarrow \varphi_0$ in H , then $S(t)\varphi_0 \in U, \forall t \geq 0$. Thus, by (A.2.7) it follows that

$$\lim_{t \rightarrow \infty} S(t)\varphi_0 = u_0,$$

i.e., $\varphi_0 \in W_0^s$. Hence, W_0^s is a relatively closed set in U . Therefore $D = U/(W_0^u + W_0^s)$ is an open set. The proof is complete. \square

Let $u_0 \in H$ be an isolated singular point of (A.2.1). The eigenvalues of $L + DG(u_0)$ satisfy

$$\begin{aligned} \operatorname{Re} \lambda_i &= 0 && \text{for } 1 \leq i \leq n, \\ \operatorname{Re} \lambda_j &> 0 && \text{for } n+1 \leq j \leq n+m, \\ \operatorname{Re} \lambda_k &< 0 && \text{for } k \geq n+m+1. \end{aligned} \tag{A.2.14}$$

Equation (A.2.1) has three invariant manifolds at u_0 : the stable manifold M^s , the unstable manifold M^u , and the center manifold M^c such that

$$\dim M^c = n, \quad \dim M^u = m, \quad \dim M^s = \infty. \tag{A.2.15}$$

If $n = 0$ in (A.2.15), then the local topological structure of u_0 in its neighborhood U is simple, which is characterized by

$$W^u = M^u, \quad W^s = M^s, \quad D = U/(M^u + M^s).$$

Moreover, M^u is homeomorphic to an m -dimensional open ball, and

$$\text{ind}(-(L + G), u_0) = (-1)^m.$$

Especially, if $n = m = 0$, then u_0 is a local attractor.

As $n \neq 0$, the local topological structure of u_0 in the center manifold is complex. By the reduction theorem of index (Nirenberg, 2001; Chen, 1981), we derive that

$$\text{ind}(-(L + G), u_0) = (-1)^m \text{ind}(g, u_0), \quad (\text{A.2.16})$$

where $g = -(L + G)|_{M^c}$ is the restriction of $-(L + G)$ on M^c .

We remark that it is useful to study the locally topological structure of singular points in the center manifolds M^c for the bifurcation and transition theory developed in this book.

Let $B_r^c \subset M^c$ be the open ball with center u_0 and radius r , and

$$W_r^u = W^u \bigcap B_r^c, \quad W_r^s = W^s \bigcap B_r^c, \quad D_r = D \bigcap B_r^c.$$

Lemma A.2.8 *If $D_r = \emptyset$, then $B_r^c = W_r^u$ or $B_r^c = W_r^s$. If $D_r \neq \emptyset$, then*

$$D_r = \sum_{j=1}^k D_j \quad (1 \leq k < \infty), \quad (\text{A.2.17})$$

where each D_j is a connected component, and $\partial D_j \setminus \partial B_r^c$ consists of only two connected components Γ_j^+ and Γ_j^- :

$$\partial D_j \setminus \partial B_r^c = \Gamma_j^+ + \Gamma_j^-,$$

and $\Gamma_j^+ \subset W_r^u, \Gamma_j^- \subset W_r^s$. Moreover, Γ_j^+ (resp. Γ_j^-) is a deformation retract of $D_j + \Gamma_j^+$ (resp. $D_j + \Gamma_j^-$).

Proof. Obviously, if the hyperbolic set $D_r = \emptyset$, then the ball B_r^c in center manifold is either a stable manifold or an unstable manifold, i.e., either $B_r^c = W_r^s$ or $B_r^c = W_r^u$.

Let $D_r \neq \emptyset$. We consider the intersections of $\overline{W}_r^s, \overline{W}_r^u$ and \overline{D}_r with the sphere ∂B_r^c . Denote by

$$I = \overline{W}_r^s \bigcap \partial B_r^c, \quad O = \overline{W}_r^u \bigcap \partial B_r^c, \quad U = \overline{D}_r \bigcap \partial B_r^c.$$

Let $S(t)$ be the operator semigroup generated by (A.2.1). It is clear that

$$\begin{aligned}
F(\varphi) > F(\psi) & \quad \forall \varphi \in I \text{ and } \psi \in O \\
S(t)\varphi \in B_r^c & \quad \forall t > 0, \quad \text{as } \varphi \in I, \\
S(-t)\varphi \in B_r^c, \quad S(t)\varphi \notin B_r^c & \quad \forall t > 0, \quad \text{as } \varphi \in O;
\end{aligned} \tag{A.2.18}$$

namely the flows passing through the points $\varphi \in I$ are into B_r^c and the flows passing through $\varphi \in O$ are out B_r^c forever.

To complete this proof, it suffices to only consider the case where D_r is connected, for other cases the proof is the same.

Let D_r be connected. Then $U = \overline{D_r} \cap \partial B_r^c$ is also connected in ∂B_r^c . It is clear that $D_r \neq B_r^c$, i.e., $I \neq \emptyset, O \neq \emptyset$, otherwise, from (A.2.7) one can deduce that there are no singular points in B_r^c . Denote by

$$\begin{aligned}
U^- &= \{\varphi \in U \mid S(t)\varphi \in B_r^c, \forall 0 < t < t_\varphi \quad \text{and } S(t_\varphi)\varphi \in \partial B_r^c\}, \\
U^+ &= \{\varphi \in U \mid S(-t)\varphi \in B_r^c, \quad \forall 0 < t < t_\varphi, \quad \text{and } S(-t_\varphi)\varphi \in \partial B_r^c\}.
\end{aligned}$$

Namely, the flow $S(t)U^-$ enters B_r^c from U^- and leaves B_r^c from U^+ . It is clear that U^- is homeomorphic to U^+ by the mapping $h : U^- \rightarrow U^+$ defined by

$$h(\varphi) = \psi, \quad \varphi \in U^-, \quad \text{and } \psi = S(t_\varphi)\varphi \in U^+. \tag{A.2.19}$$

Let $I_0 \subset I$ be a connected component. We denote

$$T_\rho(I_0) = \{\varphi \in U \mid \text{dist}(\varphi, I_0) < \rho\}.$$

Then $T_\rho(I_0) + I_0$ is a collar neighborhood of I_0 in $U \subset \partial B_r^c$. It is known that I_0 is a deformation retract of $T_\rho(I_0) + I_0$. Obviously, for $\rho > 0$ sufficiently small, we have

$$T_\rho(I_0) \subset U^-.$$

On the other hand, by (A.2.18) we have

$$\begin{aligned}
\lim_{t \rightarrow \infty} S(t)\varphi &= u_0 \quad \forall \varphi \in I, \\
\lim_{t \rightarrow -\infty} S(-t)\varphi &= u_0 \quad \forall \varphi \in O.
\end{aligned} \tag{A.2.20}$$

By the stability of extended orbits, Theorem 2.1.15 in Ma and Wang (2005d), it follows from (A.2.19) to (A.2.20) that there is a connected component O_0 of O , such that

$$\begin{aligned}
h(T_\rho(I_0)) + O_0 &\text{ is a collar neighborhood of } O_0, \text{ and} \\
\text{dist}(h(T_\rho(I)), O_0) &\rightarrow 0 \text{ as } \rho \rightarrow 0.
\end{aligned} \tag{A.2.21}$$

Hence, O_0 is a deformation retract of $h(T_\rho(I^-)) + O_0$, and

$$h(T_\rho(I_0)) \subset U^+.$$

Let U_0^- and U_0^+ be the connected components of U^- and U^+ containing $T_\rho(I_0)$ and $h(T_\rho(I_0))$ respectively. Then, by the stability of extended orbits, Theorem 2.1.15

in Ma and Wang (2005d), we can deduce that $h : U_0^- \rightarrow U_0^+$ is a homeomorphism, where h is defined by (A.2.19), which implies that

$$\partial U_0^- / I_0 = \partial U_0^+ / O_0 = \partial U_0^- \cap \partial U_0^+,$$

and I_0 and O_0 are the deformation retracts of $U_0^- + I_0$ and $U_0^+ + O_0$ respectively. Hence $\partial U_0^- / I_0$ and $\partial U_0^+ / O_0$ are connected. Thus, we deduce that

$$U = U_0^- + U_0^+ + \partial U_0^- \cap \partial U_0^+.$$

It means that

$$\partial U = A^- + A^+, \quad A^- \subset I_0 = I, \quad A^+ \subset O_0 = 0. \quad (\text{A.2.22})$$

We note that D_r is homeomorphic to $U \times (0, r)$. Then, by (A.2.22) we have

$$\partial D_r = \Gamma^- + \Gamma^+, \quad \Gamma^- \subset W_r^s = I \times (0, r), \quad \Gamma^+ \subset W_r^u = O \times (0, r).$$

Moreover $\Gamma^- = I \times (0, r)$ (resp. $\Gamma^+ = O \times (0, r)$) is a deformation retract of $D_r + \Gamma^-$ (resp. $D_r + \Gamma^+$).

By (A.2.20) we can deduce that the number of the connected components of I and O must be finite, otherwise the singular point u_0 is not isolated. Hence, (A.2.19) holds true.

Thus, the lemma is proved. \square

A.3 Constraint Variation

A.3.1 General Orthogonal Decomposition Theorem

The following is a general orthogonal decomposition theorem. Let X be a linear space, H be a Hilbert space, and

$$L : X \rightarrow H \text{ be a linear map.}$$

Let

$$\mathcal{N} = \{\varphi \in X \mid L\varphi = 0\},$$

be the kernel of L , and let H_1 be the completion of X/\mathcal{N} with the following norm

$$\|\varphi\|_{H_1}^2 = \langle L\varphi, L\varphi \rangle_H, \quad \varphi \in X/\mathcal{N}. \quad (\text{A.3.1})$$

It is clear that H_1 is a Hilbert space, and

$$\begin{aligned} L : H_1 &\rightarrow H \text{ is bounded,} \\ L^* : H &\rightarrow H_1^* \text{ is the dual operator of } L. \end{aligned} \quad (\text{A.3.2})$$

Then we have the L -orthogonal decomposition theorem as follows.

Theorem A.3.1 (L-Orthogonal Decomposition Theorem) *For the linear operators L and L^* in (A.3.2) and any $u \in H$, there exists a $\varphi \in H_1$ and $v \in H$, such that u can be decomposed as*

$$u = L\varphi + v, \quad L^*v = 0, \quad \langle L\varphi, v \rangle_H = 0. \quad (\text{A.3.3})$$

Proof. For a given $u \in H$, consider the existence of the equation

$$L^*L\varphi = L^*u, \quad (\text{A.3.4})$$

for solution $\varphi \in H_1$. By (A.3.1) we can see that the operator

$$A = L^*L : H_1 \rightarrow H_1^*,$$

is positive definition, i.e., for any $\psi \in H_1$ we have

$$\langle A\psi, \psi \rangle_{H_1} = \langle L\psi, L\psi \rangle_H = \|\psi\|_{H_1}^2.$$

Therefore, by the Lax-Milgram theorem, the Eq. (A.3.4) has a unique solution $\varphi \in H_1$. Then we take

$$v = u - L\varphi \in H. \quad (\text{A.3.5})$$

By (A.3.4) we see that

$$L^*v = 0. \quad (\text{A.3.6})$$

Thus, (A.3.5) is the decomposition of (A.3.4), and by (A.3.6) we have

$$\langle L\varphi, v \rangle_H = \langle \varphi, L^*v \rangle_{H_1} = 0,$$

i.e., $L\varphi$ and v are orthogonal in H . The proof is complete.

We demonstrate the applications of the above theorem with two examples.

Example A.3.2 Let $H = L^2(\Omega, \mathbb{R}^n)$ be the space of all n-dimensional vector fields defined on an open set $\Omega \subset \mathbb{R}^n$, and

$$H_1 = H^1(\Omega)/\mathbb{R} = \left\{ \varphi \in H^1(\Omega) \mid \int_{\Omega} \varphi dx = 0 \right\}.$$

Let

$$L = \nabla : H_1 \rightarrow H \text{ be the gradient operator,} \quad (\text{A.3.7})$$

and the duality $L^* : H \rightarrow H_1^*$ is as follows

$$\langle L^*v, \varphi \rangle_{H_1} = \int_{\Omega} (-\operatorname{div} v) \varphi dx + \int_{\partial\Omega} \varphi v \cdot n ds \quad (\text{A.3.8})$$

$\forall \varphi \in H_1$. By (A.3.8) we can see that

$$L^*v = 0 \iff \operatorname{div} v = 0, v \cdot n|_{\partial\Omega} = 0. \quad (\text{A.3.9})$$

Therefore, for the operators (A.3.7) and (A.3.9) the vector fields u in $H = L^2(\Omega, \mathbb{R}^n)$ can be decomposed as

$$u = \nabla\varphi + v, \operatorname{div} v = 0, v \cdot n|_{\partial\Omega} = 0. \quad (\text{A.3.10})$$

This is the Leray decomposition. \square

Example A.3.3 The decomposition of the tensor in Riemannian manifolds. Let M be a n -dimensional manifold without boundary, and $H = L^2(T_r^k M)$ be the space consisting of square integrable (k, r) -tensors on M , and

$$H_1 = H^1(T_r^{k-1} M) \quad (\text{or } H^1(T_{r-1}^k M)).$$

For the gradient operator

$$L = \nabla : H_1 \rightarrow H,$$

the dual operator is the divergent operator

$$L^* = \operatorname{div} : H \rightarrow H_1^*.$$

By Theorem A.3.1, for $u \in H = L^2(T_r^k M)$ we have

$$u = \nabla\varphi + v, \operatorname{div} v = 0, \varphi \in H_1. \quad (\text{A.3.11})$$

A.3.2 Variations with Constrained-Infinitesimals

Let H, H_1 be two Hilbert spaces, and

$$L : H_1 \rightarrow H, \quad L^* : H \rightarrow H_1^* \quad (\text{A.3.12})$$

be a pair of dual linear bounded operators. Consider the functionals defined on H , i.e.,

$$F : H \rightarrow \mathbb{R}^1. \quad (\text{A.3.13})$$

Then we introduce the following definitions. Let

$$\mathcal{N}^* = \{v \in H \mid L^*v = 0\} \neq \{0\}, \quad (\text{A.3.14})$$

i.e., $\dim \mathcal{N}^* > 0$.

We now recall the notion of constraint variations given in Definition 1.1.4:

1. For any $u \in H$, the derivative operator of F at u with L^* -constraint, denoted by $\delta_{L^*} F(u)$, is defined as follows

$$\langle \delta_{L^*} F(u), v \rangle_H = \frac{d}{dt} \Big|_{t=0} F(u + tv) \quad \forall v \in \mathcal{N}^*. \quad (\text{A.3.15})$$

2. The derivative operator of F at u with L -constraint, denoted by $\delta_L F(u)$, is defined as follows

$$\langle \delta_L F(u), \varphi \rangle_{H_1} = \frac{d}{dt} \Big|_{t=0} F(u + tL\varphi) \quad \forall \varphi \in H_1. \quad (\text{A.3.16})$$

The following theorem is based on Theorem A.3.1.

Theorem A.3.4 (Variation with L -constraints) *For the variational derivatives with L^* and L constraints defined in (A.3.15) and (A.3.16) we have the following conclusions.*

1. For $\delta_{L^*} F(u)$, there is a $\varphi \in H_1$ such that

$$\delta_{L^*} F(u) = \delta F(u) + L\varphi. \quad (\text{A.3.17})$$

2. For $\delta_L F(u)$, we have

$$\delta_L F(u) = L^* \delta F(u). \quad (\text{A.3.18})$$

Here $\delta F(u)$ is the normal derivative operator.

Proof. The theorem follows from Definition 1.1.4 and Theorem A.3.1. In fact, by definition, we have

$$\langle \delta_{L^*} F(u), v \rangle_H = \langle \delta F(u), v \rangle \quad \forall L^* v = 0,$$

which, by Theorem A.3.1, implies (A.3.17).

Also, by definition, we have

$$\langle \delta_L F(u), \varphi \rangle_{H_1} = \langle \delta F(u), L\varphi \rangle_H = \langle L^* \delta F(u), \varphi \rangle_{H_1},$$

which leads to (A.3.18).

By Theorem A.3.4, we shall show that all physical motion systems, without coupling to other systems, have a functional

$$F : H \rightarrow \mathbb{R}^1$$

such that the dynamical equations of the systems can be expressed as

$$\frac{du}{dt} = -A \delta_L F(u), \quad (\text{A.3.19})$$

where A is a coefficient matrix, and $\delta_L F(u)$ is the derivative operator of F at u with L -constraint. Note that the normal derivative operator $\delta F(u)$ corresponds to $\delta_L F(u)$ with $L = id$ being the identity.

References

- Aftalion, A. and Q. Du (2002). The bifurcation diagrams for the Ginzburg-Landau system of superconductivity. *Phys. D* 163(1-2), 94–105.
- Anderson, M. H., J. R. Ensher, M. R. Matthews, C. E. Wieman, and E. A. Cornell (1995, July). Observation of Bose-Einstein Condensation in a Dilute Atomic Vapor. *Science* 269, 198–201.
- Andronov, A. A., E. A. Leontovich, I. I. Gordon, and A. G. Maier (1973). *Theory of bifurcations of dynamic systems on a plane*. Halsted Press [A division of John Wiley & Sons], New York-Toronto, Ont. Translated from the Russian.
- Bao, W. (2002). The random projection method for a model problem of combustion with stiff chemical reactions. *Appl. Math. Comput.* 130(2-3), 561–571.
- Bardeen, J., L. N. Cooper, and J. R. Schrieffer (1957). Theory of superconductivity. *Phys. Rev.* 108, 1175–1204.
- Bates, P. W. and P. C. Fife (1993). The dynamics of nucleation for the Cahn-Hilliard equation. *SIAM J. Appl. Math.* 53(4), 990–1008.
- Batiste, O., E. Knobloch, A. Alonso, and I. Mercader (2006). Spatially localized binary-fluid convection. *J. Fluid Mech.* 560, 149–158.
- Battisti, D. S. and A. C. Hirst (1989). Interannual variability in a tropical atmosphere-ocean model. influence of the basic state, ocean geometry and nonlinearity. *J. Atmos. Sci.* 46, 1687–1712.
- Bednorz, J. G. and K. A. Muller (1986). Possible high T_c superconductivity in the Ba-La-Cu-O system. *Z. Phys. B* 64, 189–193.
- Belousov, B. P. (1959). An oscillating reaction and its mechanism. *Sborn. referat. radiat. med.*, 145.
- Bjerknes, V. (1904). Das problem von der wettervorhersage, betrachtet vom standpunkt der mechanik un der physik. *Meteor. Z.* 21, 1–7.
- Bona, J. L., C.-H. Hsia, T. Ma, and S. Wang (2011). Hopf bifurcation for two-dimensional doubly diffusive convection. *Appl. Anal.* 90(1), 5–30.
- Bose, S. N. (1924). Plancks Gesetz und Lichtquantenhypothese. *Zeitschrift für Physik* 26, 178–181.

- Branstator, G. W. (1987). A striking example of the atmosphere's leading traveling pattern. *J. Atmos. Sci.* 44, 2310–2323.
- Brenner, M. P., L. S. Levitov, and E. O. Budrene (1998). Physical mechanisms for chemotactic pattern formation by bacteria. *Biophysical Journal* 74, 1677–1693.
- Budrene, E. O. and H. C. Berg (1991). Complex patterns formed by motile cells of *Escherichia coli*. *Nature* 349, 630–633.
- Budrene, E. O. and H. C. Berg (1995). Dynamics of formation of symmetric patterns of chemotactic bacteria. *Nature* 376, 49–53.
- Busse, F. (1978). Non-linear properties of thermal convection. *Reports on Progress in Physics* 41, 1929–1967.
- Caffarelli, L., R. Kohn, and L. Nirenberg (1982). On the regularity of the solutions of Navier-Stokes equations. *Comm. Pure Appl. Math.* 35, 771–831.
- Cahn, J. and J. E. Hillard (1957). Free energy of a nonuniform system i. interfacial energy. *J. Chemical Physics* 28, 258–267.
- Chaikin, P. M. and T. C. Lubensky (2000). *Principles of condensed matter physics*. Cambridge university press.
- Chandrasekhar, S. (1981). *Hydrodynamic and Hydromagnetic Stability*. Dover Publications, Inc.
- Charney, J. (1947). The dynamics of long waves in a baroclinic westerly current. *J. Meteorol.* 4, 135–163.
- Charney, J. (1948). On the scale of atmospheric motion. *Geophys. Publ.* 17(2), 1–17.
- Chekroun, M., H. Liu, and S. Wang (2014a). *Approximation of Invariant Manifolds: Stochastic Manifolds for Nonlinear SPDEs I*. Springer Briefs in Mathematics. Springer, New York.
- Chekroun, M., H. Liu, and S. Wang (2014b). *Parameterizing Manifolds and Non-Markovian Reduced Equations: Stochastic Manifolds for Nonlinear SPDEs II*. Springer Briefs in Mathematics. Springer, New York.
- Chen, W. (1981). *Nonlinear functional analysis*. P. R. China: Gansu Renmin Press.
- Choi, Y., T. Ha, J. Han, and D. S. Lee (2017). Bifurcation and final patterns of a modified Swift-Hohenberg equation. *Discrete & Continuous Dynamical Systems-Series B* 22(7).
- Choi, Y., J. Han, and C.-H. Hsia (2015). Bifurcation analysis of the damped Kuramoto-Sivashinsky equation with respect to the period. *Discrete Contin. Dyn. Syst. Ser. B* 20(7), 1933–1957.
- Choi, Y., J. Han, and J. Park (2015). Dynamical bifurcation of the generalized Swift-Hohenberg equation. *International Journal of Bifurcation and Chaos* 25(08), 1550095.
- Chow, S. N. and J. K. Hale (1982). *Methods of bifurcation theory*, Volume 251 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science]*. New York: Springer-Verlag.
- Crandall, M. G. and P. H. Rabinowitz (1977). The Hopf bifurcation theorem in infinite dimensions. *Arch. Rational Mech. Anal.* 67(1), 53–72.
- Cross, M. and P. Hohenberg (1993). Pattern formation outside of equilibrium. *Reviews of Modern Physics* 65(3), 851–1112.

- Davis, K. B., M. O. Mewes, M. R. Andrews, N. J. van Druten, D. S. Durfee, D. M. Kurn, and W. Ketterle (1995, Nov). Bose-Einstein condensation in a gas of sodium atoms. *Phys. Rev. Lett.* 75, 3969–3973.
- de Gennes, P. (1966). *Superconductivity of Metals and Alloys*. W. A. Benjamin.
- del Pino, M., P. L. Felmer, and P. Sternberg (2000). Boundary concentration for eigenvalue problems related to the onset of superconductivity. *Comm. Math. Phys.* 210(2), 413–446.
- Desai, R. C. and R. Kapral (2009). *Self-Organized and Self-Assembled Structures*. Cambridge University Press.
- Dijkstra, H., T. Sengul, J. Shen, and S. Wang (2015). Dynamic transitions of quasi-geostrophic channel flow. *SIAM Journal on Applied Mathematics* 75(5), 2361–2378.
- Dijkstra, H., T. Sengul, and S. Wang (2013). Dynamic transitions of surface tension driven convection. *Physica D: Nonlinear Phenomena* 247(1), 7–17.
- Dijkstra, H. A. (2000). *Nonlinear Physical Oceanography: A Dynamical Systems Approach to the Large Scale Ocean Circulation and El Niño*. Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Dijkstra, H. A. and M. Ghil (2005). Low-frequency variability of the large-scale ocean circulations: a dynamical systems approach. *Review of Geophysics* 43, 1–38.
- Dijkstra, H. A. and M. J. Molemaker (1997). Symmetry breaking and overturning oscillations in thermohaline-driven flows. *J. Fluid Mech.* 331, 195–232.
- Dijkstra, H. A. and M. J. Molemaker (1999). Imperfections of the North-Atlantic wind-driven ocean circulation: Continental geometry and wind stress shape. *J. Mar. Res.* 57, 1–28.
- Dijkstra, H. A. and J. D. Neelin (1999). Imperfections of the thermohaline circulation: Multiple equilibria and flux-correction. *J. Clim.* 12, 1382–1392.
- Dijkstra, H. A. and J. D. Neelin (2000). Imperfections of the thermohaline circulation: Latitudinal asymmetry versus asymmetric freshwater flux. *J. Clim.* 13, 366–382.
- Drazin, P. and W. Reid (1981). *Hydrodynamic Stability*. Cambridge University Press.
- Du, Q., M. D. Gunzburger, and J. S. Peterson (1992). Analysis and approximation of the Ginzburg-Landau model of superconductivity. *SIAM Rev.* 34(1), 54–81.
- Eckhardt, B., T. M. Schneider, B. Hof, and J. Westerweel (2007). Turbulence transition in pipe flow. *Annu. Rev. Fluid Mech.* 39, 447–468.
- Ehrenfest, P. (1933). *Phasenumwandlungen im ueblichen und erweiterten Sinn, classifiziert nach den entsprechenden Singularitaeten des thermodynamischen Potentiales*. NV Noord-Hollandsche Uitgevers Maatschappij.
- Einstein, A. (1924). Quantentheorie des einatomigen idealen gases. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physik- Mathematik*, 261–267.
- Einstein, A. (1925). Quantentheorie des einatomigen idealen gases. zweite abhandlung. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physik- Mathematik*, 3–14.
- Field, M. (1996). *Lectures on bifurcations, dynamics and symmetry*, Volume 356 of *Pitman Research Notes in Mathematics Series*. Harlow: Longman.

- Field, R. J., E. Körös, and R. M. Noyes (1972). Oscillations in chemical systems, Part 2. thorough analysis of temporal oscillations in the bromate-cerium-malonic acid system. *J. Am Chem. Soc.* 94, 8649–8664.
- Field, R. J. and R. M. Noyes (1974). Oscillations in chemical systems, IV. limit cycle behavior in a model of a real chemical reaction. *J. Chem. Physics.* 60, 1877–1884.
- Fisher, M. (1964). Specific heat of a gas near the critical point. *Physical Review* 136:6A, A1599–A1604.
- Fisher, M. E. (1998). Renormalization group theory: Its basis and formulation in statistical physics. *Reviews of Modern Physics* 70(2), 653.
- Foias, C., O. Manley, and R. Temam (1987). Attractors for the Bénard problem: existence and physical bounds on their fractal dimension. *Nonlinear Anal.* 11(8), 939–967.
- Foiaş, C. and R. Temam (1979). Some analytic and geometric properties of the solutions of the evolution Navier-Stokes equations. *J. Math. Pures Appl.* (9) 58(3), 339–368.
- Friedlander, S., M. Vishik, and V. Yudovich (2000). Unstable eigenvalues associated with inviscid fluid flows. *J. Math. Fluid Mech.* 2(4), 365–380.
- Getling, A. V. (1997). *Rayleigh-Benard convection: Structures and dynamics*. Advanced Series in Nonlinear Dynamics. World Scientific.
- Ghil, M. (1976). Climate stability for a sellers-type model. *J. Atmos. Sci.* 33, 3–20.
- Ghil, M. (2000). Is our climate stable? Bifurcations, transitions and oscillations in climate dynamics, in Science for survival and sustainable development, V. I. Keilis-Borok and M. Sorondo (eds.), Pontifical Academy of Sciences. pp. 163–184.
- Ghil, M. and S. Childress (1987). *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory, and Climate Dynamics*. Springer-Verlag, New York.
- Ghil, M., T. Ma, and S. Wang (2001). Structural bifurcation of 2-D incompressible flows. *Indiana Univ. Math. J.* 50(Special Issue), 159–180. Dedicated to Professors Ciprian Foias and Roger Temam (Bloomington, IN, 2000).
- Ghil, M., T. Ma, and S. Wang (2005). Structural bifurcation of 2-D nondivergent flows with Dirichlet boundary conditions: applications to boundary-layer separation. *SIAM J. Appl. Math.* 65(5), 1576–1596 (electronic).
- Ginzburg, V. L. (2004). On superconductivity and superfluidity (what i have and have not managed to do), as well as on the 'physical minimum' at the beginning of the xxi century. *Phys.-Usp.* 47, 1155–1170.
- Glansdorff, P. and I. Prigogine (1971). Structure, stability, and fluctuations. Wiley-Interscience, New York.
- Goldstein, S. (1937). The stability of viscous fluid flow between rotating cylinders. *Proc. Camb. Phil. Soc.* 33, 41–61.
- Golubitsky, M. and D. G. Schaeffer (1985). *Singularities and groups in bifurcation theory. Vol. I*, Volume 51 of *Applied Mathematical Sciences*. New York: Springer-Verlag.

- Golubitsky, M., I. Stewart, and D. G. Schaeffer (1988). *Singularities and groups in bifurcation theory. Vol. II*, Volume 69 of *Applied Mathematical Sciences*. New York: Springer-Verlag.
- Gor'kov, L. (1968). Generalization of the Ginzburg-Landau equations for non-stationary problems in the case of alloys with paramagnetic impurities. *Sov. Phys. JETP* 27, 328–334.
- Gross, E. P. (1961). Structure of a quantized vortex in boson systems. *Il Nuovo Cimento (1955–1965)* 20(3), 454–477.
- Guckenheimer, J. and P. Holmes (1990). *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, Volume 42 of *Applied Mathematical Sciences*. New York: Springer-Verlag. Revised and corrected reprint of the 1983 original.
- Guo, B. and Y. Han (2009). Attractor and spatial chaos for the Brusselator in \mathbf{R}^N . *Nonlinear Anal.* 70(11), 3917–3931.
- Guo, Y. and H. J. Hwang (2010). Pattern formation (I): the Keller-Segel model. *J. Differential Equations* 249(7), 1519–1530.
- Hale, J. (1988). Asymptotic behaviour of dissipative systems. *AMS Providence RI*.
- Han, D., M. Hernandez, and Q. Wang (2018). Dynamical transitions of a low-dimensional model for Rayleigh–Bénard convection under a vertical magnetic field. *Chaos, Solitons & Fractals* 114, 370–380.
- Han, J. and C.-H. Hsia (2012). Dynamical bifurcation of the two dimensional Swift-Hohenberg equation with odd periodic condition. *Dis. Cont. Dyn. Sys. B* 17, 2431–2449.
- Hastings, S. P. and J. D. Murray (1975). The existence of oscillatory solutions in the Field-Noyes model for the Belousov-Zhabotinskii reaction. *SIAM J. Appl. Math.* 28, 678–688.
- Held, I. M. and M. J. Suarez (1974). Simple albedo feedback models of the ice caps. *Tellus* 26, 613–629.
- Henry, D. (1981). *Geometric theory of semilinear parabolic equations*, Volume 840 of *Lecture Notes in Mathematics*. Berlin: Springer-Verlag.
- Hernández, M. and K. W. Ong (2018). Stochastic Swift-Hohenberg equation with degenerate linear multiplicative noise. *Journal of Mathematical Fluid Mechanics*, 1–20.
- Ho, T.-L. (1998). Spinor bose condensates in optical traps. *Physical review letters* 81(4), 742.
- Hopf, E. (1942). Abzweigung einer periodischen Lösung von einer stationären Lösung eines differentialsystems. *Ber. Math.-Phys. K. Sachs. Akad. Wiss. Leipzig* 94, 1–22.
- Hopf, E. (1951). Über die Anfangswertaufgabe für die hydrodynamischen Grundgleichungen. *Math. Nachr.* 4, 213–231.
- Hou, Z. and T. Ma (2013). Dynamic phase transition for the Taylor problem in the wide-gap case. *Bound. Value Probl.*, 2013:227, 13.
- Hsia, C.-H., C.-S. Lin, T. Ma, and S. Wang (2011). Tropical atmospheric circulations with humidity effects.

- Hsia, C.-H., T. Ma, and S. Wang (2007). Stratified rotating Boussinesq equations in geophysical fluid dynamics: dynamic bifurcation and periodic solutions. *J. Math. Phys.* 48(6), 065602, 20.
- Hsia, C.-H., T. Ma, and S. Wang (2008a). Attractor bifurcation of three dimensional double-diffusive convection. *ZAA* 27, 233–252.
- Hsia, C.-H., T. Ma, and S. Wang (2008b). Bifurcation and stability of two-dimensional double-diffusive convection. *Commun. Pure Appl. Anal.* 7(1), 23–48.
- Hsia, C.-H., T. Ma, and S. Wang (2010). Rotating Boussinesq equations: dynamic stability and transitions. *Discrete Contin. Dyn. Syst.* 28(1), 99–130.
- Iooss, G. and M. Adelmeyer (1998). *Topics in bifurcation theory and applications* (Second ed.), Volume 3 of *Advanced Series in Nonlinear Dynamics*. River Edge, NJ: World Scientific Publishing Co. Inc.
- Jin, F. F. (1996). Tropical ocean-atmosphere interaction, the pacific cold tongue, and the el nino southern oscillation. *Science* 274, 76–78.
- Jin, F. F., D. Neelin, and M. Ghil (1996). El niño southern oscillation and the annual cycle: subharmonic frequency locking and aperiodicity. *Physica D* 98, 442–465.
- Johnson, M. A., P. Noble, L. M. Rodrigues, Z. Yang, and K. Zumbrun (2019). Spectral stability of inviscid roll waves. *Comm. Math. Phys.* 367(1), 265–316.
- Kadanoff, L. P. (2000). *Statistical physics: statics, dynamics and renormalization*. World Scientific Publishing Co Inc.
- Kaper, H., S. Wang, and M. Yari (2009). Dynamic transitions of turing patterns. *Nonlinearity* 22, 601–626.
- Kaper, H. G. and T. J. Kaper (2002). Asymptotic analysis of two reduction methods for systems of chemical reactions. *Phys. D* 165(1-2), 66–93.
- Kapitza, P. (1938, 01). Viscosity of liquid helium below the λ -point. *Nature* 141, 74 EP –.
- Kapral, R. and K. Showalter (1995). *Chemical Waves and Patterns*. Kluwer, Dordrecht.
- Kato, T. (1995). *Perturbation theory for linear operators*. Classics in Mathematics. Berlin: Springer-Verlag. Reprint of the 1980 edition.
- Keller, E. F. and L. A. Segel (1970). Initiation of slime mold aggregation viewed as an instability. *J. Theor. Biol.* 26, 399–415.
- Kieu, C., T. Sengul, Q. Wang, and D. Yan (2018). On the Hopf (double Hopf) bifurcations and transitions of two-layer western boundary currents. *Communications in Nonlinear Science and Numerical Simulation*.
- Kirchgässner, K. (1975). Bifurcation in nonlinear hydrodynamic stability. *SIAM Rev.* 17(4), 652–683.
- Kiselev, A. A. and O. A. Ladyženskaya (1957). On the existence and uniqueness of the solution of the nonstationary problem for a viscous, incompressible fluid. *Izv. Akad. Nauk SSSR. Ser. Mat.* 21, 655–680.
- Kleman, M. and O. D. Lavrentovich (2007). *Soft matter physics: an introduction*. Springer Science & Business Media.
- Koschmieder, E. L. (1993). *Bénard cells and Taylor vortices*. Cambridge Monographs on Mechanics. Cambridge University Press.

- Kosterlitz, J. M. and D. J. Thouless (1973). Ordering, metastability and phase transitions in two-dimensional systems. *Journal of Physics C: Solid State Physics* 6(7), 1181.
- Krasnosel'skii, M. A. (1956). *Topologicheskie metody v teorii nelineinykh integralnykh uravnenii*. Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow.
- Kushnir, Y. (1987). Retrograding wintertime low-frequency disturbances over the north pacific ocean. *J. Atmos. Sci.* 44, 2727–2742.
- Kuznetsov, Y. A. (2004). *Elements of applied bifurcation theory* (Third ed.), Volume 112 of *Applied Mathematical Sciences*. Springer-Verlag, New York.
- Ladyzenskaja, O. (1991). *Attractors for Semigroups and Evolution Equations*. Cambridge Univ. Press.
- Ladyženskaya, O. A. (1958). On the nonstationary Navier-Stokes equations. *Vestnik Leningrad. Univ.* 13(19), 9–18.
- Landau, L. D. and E. M. Lifshitz (1969). *Statistical Physics: V. 5: Course of Theoretical Physics*. Pergamon press.
- Landau, L. D. and E. M. Lifshitz (1975). *Course of theoretical physics, Vol. 2* (Fourth ed.). Oxford: Pergamon Press. The classical theory of fields, Translated from the Russian by Morton Hamermesh.
- Langer, J. (1971). Theory of spinodal decomposition in alloys. *Ann. of Physics* 65, 53–86.
- Lapidus, I. and R. Schiller (1976). Model for the chemotactic response of a bacterial population. *Biophys J.* 16(7), 779–789.
- Lappa, M. (2009). *Thermal convection: Patterns, evolution and stability*. Wiley.
- Leray, J. (1933). Etude de diverses équations intégrales non linéaires et de quelques problèmes que posent l'hydrodynamique. *J. Math. Pures et Appl.* XII, 1–82.
- Li, D. and Z.-Q. Wang (2018). Local and global dynamic bifurcations of nonlinear evolution equations. *Indiana Univ. Math. J.* 67(2), 583–621.
- Li, J. (2017, Aug). Dynamic bifurcation for the granulation convection in the solar photosphere. *Boundary Value Problems* 2017(1), 110.
- Li, J. and S. Wang (2008). Some mathematical and numerical issues in geophysical fluid dynamics and climate dynamics. *Communications in Computational Physics* 3:4, 759–793.
- Li, L., M. Hernandez, and K. W. Ong (2018). Stochastic attractor bifurcation for the two-dimensional Swift-Hohenberg equation. *Mathematical Methods in the Applied Sciences* 41(5), 2105–2118.
- Li, L. and K. W. Ong (2016). Dynamic transitions of generalized Burgers equation. *J. Math. Fluid Mech.* 18(1), 89–102.
- Lifschitz, E. M. and L. P. Pitajewski (1990). *Lehrbuch der theoretischen Physik (“Landau-Lifschitz”)*. Band X (Second ed.). Akademie-Verlag, Berlin. Physikalische Kinetik. [Physical kinetics], Translated from the Russian by Gerd Röpke and Thomas Frauenheim, Translation edited and with a foreword by Paul Ziesche and Gerhard Diener.
- Lifshitz, E. and L. Pitaevskii (1980). Statistical physics part 2, Landau and Lifshitz course of theoretical physics vol. 9.

- Lin, C. and F. Shu (1964). On the spiral structure of disk galaxies. *Astrophysical Journal* 140, 646–655.
- Lions, J.-L., R. Temam, and S. Wang (1992a). New formulations of the primitive equations of atmosphere and applications. *Nonlinearity* 5(2), 237–288.
- Lions, J.-L., R. Temam, and S. Wang (1992b). On the equations of the large-scale ocean. *Nonlinearity* 5(5), 1007–1053.
- Lions, J.-L., R. Temam, and S. Wang (1993). Models for the coupled atmosphere and ocean. (CAO I,II). *Comput. Mech. Adv.* 1(1), 120.
- Liu, H., T. Sengul, and S. Wang (2011). Dynamic transitions for quasilinear systems and Cahn-Hilliard equation with Onsager mobility. *Journal of Mathematical Physics* 52(1), 013508.
- Liu, H., T. Sengul, and S. Wang (2012a). Dynamic transitions for quasilinear systems and Cahn-Hilliard equation with Onsager mobility. *Journal of Mathematical Physics* 53(2), 023518.
- Liu, H., T. Sengul, and S. Wang (2012b). Dynamic transitions for quasilinear systems and Cahn-Hilliard equation with Onsager mobility. *Journal of Mathematical Physics* 53, 023518, <https://doi.org/10.1063/1.3687414>, 1–31.
- Liu, H., T. Sengul, S. Wang, and P. Zhang (2015). Dynamic transitions and pattern formations for a Cahn–Hilliard model with long-range repulsive interactions. *Communications in Mathematical Sciences* 13(5), 1289–1315.
- Liu, R., T. Ma, S. Wang, and J. Yang (2017a). Dynamic theory of fluctuations and critical exponents of thermodynamic phase transitions. [hal-01674269](https://arxiv.org/abs/1609.04269).
- Liu, R., T. Ma, S. Wang, and J. Yang (2017b). Topological Phase Transition V: Interior Separation and Cyclone Formation Theory. *Hal preprint*: hal-01673496.
- Liu, R., T. Ma, S. Wang, and J. Yang (2019). Thermodynamical potentials of classical and quantum systems. *Discrete & Continuous Dynamical Systems - B; see also hal-01632278(2017)* 24, 1411.
- Lorenz, E. N. (1963a). Deterministic nonperiodic flow. *J. Atmos. Sci.* 20, 130–141.
- Lorenz, E. N. (1963b). The mechanics of vacillation. *J. Atmos. Sci.* 20, 448–464.
- Lorenz, E. N. (1967). *The Nature and Theory of the General Circulation of the Atmosphere*. World Meteorological Organization, Geneva, Switzerland.
- Luo, H., Q. Wang, and T. Ma (2015a). A predictable condition for boundary layer separation of 2-D incompressible fluid flows. *Nonlinear Anal. Real World Appl.* 22, 336–341.
- Luo, H., Q. Wang, and T. Ma (2015b). A predictable condition for boundary layer separation of 2-d incompressible fluid flows. *Nonlinear Anal. Real World Appl.* 22, 336–341.
- Ma, Q., S. Wang, and C. Zhong (2002). Necessary and sufficient conditions for the existence of global attractors for semigroups and applications. *Indiana Univ. Math. J.* 51(6), 1541–1559.
- Ma, T. (2011). *Theory and Methods of Partial Differential Equations (in Chinese)*. Beijing, Science Press.
- Ma, T., D. Li, R. Liu, and J. Yang (2016). Mathematical theory for quantum phase transitions. <https://arxiv.org/pdf/1610.06988.pdf>.

- Ma, T., R. Liu, and J. Yang (2018). *Physical World from the Mathematical Point of View: Statistical Physics and Critical Phase Transition Theory (in Chinese)*. Science Press, Beijing.
- Ma, T. and S. Wang (2000). Structural evolution of the Taylor vortices. *M2AN Math. Model. Numer. Anal.* 34(2), 419–437. Special issue for R. Temam’s 60th birthday.
- Ma, T. and S. Wang (2001). Structure of 2D incompressible flows with the Dirichlet boundary conditions. *Discrete Contin. Dyn. Syst. Ser. B* 1(1), 29–41.
- Ma, T. and S. Wang (2004a). Boundary layer separation and structural bifurcation for 2-D incompressible fluid flows. *Discrete Contin. Dyn. Syst.* 10(1-2), 459–472. Partial differential equations and applications.
- Ma, T. and S. Wang (2004b). Dynamic bifurcation and stability in the Rayleigh-Bénard convection. *Commun. Math. Sci.* 2(2), 159–183.
- Ma, T. and S. Wang (2004c). Interior structural bifurcation and separation of 2D incompressible flows. *J. Math. Phys.* 45(5), 1762–1776.
- Ma, T. and S. Wang (2005a). Bifurcation and stability of superconductivity. *J. Math. Phys.* 46(9), 095112, 31.
- Ma, T. and S. Wang (2005b). *Bifurcation theory and applications*, Volume 53 of *World Scientific Series on Nonlinear Science. Series A: Monographs and Treatises*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ.
- Ma, T. and S. Wang (2005c). Dynamic bifurcation of nonlinear evolution equations. *Chinese Ann. Math. Ser. B* 26(2), 185–206.
- Ma, T. and S. Wang (2005d). *Geometric theory of incompressible flows with applications to fluid dynamics*, Volume 119 of *Mathematical Surveys and Monographs*. Providence, RI: American Mathematical Society.
- Ma, T. and S. Wang (2006). Stability and bifurcation of the Taylor problem. *Arch. Ration. Mech. Anal.* 181(1), 149–176.
- Ma, T. and S. Wang (2007a). Rayleigh-Bénard convection: dynamics and structure in the physical space. *Commun. Math. Sci.* 5(3), 553–574.
- Ma, T. and S. Wang (2007b). *Stability and Bifurcation of Nonlinear Evolutions Equations*. Science Press, Beijing.
- Ma, T. and S. Wang (2008a). Dynamic model and phase transitions for liquid helium. *Journal of Mathematical Physics* 49:073304, 1–18.
- Ma, T. and S. Wang (2008b). Dynamic phase transition theory in PVT systems. *Indiana University Mathematics Journal* 57:6, 2861–2889.
- Ma, T. and S. Wang (2008c). Dynamic transitions for ferromagnetism. *Journal of Mathematical Physics* 49:053506, 1–18.
- Ma, T. and S. Wang (2008d). Exchange of stabilities and dynamic transitions. *Georgian Mathematics Journal* 15:3, 581–590.
- Ma, T. and S. Wang (2008e). Superfluidity of helium-3. *Physica A: Statistical Mechanics and its Applications* 387:24, 6013–6031.
- Ma, T. and S. Wang (2009a). Boundary-layer and interior separations in the Taylor-Couette-Poiseuille flow. *J. Math. Phys.* 50(3), 033101, 29.
- Ma, T. and S. Wang (2009b). Cahn-Hilliard equations and phase transition dynamics for binary systems. *Dist. Cont. Dyn. Systs., Ser. B* 11:3, 741–784.

- Ma, T. and S. Wang (2009c). Phase separation of binary systems. *Physica A: Statistical Physics and its Applications* 388:23, 4811–4817.
- Ma, T. and S. Wang (2009d). Phase transition and separation for mixture of liquid he-3 and he-4, in *lev davidovich landau and his impact on contemporary theoretical physics*, horizons in world physics, edited by A. Sakaji and I. Licata. 264, 107–119.
- Ma, T. and S. Wang (2010a). Dynamic transition and pattern formation in Taylor problem. *Chin. Ann. Math. Ser. B* 31(6), 953–974.
- Ma, T. and S. Wang (2010b). Dynamic transition theory for thermohaline circulation. *Phys. D* 239(3-4), 167–189.
- Ma, T. and S. Wang (2010c). Tropical atmospheric circulations: dynamic stability and transitions. *Discrete Contin. Dyn. Syst.* 26(4), 1399–1417.
- Ma, T. and S. Wang (2011a). Dynamic transition and pattern formation for chemotactic systems.
- Ma, T. and S. Wang (2011b). El Niño southern oscillation as sporadic oscillations between metastable states. *Advances in Atmospheric Sciences* 28:3, 612–622.
- Ma, T. and S. Wang (2011c). Phase transitions for Belousov-Zhabotinsky reactions. *Math. Methods Appl. Sci.* 34(11), 1381–1397.
- Ma, T. and S. Wang (2011d). Phase transitions for the Brusselator model. *J. Math. Phys.* 52(3), 033501, 23.
- Ma, T. and S. Wang (2011e). Third-order gas-liquid phase transition and the nature of Andrews critical point. *AIP Advances* 1, 042101.
- Ma, T. and S. Wang (2012). Unified field theory and principle of representation invariance. *arXiv:1212.4893; version 1 appeared in Applied Mathematics and Optimization*, <https://doi.org/10.1007/s00245-013-9226-0>, 69:3, pp 359–392.
- Ma, T. and S. Wang (2014a). Astrophysical dynamics and cosmology. *Journal of Mathematical Study* 47:4, 305–378.
- Ma, T. and S. Wang (2014b). Gravitational field equations and theory of dark matter and dark energy. *Discrete and Continuous Dynamical Systems, Ser. A* 34:2, 335–366; see also arXiv:1206.5078v2.
- Ma, T. and S. Wang (2015a). *Mathematical Principles of Theoretical Physics*. Science Press, 524 pages.
- Ma, T. and S. Wang (2015b). Unified field equations coupling four forces and principle of interaction dynamics. *Discrete and Continuous Dynamical Systems, Ser. A* 35:3, 1103–1138; see also arXiv:1210.0448.
- Ma, T. and S. Wang (2015c). Weakton model of elementary particles and decay mechanisms. *Electronic Journal of Theoretical Physics*, 139–178; see also Indiana University ISCAM Preprint 1304: <http://www.indiana.edu/~iscam/preprint/1304.pdf>.
- Ma, T. and S. Wang (2016a). Quantum rule of angular momentum. *AIMS Mathematics* 1:2, 137–143.
- Ma, T. and S. Wang (2016b). Spectral theory of differential operators and energy levels of subatomic particles. *J. Math. Study* 49, 259–292; see also Isaac Newton Institute Preprint NI14002: <https://www.newton.ac.uk/files/preprints/ni14002.pdf>.

- Ma, T. and S. Wang (2017a). Dynamic law of physical motion and potential-descending principle. *J. Math. Study* 50:3, 215–241; see also HAL preprint: hal-01558752.
- Ma, T. and S. Wang (2017b). Dynamical theory of thermodynamic phase transitions. *Hal preprint: hal-01568201*.
- Ma, T. and S. Wang (2017c). Radiations and potentials of four fundamental interactions. *Hal preprint: hal-01616874*.
- Ma, T. and S. Wang (2017d). Statistical theory of heat. *Hal preprint: hal-01578634*.
- Ma, T. and S. Wang (2017e). Topological Phase Transitions I: Quantum Phase Transitions. *Hal preprint: hal-01651908*.
- Ma, T. and S. Wang (2017f). Topological Phase Transitions II: Solar Surface Eruptions and Sunspots. *Hal preprint: hal-01672381*.
- Ma, T. and S. Wang (2017g). Topological Phase Transitions II: Spiral Structure of Galaxies. *Hal preprint: hal-01671178*.
- Ma, T. and S. Wang (2017h). Topological Phase Transitions IV: Dynamic Theory of Boundary-Layer Separations. *Hal preprint: hal-01672759*.
- Ma, T. and S. Wang (2019). Quantum Mechanism of Condensation and High T_c Superconductivity. *International Journal of Theoretical Physics B* 33, 1950139 (34 pages); see also hal-01613117 (2017).
- Malkus, W. V. R. and G. Veronis (1958). Finite amplitude cellular convection. *J. Fluid Mech.* 4, 225–260.
- Marsden, J. E. and M. McCracken (1976). *The Hopf bifurcation and its applications*. New York: Springer-Verlag. With contributions by P. Chernoff, G. Childs, S. Chow, J. R. Dorroh, J. Guckenheimer, L. Howard, N. Kopell, O. Lanford, J. Mallet-Paret, G. Oster, O. Ruiz, S. Schechter, D. Schmidt and S. Smale, Applied Mathematical Sciences, Vol. 19.
- Mezić, I. (2001). Break-up of invariant surfaces in action-angle-angle maps and flows. *Phys. D* 154(1-2), 51–67.
- Murray, J. (2002). *Mathematical Biology, II*. 3rd Ed. Springer-Verlag.
- Nadin, G., B. Perthame, and L. Ryzhik (2008). Traveling waves for the Keller-Segel system with Fisher birth terms. *Interfaces Free Bound.* 10(4), 517–538.
- Neelin, J. D. (1990a). A hybrid coupled general circulation model for el niño studies. *J. Atmos. Sci.* 47, 674–693.
- Neelin, J. D. (1990b). The slow sea surface temperature mode and the fast-wave limit: Analytic theory for tropical interannual oscillations and experiments in a hybrid coupled model. *J. Atmos. Sci.* 48, 584–606.
- Neelin, J. D., D. S. Battisti, A. C. Hirst, F.-F. Jin, Y. Wakata, T. Yamagata, and S. E. Zebiak (1998). Enso theory. *J. Geophys. Res.* 103, 14261–14290.
- Newton, P. K. (2001). *The N-vortex problem*, Volume 145 of *Applied Mathematical Sciences*. New York: Springer-Verlag. Analytical techniques.
- Nicolis, G. and I. Prigogine (1977). Self-organization in nonequilibrium systems. *Wiley-Interscience, New York*.
- Nirenberg, L. (1981). Variational and topological methods in nonlinear problems. *Bull. Amer. Math. Soc. (N.S.)* 4(3), 267–302.
- Nirenberg, L. (2001). *Topics in nonlinear functional analysis*, Volume 6 of *Courant Lecture Notes in Mathematics*. New York: New York University Courant Institute

- of Mathematical Sciences. Chapter 6 by E. Zehnder, Notes by R. A. Artino, Revised reprint of the 1974 original.
- Nishikawa, K. and T. Morita (1998). Fluid behavior at supercritical states studied by small-angle X-ray scattering. *Journal of Supercritical Fluid* 13, 143–148.
- Novick-Cohen, A. and L. A. Segel (1984). Nonlinear aspects of the Cahn-Hilliard equation. *Phys. D* 10(3), 277–298.
- Oertel, H. (2001). *Prandtl-Führer durch die Strömungslehre*. Vieweg+Teubner Verlag.
- Ohmi, T. and K. Machida (1998). Bose-Einstein condensation with internal degrees of freedom in alkali atom gases. *Journal of the Physical Society of Japan* 67(6), 1822–1825.
- Ong, K. W. (2016). Dynamic transitions of generalized Kuramoto-Sivashinsky equation. *Discrete & Continuous Dynamical Systems-Series B* 21(4).
- Onuki, O. (2002). Phase transition dynamics. Cambridge Univ. Press..
- Özer, S. and T. Şengül (2016). Stability and transitions of the second grade Poiseuille flow. *Physica D: Nonlinear Phenomena* 331, 71–80.
- Özer, S. and T. Şengül (2018, Jun). Transitions of spherical thermohaline circulation to multiple equilibria. *Journal of Mathematical Fluid Mechanics* 20(2), 499–515.
- Palm, E. (1975). Nonlinear thermal convection. *Annual Review of Fluid Mechanics* 7(1), 39–61.
- Pathria, R. K. and P. D. Beale (2011). *Statistical Mechanics* (third ed.). Elsevier.
- Pazy, A. (1983). *Semigroups of linear operators and applications to partial differential equations*, Volume 44 of *Applied Mathematical Sciences*. New York: Springer-Verlag.
- Pedlosky, J. (1987). *Geophysical Fluid Dynamics* (second ed.). New-York: Springer-Verlag.
- Peng, C. (2018). Attractor bifurcation and phase transition for liquid ^4He . *Acta Math. Appl. Sin. Engl. Ser.* 34(2), 318–329.
- Peres Hari, L., J. Rubinstein, and P. Sternberg (2013). Kinematic and dynamic vortices in a thin film driven by an applied current and magnetic field. *Phys. D* 261, 31–41.
- Perko, L. (1991). *Differential equations and dynamical systems*, Volume 7 of *Texts in Applied Mathematics*. New York: Springer-Verlag.
- Perthame, B. and A.-L. Dalibard (2009). Existence of solutions of the hyperbolic Keller-Segel model. *Trans. Amer. Math. Soc.* 361(5), 2319–2335.
- Perthame, B., C. Schmeiser, M. Tang, and N. Vauchelet (2011). Travelling plateaus for a hyperbolic Keller-Segel system with attraction and repulsion: existence and branching instabilities. *Nonlinearity* 24(4), 1253–1270.
- Philander, S. G. and A. Fedorov (2003). Is el niño sporadic or cyclic? *Annu. Rev. Earth Planet. Sci.* 31, 579–594.
- Phillips, N. A. (1956). The general circulation of the atmosphere: A numerical experiment. *Quart J Roy Meteorol Soc* 82, 123–164.
- Pismen, L. M. (2006). *Patterns and Interfaces in Dissipative Dynamics*. Springer, Berlin.

- Pitaevskii, L. (1961). Vortex lines in an imperfect bose gas. *Sov. Phys. JETP* 13(2), 451–454.
- Pitaevskii, L. P. and S. Stringari (2016). *Bose-Einstein condensation and superfluidity*, Volume 164 of *Internat. Ser. Mono. Phys.* Oxford: Clarendon Press.
- Prandtl, L. (1904). In *Verhandlungen des dritten internationalen Mathematiker-Kongresses*. Heidelberg, Leipeizig, pp. 484–491.
- Prigogine, I. and R. Lefever (1968). Symmetry breaking instabilities in dissipative systems. *II. J. Chem. Phys.* 48, 1695.
- Quon, C. and M. Ghil (1992). Multiple equilibria in thermosolutal convection due to salt-flux boundary conditions. *J. Fluid Mech.* 245, 449–484.
- Quon, C. and M. Ghil (1995). Multiple equilibria and stable oscillations in thermosolutal convection at small aspect ratio. *J. Fluid Mech.* 291, 33–56.
- Rabinowitz, P. H. (1968). Existence and nonuniqueness of rectangular solutions of the Bénard problem. *Arch. Rational Mech. Anal.* 29, 32–57.
- Rabinowitz, P. H. (1971). Some global results for nonlinear eigenvalue problems. *J. Functional Analysis* 7, 487–513.
- Raguin, L. G. and J. G. Georgiadis (2004). Kinematics of the stationary helical vortex mode in Taylor-Couette-Poiseuille flow. *J. Fluid Mech.* 516, 125–154.
- Rayleigh, L. (1916). On convection currents in a horizontal layer of fluid, when the higher temperature is on the under side. *Phil. Mag.* 32(6), 529–46.
- Reichl, L. E. (1998). *A modern course in statistical physics* (Second ed.). A Wiley-Interscience Publication. New York: John Wiley & Sons Inc.
- Richardson, L. F. (1922). *Weather Prediction by Numerical Process*. Cambridge University Press.
- Rogerson, A. M., P. D. Miller, L. J. Pratt, and C. K. R. T. Jones (1999). Lagrangian motion and fluid exchange in a barotropic meandering jet. *J. Phys. Oceanogr.* 29(10), 2635–2655.
- Rooth, C. (1982). Hydrology and ocean circulation. *Prog. Oceanogr.* 11, 131–149.
- Rossby, C.-G. (1926). On the solution of problems of atmospheric motion by means of model experiment. *Mon. Wea. Rev.* 54, 237–240.
- Rubin, V. and J. W. K. Ford (1970). Rotation of the andromeda nebula from a spectroscopic survey of emission regions. *Astrophysical Journal* 159, 379–404.
- Salby, M. L. (1996). *Fundamentals of Atmospheric Physics*. Academic Press.
- Samelson, R. M. (2008). Time-periodic flows in geophysical and classical fluid dynamics. in: *Handbook of numerical analysis, special volume on computational methods for the ocean and the atmosphere*. R. Temam and J. Tribbia, eds. Elsevier, New York. To appear.
- Samelson, R. M. and S. Wiggins (2006). *Lagrangian transport in geophysical jets and waves*, Volume 31 of *Interdisciplinary Applied Mathematics*. New York: Springer. The dynamical systems approach.
- Sardeshmukh, P. D., G. P. Compo, and C. Penland (2000). Changes of probability associated with el niño. *Journal of Climate*, 4268–4286.
- Sattinger, D. H. (1978). Group representation theory, bifurcation theory and pattern formation. *J. Funct. Anal.* 28(1), 58–101.

- Sattinger, D. H. (1979). *Group-theoretic methods in bifurcation theory*, Volume 762 of *Lecture Notes in Mathematics*. Berlin: Springer. With an appendix entitled “How to find the symmetry group of a differential equation” by Peter Oliver.
- Sattinger, D. H. (1980). Bifurcation and symmetry breaking in applied mathematics. *Bull. Amer. Math. Soc. (N.S.)* 3(2), 779–819.
- Sattinger, D. H. (1983). *Branching in the presence of symmetry*, Volume 40 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM).
- Schmid, A. (1966). A time dependent Ginzburg-Landau equation and its application to the problem of resistivity in the mixed state. *Phys. Kondens. Mater.* 5, 302–317.
- Schneider, E. K., B. P. Kirtman, D. G. DeWitt, A. Rosati, L. Ji, and J. J. Tribbia (2003). Retrospective ENSO forecasts: Sensitivity to atmospheric model and ocean resolution. *Monthly Weather Review* 131:12, 3038–3060.
- Schopf, P. S. and M. J. Suarez (1987). Vacillations in a coupled ocean-atmosphere model. *J. Atmos. Sci.* 45, 549–566.
- Sengers, J. V. and G. S. Shanks (2009). Experimental critical-exponent values for fluids. *J Stat Phys* 137, 857–877.
- Sengul, T., J. Shen, and S. Wang (2015). Pattern formations of 2d Rayleigh–Bénard convection with no-slip boundary conditions for the velocity at the critical length scales. *Mathematical Methods in the Applied Sciences* 38(17), 3792–3806.
- Sengul, T. and S. Wang (2011). Pattern formation and dynamic transition for magnetohydrodynamic convection.
- Sengul, T. and S. Wang (2013). Pattern formation in Rayleigh–Bénard convection. *Communications in Mathematical Sciences* 11(1), 315–343.
- Sengul, T. and S. Wang (2014). Pattern formation and dynamic transition for magnetohydrodynamic convection. *Communications on Pure & Applied Analysis* 13(6), 2609–2639.
- Şengül, T. and S. Wang (2018). Dynamic transitions and baroclinic instability for 3d continuously stratified Boussinesq flows. *Journal of Mathematical Fluid Mechanics*, 1–21.
- Shen, J. and X. Yang (2010). Numerical approximations of Allen-Cahn and Cahn-Hilliard equations. *Discrete Contin. Dyn. Syst.* 28(4), 1669–1691.
- Shi, J. (2009). Bifurcation in infinite dimensional spaces and applications in spatiotemporal biological and chemical models. *Front. Math. China* 4(3), 407–424.
- Smoller, J. (1983). *Shock waves and reaction-diffusion equations*, Volume 258 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science]*. New York: Springer-Verlag.
- Stanley, H. E. (1971). *Introduction to Phase Transitions and Critical Phenomena*. Oxford University Press, New York and Oxford.
- Stern, M. E. (1960). The “salt fountain” and thermohaline convection. *Tellus* 12, 172–175.
- Stommel, H. (1961). Thermohaline convection with two stable regimes of flow. *Tellus* 13, 224–230.
- Surana, A., O. Grunberg, and G. Haller (2006). Exact theory of three-dimensional flow separation. I. Steady separation. *J. Fluid Mech.* 564, 57–103.

- Swinney, H. L. (1988). Instabilities and chaos in rotating fluids. In *Nonlinear evolution and chaotic phenomena (Noto, 1987)*, Volume 176 of *NATO Adv. Sci. Inst. Ser. B Phys.*, pp. 319–326. New York: Plenum.
- Swinney, H. L., N. Kreisberg, W. D. McCormick, Z. Noszticzius, and G. Skinner (1990). Spatiotemporal patterns in reaction-diffusion systems. In *Chaos (Woods Hole, MA, 1989)*, pp. 197–204. New York: Amer. Inst. Phys.
- Tang, Q. and S. Wang (1995). Time dependent Ginzburg-Landau equations of superconductivity. *Phys. D* 88(3-4), 139–166.
- Taylor, G. I. (1923). Stability of a viscous liquid contained between two rotating cylinders. *Philos. Trans. Royl London Ser. A* 223, 289–243.
- Temam, R. (1984). *Navier-Stokes Equations, Theory and Numerical Analysis*, 3rd, rev. ed. North Holland, Amsterdam.
- Temam, R. (1997). *Infinite-dimensional dynamical systems in mechanics and physics* (Second ed.), Volume 68 of *Applied Mathematical Sciences*. New York: Springer-Verlag.
- Thual, O. and J. C. McWilliams (1992). The catastrophe structure of thermohaline convection in a two-dimensional fluid model and a comparison with low-order box models. *Geophys. Astrophys. Fluid Dyn.* 64, 67–95.
- Tinkham, M. (1996). *Introduction to Superconductivity*. McGraw-Hill, Inc.
- Tziperman, E. (1997). Inherently unstable climate behavior due to weak thermohaline ocean circulation. *Nature* 386, 592–595.
- Tziperman, E., J. R. Toggweiler, Y. Feliks, and K. Bryan (1994). Instability of the thermohaline circulation with respect to mixed boundary conditions: Is it really a problem for realistic models? *J. Phys. Oceanogr.* 24, 217–232.
- Tzou, J. C., B. J. Matkowsky, and V. A. Volpert (2009). Interaction of Turing and Hopf modes in the superdiffusive Brusselator model. *Appl. Math. Lett.* 22(9), 1432–1437.
- Velte, W. (1966). Stabilität und verzweigung stationärer lösungen der davier-stokeschen gleichungen. *Arch. Rat. Mech. Anal.* 22, 1–14.
- Veronis, G. (1963). An analysis of wind-driven ocean circulation with a limited Fourier components. *J. Atmos. Sci.* 20, 577–593.
- Veronis, G. (1966). Wind-driven ocean circulation, part ii: Numerical solution of the nonlinear problem. *Deep-Sea Res.* 13, 31–55.
- Vishik, I. M. (2012). Phase competition in trisection superconducting dome. *Proc. Natl Acad. Sci. USA* 109, 18332–18337.
- Vojta, M. (2003). Quantum phase transitions. *Reports on Progress in Physics* 66(12), 2069.
- von Neumann, J. (1960). Some remarks on the problem of forecasting climatic fluctuations. In R. L. Pfeffer (Ed.), *Dynamics of climate*, pp. 9–12. Pergamon Press.
- Wang, Q. (2014). Stability and bifurcation of a viscous incompressible plasma fluid contained between two concentric rotating cylinders. *Discrete & Continuous Dynamical Systems-Series B* 19(2).
- Wang, Q., H. Luo, and T. Ma (2015a). Boundary layer separation of 2-D incompressible Dirichlet flows. *Discrete Contin. Dyn. Syst. Ser. B* 20(2), 675–682.

- Wang, Q., H. Luo, and T. Ma (2015b). Boundary layer separation of 2-d incompressible Dirichlet flows. *Discrete Contin. Dyn. Syst. Ser. B* 20, 675–682.
- Wang, Q. and H. Wang (2016). The dynamical mechanism of jets for AGN. *Discrete Contin. Dyn. Syst. Ser. B* 21(3), 943–957.
- Wang, S. and P. Yang (2013). Remarks on the Rayleigh-Benard convection on spherical shells. *Journal of Mathematical Fluid Mechanics* 15(3), 537–552.
- Wiggins, S. (1990). *Introduction to applied nonlinear dynamical systems and chaos*, Volume 2 of *Texts in Applied Mathematics*. New York: Springer-Verlag.
- Yadome, M., Y. Nishiura, and T. Teramoto (2014). Robust pulse generators in an excitable medium with jump-type heterogeneity. *SIAM J. Appl. Dyn. Syst.* 13(3), 1168–1201.
- Yang, C. N. and R. Mills (1954). Conservation of isotopic spin and isotopic gauge invariance. *Physical Review* 96, 191–195.
- Yang, J. and R. Liu (2016). A fluid dynamical model coupling heat with application to interior separations. <https://arxiv.org/pdf/1606.07152.pdfV>.
- Yarahmadian, S. and M. Yari (2014). Phase transition analysis of the dynamic instability of microtubules. *Nonlinearity* 27(9), 2165.
- Yari, M. (2015). Transition of patterns in the cell-chemotaxis system with proliferation source. *Nonlinear Anal.* 117, 124–132.
- You, H., R. Yuan, and Z. Zhang (2013). Attractor bifurcation for extended Fisher-Kolmogorov equation. *Abstr. Appl. Anal.*, Art. ID 365436, 11.
- Yudovich, V. I. (1966). Secondary flows and fluid instability between rotating cylinders. *Appl. Math. Mech.* 30, 822–833.
- Yudovich, V. I. (1967a). Free convection and bifurcation. *J. Appl. Math. Mech.* 31, 103–114.
- Yudovich, V. I. (1967b). Stability of convection flows. *J. Appl. Math. Mech.* 31, 272–281.
- Zebiak, S. E. and M. A. Cane (1987). A model el nino southern oscillation. *Mon. Wea. Rev.* 115, 2262–2278.
- Zhabotinski, A. (1964). Periodic process of the oxidation of malonic acid in solution (study of the kinetics of Belousov's reaction). *Biofizika* 9, 306–311.
- Zhang, D. and R. Liu (2018). Dynamical transition for s-k-t biological competing model with cross-diffusion. *Mathematical Methods in the Applied Sciences* 41(12), 4641–4658.
- Zhang, H., K. Jiang, and P. Zhang (2014). Dynamic transitions for Landau-Brazovskii model. *Discrete & Continuous Dynamical Systems-Series B* 19(2).
- Zhang, Q. and H. Luo (2013). Attractor bifurcation for the extended Fisher-Kolmogorov equation with periodic boundary condition. *Bound. Value Probl.*, 2013:169, 13.

Index

- L*-constraint variation, 5
 $SO(3)$ spinor representations, 583
 $SO(3)$ spinors, 582
 $SO(N)$ symmetry, 580
 S^1 -attractor bifurcation, 41
 S^m -attractor bifurcation, 40, 48
- μ**
induction magnaton, 632
- ^4He , 237
- k -th order nondegenerate singularity, 85, 96
dynamic transition of, 99
index, 85
index formula, 86
- A**
Abrikosov vortex-Meissner transition, 647, 654
Abrikosov vortices, 235
Albert Einstein, viii, 2
Anderson-Brinkman-Morel state, 249
Andrews critical point, 26
locus, 168
Angular momentum rule, 556
Anti-diffusive effect of heat, 676, 687
Asymmetry of fluctuation, 183
Asymmetry of Fluctuations, 28
Attractor bifurcation, 40
infinite dimensional case, 50
- B**
Balian-Werthamer state, 249
Belousov-Zhabotinsky reaction, 472
Binary system, 28, 183
Boltzmann equation, 536
aims of, 537
problems in, 536
Born rule, 548
- Bose-Einstein distribution, 564
Boundary-layer separation, 24, 688
dynamic theory of, 688
phenomena, 688
predicable condition, 695
separation equation, 692, 693
theorem, 691
- Boussinesq equation, 11
Boussinesq equations, 8, 12
Bragg-Williams potential, 28, 184
BZ reaction, 472
- C**
Cahn-Hilliard equation, 28, 534
with Bragg-Williams potential, 184
Center manifold function, xxiii
approximation, 721
Center manifold reduction, 18, 719
Chaotic state, 20
Chemical potential, 533
is the Lagrangian multiplier, 534
physical property of, 534
Chemical reaction-diffusion model, 469
Chemotaxis, 502
Keller-Segel model, 503
Classification of dynamic transition, 16
Climate dynamics, 406
Completely continuous field, 31
Condensates, 597
field equations, 599, 636
quantum mechanism of, 597
quantum rules, 616
quantum system of, 629
thermodynamic potentials, 616
topological index, 649
topological structure equation, 600

- Constraint variation, 5, 732
 Convection scales, 407
 Coronal mass ejection, 683
 Coupling, ix, 3
 Critical exponents, 131, 149
 - definition of, 152
 - first theorem of, 157
 - second theorem of, 163
 Critical parameter equation, 15
 Critical parameters, 149
 Critical temperature
 - based on microscopic theory, 614
 - based on standard model, 612
 - for superconductivity, 612
 - formula for, 612
 Curie temperature, 173
- D**
- De Broglie relations, 547
 Dielectric system, 594
 Dynamic classification scheme, 16
 Dynamic law of motion, 4, 5
 Dynamic transition, 32
 - classification, 16
 - continuous, 33, 40
 - definition of, 32
 - dynamic classification, 31
 - dynamic classification scheme, 16
 - ferromagnetism, 177
 - from k -th order nondegenerate singularity, 99
 - from complex simple eigenvalue, 55
 - from eigenvalue with multiplicity two, 63
 - from simple eigenvalue, 52, 114
 - general features, 19
 - jump, 33
 - mixed, 33
 - stability, 48
 - superconductivity, 220
 - vs transition in physical space, 24
 Dynamic transition theory, 31
 - dynamic classification scheme, 16
 - key philosophy, xx

E

Ehrenfest classification, 135
 El Niño Southern Oscillation, 22, 408

 - abbreviated as ENSO, 22
 - El Niño phase, 22
 - La Niña phase, 22
 Electron

 - energy levels, 560
 - photon absorption and radiation of, 557
 - photon cloud of, 555
 PID interaction potential, 604
 Electron-pair, 606

 - binding energy, 610
 Elliptic region, 37
 Energy level formula

 - of temperature, 566
 Energy level temperature formula, 566, 676

 - physical meaning of, 571
 Energy levels, 558

 - of atom, 561
 - of electron, 560
 - of photons, 558
 Energy operator, 548
 Enthalpy system, 545
 Entropy, 572

 - photon number formula of, 574
 - physical meaning of, 572
 Entropy-ascending principle, 537
 Equilibrium phase transition, 129
 Equilibrium system, 562

 - statistics of, 562
 Essence of physics, ix, 3

 - coupling, ix, 3

F

Fermi statistics, 248
 Fermi-Dirac distribution, 565
 Ferromagnetism, 173

 - classical theory of, 173
 - magnetic susceptibility, 173
 Fick's law of diffusion, 468
 Field-Körös-Noyes equation, 472
 Fluctuation, 141

 - dynamic law of, 144
 - dynamic theory of, 141
 Fluctuation theory, 131
 Fluid dynamics, 7, 10

G

Galactic dynamics

 - model for, 666
 Galactic spiral structure, 659

 - theory of, 670
 General Relativity, viii, 2
 Geophysical fluid dynamics, 406
 Gibbs free energy

 - for BEC, 625
 - for gaseous condensates, 625
 Gibbs system, 546
 Ginzburg-Landau free energy, 211
 Global attractor

 - gradient-type system, 726
 Gradient-type system, 725

 - global attractor, 726

- Gravitation radiation, 662
Gravitational field equations, 664
Graviton thermodynamics, 664
Guiding principle of physics, 1
- H**
Hamiltonian dynamics, 9
Hamiltonian energy
 of superconductor, 631
Heat
 caloric theory of, 579
 essence of, 575
 statistical theory of, 566
Heisenberg uncertainty relation, 549
Helmholtz system, 546
Herman Weyl, 4
Hexagonal convection cells, 10
High Tc superconductivity, 597
 quantum mechanism of, 597
 theory of, 602
Hyperbolic region, 37
- I**
Incompressible flows, 639
Index formula, 63, 236
Induction magneton, 632
Interior separation, 643
 theory of, 706
 U-flow pattern for, 711
Interior separation equation, 708
Internal energy system, 544
Irreversibility, 535
 PDP is the principle of, 536
- J**
Josephson superconductor-insulator transition, 646
- K**
Keller-Segel model, 503
- L**
Lagrangian dynamics, 9
Law of entropy transfer, 573
Law of heat transfer, 676
Law of temperature, 575
Legendre transform, 542
 convexity condition, 543
Liquid ^3He A–B phase transitions, 648
Liquid helium-3, 248
 dynamic model, 250
Liquid helium-4, 237
 dynamic model, 238
 Ginzburg-Landau free energy, 238
- phase diagram, 242
Liquid-helium-4
 metastable region, 245
- M**
Magnetic system, 592
Mass generation mechanism, 554
Maxwell distribution, 537
Maxwell-Boltzmann statistics, 562
Meissner effect, 211, 652
 microscopic mechanism, 652
Metastable domain, 33
Metastable state, 20
Mixture of liquid helium-4 and helium-3, 269
Momentum operator, 548
Mott superfluid-insulator transition, 647
- N**
N-component system, 590
Navier-Stokes equations, 112, 406
Newtonian law, 7
North Atlantic subpolar gyre, 704
- O**
Observable, 548
Oceanic circulation, 701
Orthogonal decomposition theorem, 732
- P**
Parabolic region, 37
Partition function, 564
Pauli exclusion principle, 549
PDP, ix, 129, 132, 532
PEP, 129, 532
Periodic orbit, 102
Perturbed system, 113
 with complex eigenvalue, 119
PES, 32, 137
PEs, 407
Phase separation, 183
 binary system, 183
Phase transition
 dynamic classification scheme, 34
 Type-I, 17
 Type-II, 17
 Type-III, 17
Photon absorption and radiation, 557
 of electron, 557
Photon absorption and radiation mechanism, 676
Photon cloud
 of electron, 555
Photon cloud model, 555
Photon number entropy formula, 574, 676

- Physical-vapor transport system, 26, 164
 PID, 4
 PID interaction potential
 electron, 604
 Plasma fuel, 687
 Poincaré index formula, 37
 Population model of biological species, 470
 Position operator, 548
 Potential-descending principle, ix, 129, 132, 532
 leads to 1st and 2nd laws, 541
 Primitive equations, 407
 Principle of equal probability, 129, 532
 Principle of exchange of stability, 14, 32, 137
 abbreviated as PES, 14, 15
 Principle of general relativity, viii, 2
 Principle of interaction dynamics (PID), 4
 Principle of symmetry-breaking, 3
 Propagating Taylor vortices (PTV), 25
 PSB, 3
 PVT system, 164
 dynamic model, 166
 Gibbs energy for, 165
 phase transition for, 167
 time-dependent model, 165
- Q**
 QM wave function
 new field theoretic interpretation, 550
 QPT, 644
 physical characterizations, 646
 Quantum condensates, 649
 Quantum phase transition
 abbreviated as QPT, 644
 Quantum rules, 547
 Quantum system, 564
 of condensates, 629
 Quasi-geostrophic equation, 407
- R**
 Raleigh number, 281
 Rayleigh–Bénard convection, 280
 Rayleigh–Bénard convection, 10
 Reaction process, 468
- S**
 Schrödinger equation, 547
 Screening of Coulomb repulsion, 604
 Second-order nondegenerate singularity, 63
 index formula, 63, 64
 Singular separation, 107
 of periodic orbit, 111
 Solar eruption, 674
 Solar eruptions
 period of, 684
 Solar flare, 675, 683
 Solar prominence, 675, 684
 Spin interaction energy, 630
 Spiral arm pattern, 672
 Spiral pattern formation, 661
 Standard model, 132, 133
 critical exponents of, 152
 of statistical physics, 132
 of thermodynamic system, 533
 Statistical theory of heat, 566
 Structural bifurcation, 642
 Structural stability theorem, 640
 Symmetry, 4
 Sunspots, 674, 676, 682
 Sunspots and solar eruption
 formation of, 679
 Superconductivity, 597
 Abrikosov vortices, 235
 BCS theory, 603
 critical length scale, 235
 critical magnetic strength, 234
 field equations, 636
 free energy, 618
 Ginzburg–Landau free energy, 212
 physical mechanism, 602
 quantum mechanism of, 597, 599
 TDGL, 213
 TDGL coupled with entropy, 226
 Supercooled states, 140
 Supercurrent, 601
 Superfluid system, 633
 Superfluidity, 237, 248, 599
 quantum mechanism of, 599
 Superheated states, 140
- T**
 Taylor-Couette (TC) flow, 25
 Taylor-Couette-Poiseuille (TCP) flow, 25
 Temperature
 energy level formula of, 566
 law of, 575
 Theory of Andrews critical point, 27
 Thermal energy, 576
 Thermodynamic phase transition, 132, 136
 first theorem of, 136
 second theorem of, 137
 third theorem of, 140
 Thermodynamic potentials, 542, 579
 basic form of, 582
 dielectric system, 594
 for condensates, 616

- for liquid ^3He superfluidity, 622
for liquid ^4He , 620
magnetic system, 592
N-component systems, 590
PVT systems, 586
- Thermohaline circulation, 428
- Time-dependent Ginzburg-Landau equation, 213
- Topological index
of condensates, 649
- Topological phase transition, 639
abbreviated as TPT, 639
- Topological structure equation
condensates, 600
- Tornado and hurricane, 715
- TPT
topological phase transition, 639
- Transition
from an attractor, 32
from an equilibrium, 32
- Transition in physical space, 24
- Turbulent friction, 407
- V**
- Van der Waals equation, 165
- Variation with constraint-infinities, 734
- W**
- Walker circulation, 22, 409
- Wave function, 548
- Weakton model, 553
of elementary particles, 553
- Wind-driven north Atlantic gyre, 702
- Y**
- Yang-Mills functional, 4