OXFORD

MATHEMATICAL PHYSICS

with DIFFERENTIAL EQUATIONS

YISONG YANG

Mathematical Physics with Differential Equations

Mathematical Physics with Differential Equations

Yisong Yang
Professor, Courant Institute of Mathematical Sciences,
New York University

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© Yisong Yang 2023

The moral rights of the author have been asserted

All rights reserved. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted, in any form or by any means, without the
prior permission in writing of Oxford University Press, or as expressly permitted
by law, by licence or under terms agreed with the appropriate reprographics
rights organization. Enquiries concerning reproduction outside the scope of the
above should be sent to the Rights Department, Oxford University Press, at the
address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2022943690

ISBN 978–0–19–287261–6
ISBN 978–0–19–287262–3 (pbk.)

DOI: 10.1093/oso/9780192872616.001.0001

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

For Sheng,
Peter, Anna, and Julia

Contents

Preface

This book aims to present a broad range of fundamental topics in theoretical and mathematical physics in a thorough and transparent manner based on the viewpoint of differential equations. The subject areas covered include classical and quantum many-body problems, thermodynamics, electromagnetism, magnetic monopoles, special relativity, gauge field theories, general relativity, superconductivity, vortices and other topological solitons, and canonical quantization of fields, for which, differential equations are essential for comprehension and have played, and will continue to play important roles. Over the past decade, the author has used most of these topics at several universities, domestically and internationally, as courses and seminars mainly for mathematical graduate students and researchers trained and interested in differential equations. These activities and experiences convinced the author that many of the concepts, construction, structures, ideas, and insights of fundamental physics can be taught and learned effectively and productively, with emphasis on what are offered by or demanded from differential equations.

With this in mind, the book has several goals to accomplish. Firstly, the style of the presentation hopefully provides a handy and direct access to approach the subjects discussed. Secondly, it serves to render a fairly wide selection of themes that may further be tailored for a graduate-level mathematical physics curriculum out of the individual preference of the instructor or reader. Thirdly, it supplies a balanced pool of topics for upper-level or honors undergraduate seminars. Fourthly, it offers guidance and stimulation to the related contemporary research frontiers and literature.

Except for knowledge on differential equations, the prerequisite for the reader of the book is kept minimal, although certain levels of acquaintance with undergraduate general physics is helpful for the reader to proceed smoothly. Thus, the book begins with classical mechanics in canonical formalism and moves on to various advanced subjects. However, unless needed, the book excludes specialized topics of traditional classical mechanics such as fluids and elasticity theories, since they are treated extensively elsewhere in the literature. The book may be used for self-study, as a textbook, or as a supplemental source book for a course in mathematical physics with concentration and interests in quantum mechanics, field theory, and general relativity, emphasizing insights from differential equations.

While the book holds fifteen chapters, each chapter may be studied or presented separately in a more or less self-contained manner, depending on interests and readiness of the reader or audience.

In Chapter 1, we start with a presentation of the canonical formalism of classical mechanics. We then consider the classical many-body problems in three-, two-, and one-dimensional settings, subsequently. Specifically, in three dimensions, we discuss the many-body problem governed by Newton's gravity, consolidated by a thorough study of Kepler's laws of planetary motion and a derivation of Newton's law of gravitation, as a by-product; in two dimensions, we introduce the Helmholtz–Kirchhoff point-vortex model; and, in one dimension, we present a dynamical system problem in biophysics known as the DNA denaturation. For this third subject we also explain how to implement ideas of thermodynamics to study a temperature-dependent mechanical system. The goal of this chapter is to lay a Lagrangian field-theoretical foundation for field theory and enlighten the study with some exemplary applications.

In Chapter 2, we consider quantum many-body problems. We explain how the Schrödinger equation is conceptualized and the statistical interpretation of the wave function. Then, we formulate the quantum many-body problem describing an atomic system and discuss the hydrogen model as an illustration. Next, we show how the Hartree–Fock method, Thomas–Fermi approach, and density functional theory may be utilized in various situations as computational tools to find the ground state solution of a quantum many-body problem. An initial goal of this chapter is to illustrate a monumental transition from classical to quantum mechanics based on the Schrödinger equation realization of the photoelectric effect. A second goal of this chapter is to introduce some mathematical challenges presented by quantum many-body problems. Here the study of the hydrogen model serves as a motivating starting point of the quantum many-body problem, which naturally leads to the development of subsequent analytic methods of computational significance when the dimension of the problem goes up. In particular, we show that the quantum-mechanical description of a many-body problem, whose classical-mechanical behavior is governed by nonlinear ordinary differential equations, is now given by a linear partial differential equation, and that appropriate approximations of such a linear equation necessitate the formulation of various nonlinear equation problems in respectively specialized situations.

Chapter 3 is a study of the Maxwell equations and some distinguished consequences. First, we present the equations and discuss the associated electromagnetic duality phenomenon. We next formulate the Dirac monopole and Dirac strings and show how to use a gauge field to resolve the Dirac string puzzle and obtain Dirac's charge quantization formula. We demonstrate how this idea inspired Schwinger to derive a generalized quantization formula for a particle carrying both electric and magnetic charges, known as dyon. We then present the Aharonov–Bohm effect for wave interference, which demonstrates the significant roles played by the gauge field and topology of a system at quantum level. The goal of this chapter is to appreciate how some of the fundamental and rich contents of electromagnetic interaction may be investigated productively through exploring the structures of the differential equations governing the interaction.

Specifically, by complexifying the wave equation, we obtain a novel derivation of the Maxwell equations, which also embodies a clear and natural revelation of the electromagnetic duality, and, by considering the topological properties of the solutions to the Maxwell equations, we arrive at the findings of Dirac, Schwinger, and Aharonov–Bohm.

Chapter 4 is a succinct introduction to special relativity. Since most of the subjects covered in this text are concerned with relativistic field equations, some solid knowledge on special relativity is necessary. Thus we carry out a study of special relativity in this chapter. We first discuss spacetime, inertial frames, and the Lorentz transformations. We present topics that include spacetime line element, proper time, and a series of notions, including length contraction, time dilation, and simultaneity of events. Then, we study relativistic mechanics. Although this chapter is short, its goal is to serve as the foundation for many following chapters, including those on the Dirac equations, gauge field theory, general relativity, and topological solitons. In particular, the Lagrangian action for the motion of a relativistic particle will be the starting point for the formulation of the Nambu–Goto string action and the Born–Infeld theory.

In Chapter 5, we present the Abelian gauge field theory. We start with an introduction to the notions of covariance, contravariance, and invariance for quantities defined over a spacetime. We formulate the Klein–Gordon equation, which is a relativistic extension of the Schrödinger equation. We then show that a gauge field is brought up again naturally in order to promote the internal symmetry of the system from global to local such that the Maxwell equations are deduced as a consequence. Furthermore, we discuss various concepts of symmetry breaking and illustrate the ideas of the Higgs mechanism as another important application of gauge field equations. A goal achieved in this chapter is that a vista of important physical consequences may be obtained from examining some basic structural aspects of the equations of motion without sufficient knowledge about their solutions.

Chapter 6 centers around the Dirac equation. We first show how to obtain the classical Dirac equation and what immediate consequences the equation offers in contrast to the Klein–Gordon equation. We next consider the Dirac equation coupled with a gauge field and present its Schrödinger equation approximations in electrostatic and magnetostatic limits, respectively. In particular, we derive the Stern–Gerlach term, whose presence is essential for the explanation of the Zeeman effect. We then review some nonlinear Dirac equations. This study shows that sometimes profound physics may be unveiled unexpectedly through an exploration of some deeply hidden internal structures of the governing equations.

Chapter 7 covers the Ginzburg–Landau theory for superconductivity. We begin with a discussion of perfect conductors, superconductors, the Meissner effect, the London equations, and the Pippard equation. We next present the Ginzburg–Landau equations for superconductivity and show how to come up with the London equations in the uniform order-parameter limit and demonstrate the Meissner effect. We then study the classification of superconductivity in view of surface energy and discuss the appearance of mixed states in type II superconductors. We end the chapter with a review of some generalized

Ginzburg–Landau equations. This study retraces the historical path regarding how differential equations of varied subtleties have been exploited in line with real-world observations to advance the understanding of superconductivity. In particular, it also describes an unsolved two-point boundary value problem, arising in the Ginzburg–Landau theory, for the classification of superconductivity.

Chapter 8 grows out of the subjects covered in Chapter 5 and Chapter 7. Specifically, in this chapter, we focus on the static Abelian Higgs theory or the Ginzburg–Landau theory in two dimensions, which possesses a distinctive class of mixed-state solutions of a topological characteristic known as vortices. We describe such solutions in detail in view of several important facets including energy concentration, vortex-line distribution, quantization of magnetic flux or charge, and exponential decay properties. We also discuss the use of such vortex-line solutions in a linear confinement mechanism for magnetic monopoles, a topic actively pursued in quark confinement research in recent years. This study shows again the applications of solutions of gauge field equations, of topological characteristics, to fundamental physics, of both quantitative and conceptual values.

In Chapter 9, we move onto the subject of non-Abelian gauge field theory. We first present the theory on a general level, and then specialize on the Yang–Mills–Higgs theory. We discuss a series of concrete formalisms including the Georgi–Glashow model and the Weinberg–Salam electroweak theory. We also illustrate some important families of solutions such as the 't Hooft–Polyakov monopole, Julia–Zee dyon, and Bogomol'nyi–Prasad–Sommerfield explicit solution. The main goal of this chapter is to present a broad family of nonlinear partial differential equations of importance in elementary particle physics.

In Chapter 10, we study the Einstein equations of general relativity and related subjects. We begin with an introduction to the basics of Riemannian geometry and then present the Einstein tensor and the Einstein equations for gravitation. Subsequently, we unfold our discussion mainly around special solutions of the Einstein equations, categorized into time-dependent space-uniform solutions and time-independent space-symmetric solutions. In the former category, we elaborate on the cosmological consequences and implications richly contained in various solutions of the Friedmann type equations under the Robertson–Walker metric, which include the Big Bang cosmological scenario, patterns of expansion of the universe, and an estimate of the age of the universe. In the latter category, we begin with a presentation of the Schwarzschild solution and a discussion of several notions unveiled, such as the event horizon and black hole. We then present a derivation of the Reissner–Nordström solution for a black hole carrying both electric and magnetic charges and discuss its consequences. We will also discuss the Kerr solution describing a rotating black hole. Afterwards, we consider the gravitational mass problem and the Penrose bounds as additional themes. We next present a discussion of gravitational waves in the weak-field limit. We conclude the chapter with a study of the cosmological expansion of an isotropic and homogeneous universe propelled by a scalar-wave

matter known as quintessence. The main goal of this chapter is to use the Einstein equations as a key to access a broad range of gravity-related research directions of contemporary interests.

Chapter 11 is about charged vortices and the Chern–Simons equations. For conciseness, we focus on the simplest Abelian situations. We first present the Julia–Zee theorem and its proof, which states that finite-energy electrically charged vortices, which are static solutions in two dimensions, do not exist in the usual Yang–Mills–Higgs theory. Thus, some modification of the theory is to be made in order to accommodate charged vortices, and the addition of a Chern–Simons topological term to the Lagrangian action density will serve the purpose. In this chapter, our goal is to present a brief introduction to the Chern–Simons vortex equations. Besides the motivation for allowing electrically charged vortices, other applications of the Chern–Simons theory include anyon physics of condensed matters, gravity theory, and high-temperature superconductivity, where non-Abelian structures are also abundantly utilized. It is hoped that this introduction will serve to spark interest and inspiration in the study of an enormous family of partial differential equation problems of challenges, under the shared title of the Chern–Simons vortex equations.

In Chapter 12, we consider the Skyrme model and some related topics. We begin with an exploration of the well-known dimensionality constraints brought forth by the Derrick theorem and the Pohozaev identity. We then introduce the Skyrme model to maneuver around the dimensionality constraints. As a related topic, we will also discuss the Faddeev model, which may be viewed as a descent of the Skyrme model and which brings about knot-like solutions characterized by fractionally-powered growth laws relating energy to topology. In addition, we present a discussion of Coleman's Q-ball model, which also has no dimensionality restriction. Due to the difficulties associated with the structures of the nonlinearities and topological characteristics of these field-theoretical models, it has been a daunting task to consider the equations of motion directly, except in numerical studies, and one needs to focus on their variational solutions. In either situation, hopefully this chapter serves as an invitation to many related research topics.

Chapter 13 is a short discussion on strings and branes. We first revisit the relativistic motion of a free particle and subsequently formulate the Nambu–Goto string equations. We then extend the study to consider branes and their governing equations. We next present the Polyakov strings and branes and their equations of motion. Thus, our goal of this chapter is to emphasize the challenges and difficulties encountered in these highly geometric and nonlinear partial differential equations as classical field-theory equations. Except in some extremely simplified or reduced limits, these equations are not yet well understood, regarding their solutions.

In Chapter 14, we present the Born–Infeld theory of electromagnetism and some of the associated mathematical problems. To start, we recall the energy divergence problem of the point-charge model of the electron and the idea of Born and Infeld in tackling the problem based on a revision of the action density motivated by special relativity, sometimes referred to as the first formulation of

Born and Infeld. Within this formalism, we consider some interesting illustrative calculations around the electric and dyonic point charge problems. We next present the second formulation of Born–Infeld based on invariance consideration and show how to resolve the energy divergence problem associated with a dyonic point charge encountered in the first formulation of Born and Infeld. We then relate the Born–Infeld equations to the minimal surface equations and propose some generalized Bernstein problems. Subsequently, we conduct a discussion of an integer-squared law of a universal nature regarding the global vortex solutions of the Born–Infeld equations in two dimensions. Furthermore, we also present a series of electrically and dyonically charged black hole solutions of the Einstein equations coupled with the Born–Infeld equations. Thereafter, we consider the generalized Born–Infeld theories and present some interesting applications, including a nonlinear mechanism for an exclusion of monopoles as finite-energy magnetically charged point particles, relegation and removal of curvature singularities of charged black holes of the Reissner–Nordström type, and theoretical realizations of cosmological expansion and equations of state of cosmological fluids through appropriate Born–Infeld scalar-wave matters in the form of k-essence. In some sense, this chapter may be regarded as a gauge-field or scalar-field extension of the subjects discussed in Chapter 13. Therefore, the difficulties we encounter here are similar to those there. On the other hand, within the limitation of the Born–Infeld theory, here we are able to see how real progress is made for many important issues of concern, such as the resolution of an electric point charge of divergent energy, electromagnetic asymmetry, singularity relegation for charged black holes, and k-essence scalar field cosmology, all based on pursuing special solutions of the governing equations in various situations.

Chapter 15 is the final chapter and provides some taste of field quantization and a further view expansion. For clarity and conciseness, our discussion will be clustered around harmonic oscillators. We start with a study of the quantum mechanics of harmonic oscillators based on canonical quantization. We next consider the Hamiltonian formalism of general field equations in terms of functional derivative and commutators. We then show how to quantize the Klein–Gordon equation and the Schrödinger equation. In doing so, we encounter the well-known infinity problem arising from a divergent zero-point energy, which gives us an opportunity to explain the concept of renormalization. We then move on to quantize the Maxwell equations that govern electromagnetic fields propagating in free space. We focus our attention on the quantization of energy, momentum, and spin angular momentum directly, rather than the electromagnetic fields themselves, and derive the Planck–Einstein and Compton–Debye formulas for the photoelectric effect and photon spin in the context of quantum field theory. We conclude the chapter with a discussion of the thermodynamics of a harmonic oscillator, both classically and quantum mechanically, such that we are able to come up with a picture about the relation, ranges of applicability and limitation, and transition with regard to temperature, of classical and quantum-mechanical descriptions of a physical system, in general. Thus, part of the goal of this chapter is to show in view of quantum field theory

what may be expected beyond classical field equations both in sense of differential equations and meaning of quantum physics.

Exercises appear at the end of each chapter. These mostly straightforward problems serve either to supplement the details or expand the scope of the materials of the text. Working out some of the problems may be useful for checking the understanding of the subjects covered but omitting this process should not compromise the quality of learning too much since throughout the text the materials are presented in sufficient details and elaboration.

An ideal reader of this book is a person well versed in college-level differential equations who is motivated by physical applications and is interested in gaining insights into field-theoretical physics through differential equations. In order to keep the volume of the book to a reasonable size, we leave out introductory materials about basic physical concepts commonly covered in an undergraduate course in general physics. For example, when we discuss the Ginzburg–Landau theory of superconductivity, we assume the reader knows what a superconductor is and how it behaves. Thus, if this book is used as a textbook for a short or extensive course, it will serve the purpose better if it is supplemented with some extra conceptual nontechnical reading materials, which should be easily available.

In addition to serving for self-study, the materials covered in the book are planned in such a way that each of the chapters may be used for a short concentrated topic course ranging from two to seven weeks or longer, with about two to three hours of lectures per week. Specifically, Chapters 1, 3–9, 11, and 13 may be candidates for a two-week course, Chapters 2, 12, and 15 for a three- to four-week course, and Chapters 10 and 14 for a six- to seven-week course. For a one-semester course, the author suggests picking a collection of about six to seven chapters depending on the interests of the instructor and students. At an elementary level, a choice may be Chapters 1–4, Chapter 7, Chapter 8, and Chapter 11, supplemented with Section 5.1 if necessary. At a more advanced level, a choice may be Chapters 5–10 and Chapter 15. The materials of the full book are more than enough for a year-long course. Moreover, except for Chapters 4 and 15, all other chapters may be studied for research topics and projects of differential equations and nonlinear analysis in theoretical and mathematical physics.

We supplement the book with an Appendices chapter of six sections, which cover some concepts and subjects encountered and used elsewhere in the main text. In the first section, we give a full introduction to the notions of indices of vector fields and topological degrees of maps, in the context of the Euclidean spaces. We begin our discussion from the argument principle in complex analysis and then extend the construction to real situations, highlighted with some applications as examples, including a proof of the fundamental theorem of algebra and a study of the issue of existence and non-existence of periodic orbits of some dynamical systems. Subsequently, we develop the concepts in higher dimensions and conclude the discussion with a proof of the Brouwer fixed-point theorem. In the second section, we consider the concepts of linking number and the Hopf invariant based on our knowledge on the topological degree of a map. We then consider these constructions in view of the concepts of the

helicity of a vector field, the Chern–Simons invariant, and the classical integral representation of the Hopf invariant by Whitehead. In the third section, we present a comprehensive discussion of the Noether theorem, which associates continuous symmetries of a Lagrangian mechanical or field-theoretical system with its conserved quantities schematically. As illustrations, we first consider the motion of a point mass and derive its energy, linear momenta, and angular momenta, as consequences of time- and space-translation invariance and rotation-invariance. We then develop the formalism in the setting of a general Lagrangian field-theoretical framework and show how to construct the associated energy-momentum tensor and various Noether charges and currents. In the fourth section, we describe the possible eigenvalues of the angular momentum operators of a particle in non-relativistic quantum-mechanical motion based on the associated commutation relations of these operators. As a by-product, we explain how to deduce Dirac's charge quantization formula using Saha's method without resorting to a treatment of the Dirac strings. In the fifth section, we show how the concept of the intrinsic spin of a particle in quantum-mechanical motion arises as a result of "correcting" a "deficiency" in the spectra of orbital quantum momentum operators. As a consequence, we are naturally led to the introduction of spin matrices and spinors. In particular, we show how the Pauli spin matrices are called upon, and then explain how the particle spins are related to particle statistics and classification by virtue of the spin-statistics theorem. In the sixth section, we present a comprehensive discussion on the problem of gravitational deflection of light near a massive celestial body. We begin by considering the light deflection problem in the context of Newtonian gravity and derive the associated bending angle. We then study the geodesic equations for the motion of a photon subject to the Schwarzschild black hole metric and deduce Einstein's deflection angle, that exactly doubles that of Newton and was famously confirmed by Dyson, Eddington, and Davidson in 1919.

Thus, these sections may be clustered into four subgroups. The first subgroup consists of the first two sections and concerns with some topological concepts and constructions; the second subgroup is made of the subsequent section that focuses on conservation laws in relation to continuous symmetries in a system; the third subgroup is comprised of the next two sections and addresses issues around the eigenvalues of angular momentum operators and spins of particles; the fourth subgroup, which is the last section of this chapter and supplements Chapter 1 loosely and Chapter 10 tightly, is a study of the gravitational light deflection phenomenon. Each of these four subgroups of subjects may be of independent interest to some readers. As in the rest of the book, exercises appear at the end of the chapter.

The author thanks Sven Bjarke Gudnason, Luciano Medina, Wenxuan Tao, and Tigran Tchrakian for some constructive comments and suggestions, and Dan Taber of Oxford University Press for valuable editorial suggestions and advice, which helped improve the presentation and organization of the contents of the book.

Author
West Windsor, New Jersey

Notation and Convention

We use $\mathbb{N}$ to denote the set of all natural numbers,

$$\mathbb{N} = \{0, 1, 2, \dots\},$$

and $\mathbb{Z}$ the set of all integers,

$$\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}.$$

We use $\mathbb{R}$ and $\mathbb{C}$ to denote the sets of real and complex numbers, respectively.

We use the roman type letter i to denote the imaginary unit $\sqrt{-1}$. For a complex number $c = a + \mathrm{i}b$ where a and b are real numbers we use

$$\overline{c} = a - \mathrm{i}b$$

to denote the complex conjugate of c. We use $\mathrm{Re}\{c\}$ and $\mathrm{Im}\{c\}$ to denote the real and imaginary parts of the complex number $c = a + \mathrm{i}b$. That is,

$$\mathrm{Re}\{c\} = a, \quad \mathrm{Im}\{c\} = b.$$

The signature of an $(n+1)$-dimensional Minkowski spacetime is always $(+ - \cdots -)$. The $(n+1)$-dimensional flat Minkowski spacetime is denoted by $\mathbb{R}^{n,1}$ and is equipped with the scalar product

$$xy = x^0 y^0 - x^1 y^1 - \cdots - x^n y^n,$$

where $x = (x^0, x^1, \dots, x^n)$ and $y = (y^0, y^1, \dots, y^n) \in \mathbb{R}^{n,1}$ are spacetime vectors.

Unless otherwise stated, we always use the Greek letters α, β, μ, ν to denote the spacetime indices,

$$\alpha, \beta, \mu, \nu = 0, 1, 2, \dots, n,$$

and the Latin letters i, j, k, l to denote the space indices,

$$i, j, k, l = 1, 2, \dots, n.$$

We use t to denote the variable in a polynomial or a function or the transpose operation on a vector or a matrix.

When an expression, say X or Y, is given, we use $X \equiv Y$ to denote that Y, or X, is defined to be X, or Y, respectively.

Occasionally, we use the symbol $\forall$ to express "for all", and $\exists$, to express "there exists".

We use $[,]$ to denote the commutators operated on suitable "quantities" so that

$$[A, B] = AB - BA, \quad A_{[a}B_{b]} = A_aB_b - A_bB_a,$$

and so on.

We observe the summation convention over repeated indices unless otherwise stated. For example,

$$A_iB_i = \sum_{i=1}^{n} A_iB_i, \quad A_iB^i = \sum_{i=1}^{n} A_iB^i,$$

$$E_{ij}F^{ij} = \sum_{i,j=1}^{n} E_{ij}F^{ij}, \quad E_{ij}^2 = \sum_{i,j=1}^{n} E_{ij}^2.$$

The roman type letter e is reserved to denote the Euler number or the base of the natural logarithmic system and the italic type letter e to denote an irrelevant physical coupling constant such as the charge of the positron ($-e$ will then be the charge of the electron). The roman type letter d denotes the differential and the italic type letter d denotes a "quantity", in mathematical display mode.

The references in bibliography and their citations in text follow alphabetic orders by the last names of the authors.

Although the Greek letters μ, ν, etc., denote the indices of the spacetime coordinates, occasionally, they are also used to denote the Radon measures or some parameters in other contexts, when there is no risk of confusion but there is a need to be consistent with literature. Furthermore, sometimes ν is used to denote the outnormal to the boundary of a bounded domain.

The unit sphere centered at the origin, in $\mathbb{R}^n$ $(n = 2, 3, \dots)$, is denoted by S^{n-1}.

The area element of a surface such as the boundary of a spatial domain is often denoted by $\mathrm{d}\sigma$. However, the Lebesgue measure of a domain for integration is sometimes omitted to save space when there is no risk of confusion.

Let S be a set of finitely many points. We use $|S|$ or $\#S$ to denote the number of points in S.

Let S be a subset of the set T in a certain space. We use $T \setminus S$ to denote the complement of S in T, or simply S^c when T is the full space.

We use the roman type abbreviation supp to denote the support of a function.

The letter C will be used to denote a positive constant which may assume different values at different places.

For a complex matrix A, we use $A^\dagger$ to denote its Hermitian conjugate, which consists of a matrix transposition and a complex conjugation.

The symbol $W^{k,p}$ denotes the Sobolev space of functions whose distributional derivatives up to the kth order are all in the space L^p.

By convention, various matrix Lie algebras are denoted by lowercase letters. For example, the Lie algebras of the Lie groups $SO(N)$ and $SU(N)$ are denoted by $so(N)$ and $su(N)$, respectively.

The notation for various derivatives is as follows,

$$\partial_\mu = \frac{\partial}{\partial x^\mu}, \quad \partial_\pm = \partial_1 \pm \mathrm{i}\partial_2, \quad \partial = \frac{1}{2}(\partial_1 - \mathrm{i}\partial_2), \quad \overline{\partial} = \frac{1}{2}(\partial_1 + \mathrm{i}\partial_2).$$

Besides, with the complex variable $z = x^1 + \mathrm{i}x^2$, we always understand that $\partial_z = \frac{\partial}{\partial z} = \partial, \partial_{\overline{z}} = \frac{\partial}{\partial \overline{z}} = \overline{\partial}$. Thus, for any function f that only has partial derivatives with respect to x^1 and x^2, the quantities $\partial_z f = \frac{\partial f}{\partial z}$ and $\partial_{\overline{z}} = \frac{\partial f}{\partial \overline{z}}$ are well defined.

Vectors and tensors are often simply denoted by their general components, respectively, following physics literature. For example, it is understood that

$$A_\mu \equiv (A_\mu) = (A_0, A_1, A_2, A_3), \quad g_{\mu\nu} \equiv (g_{\mu\nu}).$$

In a volume of this scope, it is inevitable to have a letter to carry different but standard meanings in different contexts, although such a multiple usage of letters has been kept to a minimum. Here are some examples. The Greek letter ν usually denotes a spacetime coordinate index but also stands for a unit normal vector to a surface; δ may stand for a small positive number, variation of a functional, or the Dirac distribution, and δ_{ij} is the Kronecker symbol; g may stand for a coupling constant such as a magnetic charge, a metric tensor or its determinant, or a function; x may denote the coordinate of a point in the real axis or a point in space or spacetime; P may denote a magnetic charge or a momentum vector; G usually stands for Newton's universal gravitational constant but may also denotes a Lie group or a function; ρ usually stands for a charge, mass, probability, or energy density, but may also denotes a radial variable or radial coordinate under consideration; the lower case letter c usually denotes the speed of light in vacuum but also occasionally a constant that should be made clear in the context.

For convenience, we sometimes use $\mathbf{x}$ to denote a point in $\mathbb{R}^3$ or $\mathbb{R}^n$ in general. We use $\nabla \times \mathbf{F}$ and $\operatorname{curl} \mathbf{F}$, and, $\nabla \cdot \mathbf{F}$ and $\operatorname{div} \mathbf{F}$, interchangeably for the curl and divergence operations, respectively, on a vector field $\mathbf{F}$ over $\mathbb{R}^3$. For an $\mathbb{R}^n$-valued vector, say $\mathbf{A}$, we use $\|\mathbf{A}\|$ and $|\mathbf{A}|$ alternatively to denote the length or norm of $\mathbf{A}$ with respect to the Euclidean scalar product $\cdot$ of $\mathbb{R}^n$ such that $\mathbf{A} \cdot \mathbf{A} = |\mathbf{A}|^2$ is also rewritten as $\mathbf{A}^2$ concisely.

For a positive quantity or variable, say, r, we use $r \ll 1$ or $r \gg 1$ to denote the assumption that r is sufficiently small or large, respectively.

We use the overdot $\dot{}$ to denote differentiation with respect to a "time variable", t, and the prime $'$ differentiation with respect to some other variable, which should be self-evident from the context. Alternatively, we also use $'$ to denote a quantity that is a variation of an original one following a specific rule or understanding.

All displayed mathematical expressions are numbered regardless whether they are referred to in the text for the sake of convenience of the reader in case any need of their reference is called on while using the book.

In some chapters, the first sections also serve to briefly survey the subjects to be covered in the subsequent sections.

1

Hamiltonian systems and applications

This chapter first introduces the Hamiltonian or Lagrangian formalism of classical mechanics, which is the conceptual foundation of all later developments. As illustrations of applications, it then presents a few important many-body problems. Among these, it first discusses the classical N-body problem in $\mathbb{R}^3$ and next considers Kepler's laws of planetary motion as an important application of the formalism and derive Newton's law of gravitation as a by-product. It then presents the Helmholtz–Kirchhoff point vortex problem, which may be regarded as an N-body problem in $\mathbb{R}^2$. The chapter ends with a study of an N-body problem in $\mathbb{R}$ modeling an over-simplified DNA system. In order to understand a thermodynamical phenomenon of the system known as DNA denaturation, it takes this opportunity to make a short introduction to some basic concepts of statistical mechanics.

1.1 Motion of massive particle

The Hamiltonian or Lagrangian formalism of classical mechanics lays the foundation of classical and quantum field theories and grows out of Newtonian mechanics describing the interaction of point masses. In this section, we formulate the Hamilton–Lagrange mechanics from Newton's law of motion.

Equations of motion of Newton

Consider the motion of a point particle of mass m and coordinates $(q^i) = q$ in a potential field $V(q,t)$ described by Newtonian mechanics in the n-dimensional

Mathematical Physics with Differential Equations. Yisong Yang, Oxford University Press.
© Yisong Yang (2023). DOI: 10.1093/oso/9780192872616.003.0001

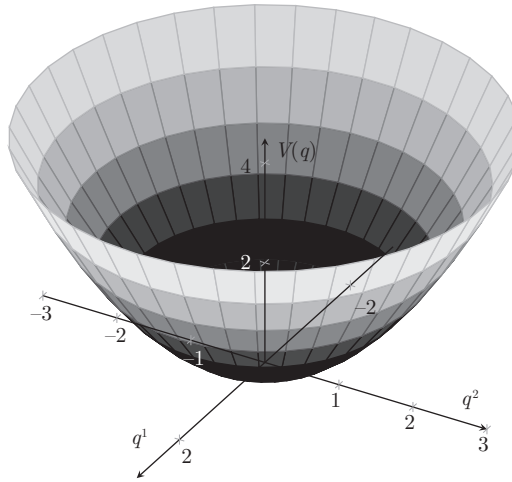

Figure 1.1 An illustrative example of a potential well of that of a two-dimensional harmonic oscillator defined by the quadratic potential function $V(q) = \frac{1}{2}([q^1]^2 + [q^2]^2)$ for which the equations of motion are seen to "drive" the particle to the equilibrium state given by $q^1 = q^2 = 0$, which minimizes the potential energy.

Euclidean space $\mathbb{R}^n$. The equations of motion are

$$m\ddot{q}^i = -\frac{\partial V}{\partial q^i}, \quad i = 1, 2, \ldots, n, \tag{1.1.1}$$

where the (double) overdot $(\ddot{\ }) \dot{\ }$ denotes (second) first-order time derivative. Since

$$-\nabla V = -\nabla_q V = -\left(\frac{\partial V}{\partial q^i}\right) \tag{1.1.2}$$

defines the direction along which the potential energy V decreases most rapidly, the equation (1.1.1) says that the particle is accelerated along the direction of the flow of the *steepest descent* of V. Figure 1.1 shows a typical profile of the potential energy function V in the form of a "potential well."

Lagrangian formalism

With the *Lagrangian function*

$$L(q, \dot{q}, t) = \frac{1}{2}m\sum_{i=1}^{n}(\dot{q}^i)^2 - V(q, t), \tag{1.1.3}$$

which is simply the difference of the kinetic and potential energies of the moving particle, the equations in (1.1.1) are the *Euler–Lagrange equations* of the action

$$\int_{t_1}^{t_2} L(q(t), \dot{q}(t), t)\,\mathrm{d}t, \tag{1.1.4}$$

over the admissible space of trajectories $\{q(t)\,|\,t_1 < t < t_2\}$ starting and terminating at fixed points at $t = t_1$ and $t = t_2$, respectively.

Hamiltonian formalism

The *Hamiltonian function* or *energy* at any time t is the sum of kinetic and potential energies given by

$$H = \frac{1}{2}m\sum_{i=1}^{n}(\dot{q}^i)^2 + V(q,t) = m\sum_{i=1}^{n}(\dot{q}^i)^2 - L. \tag{1.1.5}$$

Introduce the *momentum vector* $p = (p_i)$,

$$p_i = m\dot{q}^i = \frac{\partial L}{\partial \dot{q}^i}, \quad i = 1, 2, \ldots, n. \tag{1.1.6}$$

Then, in view of (1.1.5), H is defined from L, through a *Legendre transformation*, by

$$H(q,p,t) = \sum_{i=1}^{n} p_i\dot{q}^i - L(q,\dot{q},t), \tag{1.1.7}$$

with $p = (p_i)$, and the equations of motion, (1.1.1), are a *Hamiltonian system*,

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q^i}, \quad i = 1, 2, \ldots, n. \tag{1.1.8}$$

In a compressed fashion, the system (1.1.8) reads

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}\begin{pmatrix} \nabla_q H \\ \nabla_p H \end{pmatrix} = J\nabla_{q,p}H, \tag{1.1.9}$$

where I_n denotes the $n\times n$ identity matrix and J is a *symplectic matrix* satisfying $J^2 = -I_{2n}$.

General formalism

For general applications, it is important to consider when the Lagrangian function L is an arbitrary function of $q, \dot{q}$, and t. The equations of motion are the Euler–Lagrange equations of (1.1.4),

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}^i}\right) = \frac{\partial L}{\partial q^i}, \quad i = 1, 2, \ldots, n. \tag{1.1.10}$$

To make a similar Hamiltonian formulation, we are motivated from (1.1.6) to introduce the *generalized momentum vector* $p = (p_i)$ by setting

$$p_i = \frac{\partial L}{\partial \dot{q}^i}, \quad i = 1, 2, \ldots, n. \tag{1.1.11}$$

We still use the Legendre transformation (1.1.7) to define the corresponding Hamiltonian function H. A direct calculation shows that the system (1.1.10) is now equivalent to the Hamiltonian system (1.1.8) in the present general framework.

We note that an important property of a Hamiltonian function is that it is independent of the variables $\dot{q}^i$ $(i = 1, 2, \ldots, n)$. In fact, from the definition of the generalized momentum vector given by (1.1.11), we have

$$\frac{\partial H}{\partial \dot{q}^i} = p_i - \frac{\partial L}{\partial \dot{q}^i} = 0, \quad i = 1, 2, \ldots, n. \tag{1.1.12}$$

This fact justifies our notation of $H(q, p, t)$ in (1.1.7) instead of $H(q, p, \dot{q}, t)$.

Evolution equation

Let F be a dynamical quantity that is an arbitrary function depending on q^i, p_i $(i = 1, 2, \ldots, n)$ and time t. We see that F varies its value along a trajectory of the equations of motion, (1.1.8), according to

$$\begin{aligned} \frac{\mathrm{d}F}{\mathrm{d}t} &= \frac{\partial F}{\partial t} + \frac{\partial F}{\partial q^i}\dot{q}^i + \frac{\partial F}{\partial p_i}\dot{p}_i \\ &= \frac{\partial F}{\partial t} + \frac{\partial F}{\partial q^i}\frac{\partial H}{\partial p_i} - \frac{\partial F}{\partial p_i}\frac{\partial H}{\partial q^i}, \end{aligned} \tag{1.1.13}$$

where, and in the sequel, we observe the *summation convention* over repeated indices, although occasionally we also spell out the summation explicitly. Thus, we are motivated to use the *Poisson bracket* $\{\cdot, \cdot\}$ defined by

$$\begin{aligned} \{f, g\} &= \frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i} \\ &= (\nabla_q f, \nabla_p f) \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix} \begin{pmatrix} \nabla_q g \\ \nabla_p g \end{pmatrix}, \end{aligned} \tag{1.1.14}$$

to rewrite the rate of change of F with respect to time t as

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \frac{\partial F}{\partial t} + \{F, H\}. \tag{1.1.15}$$

In particular, when the Hamiltonian H does not depend on time t explicitly, $H = H(q, p)$, then (1.1.15) implies that

$$\frac{\mathrm{d}H}{\mathrm{d}t} = 0, \tag{1.1.16}$$

which gives the fact that energy is conserved and the mechanical system is thus called *conservative*.

Use of complexified coordinates

It will be useful to "*complexify*" our formulation of classical mechanics. For this purpose we introduce the complex variables

$$u_i = \frac{1}{\sqrt{2}}(q^i + \mathrm{i}p_i), \quad \overline{u}_i = \frac{1}{\sqrt{2}}(q^i - \mathrm{i}p_i), \quad i = 1, 2, \dots, n, \quad \mathrm{i} = \sqrt{-1}. \tag{1.1.17}$$

Here the normalization factor $\frac{1}{\sqrt{2}}$ is introduced in order to make the transformation $(q^i, p_i) \to (u_i, \overline{u}_i)$ *isometric* or *unitary*, $|u_i|^2 + |\overline{u}_i|^2 = |q^i|^2 + |p_i|^2$, in the domain of complex quantities. Then the Hamiltonian function H depends only on $u = (u_i)$ and $\overline{u} = (\overline{u}_i)$,

$$H = H(u, \overline{u}, t). \tag{1.1.18}$$

Hence, in terms of differential operators, there hold

$$\frac{\partial}{\partial u_i} = \frac{1}{\sqrt{2}}\left(\frac{\partial}{\partial q^i} - \mathrm{i}\frac{\partial}{\partial p_i}\right), \quad \frac{\partial}{\partial \overline{u}_i} = \frac{1}{\sqrt{2}}\left(\frac{\partial}{\partial q^i} + \mathrm{i}\frac{\partial}{\partial p_i}\right), \tag{1.1.19}$$

and the Hamiltonian system (1.1.8) takes the concise form

$$\mathrm{i}\dot{u}_i = \frac{\partial H}{\partial \overline{u}_i}, \quad i = 1, 2, \dots, n. \tag{1.1.20}$$

Again, let F be a function depending on $u, \overline{u}$, and t. Then (1.1.20) gives us

$$\begin{aligned} \frac{\mathrm{d}F}{\mathrm{d}t} &= \frac{\partial F}{\partial t} + \frac{\partial F}{\partial u_i}\dot{u}_i + \frac{\partial F}{\partial \overline{u}_i}\dot{\overline{u}}_i \\ &= \frac{\partial F}{\partial t} - \mathrm{i}\frac{\partial F}{\partial u_i}\frac{\partial H}{\partial \overline{u}_i} + \mathrm{i}\frac{\partial F}{\partial \overline{u}_i}\frac{\partial H}{\partial u_i}. \end{aligned} \tag{1.1.21}$$

Thus, with the notation

$$\{f, g\} = \frac{\partial f}{\partial u_i}\frac{\partial g}{\partial \overline{u}_i} - \frac{\partial f}{\partial \overline{u}_i}\frac{\partial g}{\partial u_i} \tag{1.1.22}$$

for the Poisson bracket, we have

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \frac{\partial F}{\partial t} + \frac{1}{\mathrm{i}}\{F, H\}. \tag{1.1.23}$$

In particular, the complexified Hamiltonian system (1.1.20) becomes

$$\dot{u}_i = \frac{1}{\mathrm{i}}\{u_i, H\}, \quad i = 1, 2, \dots, n, \tag{1.1.24}$$

which closely resembles the *Heisenberg equation*, in the *Heisenberg representation* of quantum mechanics, which we discuss later.

1.2 Many-body problem

The *many-body*, or more precisely, *N-body problem* stems from *celestial mechanics*, which treats celestial bodies as point particles interacting through Newton's law of gravitation.

We start from considering the gravitational force between a point mass M fixed at the origin of $\mathbb{R}^3$ and another point mass m at $\mathbf{x} \in \mathbb{R}^3 \setminus \{\mathbf{0}\}$. The potential field that induces the gravitational force is

$$U(\mathbf{x}) = -G\frac{Mm}{|\mathbf{x}|}, \quad \mathbf{x} \in \mathbb{R}^3 \setminus \{\mathbf{0}\}, \tag{1.2.1}$$

where $G > 0$ is Newton's universal gravitational constant, so that it exerts the force

$$-\nabla_{\mathbf{x}} U = -G\frac{Mm}{|\mathbf{x}|^3}\mathbf{x} \tag{1.2.2}$$

to the point mass m at $\mathbf{x}$, leading to the equation of motion,

$$m\ddot{\mathbf{x}} = -G\frac{Mm}{|\mathbf{x}|^3}\mathbf{x}. \tag{1.2.3}$$

Equations of motion of N-body problem

Now consider N point particles, each of mass m_i, located at $\mathbf{x}_i \in \mathbb{R}^3$, $i = 1, \ldots, N$. Then the equations governing the motion of these masses are

$$m_i\ddot{\mathbf{x}}_i = -G\sum_{j\neq i}^{N} \frac{m_i m_j(\mathbf{x}_i - \mathbf{x}_j)}{|\mathbf{x}_i - \mathbf{x}_j|^3} = -\nabla_{\mathbf{x}_i} U, \quad i = 1, \ldots, N, \tag{1.2.4}$$

where

$$U(\mathbf{x}_1, \ldots, \mathbf{x}_n) = -G\sum_{1\leq i<j\leq N}^{N} \frac{m_i m_j}{|\mathbf{x}_i - \mathbf{x}_j|}, \tag{1.2.5}$$

is the total gravitational potential of the N-particle system. In particular, the motion follows the principle that the particles are accelerated along the directions of the fastest descendants that would lower the gravitational potential.

Hamiltonian system

In order to recast the system into a Hamiltonian system, we relabel the coordinate variables and masses according to

$$\begin{aligned}(\mathbf{x}_1, \ldots, \mathbf{x}_N) &\mapsto (q^1, q^2, \ldots, q^{3N}),\\ (m_1, m_1, m_1, \ldots, m_N, m_N, m_N) &\mapsto (m_1, m_2, \ldots, m_{3N}),\end{aligned} \tag{1.2.6}$$

which allows us to introduce the momentum variables

$$p_i = m_i\dot{q}^i, \quad i = 1, 2, \ldots, 3N. \tag{1.2.7}$$

It can be seen that the equations of motion now take the form of a Hamiltonian system

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q^i}, \quad i = 1, 2, \ldots, 3N, \tag{1.2.8}$$

where the Hamiltonian function H is defined by

$$H = \sum_{i=1}^{3N} \frac{p_i^2}{2m_i} + U = \sum_{i=1}^{3N} \frac{1}{2}\dot{q}^i p_i + U. \tag{1.2.9}$$

Conserved quantities

As a general discussion, we begin with considering the first-order differential equations

$$\dot{x}^i = f^i(x), \quad x = (x^i) \in \mathbb{R}^n, \quad i = 1, \ldots, n, \tag{1.2.10}$$

where the functions $f^1(x), \ldots, f^n(x)$ do not depend on time t explicitly such that the equations are referred to as autonomous. A *first integration* of (1.2.10) is a function depending on $x^1, \ldots, x^n$, say $F(x)$, which is constant along any solution of (1.2.10). That is,

$$\frac{\mathrm{d}F(x(t))}{\mathrm{d}t} = \frac{\partial F}{\partial x^i}\dot{x}^i = f^i \partial_i F = 0, \tag{1.2.11}$$

where $x = x(t) = (x^i(t))$ is a solution to (1.2.10). In other words, a first integral is a conserved quantity of time t with respect to the equations of motion. On the other hand, since any solution to (1.2.10) is considered as a curve in $\mathbb{R}^n$ which may in turn be interpreted as the intersection of $n-1$ hypersurfaces, thus the general solution of (1.2.10) may assume the form

$$F_1(x) = C_1, \quad \ldots, \quad F_{n-1}(x) = C_{n-1}, \tag{1.2.12}$$

where $F_1, \ldots, F_n$ are *functionally independent* first integrals of (1.2.10), satisfying the condition that the Jacobian matrix

$$J(F, x) = (\partial_i F_j) \tag{1.2.13}$$

is of full rank. That is, $J(F, x)$ is of rank $n-1$. In this situation, we say that the autonomous system (1.2.10) is *integrable* or *completely integrable* or has a *complete integration*. Therefore, a system is integrable if and only if it possesses the maximum possible number of independent conserved quantities. In particular, an autonomous Hamiltonian system is integrable if and only if it possesses the maximum number of functionally independent quantities, each of which is commutative with respect to the underlying Hamiltonian function of the system and the induced Poisson bracket.

Since the system (1.2.8) consists of $6N$ first-order autonomous equations, its complete integration (solution) requires obtaining $6N-1$ independent integrals. By exploring the mechanical properties of the system, we have the following

immediate integrals (or conserved quantities), namely, the center of masses $\mathbf{x}_0$ determined by

$$\mathbf{x}_0 \sum_{i=1}^{N} m_i = \sum_{i=1}^{N} m_i \mathbf{x}_i, \tag{1.2.14}$$

which moves at a constant velocity; the total (linear) momentum $\mathbf{L}_0$ given by

$$\mathbf{L}_0 = \sum_{i=1}^{N} m_i \dot{\mathbf{x}}_i = \sum_{i=1}^{N} \mathbf{p}_i; \tag{1.2.15}$$

the total angular momentum $\mathbf{a}_0$ expressed as

$$\mathbf{a}_0 = \sum_{i=1}^{N} m_i \mathbf{x}_i \times \dot{\mathbf{x}}_i = \sum_{i=1}^{N} \mathbf{x}_i \times \mathbf{p}_i; \tag{1.2.16}$$

and the conserved total energy H stated in (1.2.9). Thus, we have a total of ten obvious first integrals. This number count indicates that the N-body problem quickly becomes highly nontrivial when N increases. Indeed, the $N = 3$ situation is already notoriously hard, since its integration requires a total of $6 \cdot 3 - 1 = 17$ independent integrals, and is known as the three-body problem. In general, it is believed that the N-body problem is not integrable. The best understood situation is the two-body problem [394, 453], also known as *Kepler's problem* [20]. Here we only mention that the two-body problem is completely integrable for the following reasons:

(i) Using the center of mass coordinate frame it can be shown that the two-body problem is actually planar.

(ii) Notice that, as a planar problem, an N-body Hamiltonian system consists of $4N$ first-order equations.

(iii) The complete integration of an autonomous Hamiltonian system of $4N$ equations requires $4N - 1$ independent integrals.

(iv) The same collection of mechanical quantities give us $2 + 2 + 2 + 1 = 7$ independent integrals.

(v) When $N = 2$ these 7 integrals render the required number of independent integrals for a complete solution of the problem.

When the masses are replaced by charges so that Newton's gravitation is placed by Coulomb's law of electrostatics, we can study the N-body problem of charged particles. The quantum mechanical version of this is called the quantum N-body problem [303], which has important applications in the theory of atoms and molecules and is of contemporary research interest [357]. We consider this problem in Chapter 2.

1.3 Kepler's laws of planetary motion

As an important example of applications of Hamiltonian systems, this section presents a thorough study of Kepler's laws, which describe the motion of a planet around the sun.

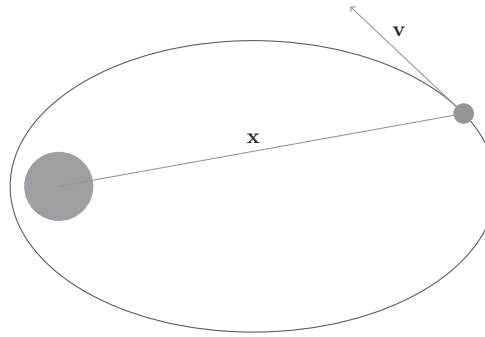

Figure 1.2 The planar motion of a planet around the sun. The orbit is an ellipse with the sun sitting at one of the foci of the ellipse.

Based on his study of then-available observed astronomical data for the motion of planets around the sun, Kepler published between 1609 and 1619 three fundamental laws, known as *Kepler's laws of planetary motion*, which may be stated as follows.

(i) The motion of a planet about the sun is planar and the orbit is an ellipse with the sun sitting at one of the foci of the ellipse. This is *Kepler's first law*.

(ii) The line segment connecting the sun and planet sweeps out equal areas during equal time lapses. This is *Kepler's second law*.

(iii) The square of the time period of the orbit of ellipse is proportional to the cube of the length of the semi-major axis of the ellipse and the proportionality constant is independent of the mass of the planet. This is *Kepler's third law*.

Figure 1.2 illustrates such planetary motion with the sun resting at a focus of the elliptical trajectory orbited by a planet with the position vector $\mathbf{x}$, measured from the focus, and the velocity vector $\mathbf{v} = \dot{\mathbf{x}}$.

Note that, strictly speaking, this is not a two-body problem since the sun is *fixed*.

Polar-variable representation of ellipse

In Cartesian coordinates x, y, the ellipse of semi-major axis $a > 0$ and semi-minor axis $b > 0$ with $a \geq b$ and foci at $(0, 0)$ and $(2c, 0)$ where

$$c = \sqrt{a^2 - b^2}, \tag{1.3.1}$$

centered at $(c, 0)$, is given by the equation

$$\frac{(x-c)^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{1.3.2}$$

such that the quantity

$$e = \frac{c}{a} = \sqrt{1 - \frac{b^2}{a^2}}, \tag{1.3.3}$$

is referred to as the eccentricity of the ellipse. Of course, $0 \leq e < 1$. Thus, in terms of the polar coordinates r, θ, or

$$x = r\cos\theta, \quad y = r\sin\theta, \quad r > 0, \quad 0 \leq \theta \leq 2\pi, \tag{1.3.4}$$

we can recast (1.3.2) into

$$r^2 = \left([1 - e^2]a + er\cos\theta\right)^2. \tag{1.3.5}$$

Therefore we may resolve (1.3.5) to arrive at the polar-variable form of the ellipse (1.3.2):

$$r = r(\theta) = \frac{a(1 - e^2)}{1 - e\cos\theta}. \tag{1.3.6}$$

More generally, if the major axis of the ellipse is tilted up from the x-axis at an angle θ_0 about the origin, then (1.3.6) assumes the modified form

$$r = r(\theta) = \frac{a(1 - e^2)}{1 - e\cos(\theta - \theta_0)}. \tag{1.3.7}$$

Planar motion

Let $\mathbf{x}$ denote the position vector of a planet of mass m under consideration which is attracted toward the sun of mass M resting at the origin. Thus the second law of Newton leads to the equation of motion

$$m\mathbf{a} = -F(r)\mathbf{u}, \quad \mathbf{a} = \dot{\mathbf{v}}, \quad \mathbf{v} = \dot{\mathbf{x}}, \quad \mathbf{u} = \frac{\mathbf{x}}{r}, \quad r = |\mathbf{x}|, \quad \mathbf{x} \neq \mathbf{0}. \tag{1.3.8}$$

An immediate consequence of (1.3.8) is that the vector

$$\mathbf{w} = \mathbf{x} \times \mathbf{v} \tag{1.3.9}$$

is a constant vector. If $\mathbf{w} = \mathbf{0}$, then the planet moves along the radial direction, away or toward the sun. Eventually, a collision will occur, and the motion is not sustainable. Such a situation should be excluded. Thus we may now assume $\mathbf{w} \neq \mathbf{0}$. In this case, we conclude that the motion is confined to a plane that is perpendicular to $\mathbf{w}$.

Proof of Kepler's first law

To determine the orbit of the motion of the planet, we need to integrate the second-order differential equation (1.3.8). We have seen that (1.3.9) is an integral.

So we are to obtain another one. For this purpose, we form the vector $\mathbf{v}\times\mathbf{w}$ and investigate how this vector evolves with time. Hence, with $\mathbf{v}=\dot{r}\mathbf{u}+r\dot{\mathbf{u}}$, we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{v}\times\mathbf{w}) &= \mathbf{a}\times\mathbf{w}\\ &= -\frac{F(r)}{m}\,\mathbf{u}\times(r\mathbf{u}\times[\dot{r}\mathbf{u}+r\dot{\mathbf{u}}])\\ &= -\frac{F(r)r^2}{m}\,\mathbf{u}\times(\mathbf{u}\times\dot{\mathbf{u}}).\end{aligned}\tag{1.3.10}$$

Now recall the vector cross-product identity

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=(\mathbf{A}\cdot\mathbf{C})\mathbf{B}-(\mathbf{A}\cdot\mathbf{B})\mathbf{C}.\tag{1.3.11}$$

Using (1.3.11) in (1.3.10) and applying the properties $\mathbf{u}\cdot\mathbf{u}=1$ and $\mathbf{u}\cdot\dot{\mathbf{u}}=0$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{v}\times\mathbf{w})=\frac{F(r)r^2}{m}\,\dot{\mathbf{u}}.\tag{1.3.12}$$

The left-hand side of (1.3.12) is a total derivative of t. Therefore, in order to render a total derivative for the right-hand side of (1.3.12) to maintain consistency with its left-hand side, it suffices to take

$$F(r)r^2=K=\text{constant}.\tag{1.3.13}$$

In other words, we can draw the conclusion that the function $F(r)$ in (1.3.8) may be taken to follow an inverse-square law,

$$F(r)=\frac{K}{r^2}.\tag{1.3.14}$$

Moreover, consistency in (1.3.12) indicates that K should contain m as a factor such that the right-hand side of (1.3.12) is independent of m as its left-hand side. Besides, by reciprocal symmetry, we then conclude that K should contain M as a factor as well. Consequently we have

$$K=GmM,\tag{1.3.15}$$

where $G>0$ is a proportionality constant independent of m and M which is in fact Newton's universal gravitational constant.

In view of (1.3.8), (1.3.14), and (1.3.15), we have somehow derived Newton's law for gravitation. However, this derivation is heuristic or plausible and its justification is yet to be made through a full examination of Kepler's laws, discussed next.

We now proceed to prove Kepler's first law.

Inserting (1.3.14) and (1.3.15) into (1.3.12), we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\,(\mathbf{v}\times\mathbf{w}-GM\mathbf{u})=\mathbf{0}.\tag{1.3.16}$$

Hence

$$\mathbf{v} \times \mathbf{w} - GM\mathbf{u} = \text{constant} = -\mathbf{A} \text{ (say)}, \tag{1.3.17}$$

which is a second integral of the equation (1.3.8) as desired, in addition to the first integral (1.3.9). Thus the equation (1.3.8) is integrated and it remains to see what its integration or solution looks like.

In fact, from (1.3.9) and (1.3.17), we get

$$\begin{aligned} w^2 \equiv |\mathbf{w}|^2 &= (\mathbf{x} \times \mathbf{v}) \cdot \mathbf{w} = \mathbf{x} \cdot (\mathbf{v} \times \mathbf{w}) \\ &= r\mathbf{u} \cdot (GM\mathbf{u} - \mathbf{A}) \\ &= GMr - r\mathbf{A} \cdot \mathbf{u}. \end{aligned} \tag{1.3.18}$$

Note that, since $\mathbf{v} \times \mathbf{w}$ lies in the plane of the motion of the planet, so does the constant vector $\mathbf{A}$ in view of (1.3.17). Besides, in view of the relation (1.3.18) and the condition $\mathbf{w} \neq \mathbf{0}$, we have

$$\mathbf{A} \cdot \mathbf{u} < GM. \tag{1.3.19}$$

Thus, if the orbit of the planet is a closed curve so that the unit vector $\mathbf{u}$ may assume all possible directions in the plane of the motion, then (1.3.19) is equivalent to

$$|\mathbf{A}| < GM. \tag{1.3.20}$$

If (1.3.20) is violated, the orbit of the planet will not be a closed curve. This situation is not our interest here.

Assume the condition (1.3.20). There are two cases to consider.

(i) $\mathbf{A} = \mathbf{0}$. Then (1.3.18) gives us

$$r = \frac{w^2}{GM}. \tag{1.3.21}$$

Thus the orbit is a circle. This is clearly an exceptional case.

(ii) $\mathbf{A} \neq \mathbf{0}$. Let θ be the angle between $\mathbf{A}$ and $\mathbf{u}$. Then (1.3.18) leads to

$$r = \frac{w^2}{GM - |\mathbf{A}| \cos\theta} \equiv r(\theta). \tag{1.3.22}$$

Comparing (1.3.22) with (1.3.6), we see that the orbit of the planet is an ellipse with eccentricity and semi-major axis given to be

$$e = \frac{|\mathbf{A}|}{GM}, \quad a = \frac{w^2}{GM\left(1 - \left[\frac{|\mathbf{A}|}{GM}\right]^2\right)}, \tag{1.3.23}$$

respectively, and a focus at $r = 0$. This is clearly a generic case.

Thus we have established Kepler's first law.

From (1.3.22), we see that the distance r is maximized at $\theta = 0$, indicating that the planet is at *aphelion*, and minimized at $\theta = \pi$, at *perihelion*, with the values

$$r_a = r(0) = \frac{w^2}{GM - |\mathbf{A}|}, \quad r_p = r(\pi) = \frac{w^2}{GM + |\mathbf{A}|}, \tag{1.3.24}$$

respectively. As a consequence of (1.3.23) and (1.3.24), we get

$$e = \frac{r_a - r_p}{r_a + r_p}. \tag{1.3.25}$$

For the earth, this quantity is about 0.016710218.

Proof of Kepler's second law

Use (r, θ) to denote the polar variables given in (1.3.22) which describe the elliptical motion of a planet about the sun. The area swept out by the line segment connecting the sun and the planet over the span of the angle θ between $\theta = 0$ (say) and $\theta > 0$ is given by the integral

$$\mathcal{A}(\theta) = \int_0^\theta \frac{1}{2} r^2(\varphi) \mathrm{d}\varphi. \tag{1.3.26}$$

Hence, we have

$$\frac{\mathrm{d}\mathcal{A}(\theta)}{\mathrm{d}t} = \frac{1}{2} r^2(\theta)\dot{\theta}. \tag{1.3.27}$$

On the other hand, let $\mathbf{e}_1$ and $\mathbf{e}_2$ be two fixed orthonormal vectors in the plane of the orbit of the planet and write the radial unit vector $\mathbf{u}$ as $\mathbf{u} = \cos\theta\, \mathbf{e}_1 + \sin\theta\, \mathbf{e}_2$. Then we have

$$\begin{aligned} \mathbf{w} = \mathbf{x} \times \mathbf{v} &= r\mathbf{u} \times \dot{\mathbf{x}} = r\mathbf{u} \times (\dot{r}\mathbf{u} + r\dot{\mathbf{u}}) = r^2 \mathbf{u} \times \dot{\mathbf{u}} \\ &= (r^2\dot{\theta})\,(\cos\theta\, \mathbf{e}_1 + \sin\theta\, \mathbf{e}_2) \times (-\sin\theta\, \mathbf{e}_1 + \cos\theta\, \mathbf{e}_2) \\ &= (r^2\dot{\theta})(\mathbf{e}_1 \times \mathbf{e}_2), \end{aligned} \tag{1.3.28}$$

resulting in $w^2 = r^4(\theta)\,\dot{\theta}^2$. We may assume that θ increases with respect to time t. Thus we have

$$\dot{\theta} = \frac{w}{r^2(\theta)}. \tag{1.3.29}$$

Inserting (1.3.29) into (1.3.27), we arrive at

$$\frac{\mathrm{d}\mathcal{A}(\theta)}{\mathrm{d}t} = \frac{w}{2} = \text{constant}, \tag{1.3.30}$$

which establishes Kepler's second law.

A by-product of (1.3.29) is that it gives a description how the angular velocity of the planet depends on its distance from or location with respect to the sun. Substituting (1.3.22) into (1.3.29), we have

$$\dot{\theta} = \frac{(GM - |\mathbf{A}|\cos\theta)^2}{w^3}. \tag{1.3.31}$$

Furthermore, using $\mathbf{u} \cdot \dot{\mathbf{u}} = 0$ in $\mathbf{v} = \dot{r}\mathbf{u} + r\dot{\mathbf{u}}$, we obtain the linear speed of the motion of the planet to be

$$\begin{aligned} v = |\mathbf{v}| &= \sqrt{\dot{r}^2 + r^2\dot{\theta}^2} \\ &= \frac{1}{w}\sqrt{(|\mathbf{A}| - GM)^2 + 2GM|\mathbf{A}|(1 - \cos\theta)}, \end{aligned} \tag{1.3.32}$$

by virtue of (1.3.22) and (1.3.29). In particular, we see that the linear speed v is maximized or minimized wherever the angular velocity $\dot{\theta}$ of the motion is.

Proof of Kepler's third law

Use $T > 0$ to denote the time period for the orbiting planet. When it completes one round of its trip along its elliptical orbit of semi-major axis a and semi-minor axis $b = a\sqrt{1 - e^2}$ determined by (1.3.23) such that the area swept out by the line segment connecting the planet to the sun equals πab, we have by applying (1.3.30) the relation

$$\pi ab = \int_0^T \frac{\mathrm{d}A(\theta)}{\mathrm{d}t}\,\mathrm{d}t = \frac{wT}{2}. \tag{1.3.33}$$

Using (1.3.23) in (1.3.33), we may eliminate the eccentricity to obtain the following neat expression,

$$\begin{aligned} T^2 &= \frac{4\pi^2 a^2 b^2}{w^2} \\ &= \left(\frac{4\pi^2}{GM}\right) a^3, \end{aligned} \tag{1.3.34}$$

which establishes Kepler's third law.

The successful establishment of Kepler's three laws justifies the earlier plausible argument leading to the expression (1.3.15) and, as a consequence, also renders Newton's law for gravitation.

Note that, when Kepler's problem is treated as a two-body system such that the planet is considered to orbit around the center of mass, instead, of a joint or effective total mass, $M + m$, the details of the results obtained are to be correspondingly modified. For example, in the two-body context, the formula (1.3.34) is updated by the analogous expression

$$\frac{a^3}{T^2} = \frac{G(M + m)}{4\pi^2}. \tag{1.3.35}$$

1.4 Helmholtz–Kirchhoff vortex model

In this section, we consider the motion of the Helmholtz–Kirchhoff point vortices in a planar fluid, which may naturally be modeled by an N-body problem in the plane for which a quantity called the *vortex strength* or *vortex charge* serves the

role of mass as in the classical N-body problem in the Euclidean space discussed in Section 1.2. At first sight, the study of fluid motion is concerned with the dynamical properties of continuous media and does not seem to be a subject that would lend itself to a Hamiltonian description of a discrete N-body system. After all, it is a very different phenomenon from the study of orbits of planets moving in empty space subject to gravity as seen in Section 1.3. However, in this section, we will show that, in the context of the formalism by Helmholtz [286] and Kirchhoff [332, 349], the motion of point vortices in a planar fluid can indeed be described by a very simple Hamiltonian N-body system in the plane.

Vorticity field and strength of vorticity

It will be instructive to start from a general discussion. Let $\mathbf{v}$ be the velocity field of a fluid. Then

$$\mathbf{w} = \nabla \times \mathbf{v} \tag{1.4.1}$$

describes the tendency that the fluid swirls itself, which is commonly called the *vorticity field*. Imagine that we form a vortex tube by vortex lines, similar to streamlines induced from the velocity field. Then cut off two cross-sections, say S_1 and S_2, to form a cylindrically shaped finite vortex tube, say T. Then the divergence theorem says that

$$\int_T \nabla \cdot \mathbf{w}\, \mathrm{d}x = \int_T \nabla \cdot (\nabla \times \mathbf{v})\, \mathrm{d}x = 0, \tag{1.4.2}$$

which then implies

$$\int_{S_1} \mathbf{w} \cdot \mathrm{d}\mathbf{S} = \int_{S_2} \mathbf{w} \cdot \mathrm{d}\mathbf{S}, \tag{1.4.3}$$

where the orientations on S_1 and S_2 are chosen in an obviously compatible way. In other words, the flux of vortex lines across the vortex tube is constant along the tube. This common flux is called the *strength* or *tension of the vortex tube*. On the other hand, the circulation of a vector field $\mathbf{v}$ along a closed curve C is defined to be

$$\oint_C \mathbf{v} \cdot \mathrm{d}\mathbf{s}. \tag{1.4.4}$$

Thus, if C is the boundary curve of a cross-section of a vortex tube of the fluid with velocity field $\mathbf{v}$, the above discussion indicates that the strength of the vortex tube may be expressed as the circulation of the fluid around the vortex tube.

Planar situation

Now consider the motion confined in a horizontal plane so that $\mathbf{v} = (v_1, v_2, 0)$. Then the vorticity field $\mathbf{w}$ is always along the vertical direction so that we may express it as a scalar field given by

$$w = \partial_1 v_2 - \partial_2 v_1. \tag{1.4.5}$$

Of course, vortex lines are all vertical to the plane.

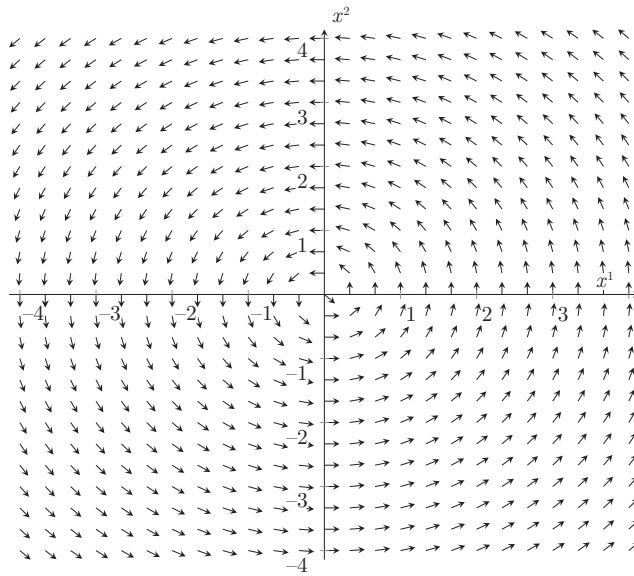

Figure 1.3 A vector field that generates concentric flow-lines plotted in the (x^1, x^2)-coordinate plane.

Helmholtz–Kirchhoff point vortices

A *Helmholtz–Kirchhoff point vortex* centered at the origin of $\mathbb{R}^2$ is an idealized situation where the velocity field is centrally generated from a specified scalar potential function according to the relations

$$v_j = \epsilon_{jk}\partial_k U, \quad j,k = 1,2, \quad U(\mathbf{x}) = -\frac{\gamma}{2\pi}\ln|\mathbf{x}|, \quad \mathbf{x} = (x^1, x^2) \in \mathbb{R}^2, \tag{1.4.6}$$

where ϵ_{jk} is the standard skew-symmetric Kronecker symbol with $\epsilon_{12} = 1$ and $\gamma > 0$ is a parameter. It is clear that the flow-lines are concentric circles around the origin (Figure 1.3).

Let C_r be any one of such circles of radius $r > 0$. Then, we have

$$\oint_{C_r} \mathbf{v} \cdot \mathrm{d}\mathbf{s} = \gamma, \tag{1.4.7}$$

which says the circulation along any flow line or the strength of any vortex tube containing the center of the vortex takes the constant value γ. Here we verify these facts.

Indeed, along C_r, the line element in terms of the polar coordinates r, θ with

$$x^1 = r\cos\theta, \quad x^2 = r\sin\theta, \quad r > 0, \quad 0 \le \theta \le 2\pi, \tag{1.4.8}$$

is

$$\mathrm{d}\mathbf{s} = \mathrm{d}(r\cos\theta, r\sin\theta) = (-r\sin\theta, r\cos\theta)\,\mathrm{d}\theta, \tag{1.4.9}$$

and the velocity vector field $\mathbf{v}$ assumes the form

$$\mathbf{v} = (\partial_2 U, -\partial_1 U) = \left(U_r \frac{x^2}{r}, -U_r \frac{x^1}{r} \right) = U_r(\sin\theta, -\cos\theta), \tag{1.4.10}$$

which is seen to generate concentric flowlines. Therefore, we have

$$\oint_{C_r} \mathbf{v} \cdot \mathrm{d}\mathbf{s} = -\int_0^{2\pi} U_r \, r \, \mathrm{d}\theta = -2\pi r U_r. \tag{1.4.11}$$

Hence, the constant circulation assumption leads to the differential equation

$$U_r = -\frac{\gamma}{2\pi} \frac{1}{r}, \tag{1.4.12}$$

which may be integrated to yield the result

$$U(r) = -\frac{\gamma}{2\pi} \ln r, \tag{1.4.13}$$

as anticipated.

The quantity γ gives the r-independent circulation or strength of the vortex centered at the origin. Furthermore, we can also compute the vorticity field directly,

$$w = -\Delta U = -(\partial_1^2 + \partial_2^2) U = \gamma \delta(\mathbf{x}), \tag{1.4.14}$$

in the sense of *distributions* or *weak derivatives*, which clearly reveals a *point vortex* at the origin given by the Dirac function and justifies again the quantity γ as the *strength of the point vortex*.

It may be instructive to make a note on how (1.4.13) is constructed. First recall that it is clear that a *radial vector field*, $\mathbf{u}$, can be realized as the gradient of a radially symmetric scalar field, say $U(r)$. That is,

$$\mathbf{u} = \nabla_{\mathbf{x}} U = \frac{U'(r)}{r} \mathbf{x}, \tag{1.4.15}$$

as shown in Figure 1.4. Next we can obtain a vector field that generates concentric flow-lines by a 90° counterclockwise rotation of the vector field $\mathbf{u}$ given in (1.4.15). Hence, we arrive at (1.4.10) and all the rest follows naturally.

Planar N-body problem of point vortices

Following the vortex model of Helmholtz [286] and the Hamiltonian formalism of Kirchhoff [332, 349], the dynamical interaction of N point vortices located at $\mathbf{x}_i = \mathbf{x}_i(t) \in \mathbb{R}^2$ of respective strengths γ_i's $(i = 1, \ldots, N)$ at time t is governed by the interaction potential

$$U(\mathbf{x}_1, \ldots, \mathbf{x}_N) = -\frac{1}{2\pi} \sum_{1 \le i < j \le N} \gamma_i \gamma_j \ln |\mathbf{x}_i - \mathbf{x}_j|, \tag{1.4.16}$$

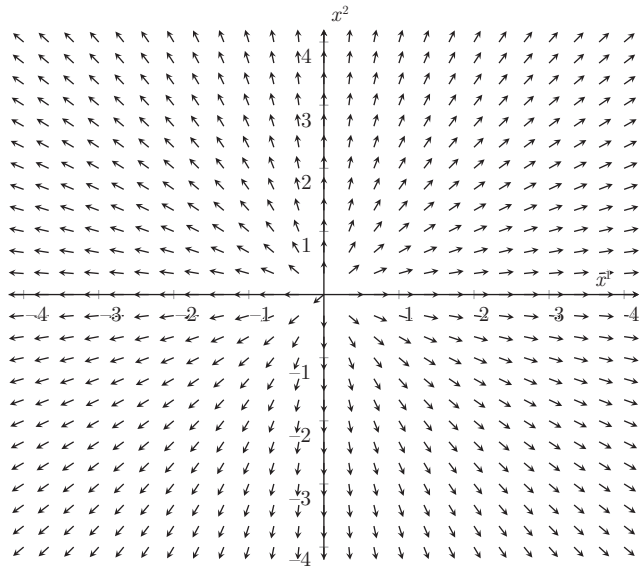

Figure 1.4 A radial vector field that generates co-centered ray-like flow-lines plotted in the (x^1, x^2)-coordinate plane.

and the equations of motion

$$\gamma_i \dot{\mathbf{x}}_i = J\nabla_{\mathbf{x}_i} U, \quad i = 1, \ldots, N, \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{1.4.17}$$

Rewriting $\mathbf{x}$ in the coordinate form with $\mathbf{x} = (x^1, x^2) \in \mathbb{R}^2$ and setting

$$x_i^1 = q^i, \quad \gamma_i x_i^2 = p_i, \quad i = 1, \ldots, N, \tag{1.4.18}$$

we arrive at

$$\dot{q}^i = \frac{\partial U}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial U}{\partial q^i}, \quad i = 1, \ldots, N, \tag{1.4.19}$$

which is a Hamiltonian system. Note that the "momenta" p_i's appear in the Hamiltonian function U in a "non-standard" way. The reason for this odd appearance is that the p_i's actually do not have a mechanical meaning as momenta and are artificially identified as the momentum variables. However, it is interesting to note how the circulations γ_i's are being absorbed into these momenta so consistently.

See [64, 65, 360, 361, 415, 564] and references therein for further developments of the subject of the Helmholtz–Kirchhoff point vortices.

1.5 Partition function and thermodynamics

This section considers some *thermodynamical properties* of a Hamiltonian system in the formalism of *statistical mechanics*. The key to our study is the notion of *partition function*.

Partition function

For simplicity, consider a closed system that can occupy a countable set of states indexed by $s \in \mathbb{N}$ (the set of non-negative integers) and of distinct energies E_s ($s \in \mathbb{N}$). Then the partition function of the system is defined by

$$Z = \sum_{s=0}^{\infty} \mathrm{e}^{-\beta E_s}, \tag{1.5.1}$$

where

$$\beta = \frac{1}{k_{\mathrm{B}} T} \tag{1.5.2}$$

is the *inverse temperature* for which $k_{\mathrm{B}} > 0$ is the *Boltzmann constant* and T the *absolute temperature*. Thus, in order that (1.5.1) makes sense, the sequence $\{E_s\}$ cannot have a limiting point and has to diverge sufficiently rapidly as $s \to \infty$.

Boltzmann factor

Assuming all conditions are valid so that $Z < \infty$ in (1.5.1), we see that

$$P_s = \frac{1}{Z} \mathrm{e}^{-\beta E_s}, \quad s \in \mathbb{N}, \tag{1.5.3}$$

may naturally be interpreted as the probability that the system occupies the state s so that its energy is $E = E_s$ ($s \in \mathbb{N}$). The quantity $\mathrm{e}^{-\beta E_s}$ is also called the *Boltzmann factor*. With such an understanding, the partition function Z may be regarded as the normalization factor of the sequence of the Boltzmann factors which give rise to the probability distribution of the random energy, E, of the system.

Thermodynamic quantities

We now illustrate how to use Z to obtain statistical information of the system.

First, the expected value of the energy (the *thermodynamic value* of the energy) is

$$\begin{aligned} \langle E \rangle &= \sum_{s=0}^{\infty} E_s P_s = \frac{1}{Z} \sum_{s=0}^{\infty} E_s \mathrm{e}^{-\beta E_s} \\ &= -\frac{\partial \ln Z}{\partial \beta} = k_{\mathrm{B}} T^2 \frac{\partial \ln Z}{\partial T}, \end{aligned} \tag{1.5.4}$$

which is also commonly denoted as U. Next, in view of (1.5.4), the variance is

$$\begin{aligned}\sigma_E^2 &= \langle (E - \langle E\rangle)^2\rangle = \langle E^2\rangle - \langle E\rangle^2 \\ &= \frac{1}{Z}\sum_{s=0}^{\infty} E_s^2 \mathrm{e}^{-\beta E_s} - \left(\frac{\partial \ln Z}{\partial \beta}\right)^2 \\ &= \frac{1}{Z}\frac{\partial^2 Z}{\partial \beta^2} - \frac{1}{Z^2}\left(\frac{\partial Z}{\partial \beta}\right)^2 = \frac{\partial^2 \ln Z}{\partial \beta^2},\end{aligned} \tag{1.5.5}$$

which, in view of (1.5.4) and (1.5.5), gives rise to the *heat capacity*

$$C_v = \frac{\partial U}{\partial T} = \frac{\partial \langle E\rangle}{\partial T} = \frac{1}{k_{\mathrm{B}} T^2}\frac{\partial^2 \ln Z}{\partial \beta^2} = \frac{1}{k_{\mathrm{B}} T^2}\sigma_E^2. \tag{1.5.6}$$

Besides, the *entropy* or the *Gibbs entropy*, also often called the *Shannon entropy*, of the system, S, which measures the *disorder* or *uncertainty* of the system, is given in view of (1.5.4) by

$$\begin{aligned}S &= -k_{\mathrm{B}} \sum_{s=0}^{\infty} P_s \ln P_s = -k_{\mathrm{B}} \sum_{s=0}^{\infty} P_s \ln \frac{\mathrm{e}^{-\beta E_s}}{Z} \\ &= k_{\mathrm{B}} \sum_{s=0}^{\infty} (P_s \ln Z + \beta E_s P_s) \\ &= k_{\mathrm{B}}(\ln Z + \beta \langle E\rangle) = k_{\mathrm{B}} \ln Z + \frac{1}{T}\langle E\rangle \\ &= \frac{\partial}{\partial T}(k_{\mathrm{B}} T \ln Z) \equiv -\frac{\partial A}{\partial T},\end{aligned} \tag{1.5.7}$$

where

$$\begin{aligned}A &= -k_{\mathrm{B}} T \ln Z \\ &= U - TS \quad (U = \langle E\rangle),\end{aligned} \tag{1.5.8}$$

in view of the second line in (1.5.7), is the *Helmholtz free energy*. These examples show the usefulness of the partition function.

Derivation of partition function

After seeing the importance of the partition function (1.5.1), here we show how it arises naturally from the laws of thermodynamics.

First, when a system is at its thermodynamic equilibrium, the *second law of thermodynamics* asserts a *maximized entropy*. Thus, mathematically, the system follows the probability distribution

$$P_s = P(\{E = E_s\}), \tag{1.5.9}$$

that maximizes the entropy

$$S = -k_{\mathrm{B}} \sum_{s=0}^{\infty} P_s \ln P_s, \tag{1.5.10}$$

subject to the constraints

$$\sum_{s=0}^{\infty} P_s = 1, \quad \sum_{s=0}^{\infty} E_s P_s = U, \tag{1.5.11}$$

imposed to the total probability and average energy. Therefore, we are to extremize the Lagrange function

$$L = -k_{\mathrm{B}} \sum_{s=0}^{\infty} P_s \ln P_s + \lambda_1 \left(\sum_{s=0}^{\infty} P_s - 1 \right) + \lambda_2 \left(\sum_{s=0}^{\infty} E_s P_s - U \right), \tag{1.5.12}$$

where λ_1, λ_2 are the Lagrange multipliers. So we are led to setting up the equations

$$\frac{\partial L}{\partial P_s} = -k_{\mathrm{B}} (\ln P_s + 1) + \lambda_1 + \lambda_2 E_s = 0, \quad \forall s \in \mathbb{N}, \tag{1.5.13}$$

which render the solution

$$P_s = a \mathrm{e}^{\frac{\lambda_2 E_s}{k_{\mathrm{B}}}}, \quad a \equiv \mathrm{e}^{-1+\frac{\lambda_2}{k_{\mathrm{B}}}}. \tag{1.5.14}$$

Next, multiplying (1.5.13) by P_s, summing up, and using (1.5.11), we have

$$S - k_{\mathrm{B}} + \lambda_1 + \lambda_2 U = 0, \tag{1.5.15}$$

which yields the differential relation

$$\mathrm{d}S + \lambda_2 \mathrm{d}U = 0. \tag{1.5.16}$$

On the other hand, recall the *first law of thermodynamics*, which says the increment of heat to the system, $\mathrm{d}Q = T\mathrm{d}S$, is the result of increment of average energy, $\mathrm{d}U$, and the extra mechanical work done to the system $\mathrm{d}W$, namely, $T\mathrm{d}S = \mathrm{d}U + \mathrm{d}W$. However, in the present equilibrium situation, $\mathrm{d}W = 0$. Hence, $T\mathrm{d}S = \mathrm{d}U$. Inserting this into (1.5.16), we find

$$\lambda_2 = -\frac{1}{T}. \tag{1.5.17}$$

Then, substituting (1.5.17) into (1.5.14), we obtain

$$P_s = a \mathrm{e}^{-\beta E_s}, \quad \beta = \frac{1}{k_{\mathrm{B}} T}. \tag{1.5.18}$$

Finally, summing up P_s in (1.5.18) and using (1.5.11), we have

$$\frac{1}{a} = \sum_{s=0}^{\infty} \mathrm{e}^{-\beta E_s}, \tag{1.5.19}$$

which gives us the partition function defined in (1.5.1).

Hamiltonian system

For a classical Hamiltonian system with generalized coordinates $q = (q_1, \ldots, q_n)$ and momenta $p = (p_1, \cdots, p_n)$, governed by the Hamiltonian function $H(q, p)$, the partition function is expressed by

$$Z = \int \mathrm{e}^{-\beta H(q,p)} \, \mathrm{d}q\mathrm{d}p, \tag{1.5.20}$$

where q, p take over the role of the state index s and the integral replaces the summation in our earlier discussion. Therefore, a similar collection of knowledge can be gathered as before. For example, if $F(q, p)$ is a mechanical quantity of interest, then its expected or thermodynamic value is given by

$$\langle F \rangle = \frac{1}{Z} \int F(q, p) \mathrm{e}^{-\beta H(q,p)} \, \mathrm{d}q\mathrm{d}p. \tag{1.5.21}$$

A fairly thorough treatment of statistical mechanics may be found in [252, 297, 350, 391]. The next section applies these ideas to study a thermodynamic property of DNA.

1.6 Dynamic modeling of DNA denaturation

DNA, the short name for *deoxyribonucleic acid*, is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms. Chemically, a DNA consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups. These two strands run in parallel and form a double helix. Attached to each sugar is one of four types of nucleotide molecules, also called *bases*, named by letters A (adenine), C (cytosine), G (guanine), and, T (thymine), so that only A and T, C and G, from opposite strands may bind to form pairs. During the last four decades, biologists and physicists have carried out extensive research on the dynamics of DNA, using mathematical modeling and computer simulation, and obtained profound knowledge about DNA and its function.

Mathematical modeling of dynamics of DNA was initiated in 1980 by Englander and colleagues [192], who presented a discrete *sine–Gordon soliton interpretation of the DNA* of n pairs of bases and used the solitary wave in the continuous limit as an approximation in the limit $n \to \infty$ to obtain some qualitative behavior of DNA. In 1989, Peyrard and Bishop [446] published their pioneering work on DNA dynamical modeling in which the base pairing due to hydrogen bonding is recognized, the discreteness of the model is maintained, and a statistical mechanics study is fully carried out which describes the inter-strand separation in the double helix as a function of temperature, leading to a mathematical formulation of *DNA denaturation*, that is, the phenomenon that the two DNA strands in the double helix become separated when heated.

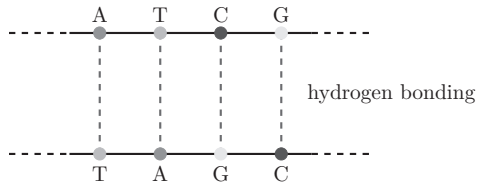

Figure 1.5 An oversimplified DNA chain model.

Chain model

Following [446], here we initially allow two degrees of freedom for each pair of bases and use u_i and v_i to denote the transverse displacements of the bases from their equilibrium positions along the direction of the hydrogen bonds that connect the two bases in a pair. The governing Hamiltonian for the *double helix model* containing a longitudinal *harmonic coupling* between neighboring bases due to *stacking* and n pairs of bases is given as

$$H = \sum_{i=1}^{n} \left(\frac{1}{2} m(\dot{u}_i^2 + \dot{v}_i^2) + \frac{1}{2}\kappa([u_i - u_{i-1}]^2 + [v_i - v_{i-1}]^2) + V(u_i - v_i) \right), \tag{1.6.1}$$

where a common mass m is taken for all bases, a uniform "elastic" (*stacking force*) constant κ is assumed for simplicity, and the potential energy V is defined by

$$V(u) = D(\mathrm{e}^{-au} - 1)^2, \tag{1.6.2}$$

which accounts for the hydrogen bonding, with a, D some positive constants, and is of the *Morse type* [401]. Figure 1.5 illustrates just such an oversimplified model. Here we note that the definitions of u_0 and v_0 depend on the specific boundary condition to be considered.

In terms of the new variables x_i, y_i and the associated momenta p_i, P_i, defined by

$$x_i = \frac{(u_i + v_i)}{\sqrt{2}}, \quad y_i = \frac{(u_i - v_i)}{\sqrt{2}}, \quad p_i = m\dot{x}_i, \quad P_i = m\dot{y}_i, \tag{1.6.3}$$

the Hamiltonian (1.6.1) is normalized into the form

$$\begin{aligned} H = &\sum_{i=1}^{n} \left(\frac{p_i^2}{2m} + \frac{1}{2}\kappa(x_i - x_{i-1})^2 \right) \\ &+ \sum_{i=1}^{n} \left(\frac{P_i^2}{2m} + \frac{1}{2}\kappa(y_i - y_{i-1})^2 + D(\mathrm{e}^{-a\sqrt{2}y_i} - 1)^2 \right). \end{aligned} \tag{1.6.4}$$

It is important to realize that the variable y_i measures the stretching distance between the bases in a pair of bases.

Factorization of partition function

To understand the thermal dynamics of stretching, note that the partition function Z is seen to be factored as

$$Z = \int \mathrm{e}^{-\beta H(p,x,P,y)}\,\mathrm{d}x\mathrm{d}y\mathrm{d}p\mathrm{d}P = Z_p Z_x Z_P Z_y, \tag{1.6.5}$$

where $\beta = (k_{\mathrm{B}}T)^{-1}$, with T the absolute temperature and k_{B} the Boltzmann constant, and, to save space, we use $\mathrm{d}x$ (say) to denote $\mathrm{d}x^1 \cdots \mathrm{d}x^n$ and use x to denote the vector coordinates (x^i) or a single variable interchangeably if there is no risk of confusion in the context. From this, Peyrard and Bishop [446] recognized that the mean stretching $\langle y_\ell \rangle$ of the bases at the position $\ell = 1, \ldots, n$, due to the hydrogen bonding, is given by

$$\langle y_\ell \rangle = \frac{1}{Z}\int y_\ell \mathrm{e}^{-\beta H}\,\mathrm{d}x\mathrm{d}y\mathrm{d}p\mathrm{d}P = \frac{1}{Z_y}\int y_\ell \mathrm{e}^{-\beta\sum_{i=1}^{n} f(y_i, y_{i-1})}\,\mathrm{d}y, \tag{1.6.6}$$

where the factors involving x, p, P are dropped as a consequence of the decomposed Hamiltonian (1.6.4) and $f(y, y')$ is the reduced potential given by the y-dependent terms in (1.6.4) as

$$f(y, y') = \frac{1}{2}\kappa(y - y')^2 + D\left(\mathrm{e}^{-a\sqrt{2}y} - 1\right)^2. \tag{1.6.7}$$

It is still rather difficult to analyze the quantity (1.6.6) as a function of the temperature T without further simplification. In [446], Peyrard and Bishop take $n \to \infty$ in (1.6.6) to arrive at the position-independent mean stretching

$$\langle y \rangle = \langle \varphi_0 | y | \varphi_0 \rangle = \int \varphi_0^2(y) y\,\mathrm{d}y, \tag{1.6.8}$$

where

$$\varphi_0(y) = \frac{(\sqrt{2}a)^{\frac{1}{2}}(2d)^{d-\frac{1}{2}}}{\Gamma(2d-1)^{\frac{1}{2}}}\exp(-d\mathrm{e}^{-\sqrt{2}ay})\mathrm{e}^{-(d-\frac{1}{2})\sqrt{2}ay}, \tag{1.6.9}$$

$$d = \frac{1}{a}\beta(\kappa D)^{\frac{1}{2}} > \frac{1}{2}. \tag{1.6.10}$$

Based on this formalism, Peyrard and Bishop [446] succeeded in finding a thermodynamical description of the DNA denaturation phenomenon. Using (1.6.9) with (1.6.10) and numerical evaluation, they showed that the base mean stretching $\langle y_\ell \rangle$ increases significantly as the temperature climbs to a particular level which is an unambiguous indication of DNA denaturation. Another interesting by-product of such a calculation is that, since the dependence of the ground state on the absolute temperature $T = (k_{\mathrm{B}}\beta)^{-1}$ is through the parameter d given earlier, a greater value of the elastic constant κ leads to a higher DNA denaturation temperature, which is what was observed [214, 446] in the laboratory.

Hamiltonian of out-of-phase motion

In particular, we see that the dynamics of the DNA molecule is effectively described by the reduced Hamiltonian that contains the "out-of-phase" motion of the bases only given in terms of the y-variables as

$$H = \sum_{i=1}^{n} \left(\frac{1}{2} m \dot{y}_i^2 + \frac{1}{2} \kappa (y_i - y_{i-1})^2 + D \left(\mathrm{e}^{-a\sqrt{2} y_i} - 1 \right)^2 \right). \tag{1.6.11}$$

See [445] for a review of related topics and directions. This example shows how a simple system of ordinary differential equations may be used to investigate a fundamental problem in biophysics.

Exercises

1. Let $q = (q^1, \dots, q^n) \in \mathbb{R}^n$ be the position coordinate vector of a particle that passes the points $q = q_1$ and $q = q_2$ at the times $t = t_1$ and $t = t_2$, respectively. Assume that the motion of such a particle is governed by the action
$$\mathcal{A}(q) = \int_{t_1}^{t_2} L(q(t), \dot{q}(t), t)\, \mathrm{d}t \tag{1.E.1}$$
so that the trajectory of the motion $q = q(t)$ is a critical point of the action. Show that $q(t)$ solves the equation of motion
$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial L}{\partial \dot{q}^i} \right) = \frac{\partial L}{\partial q^i}, \quad i = 1, 2, \dots, n. \tag{1.E.2}$$
2. Let (x, y) be a point on the ellipse defined by the equation (1.3.2). Show that the sum of the distances from (x, y) to the two foci, $(0, 0)$ and $(2c, 0)$, is $2a$.
3. Consider the equation of an ellipse given by (1.3.6). It is clear that the maximum $r_{\max}$ and minimum $r_{\min}$ of the radial distance $r = r(\theta)$ from a point on the ellipse to its focus at the origin are attained at the points where $\theta = 0$ and $\theta = \pi$, called the aphelion and perihelion, respectively. Show that the eccentricity e of the ellipse can be represented in terms of $r_{\max}$ and $r_{\min}$ by
$$e = \frac{r_{\max} - r_{\min}}{r_{\max} + r_{\min}}. \tag{1.E.3}$$
4. Formulate the equation of motion (1.3.8) in the form of a Hamiltonian system.
5. Consider the gravitational interaction of two point masses μ and m, fixed at the origin and moving around the origin, respectively, in spherical coordinates
$$x = r \sin\theta \cos\phi, \quad y = r \sin\theta \sin\phi, \quad z = r \cos\theta. \tag{1.E.4}$$

(a) Show that the kinetic energy of the moving point mass m is given by

$$\mathcal{K} = \frac{1}{2} m \left(\dot{r}^2 + r^2 \dot{\theta}^2 + r^2 \sin^2 \theta \, \dot{\phi}^2 \right), \tag{1.E.5}$$

and obtain the associated Lagrangian function.

(b) Show that the momenta of the moving mass associated with the cooordinates r, θ, ϕ are

$$p_r = m\dot{r}, \quad p_\theta = mr^2 \dot{\theta}, \quad p_\phi = mr^2 \sin^2 \theta \, \dot{\phi}. \tag{1.E.6}$$

(c) Use (b) to show that the Hamiltonian of the problem is

$$H(r, \theta, \phi, p_r, p_\theta, p_\phi) = \frac{1}{2m} \left(p_r^2 + \frac{p_\theta^2}{r^2} + \frac{p_\phi^2}{r^2 \sin^2 \theta} \right) - \frac{\mu m}{r}, \tag{1.E.7}$$

(d) Derive the Hamiltonian equation from (c) governing the mechanical variables $r, \theta, \phi, p_r, p_\theta, p_\phi$.

6. Consider the classical central-force motion of a particle of mass m governed by the equation

$$m\ddot{\mathbf{x}} = f(x, y, z)\mathbf{x}, \quad \mathbf{x} = (x, y, z) \in \mathbb{R}^3 \setminus \{\mathbf{0}\}, \tag{1.E.8}$$

where f is a real-valued continuous function.

(a) Establish the law of conservation for the angular momentum

$$\frac{\mathrm{d}}{\mathrm{d}t}(m\mathbf{x} \times \dot{\mathbf{x}}) = \mathbf{0}. \tag{1.E.9}$$

(b) Use (a) to show that the motion of the particle is planar. That is, its orbit is confined to a plane, say P.

(c) Prove that the areas in P swept out by the position vector $\mathbf{x}$ during equal time intervals are equal. In other words, Kepler's second law is valid for central-force motion, in general, governed by the equation (1.E.8).

7. Derive (1.3.32).

8. Consider the time-dependent Hamiltonian

$$H(q, p, t) = \frac{p^2}{2m} + \frac{kq^2}{2} - \mu \, q \cos(\omega t), \tag{1.E.10}$$

where $m, k, \mu, \omega > 0$ are constants.

(a) Write the Hamiltonian equations.

(b) Find the general solution of the Hamiltonian equations.

(c) Obtain the Lagrangian function and Lagrange equation.

9. Consider the nonlinear equation

$$\ddot{x} = -x + \frac{\lambda}{(1-x)^2}, \tag{1.E.11}$$

where $\lambda > 0$ is a constant, governing the dynamics of the moving plate of an electrostatic actuator in a microelectromechanical system (MEMS).

(a) Find the Lagrangian and Hamiltonian functions of the equation.

(b) Show that when $\lambda > 0$ is small the solution with the initial condition

$$x(0) = 0, \quad \dot{x}(0) = 0 \tag{1.E.12}$$

is periodic. See [609].

10. Consider the motion of a point particle of mass m and electric charge Q in an electromagnetic field of electric potential V and magnetic potential $\mathbf{A}$. If the spatial position vector $\mathbf{x}$ of the particle at time t is $\mathbf{x}(t)$, then $\mathbf{x}(t)$ is governed by the Lagrangian function

$$L(\mathbf{x}, \dot{\mathbf{x}}, t) = \frac{1}{2} m\dot{\mathbf{x}}^2 - QV(\mathbf{x}, t) + Q\dot{\mathbf{x}} \cdot \mathbf{A}(\mathbf{x}, t). \tag{1.E.13}$$

(a) Obtain the equation of motion or the Lagrange equation of the particle.

(b) Use (a) to show that the equation of motion of the charged particle is of the form of the Newton law $m\ddot{\mathbf{x}} = \mathbf{F}$ where $\mathbf{F}$ is the *Lorentz force* given by

$$\mathbf{F} = Q(\mathbf{E} + \dot{\mathbf{x}} \times \mathbf{B}), \tag{1.E.14}$$

with

$$\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla \times \mathbf{A}, \tag{1.E.15}$$

being the electric and magnetic fields induced from the potential fields V and $\mathbf{A}$.

(c) Find the Hamiltonian function of the system.

(d) Recover the result in (a) by obtaining the Hamiltonian equations for the motion of the particle.

11. As in the discussion of the Kirchhoff vortex model in $\mathbb{R}^2$, construct a centralized (radial) vector field $\mathbf{F}$ in $\mathbb{R}^3$ such that the flux

$$\Phi = \int_{\partial\Omega} \mathbf{F} \cdot \mathrm{d}\mathbf{S} \tag{1.E.16}$$

is independent of the choice of the bounded domain Ω around the origin of $\mathbb{R}^3$. (You may start your study with the situation when Ω is an arbitrary ball.)

12. Let the energies of a system be $E_s = \varepsilon s, s = 0, 1, 2, \ldots$, where $\varepsilon > 0$ is fixed.

(a) Find the partition function Z.

(b) Compute the Helmholtz free energy.

(c) Compute the entropy of the system.

13. Let the energy spectrum of a thermodynamic system be a continuum given by $\{E_s = \varepsilon s \,|\, s > 0\}$ where $\varepsilon > 0$ is fixed.

(a) Find the partition function Z.

(b) Find the distribution function of the energy.

(c) Find the thermodynamic energy U.

14. Let the partition function of a system be $Z = 1 + \mathrm{e}^{-\beta\varepsilon}$ where $\varepsilon > 0$.

(a) Compute the thermodynamic energy U and find how it depends on the absolute temperature T. In particular, find

$$U(0) = \lim_{T\to 0} U, \quad U(\infty) = \lim_{T\to\infty} U. \tag{1.E.17}$$

(b) Use (a) to find the heat capacity of the system.

15. Derive the Hamiltonian equations of the model (1.6.11) assuming $y_0 = 0$ and $y_0 = y_n$, respectively.

16. Consider an over-simplified model of one-base DNA dynamics governed by the equation

$$\ddot{x} + kx + D\mathrm{e}^{x}(\mathrm{e}^{x} - 1) = f(t), \tag{1.E.18}$$

where $k, D > 0$ are constants and $f(t)$ is a forcing term.

(a) Find the Lagrangian and Hamiltonian functions of the equation.

(b) Let $f(t)$ be of period $T > 0$. Investigate whether the equation has a solution of period T under some appropriate conditions.

2

Schrödinger equation and quantum mechanics

Quantum mechanics, developed at the beginning of the twentiety century, attempted to explain a broad range of physical phenomena in microscopic scales based upon a series of celebrated experiments that could not be explained within the conceptual framework of classical physics. This chapter focuses on the Schrödinger equation, which is the foundation of quantum mechanics, and aims to understand some fundamental features of it. It then presents a few methodological approaches to analyzing the quantum many-body problem, which is of contemporary research interest. In doing so, it shows that the study of the quantum aspects of an N-body system, which is governed by a set of nonlinear ordinary differential equations, renders the problem linear through the Schrödinger equation, which is a partial differential equation, governing the state of the system. Furthermore, the search for approximate solutions of the Schrödinger equation leads us to considering nonlinear equations of combined differential-integral type and problems of calculus of variations.

2.1 Path to quantum mechanics

The core or essence of quantum mechanics is the Schrödinger equation. Thus, the path to quantum mechanics lies in understanding how the Schrödinger equation came about. The goal of the first two sections of this chapter is to understand how the Schrödinger equation is introduced. To this goal, in this section, we first recall some milestone early-day discoveries which revealed the wave-particle duality nature of matter interaction in microscopic world and how these discoveries were given precise mathematical perceptions or formulations successfully by the pioneers of quantum mechanics. Then, the next section shows

Mathematical Physics with Differential Equations. Yisong Yang, Oxford University Press.
© Yisong Yang (2023). DOI: 10.1093/oso/9780192872616.003.0002

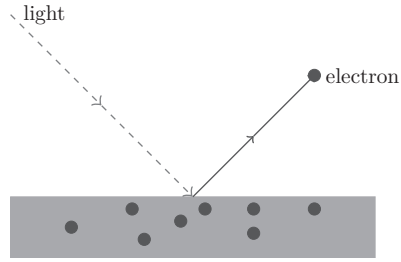

Figure 2.1 An illustration of the photoelectric effect that electrons in a piece of metal may be energized by a light beam to escape from the metal.

how the Schrödinger equation may be "derived," such that its distinguished wave-equation characteristics serve to unify these formulations in a single framework.

Photoelectric effect

Place a piece of metal in a vacuum tube and shoot a beam of light onto it. The electrons in the metal may become sufficiently energized to be emitted from the metal (Figure 2.1). This is known as the *photoelectric effect*, which has found a wide range of applications in today's electronics. Now measure the energy carried by an emitted electron and denote it by E_e. It is known that E_e may be written as the difference of two quantities, one is proportional to the frequency, ν, of the light beam so that the proportionality constant, h, is universal and independent of the metal, the other, ϕ, depends on the metal but is independent of the light frequency. Therefore, we have

$$E_e = h\nu - \phi. \tag{2.1.1}$$

Einstein's postulate

In 1905, Einstein postulated that light, a special form of electromagnetic waves, is composed of particles called photons. Each photon carries an amount of energy equal to $h\nu$. That is,

$$E = h\nu. \tag{2.1.2}$$

When the photon hits an electron in a metal, the electron receives this amount of energy, consumes the amount of the binding energy of the metal to the electron to escape from the metal, and becomes an emitted electron of the energy given by (2.1.1). The equation (2.1.2) is also known as the *Planck–Einstein relation*. Section 15.6 presents a derivation of the formula (2.1.2) as a consequence of quantization of electromagnetic fields.

In terms of angular frequency

In physics, frequency ν is measured in hertz with unit second^{-1} (times per second), and angular frequency ω is related to ν by $\omega = 2\pi\nu$ (radians per second). Hence, in terms of ω, the Einstein formula becomes

$$E = \hbar\omega, \quad \hbar = \frac{h}{2\pi}. \tag{2.1.3}$$

Recall that energy is measured in unit of Joules and one Joule is equal to one Newton×meter. The constant $\hbar$ in (2.1.3), called the *Planck constant*, is a tiny number of the unit of Joules × second and accepted to be

$$\hbar = 1.05457 \times 10^{-34}. \tag{2.1.4}$$

Historically, h is called the Planck constant, and $\hbar$ the *Dirac constant* or the *extended Planck constant*.

Compton effect

After Einstein's 1905 postulate that light is composed of photons, physicists began to wonder whether a photon might exhibit its kinetic momentum in interaction, that is, when colliding with another particle. In 1922, Compton and Debye came with a very simple mathematical description of this behavior, which was then experimentally observed by Compton himself in 1923 and further confirmed by Woo, then Compton's graduate student. In simple terms, when a photon hits an electron, it behaves like a particle when it collides with another particle so that one observes energy as well as momentum conservation relations, which is evidenced by a wavelength shift after the collision.

Intuitively and mathematically, we may write the energy of a photon by the *Einstein formula*, $E = mc^2$, where c is the speed of light in vacuum and m is the "virtual rest mass" of photon (note that a photon in fact has no rest mass so that the connection made this way is only intuitive, and a completely rigorous treatment along the same line may be formulated with the full relativistic energy-mass-momentum formula (4.3.22) to be derived in Chapter 4). Thus, in view of (2.1.3), we have

$$E = mc^2 = \hbar\omega. \tag{2.1.5}$$

On the other hand, recall that the wavenumber (also called the *angular wavenumber*) k, wavelength λ, frequency ν, angular frequency ω, and speed c of a photon are related by

$$k = \frac{2\pi}{\lambda}, \quad c = \lambda\nu = \lambda\,\frac{\omega}{2\pi} = \frac{\omega}{k}. \tag{2.1.6}$$

Consequently, the momentum of the photon is given by

$$\begin{aligned} p = mc &= \frac{E}{c} = \hbar\,\frac{\omega}{c} \\ &= \hbar\, k. \end{aligned} \tag{2.1.7}$$

De Broglie's wave-particle duality hypothesis

In 1924, de Broglie formulated his celebrated *wave-particle duality* hypothesis in his PhD thesis, which equalizes waves and particles, takes the Einstein formula (2.1.3) and the *Compton–Debye formula* (2.1.7) as two axioms, and reiterates the wave and particle characteristics of all interactions in nature:

$$E = \hbar\,\omega, \tag{2.1.8}$$

$$p = \hbar\,k. \tag{2.1.9}$$

In other words, a particle of energy E and momentum p behaves like a wave of wavenumber k and a wave of wavenumber k behaves like a particle of energy E and momentum p such that E, p, and k are related through (2.1.8) and (2.1.9). In Section 15.6, we show how the Compton–Debye formula (2.1.7) may be derived when we quantize electromagnetic fields.

2.2 Schrödinger equation

Based on de Broglie's wave-particle duality, we now derive the *Schrödinger equation*, first published by Schrödinger in 1926. The wave-equation features of this equation and the statistical interpretation of its solution enable us to perceive and understand some of the most profound physical properties of nature only available or observable at microscopic scales. Interestingly, we will see how classical and quantum mechanics are statistically linked, also through the Schrödinger equation.

Wave motion in terms of angular wavenumber and frequency

Consider a stationary wave distributed over the x-axis of wavenumber k (the wave has k repeated cycles over the standard angular (cell) interval $[0, 2\pi]$) whose simplest form is given by

$$C\mathrm{e}^{\mathrm{i}kx}. \tag{2.2.1}$$

Switch on time-dependence so that the wave moves to right (say) at velocity $v > 0$. We see from (2.2.1) that the wave is represented by

$$\phi(x,t) = C\mathrm{e}^{\mathrm{i}k(x-vt)}. \tag{2.2.2}$$

Notice that we can extend (2.1.6) as

$$k = \frac{2\pi}{\lambda}, \quad v = \lambda\nu = \lambda\,\frac{\omega}{2\pi}. \tag{2.2.3}$$

Combining (2.2.2) and (2.2.3), we have

$$\phi(x,t) = C\mathrm{e}^{\mathrm{i}(kx-\omega t)}. \tag{2.2.4}$$

Momentum and energy as eigenvalues

Formally, in view of (2.2.4), the de Broglie momentum (2.1.9) can be read off as an eigenvalue of the operator $-\mathrm{i}\hbar\frac{\partial}{\partial x}$. That is,

$$\left(-\mathrm{i}\hbar\frac{\partial}{\partial x}\right)\phi = (\hbar\, k)\phi = p\phi, \tag{2.2.5}$$

so that

$$\left(-\mathrm{i}\hbar\frac{\partial}{\partial x}\right)^2\phi = (\hbar\, k)^2\phi = p^2\phi. \tag{2.2.6}$$

Similarly, the de Broglie energy (2.1.8) can be read off as an eigenvalue of the operator $\mathrm{i}\hbar\frac{\partial}{\partial t}$. That is,

$$\left(\mathrm{i}\hbar\frac{\partial}{\partial t}\right)\phi = (\hbar\,\omega)\phi = E\phi. \tag{2.2.7}$$

Schrödinger equation

For a free particle of mass $m > 0$, we know that there holds the classical relation

$$E = \frac{p^2}{2m}. \tag{2.2.8}$$

In view of (2.2.6)–(2.2.8), we arrive at the *free Schrödinger equation*

$$\mathrm{i}\hbar\frac{\partial\phi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\phi}{\partial x^2}. \tag{2.2.9}$$

For a particle moving in a potential field $V = V(x,t)$, the energy-momentum relation (2.2.8) becomes

$$E = \frac{p^2}{2m} + V. \tag{2.2.10}$$

Therefore the Schrödinger equation (2.2.9) for a free particle is modified into form

$$\mathrm{i}\hbar\frac{\partial\phi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\phi}{\partial x^2} + V\phi. \tag{2.2.11}$$

This is called the *Schrödinger wave equation* whose solution, ϕ, is called a *wave function*.

Born's statistical interpretation of wave function

Consider the Schrödinger equation (2.2.11) describing a particle of mass m and assume that ϕ is a "normalized" solution of (2.2.11) which satisfies the normalization condition

$$\int |\phi(x,t)|^2\,\mathrm{d}x = 1, \tag{2.2.12}$$

and characterizes the "state" of the particle. According to Born, the mathematical meaning of such a wave function is that $\rho(x,t) = |\phi(x,t)|^2$ gives the *probability density* for the location of the particle at time t. In other words, the probability of finding the particle in an interval (a,b) at time t is

$$P(\{a < x(t) < b\}) = \int_a^b |\phi(x,t)|^2 \,\mathrm{d}x. \tag{2.2.13}$$

With this interpretation, we see that the expected location of the particle at time t is

$$\langle x\rangle(t) = \int x|\phi(x,t)|^2 \,\mathrm{d}x = \int \overline{\phi}(x,t)\, x\, \phi(x,t)\,\mathrm{d}x. \tag{2.2.14}$$

Operator representations of physical quantities

Naturally, the expected value of the momentum of the particle should be equal to the product of the particle mass and the expected value of the particle velocity. Therefore, in view of (2.2.11), we have

$$\begin{aligned}\langle \hat{p}\rangle(t) &= m\frac{\mathrm{d}\langle x\rangle(t)}{\mathrm{d}t}\\ &= m\int (\overline{\phi}_t(x,t)\, x\, \phi(x,t) + \overline{\phi}(x,t)\, x\, \phi_t(x,t))\,\mathrm{d}x\\ &= \mathrm{i}\,\frac{\hbar}{2}\int (x\overline{\phi}\phi_{xx} - x\overline{\phi}_{xx}\phi)\,\mathrm{d}x\\ &= \int \overline{\phi}(x,t)\Big(-\mathrm{i}\hbar\frac{\partial}{\partial x}\Big)\phi(x,t)\,\mathrm{d}x.\end{aligned} \tag{2.2.15}$$

Hence, formally, the expected value of the momentum is the "expected value" of the operator

$$\hat{p} = -\mathrm{i}\hbar\frac{\partial}{\partial x}. \tag{2.2.16}$$

In other words, within the framework of Born's statistical interpretation of the wave function, momentum has its elegant operator representation (2.2.16).

In this manner, we have the trivial operator representations

$$\hat{x} = x, \quad \hat{f}(x) = f(x), \tag{2.2.17}$$

for the particle coordinate x and its functions. Besides, (2.2.10) gives us the energy representation

$$\hat{E} = \frac{1}{2m}\hat{p}^2 + V, \tag{2.2.18}$$

which is the quantum-mechanical Hamiltonian. Thus,

$$\langle \hat{E}\rangle = \int \overline{\phi}(x,t)\Big(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V\Big)\phi(x,t)\,\mathrm{d}x. \tag{2.2.19}$$

Using (2.2.11) in (2.2.19), we have

$$\langle \hat{E} \rangle = \int \overline{\phi}(x,t) \left(\mathrm{i}\hbar \frac{\partial}{\partial t} \right) \phi(x,t)\, \mathrm{d}x. \tag{2.2.20}$$

In other words, energy should be represented by the operator

$$\hat{E} = \mathrm{i}\hbar \frac{\partial}{\partial t}. \tag{2.2.21}$$

These fundamental *operator* representations of various physical quantities composed from their *classical counterparts*, collectively carried out in a procedure known as the *first quantization*, form the foundation of quantum mechanics.

Conservation law and probability current

It is easily checked that the normalization condition is well posed because

$$\frac{\mathrm{d}}{\mathrm{d}t} \int |\phi(x,t)|^2\, \mathrm{d}x = 0, \tag{2.2.22}$$

by virtue of the equation (2.2.11) so that it suffices to require the condition

$$\int |\phi(x,0)|^2\, \mathrm{d}x = 1, \tag{2.2.23}$$

initially. Here, we look for some additional consequences from the *global conservation law* (2.2.22). For this purpose, we differentiate the probability density ρ to get

$$\begin{aligned} \rho_t &= \overline{\phi}_t \phi + \overline{\phi} \phi_t \\ &= -\mathrm{i} \frac{\hbar}{2m} (\phi \overline{\phi}_x - \overline{\phi} \phi_x)_x, \end{aligned} \tag{2.2.24}$$

where we have used (2.2.11) again. It is interesting to view ρ as a "*charge*" density and rewrite (2.2.24) in the form of a *conservation law*,

$$\frac{\partial}{\partial t} \rho + \frac{\partial}{\partial x} j = 0, \tag{2.2.25}$$

where j may be viewed as a "*current*" density, which is identified to be

$$j = \mathrm{i}\, \frac{\hbar}{2m} (\phi \overline{\phi}_x - \overline{\phi} \phi_x). \tag{2.2.26}$$

We note that it is crucial that ϕ is complex-valued: if it is real-valued, the current density will be zero identically and ρ will be time-independent.

Furthermore, differentiating (2.2.13) and using (2.2.26), we have

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} P(\{a < x(t) < b\}) &= \frac{\mathrm{d}}{\mathrm{d}t} Q(a,b)(t) \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \int_a^b \rho(x,t)\, \mathrm{d}x \\ &= j(a,t) - j(b,t), \end{aligned} \tag{2.2.27}$$

where $Q(a,b)$ may be interpreted as the charge contained in the interval (a,b) at time t so that its rate of change is equal to the net current following into such an interval. Or more correctly, we may call Q the "*probability charge*" and j the "*probability current.*"

Ehrenfest theorem

Differentiating (2.2.15) and using (2.2.11), we have

$$\begin{aligned}\frac{\mathrm{d}\langle\hat{p}\rangle}{\mathrm{d}t} &= \int\left(\overline{\phi}_t\left[-\mathrm{i}\hbar\frac{\partial}{\partial x}\right]\phi+\overline{\phi}\left[-\mathrm{i}\hbar\frac{\partial}{\partial x}\right]\phi_t\right)\mathrm{d}x\\ &= -\mathrm{i}\hbar\int\left(\left[-\mathrm{i}\,\frac{\hbar}{2m}\overline{\phi}_{xx}+\mathrm{i}\,\frac{V}{\hbar}\overline{\phi}\right]\phi_x+\overline{\phi}\frac{\partial}{\partial x}\left[\mathrm{i}\,\frac{\hbar}{2m}\phi_{xx}-\mathrm{i}\,\frac{V}{\hbar}\phi\right]\right)\mathrm{d}x\\ &= -\frac{\hbar^2}{2m}\int([\overline{\phi}_x\phi_x]_x-[\overline{\phi}\phi_{xx}]_x)\,\mathrm{d}x-\int V_x|\phi|^2\,\mathrm{d}x\\ &= -\langle V_x\rangle,\end{aligned}\tag{2.2.28}$$

which may be compared with the equation of motion in the classical Newtonian mechanics,

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2}=\frac{\mathrm{d}p}{\mathrm{d}t}=-V_x.\tag{2.2.29}$$

In other words, in quantum mechanics, in sense of expected value, quantum operators obey the equation of motion of Newtonian mechanics. This statement is known as the *Ehrenfest theorem.*

Complex potential and unstable particles

The profound meaning of the conservation law (2.2.12) is that a particle can never disappear once it is present. Here we show that a small modification may be made so that we are able to describe unstable particles which may disappear after some time lapse. We will not justify whether such a modification is physically correct but will only be content to know that there is room in the Schrödinger equation to accommodate theoretical explorations. To this end, we assume that the potential energy V in (2.2.11) is perturbed by an imaginary quantity,

$$V=V_1+\mathrm{i}V_2,\quad V_1 \text{ and } V_2 \text{ are both real-valued,}\tag{2.2.30}$$

which is allowed in the mathematical setting of the equation. Hence, (2.2.11) becomes

$$\mathrm{i}\hbar\frac{\partial\phi}{\partial t}=-\frac{\hbar^2}{2m}\frac{\partial^2\phi}{\partial x^2}+(V_1+\mathrm{i}V_2)\phi.\tag{2.2.31}$$

In view of (2.2.31), we see that the probability that there is a particle present at time t, that is,

$$P(t)=\int|\phi(x,t)|^2\,\mathrm{d}x,\tag{2.2.32}$$

satisfies the equation

$$P'(t) = \frac{2}{\hbar}\int V_2(x,t)|\phi(x,t)|^2\,\mathrm{d}x. \tag{2.2.33}$$

This clearly indicates that it is the presence of V_2 that breaks down the probability conservation law (2.2.22).

For simplicity, we further assume that there is a constant $K > 0$ such that

$$V_2(x,t) \le -K, \quad \forall x,t. \tag{2.2.34}$$

Then (2.2.33) and (2.2.34) lead us to

$$P'(t) \le -\frac{2K}{\hbar}P(t). \tag{2.2.35}$$

Thus, if a particle is present initially, then $P(0) = 1$. Consequently, we can integrate (2.2.35) to infer that

$$P(t) \le \mathrm{e}^{-\frac{2K}{\hbar}t}, \quad t > 0. \tag{2.2.36}$$

In other words, in a bulk situation, we will observe loss of particles as time elapses, suggesting that we encounter *unstable particles*.

Equation in higher dimensions

Our discussion about the one-dimensional Schrödinger equations can be extended to arbitrarily high dimensions. For this purpose, we consider the spacetime of dimension $(n+1)$ with coordinates $t = x^0, \mathbf{x} = (x^1, \ldots, x^n)$, for time and space, respectively. We use the Greek letters μ, ν, etc., to denote the spacetime indices, $\mu, \nu = 0, 1, \ldots, n$, the Latin letters i, j, k, etc., the space indices, $i, j, k = 1, \ldots, n$, and ∇ the gradient operator on functions depending on $x^1, \ldots, x^n$.

The Schrödinger equation that quantum-mechanically governs a particle of mass m in $\mathbb{R}^n$ is given by

$$\mathrm{i}\hbar\frac{\partial\phi}{\partial t} = -\frac{\hbar^2}{2m}\Delta\phi + V\phi, \tag{2.2.37}$$

since the energy and momentum operators are, respectively, given by

$$\hat{E} = \mathrm{i}\hbar\frac{\partial}{\partial t}, \quad \hat{\mathbf{p}} = -\mathrm{i}\hbar\nabla, \tag{2.2.38}$$

and the total energy operator, or the Hamiltonian, is

$$\hat{H} = \frac{1}{2m}\hat{\mathbf{p}}^2 + V. \tag{2.2.39}$$

Subsequently, the associated probability current $j = (j^\mu) = (j^0, \mathbf{j}) = (\rho, j^i)$ is defined by

$$\rho = |\phi|^2, \quad j^i = \mathrm{i}\,\frac{\hbar}{2m}(\phi\partial_i\overline{\phi} - \overline{\phi}\partial_i\phi), \quad i = 1, \ldots, n, \tag{2.2.40}$$