

Principles of Quantum Mechanics

SECOND EDITION

R. Shankar

Principles of Quantum Mechanics

SECOND EDITION

Principles of Quantum Mechanics

SECOND EDITION

R. Shankar

*Yale University
New Haven, Connecticut*



Springer

Library of Congress Cataloging-in-Publication Data

Shankar, Ramamurti.

Principles of quantum mechanics / R. Shankar. 2nd ed.
p. cm.

Includes bibliographical references and index.
ISBN 0-306-44790-8

1. Quantum theory. I. Title.
QC174. 12.S52 1994

530. 1'2--dc20

94-26837
CIP

ISBN 978-1-4757-0578-2

ISBN 978-1-4757-0576-8 (eBook)

DOI: 10.1007/978-1-4757-0576-8

© 1994, 1980 Springer Science+Business Media, LLC

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America.

19 18 (corrected printing, 2008)

springer.com

To
My Parents
and to
Uma, Umesh, Ajeet, Meera, and Maya

Preface to the Second Edition

Over the decade and a half since I wrote the first edition, nothing has altered my belief in the soundness of the overall approach taken here. This is based on the response of teachers, students, and my own occasional rereading of the book. I was generally quite happy with the book, although there were portions where I felt I could have done better and portions which bothered me by their absence. I welcome this opportunity to rectify all that.

Apart from small improvements scattered over the text, there are three major changes. First, I have rewritten a big chunk of the mathematical introduction in Chapter 1. Next, I have added a discussion of time-reversal invariance. I don't know how it got left out the first time—I wish I could go back and change it. The most important change concerns the inclusion of Chapter 21, "Path Integrals: Part II." The first edition already revealed my partiality for this subject by having a chapter devoted to it, which was quite unusual in those days. In this one, I have cast off all restraint and gone all out to discuss many kinds of path integrals and their uses. Whereas in Chapter 8 the path integral recipe was simply given, here I start by deriving it. I derive the configuration space integral (the usual Feynman integral), phase space integral, and (oscillator) coherent state integral. I discuss two applications: the derivation and application of the Berry phase and a study of the lowest Landau level with an eye on the quantum Hall effect. The relevance of these topics is unquestionable. This is followed by a section of imaginary time path integrals—its description of tunneling, instantons, and symmetry breaking, and its relation to classical and quantum statistical mechanics. An introduction is given to the transfer matrix. Then I discuss spin coherent state path integrals and path integrals for fermions. These were thought to be topics too advanced for a book like this, but I believe this is no longer true. These concepts are extensively used and it seemed a good idea to provide the students who had the wisdom to buy this book with a head start.

How are instructors to deal with this extra chapter given the time constraints? I suggest omitting some material from the earlier chapters. (No one I know, myself included, covers the whole book while teaching any fixed group of students.) A realistic option is for the instructor to teach part of Chapter 21 and assign the rest as reading material, as topics for take-home exams, term papers, etc. To ignore it,

I think, would be to lose a wonderful opportunity to expose the student to ideas that are central to many current research topics and to deny them the attendant excitement. Since the aim of this chapter is to guide students toward more frontline topics, it is more concise than the rest of the book. Students are also expected to consult the references given at the end of the chapter.

Over the years, I have received some very useful feedback and I thank all those students and teachers who took the time to do so. I thank Howard Haber for a discussion of the Born approximation; Harsh Mathur and Ady Stern for discussions of the Berry phase; Alan Chodos, Steve Girvin, Ilya Gruzberg, Martin Gutzwiller, Ganpathy Murthy, Charlie Sommerfeld, and Senthil Todari for many useful comments on Chapter 21. I am most grateful to Captain Richard F. Malm, U.S.C.G. (Retired), Professor Dr. D. Schlüter of the University of Kiel, and Professor V. Yakovenko of the University of Maryland for detecting numerous errors in the first printing and taking the trouble to bring them to my attention. I thank Amelia McNamara of Plenum for urging me to write this edition and Plenum for its years of friendly and warm cooperation. I thank Ron Johnson, Editor at Springer for his tireless efforts on behalf of this book, and Chris Bostock, Daniel Keren and Jimmy Snyder for their generous help in correcting errors in the 14th printing. Finally, I thank my wife Uma for shielding me as usual from real life so I could work on this edition, and my battery of kids (revised and expanded since the previous edition) for continually charging me up.

R. Shankar

New Haven, Connecticut

Preface to the First Edition

Publish and perish—*Giordano Bruno*

Given the number of books that already exist on the subject of quantum mechanics, one would think that the public needs one more as much as it does, say, the latest version of the Table of Integers. But this does not deter me (as it didn't my predecessors) from trying to circulate my own version of how it ought to be taught. The approach to be presented here (to be described in a moment) was first tried on a group of Harvard undergraduates in the summer of '76, once again in the summer of '77, and more recently at Yale on undergraduates ('77-'78) and graduates ('78-'79) taking a year-long course on the subject. In all cases the results were very satisfactory in the sense that the students seemed to have learned the subject well and to have enjoyed the presentation. It is, in fact, their enthusiastic response and encouragement that convinced me of the soundness of my approach and impelled me to write this book.

The basic idea is to develop the subject from its postulates, after addressing some indispensable preliminaries. Now, most people would agree that the best way to teach any subject that has reached the point of development where it can be reduced to a few postulates is to start with the latter, for it is this approach that gives students the fullest understanding of the foundations of the theory and how it is to be used. But they would also argue that whereas this is all right in the case of special relativity or mechanics, a typical student about to learn quantum mechanics seldom has any familiarity with the mathematical language in which the postulates are stated. I agree with these people that this problem is real, but I differ in my belief that it should and can be overcome. This book is an attempt at doing just this.

It begins with a rather lengthy chapter in which the relevant mathematics of vector spaces developed from simple ideas on vectors and matrices the student is assumed to know. The level of rigor is what I think is needed to make a practicing quantum mechanic out of the student. This chapter, which typically takes six to eight lecture hours, is filled with examples from physics to keep students from getting too fidgety while they wait for the “real physics.” Since the math introduced has to be taught sooner or later, I prefer sooner to later, for this way the students, when they get to it, can give quantum theory their fullest attention without having to

battle with the mathematical theorems at the same time. Also, by segregating the mathematical theorems from the physical postulates, any possible confusion as to which is which is nipped in the bud.

This chapter is followed by one on classical mechanics, where the Lagrangian and Hamiltonian formalisms are developed in some depth. It is for the instructor to decide how much of this to cover; the more students know of these matters, the better they will understand the connection between classical and quantum mechanics. Chapter 3 is devoted to a brief study of idealized experiments that betray the inadequacy of classical mechanics and give a glimpse of quantum mechanics.

Having trained and motivated the students I now give them the postulates of quantum mechanics of a single particle in one dimension. I use the word “postulate” here to mean “that which cannot be deduced from pure mathematical or logical reasoning, and given which one can formulate and solve quantum mechanical problems and interpret the results.” This is not the sense in which the true axiomatist would use the word. For instance, where the true axiomatist would just postulate that the dynamical variables are given by Hilbert space operators, I would add the operator identifications, i.e., specify the operators that represent coordinate and momentum (from which others can be built). Likewise, I would not stop with the statement that there is a Hamiltonian operator that governs the time evolution through the equation $i\hbar\partial|\psi\rangle/\partial t = H|\psi\rangle$; I would say the H is obtained from the classical Hamiltonian by substituting for x and p the corresponding operators. While the more general axioms have the virtue of surviving as we progress to systems of more degrees of freedom, with or without classical counterparts, students given just these will not know how to calculate anything such as the spectrum of the oscillator. Now one can, of course, try to “derive” these operator assignments, but to do so one would have to appeal to ideas of a postulatory nature themselves. (The same goes for “deriving” the Schrödinger equation.) As we go along, these postulates are generalized to more degrees of freedom and it is for pedagogical reasons that these generalizations are postponed. Perhaps when students are finished with this book, they can free themselves from the specific operator assignments and think of quantum mechanics as a general mathematical formalism obeying certain postulates (in the strict sense of the term).

The postulates in Chapter 4 are followed by a lengthy discussion of the same, with many examples from fictitious Hilbert spaces of three dimensions. Nonetheless, students will find it hard. It is only as they go along and see these postulates used over and over again in the rest of the book, in the setting up of problems and the interpretation of the results, that they will catch on to how the game is played. It is hoped they will be able to do it on their own when they graduate. I think that any attempt to soften this initial blow will be counterproductive in the long run.

Chapter 5 deals with standard problems in one dimension. It is worth mentioning that the scattering off a step potential is treated using a wave packet approach. If the subject seems too hard at this stage, the instructor may decide to return to it after Chapter 7 (oscillator), when students have gained more experience. But I think that sooner or later students must get acquainted with this treatment of scattering.

The classical limit is the subject of the next chapter. The harmonic oscillator is discussed in detail in the next. It is the first realistic problem and the instructor may be eager to get to it as soon as possible. If the instructor wants, he or she can discuss the classical limit after discussing the oscillator.

We next discuss the path integral formulation due to Feynman. Given the intuitive understanding it provides, and its elegance (not to mention its ability to give the full propagator in just a few minutes in a class of problems), its omission from so many books is hard to understand. While it is admittedly hard to actually evaluate a path integral (one example is provided here), the notion of expressing the propagator as a sum over amplitudes from various paths is rather simple. The importance of this point of view is becoming clearer day by day to workers in statistical mechanics and field theory. I think every effort should be made to include at least the first three (and possibly five) sections of this chapter in the course.

The content of the remaining chapters is standard, in the first approximation. The style is of course peculiar to this author, as are the specific topics. For instance, an entire chapter (11) is devoted to symmetries and their consequences. The chapter on the hydrogen atom also contains a section on how to make numerical estimates starting with a few mnemonics. Chapter 15, on addition of angular momenta, also contains a section on how to understand the “accidental” degeneracies in the spectra of hydrogen and the isotropic oscillator. The quantization of the radiation field is discussed in Chapter 18, on time-dependent perturbation theory. Finally the treatment of the Dirac equation in the last chapter (20) is intended to show that several things such as electron spin, its magnetic moment, the spin-orbit interaction, etc. which were introduced in an ad hoc fashion in earlier chapters, emerge as a coherent whole from the Dirac equation, and also to give students a glimpse of what lies ahead. This chapter also explains how Feynman resolves the problem of negative-energy solutions (in a way that applies to bosons and fermions).

For Whom Is this Book Intended?

In writing it, I addressed students who are trying to learn the subject by themselves; that is to say, I made it as self-contained as possible, included a lot of exercises and answers to most of them, and discussed several tricky points that trouble students when they learn the subject. But I am aware that in practice it is most likely to be used as a class text. There is enough material here for a full year graduate course. It is, however, quite easy to adapt it to a year-long undergraduate course. Several sections that may be omitted without loss of continuity are indicated. The sequence of topics may also be changed, as stated earlier in this preface. I thought it best to let the instructor skim through the book and chart the course for his or her class, given their level of preparation and objectives. Of course the book will not be particularly useful if the instructor is not sympathetic to the broad philosophy espoused here, namely, that first comes the mathematical training and then the development of the subject from the postulates. To instructors who feel that this approach is all right in principle but will not work in practice, I reiterate that it has been found to work in practice, not just by me but also by teachers elsewhere.

The book may be used by nonphysicists as well. (I have found that it goes well with chemistry majors in my classes.) *Although I wrote it for students with no familiarity with the subject, any previous exposure can only be advantageous.*

Finally, I invite instructors and students alike to communicate to me any suggestions for improvement, whether they be pedagogical or in reference to errors or misprints.

Acknowledgments

As I look back to see who all made this book possible, my thoughts first turn to my brother R. Rajaraman and friend Rajaram Nityananda, who, around the same time, introduced me to physics in general and quantum mechanics in particular. Next come my students, particularly Doug Stone, but for whose encouragement and enthusiastic response I would not have undertaken this project. I am grateful to Professor Julius Kovacs of Michigan State, whose kind words of encouragement assured me that the book would be as well received by my peers as it was by my students. More recently, I have profited from numerous conversations with my colleagues at Yale, in particular Alan Chodos and Peter Mohr. My special thanks go to Charles Sommerfield, who managed to make time to read the manuscript and made many useful comments and recommendations. The detailed proofreading was done by Tom Moore. I thank you, the reader, in advance, for drawing to my notice any errors that may have slipped past us.

The bulk of the manuscript production cost were borne by the J. W. Gibbs fellowship from Yale, which also supported me during the time the book was being written. Ms. Laurie Liptak did a fantastic job of typing the first 18 chapters and Ms. Linda Ford did the same with Chapters 19 and 20. The figures are by Mr. J. Brosious. Mr. R. Badrinath kindly helped with the index.[‡]

On the domestic front, encouragement came from my parents, my in-laws, and most important of all from my wife, Uma, who cheerfully donated me to science for a year or so and stood by me throughout. Little Umesh did his bit by tearing up all my books on the subject, both as a show of support and to create a need for this one.

R. Shankar

New Haven, Connecticut

[‡] It is a pleasure to acknowledge the help of Mr. Richard Hatch, who drew my attention to a number of errors in the first printing.

Prelude

Our description of the physical world is dynamic in nature and undergoes frequent change. At any given time, we summarize our knowledge of natural phenomena by means of certain laws. These laws adequately describe the phenomenon studied up to that time, to an accuracy then attainable. As time passes, we enlarge the domain of observation and improve the accuracy of measurement. As we do so, we constantly check to see if the laws continue to be valid. Those laws that do remain valid gain in stature, and those that do not must be abandoned in favor of new ones that do.

In this changing picture, the laws of classical mechanics formulated by Galileo, Newton, and later by Euler, Lagrange, Hamilton, Jacobi, and others, remained unaltered for almost three centuries. The expanding domain of classical physics met its first obstacles around the beginning of this century. The obstruction came on two fronts: at large velocities and small (atomic) scales. The problem of large velocities was successfully solved by Einstein, who gave us his relativistic mechanics, while the founders of quantum mechanics—Bohr, Heisenberg, Schrödinger, Dirac, Born, and others—solved the problem of small-scale physics. The union of relativity and quantum mechanics, needed for the description of phenomena involving simultaneously large velocities and small scales, turns out to be very difficult. Although much progress has been made in this subject, called quantum field theory, there remain many open questions to this date. We shall concentrate here on just the small-scale problem, that is to say, on non-relativistic quantum mechanics.

The passage from classical to quantum mechanics has several features that are common to all such transitions in which an old theory gives way to a new one:

- (1) There is a domain D_n of phenomena described by the new theory and a subdomain D_o wherein the old theory is reliable (to a given accuracy).
- (2) Within the subdomain D_o either theory may be used to make quantitative predictions. It might often be more expedient to employ the old theory.
- (3) In addition to numerical accuracy, the new theory often brings about radical conceptual changes. Being of a qualitative nature, these will have a bearing on all of D_n .

For example, in the case of relativity, D_o and D_n represent (macroscopic) phenomena involving small and arbitrary velocities, respectively, the latter, of course,

being bounded by the velocity of light. In addition to giving better numerical predictions for high-velocity phenomena, relativity theory also outlaws several cherished notions of the Newtonian scheme, such as absolute time, absolute length, unlimited velocities for particles, etc.

In a similar manner, quantum mechanics brings with it not only improved numerical predictions for the microscopic world, but also conceptual changes that rock the very foundations of classical thought.

This book introduces you to this subject, starting from its postulates. Between you and the postulates there stand three chapters wherein you will find a summary of the mathematical ideas appearing in the statement of the postulates, a review of classical mechanics, and a brief description of the empirical basis for the quantum theory. In the rest of the book, the postulates are invoked to formulate and solve a variety of quantum mechanical problems. It is hoped that, by the time you get to the end of the book, you will be able to do the same yourself.

Note to the Student

Do as many exercises as you can, especially the ones marked * or whose results carry equation numbers. The answer to each exercise is given either with the exercise or at the end of the book.

The first chapter is very important. Do not rush through it. Even if you know the math, read it to get acquainted with the notation.

I am not saying it is an easy subject. But I hope this book makes it seem reasonable.

Good luck.

Contents

1. Mathematical Introduction	1
1.1. Linear Vector Spaces: Basics	1
1.2. Inner Product Spaces	7
1.3. Dual Spaces and the Dirac Notation	11
1.4. Subspaces	17
1.5. Linear Operators	18
1.6. Matrix Elements of Linear Operators	20
1.7. Active and Passive Transformations	29
1.8. The Eigenvalue Problem	30
1.9. Functions of Operators and Related Concepts	54
1.10. Generalization to Infinite Dimensions	57
2. Review of Classical Mechanics	75
2.1. The Principle of Least Action and Lagrangian Mechanics	78
2.2. The Electromagnetic Lagrangian	83
2.3. The Two-Body Problem	85
2.4. How Smart Is a Particle?	86
2.5. The Hamiltonian Formalism	86
2.6. The Electromagnetic Force in the Hamiltonian Scheme	90
2.7. Cyclic Coordinates, Poisson Brackets, and Canonical Transformations	91
2.8. Symmetries and Their Consequences	98
3. All Is Not Well with Classical Mechanics	107
3.1. Particles and Waves in Classical Physics	107
3.2. An Experiment with Waves and Particles (Classical)	108
3.3. The Double-Slit Experiment with Light	110
3.4. Matter Waves (de Broglie Waves)	112
3.5. Conclusions	112

4. The Postulates—a General Discussion	115
4.1. The Postulates	115
4.2. Discussion of Postulates I–III	116
4.3. The Schrödinger Equation (Dotting Your <i>i</i> 's and Crossing your \hbar 's)	143
5. Simple Problems in One Dimension	151
5.1. The Free Particle	151
5.2. The Particle in a Box	157
5.3. The Continuity Equation for Probability	164
5.4. The Single-Step Potential: a Problem in Scattering	167
5.5. The Double-Slit Experiment	175
5.6. Some Theorems	176
6. The Classical Limit	179
7. The Harmonic Oscillator	185
7.1. Why Study the Harmonic Oscillator?	185
7.2. Review of the Classical Oscillator	188
7.3. Quantization of the Oscillator (Coordinate Basis)	189
7.4. The Oscillator in the Energy Basis	202
7.5. Passage from the Energy Basis to the <i>X</i> Basis	216
8. The Path Integral Formulation of Quantum Theory	223
8.1. The Path Integral Recipe	223
8.2. Analysis of the Recipe	224
8.3. An Approximation to $U(t)$ for the Free Particle	225
8.4. Path Integral Evaluation of the Free-Particle Propagator	226
8.5. Equivalence to the Schrödinger Equation	229
8.6. Potentials of the Form $V=a+bx+cx^2+d\dot{x}+e\dot{x}\dot{x}$	231
9. The Heisenberg Uncertainty Relations	237
9.1. Introduction	237
9.2. Derivation of the Uncertainty Relations	237
9.3. The Minimum Uncertainty Packet	239
9.4. Applications of the Uncertainty Principle	241
9.5. The Energy–Time Uncertainty Relation	245
10. Systems with <i>N</i> Degrees of Freedom	247
10.1. <i>N</i> Particles in One Dimension	247
10.2. More Particles in More Dimensions	259
10.3. Identical Particles	260

11. Symmetries and Their Consequences	279
11.1. Overview	279
11.2. Translational Invariance in Quantum Theory	279
11.3. Time Translational Invariance	294
11.4. Parity Invariance	297
11.5. Time-Reversal Symmetry	301
12. Rotational Invariance and Angular Momentum	305
12.1. Translations in Two Dimensions	305
12.2. Rotations in Two Dimensions	306
12.3. The Eigenvalue Problem of L_z	313
12.4. Angular Momentum in Three Dimensions	318
12.5. The Eigenvalue Problem of L^2 and L_z	321
12.6. Solution of Rotationally Invariant Problems	339
13. The Hydrogen Atom	353
13.1. The Eigenvalue Problem	353
13.2. The Degeneracy of the Hydrogen Spectrum	359
13.3. Numerical Estimates and Comparison with Experiment	361
13.4. Multielectron Atoms and the Periodic Table	369
14. Spin	373
14.1. Introduction	373
14.2. What is the Nature of Spin?	373
14.3. Kinematics of Spin	374
14.4. Spin Dynamics	385
14.5. Return of Orbital Degrees of Freedom	397
15. Addition of Angular Momenta	403
15.1. A Simple Example	403
15.2. The General Problem	408
15.3. Irreducible Tensor Operators	416
15.4. Explanation of Some “Accidental” Degeneracies	421
16. Variational and WKB Methods	429
16.1. The Variational Method	429
16.2. The Wentzel–Kramers–Brillouin Method	435
17. Time-Independent Perturbation Theory	451
17.1. The Formalism	451
17.2. Some Examples	454
17.3. Degenerate Perturbation Theory	464

18. Time-Dependent Perturbation Theory	473
18.1. The Problem	473
18.2. First-Order Perturbation Theory	474
18.3. Higher Orders in Perturbation Theory	484
18.4. A General Discussion of Electromagnetic Interactions	492
18.5. Interaction of Atoms with Electromagnetic Radiation	499
19. Scattering Theory	523
19.1. Introduction	523
19.2. Recapitulation of One-Dimensional Scattering and Overview	524
19.3. The Born Approximation (Time-Dependent Description)	529
19.4. Born Again (The Time-Independent Approximation)	534
19.5. The Partial Wave Expansion	545
19.6. Two-Particle Scattering	555
20. The Dirac Equation	563
20.1. The Free-Particle Dirac Equation	563
20.2. Electromagnetic Interaction of the Dirac Particle	566
20.3. More on Relativistic Quantum Mechanics	574
21. Path Integrals—II	581
21.1. Derivation of the Path Integral	582
21.2. Imaginary Time Formalism	613
21.3. Spin and Fermion Path Integrals	636
21.4. Summary	652
Appendix	655
A.1. Matrix Inversion	655
A.2. Gaussian Integrals	659
A.3. Complex Numbers	660
A.4. The $i\epsilon$ Prescription	661
ANSWERS TO SELECTED EXERCISES	665
TABLE OF CONSTANTS	669
INDEX	671

1

Mathematical Introduction

The aim of this book is to provide you with an introduction to quantum mechanics, starting from its axioms. It is the aim of this chapter to equip you with the necessary mathematical machinery. All the math you will need is developed here, starting from some basic ideas on vectors and matrices that you are assumed to know. Numerous examples and exercises related to classical mechanics are given, both to provide some relief from the math and to demonstrate the wide applicability of the ideas developed here. The effort you put into this chapter will be well worth your while: not only will it prepare you for this course, but it will also unify many ideas you may have learned piecemeal. To really learn this chapter, you must, as with any other chapter, work out the problems.

1.1. Linear Vector Spaces: Basics

In this section you will be introduced to *linear vector spaces*. You are surely familiar with the arrows from elementary physics encoding the magnitude and direction of velocity, force, displacement, torque, etc. You know how to add them and multiply them by scalars and the rules obeyed by these operations. For example, you know that scalar multiplication is distributive: the multiple of a sum of two vectors is the sum of the multiples. What we want to do is abstract from this simple case a set of basic features or axioms, and say that any set of objects obeying the same forms a linear vector space. The cleverness lies in deciding which of the properties to keep in the generalization. If you keep too many, there will be no other examples; if you keep too few, there will be no interesting results to develop from the axioms.

The following is the list of properties the mathematicians have wisely chosen as requisite for a vector space. As you read them, please compare them to the world of arrows and make sure that these are indeed properties possessed by these familiar vectors. But note also that conspicuously missing are the requirements that every vector have a magnitude and direction, which was the first and most salient feature drilled into our heads when we first heard about them. So you might think that in dropping this requirement, the baby has been thrown out with the bath water. However, you will have ample time to appreciate the wisdom behind this choice as

you go along and see a great unification and synthesis of diverse ideas under the heading of vector spaces. You will see examples of vector spaces that involve entities that you cannot intuitively perceive as having either a magnitude or a direction. While you should be duly impressed with all this, remember that it does not hurt at all to think of these generalizations in terms of arrows and to use the intuition to prove theorems or at the very least anticipate them.

Definition 1. A linear vector space \mathbb{V} is a collection of objects $|1\rangle, |2\rangle, \dots, |V\rangle, \dots, |W\rangle, \dots$, called vectors, for which there exists

1. A definite rule for forming the vector sum, denoted $|V\rangle + |W\rangle$
2. A definite rule for multiplication by scalars a, b, \dots , denoted $a|V\rangle$ with the following features:

- The result of these operations is another element of the space, a feature called *closure*: $|V\rangle + |W\rangle \in \mathbb{V}$.
- Scalar multiplication is *distributive in the vectors*: $a(|V\rangle + |W\rangle) = a|V\rangle + a|W\rangle$.
- Scalar multiplication is *distributive in the scalars*: $(a+b)|V\rangle = a|V\rangle + b|V\rangle$.
- Scalar multiplication is *associative*: $a(b|V\rangle) = ab|V\rangle$.
- Addition is *commutative*: $|V\rangle + |W\rangle = |W\rangle + |V\rangle$.
- Addition is *associative*: $|V\rangle + (|W\rangle + |Z\rangle) = (|V\rangle + |W\rangle) + |Z\rangle$.
- There exists a *null vector* $|0\rangle$ obeying $|V\rangle + |0\rangle = |V\rangle$.
- For every vector $|V\rangle$ there exists an *inverse under addition*, $|-V\rangle$, such that $|V\rangle + |-V\rangle = |0\rangle$.

There is a good way to remember all of these; *do what comes naturally*.

Definition 2. The numbers a, b, \dots are called the *field* over which the vector space is defined.

If the field consists of all real numbers, we have a *real vector space*, if they are complex, we have a *complex vector space*. The vectors themselves are neither real nor complex; the adjective applies only to the scalars.

Let us note that the above axioms imply

- $|0\rangle$ is unique, i.e., if $|0'\rangle$ has all the properties of $|0\rangle$, then $|0\rangle = |0'\rangle$.
- $|0|V\rangle = |0\rangle$.
- $|-V\rangle = -|V\rangle$.
- $|-V\rangle$ is the unique additive inverse of $|V\rangle$.

The proofs are left as to the following exercise. You don't have to know the proofs, but you do have to know the statements.

Exercise 1.1.1. Verify these claims. For the first consider $|0\rangle + |0'\rangle$ and use the advertised properties of the two null vectors in turn. For the second start with $|0\rangle = (0+1)|V\rangle + |-V\rangle$. For the third, begin with $|V\rangle + (-|V\rangle) = 0|V\rangle = |0\rangle$. For the last, let $|W\rangle$ also satisfy $|V\rangle + |W\rangle = |0\rangle$. Since $|0\rangle$ is unique, this means $|V\rangle + |W\rangle = |V\rangle + |-V\rangle$. Take it from here.

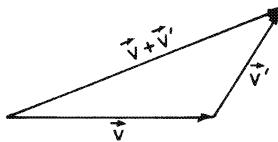


Figure 1.1. The rule for vector addition. Note that it obeys axioms (i)–(iii).

Exercise 1.1.2. Consider the set of all entities of the form (a, b, c) where the entries are real numbers. Addition and scalar multiplication are defined as follows:

$$(a, b, c) + (d, e, f) = (a+d, b+e, c+f)$$

$$\alpha(a, b, c) = (\alpha a, \alpha b, \alpha c).$$

Write down the null vector and inverse of (a, b, c) . Show that vectors of the form $(a, b, 1)$ do not form a vector space.

Observe that we are using a new symbol $|V\rangle$ to denote a generic vector. This object is called *ket V* and this nomenclature is due to Dirac whose notation will be discussed at some length later. We do not purposely use the symbol \vec{V} to denote the vectors as the first step in weaning you away from the limited concept of the vector as an arrow. You are however not discouraged from associating with $|V\rangle$ the arrowlike object till you have seen enough vectors that are not arrows and are ready to drop the crutch.

You were asked to verify that the set of arrows qualified as a vector space as you read the axioms. Here are some of the key ideas you should have gone over. The vector space consists of arrows, typical ones being \vec{V} and \vec{V}' . The rule for addition is familiar: take the tail of the second arrow, put it on the tip of the first, and so on as in Fig. 1.1.

Scalar multiplication by a corresponds to stretching the vector by a factor a . This is a real vector space since stretching by a complex number makes no sense. (If a is negative, we interpret it as changing the direction of the arrow as well as rescaling it by $|a|$.) Since these operations acting on arrows give more arrows, we have closure. Addition and scalar multiplication clearly have all the desired associative and distributive features. The null vector is the arrow of zero length, while the inverse of a vector is the vector reversed in direction.

So the set of all arrows qualifies as a vector space. But we cannot tamper with it. For example, the set of all arrows with positive z -components do not form a vector space: there is no inverse.

Note that so far, no reference has been made to magnitude or direction. The point is that while the arrows have these qualities, members of a vector space need not. This statement is pointless unless I can give you examples, so here are two.

Consider the set of all 2×2 matrices. We know how to add them and multiply them by scalars (multiply all four matrix elements by that scalar). The corresponding rules obey closure, associativity, and distributive requirements. The null matrix has all zeros in it and the inverse under *addition* of a matrix is the matrix with all elements negated. You must agree that here we have a genuine vector space consisting of things which don't have an obvious length or direction associated with them. When we want to highlight the fact that the matrix M is an element of a vector space, we may want to refer to it as, say, ket number 4 or: $|4\rangle$.

As a second example, consider all functions $f(x)$ defined in an interval $0 \leq x \leq L$. We define scalar multiplication by a simply as $af(x)$ and addition as pointwise addition: the sum of two functions f and g has the value $f(x) + g(x)$ at the point x . The null function is zero everywhere and the additive inverse of f is $-f$.

Exercise 1.1.3. Do functions that vanish at the end points $x=0$ and $x=L$ form a vector space? How about *periodic functions* obeying $f(0)=f(L)$? How about functions that obey $f(0)=4$? If the functions do not qualify, list the things that go wrong.

The next concept is that of *linear independence* of a set of vectors $|1\rangle, |2\rangle \dots |n\rangle$. First consider a linear relation of the form

$$\sum_{i=1}^n a_i|i\rangle = |0\rangle \quad (1.1.1)$$

We may assume without loss of generality that the left-hand side does not contain any multiple of $|0\rangle$, for if it did, it could be shifted to the right, and combined with the $|0\rangle$ there to give $|0\rangle$ once more. (We are using the fact that any multiple of $|0\rangle$ equals $|0\rangle$.)

Definition 3. The set of vectors is said to be *linearly independent* if the only such linear relation as Eq. (1.1.1) is the trivial one with all $a_i=0$. If the set of vectors is not linearly independent, we say they are *linearly dependent*.

Equation (1.1.1) tells us that it is not possible to write any member of the linearly independent set in terms of the others. On the other hand, if the set of vectors is linearly dependent, such a relation will exist, and it must contain at least two nonzero coefficients. Let us say $a_3 \neq 0$. Then we could write

$$|3\rangle = \sum_{i=1, i \neq 3}^n \frac{-a_i}{a_3} |i\rangle \quad (1.1.2)$$

thereby expressing $|3\rangle$ in terms of the others.

As a concrete example, consider two nonparallel vectors $|1\rangle$ and $|2\rangle$ in a plane. These form a linearly independent set. There is no way to write one as a multiple of the other, or equivalently, no way to combine them to get the null vector. On the other hand, if the vectors are parallel, we can clearly write one as a multiple of the other or equivalently play them against each other to get 0.

Notice I said 0 and not $|0\rangle$. This is, strictly speaking, incorrect since a set of vectors can only add up to a vector and not a number. It is, however, common to represent the null vector by 0.

Suppose we bring in a third vector $|3\rangle$ also in the plane. If it is parallel to either of the first two, we already have a linearly dependent set. So let us suppose it is not. But even now the three of them are *linearly dependent*. This is because we can write one of them, say $|3\rangle$, as a linear combination of the other two. To find the combination, draw a line from the tail of $|3\rangle$ in the direction of $|1\rangle$. Next draw a line antiparallel to $|2\rangle$ from the tip of $|3\rangle$. These lines will intersect since $|1\rangle$ and $|2\rangle$ are

not parallel by assumption. The intersection point P will determine how much of $|1\rangle$ and $|2\rangle$ we want: we go from the tail of $|3\rangle$ to P using the appropriate multiple of $|1\rangle$ and go from P to the tip of $|3\rangle$ using the appropriate multiple of $|2\rangle$.

Exercise 1.1.4. Consider three elements from the vector space of real 2×2 matrices:

$$|1\rangle = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad |2\rangle = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad |3\rangle = \begin{bmatrix} -2 & -1 \\ 0 & -2 \end{bmatrix}$$

Are they linearly independent? Support your answer with details. (Notice we are calling these matrices vectors and using kets to represent them to emphasize their role as elements of a vector space.)

Exercise 1.1.5. Show that the following row vectors are linearly dependent: $(1, 1, 0)$, $(1, 0, 1)$, and $(3, 2, 1)$. Show the opposite for $(1, 1, 0)$, $(1, 0, 1)$, and $(0, 1, 1)$.

Definition 4. A vector space has *dimension n* if it can accommodate a maximum of n linearly independent vectors. It will be denoted by $\mathbb{V}^n(R)$ if the field is real and by $\mathbb{V}^n(C)$ if the field is complex.

In view of the earlier discussions, the plane is two-dimensional and the set of all arrows not limited to the plane define a three-dimensional vector space. How about 2×2 matrices? They form a four-dimensional vector space. Here is a proof. The following vectors are linearly independent:

$$|1\rangle = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad |2\rangle = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad |3\rangle = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \quad |4\rangle = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

since it is impossible to form linear combinations of any three of them to give the fourth any three of them will have a zero in the one place where the fourth does not. So the space is at least four-dimensional. Could it be bigger? No, since any arbitrary 2×2 matrix can be written in terms of them:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = a|1\rangle + b|2\rangle + c|3\rangle + d|4\rangle$$

If the scalars a, b, c, d are real, we have a *real four-dimensional space*, if they are complex we have a *complex four-dimensional space*.

Theorem 1. Any vector $|V\rangle$ in an n -dimensional space can be written as a linear combination of n linearly independent vectors $|1\rangle \dots |n\rangle$.

The proof is as follows: if there were a vector $|V\rangle$ for which this were not possible, it would join the given set of vectors and form a set of $n+1$ linearly independent vectors, which is not possible in an n -dimensional space by definition.

Definition 5. A set of n linearly independent vectors in an n -dimensional space is called a *basis*.

Thus we can write, on the strength of the above

$$|V\rangle = \sum_{i=1}^n v_i |i\rangle \quad (1.1.3)$$

where the vectors $|i\rangle$ form a basis.

Definition 6. The coefficients of expansion v_i of a vector in terms of a linearly independent basis ($|i\rangle$) are called the *components of the vector in that basis*.

Theorem 2. The expansion in Eq. (1.1.3) is unique.

Suppose the expansion is not unique. We must then have a second expansion:

$$|V\rangle = \sum_{i=1}^n v'_i |i\rangle \quad (1.1.4)$$

Subtracting Eq. (1.1.4) from Eq. (1.1.3) (i.e., multiplying the second by the scalar -1 and adding the two equations) we get

$$|0\rangle = \sum_i (v_i - v'_i) |i\rangle \quad (1.1.5)$$

which implies that

$$v_i = v'_i \quad (1.1.6)$$

since the basis vectors are linearly independent and only a trivial linear relation between them can exist. Note that given a basis the components are unique, but if we change the basis, the components will change. We refer to $|V\rangle$ as the vector in the abstract, having an existence of its own and satisfying various relations involving other vectors. When we choose a basis the vectors assume concrete forms in terms of their components and the relation between vectors is satisfied by the components. Imagine for example three arrows in the plane, \vec{A} , \vec{B} , \vec{C} satisfying $\vec{A} + \vec{B} = \vec{C}$ according to the laws for adding arrows. So far no basis has been chosen and we do not need a basis to make the statement that the vectors from a closed triangle. Now we choose a basis and write each vector in terms of the components. The components will satisfy $C_i = A_i + B_i$, $i = 1, 2$. If we choose a different basis, the components will change in numerical value, but the relation between them expressing the equality of \vec{C} to the sum of the other two will still hold between the new set of components.

In the case of nonarrow vectors, adding them in terms of components proceeds as in the elementary case thanks to the axioms. If

$$|V\rangle = \sum_i v_i |i\rangle \quad \text{and} \quad (1.1.7)$$

$$|W\rangle = \sum_i w_i |i\rangle \quad \text{then} \quad (1.1.8)$$

$$|V\rangle + |W\rangle = \sum_i (v_i + w_i) |i\rangle \quad (1.1.9)$$

where we have used the axioms to carry out the regrouping of terms. Here is the conclusion:

To add two vectors, add their components.

There is no reference to taking the tail of one and putting it on the tip of the other, etc., since in general the vectors have no head or tail. Of course, if we are dealing with arrows, we can add them either using the tail and tip routine or by simply adding their components in a basis.

In the same way, we have:

$$a|V\rangle = a \sum_i v_i |i\rangle = \sum_i av_i |i\rangle \quad (1.1.10)$$

In other words,

To multiply a vector by a scalar, multiply all its components by the scalar.

1.2. Inner Product Spaces

The matrix and function examples must have convinced you that we can have a vector space with no preassigned definition of length or direction for the elements. However, we can make up quantities that have the same properties that the lengths and angles do in the case of arrows. The first step is to define a sensible analog of the dot product, for in the case of arrows, from the dot product

$$\vec{A} \cdot \vec{B} = |A||B| \cos \theta \quad (1.2.1)$$

we can read off the length of say \vec{A} as $\sqrt{|A| \cdot |A|}$ and the cosine of the angle between two vectors as $\vec{A} \cdot \vec{B} / |A||B|$. Now you might rightfully object: how can you use the dot product to define the length and angles, if the dot product itself requires knowledge of the lengths and angles? The answer is this. Recall that the dot product has a second

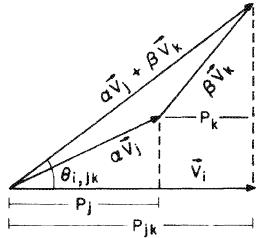


Figure 1.2. Geometrical proof that the dot product obeys axiom (3) for an inner product. The axiom requires that the projections obey $P_k + P_j = P_{jk}$.

equivalent expression in terms of the components:

$$\vec{A} \cdot \vec{B} = A_x B_x + A_y B_y + A_z B_z \quad (1.2.2)$$

Our goal is to define a similar formula for the general case where we *do* have the notion of components in a basis. To this end we recall the main features of the above dot product :

1. $\vec{A} \cdot \vec{B} = \vec{B} \cdot \vec{A}$ (symmetry)
2. $\vec{A} \cdot \vec{A} \geq 0 \quad 0 \text{ iff } \vec{A} = 0$ (positive semidefiniteness)
3. $\vec{A} \cdot (b\vec{B} + c\vec{C}) = b\vec{A} \cdot \vec{B} + c\vec{A} \cdot \vec{C}$ (linearity)

The linearity of the dot product is illustrated in Fig. 1.2.

We want to invent a generalization called the *inner product* or *scalar product* between any two vectors $|V\rangle$ and $|W\rangle$. We denote it by the symbol $\langle V|W\rangle$. It is once again a number (generally complex) dependent on the two vectors. We demand that it obey the following axioms:

- $\langle V|W\rangle = \langle W|V\rangle^*$ (skew-symmetry)
- $\langle V|V\rangle \geq 0 \quad 0 \text{ iff } |V\rangle = |0\rangle$ (positive semidefiniteness)
- $\langle V|(a|W\rangle + b|Z\rangle) \equiv \langle V|aW + bZ\rangle = a\langle V|W\rangle + b\langle V|Z\rangle$ (linearity in ket)

Definition 7. A vector space with an inner product is called an *inner product space*.

Notice that we have not yet given an explicit rule for actually evaluating the scalar product, we are merely demanding that any rule we come up with must have these properties. With a view to finding such a rule, let us familiarize ourselves with the axioms. The first differs from the corresponding one for the dot product and makes the inner product sensitive to the order of the two factors, with the two choices leading to complex conjugates. In a real vector space this axioms states the symmetry of the dot product under exchange of the two vectors. For the present, let us note that this axiom ensures that $\langle V|V\rangle$ is real.

The second axiom says that $\langle V|V\rangle$ is not just real but also positive semidefinite, vanishing only if the vector itself does. If we are going to define the length of the vector as the square root of its inner product with itself (as in the dot product) this quantity had better be real and positive for all nonzero vectors.

The last axiom expresses the linearity of the inner product when a linear superposition $a|W\rangle + b|Z\rangle \equiv |aW+bZ\rangle$ appears as the second vector in the scalar product. We have discussed its validity for the arrows case (Fig. 1.2).

What if the first factor in the product is a linear superposition, i.e., what is $\langle aW+bZ|V\rangle$? This is determined by the first axiom:

$$\begin{aligned}\langle aW+bZ|V\rangle &= \langle V|aW+bZ\rangle^* \\ &= (a\langle V|W\rangle + b\langle V|Z\rangle)^* \\ &= a^*\langle V|W\rangle^* + b^*\langle V|Z\rangle^* \\ &= a^*\langle W|V\rangle + b^*\langle Z|V\rangle\end{aligned}\tag{1.2.3}$$

which expresses the *antilinearity* of the inner product with respect to the first factor in the inner product. In other words, the inner product of a linear superposition with another vector is the corresponding superposition of inner products if the superposition occurs in the second factor, while it is the superposition with all coefficients conjugated if the superposition occurs in the first factor. This asymmetry, unfamiliar in real vector spaces, is here to stay and you will get used to it as you go along.

Let us continue with inner products. Even though we are trying to shed the restricted notion of a vector as an arrow and seeking a corresponding generalization of the dot product, we still use some of the same terminology.

Definition 8. We say that two vectors are *orthogonal* or perpendicular if their inner product vanishes.

Definition 9. We will refer to $\sqrt{\langle V|V\rangle} \equiv |V|$ as the *norm* or length of the vector. A *normalized vector* has unit norm.

Definition 10. A set of basis vectors all of unit norm, which are pairwise orthogonal will be called an *orthonormal basis*.

We will also frequently refer to the inner or scalar product as the dot product.

We are now ready to obtain a concrete formula for the inner product in terms of the components. Given $|V\rangle$ and $|W\rangle$

$$\begin{aligned}|V\rangle &= \sum_i v_i|i\rangle \\ |W\rangle &= \sum_j w_j|j\rangle\end{aligned}$$

we follow the axioms obeyed by the inner product to obtain:

$$\langle V|W\rangle = \sum_i \sum_j v_i^* w_j \langle i|j\rangle\tag{1.2.4}$$

To go any further we have to know $\langle i|j\rangle$, the inner product between basis vectors. That depends on the details of the basis vectors and all we know for sure is that

they are linearly independent. This situation exists for arrows as well. Consider a two-dimensional problem where the basis vectors are two linearly independent but nonperpendicular vectors. If we write all vectors in terms of this basis, the dot product of any two of them will likewise be a double sum with four terms (determined by the four possible dot products between the basis vectors) as well as the vector components. However, if we use an orthonormal basis such as \hat{i}, \hat{j} , only diagonal terms like $\langle i | i \rangle$ will survive and we will get the familiar result $\hat{A} \cdot \hat{B} = A_x B_x + A_y B_y$ depending only on the components.

For the more general nonarrow case, we invoke Theorem 3.

Theorem 3 (Gram-Schmidt). Given a linearly independent basis we can form linear combinations of the basis vectors to obtain an orthonormal basis.

Postponing the proof for a moment, let us assume that the procedure has been implemented and that the current basis is orthonormal:

$$\langle i | j \rangle = \begin{cases} 1 & \text{for } i=j \\ 0 & \text{for } i \neq j \end{cases} \equiv \delta_{ij}$$

where δ_{ij} is called the *Kronecker delta symbol*. Feeding this into Eq. (1.2.4) we find the double sum collapses to a single one due to the Kronecker delta, to give

$$\langle V | W \rangle = \sum_i v_i^* w_i \quad (1.2.5)$$

This is the form of the inner product we will use from now on.

You can now appreciate the first axiom; but for the complex conjugation of the components of the first vector, $\langle V | V \rangle$ would not even be real, not to mention positive. But now it is given by

$$\langle V | V \rangle = \sum_i |v_i|^2 \geq 0 \quad (1.2.6)$$

and vanishes only for the null vector. This makes it sensible to refer to $\langle V | V \rangle$ as the length or norm squared of a vector.

Consider Eq. (1.2.5). Since the vector $|V\rangle$ is uniquely specified by its components in a given basis, we may, in this basis, write it as a column vector:

$$|V\rangle \rightarrow \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \quad \text{in this basis} \quad (1.2.7)$$

Likewise

$$|W\rangle \rightarrow \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \text{ in this basis} \quad (1.2.8)$$

The inner product $\langle V|W\rangle$ is given by the matrix product of the transpose conjugate of the column vector representing $|V\rangle$ with the column vector representing $|W\rangle$:

$$\langle V|W\rangle = [v_1^*, v_2^*, \dots, v_n^*] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \quad (1.2.9)$$

1.3. Dual Spaces and the Dirac Notation

There is a technical point here. The inner product is a number we are trying to generate from two kets $|V\rangle$ and $|W\rangle$, which are both represented by column vectors in some basis. Now there is no way to make a number out of two columns by direct matrix multiplication, but there is a way to make a number by matrix multiplication of a row times a column. Our trick for producing a number out of two columns has been to associate a unique row vector with one column (its transpose conjugate) and form its matrix product with the column representing the other. This has the feature that the answer depends on which of the two vectors we are going to convert to the row, the two choices ($\langle V|W\rangle$ and $\langle W|V\rangle$) leading to answers related by complex conjugation.

But one can also take the following alternate view. Column vectors are concrete manifestations of an abstract vector $|V\rangle$ or ket in a basis. We can also work backward and go from the column vectors to the abstract kets. But then it is similarly possible to work backward and associate with each *row vector* an abstract object $\langle W|$, called *bra-W*. Now we can name the bras as we want but let us do the following. Associated with every ket $|V\rangle$ is a column vector. Let us take its *adjoint*, or transpose conjugate, and form a row vector. The abstract bra associated with this will bear the same label, i.e., it will be called $\langle V|$. Thus there are two vector spaces, the space of kets and a dual space of bras, with a ket for every bra and vice versa (the components being related by the adjoint operation). Inner products are really defined only between bras and kets and hence from elements of two distinct but related vector spaces. There is a basis of vectors $|i\rangle$ for expanding kets and a similar basis $\langle i|$ for expanding bras. The basis ket $|i\rangle$ is represented in the basis we are using by a column vector with all zeros except for a 1 in the i th row, while the basis bra $\langle i|$ is a row vector with all zeros except for a 1 in the i th column.

All this may be summarized as follows:

$$|V\rangle \leftrightarrow \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \leftrightarrow [v_1^*, v_2^*, \dots, v_n^*] \leftrightarrow \langle V | \quad (1.3.1)$$

where \leftrightarrow means “within a basis.”

There is, however, nothing wrong with the first viewpoint of associating a scalar product with a pair of columns or kets (making no reference to another dual space) and living with the asymmetry between the first and second vector in the inner product (which one to transpose conjugate?). If you found the above discussion heavy going, you can temporarily ignore it. The only thing you must remember is that in the case of a general nonarbitrary vector space:

- Vectors can still be assigned components in some orthonormal basis, just as with arrows, but these may be complex.
- The inner product of any two vectors is given in terms of these components by Eq. (1.2.5). This product obeys all the axioms.

1.3.1. Expansion of Vectors in an Orthonormal Basis

Suppose we wish to expand a vector $|V\rangle$ in an orthonormal basis. To find the components that go into the expansion we proceed as follows. We take the dot product of both sides of the assumed expansion with $|j\rangle$: (or $\langle j|$ if you are a purist)

$$|V\rangle = \sum_i v_i |i\rangle \quad (1.3.2)$$

$$\langle j | V \rangle = \sum_i v_i \langle j | i \rangle \quad (1.3.3)$$

$$= v_j \quad (1.3.4)$$

i.e., to find the j th component of a vector we take the dot product with the j th unit vector, exactly as with arrows. Using this result we may write

$$|V\rangle = \sum_i |i\rangle \langle i | V \rangle \quad (1.3.5)$$

Let us make sure the basis vectors look as they should. If we set $|V\rangle = |j\rangle$ in Eq. (1.3.5), we find the correct answer: the i th component of the j th basis vector is δ_{ij} . Thus for example the column representing basis vector number 4 will have a 1 in the 4th row and zero everywhere else. The abstract relation

$$|V\rangle = \sum_i v_i |i\rangle \quad (1.3.6)$$

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + v_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \quad (1.3.7)$$

1.3.2. Adjoint Operation

We have seen that we may pass from the column representing a ket to the row representing the corresponding bra by the adjoint operation, i.e., transpose conjugation. Let us now ask: if $\langle V |$ is the bra corresponding to the ket $|V\rangle$ what bra corresponds to $a|V\rangle$ where a is some scalar? By going to any basis it is readily found that

$$a|V\rangle \rightarrow \begin{bmatrix} av_1 \\ av_2 \\ \vdots \\ av_n \end{bmatrix} \rightarrow [a^*v_1^*, a^*v_2^*, \dots, a^*v_n^*] \rightarrow \langle V|a^* \quad (1.3.8)$$

It is customary to write $a|V\rangle$ as $|aV\rangle$ and the corresponding bra as $\langle aV|$. What we have found is that

$$\langle aV| = \langle V|a^* \quad (1.3.9)$$

Since the relation between bras and kets is linear we can say that if we have an equation among kets such as

$$a|V\rangle = b|W\rangle + c|Z\rangle + \dots \quad (1.3.10)$$

this implies another one among the corresponding bras:

$$\langle V|a^* = \langle W|b^* + \langle Z|c^* + \dots \quad (1.3.11)$$

The two equations above are said to be *adjoints of each other*. Just as any equation involving complex numbers implies another obtained by taking the complex conjugates of both sides, an equation between (bras) kets implies another one between (kets) bras. If you think in a basis, you will see that this follows simply from the fact that if two columns are equal, so are their transpose conjugates.

Here is the rule for taking the adjoint:

To take the adjoint of a linear equation relating kets (bras), replace every ket (bra) by its bra (ket) and complex conjugate all coefficients.

We can extend this rule as follows. Suppose we have an expansion for a vector:

$$|V\rangle = \sum_{i=1} v_i |i\rangle \quad (1.3.12)$$

in terms of basis vectors. The adjoint is

$$\langle V| = \sum_{i=1} \langle i| v_i^*$$

Recalling that $v_i = \langle i|V\rangle$ and $v_i^* = \langle V|i\rangle$, it follows that the adjoint of

$$|V\rangle = \sum_{i=1} |i\rangle \langle i| V \rangle \quad (1.3.13)$$

is

$$\langle V| = \sum_{i=1} \langle V| i \rangle \langle i| \quad (1.3.14)$$

from which comes the rule:

To take the adjoint of an equation involving bras and kets and coefficients, reverse the order of all factors, exchanging bras and kets and complex conjugating all coefficients.

Gram–Schmidt Theorem

Let us now take up the Gram–Schmidt procedure for converting a linearly independent basis into an orthonormal one. The basic idea can be seen by a simple example. Imagine the two-dimensional space of arrows in a plane. Let us take two nonparallel vectors, which qualify as a basis. To get an orthonormal basis out of these, we do the following:

- Rescale the first by its own length, so it becomes a unit vector. This will be the first basis vector.
- Subtract from the second vector its projection along the first, leaving behind only the part perpendicular to the first. (Such a part will remain since by assumption the vectors are nonparallel.)
- Rescale the left over piece by its own length. We now have the second basis vector: it is orthogonal to the first and of unit length.

This simple example tells the whole story behind this procedure, which will now be discussed in general terms in the Dirac notation.

Let $|I\rangle, |II\rangle, \dots$ be a linearly independent basis. The first vector of the orthonormal basis will be

$$|1\rangle = \frac{|I\rangle}{|I|} \quad \text{where} \quad |I| = \sqrt{\langle I|I\rangle}$$

Clearly

$$\langle 1|1\rangle = \frac{\langle I|I\rangle}{|I|^2} = 1$$

As for the second vector in the basis, consider

$$|2'\rangle = |II\rangle - |1\rangle\langle 1|II\rangle$$

which is $|II\rangle$ minus the part pointing along the first unit vector. (Think of the arrow example as you read on.) Not surprisingly it is orthogonal to the latter:

$$\langle 1|2'\rangle = \langle 1|II\rangle - \langle 1|1\rangle\langle 1|II\rangle = 0$$

We now divide $|2'\rangle$ by its norm to get $|2\rangle$ which will be orthogonal to the first and normalized to unity. Finally, consider

$$|3'\rangle = |III\rangle - |1\rangle\langle 1|III\rangle - |2\rangle\langle 2|III\rangle$$

which is orthogonal to both $|1\rangle$ and $|2\rangle$. Dividing by its norm we get $|3\rangle$, the third member of the orthogonal basis. There is nothing new with the generation of the rest of the basis.

Where did we use the linear independence of the original basis? What if we had started with a linearly dependent basis? Then at some point a vector like $|2'\rangle$ or $|3'\rangle$ would have vanished, putting a stop to the whole procedure. On the other hand, linear independence will assure us that such a thing will never happen since it amounts to having a nontrivial linear combination of linearly independent vectors that adds up the null vector. (Go back to the equations for $|2'\rangle$ or $|3'\rangle$ and satisfy yourself that these are linear combinations of the old basis vectors.)

Exercise 1.3.1. Form an orthonormal basis in two dimensions starting with $\vec{A} = 3\vec{i} + 4\vec{j}$ and $\vec{B} = 2\vec{i} - 6\vec{j}$. Can you generate another orthonormal basis starting with these two vectors? If so, produce another.

16

CHAPTER 1

Exercise 1.3.2. Show how to go from the basis

$$|I\rangle = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} \quad |II\rangle = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \quad |III\rangle = \begin{bmatrix} 0 \\ 2 \\ 5 \end{bmatrix}$$

to the orthonormal basis

$$|1\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad |2\rangle = \begin{bmatrix} 0 \\ 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \quad |3\rangle = \begin{bmatrix} 0 \\ -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$$

When we first learn about dimensionality, we associate it with the number of perpendicular directions. In this chapter we defined it in terms of the maximum number of linearly independent vectors. The following theorem connects the two definitions.

Theorem 4. The dimensionality of a space equals n_{\perp} , the maximum number of mutually orthogonal vectors in it.

To show this, first note that any mutually orthogonal set is also linearly independent. Suppose we had a linear combination of orthogonal vectors adding up to zero. By taking the dot product of both sides with any one member and using the orthogonality we can show that the coefficient multiplying that vector had to vanish. This can clearly be done for all the coefficients, showing the linear combination is trivial.

Now n_{\perp} can only be equal to, greater than or lesser than n , the dimensionality of the space. The Gram–Schmidt procedure eliminates the last case by explicit construction, while the linear independence of the perpendicular vectors rules out the penultimate option.

Schwarz and Triangle Inequalities

Two powerful theorems apply to any inner product space obeying our axioms:

Theorem 5. The Schwarz Inequality

$$|\langle V | W \rangle| \leq |V| |W| \tag{1.3.15}$$

Theorem 6. The Triangle Inequality

$$|V + W| \leq |V| + |W| \tag{1.3.16}$$

The proof of the first will be provided so you can get used to working with bras and kets. The second will be left as an exercise.

Before proving anything, note that the results are obviously true for arrows: the *Schwarz inequality* says that the dot product of two vectors cannot exceed the product of their lengths and the *triangle inequality* says that the length of a sum cannot exceed the sum of the lengths. This is an example which illustrates the merits of thinking of abstract vectors as arrows and guessing what properties they might share with arrows. The proof will of course have to rely on just the axioms.

To prove the Schwarz inequality, consider axiom $\langle Z|Z\rangle \geq 0$ applied to

$$|Z\rangle = |V\rangle - \frac{\langle W|V\rangle}{|W|^2} |W\rangle \quad (1.3.17)$$

We get

$$\begin{aligned} \langle Z|Z\rangle &= \langle V - \frac{\langle W|V\rangle}{|W|^2} W | V - \frac{\langle W|V\rangle}{|W|^2} W \rangle \\ &= \langle V|V\rangle - \frac{\langle W|V\rangle\langle V|W\rangle}{|W|^2} - \frac{\langle W|V\rangle^*\langle W|V\rangle}{|W|^2} \\ &\quad + \frac{\langle W|V\rangle^*\langle W|V\rangle\langle W|W\rangle}{|W|^4} \\ &\geq 0 \end{aligned} \quad (1.3.18)$$

where we have used the antilinearity of the inner product with respect to the bra. Using

$$\langle W|V\rangle^* = \langle V|W\rangle$$

we find

$$\langle V|V\rangle \geq \frac{\langle W|V\rangle\langle V|W\rangle}{|W|^2} \quad (1.3.19)$$

Cross-multiplying by $|W|^2$ and taking square roots, the result follows.

Exercise 1.3.3. When will this equality be satisfied? Does this agree with your experience with arrows?

Exercise 1.3.4. Prove the triangle inequality starting with $|V+W|^2$. You must use $\text{Re}\langle V|W\rangle \leq |\langle V|W\rangle|$ and the Schwarz inequality. Show that the final inequality becomes an equality only if $|V\rangle = a|W\rangle$ where a is a real positive scalar.

1.4. Subspaces

Definition 11. Given a vector space \mathbb{V} , a subset of its elements that form a vector space among themselves‡ is called a *subspace*. We will denote a particular subspace i of dimensionality n_i by $\mathbb{V}_i^{n_i}$.

‡ Vector addition and scalar multiplication are defined the same way in the subspace as in \mathbb{V} .

Example 1.4.1. In the space $\mathbb{V}^3(R)$, the following are some examples of subspaces: (a) all vectors along the x axis, the space \mathbb{V}_x^1 ; (b) all vectors along the y axis, the space \mathbb{V}_y^1 ; (c) all vectors in the $x-y$ plane, the space \mathbb{V}_{xy}^2 . Notice that all subspaces contain the null vector and that each vector is accompanied by its inverse to fulfill axioms for a vector space. Thus the set of all vectors along the positive x axis alone do not form a vector space. \square

Definition 12. Given two subspaces $\mathbb{V}_i^{n_i}$ and $\mathbb{V}_j^{m_j}$, we define their sum $\mathbb{V}_i^{n_i} \oplus \mathbb{V}_j^{m_j} = \mathbb{V}_k^{n_i+m_j}$ as the set containing (1) all elements of $\mathbb{V}_i^{n_i}$, (2) all elements of $\mathbb{V}_j^{m_j}$, (3) all possible linear combinations of the above. But for the elements (3), closure would be lost.

Example 1.4.2. If, for example, $\mathbb{V}_x^1 \oplus \mathbb{V}_y^1$ contained only vectors along the x and y axes, we could, by adding two elements, one from each direction, generate one along neither. On the other hand, if we also included all linear combinations, we would get the correct answer, $\mathbb{V}_x^1 \oplus \mathbb{V}_y^1 = \mathbb{V}_{xy}^2$. \square

*Exercise 1.4.1.** In a space \mathbb{V}^n , prove that the set of all vectors $\{|V_\perp^1\rangle, |V_\perp^2\rangle, \dots\}$, orthogonal to any $|V\rangle \neq 0\rangle$, form a subspace \mathbb{V}^{n-1} .

Exercise 1.4.2. Suppose $\mathbb{V}_1^{n_1}$ and $\mathbb{V}_2^{n_2}$ are two subspaces such that any element of \mathbb{V}_1 is orthogonal to any element of \mathbb{V}_2 . Show that the dimensionality of $\mathbb{V}_1 \oplus \mathbb{V}_2$ is $n_1 + n_2$. (Hint: Theorem 4.)

1.5. Linear Operators

An operator Ω is an instruction for transforming any given vector $|V\rangle$ into another, $|V'\rangle$. The action of the operator is represented as follows:

$$\Omega|V\rangle = |V'\rangle \quad (1.5.1)$$

One says that the operator Ω has transformed the ket $|V\rangle$ into the ket $|V'\rangle$. We will restrict our attention throughout to operators Ω that do not take us out of the vector space, i.e., if $|V\rangle$ is an element of a space \mathbb{V} , so is $|V'\rangle = \Omega|V\rangle$.

Operators can also act on bras:

$$\langle V'|\Omega = \langle V''| \quad (1.5.2)$$

We will only be concerned with *linear operators*, i.e., ones that obey the following rules:

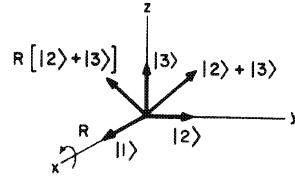
$$\Omega\alpha|V_i\rangle = \alpha\Omega|V_i\rangle \quad (1.5.3a)$$

$$\Omega\{\alpha|V_i\rangle + \beta|V_j\rangle\} = \alpha\Omega|V_i\rangle + \beta\Omega|V_j\rangle \quad (1.5.3b)$$

$$\langle V_i|\alpha\Omega = \langle V_i|\Omega\alpha \quad (1.5.4a)$$

$$(\langle V_i|\alpha + \langle V_j|\beta)\Omega = \alpha\langle V_i|\Omega + \beta\langle V_j|\Omega \quad (1.5.4b)$$

Figure 1.3. Action of the operator $R(\frac{1}{2}\pi\mathbf{i})$. Note that $R(|2\rangle + |3\rangle) = R|2\rangle + R|3\rangle$ as expected of a linear operator. (We will often refer to $R(\frac{1}{2}\pi\mathbf{i})$ as R if no confusion is likely.)



Example 1.5.1. The simplest operator is the identity operator, I , which carries the instruction:

$$I \rightarrow \text{Leave the vector alone!}$$

Thus,

$$I|V\rangle = |V\rangle \quad \text{for all kets } |V\rangle \quad (1.5.5)$$

and

$$\langle V|I = \langle V| \quad \text{for all bras } \langle V| \quad (1.5.6)$$

We next pass on to a more interesting operator on $\mathbb{V}^3(R)$:

$$R(\frac{1}{2}\pi\mathbf{i}) \rightarrow \text{Rotate vector by } \frac{1}{2}\pi \text{ about the unit vector } \mathbf{i}$$

[More generally, $R(\theta)$ stands for a rotation by an angle $\theta = |\theta|$ about the axis parallel to the unit vector $\hat{\theta} = \theta/\theta$.] Let us consider the action of this operator on the three unit vectors \mathbf{i} , \mathbf{j} , and \mathbf{k} , which in our notation will be denoted by $|1\rangle$, $|2\rangle$, and $|3\rangle$ (see Fig. 1.3). From the figure it is clear that

$$R(\frac{1}{2}\pi\mathbf{i})|1\rangle = |1\rangle \quad (1.5.7a)$$

$$R(\frac{1}{2}\pi\mathbf{i})|2\rangle = |3\rangle \quad (1.5.7b)$$

$$R(\frac{1}{2}\pi\mathbf{i})|3\rangle = -|2\rangle \quad (1.5.7c)$$

Clearly $R(\frac{1}{2}\pi\mathbf{i})$ is linear. For instance, it is clear from the same figure that $R(|2\rangle + |3\rangle) = R|2\rangle + R|3\rangle$. \square

The nice feature of linear operators is that once their action on the basis vectors is known, their action on any vector in the space is determined. If

$$\Omega|i\rangle = |i'\rangle$$

for a basis $|1\rangle, |2\rangle, \dots, |n\rangle$ in \mathbb{V}^n , then for any $|V\rangle = \sum v_i|i\rangle$

$$\Omega|V\rangle = \sum_i \Omega v_i|i\rangle = \sum_i v_i \Omega|i\rangle = \sum_i v_i|i'\rangle \quad (1.5.8)$$

This is the case in the example $\Omega = R(\frac{1}{2}\pi\mathbf{i})$. If

$$|V\rangle = v_1|1\rangle + v_2|2\rangle + v_3|3\rangle$$

is any vector, then

$$R|V\rangle = v_1R|1\rangle + v_2R|2\rangle + v_3R|3\rangle = v_1|1\rangle + v_2|3\rangle - v_3|2\rangle$$

The *product of two operators* stands for the instruction that the instructions corresponding to the two operators be carried out in sequence

$$\Lambda\Omega|V\rangle = \Lambda(\Omega|V\rangle) = \Lambda|\Omega V\rangle \quad (1.5.9)$$

where $|\Omega V\rangle$ is the ket obtained by the action of Ω on $|V\rangle$. The order of the operators in a product is very important: in general,

$$\Omega\Lambda - \Lambda\Omega \equiv [\Omega, \Lambda]$$

called the *commutator* of Ω and Λ isn't zero. For example $R(\frac{1}{2}\pi\mathbf{i})$ and $R(\frac{1}{2}\pi\mathbf{j})$ do not commute, i.e., their commutator is nonzero.

Two useful identities involving commutators are

$$[\Omega, \Lambda\theta] = \Lambda[\Omega, \theta] + [\Omega, \Lambda]\theta \quad (1.5.10)$$

$$[\Lambda\Omega, \theta] = \Lambda[\Omega, \theta] + [\Lambda, \theta]\Omega \quad (1.5.11)$$

Notice that apart from the emphasis on ordering, these rules resemble the chain rule in calculus for the derivative of a product.

The *inverse* of Ω , denoted by Ω^{-1} , satisfies[‡]

$$\Omega\Omega^{-1} = \Omega^{-1}\Omega = I \quad (1.5.12)$$

Not every operator has an inverse. The condition for the existence of the inverse is given in Appendix A.1. The operator $R(\frac{1}{2}\pi\mathbf{i})$ has an inverse: it is $R(-\frac{1}{2}\pi\mathbf{i})$. The inverse of a product of operators is the product of the inverses in reverse:

$$(\Omega\Lambda)^{-1} = \Lambda^{-1}\Omega^{-1} \quad (1.5.13)$$

for only then do we have

$$(\Omega\Lambda)(\Omega\Lambda)^{-1} = (\Omega\Lambda)(\Lambda^{-1}\Omega^{-1}) = \Omega\Lambda\Lambda^{-1}\Omega^{-1} = \Omega\Omega^{-1} = I$$

1.6. Matrix Elements of Linear Operators

We are now accustomed to the idea of an abstract vector being represented in a basis by an n -tuple of numbers, called its components, in terms of which all vector

[‡] In $\mathbb{V}^n(C)$ with n finite, $\Omega^{-1}\Omega = I \Leftrightarrow \Omega\Omega^{-1} = I$. Prove this using the ideas introduced toward the end of Theorem A.1.1., Appendix A.1.

operations can be carried out. We shall now see that in the same manner a linear operator can be represented in a basis by a set of n^2 numbers, written as an $n \times n$ matrix, and called its *matrix elements* in that basis. Although the matrix elements, just like the vector components, are basis dependent, they facilitate the computation of all basis-independent quantities, by rendering the abstract operator more tangible.

Our starting point is the observation made earlier, that the action of a linear operator is fully specified by its action on the basis vectors. If the basis vectors suffer a change

$$\Omega|i\rangle = |i'\rangle$$

(where $|i'\rangle$ is known), then any vector in this space undergoes a change that is readily calculable:

$$\Omega|V\rangle = \Omega \sum_i v_i |i\rangle = \sum_i v_i \Omega|i\rangle = \sum_i v_i |i'\rangle$$

When we say $|i'\rangle$ is known, we mean that its components in the original basis

$$\langle j|i'\rangle = \langle j|\Omega|i\rangle \equiv \Omega_{ji} \quad (1.6.1)$$

are known. The n^2 numbers, Ω_{ij} , are the *matrix elements* of Ω in this basis. If

$$\Omega|V\rangle = |V'\rangle$$

then the components of the transformed ket $|V'\rangle$ are expressable in terms of the Ω_{ij} and the components of $|V\rangle$:

$$\begin{aligned} v'_i &= \langle i|V'\rangle = \langle i|\Omega|V\rangle = \langle i|\Omega\left(\sum_j v_j |j\rangle\right) \\ &= \sum_j v_j \langle i|\Omega|j\rangle \\ &= \sum_j \Omega_{ij} v_j \end{aligned} \quad (1.6.2)$$

Equation (1.6.2) can be cast in matrix form:

$$\begin{bmatrix} v'_1 \\ v'_2 \\ \vdots \\ v'_n \end{bmatrix} = \begin{bmatrix} \langle 1|\Omega|1\rangle & \langle 1|\Omega|2\rangle & \cdots & \langle 1|\Omega|n\rangle \\ \langle 2|\Omega|1\rangle & & & \\ \vdots & & & \vdots \\ \langle n|\Omega|1\rangle & \cdots & & \langle n|\Omega|n\rangle \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \quad (1.6.3)$$

A mnemonic: the elements of the first column are simply the components of the first transformed basis vector $|1'\rangle = \Omega|1\rangle$ in the given basis. Likewise, the elements of the j th column represent the image of the j th basis vector after Ω acts on it.

Convince yourself that the same matrix Ω_{ij} acting to the *left* on the row vector corresponding to any $\langle v' |$ gives the row vector corresponding to $\langle v'' | = \langle v' | \Omega$.

Example 1.6.1. Combining our mnemonic with the fact that the operator $R(\frac{1}{2}\pi\mathbf{i})$ has the following effect on the basis vectors:

$$R(\frac{1}{2}\pi\mathbf{i})|1\rangle = |1\rangle$$

$$R(\frac{1}{2}\pi\mathbf{i})|2\rangle = |3\rangle$$

$$R(\frac{1}{2}\pi\mathbf{i})|3\rangle = -|2\rangle$$

we can write down the matrix that represents it in the $|1\rangle, |2\rangle, |3\rangle$ basis:

$$R(\frac{1}{2}\pi\mathbf{i}) \leftrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \quad (1.6.4)$$

For instance, the -1 in the third column tells us that R rotates $|3\rangle$ into $-|2\rangle$. One may also ignore the mnemonic altogether and simply use the definition $R_{ij} = \langle i | R | j \rangle$ to compute the matrix. \square

Exercise 1.6.1. An operator Ω is given by the matrix

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

What is its action?

Let us now consider certain specific operators and see how they appear in matrix form.

(1) The Identity Operator I .

$$I_{ij} = \langle i | I | j \rangle = \langle i | j \rangle = \delta_{ij} \quad (1.6.5)$$

Thus I is represented by a diagonal matrix with 1's along the diagonal. You should verify that our mnemonic gives the same result.

(2) The Projection Operators. Let us first get acquainted with *projection operators*. Consider the expansion of an arbitrary ket $|V\rangle$ in a basis:

$$|V\rangle = \sum_{i=1}^n |i\rangle \langle i | V \rangle$$

In terms of the objects $|i\rangle\langle i|$, which are linear operators, and which, by definition, act on $|V\rangle$ to give $|i\rangle\langle i|V\rangle$, we may write the above as

$$|V\rangle = \left(\sum_{i=1}^n |i\rangle\langle i| \right) |V\rangle \quad (1.6.6)$$

Since Eq. (1.6.6) is true for all $|V\rangle$, the object in the brackets must be identified with the identity (operator)

$$I = \sum_{i=1}^n |i\rangle\langle i| = \sum_{i=1}^n \mathbb{P}_i \quad (1.6.7)$$

The object $\mathbb{P}_i = |i\rangle\langle i|$ is called the *projection operator* for the ket $|i\rangle$. Equation (1.6.7), which is called the *completeness relation*, expresses the identity as a sum over projection operators and will be invaluable to us. (If you think that any time spent on the identity, which seems to do nothing, is a waste of time, just wait and see.)

Consider

$$\mathbb{P}_i |V\rangle = |i\rangle\langle i|V\rangle = |i\rangle v_i \quad (1.6.8)$$

Clearly \mathbb{P}_i is linear. Notice that whatever $|V\rangle$ is, $\mathbb{P}_i |V\rangle$ is a multiple of $|i\rangle$ with a coefficient (v_i) which is the component of $|V\rangle$ along $|i\rangle$. Since \mathbb{P}_i projects out the component of any ket $|V\rangle$ along the direction $|i\rangle$, it is called a *projection operator*. The completeness relation, Eq. (1.6.7), says that the sum of the projections of a vector along all the n directions equals the vector itself. Projection operators can also act on bras in the same way:

$$\langle V|\mathbb{P}_i = \langle V|i\rangle\langle i| = v_i^* \langle i| \quad (1.6.9)$$

Projection operators corresponding to the basis vectors obey

$$\mathbb{P}_i \mathbb{P}_j = |i\rangle\langle i|j\rangle\langle j| = \delta_{ij} \mathbb{P}_i \quad (1.6.10)$$

This equation tells us that (1) once \mathbb{P}_i projects out the part of $|V\rangle$ along $|i\rangle$, further applications of \mathbb{P}_i make no difference; and (2) the subsequent application of \mathbb{P}_j ($j \neq i$) will result in zero, since a vector entirely along $|i\rangle$ cannot have a projection along a perpendicular direction $|j\rangle$.

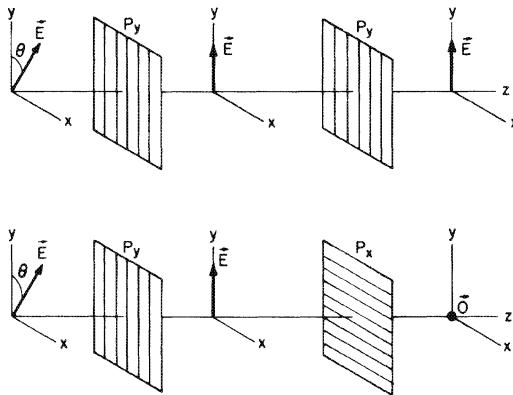


Figure 1.4. P_x and P_y are polarizers placed in the way of a beam traveling along the z axis. The action of the polarizers on the electric field \mathbf{E} obeys the law of combination of projection operators: $P_i P_j = \delta_{ij} P_j$.

The following example from optics may throw some light on the discussion. Consider a beam of light traveling along the z axis and polarized in the $x-y$ plane at an angle θ with respect to the y axis (see Fig. 1.4). If a polarizer P_y , that only admits light polarized along the y axis, is placed in the way, the projection $E \cos \theta$ along the y axis is transmitted. An additional polarizer P_y placed in the way has no further effect on the beam. We may equate the action of the polarizer to that of a projection operator P_y that acts on the electric field vector \mathbf{E} . If P_y is followed by a polarizer P_x the beam is completely blocked. Thus the polarizers obey the equation $P_i P_j = \delta_{ij} P_j$ expected of projection operators.

Let us next turn to the matrix elements of \mathbb{P}_i . There are two approaches. The first one, somewhat indirect, gives us a feeling for what kind of an object $|i\rangle\langle i|$ is. We know

$$|i\rangle \leftrightarrow \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

and

$$\langle i | \leftrightarrow [0, 0, \dots, 1, 0, 0, \dots, 0]$$

so that

$$|i\rangle\langle i| \leftrightarrow \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} [0, 0, \dots, 1, 0, \dots, 0] = \begin{bmatrix} 0 & & & & & & 0 \\ & \ddots & & & & & \\ & & 0 & & & & \\ & & & 1 & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & 0 \end{bmatrix} \quad (1.6.11)$$

by the rules of matrix multiplication. Whereas $\langle V|V' \rangle = (1 \times n \text{ matrix}) \times (n \times 1 \text{ matrix}) = (1 \times 1 \text{ matrix})$ is a scalar, $|V\rangle\langle V'| = (n \times 1 \text{ matrix}) \times (1 \times n \text{ matrix}) = (n \times n \text{ matrix})$ is an operator. The inner product $\langle V|V' \rangle$ represents a bra and ket which have found each other, while $|V\rangle\langle V'|$, sometimes called the *outer product*, has the two factors looking the other way for a bra or a ket to dot with.

The more direct approach to the matrix elements gives

$$(\mathbb{P}_i)_{kl} = \langle k|i\rangle\langle i|l\rangle = \delta_{ki}\delta_{il} = \delta_{kl}\delta_{ii} \quad (1.6.12)$$

which is of course identical to Eq. (1.6.11). The same result also follows from mnemonic. Each projection operator has only one nonvanishing matrix element, a 1 at the i th element on the diagonal. The completeness relation, Eq. (1.6.7), says that when all the \mathbb{P}_i are added, the diagonal fills out to give the identity. If we form the sum over just some of the projection operators, we get the operator which projects a given vector into the subspace spanned by just the corresponding basis vectors.

Matrices Corresponding to Products of Operators

Consider next the matrices representing a product of operators. These are related to the matrices representing the individual operators by the application of Eq. (1.6.7):

$$\begin{aligned} (\Omega\Lambda)_{ij} &= \langle i|\Omega\Lambda|j\rangle = \langle i|\Omega I\Lambda|j\rangle \\ &= \sum_k \langle i|\Omega|k\rangle\langle k|\Lambda|j\rangle = \sum_k \Omega_{ik}\Lambda_{kj} \end{aligned} \quad (1.6.13)$$

Thus the matrix representing the product of operators is the product of the matrices representing the factors.

The Adjoint of an Operator

Recall that given a ket $\alpha|V\rangle \equiv |\alpha V\rangle$ the corresponding bra is

$$\langle \alpha V | = \langle V | \alpha^* \quad (\text{and not } \langle V | \alpha)$$

In the same way, given a ket

$$\Omega|V\rangle = |\Omega V\rangle$$

the corresponding bra is

$$\langle \Omega V | = \langle V | \Omega^\dagger \quad (1.6.14)$$

which *defines* the operator Ω^\dagger . One may state this equation in words: if Ω turns a ket $|V\rangle$ to $|\Omega V\rangle$, then Ω^\dagger turns the bra $\langle V |$ into $\langle \Omega V |$. Just as α and α^* , $|V\rangle$ and $\langle V |$ are related but distinct objects, so are Ω and Ω^\dagger . The relation between Ω , and Ω^\dagger , called the *adjoint* of Ω or “omega dagger,” is best seen in a basis:

$$\begin{aligned} (\Omega^\dagger)_{ij} &= \langle i | \Omega^\dagger | j \rangle = \langle \Omega i | j \rangle \\ &= \langle j | \Omega i \rangle^* = \langle j | \Omega | i \rangle^* \end{aligned}$$

so

$$\Omega_j^\dagger = \Omega_{ji}^* \quad (1.6.15)$$

In other words, the matrix representing Ω^\dagger is the transpose conjugate of the matrix representing Ω . (Recall that the row vector representing $\langle V |$ is the transpose conjugate of the column vector representing $|V\rangle$. *In a given basis, the adjoint operation is the same as taking the transpose conjugate.*)

The adjoint of a product is the product of the adjoints in reverse:

$$(\Omega\Lambda)^\dagger = \Lambda^\dagger \Omega^\dagger \quad (1.6.16)$$

To prove this we consider $\langle \Omega\Lambda V |$. First we treat $\Omega\Lambda$ as one operator and get

$$\langle \Omega\Lambda V | = \langle (\Omega\Lambda) V | = \langle V | (\Omega\Lambda)^\dagger$$

Next we treat (ΛV) as just another vector, and write

$$\langle \Omega\Lambda V | = \langle \Omega(\Lambda V) | = \langle \Lambda V | \Omega^\dagger$$

We next pull out Λ , pushing Ω^\dagger further out:

$$\langle \Lambda V | \Omega^\dagger = \langle V | \Lambda^\dagger \Omega^\dagger$$

Comparing this result with the one obtained a few lines above, we get the desired result.

Consider now an equation consisting of kets, scalars, and operators, such as

$$\alpha_1 |V_1\rangle = \alpha_2 |V_2\rangle + \alpha_3 |V_3\rangle \langle V_4 | V_5 \rangle + \alpha_4 \Omega \Lambda |V_6\rangle \quad (1.6.17a)$$

What is its adjoint? Our old rule tells us that it is

$$\langle V_1 | \alpha_1^* = \langle V_2 | \alpha_2^* + \langle V_3 | V_4 \rangle \langle V_3 | \alpha_3^* + \langle \Omega \Lambda V_6 | \alpha_4^*$$

In the last term we can replace $\langle \Omega \Lambda V_6 |$ by

$$\langle V_6 | (\Omega \Lambda)^\dagger = \langle V_6 | \Lambda^\dagger \Omega^\dagger$$

so that finally we have the adjoint of Eq. (1.6.17a):

$$\langle V_1 | \alpha_1^* = \langle V_2 | \alpha_2^* + \langle V_3 | V_4 \rangle \langle V_3 | \alpha_3^* + \langle V_6 | \Lambda^\dagger \Omega^\dagger \alpha_4^* \quad (1.6.17b)$$

The final rule for taking the adjoint of the most general equation we will ever encounter is this:

When a product of operators, bras, kets, and explicit numerical coefficients is encountered, reverse the order of all factors and make the substitutions $\Omega \leftrightarrow \Omega^\dagger$, $| \rangle \leftrightarrow \langle |$, $\alpha \leftrightarrow \alpha^*$.

(Of course, there is no real need to reverse the location of the scalars α except in the interest of uniformity.)

Hermitian, Anti-Hermitian, and Unitary Operators

We now turn our attention to certain special classes of operators that will play a major role in quantum mechanics.

Definition 13. An operator Ω is *Hermitian* if $\Omega^\dagger = \Omega$.

Definition 14. An operator Ω is *anti-Hermitian* if $\Omega^\dagger = -\Omega$.

The adjoint is to an operator what the complex conjugate is to numbers. Hermitian and anti-Hermitian operators are like pure real and pure imaginary numbers. Just as every number may be decomposed into a sum of pure real and pure imaginary parts,

$$\alpha = \frac{\alpha + \alpha^*}{2} + \frac{\alpha - \alpha^*}{2}$$

we can decompose every operator into its Hermitian and anti-Hermitian parts:

$$\Omega = \frac{\Omega + \Omega^\dagger}{2} + \frac{\Omega - \Omega^\dagger}{2} \quad (1.6.18)$$

*Exercise 1.6.2.** Given Ω and Λ are Hermitian what can you say about (1) $\Omega \Lambda$; (2) $\Omega \Lambda + \Lambda \Omega$; (3) $[\Omega, \Lambda]$; and (4) $i[\Omega, \Lambda]$?

Definition 15. An operator U is *unitary* if

$$UU^\dagger = I \quad (1.6.19)$$

This equation tells us that U and U^\dagger are inverses of each other. Consequently, from Eq. (1.5.12),

$$U^\dagger U = I \quad (1.6.20)$$

Following the analogy between operators and numbers, unitary operators are like complex numbers of unit modulus, $u = e^{i\theta}$. Just as $u^* u = 1$, so is $U^\dagger U = I$.

*Exercise 1.6.3.** Show that a product of unitary operators is unitary.

Theorem 7. Unitary operators preserve the inner product between the vectors they act on.

Proof. Let

$$|V'_1\rangle = U|V_1\rangle$$

and

$$|V'_2\rangle = U|V_2\rangle$$

Then

$$\begin{aligned} \langle V'_2 | V'_1 \rangle &= \langle UV_2 | UV_1 \rangle \\ &= \langle V_2 | U^\dagger U | V_1 \rangle = \langle V_2 | V_1 \rangle \end{aligned} \quad (1.6.21)$$

(Q.E.D.)

Unitary operators are the generalizations of rotation operators from $\mathbb{V}^3(R)$ to $\mathbb{V}^n(C)$, for just like rotation operators in three dimensions, they preserve the lengths of vectors and their dot products. In fact, on a real vector space, the unitarity condition becomes $U^{-1} = U^T$ (T means transpose), which defines an *orthogonal* or rotation matrix. [$R(\frac{1}{2}\pi\mathbf{i})$ is an example.]

Theorem 8. If one treats the columns of an $n \times n$ unitary matrix as components of n vectors, these vectors are orthonormal. In the same way, the rows may be interpreted as components of n orthonormal vectors.

Proof 1. According to our mnemonic, the j th column of the matrix representing U is the image of the j th basis vector after U acts on it. Since U preserves inner products, the rotated set of vectors is also orthonormal. Consider next the rows. We now use the fact that U^\dagger is also a rotation. (How else can it neutralize U to give $U^\dagger U = I$?) Since the rows of U are the columns of U^\dagger (but for an overall complex

conjugation which does not affect the question of orthonormality), the result we already have for the columns of a unitary matrix tells us the rows of U are orthonormal.

Proof 2. Since $U^\dagger U = I$,

$$\begin{aligned}\delta_{ij} &= \langle i | I | j \rangle = \langle i | U^\dagger U | j \rangle \\ &= \sum_k \langle i | U^\dagger | k \rangle \langle k | U | j \rangle \\ &= \sum_k U_{ik}^* U_{kj} = \sum_k U_{ki}^* U_{kj} \end{aligned} \quad (1.6.22)$$

which proves the theorem for the columns. A similar result for the rows follows if we start with the equation $UU^\dagger = I$. Q.E.D.

Note that $U^\dagger U = I$ and $UU^\dagger = I$ are not independent conditions.

*Exercise 1.6.4.** It is assumed that you know (1) what a *determinant* is, (2) that $\det \Omega^T = \det \Omega$ (T denotes transpose), (3) that the determinant of a product of matrices is the product of the determinants. [If you do not, verify these properties for a two-dimensional case

$$\Omega = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

with $\det \Omega = (\alpha\delta - \beta\gamma)$.] Prove that the determinant of a unitary matrix is a complex number of unit modulus.

*Exercise 1.6.5.** Verify that $R(\frac{1}{2}\pi\mathbf{i})$ is unitary (orthogonal) by examining its matrix.

Exercise 1.6.6. Verify that the following matrices are unitary:

$$\frac{1}{2^{1/2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}, \quad \frac{1}{2} \begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix}$$

Verify that the determinant is of the form $e^{i\theta}$ in each case. Are any of the above matrices Hermitian?

1.7. Active and Passive Transformations

Suppose we subject all the vectors $|V\rangle$ in a space to a unitary transformation

$$|V\rangle \rightarrow U|V\rangle \quad (1.7.1)$$

Under this transformation, the matrix elements of any operator Ω are modified as follows:

$$\langle V' | \Omega | V \rangle \rightarrow \langle UV' | \Omega | UV \rangle = \langle V' | U^\dagger \Omega U | V \rangle \quad (1.7.2)$$

It is clear that the same change would be effected if we left the vectors alone and subjected all operators to the change

$$\Omega \rightarrow U^\dagger \Omega U \quad (1.7.3)$$

The first case is called an *active transformation* and the second a *passive transformation*. The present nomenclature is in reference to the vectors: they are affected in an active transformation and left alone in the passive case. The situation is exactly the opposite from the point of view of the operators.

Later we will see that the physics in quantum theory lies in the matrix elements of operators, and that active and passive transformations provide us with two equivalent ways of describing the same physical transformation.

*Exercise 1.7.1.** The *trace* of a matrix is defined to be the sum of its diagonal matrix elements

$$\text{Tr } \Omega = \sum_i \Omega_{ii}$$

Show that

- (1) $\text{Tr}(\Omega\Lambda) = \text{Tr}(\Lambda\Omega)$
- (2) $\text{Tr}(\Omega\Lambda\theta) = \text{Tr}(\Lambda\theta\Omega) = \text{Tr}(\theta\Omega\Lambda)$ (The permutations are *cyclic*).
- (3) The trace of an operator is unaffected by a unitary change of basis $|i\rangle \rightarrow U|i\rangle$. [Equivalently, show $\text{Tr } \Omega = \text{Tr}(U^\dagger \Omega U)$.]

Exercise 1.7.2. Show that the determinant of a matrix is unaffected by a unitary change of basis. [Equivalently show $\det \Omega = \det(U^\dagger \Omega U)$.]

1.8. The Eigenvalue Problem

Consider some linear operator Ω acting on an arbitrary *nonzero* ket $|V\rangle$:

$$\Omega|V\rangle = |V'\rangle \quad (1.8.1)$$

Unless the operator happens to be a trivial one, such as the identity or its multiple, the ket will suffer a nontrivial change, i.e., $|V'\rangle$ will not be simply related to $|V\rangle$. So much for an arbitrary ket. Each operator, however, has certain kets of its own, called its *eigenkets*, on which its action is simply that of rescaling:

$$\Omega|V\rangle = \omega|V\rangle \quad (1.8.2)$$

Equation (1.8.2) is an eigenvalue equation: $|V\rangle$ is an *eigenket* of Ω with *eigenvalue* ω . In this chapter we will see how, given an operator Ω , one can systematically determine all its eigenvalues and eigenvectors. How such an equation enters physics will be illustrated by a few examples from mechanics at the end of this section, and once we get to quantum mechanics proper, it will be eigen, eigen, eigen all the way.

Example 1.8.1. To illustrate how easy the eigenvalue problem really is, we will begin with a case that will be completely solved: the case $\Omega = I$. Since

$$I|V\rangle = |V\rangle$$

for all $|V\rangle$, we conclude that

- (1) the only eigenvalue of I is 1;
- (2) all vectors are its eigenvectors with this eigenvalue. \square

Example 1.8.2. After this unqualified success, we are encouraged to take on a slightly more difficult case: $\Omega = \mathbb{P}_V$, the projection operator associated with a *normalized* ket $|V\rangle$. Clearly

- (1) any ket $\alpha|V\rangle = |\alpha V\rangle$, parallel to $|V\rangle$ is an eigenket with eigenvalue 1:

$$\mathbb{P}_V|\alpha V\rangle = |V\rangle\langle V|\alpha V\rangle = \alpha|V\rangle|V|^2 = 1 \cdot |\alpha V\rangle$$

- (2) any ket $|V_\perp\rangle$, perpendicular to $|V\rangle$, is an eigenket with eigenvalue 0:

$$\mathbb{P}_V|V_\perp\rangle = |V\rangle\langle V|V_\perp\rangle = 0 = 0|V_\perp\rangle$$

- (3) kets that are neither, i.e., kets of the form $\alpha|V\rangle + \beta|V_\perp\rangle$, are simply not eigenkets:

$$\mathbb{P}_V(\alpha|V\rangle + \beta|V_\perp\rangle) = |\alpha V\rangle \neq \gamma(\alpha|V\rangle + \beta|V_\perp\rangle)$$

Since every ket in the space falls into one of the above classes, we have found all the eigenvalues and eigenvectors. \square

Example 1.8.3. Consider now the operator $R(\frac{1}{2}\pi\mathbf{i})$. We already know that it has one eigenket, the basis vector $|1\rangle$ along the x axis:

$$R(\frac{1}{2}\pi\mathbf{i})|1\rangle = |1\rangle$$

Are there others? Of course, any vector $\alpha|1\rangle$ along the x axis is also unaffected by the x rotation. This is a general feature of the eigenvalue equation and reflects the linearity of the operator:

if

$$\Omega|V\rangle = \omega|V\rangle$$

then

$$\Omega\alpha|V\rangle = \alpha\Omega|V\rangle = \alpha\omega|V\rangle = \omega\alpha|V\rangle$$

for any multiple α . Since the eigenvalue equation fixes the eigenvector only up to an overall scale factor, we will not treat the multiples of an eigenvector as distinct eigenvectors. With this understanding in mind, let us ask if $R(\frac{1}{2}\pi\mathbf{i})$ has any eigenvectors besides $|1\rangle$. Our intuition says no, for any vector not along the x axis necessarily gets rotated by $R(\frac{1}{2}\pi\mathbf{i})$ and cannot possibly transform into a multiple of itself. Since every vector is either parallel to $|1\rangle$ or isn't, we have fully solved the eigenvalue problem.

The trouble with this conclusion is that it is wrong! $R(\frac{1}{2}\pi\mathbf{i})$ has two other eigenvectors besides $|1\rangle$. But our intuition is not to be blamed, for these vectors are in $\mathbb{V}^3(C)$ and not $\mathbb{V}^3(R)$. It is clear from this example that we need a reliable and systematic method for solving the eigenvalue problem in $\mathbb{V}^n(C)$. We now turn our attention to this very question. \square

The Characteristic Equation and the Solution to the Eigenvalue Problem

We begin by rewriting Eq. (1.8.2) as

$$(\Omega - \omega I)|V\rangle = |0\rangle \quad (1.8.3)$$

Operating both sides with $(\Omega - \omega I)^{-1}$, assuming it exists, we get

$$|V\rangle = (\Omega - \omega I)^{-1}|0\rangle \quad (1.8.4)$$

Now, any finite operator (an operator with finite matrix elements) acting on the null vector can only give us a null vector. It therefore seems that in asking for a nonzero eigenvector $|V\rangle$, we are trying to get something for nothing out of Eq. (1.8.4). This is impossible. It follows that our assumption that the operator $(\Omega - \omega I)^{-1}$ exists (as a finite operator) is false. So we ask when this situation will obtain. Basic matrix theory tells us (see Appendix A.1) that the inverse of any matrix M is given by

$$M^{-1} = \frac{\text{cofactor } M^T}{\det M} \quad (1.8.5)$$

Now the cofactor of M is finite if M is. Thus what we need is the vanishing of the determinant. The condition for nonzero eigenvectors is therefore

$$\det(\Omega - \omega I) = 0 \quad (1.8.6)$$

This equation will determine the eigenvalues ω . To find them, we project Eq. (1.8.3) onto a basis. Dotting both sides with a basis bra $\langle i|$, we get

$$\langle i|\Omega - \omega I|V\rangle = 0$$

and upon introducing the representation of the identity [Eq. (1.6.7)], to the left of $|V\rangle$, we get the following image of Eq. (1.8.3):

$$\sum_j (\Omega_{ij} - \omega \delta_{ij}) v_j = 0 \quad (1.8.7)$$

Setting the determinant to zero will give us an expression of the form

$$\sum_{m=0}^n c_m \omega^m = 0 \quad (1.8.8)$$

Equation (1.8.8) is called the *characteristic equation* and

$$P^n(\omega) = \sum_{m=0}^n c_m \omega^m \quad (1.8.9)$$

is called the *characteristic polynomial*. Although the polynomial is being determined in a particular basis, the eigenvalues, which are its roots, are basis independent, for they are defined by the abstract Eq. (1.8.3), which makes no reference to any basis.

Now, a fundamental result in analysis is that every n th-order polynomial has n roots, not necessarily distinct and not necessarily real. Thus every operator in $\mathbb{V}^n(C)$ has n eigenvalues. Once the eigenvalues are known, the eigenvectors may be found, at least for Hermitian and unitary operators, using a procedure illustrated by the following example. [Operators on $\mathbb{V}^n(C)$ that are not of the above variety may not have n eigenvectors—see Exercise 1.8.4. Theorems 10 and 12 establish that Hermitian and unitary operators on $\mathbb{V}^n(C)$ will have n eigenvectors.]

Example 1.8.4. Let us use the general techniques developed above to find all the eigenvectors and eigenvalues of $R(\frac{1}{2}\pi\mathbf{i})$. Recall that the matrix representing it is

$$R(\frac{1}{2}\pi\mathbf{i}) \leftrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

Therefore the characteristic equation is

$$\det(R - \omega I) = \begin{vmatrix} 1 - \omega & 0 & 0 \\ 0 & -\omega & -1 \\ 0 & 1 & -\omega \end{vmatrix} = 0$$

i.e.,

$$(1 - \omega)(\omega^2 + 1) = 0 \quad (1.8.10)$$

with roots $\omega = 1, \pm i$. We know that $\omega = 1$ corresponds to $|1\rangle$. Let us see this come out of the formalism. Feeding $\omega = 1$ into Eq. (1.8.7) we find that the components x_1 , x_2 , and x_3 of the corresponding eigenvector must obey the equations

$$\begin{bmatrix} 1 - 1 & 0 & 0 \\ 0 & 0 - 1 & -1 \\ 0 & 1 & 0 - 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \rightarrow \begin{cases} 0 = 0 \\ -x_2 - x_3 = 0 \\ x_2 - x_3 = 0 \end{cases} \rightarrow x_2 = x_3 = 0$$

Thus any vector of the form

$$x_1|1\rangle \leftrightarrow \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix}$$

is acceptable, as expected. It is conventional to use the freedom in scale to normalize the eigenvectors. Thus in this case a choice is

$$|\omega = 1\rangle = |1\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

I say *a* choice, and not *the* choice, since the vector may be multiplied by a number of modulus unity without changing the norm. There is no universally accepted convention for eliminating this freedom, except perhaps to choose the vector with real components when possible.

Note that of the three simultaneous equations above, the first is not a real equation. In general, there will be only $(n-1)$ LI equations. This is the reason the norm of the vector is not fixed and, as shown in Appendix A.1, the reason the determinant vanishes.

Consider next the equations corresponding to $\omega = i$. The components of the eigenvector obey the equations

$$(1 - i)x_1 = 0 \quad (\text{i.e., } x_1 = 0)$$

$$-ix_2 - x_3 = 0 \quad (\text{i.e., } x_2 = ix_3)$$

$$x_2 - ix_3 = 0 \quad (\text{i.e., } x_2 = ix_3)$$

Notice once again that we have only $n-1$ useful equations. A properly normalized solution to the above is

$$|\omega = i\rangle \leftrightarrow \frac{1}{2^{1/2}} \begin{bmatrix} 0 \\ i \\ 1 \end{bmatrix}$$

A similar procedure yields the third eigenvector:

$$|\omega = -i\rangle \leftrightarrow \frac{1}{2^{1/2}} \begin{bmatrix} 0 \\ -i \\ 1 \end{bmatrix} \quad \square$$

In the above example we have introduced a popular convention: labeling the eigenvectors by the eigenvalue. For instance, the ket corresponding to $\omega = \omega_i$ is labeled $|\omega = \omega_i\rangle$ or simply $|\omega_i\rangle$. This notation presumes that to each ω_i there is just one vector labeled by it. Though this is not always the case, only a slight change in this notation will be needed to cover the general case.

The phenomenon of a single eigenvalue representing more than one eigenvector is called *degeneracy* and corresponds to repeated roots for the characteristic polynomial. In the face of degeneracy, we need to modify not just the labeling, but also the procedure used in the example above for finding the eigenvectors. Imagine that instead of $R(\frac{1}{2}\pi i)$ we were dealing with another operator Ω on $\mathbb{V}^3(R)$ with roots ω_1 and $\omega_2 = \omega_3$. It appears as if we can get two eigenvectors, by the method described above, one for each distinct ω . How do we get a third? Or is there no third? These questions will be answered in all generality shortly when we examine the question of degeneracy in detail. We now turn our attention to two central theorems on Hermitian operators. These play a vital role in quantum mechanics.

Theorem 9. The eigenvalues of a Hermitian operator are real.

Proof. Let

$$\Omega|\omega\rangle = \omega|\omega\rangle$$

Dot both sides with $\langle\omega|$:

$$\langle\omega|\Omega|\omega\rangle = \omega\langle\omega|\omega\rangle \quad (1.8.11)$$

Take the adjoint to get

$$\langle\omega|\Omega^\dagger|\omega\rangle = \omega^*\langle\omega|\omega\rangle$$

Since $\Omega = \Omega^\dagger$, this becomes

$$\langle\omega|\Omega|\omega\rangle = \omega^*\langle\omega|\omega\rangle$$

Subtracting from Eq. (1.8.11)

$$0 = (\omega - \omega^*)\langle\omega|\omega\rangle$$

$$\omega = \omega^* \quad \text{Q.E.D.}$$

Theorem 10. To every Hermitian operator Ω , there exists (at least) a basis consisting of its orthonormal eigenvectors. It is diagonal in this eigenbasis and has its eigenvalues as its diagonal entries.

Proof. Let us start with the characteristic equation. It must have at least one root, call it ω_1 . Corresponding to ω_1 there must exist at least one nonzero eigenvector $|\omega_1\rangle$. [If not, Theorem (A.1.1) would imply that $(\Omega - \omega_1 I)$ is invertible.] Consider the subspace $\mathbb{V}_{\perp 1}^{n-1}$ of all vectors orthogonal to $|\omega_1\rangle$. Let us choose as our basis the vector $|\omega_1\rangle$ (normalized to unity) and any $n-1$ orthonormal vectors $\{V_{\perp 1}^1, V_{\perp 1}^2, \dots, V_{\perp 1}^{n-1}\}$ in $\mathbb{V}_{\perp 1}^{n-1}$. In this basis Ω has the following form:

$$\Omega \leftrightarrow \begin{bmatrix} \omega_1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \boxed{} & & & & & \\ 0 & & \boxed{} & & & & \\ \vdots & & & \boxed{} & & & \\ 0 & & & & \boxed{} & & \end{bmatrix} \quad (1.8.12)$$

The first column is just the image of $|\omega_1\rangle$ after Ω has acted on it. Given the first column, the first row follows from the Hermiticity of Ω .

The characteristic equation now takes the form

$$(\omega_1 - \omega) \cdot (\text{determinant of boxed submatrix}) = 0$$

$$(\omega_1 - \omega) \sum_0^{n-1} c_m \omega^m = (\omega_1 - \omega) P^{n-1}(\omega) = 0$$

Now the polynomial P^{n-1} must also generate one root, ω_2 , and a normalized eigenvector $|\omega_2\rangle$. Define the subspace $\mathbb{V}_{\perp 1,2}^{n-2}$ of vectors in $\mathbb{V}_{\perp 1}^{n-1}$ orthogonal to $|\omega_2\rangle$ (and automatically to $|\omega_1\rangle$) and repeat the same procedure as before. Finally, the matrix Ω becomes, in the basis $|\omega_1\rangle, |\omega_2\rangle, \dots, |\omega_n\rangle$,

$$\Omega \leftrightarrow \begin{bmatrix} \omega_1 & 0 & 0 & \cdots & 0 \\ 0 & \omega_2 & 0 & & 0 \\ 0 & 0 & \omega_3 & & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & & \omega_n \end{bmatrix}$$

Since every $|\omega_i\rangle$ was chosen from a space that was orthogonal to the previous ones, $|\omega_1\rangle, |\omega_2\rangle, \dots, |\omega_{i-1}\rangle$; the basis of eigenvectors is orthonormal. (Notice that nowhere did we have to assume that the eigenvalues were all distinct.) Q.E.D.

[The analogy between real numbers and Hermitian operators is further strengthened by the fact that in a certain basis (of eigenvectors) the Hermitian operator can be represented by a matrix with all real elements.]

In stating Theorem 10, it was indicated that there might exist more than one basis of eigenvectors that diagonalized Ω . This happens if there is any degeneracy. Suppose $\omega_1 = \omega_2 = \omega$. Then we have two orthonormal vectors obeying

$$\Omega|\omega_1\rangle = \omega|\omega_1\rangle$$

$$\Omega|\omega_2\rangle = \omega|\omega_2\rangle$$

It follows that

$$\Omega[\alpha|\omega_1\rangle + \beta|\omega_2\rangle] = \alpha\omega|\omega_1\rangle + \beta\omega|\omega_2\rangle = \omega[\alpha|\omega_1\rangle + \beta|\omega_2\rangle]$$

for any α and β . Since the vectors $|\omega_1\rangle$ and $|\omega_2\rangle$ are orthogonal (and hence LI), we find that there is a whole two-dimensional subspace spanned by $|\omega_1\rangle$ and $|\omega_2\rangle$, the elements of which are eigenvectors of Ω with eigenvalue ω . One refers to this space as an *eigenspace* of Ω with eigenvalue ω . Besides the vectors $|\omega_1\rangle$ and $|\omega_2\rangle$, there exists an infinity of orthonormal pairs $|\omega'_1\rangle, |\omega'_2\rangle$, obtained by a rigid rotation of $|\omega_1\rangle, |\omega_2\rangle$, from which we may select any pair in forming the eigenbasis of Ω . In general, if an eigenvalue occurs m_i times, that is, if the characteristic equation has m_i of its roots equal to some ω_i , there will be an eigenspace $\mathbb{V}_{\omega_i}^{m_i}$ from which we may choose any m_i orthonormal vectors to form the basis referred to in Theorem 10.

In the absence of degeneracy, we can prove Theorem 9 and 10 very easily. Let us begin with two eigenvectors:

$$\Omega|\omega_i\rangle = \omega_i|\omega_i\rangle \quad (1.8.13a)$$

$$\Omega|\omega_j\rangle = \omega_j|\omega_j\rangle \quad (1.8.13b)$$

Dotting the first with $\langle\omega_j|$ and the second with $\langle\omega_i|$, we get

$$\langle\omega_j|\Omega|\omega_i\rangle = \omega_i\langle\omega_j|\omega_i\rangle \quad (1.8.14a)$$

$$\langle\omega_i|\Omega|\omega_j\rangle = \omega_j\langle\omega_i|\omega_j\rangle \quad (1.8.14b)$$

Taking the adjoint of the last equation and using the Hermitian nature of Ω , we get

$$\langle\omega_j|\Omega|\omega_i\rangle = \omega_j^*\langle\omega_j|\omega_i\rangle$$

Subtracting this equation from Eq. (1.8.14a), we get

$$0 = (\omega_i - \omega_j^*)\langle\omega_j|\omega_i\rangle \quad (1.8.15)$$

If $i=j$, we get, since $\langle\omega_i|\omega_i\rangle \neq 0$,

$$\omega_i = \omega_i^* \quad (1.8.16)$$

If $i \neq j$, we get

$$\langle \omega_i | \omega_j \rangle = 0 \quad (1.8.17)$$

since $\omega_i - \omega_j^* = \omega_i - \omega_j \neq 0$ by assumption. That the proof of orthogonality breaks down for $\omega_i = \omega_j$ is not surprising, for two vectors labeled by a degenerated eigenvalue could be any two members of the degenerate space which need not necessarily be orthogonal. The modification of this proof in this case of degeneracy calls for arguments that are essentially the ones used in proving Theorem 10. The advantage in the way Theorem 10 was proved first is that it suffers no modification in the degenerate case.

Degeneracy

We now address the question of degeneracy as promised earlier. Now, our general analysis of Theorem 10 showed us that in the face of degeneracy, we have not one, but an infinity of orthonormal eigenbases. Let us see through an example how this variety manifests itself when we look for eigenvectors and how it is to be handled.

Example 1.8.5. Consider an operator Ω with matrix elements

$$\Omega \leftrightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

in some basis. The characteristic equation is

$$(\omega - 2)^2 \omega = 0$$

i.e.,

$$\omega = 0, 2, 2$$

The vector corresponding to $\omega = 0$ is found by the usual means to be

$$|\omega = 0\rangle \leftrightarrow \frac{1}{2^{1/2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

The case $\omega = 2$ leads to the following equations for the components of the eigenvector:

$$-x_1 + x_3 = 0$$

$$0 = 0$$

$$x_1 - x_3 = 0$$

Now we have just one equation, instead of the two ($n - 1$) we have grown accustomed to! This is a reflection of the degeneracy. For every extra appearance (besides the first) a root makes, it takes away one equation. Thus degeneracy permits us extra degrees of freedom besides the usual one (of normalization). The conditions

$$x_1 = x_3$$

$$x_2 \text{ arbitrary}$$

define an ensemble of vectors that are perpendicular to the first, $|\omega=0\rangle$, i.e., lie in a plane perpendicular to $|\omega=0\rangle$. This is in agreement with our expectation that a twofold degeneracy should lead to a two-dimensional eigenspace. The freedom in x_2 (or more precisely, the ratio x_2/x_3) corresponds to the freedom of orientation in this plane. Let us arbitrarily choose $x_2 = 1$, to get a normalized eigenvector corresponding to $\omega = 2$:

$$|\omega=2\rangle \leftrightarrow \frac{1}{3^{1/2}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

The third vector is now chosen to lie in this plane and to be orthogonal to the second (being in this plane automatically makes it perpendicular to the first $|\omega=0\rangle$):

$$|\omega=2, \text{ second one}\rangle \leftrightarrow \frac{1}{6^{1/2}} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

Clearly each distinct choice of the ratio, x_2/x_3 , gives us a distinct doublet of orthonormal eigenvectors with eigenvalue 2. \square

Notice that in the face of degeneracy, $|\omega_i\rangle$ no longer refers to a single ket but to a generic element of the eigenspace $\mathbb{V}_{\omega_i}^{m_i}$. To refer to a particular element, we must use the symbol $|\omega_i, \alpha\rangle$, where α labels the ket within the eigenspace. A natural choice of the label α will be discussed shortly.

We now consider the analogs of Theorems 9 and 10 for unitary operators.

Theorem 11. The eigenvalues of a unitary operator are complex numbers of unit modulus.

Theorem 12. The eigenvectors of a unitary operator are mutually orthogonal. (We assume there is no degeneracy.)

Proof of Both Theorems (assuming no degeneracy). Let

$$U|u_i\rangle = u_i|u_i\rangle \quad (1.8.18a)$$

and

$$U|u_j\rangle = u_j|u_j\rangle \quad (1.8.18b)$$

If we take the adjoint of the second equation and dot each side with the corresponding side of the first equation, we get

$$\langle u_j | U^\dagger U | u_i \rangle = u_i u_j^* \langle u_j | u_i \rangle$$

so that

$$(1 - u_i u_j^*) \langle u_j | u_i \rangle = 0 \quad (1.8.19)$$

If $i=j$, we get, since $\langle u_i | u_i \rangle \neq 0$,

$$u_i u_i^* = 1 \quad (1.8.20a)$$

while if $i \neq j$,

$$\langle u_i | u_j \rangle = 0 \quad (1.8.20b)$$

since $|u_i\rangle \neq |u_j\rangle \Rightarrow u_i \neq u_j \Rightarrow u_i u_j^* \neq u_i u_i^* \Rightarrow u_i u_j^* \neq 1$. (Q.E.D.)

If U is degenerate, we can carry out an analysis parallel to that for the Hermitian operator Ω , with just one difference. Whereas in Eq. (1.8.12), the zeros of the first row followed from the zeros of the first column and $\Omega^\dagger = \Omega$, here they follow from the requirement that the sum of the modulus squared of the elements in each row adds up to 1. Since $|u_1| = 1$, all the other elements in the first row must vanish.

Diagonalization of Hermitian Matrices

Consider a Hermitian operator Ω on $\mathbb{V}''(C)$ represented as a matrix in some orthonormal basis $|1\rangle, \dots, |i\rangle, \dots, |n\rangle$. If we trade this basis for the eigenbasis $|\omega_1\rangle, \dots, |\omega_i\rangle, \dots, |\omega_n\rangle$, the matrix representing Ω will become diagonal. Now the operator U inducing the change of basis

$$|\omega_i\rangle = U|i\rangle \quad (1.8.21)$$

is clearly unitary, for it “rotates” one orthonormal basis into another. (If you wish you may apply our mnemonic to U and verify its unitary nature: its columns contain the components of the eigenvectors $|\omega_i\rangle$ that are orthonormal.) This result is often summarized by the statement:

Every Hermitian matrix on $\mathbb{V}''(C)$ may be diagonalized by a unitary change of basis.

We may restate this result in terms of passive transformations as follows:

If Ω is a Hermitian matrix, there exists a unitary matrix U (built out of the eigenvectors of Ω) such that $U^\dagger \Omega U$ is diagonal.

Thus the problem of finding a basis that diagonalizes Ω is equivalent to solving its eigenvalue problem.

Exercise 1.8.1. (1) Find the eigenvalues and normalized eigenvectors of the matrix

$$\Omega = \begin{bmatrix} 1 & 3 & 1 \\ 0 & 2 & 0 \\ 0 & 1 & 4 \end{bmatrix}$$

(2) Is the matrix Hermitian? Are the eigenvectors orthogonal?

*Exercise 1.8.2.** Consider the matrix

$$\Omega = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

- (1) Is it Hermitian?
- (2) Find its eigenvalues and eigenvectors.
- (3) Verify that $U^\dagger \Omega U$ is diagonal, U being the matrix of eigenvectors of Ω .

*Exercise 1.8.3.** Consider the Hermitian matrix

$$\Omega = \frac{1}{2} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & -1 \\ 0 & -1 & 3 \end{bmatrix}$$

- (1) Show that $\omega_1 = \omega_2 = 1$; $\omega_3 = 2$.
- (2) Show that $|\omega=2\rangle$ is any vector of the form

$$\frac{1}{(2a^2)^{1/2}} \begin{bmatrix} 0 \\ a \\ -a \end{bmatrix}$$

- (3) Show that the $\omega=1$ eigenspace contains all vectors of the form

$$\frac{1}{(b^2 + 2c^2)^{1/2}} \begin{bmatrix} b \\ c \\ c \end{bmatrix}$$

either by feeding $\omega=1$ into the equations or by requiring that the $\omega=1$ eigenspace be orthogonal to $|\omega=2\rangle$.

Exercise 1.8.4. An arbitrary $n \times n$ matrix need not have n eigenvectors. Consider as an example

$$\Omega = \begin{bmatrix} 4 & 1 \\ -1 & 2 \end{bmatrix}$$

- (1) Show that $\omega_1 = \omega_2 = 3$.
- (2) By feeding in this value show we get only one eigenvector of the form

$$\frac{1}{(2a^2)^{1/2}} \begin{bmatrix} +a \\ -a \end{bmatrix}$$

We cannot find another one that is LI.

*Exercise 1.8.5.** Consider the matrix

$$\Omega = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$$

- (1) Show that it is unitary.
- (2) Show that its eigenvalues are $e^{i\theta}$ and $e^{-i\theta}$.
- (3) Find the corresponding eigenvectors; show that they are orthogonal.
- (4) Verify that $U^\dagger \Omega U = (\text{diagonal matrix})$, where U is the matrix of eigenvectors of Ω .

*Exercise 1.8.6.** (1) We have seen that the determinant of a matrix is unchanged under a unitary change of basis. Argue now that

$$\det \Omega = \text{product of eigenvalues of } \Omega = \prod_{i=1}^n \omega_i$$

for a Hermitian or unitary Ω .

- (2) Using the invariance of the trace under the same transformation, show that

$$\text{Tr } \Omega = \sum_{i=1}^n \omega_i$$

Exercise 1.8.7. By using the results on the trace and determinant from the last problem, show that the eigenvalues of the matrix

$$\Omega = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

are 3 and -1. Verify this by explicit computation. Note that the Hermitian nature of the matrix is an essential ingredient.

*Exercise 1.8.8.** Consider Hermitian matrices M^1, M^2, M^3, M^4 that obey

$$M^i M^j + M^j M^i = 2\delta^{ij} I, \quad i, j = 1, \dots, 4$$

(1) Show that the eigenvalues of M^i are ± 1 . (Hint: go to the eigenbasis of M^i , and use the equation for $i=j$.)

(2) By considering the relation

$$M^i M^i = -M^j M^i \quad \text{for } i \neq j$$

show that M^i are traceless. [Hint: $\text{Tr}(ACB) = \text{Tr}(CBA)$.]

(3) Show that they cannot be odd-dimensional matrices.

Exercise 1.8.9. A collection of masses m_α , located at \mathbf{r}_α and rotating with angular velocity $\boldsymbol{\omega}$ around a common axis has an angular momentum

$$\mathbf{l} = \sum_{\alpha} m_{\alpha} (\mathbf{r}_{\alpha} \times \mathbf{v}_{\alpha})$$

where $\mathbf{v}_{\alpha} = \boldsymbol{\omega} \times \mathbf{r}_{\alpha}$ is the velocity of m_{α} . By using the identity

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

show that each Cartesian component l_i of \mathbf{l} is given by

$$l_i = \sum_j M_{ij} \omega_j$$

where

$$M_{ij} = \sum_{\alpha} m_{\alpha} [r_{\alpha}^2 \delta_{ij} - (\mathbf{r}_{\alpha})_i (\mathbf{r}_{\alpha})_j]$$

or in Dirac notation

$$|l\rangle = M|\omega\rangle$$

(1) Will the angular momentum and angular velocity always be parallel?

(2) Show that the moment of inertia matrix M_{ij} is Hermitian.

(3) Argue now that there exist three directions for $\boldsymbol{\omega}$ such that \mathbf{l} and $\boldsymbol{\omega}$ will be parallel.

How are these directions to be found?

(4) Consider the moment of inertia matrix of a sphere. Due to the complete symmetry of the sphere, it is clear that every direction is its eigendirection for rotation. What does this say about the three eigenvalues of the matrix M ?

Simultaneous Diagonalization of Two Hermitian Operators

Let us consider next the question of simultaneously diagonalizing two Hermitian operators.

Theorem 13. If Ω and Λ are two commuting Hermitian operators, there exists (at least) a basis of common eigenvectors that diagonalizes them both.

Proof. Consider first the case where at least one of the operators is nondegenerate, i.e., to a given eigenvalue, there is just one eigenvector, up to a scale. Let us assume Ω is nondegenerate. Consider any one of its eigenvectors:

$$\Omega|\omega_i\rangle = \omega_i|\omega_i\rangle$$

$$\Lambda\Omega|\omega_i\rangle = \omega_i\Lambda|\omega_i\rangle$$

Since $[\Lambda, \Omega] = 0$,

$$\Omega\Lambda|\omega_i\rangle = \omega_i\Lambda|\omega_i\rangle \quad (1.8.22)$$

i.e., $\Lambda|\omega_i\rangle$ is an eigenvector of Ω with eigenvalue ω_i . Since this vector is unique up to a scale,

$$\Lambda|\omega_i\rangle = \lambda_i|\omega_i\rangle \quad (1.8.23)$$

Thus $|\omega_i\rangle$ is also an eigenvector of Λ with eigenvalue λ_i . Since every eigenvector of Ω is an eigenvector of Λ , it is evident that the basis $|\omega_i\rangle$ will diagonalize both operators. Since Ω is nondegenerate, there is only one basis with this property.

What if both operators are degenerate? By ordering the basis vectors such that the elements of each eigenspace are adjacent, we can get one of them, say Ω , into the form (Theorem 10)

$$\Omega \leftrightarrow \begin{bmatrix} \omega_1 & & & & \\ & \ddots & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \omega_m \end{bmatrix}$$

Now this basis is not unique: in every eigenspace $\mathbb{V}_{\omega_i}^{m_i} \equiv \mathbb{V}_i^{m_i}$ corresponding to the eigenvalue ω_i , there exists an infinity of bases. Let us arbitrarily pick in $\mathbb{V}_{\omega_i}^{m_i}$ a set $|\omega_i, \alpha\rangle$ where the additional label α runs from 1 to m_i .

How does Λ appear in the basis? Although we made no special efforts to get Λ into a simple form, it already has a simple form by virtue of the fact that it commutes with Ω . Let us start by mimicking the proof in the nondegenerate case:

$$\Omega\Lambda|\omega_i, \alpha\rangle = \Lambda\Omega|\omega_i, \alpha\rangle = \omega_i\Lambda|\omega_i, \alpha\rangle$$

However, due to the degeneracy of Ω , we can only conclude that

$$\Lambda|\omega_i, \alpha\rangle \text{ lies in } \mathbb{V}_i^{m_i}$$

Now, since vectors from different eigenspaces are orthogonal [Eq. (1.8.15)],

$$\langle \omega_j, \beta | \Lambda | \omega_i, \alpha \rangle = 0$$

if $|\omega_i, \alpha\rangle$ and $|\omega_j, \beta\rangle$ are basis vectors such that $\omega_i \neq \omega_j$. Consequently, in this basis,

$$\Lambda \leftrightarrow \begin{bmatrix} \boxed{\Lambda_1} & & & & 0 \\ & \boxed{\Lambda_2} & & & \\ & & \ddots & & \\ 0 & & & & \boxed{\Lambda_k} \end{bmatrix}$$

which is called a *block diagonal matrix* for obvious reasons. The block diagonal form of Λ reflects the fact that when Λ acts on some element $|\omega_i, \alpha\rangle$ of the eigenspace $\mathbb{V}_i^{m_i}$, it turns it into another element of $\mathbb{V}_i^{m_i}$. Within each subspace i , Λ is given by a matrix Λ_i , which appears as a block in the equation above. Consider a matrix Λ_i in $\mathbb{V}_i^{m_i}$. It is Hermitian since Λ is. It can obviously be diagonalized by trading the basis $|\omega_i, 1\rangle, |\omega_i, 2\rangle, \dots, |\omega_i, m_i\rangle$ in $\mathbb{V}_i^{m_i}$ that we started with, for the eigenbasis of Λ_i . Let us make such a change of basis in each eigenspace, thereby rendering Λ diagonal. Meanwhile what of Ω ? It remains diagonal of course, since it is indifferent to the choice of orthonormal basis in each degenerate eigenspace. If the eigenvalues of Λ_i are $\lambda_i^{(1)}, \lambda_i^{(2)}, \dots, \lambda_i^{(m_i)}$ then we end up with

$$\Lambda \leftrightarrow \begin{bmatrix} \lambda_1^{(1)} & & & & \\ & \lambda_1^{(2)} & & & \\ & & \ddots & & \\ & & & \lambda_1^{(m_1)} & \\ & & & & \lambda_2^{(1)} \\ & & & & & \ddots \\ & & & & & & \lambda_k^{(m_k)} \end{bmatrix},$$

$$\Omega \leftrightarrow \begin{bmatrix} \omega_1 & & & & \\ & \omega_1 & & & \\ & & \ddots & & \\ & & & \omega_1 & \\ & & & & \omega_2 \\ & & & & & \ddots \\ & & & & & & \omega_m \end{bmatrix}$$

Q.E.D.

If Λ is not degenerate *within any given subspace*, $\lambda_i^{(k)} \neq \lambda_i^{(l)}$, for any k, l , and i , the basis we end up with is unique: the freedom Ω gave us in each eigenspace is fully eliminated by Λ . The elements of this basis may be named uniquely by the pair of indices ω and λ as $|\omega, \lambda\rangle$, with λ playing the role of the extra label α . If Λ is degenerate within an eigenspace of Ω , if say $\lambda_i^{(1)} = \lambda_i^{(2)}$, there is a two-dimensional eigenspace from which we can choose any two orthonormal vectors for the common basis. It is then necessary to bring in a third operator Γ , that commutes with both Ω and Λ , and which will be nondegenerate in this subspace. In general, one can always find, for finite n , a set of operators $\{\Omega, \Lambda, \Gamma, \dots\}$ that commute with each other and that nail down a unique, common, eigenbasis, the elements of which may be labeled unambiguously as $|\omega, \lambda, \gamma, \dots\rangle$. In our study of quantum mechanics it will be assumed that such a *complete set of commuting operators* exists if n is infinite.

*Exercise 1.8.10.** By considering the commutator, show that the following Hermitian matrices may be simultaneously diagonalized. Find the eigenvectors common to both and verify that under a unitary transformation to this basis, both matrices are diagonalized.

$$\Omega = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 2 \end{bmatrix}$$

Since Ω is degenerate and Λ is not, you must be prudent in deciding which matrix dictates the choice of basis.

Example 1.8.6. We will now discuss, in some detail, the complete solution to a problem in mechanics. It is important that you understand this example thoroughly, for it not only illustrates the use of the mathematical techniques developed in this chapter but also contains the main features of the central problem in quantum mechanics.

The mechanical system in question is depicted in Fig. 1.5. The two masses m are coupled to each other and the walls by springs of force constant k . If x_1 and x_2 measure the displacements of the masses from their equilibrium points, these coordinates obey the following equations, derived through an elementary application of Newton's laws:

$$\ddot{x}_1 = -\frac{2k}{m}x_1 + \frac{k}{m}x_2 \quad (1.8.24a)$$

$$\ddot{x}_2 = \frac{k}{m}x_1 - \frac{2k}{m}x_2 \quad (1.8.24b)$$

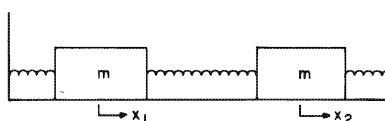


Figure 1.5. The coupled mass problem. All masses are m , all spring constants are k , and the displacements of the masses from equilibrium are x_1 and x_2 .

The problem is to find $x_1(t)$ and $x_2(t)$ given the initial-value data, which in this case consist of the initial positions and velocities. If we restrict ourselves to the case of zero initial velocities, our problem is to find $x_1(t)$ and $x_2(t)$, given $x_1(0)$ and $x_2(0)$.

In what follows, we will formulate the problem in the language of linear vector spaces and solve it using the machinery developed in this chapter. As a first step, we rewrite Eq. (1.8.24) in matrix form:

$$\begin{bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (1.8.25a)$$

where the elements of the *Hermitian* matrix Ω_{ij} are

$$\Omega_{11} = \Omega_{22} = -2k/m, \quad \Omega_{12} = \Omega_{21} = k/m \quad (1.8.25b)$$

We now view x_1 and x_2 as components of an abstract vector $|x\rangle$, and Ω_{ij} as the matrix elements of a Hermitian operator Ω . Since the vector $|x\rangle$ has two real components, it is an element of $\mathbb{V}^2(R)$, and Ω is a Hermitian operator on $\mathbb{V}^2(R)$. The abstract form of Eq. (1.8.25a) is

$$|\ddot{x}(t)\rangle = \Omega|x(t)\rangle \quad (1.8.26)$$

Equation (1.8.25a) is obtained by projecting Eq. (1.8.26) on the basis vectors $|1\rangle$, $|2\rangle$, which have the following physical significance:

$$|1\rangle \leftrightarrow \begin{bmatrix} 1 \\ 0 \end{bmatrix} \leftrightarrow \begin{bmatrix} \text{first mass displaced by unity} \\ \text{second mass undisplaced} \end{bmatrix} \quad (1.8.27a)$$

$$|2\rangle \leftrightarrow \begin{bmatrix} 0 \\ 1 \end{bmatrix} \leftrightarrow \begin{bmatrix} \text{first mass undisplaced} \\ \text{second mass displaced by unity} \end{bmatrix} \quad (1.8.27b)$$

An arbitrary state, in which the masses are displaced by x_1 and x_2 , is given in this basis by

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} x_1 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} x_2 \quad (1.8.28)$$

The abstract counterpart of the above equation is

$$|x\rangle = |1\rangle x_1 + |2\rangle x_2 \quad (1.8.29)$$

It is in this $|1\rangle$, $|2\rangle$ basis that Ω is represented by the matrix appearing in Eq. (1.8.25), with elements $-2k/m$, k/m , etc.

The basis $|1\rangle$, $|2\rangle$ is very desirable physically, for the components of $|x\rangle$ in this basis (x_1 and x_2) have the simple interpretation as displacements of the masses. However, from the standpoint of finding a mathematical solution to the initial-value problem, it is not so desirable, for the components x_1 and x_2 obey the *coupled*

differential equations (1.8.24a) and (1.8.24b). The coupling is mediated by the off-diagonal matrix elements $\Omega_{12} = \Omega_{21} = k/m$.

Having identified the problem with the $|1\rangle, |2\rangle$ basis, we can now see how to get around it: we must switch to a basis in which Ω is diagonal. The components of $|x\rangle$ in this basis will then obey another uncoupled differential equations which may be readily solved. Having found the solution, we can return to the physically preferable $|1\rangle, |2\rangle$ basis. This, then, is our broad strategy and we now turn to the details.

From our study of Hermitian operators we know that the basis that diagonalizes Ω is the basis of its normalized eigenvectors. Let $|I\rangle$ and $|II\rangle$ be its eigenvectors defined by

$$\Omega|I\rangle = -\omega_I^2|I\rangle \quad (1.8.30a)$$

$$\Omega|II\rangle = -\omega_{II}^2|II\rangle \quad (1.8.30b)$$

We are departing here from our usual notation: the eigenvalue of Ω is written as $-\omega^2$ rather than as ω in anticipation of the fact that Ω has eigenvalues of the form $-\omega^2$, with ω real. We are also using the symbols $|I\rangle$ and $|II\rangle$ to denote what should be called $|- \omega_I^2\rangle$ and $|- \omega_{II}^2\rangle$ in our convention.

It is a simple exercise (which you should perform) to solve the eigenvalue problem of Ω in the $|1\rangle, |2\rangle$ basis (in which the matrix elements of Ω are known) and to obtain

$$\omega_I = \left(\frac{k}{m}\right)^{1/2}, \quad |I\rangle \leftrightarrow \frac{1}{2^{1/2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (1.8.31a)$$

$$\omega_{II} = \left(\frac{3k}{m}\right)^{1/2}, \quad |II\rangle \leftrightarrow \frac{1}{2^{1/2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad (1.8.31b)$$

If we now expand the vector $|x(t)\rangle$ in this new basis as

$$|x(t)\rangle = |I\rangle x_I(t) + |II\rangle x_{II}(t) \quad (1.8.32)$$

[in analogy with Eq. (1.8.29)], the components x_I and x_{II} will evolve as follows:

$$\begin{aligned} \begin{bmatrix} \dot{x}_I \\ \dot{x}_{II} \end{bmatrix} &= \begin{bmatrix} -\omega_I^2 & 0 \\ 0 & -\omega_{II}^2 \end{bmatrix} \begin{bmatrix} x_I \\ x_{II} \end{bmatrix} \\ &= \begin{bmatrix} -\omega_I^2 x_I \\ -\omega_{II}^2 x_{II} \end{bmatrix} \end{aligned} \quad (1.8.33)$$

We obtain this equation by rewriting Eq. (1.8.26) in the $|I\rangle, |II\rangle$ basis in which Ω has its eigenvalues as the diagonal entries, and in which $|x\rangle$ has components x_I and

x_{II} . Alternately we can apply the operator

$$\frac{d^2}{dt^2} - \Omega^2$$

to both sides of the expansion of Eq. (1.8.32), and get

$$|0\rangle = |I\rangle(\ddot{x}_I + \omega_I^2 x_I) + |II\rangle(\ddot{x}_{II} + \omega_{II}^2 x_{II}) \quad (1.8.34)$$

Since $|I\rangle$ and $|II\rangle$ are orthogonal, each coefficient is zero.

The solution to the *decoupled* equations

$$\ddot{x}_i + \omega_i^2 x_i = 0, \quad i = I, II \quad (1.8.35)$$

subject to the condition of vanishing initial velocities, is

$$x_i(t) = x_i(0) \cos \omega_i t, \quad i = I, II \quad (1.8.36)$$

As anticipated, the components of $|x\rangle$ in the $|I\rangle, |II\rangle$ basis obey decoupled equations that can be readily solved. Feeding Eq. (1.8.36) into Eq. (1.8.32) we get

$$|x(t)\rangle = |I\rangle x_I(0) \cos \omega_I t + |II\rangle x_{II}(0) \cos \omega_{II} t \quad (1.8.37a)$$

$$= |I\rangle \langle I|x(0)\rangle \cos \omega_I t + |II\rangle \langle II|x(0)\rangle \cos \omega_{II} t \quad (1.8.37b)$$

Equation (1.8.37) provides the explicit solution to the initial-value problem. It corresponds to the following algorithm for finding $|x(t)\rangle$ given $|x(0)\rangle$.

Step (1). Solve the eigenvalue problem of Ω .

Step (2). Find the coefficients $x_I(0) = \langle I|x(0)\rangle$ and $x_{II}(0) = \langle II|x(0)\rangle$ in the expansion

$$|x(0)\rangle = |I\rangle x_I(0) + |II\rangle x_{II}(0)$$

Step (3). Append to each coefficient $x_i(0)$ ($i = I, II$) a time dependence $\cos \omega_i t$ to get the coefficients in the expansion of $|x(t)\rangle$.

Let me now illustrate this algorithm by solving the following (general) initial-value problem: Find the future state of the system given that at $t=0$ the masses are displaced by $x_1(0)$ and $x_2(0)$.

Step (1). We can ignore this step since the eigenvalue problem has been solved [Eq. (1.8.31)].

Step (2).

$$x_{\text{I}}(0) = \langle \text{I} | x(0) \rangle = \frac{1}{2^{1/2}} (1, 1) \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \frac{x_1(0) + x_2(0)}{2^{1/2}}$$

$$x_{\text{II}}(0) = \langle \text{II} | x(0) \rangle = \frac{1}{2^{1/2}} (1, -1) \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \frac{x_1(0) - x_2(0)}{2^{1/2}}$$

Step (3).

$$|x(t)\rangle = |\text{I}\rangle \frac{x_1(0) + x_2(0)}{2^{1/2}} \cos \omega_{\text{I}} t + |\text{II}\rangle \frac{x_1(0) - x_2(0)}{2^{1/2}} \cos \omega_{\text{II}} t$$

The explicit solution above can be made even more explicit by projecting $|x(t)\rangle$ onto the $|1\rangle, |2\rangle$ basis to find $x_1(t)$ and $x_2(t)$, the displacements of the masses. We get (feeding in the explicit formulas for ω_{I} and ω_{II})

$$\begin{aligned} x_1(t) &= \langle 1 | x(t) \rangle \\ &= \langle 1 | \text{I} \rangle \frac{x_1(0) + x_2(0)}{2^{1/2}} \cos \left[\left(\frac{k}{m} \right)^{1/2} t \right] + \langle 1 | \text{II} \rangle \frac{x_1(0) - x_2(0)}{2^{1/2}} \cos \left[\left(\frac{3k}{m} \right)^{1/2} t \right] \\ &= \frac{1}{2} [x_1(0) + x_2(0)] \cos \left[\left(\frac{k}{m} \right)^{1/2} t \right] + \frac{1}{2} [x_1(0) - x_2(0)] \cos \left[\left(\frac{3k}{m} \right)^{1/2} t \right] \end{aligned} \quad (1.8.38a)$$

using the fact that

$$\langle 1 | \text{I} \rangle = \langle 1 | \text{II} \rangle = 1/2^{1/2}$$

It can likewise be shown that

$$x_2(t) = \frac{1}{2} [x_1(0) + x_2(0)] \cos \left[\left(\frac{k}{m} \right)^{1/2} t \right] - \frac{1}{2} [x_1(0) - x_2(0)] \cos \left[\left(\frac{3k}{m} \right)^{1/2} t \right] \quad (1.8.38b)$$

We can rewrite Eq. (1.8.38) in matrix form as

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \cos[(k/m)^{1/2}t] + \cos[(3k/m)^{1/2}t] & \cos[(k/m)^{1/2}t] - \cos[(3k/m)^{1/2}t] \\ 2 & 2 \\ \cos[(k/m)^{1/2}t] - \cos[(3k/m)^{1/2}t] & \cos[(k/m)^{1/2}t] + \cos[(3k/m)^{1/2}t] \\ 2 & 2 \end{bmatrix} \times \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} \quad (1.8.39)$$

This completes our determination of the future state of the system given the initial state.

The Propagator

There are two remarkable features in Eq. (1.8.39):

- (1) The final-state vector is obtained from the initial-state vector upon multiplication by a matrix.
- (2) This matrix is independent of the initial state. We call this matrix the *propagator*. Finding the propagator is tantamount to finding the complete solution to the problem, for given any other initial state with displacements $\tilde{x}_1(0)$ and $\tilde{x}_2(0)$, we get $\tilde{x}_1(t)$ and $\tilde{x}_2(t)$ by applying the same matrix to the initial-state vector.

We may view Eq. (1.8.39) as the image in the $|1\rangle$, $|2\rangle$ basis of the abstract relation

$$|x(t)\rangle = U(t)|x(0)\rangle \quad (1.8.40)$$

By comparing this equation with Eq. (1.8.37b), we find the abstract representation of U :

$$U(t) = |I\rangle\langle I| \cos \omega_I t + |II\rangle\langle II| \cos \omega_{II} t \quad (1.8.41a)$$

$$= \sum_{i=1}^{II} |i\rangle\langle i| \cos \omega_i t \quad (1.8.41b)$$

You may easily convince yourself that if we take the matrix elements of this operator in the $|1\rangle$, $|2\rangle$ basis, we regain the matrix appearing in Eq. (1.8.39). For example

$$\begin{aligned} U_{11} &= \langle 1| U | 1 \rangle \\ &= \langle 1| \left\{ |I\rangle\langle I| \cos \left[\left(\frac{k}{m} \right)^{1/2} t \right] + |II\rangle\langle II| \cos \left[\left(\frac{3k}{m} \right)^{1/2} t \right] \right\} |1\rangle \\ &= \langle 1| I \rangle \langle I | 1 \rangle \cos \left[\left(\frac{k}{m} \right)^{1/2} t \right] + \langle 1| II \rangle \langle II | 1 \rangle \cos \left[\left(\frac{3k}{m} \right)^{1/2} t \right] \\ &= \frac{1}{2} \left\{ \cos \left[\left(\frac{k}{m} \right)^{1/2} t \right] + \cos \left[\left(\frac{3k}{m} \right)^{1/2} t \right] \right\} \end{aligned}$$

Notice that $U(t)$ [Eq. (1.8.41)] is determined completely by the eigenvectors and eigenvalues of Ω . We may then restate our earlier algorithm as follows. To solve the equation

$$|\ddot{x}\rangle = \Omega|x\rangle$$

- (1) Solve the eigenvalue problem of Ω .
- (2) Construct the propagator U in terms of the eigenvalues and eigenvectors.
- (3) $|x(t)\rangle = U(t)|x(0)\rangle$.

The Normal Modes

There are two initial states $|x(0)\rangle$ for which the time evolution is particularly simple. Not surprisingly, these are the eigenkets $|I\rangle$ and $|II\rangle$. Suppose we have $|x(0)\rangle = |I\rangle$. Then the state at time t is

$$\begin{aligned} |I(t)\rangle &\equiv U(t)|I\rangle \\ &= (|I\rangle\langle I| \cos \omega_1 t + |II\rangle\langle II| \cos \omega_{II} t)|I\rangle \\ &= |I\rangle \cos \omega_1 t \end{aligned} \quad (1.8.42)$$

Thus the system starting off in $|I\rangle$ is only modified by an overall factor $\cos \omega_1 t$. A similar remark holds with $I \rightarrow II$. These two modes of vibration, in which all (two) components of a vector oscillate in step are called *normal modes*.

The physics of the normal modes is clear in the $|1\rangle$, $|2\rangle$ basis. In this basis

$$|I\rangle \leftrightarrow \frac{1}{2^{1/2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and corresponds to a state in which both masses are displaced by equal amounts. The middle spring is then a mere spectator and each mass oscillates with a frequency $\omega_1 = (k/m)^{1/2}$ in response to the end spring nearest to it. Consequently

$$|I(t)\rangle \leftrightarrow \frac{1}{2^{1/2}} \begin{bmatrix} \cos[(k/m)^{1/2}t] \\ \cos[(k/m)^{1/2}t] \end{bmatrix}$$

On the other hand, if we start with

$$|II\rangle \leftrightarrow \frac{1}{2^{1/2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

the masses are displaced by equal and opposite amounts. In this case the middle spring is distorted by *twice* the displacement of each mass. If the masses are adjusted by Δ and $-\Delta$, respectively, each mass feels a restoring force of $3k\Delta$ ($2k\Delta$ from the middle spring and $k\Delta$ from the end spring nearest to it). Since the effective force constant is $k_{\text{eff}} = 3k\Delta/\Delta = 3k$, the vibrational frequency is $(3k/m)^{1/2}$ and

$$|II(t)\rangle \leftrightarrow \frac{1}{2^{1/2}} \begin{bmatrix} \cos[(3k/m)^{1/2}t] \\ -\cos[(3k/m)^{1/2}t] \end{bmatrix}$$

If the system starts off in a linear combination of $|I\rangle$ and $|II\rangle$ it evolves into the corresponding linear combination of the normal modes $|I(t)\rangle$ and $|II(t)\rangle$. This

$$\begin{aligned}
 |x(t)\rangle &= U(t)|x(0)\rangle \\
 &= |I\rangle\langle I|x(0)\rangle \cos \omega_I t + |II\rangle\langle II|x(0)\rangle \cos \omega_{II} t \\
 &= |I(t)\rangle\langle I|x(0)\rangle + |II(t)\rangle\langle II|x(0)\rangle
 \end{aligned}$$

Another way to see the simple evolution of the initial states $|I\rangle$ and $|II\rangle$ is to determine the matrix representing U in the $|I\rangle$, $|II\rangle$ basis:

$$U \xleftarrow[\substack{\text{I,II} \\ \text{basis}}]{} \begin{bmatrix} \cos \omega_I t & 0 \\ 0 & \cos \omega_{II} t \end{bmatrix} \quad (1.8.43)$$

You should verify this result by taking the appropriate matrix elements of $U(t)$ in Eq. (1.8.41b). Since each column above is the image of the corresponding basis vectors ($|I\rangle$ or $|II\rangle$) after the action of $U(t)$, (which is to say, after time evolution), we see that the initial states $|I\rangle$ and $|II\rangle$ evolve simply in time.

The central problem in quantum mechanics is very similar to the simple example that we have just discussed. The state of the system is described in quantum theory by a ket $|\psi\rangle$ which obeys the Schrödinger equation

$$i\hbar|\dot{\psi}\rangle = H|\psi\rangle$$

where \hbar is a constant related to Planck's constant h by $\hbar = h/2\pi$, and H is a *Hermitian* operator called the Hamiltonian. The problem is to find $|\psi(t)\rangle$ given $|\psi(0)\rangle$. [Since the equation is first order in t , no assumptions need be made about $|\dot{\psi}(0)\rangle$, which is determined by the Schrödinger equation to be $(-i/\hbar)H|\psi(0)\rangle$.]

In most cases, H is a time-independent operator and the algorithm one follows in solving this initial-value problem is completely analogous to the one we have just seen:

Step (1). Solve the eigenvalue problem of H .

Step (2). Find the propagator $U(t)$ in terms of the eigenvectors and eigenvalues of H .

Step (3). $|\psi(t)\rangle = U(t)|\psi(0)\rangle$.

You must of course wait till Chapter 4 to find out the physical interpretation of $|\psi\rangle$, the actual form of the operator H , and the precise relation between $U(t)$ and the eigenvalues and eigenvectors of H . \square

Exercise 1.8.11. Consider the coupled mass problem discussed above.

(1) Given that the initial state is $|1\rangle$, in which the first mass is displaced by unity and the second is left alone, calculate $|1(t)\rangle$ by following the algorithm.

(2) Compare your result with that following from Eq. (1.8.39).

Exercise 1.8.12. Consider once again the problem discussed in the previous example.

(1) Assuming that

$$|\dot{x}\rangle = \Omega|x\rangle$$

has a solution

$$|x(t)\rangle = U(t)|x(0)\rangle$$

find the differential equation satisfied by $U(t)$. Use the fact that $|x(0)\rangle$ is arbitrary.

(2) Assuming (as is the case) that Ω and U can be simultaneously diagonalized, solve for the elements of the matrix U in this common basis and regain Eq. (1.8.43). Assume $|\dot{x}(0)\rangle = 0$.

1.9. Functions of Operators and Related Concepts

We have encountered two types of objects that act on vectors: scalars, which commute with each other and with all operators; and operators, which do not generally commute with each other. It is customary to refer to the former as c numbers and the latter as q numbers. Now, we are accustomed to functions of c numbers such as $\sin(x)$, $\log(x)$, etc. We wish to examine the question whether functions of q numbers can be given a sensible meaning. We will restrict ourselves to those functions that can be written as a power series. Consider a series

$$f(x) = \sum_{n=0}^{\infty} a_n x^n \quad (1.9.1)$$

where x is a c number. We define the same function of an operator or q number to be

$$f(\Omega) = \sum_{n=0}^{\infty} a_n \Omega^n \quad (1.9.2)$$

This definition makes sense only if the sum converges to a definite limit. To see what this means, consider a common example:

$$e^{\Omega} = \sum_{n=1}^{\infty} \frac{\Omega^n}{n!} \quad (1.9.3)$$

Let us restrict ourselves to Hermitian Ω . By going to the eigenbasis of Ω we can readily perform the sum of Eq. (1.9.3). Since

$$\Omega = \begin{bmatrix} \omega_1 & & & \\ & \omega_2 & & \\ & & \ddots & \\ & & & \omega_n \end{bmatrix} \quad (1.9.4)$$

and

$$\Omega^m = \begin{bmatrix} \omega_1^m & & & \\ & \omega_2^m & & \\ & & \ddots & \\ & & & \omega_n^m \end{bmatrix} \quad (1.9.5)$$

$$e^{\Omega} = \begin{bmatrix} \sum_{m=0}^{\infty} \frac{\omega_1^m}{m!} & & & \\ & \ddots & & \\ & & \sum_{m=0}^{\infty} \frac{\omega_n^m}{m!} & \end{bmatrix} \quad (1.9.6)$$

Since each sum converges to the familiar limit e^{ω_i} , the operator e^{Ω} is indeed well defined by the power series in this basis (and therefore in any other).

*Exercise 1.9.1.** We know that the series

$$f(x) = \sum_{n=0}^{\infty} x^n$$

may be equated to the function $f(x) = (1 - x)^{-1}$ if $|x| < 1$. By going to the eigenbasis, examine when the q number power series

$$f(\Omega) = \sum_{n=0}^{\infty} \Omega^n$$

of a Hermitian operator Ω may be identified with $(1 - \Omega)^{-1}$.

*Exercise 1.9.2.** If H is a Hermitian operator, show that $U = e^{iH}$ is unitary. (Notice the analogy with c numbers: if θ is real, $u = e^{i\theta}$ is a number of unit modulus.)

Exercise 1.9.3. For the case above, show that $\det U = e^{i\text{Tr } H}$.

Derivatives of Operators with Respect to Parameters

Consider next an operator $\theta(\lambda)$ that depends on a parameter λ . Its derivative with respect to λ is defined to be

$$\frac{d\theta(\lambda)}{d\lambda} = \lim_{\Delta\lambda \rightarrow 0} \left[\frac{\theta(\lambda + \Delta\lambda) - \theta(\lambda)}{\Delta\lambda} \right]$$

If $\theta(\lambda)$ is written as a matrix in some basis, then the matrix representing $d\theta(\lambda)/d\lambda$ is obtained by differentiating the matrix elements of $\theta(\lambda)$. A special case of $\theta(\lambda)$ we

are interested in is

$$\theta(\lambda) = e^{\lambda\Omega}$$

where Ω is Hermitian. We can show, by going to the eigenbasis of Ω , that

$$\frac{d\theta(\lambda)}{d\lambda} = \Omega e^{\lambda\Omega} = e^{\lambda\Omega}\Omega = \theta(\lambda)\Omega \quad (1.9.7)$$

The same result may be obtained, even if Ω is not Hermitian, by working with the power series, provided it exists:

$$\frac{d}{d\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n \Omega^n}{n!} = \sum_{n=1}^{\infty} \frac{n\lambda^{n-1} \Omega^n}{n!} = \Omega \sum_{n=1}^{\infty} \frac{\lambda^{n-1} \Omega^{n-1}}{(n-1)!} = \Omega \sum_{m=0}^{\infty} \frac{\lambda^m \Omega^m}{m!} = \Omega e^{\lambda\Omega}$$

Conversely, we can say that if we are confronted with the differential Eq. (1.9.7), its solution is given by

$$\theta(\lambda) = c \exp\left(\int_0^\lambda \Omega d\lambda'\right) = c \exp(\Omega\lambda)$$

(It is assumed here that the exponential exists.) In the above, c is a constant (operator) of integration. The solution $\theta = e^{\lambda\Omega}$ corresponds to the choice $c = I$.

In all the above operations, we see that Ω behaves as if it were just a c number. Now, the real difference between c numbers and q numbers is that the latter do not generally commute. However, if only one q number (or powers of it) enter the picture, everything commutes and we can treat them as c numbers. If one remembers this mnemonic, one can save a lot of time.

If, on the other hand, more than one q number is involved, the order of the factors is all important. For example, it is true that

$$e^{\alpha\Omega} e^{\beta\Omega} = e^{(\alpha + \beta)\Omega}$$

as may be verified by a power-series expansion, while it is not true that

$$e^{\alpha\Omega} e^{\beta\theta} = e^{\alpha\Omega + \beta\theta}$$

or that

$$e^{\alpha\Omega} e^{\beta\theta} e^{-\alpha\Omega} = e^{\beta\theta}$$

unless $[\Omega, \theta] = 0$. Likewise, in differentiating a product, the chain rule is

$$\frac{d}{d\lambda} e^{\lambda\Omega} e^{\lambda\theta} = \Omega e^{\lambda\Omega} e^{\lambda\theta} + e^{\lambda\Omega} e^{\lambda\theta} \theta \quad (1.9.8)$$

We are free to move Ω through $e^{\lambda\Omega}$ and write the first term as

$$e^{\lambda\Omega}\Omega e^{\lambda\theta}$$

but not as

$$e^{\lambda\Omega}e^{\lambda\theta}\Omega$$

unless $[\Omega, \theta] = 0$.

1.10. Generalization to Infinite Dimensions

In all of the preceding discussions, the dimensionality (n) of the space was unspecified but assumed to be some finite number. We now consider the generalization of the preceding concepts to infinite dimensions.

Let us begin by getting acquainted with an infinite-dimensional vector. Consider a function defined in some interval, say, $a \leq x \leq b$. A concrete example is provided by the displacement $f(x, t)$ of a string clamped at $x=0$ and $x=L$ (Fig. 1.6).

Suppose we want to communicate to a person on the moon the string's displacement $f(x)$, at some time t . One simple way is to divide the interval $0-L$ into 20 equal parts, measure the displacement $f(x_i)$ at the 19 points $x=L/20, 2L/20, \dots, 19L/20$, and transmit the 19 values on the wireless. Given these $f(x_i)$, our friend on the moon will be able to reconstruct the approximate picture of the string shown in Fig. 1.7.

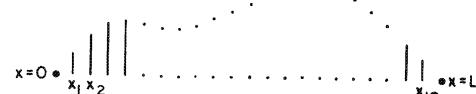
If we wish to be more accurate, we can specify the values of $f(x)$ at a larger number of points. Let us denote by $f_n(x)$ the discrete approximation to $f(x)$ that coincides with it at n points and vanishes in between. Let us now interpret the ordered n -tuple $\{f_n(x_1), f_n(x_2), \dots, f_n(x_n)\}$ as components of a ket $|f_n\rangle$ in a vector space $\mathbb{V}^n(R)$:

$$|f_n\rangle \leftrightarrow \begin{bmatrix} f_n(x_1) \\ f_n(x_2) \\ \vdots \\ f_n(x_n) \end{bmatrix} \quad (1.10.1)$$

Figure 1.6. The string is clamped at $x=0$ and $x=L$. It is free to oscillate in the plane of the paper.



Figure 1.7. The string as reconstructed by the person on the moon.



The basis vectors in this space are

$$|x_i\rangle \leftrightarrow \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i\text{th place} \quad (1.10.2)$$

corresponding to the discrete function which is unity at $x=x_i$ and zero elsewhere. The basis vectors satisfy

$$\langle x_i | x_j \rangle = \delta_{ij} \text{ (orthogonality)} \quad (1.10.3)$$

$$\sum_{i=1}^n |x_i\rangle \langle x_i| = I \text{ (completeness)} \quad (1.10.4)$$

Try to imagine a space containing n mutually perpendicular axes, one for each point x_i . Along each axis is a unit vector $|x_i\rangle$. The function $f_n(x)$ is represented by a vector whose projection along the i th direction is $f_n(x_i)$:

$$|f_n\rangle = \sum_{i=1}^n f_n(x_i) |x_i\rangle \quad (1.10.5)$$

To every possible discrete approximation $g_n(x)$, $h_n(x)$, etc., there is a corresponding ket $|g_n\rangle$, $|h_n\rangle$, etc., and vice versa. You should convince yourself that if we define vector addition as the addition of the components, and scalar multiplication as the multiplication of each component by the scalar, then the set of all kets representing discrete functions that vanish at $x=0, L$ and that are specified at n points in between, forms a vector space.

We next define the inner product in this space:

$$\langle f_n | g_n \rangle = \sum_{i=1}^n f_n(x_i) g_n(x_i) \quad (1.10.6)$$

Two functions $f_n(x)$ and $g_n(x)$ will be said to be orthogonal if $\langle f_n | g_n \rangle = 0$.

Let us now forget the man on the moon and consider the maximal specification of the string's displacement, by giving its value at every point in the interval $0-L$. In this case $f_\infty(x) \equiv f(x)$ is specified by an ordered infinity of numbers: an $f(x)$ for each point x . Each function is now represented by a ket $|f_\infty\rangle$ in an infinite-dimensional vector space and vice versa. Vector addition and scalar multiplication are defined just as before. Consider, however, the inner product. For finite n it was

defined as

$$\langle f_n | g_n \rangle = \sum_{i=1}^n f_n(x_i) g_n(x_i)$$

in particular

$$\langle f_n | f_n \rangle = \sum_{i=1}^n [f_n(x_i)]^2$$

If we now let n go to infinity, so does the sum, for practically any function. What we need is the redefinition of the inner product for finite n in such a way that as n tends to infinity, a smooth limit obtains. The natural choice is of course

$$\langle f_n | g_n \rangle = \sum_{i=1}^n f_n(x_i) g_n(x_i) \Delta, \quad \Delta = L/(n+1) \quad (1.10.6')$$

If we now let n go to infinity, we get, by the usual definition of the integral,

$$\langle f | g \rangle = \int_0^L f(x) g(x) dx \quad (1.10.7)$$

$$\langle f | f \rangle = \int_0^L f^2(x) dx \quad (1.10.8)$$

If we wish to go beyond the instance of the string and consider complex functions of x as well, in some interval $a \leq x \leq b$, the only modification we need is in the inner product:

$$\langle f | g \rangle = \int_a^b f^*(x) g(x) dx \quad (1.10.9)$$

What are the basis vectors in this space and how are they normalized? We know that each point x gets a basis vector $|x\rangle$. The orthogonality of two different axes requires that

$$\langle x | x' \rangle = 0, \quad x \neq x' \quad (1.10.10)$$

What if $x=x'$? Should we require, as in the finite-dimensional case, $\langle x | x \rangle = 1$? The answer is no, and the best way to see it is to deduce the correct normalization. We start with the natural generalization of the completeness relation Eq. (1.10.4) to the case where the kets are labeled by a continuous index x' :

$$\int_a^b |x'\rangle \langle x'| dx' = I \quad (1.10.11)$$

where, as always, the identity is required to leave each ket unchanged. Dotting both sides of Eq. (1.10.11) with some arbitrary ket $|f\rangle$ from the right and the basis bra $\langle x|$ from the left,

$$\int_a^b \langle x|x'\rangle \langle x'|f\rangle dx' = \langle x|I|f\rangle = \langle x|f\rangle \quad (1.10.12)$$

Now, $\langle x|f\rangle$, the projection of $|f\rangle$ along the basis ket $|x\rangle$, is just $f(x)$. Likewise $\langle x'|f\rangle = f(x')$. Let the inner product $\langle x|x'\rangle$ be some unknown function $\delta(x, x')$. Since $\delta(x, x')$ vanishes if $x \neq x'$ we can restrict the integral to an infinitesimal region near $x' = x$ in Eq. (1.10.12):

$$\int_{x-\varepsilon}^{x+\varepsilon} \delta(x, x') f(x') dx' = f(x) \quad (1.10.13)$$

In this infinitesimal region, $f(x')$ (for any reasonably smooth f) can be approximated by its value at $x' = x$, and pulled out of the integral:

$$f(x) \int_{x-\varepsilon}^{x+\varepsilon} \delta(x, x') dx' = f(x) \quad (1.10.14)$$

so that

$$\int_{x-\varepsilon}^{x+\varepsilon} \delta(x, x') dx' = 1 \quad (1.10.15)$$

Clearly $\delta(x, x')$ cannot be finite at $x' = x$, for then its integral over an infinitesimal region would also be infinitesimal. In fact $\delta(x, x')$ should be infinite in such a way that its integral is unity. Since $\delta(x, x')$ depends only on the difference $x - x'$, let us write it as $\delta(x - x')$. The “function,” $\delta(x - x')$, with the properties

$$\begin{aligned} \delta(x - x') &= 0, & x \neq x' \\ \int_a^b \delta(x - x') dx' &= 1, & a < x < b \end{aligned} \quad (1.10.16)$$

is called the *Dirac delta function* and fixes the normalization of the basis vectors:

$$\langle x|x'\rangle = \delta(x - x') \quad (1.10.17)$$

It will be needed any time the basis kets are labeled by a continuous index such as x . Note that it is defined only in the context of an integration: the integral of the delta function $\delta(x - x')$ with any smooth function $f(x')$ is $f(x)$. One sometimes calls

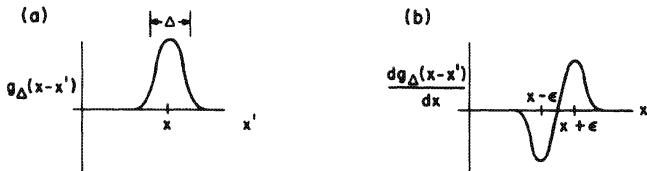


Figure 1.8. (a) The Gaussian g_Δ approaches the delta function as $\Delta \rightarrow 0$. (b) Its derivative $(dg/dx)(x - x')$ approaches $\delta'(x - x')$ as $\Delta \rightarrow 0$.

the delta function the sampling function, since it samples the value of the function $f(x')$ at one point‡

$$\int \delta(x - x') f(x') dx' = f(x) \quad (1.10.18)$$

The delta function does not look like any function we have seen before, its values being either infinite or zero. It is therefore useful to view it as the limit of a more conventional function. Consider a Gaussian

$$g_\Delta(x - x') = \frac{1}{(\pi\Delta^2)^{1/2}} \exp\left[-\frac{(x - x')^2}{\Delta^2}\right] \quad (1.10.19)$$

as shown in Fig. 1.8a. The Gaussian is centered at $x' = x$, has width Δ , maximum height $(\pi\Delta^2)^{-1/2}$, and unit area, independent of Δ . As Δ approaches zero, g_Δ becomes a better and better approximation to the delta function.§

It is obvious from the Gaussian model that the delta function is even. This may be verified as follows:

$$\delta(x - x') = \langle x | x' \rangle = \langle x' | x \rangle^* = \delta(x' - x)^* = \delta(x' - x)$$

since the delta function is real.

Consider next an object that is even more peculiar than the delta function: its derivative with respect to the *first* argument x :

$$\delta'(x - x') = \frac{d}{dx} \delta(x - x') = -\frac{d}{dx'} \delta(x - x') \quad (1.10.20)$$

What is the action of this function under the integral? The clue comes from the Gaussian model. Consider $dg_\Delta(x - x')/dx = -dg_\Delta(x - x')/dx'$ as a function of x' . As g_Δ shrinks, each bump at $\pm \varepsilon$ will become, up to a scale factor, the δ function. The

‡ We will often omit the limits of integration if they are unimportant.

§ A fine point that will not concern you till Chapter 8: This formula for the delta function is valid even if Δ^2 is pure imaginary, say, equal to $i\beta^2$. First we see from Eq. (A.2.5) that g has unit area. Consider next the integral of g times $f(x')$ over a region in x' that includes x . For the most part, we get zero because f is smooth and g is wildly oscillating as $\beta \rightarrow 0$. However, at $x = x'$, the derivative of the phase of g vanishes and the oscillations are suspended. Pulling $f(x' = x)$ out of the integral, we get the desired result.

first one will sample $-f(x - \varepsilon)$ and the second one $+f(x + \varepsilon)$, again up to a scale, so that

$$\int \delta'(x - x') f(x') dx' \propto f(x + \varepsilon) - f(x - \varepsilon) = 2\varepsilon \frac{df}{dx'} \Big|_{x'=x}$$

The constant of proportionality happens to be $1/2\varepsilon$ so that

$$\int \delta'(x - x') f(x') dx' = \frac{df}{dx'} \Big|_{x'=x} = \frac{df(x)}{dx} \quad (1.10.21)$$

This result may be verified as follows:

$$\begin{aligned} \int \delta'(x - x') f(x') dx' &= \int \frac{d\delta(x - x')}{dx} f(x') dx = \frac{d}{dx} \int \delta(x - x') f(x') dx' \\ &= \frac{df(x)}{dx} \end{aligned}$$

Note that $\delta'(x - x')$ is an odd function. This should be clear from Fig. 1.8b or Eq. (1.10.20). An equivalent way to describe the action of the δ' function is by the equation

$$\delta'(x - x') = \delta(x - x') \frac{d}{dx'} \quad (1.10.22)$$

where it is understood that both sides appear in an integral over x' and that the differential operator acts on any function that accompanies the δ' function in the integrand. In this notation we can describe the action of higher derivatives of the delta function:

$$\frac{d^n \delta(x - x')}{dx^n} = \delta(x - x') \frac{d^n}{dx'^n} \quad (1.10.23)$$

We will now develop an alternate representation of the delta function. We know from basic Fourier analysis that, given a function $f(x)$, we may define its transform

$$f(k) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} e^{-ikx} f(x) dx \quad (1.10.24)$$

and its inverse

$$f(x') = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} e^{ikx'} f(k) dk \quad (1.10.25)$$

Feeding Eq. (1.10.24) into Eq. (1.10.25), we get

$$f(x') = \int_{-\infty}^{\infty} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} dk e^{ik(x' - x)} \right) f(x) dx$$

Comparing this result with Eq. (1.10.18), we see that

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} dk e^{ik(x' - x)} = \delta(x' - x) \quad (1.10.26)$$

*Exercise 1.10.1.** Show that $\delta(ax) = \delta(x)/|a|$. [Consider $\int \delta(ax) d(ax)$. Remember that $\delta(x) = \delta(-x)$.]

*Exercise 1.10.2.** Show that

$$\delta(f(x)) = \sum_i \frac{\delta(x_i - x)}{|df/dx_i|}$$

where x_i are the zeros of $f(x)$. Hint: Where does $\delta(f(x))$ blow up? Expand $f(x)$ near such points in a Taylor series, keeping the first nonzero term.

*Exercise 1.10.3.** Consider the *theta function* $\theta(x - x')$ which vanishes if $x - x'$ is negative and equals 1 if $x - x'$ is positive. Show that $\delta(x - x') = d/dx \theta(x - x')$.

Operators in Infinite Dimensions

Having acquainted ourselves with the elements of this function space, namely, the kets $|f\rangle$ and the basis vectors $|x\rangle$, let us turn to the (linear) operators that act on them. Consider the equation

$$\Omega|f\rangle = |\tilde{f}\rangle$$

Since the kets are in correspondence with the functions, Ω takes the function $f(x)$ into another, $\tilde{f}(x)$. Now, one operator that does such a thing is the familiar differential operator, which, acting on $f(x)$, gives $\tilde{f}(x) = df(x)/dx$. In the function space we can describe the action of this operator as

$$D|f\rangle = |df/dx\rangle$$

where $|df/dx\rangle$ is the ket corresponding to the function df/dx . What are the matrix elements of D in the $|x\rangle$ basis? To find out, we dot both sides of the above equation

with $\langle x|$,

$$\langle x| D| f \rangle = \left\langle x \left| \frac{df}{dx} \right. \right\rangle = \frac{df(x)}{dx}$$

and insert the resolution of identity at the right place

$$\int \langle x| D| x' \rangle \langle x'| f \rangle dx' = \frac{df}{dx} \quad (1.10.27)$$

Comparing this to Eq. (1.10.21), we deduce that

$$\langle x| D| x' \rangle = D_{xx'} = \delta'(x - x') = \delta(x - x') \frac{d}{dx'} \quad (1.10.28)$$

It is worth remembering that $D_{xx'} = \delta'(x - x')$ is to be integrated over the second index (x') and pulls out the derivative of f at the first index (x). Some people prefer to integrate $\delta'(x - x')$ over the first index, in which case it pulls out $-df/dx'$. Our convention is more natural if one views $D_{xx'}$ as a matrix acting to the right on the components $f_{x'} \equiv f(x')$ of a vector $|f\rangle$. Thus the familiar differential operator is an infinite-dimensional matrix with the elements given above. Normally one doesn't think of D as a matrix for the following reason. Usually when a matrix acts on a vector, there is a sum over a common index. In fact, Eq. (1.10.27) contains such a sum over the index x' . If, however, we feed into this equation the value of $D_{xx'}$, the delta function renders the integration trivial:

$$\int \delta(x - x') \frac{d}{dx'} f(x') dx' = \frac{df}{dx'} \Big|_{x'=x} = \frac{df}{dx}$$

Thus the action of D is simply to apply d/dx to $f(x)$ with no sum over a common index in sight. Although we too will drop the integral over the common index ultimately, we will continue to use it for a while to remind us that D , like all linear operators, is a matrix.

Let us now ask if D is Hermitian and examine its eigenvalue problem. If D were Hermitian, we would have

$$D_{xx'} = D_{x'x}^*$$

But this is not the case:

$$D_{xx'} = \delta'(x - x')$$

while

$$D_{x'x}^* = \delta'(x' - x)^* = \delta'(x' - x) = -\delta'(x - x')$$

But we can easily convert D to a Hermitian matrix by multiplying it with a pure imaginary number. Consider

$$K = -iD$$

which satisfies

$$K_{x'x}^* = [-i\delta'(x' - x)]^* = +i\delta'(x' - x) = -i\delta'(x - x') = K_{xx'}$$

It turns out that despite the above, the operator K is not guaranteed to be Hermitian, as the following analysis will indicate. Let $|f\rangle$ and $|g\rangle$ be two kets in the function space, whose images in the X basis are two functions $f(x)$ and $g(x)$ in the interval $a - b$. If K is Hermitian, it must also satisfy

$$\langle g|K|f\rangle = \langle g|Kf\rangle = \langle Kf|g\rangle^* = \langle f|K^\dagger|g\rangle^* = \langle f|K|g\rangle^*$$

So we ask

$$\begin{aligned} & \int_a^b \int_a^b \langle g|x\rangle \langle x|K|x'\rangle \langle x'|f\rangle dx dx' \\ & \stackrel{?}{=} \left(\int_a^b \int_a^b \langle f|x\rangle \langle x|K|x'\rangle \langle x'|g\rangle dx dx' \right)^* \\ & \int_a^b g^*(x) \left[-\frac{i df(x)}{dx} \right] dx \stackrel{?}{=} \left\{ \int_a^b f^*(x) \left[-\frac{i dg(x)}{dx} \right] dx \right\}^* = i \int_a^b \frac{dg^*}{dx} f(x) dx \end{aligned}$$

Integrating the left-hand side by parts gives

$$-ig^*(x)f(x) \Big|_a^b + i \int_a^b \frac{dg^*(x)}{dx} f(x) dx$$

So K is Hermitian only if the *surface term* vanishes:

$$-ig^*(x)f(x) \Big|_a^b = 0 \quad (1.10.29)$$

In contrast to the finite-dimensional case, $K_{xx'} = K_{x'x}^*$ is not a sufficient condition for K to be Hermitian. One also needs to look at the behavior of the functions at the end points a and b . Thus K is Hermitian if the space consists of functions that obey Eq. (1.10.29). One set of functions that obey this condition are the possible configurations $f(x)$ of the string clamped at $x=0, L$, since $f(x)$ vanishes at the end points. But condition (1.10.29) can also be fulfilled in another way. Consider functions in our own three-dimensional space, parametrized by r, θ , and ϕ (ϕ is the angle measured around the z axis). Let us require that these functions be single

valued. In particular, if we start at a certain point and go once around the z axis, returning to the original point, the function must take on its original value, i.e.,

$$f(\phi) = f(\phi + 2\pi)$$

In the space of such periodic functions, $K = -i d/d\phi$ is a Hermitian operator. The surface term vanishes because the contribution from one extremity cancels that from the other:

$$-ig^*(\phi)f(\phi) \Big|_0^{2\pi} = -i[g^*(2\pi)f(2\pi) - g^*(0)f(0)] = 0$$

In the study of quantum mechanics, we will be interested in functions defined over the full interval $-\infty \leq x \leq +\infty$. They fall into two classes, those that vanish as $|x| \rightarrow \infty$, and those that do not, the latter behaving as e^{ikx} , k being a real parameter that labels these functions. It is clear that $K = -i d/dx$ is Hermitian when sandwiched between two functions of the first class or a function from each, since in either case the surface term vanishes. When sandwiched between two functions of the second class, the Hermiticity hinges on whether

$$e^{ikx} e^{-ik'x} \Big|_{-\infty}^{\infty} \stackrel{?}{=} 0$$

If $k = k'$, the contribution from one end cancels that from the other. If $k \neq k'$, the answer is unclear since $e^{i(k-k')x}$ oscillates, rather than approaching a limit as $|x| \rightarrow \infty$. Now, there exists a way of defining a limit for such functions that cannot make up their minds: the limit as $|x| \rightarrow \infty$ is defined to be the average over a large interval. According to this prescription, we have, say as $x \rightarrow \infty$,

$$\lim_{x \rightarrow \infty} e^{ikx} e^{-ik'x} = \lim_{\substack{L \rightarrow \infty \\ \Delta \rightarrow 0}} \frac{1}{\Delta} \int_0^{L+\Delta} e^{i(k-k')x} dx = 0 \quad \text{if } k \neq k'$$

and so K is Hermitian in this space.

We now turn to the eigenvalue problem of K . The task seems very formidable indeed, for we have now to find the roots of an infinite-order characteristic polynomial and get the corresponding eigenvectors. It turns out to be quite simple and you might have done it a few times in the past without giving yourself due credit. Let us begin with

$$K|k\rangle = k|k\rangle \tag{1.10.30}$$

Following the standard procedure,

67

MATHEMATICAL
INTRODUCTION

$$\begin{aligned} \langle x|K|k\rangle &= k\langle x|k\rangle \\ \int \langle x|K|x'\rangle \langle x'|k\rangle dx' &= k\psi_k(x) \\ -i \frac{d}{dx} \psi_k(x) &= k\psi_k(x) \end{aligned} \quad (1.10.31)$$

where by definition $\psi_k(x) = \langle x|k\rangle$. This equation could have been written directly had we made the immediate substitution $K = -i d/dx$ in the X basis. From now on we shall resort to this shortcut unless there are good reasons for not doing so.

The solution to the above equation is simply

$$\psi_k(x) = A e^{ikx} \quad (1.10.32)$$

where A , the overall scale, is a free parameter, unspecified by the eigenvalue problem. So the eigenvalue problem of K is fully solved: any real number k is an eigenvalue, and the corresponding eigenfunction is given by $A e^{ikx}$. As usual, the freedom in scale will be used to normalize the solution. We choose A to be $(1/2\pi)^{-1/2}$ so that

$$|k\rangle \leftrightarrow \frac{1}{(2\pi)^{1/2}} e^{ikx}$$

and

$$\langle k|k'\rangle = \int_{-\infty}^{\infty} \langle k|x\rangle \langle x|k'\rangle dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i(k-k')x} dx = \delta(k-k') \quad (1.10.33)$$

(Since $\langle k|k\rangle$ is infinite, no choice of A can normalize $|k\rangle$ to unity. The delta function normalization is the natural one when the eigenvalue spectrum is continuous.)

The attentive reader may have a question at this point.

“Why was it assumed that the eigenvalue k was real? It is clear that the function $A e^{ikx}$ with $k = k_1 + ik_2$ also satisfies Eq. (1.10.31).”

The answer is, yes, there are eigenfunctions of K with complex eigenvalues. If, however, our space includes such functions, K must be classified a non-Hermitian operator. (The surface term no longer vanishes since e^{ikx} blows up exponentially as x tends to either $+\infty$ or $-\infty$, depending on the sign of the imaginary part k_2 .) In restricting ourselves to real k we have restricted ourselves to what we will call the *physical Hilbert space*, which is of interest in quantum mechanics. This space is defined as the space of functions that can be either normalized to unity or to the Dirac delta function and plays a central role in quantum mechanics. (We use the qualifier “physical” to distinguish it from the Hilbert space as defined by mathematicians, which contains only *proper vectors*, i.e., vectors normalizable to unity. The role of the *improper vectors* in quantum theory will be clear later.)

We will assume that the theorem proved for finite dimensions, namely, that the eigenfunctions of a Hermitian operator form a complete basis, holds in the Hilbert‡ space. (The trouble with infinite-dimensional spaces is that even if you have an infinite number of orthonormal eigenvectors, you can never be sure you have them all, since adding or subtracting a few still leaves you with an infinite number of them.)

Since K is a Hermitian operator, functions that were expanded in the X basis with components $f(x) = \langle x | f \rangle$ must also have an expansion in the K basis. To find the components, we start with a ket $|f\rangle$, and do the following:

$$f(k) = \langle k | f \rangle = \int_{-\infty}^{\infty} \langle k | x \rangle \langle x | f \rangle dx = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} e^{-ikx} f(x) dx \quad (1.10.34)$$

The passage back to the X basis is done as follows:

$$f(x) = \langle x | f \rangle = \int_{-\infty}^{\infty} \langle k | x \rangle \langle k | f \rangle dk = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} e^{ikx} f(k) dk \quad (1.10.35)$$

Thus the familiar Fourier transform is just the passage from one complete basis $|x\rangle$ to another, $|k\rangle$. Either basis may be used to expand functions that belong to the Hilbert space.

The matrix elements of K are trivial in the K basis:

$$\langle k | K | k' \rangle = k' \langle k | k' \rangle = k' \delta(k - k') \quad (1.10.36)$$

Now, we know where the K basis came from: it was generated by the Hermitian operator K . Which operator is responsible for the orthonormal X basis? Let us call it the operator X . The kets $|x\rangle$ are its eigenvectors with eigenvalue x :

$$X|x\rangle = x|x\rangle \quad (1.10.37)$$

Its matrix elements in the X basis are

$$\langle x' | X | x \rangle = x \delta(x' - x) \quad (1.10.38)$$

To find its action on functions, let us begin with

$$X|f\rangle = |\tilde{f}\rangle$$

and follow the routine:

$$\begin{aligned} \langle x | X | f \rangle &= \int \langle x | X | x' \rangle \langle x' | f \rangle dx' = xf(x) = \langle x | \tilde{f} \rangle = \tilde{f}(x) \\ \therefore \tilde{f}(x) &= xf(x) \end{aligned}$$

‡ Hereafter we will omit the qualifier “physical.”

Thus the effect of X is to multiply $f(x)$ by x . As in the case of the K operator, one generally suppresses the integral over the common index since it is rendered trivial by the delta function. We can summarize the action of X in Hilbert space as

$$X|f(x)\rangle = |xf(x)\rangle \quad (1.10.39)$$

where as usual $|xf(x)\rangle$ is the ket corresponding to the function $xf(x)$.

There is a nice reciprocity between the X and K operators which manifests itself if we compute the matrix elements of X in the K basis:

$$\begin{aligned} \langle k|X|k'\rangle &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikx} x e^{ik'x} dx \\ &= +i \frac{d}{dk} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(k'-k)x} dx \right) = i\delta'(k-k')^\ddagger \end{aligned}$$

Thus if $|g(k)\rangle$ is a ket whose image in the k basis is $g(k)$, then

$$X|g(k)\rangle = \left| \frac{i dg(k)}{dk} \right\rangle \quad (1.10.40)$$

In summary then, in the X basis, X acts as x and K as $-i d/dx$ [on the functions $f(x)$], while in the K basis, K acts like k and X like $i d/dk$ [on $f(k)$]. Operators with such an interrelationship are said to be *conjugate* to each other.

The conjugate operators X and K do not commute. Their commutator may be calculated as follows. Let us operate X and K in both possible orders on some ket $|f\rangle$ and follow the action in the X basis:

$$\begin{aligned} X|f\rangle &\rightarrow xf(x) \\ K|f\rangle &\rightarrow -i \frac{df(x)}{dx} \end{aligned}$$

So

$$\begin{aligned} XK|f\rangle &\rightarrow -ix \frac{df(x)}{dx} \\ KX|f\rangle &\rightarrow -i \frac{d}{dx} xf(x) \end{aligned}$$

Therefore

$$[X, K]|f\rangle \rightarrow -ix \frac{df}{dx} + ix \frac{df}{dx} + if = if \rightarrow i|f\rangle$$

[†] In the last step we have used the fact that $\delta(k'-k) = \delta(k-k')$.

Since $|f\rangle$ is an arbitrary ket, we now have the desired result:

$$[X, K] = iI \quad (1.10.41)$$

This brings us to the end of our discussion on Hilbert space, except for a final example. Although there are many other operators one can study in this space, we restricted ourselves to X and K since almost all the operators we will need for quantum mechanics are functions of X and $P = \hbar K$, where \hbar is a constant to be defined later.

Example 1.10.1: A Normal Mode Problem in Hilbert Space. Consider a string of length L clamped at its two ends $x=0$ and L . The displacement $\psi(x, t)$ obeys the differential equation

$$\frac{\partial^2 \psi}{\partial t^2} = \frac{\partial^2 \psi}{\partial x^2} \quad (1.10.42)$$

Given that at $t=0$ the displacement is $\psi(x, 0)$ and the velocity $\dot{\psi}(x, 0)=0$, we wish to determine the time evolution of the string.

But for the change in dimensionality, the problem is identical to that of the two coupled masses encountered at the end of Section 1.8 [see Eq. (1.8.26)]. It is recommended that you go over that example once to refresh your memory before proceeding further.

We first identify $\psi(x, t)$ as components of a vector $|\psi(t)\rangle$ in a Hilbert space, the elements of which are in correspondence with possible displacements ψ , i.e., functions that are continuous in the interval $0 \leq x \leq L$ and vanish at the end points. You may verify that these functions do form a vector space.

The analog of the operator Ω in Eq. (1.8.26) is the operator $\partial^2/\partial x^2$. We recognize this to be minus the square of the operator $K \leftrightarrow -i\partial/\partial x$. Since K acts on a space in which $\psi(0)=\psi(L)=0$, it is Hermitian, and so is K^2 . Equation (1.10.42) has the abstract counterpart

$$|\dot{\psi}(t)\rangle = -K^2 |\psi(t)\rangle \quad (1.10.43)$$

We solve the initial-value problem by following the algorithm developed in Example 1.8.6:

Step (1). Solve the eigenvalue problem of $-K^2$.

Step (2). Construct the propagator $U(t)$ in terms of the eigenvectors and eigenvalues.

Step (3).

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle \quad (1.10.44)$$

The equation to solve is

$$K^2|\psi\rangle = k^2|\psi\rangle \quad (1.10.45)$$

In the X basis, this becomes

$$-\frac{d^2}{dx^2} \psi_k(x) = k^2 \psi_k(x) \quad (1.10.46)$$

the general solution to which is

$$\psi_k(x) = A \cos kx + B \sin kx \quad (1.10.47)$$

where A and B are arbitrary. However, not all these solutions lie in the Hilbert space we are considering. We want only those that vanish at $x=0$ and $x=L$. At $x=0$ we find

$$\psi_k(0) = 0 = A \quad (1.10.48a)$$

while at $x=L$ we find

$$0 = B \sin kL \quad (1.10.48b)$$

If we do not want a trivial solution ($A=B=0$) we must demand

$$\sin kL = 0, \quad kL = m\pi, \quad m = 1, 2, 3, \dots \quad (1.10.49)$$

We do not consider negative m since it doesn't lead to any further LI solutions [$\sin(-x) = -\sin x$]. The allowed eigenvectors thus form a discrete set labeled by an integer m :

$$\psi_m(x) = \left(\frac{2}{L}\right)^{1/2} \sin\left(\frac{m\pi x}{L}\right) \quad (1.10.50)$$

where we have chosen $B=(2/L)^{1/2}$ so that

$$\int_0^L \psi_m(x) \psi_{m'}(x) dx = \delta_{mm'} \quad (1.10.51)$$

Let us associate with each solution labeled by the integer m an abstract ket $|m\rangle$:

$$|m\rangle \xrightarrow[X \text{ basis}]{} (2/L)^{1/2} \sin\left(\frac{m\pi x}{L}\right) \quad (1.10.52)$$

If we project $|\psi(t)\rangle$ on the $|m\rangle$ basis, in which K is diagonal with eigenvalues $(m\pi/L)^2$, the components $\langle m| \psi(t)\rangle$ will obey the decoupled equations

$$\frac{d^2}{dt^2} \langle m| \psi(t)\rangle = -\left(\frac{m^2\pi^2}{L^2}\right) \langle m| \psi(t)\rangle, \quad m=1, 2, \dots \quad (1.10.53)$$

in analogy with Eq. (1.8.33). These equations may be readily solved (subject to the condition of vanishing initial velocities) as

$$\langle m| \psi(t)\rangle = \langle m| \psi(0)\rangle \cos\left(\frac{m\pi t}{L}\right) \quad (1.10.54)$$

Consequently

$$\begin{aligned} |\psi(t)\rangle &= \sum_{m=1}^{\infty} |m\rangle \langle m| \psi(t)\rangle \\ &= \sum_{m=1}^{\infty} |m\rangle \langle m| \psi(0)\rangle \cos \omega_m t, \quad \omega_m = \frac{m\pi}{L} \end{aligned} \quad (1.10.55)$$

or

$$U(t) = \sum_{m=1}^{\infty} |m\rangle \langle m| \cos \omega_m t, \quad \omega_m = \frac{m\pi}{L} \quad (1.10.56)$$

The propagator equation

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle$$

becomes in the $|x\rangle$ basis

$$\begin{aligned} \langle x| \psi(t)\rangle &= \psi(x, t) \\ &= \langle x| U(t)|\psi(0)\rangle \\ &= \int_0^L \langle x| U(t)|x'\rangle \langle x'| \psi(0)\rangle dx' \end{aligned} \quad (1.10.57)$$

It follows from Eq. (1.10.56) that

$$\begin{aligned} \langle x| U(t)|x'\rangle &= \sum_m \langle x| m\rangle \langle m| x'\rangle \cos \omega_m t \\ &= \sum_m \left(\frac{2}{L}\right) \sin\left(\frac{m\pi x}{L}\right) \sin\left(\frac{m\pi x'}{L}\right) \cos \omega_m t \end{aligned} \quad (1.10.58)$$

Thus, given any $\psi(x', 0)$, we can get $\psi(x, t)$ by performing the integral in Eq. (1.10.57), using $\langle x | U(t) | x' \rangle$ from Eq. (1.10.58). If the propagator language seems too abstract, we can begin with Eq. (1.10.55). Dotting both sides with $\langle x |$, we get

$$\begin{aligned}\psi(x, t) &= \sum_{m=1}^{\infty} \langle x | m \rangle \langle m | \psi(0) \rangle \cos \omega_m t \\ &= \sum_{m=1}^{\infty} \left(\frac{2}{L}\right)^{1/2} \sin\left(\frac{m\pi x}{L}\right) \cos \omega_m t \langle m | \psi(0) \rangle\end{aligned}\quad (1.10.59)$$

Given $|\psi(0)\rangle$, one must then compute

$$\langle m | \psi(0) \rangle = \left(\frac{2}{L}\right)^{1/2} \int_0^L \sin\left(\frac{m\pi x}{L}\right) \psi(x, 0) dx$$

Usually we will find that the coefficients $\langle m | \psi(0) \rangle$ fall rapidly with m so that a few leading terms may suffice to get a good approximation. \square

Exercise 1.10.4. A string is displaced as follows at $t=0$:

$$\begin{aligned}\psi(x, 0) &= \frac{2xh}{L}, \quad 0 \leq x \leq \frac{L}{2} \\ &= \frac{2h}{L}(L-x), \quad \frac{L}{2} \leq x \leq L\end{aligned}$$

Show that

$$\psi(x, t) = \sum_{m=1}^{\infty} \sin\left(\frac{m\pi x}{L}\right) \cos \omega_m t \cdot \left(\frac{8h}{\pi^2 m^2}\right) \sin\left(\frac{\pi m}{2}\right)$$

2

Review of Classical Mechanics

In this chapter we will develop the Lagrangian and Hamiltonian formulations of mechanics starting from Newton's laws. These subsequent reformulations of mechanics bring with them a great deal of elegance and computational ease. But our principal interest in them stems from the fact that they are the ideal springboards from which to make the leap to quantum mechanics. The passage from the Lagrangian formulation to quantum mechanics was carried out by Feynman in his path integral formalism. A more common route to quantum mechanics, which we will follow for the most part, has as its starting point the Hamiltonian formulation, and it was discovered mainly by Schrödinger, Heisenberg, Dirac, and Born.

It should be emphasized, and it will soon become apparent, that all three formulations of mechanics are essentially the same theory, in that their domains of validity and predictions are identical. Nonetheless, in a given context, one or the other may be more inviting for conceptual, computational, or simply aesthetic reasons.

2.1. The Principle of Least Action and Lagrangian Mechanics

Let us take as our prototype of the Newtonian scheme a point particle of mass m moving along the x axis under a potential $V(x)$. According to Newton's Second Law,

$$m \frac{d^2x}{dt^2} = -\frac{dV}{dx} \quad (2.1.1)$$

If we are given the initial state variables, the position $x(t_i)$ and velocity $\dot{x}(t_i)$, we can calculate the classical trajectory $x_{cl}(t)$ as follows. Using the initial velocity and acceleration [obtained from Eq. (2.1.1)] we compute the position and velocity at a time $t_i + \Delta t$. For example,

$$x_{cl}(t_i + \Delta t) = x(t_i) + \dot{x}(t_i)\Delta t$$

Having updated the state variables to the time $t_i + \Delta t$, we can repeat the process again to inch forward to $t_i + 2\Delta t$ and so on.

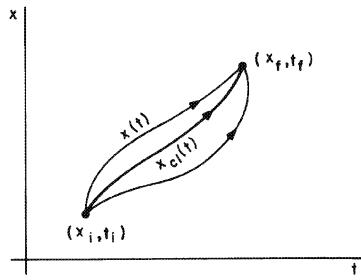


Figure 2.1. The Lagrangian formalism asks what distinguishes the actual path $x_{\text{cl}}(t)$ taken by the particle from all possible paths connecting the end points (x_i, t_i) and (x_f, t_f) .

The equation of motion being second order in time, two pieces of data, $x(t_i)$ and $\dot{x}(t_i)$, are needed to specify a unique $x_{\text{cl}}(t)$. An equivalent way to do the same, and one that we will have occasion to employ, is to specify two space-time points (x_i, t_i) and (x_f, t_f) on the trajectory.

The above scheme readily generalizes to more than one particle and more than one dimension. If we use n Cartesian coordinates (x_1, x_2, \dots, x_n) to specify the positions of the particles, the spatial configuration of the system may be visualized as a point in an n -dimensional *configuration space*. (The term “configuration space” is used even if the n coordinates are not Cartesian.) The motion of the representative point is given by

$$m_j \frac{d^2 x_j}{dt^2} = -\frac{\partial V}{\partial x_j} \quad (2.1.2)$$

where m_j stands for the mass of the particle whose coordinate is x_j . These equations can be integrated step by step, just as before, to determine the trajectory.

In the Lagrangian formalism, the problem of a single particle in a potential $V(x)$ is posed in a different way: given that the particle is at x_i and x_f at times t_i and t_f , respectively, what is it that distinguishes the actual trajectory $x_{\text{cl}}(t)$ from all other trajectories or paths that connect these points? (See Fig. 2.1.)

The Lagrangian approach is thus global, in that it tries to determine at one stroke the entire trajectory $x_{\text{cl}}(t)$, in contrast to the local approach of the Newtonian scheme, which concerns itself with what the particle is going to do in the next infinitesimal time interval.

The answer to the question posed above comes in three parts:

(1) Define a function \mathcal{L} , called the *Lagrangian*, given by $\mathcal{L} = T - V$, T and V being the kinetic and potential energies of the particle. Thus $\mathcal{L} = \mathcal{L}(x, \dot{x}, t)$. The explicit t dependence may arise if the particle is in an external time-dependent field. We will, however, assume the absence of this t dependence.

(2) For each path $x(t)$ connecting (x_i, t_i) and (x_f, t_f) , calculate the *action* $S[x(t)]$ defined by

$$S[x(t)] = \int_{t_i}^{t_f} \mathcal{L}(x, \dot{x}) dt \quad (2.1.3)$$

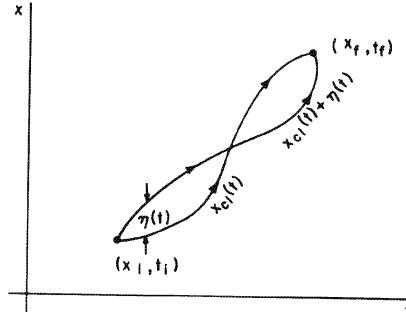


Figure 2.2. If $x_{cl}(t)$ minimizes S , then $\delta S^{(1)}=0$ if we go to any nearby path $x_{cl}(t)+\eta(t)$.

We use square brackets to enclose the argument of S to remind us that the function S depends on an entire path or function $x(t)$, and not just the value of x at some time t . One calls S a *functional* to signify that it is a function of a function.

(3) The classical path is one on which S is a minimum. (Actually we will only require that it be an extremum. It is, however, customary to refer to this condition as the *principle of least action*.)

We will now verify that this principle reproduces Newton's Second Law.

The first step is to realize that a functional $S[x(t)]$ is just a function of n variables as $n \rightarrow \infty$. In other words, the function $x(t)$ simply specifies an infinite number of values $x(t_1), \dots, x(t), \dots, x(t_f)$, one for each instant in time t in the interval $t_i \leq t \leq t_f$, and S is a function of these variables. To find its minimum we simply generalize the procedure for the finite n case. Let us recall that if $f=f(x_1, \dots, x_n)=f(\mathbf{x})$; the minimum \mathbf{x}^0 is characterized by the fact that if we move away from it by a small amount $\mathbf{\eta}$ in any direction, the first-order change $\delta f^{(1)}$ in f vanishes. That is, if we make a Taylor expansion:

$$f(\mathbf{x}^0 + \mathbf{\eta}) = f(\mathbf{x}^0) + \sum_{i=1}^n \left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}^0} \eta_i + \text{higher-order terms in } \eta \quad (2.1.4)$$

then

$$\delta f^{(1)} \equiv \sum_{i=1}^n \left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}^0} \eta_i = 0 \quad (2.1.5)$$

From this condition we can deduce an equivalent and perhaps more familiar expression of the minimum condition: every first-order partial derivative vanishes at \mathbf{x}^0 . To prove this, for say, $\partial f / \partial x_i$, we simply choose $\mathbf{\eta}$ to be along the i th direction. Thus

$$\left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}^0} = 0, \quad i = 1, \dots, n \quad (2.1.6)$$

Let us now mimic this procedure for the action S . Let $x_{cl}(t)$ be the path of least action and $x_{cl}(t) + \eta(t)$ a "nearby" path (see Fig. 2.2). The requirement that all paths coincide at t_i and t_f means

$$\eta(t_i) = \eta(t_f) = 0 \quad (2.1.7)$$

Now

$$\begin{aligned}
 S[x_{\text{cl}}(t) + \eta(t)] &= \int_{t_i}^{t_f} \mathcal{L}(x_{\text{cl}}(t) + \eta(t); \dot{x}_{\text{cl}}(t) + \dot{\eta}(t)) dt \\
 &= \int_{t_i}^{t_f} \left[\mathcal{L}(x_{\text{cl}}(t), \dot{x}_{\text{cl}}(t)) + \frac{\partial \mathcal{L}}{\partial x(t)} \Big|_{x_{\text{cl}}} \cdot \eta(t) \right. \\
 &\quad \left. + \frac{\partial \mathcal{L}}{\partial \dot{x}(t)} \Big|_{x_{\text{cl}}} \cdot \dot{\eta}(t) + \dots \right] dt \\
 &= S[x_{\text{cl}}(t)] + \delta S^{(1)} + \text{higher-order terms}
 \end{aligned}$$

We set $\delta S^{(1)} = 0$ in analogy with the finite variable case:

$$0 = \delta S^{(1)} = \int_{t_i}^{t_f} \left[\frac{\partial \mathcal{L}}{\partial x(t)} \Big|_{x_{\text{cl}}} \cdot \eta(t) + \frac{\partial \mathcal{L}}{\partial \dot{x}(t)} \Big|_{x_{\text{cl}}} \cdot \dot{\eta}(t) \right] dt$$

If we integrate the second term by parts, it turns into

$$\frac{\partial \mathcal{L}}{\partial \dot{x}(t)} \Big|_{x_{\text{cl}}} \cdot \eta(t) \Big|_{t_i}^{t_f} - \int_{t_i}^{t_f} \left[\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{x}(t)} \right]_{x_{\text{cl}}} \cdot \eta(t) dt$$

The first of these terms vanishes due to Eq. (2.1.7). So that

$$0 = \delta S^{(1)} = \int_{t_i}^{t_f} \left[\frac{\partial \mathcal{L}}{\partial x(t)} - \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{x}(t)} \right]_{x_{\text{cl}}} \cdot \eta(t) dt \quad (2.1.8)$$

Note that the condition $\delta S^{(1)} = 0$ implies that S is extremized and not necessarily minimized. We shall, however, continue the tradition of referring to this extremum as the minimum. This equation is the analog of Eq. (2.1.5): the discrete variable η_i is replaced by $\eta(t)$; the sum over i is replaced by an integral over t , and $\partial f / \partial x_i$ is replaced by

$$\frac{\partial \mathcal{L}}{\partial x(t)} - \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{x}(t)}$$

There are two terms here playing the role of $\partial f / \partial x_i$, since \mathcal{L} (or equivalently S) has both explicit and implicit (through the \dot{x} terms) dependence on $x(t)$. Since $\eta(t)$ is arbitrary, we may extract the analog of Eq. (2.1.6):

$$\left\{ \frac{\partial \mathcal{L}}{\partial x(t)} - \frac{d}{dt} \left[\frac{\partial \mathcal{L}}{\partial \dot{x}(t)} \right] \right\}_{x_{\text{cl}}(t)} = 0 \quad \text{for } t_i \leq t \leq t_f \quad (2.1.9)$$

To deduce this result for some specific time t_0 , we simply choose an $\eta(t)$ that vanishes everywhere except in an infinitesimal region around t_0 .

Equation (2.1.9) is the celebrated *Euler-Lagrange equation*. If we feed into it $\mathcal{L} = T - V$, $T = \frac{1}{2}m\dot{x}^2$, $V = V(x)$, we get

$$\frac{\partial \mathcal{L}}{\partial \dot{x}} = \frac{\partial T}{\partial \dot{x}} = m\dot{x}$$

and

$$\frac{\partial \mathcal{L}}{\partial x} = -\frac{\partial V}{\partial x}$$

so that the Euler-Lagrange equation becomes just

$$\frac{d}{dt}(m\dot{x}) = -\frac{\partial V}{\partial x}$$

which is just Newton's Second Law, Eq. (2.1.1).

If we consider a system described by n Cartesian coordinates, the same procedure yields

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}_i}\right) = \frac{\partial \mathcal{L}}{\partial x_i} \quad (i=1, \dots, n) \quad (2.1.10)$$

Now

$$T = \frac{1}{2} \sum_{i=1}^n m_i (\dot{x}_i)^2$$

and

$$V = V(x_1, \dots, x_n)$$

so that Eq. (2.1.10) becomes

$$\frac{d}{dt}(m_i \dot{x}_i) = -\frac{\partial V}{\partial x_i}$$

which is identical to Eq. (2.1.2). Thus the minimum (action) principle indeed reproduces Newtonian mechanics if we choose $\mathcal{L} = T - V$.

Notice that we have assumed that V is velocity-independent in the above proof. An important force, that of a magnetic field \mathbf{B} on a moving charge is excluded by this restriction, since $\mathbf{F}_B = q\mathbf{v} \times \mathbf{B}$, q being the charge of the particle and $\mathbf{v} = \dot{\mathbf{r}}$ its velocity. We will show shortly that this force too may be accommodated in the Lagrangian formalism, in the sense that we can find an \mathcal{L} that yields the correct force law when Eq. (2.1.10) is employed. But this \mathcal{L} no longer has the form $T - V$. One therefore frees oneself from the notion that $\mathcal{L} = T - V$; and views \mathcal{L} as some

function $\mathcal{L}(x_i, \dot{x}_i)$ which yields the correct Newtonian dynamics when fed into the Euler–Lagrange equations. To the reader who wonders why one bothers to even deal with a Lagrangian when all it does is yield Newtonian force laws in the end, I present a few of its main attractions besides its closeness to quantum mechanics. These will then be illustrated by means of an example.

(1) In the Lagrangian scheme one has merely to construct a single *scalar* \mathcal{L} and all the equations of motion follow by simple differentiation. This must be contrasted with the Newtonian scheme, which deals with vectors and is thus more complicated.

(2) The Euler–Lagrange equations (2.1.10) have the same *form* if we use, instead of the n Cartesian coordinates x_1, \dots, x_n , *any* general set of n independent coordinates q_1, q_2, \dots, q_n . To remind us of this fact we will rewrite Eq. (2.1.10) as

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right) = \frac{\partial \mathcal{L}}{\partial q_i} \quad (2.1.11)$$

One can either verify this by brute force, making a change of variables in Eq. (2.1.10) and seeing that an identical equation with x_i replaced by q_i follows, or one can simply go through our derivation of the minimum action condition and see that nowhere were the coordinates assumed to be Cartesian. Of course, at the next stage, in showing that the Euler–Lagrange equations were equivalent to Newton’s, Cartesian coordinates *were* used, for in these coordinates the kinetic energy T and the Newtonian equations have simple forms. But once the principle of least action is seen to generate the correct dynamics, we can forget all about Newton’s laws and use Eq. (2.1.11) as the equations of motion. What is being emphasized is that these equations, which express the condition for least action, are form invariant under an arbitrary change of coordinates. This form invariance must be contrasted with the Newtonian equation (2.1.2), which presumes that the x_i are Cartesian. If one trades the x_i for another non-Cartesian set of q_i , Eq. (2.1.2) will have a different form (see Example 2.1.1 at the end of this section).

Equation (2.1.11) can be made to resemble Newton’s Second Law if one defines a quantity

$$p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \quad (2.1.12)$$

called the *canonical momentum conjugate to q_i* and the quantity

$$F_i = \frac{\partial \mathcal{L}}{\partial q_i} \quad (2.1.13)$$

called the *generalized force conjugate to q_i* . Although the rate of change of the canonical momentum equals the generalized force, one must remember that neither is p_i always a linear momentum (mass times velocity or “ mv ” momentum), nor is F_i always a force (with dimensions of mass times acceleration). For example, if q_i is an angle θ , p_i will be an angular momentum and F_i a torque.

(3) Conservation laws are easily obtained in this formalism. Suppose the Lagrangian depends on a certain velocity \dot{q}_i but not on the corresponding coordinate q_i . The latter is then called a *cyclic coordinate*. It follows that the corresponding p_i is conserved:

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right) = \frac{dp_i}{dt} = \frac{\partial \mathcal{L}}{\partial q_i} = 0 \quad (2.1.14)$$

Although Newton's Second Law, Eq. (2.1.2), also tells us that if a Cartesian coordinate x_i is cyclic, the corresponding momentum $m_i \dot{x}_i$ is conserved, Eq. (2.1.14) is more general. Consider, for example, a potential $V(x, y)$ in two dimensions that depends only upon $\rho = (x^2 + y^2)^{1/2}$, and not on the polar angle ϕ , so that $V(\rho, \phi) = V(\rho)$. It follows that ϕ is a cyclic coordinate, as T depends only on ϕ (see Example 2.1.1 below). Consequently $\partial \mathcal{L} / \partial \dot{\phi} = p_\phi$ is conserved. In contrast, no obvious conservation law arises from the Cartesian Eqs. (2.1.2) since neither x nor y is cyclic. If one rewrites Newton's laws in polar coordinates to exploit $\partial V / \partial \dot{\phi} = 0$, the corresponding equations get complicated due to centrifugal and Coriolis terms. It is the Lagrangian formalism that allows us to choose coordinates that best reflect the symmetry of the potential, without altering the simple form of the equations.

Example 2.1.1. We now illustrate the above points through an example. Consider a particle moving in a plane. The Lagrangian, in Cartesian coordinates, is

$$\begin{aligned} \mathcal{L} &= \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) - V(x, y) \\ &= \frac{1}{2}m\mathbf{v} \cdot \mathbf{v} - V(x, y) \end{aligned} \quad (2.1.15)$$

where \mathbf{v} is the velocity of the particle, with $\mathbf{v} = \dot{\mathbf{r}}$, \mathbf{r} being its position vector. The corresponding equations of motion are

$$m\ddot{x} = -\frac{\partial V}{\partial x} \quad (2.1.16)$$

$$m\ddot{y} = -\frac{\partial V}{\partial y} \quad (2.1.17)$$

which are identical to Newton's laws. If one wants to get the same Newton's laws in terms of polar coordinates ρ and ϕ , some careful vector analysis is needed to unearth the centrifugal and Coriolis terms:

$$m\ddot{\rho} = -\frac{\partial V}{\partial \rho} + m\rho(\dot{\phi})^2 \quad (2.1.18)$$

$$m\ddot{\phi} = -\frac{1}{\rho^2} \frac{\partial V}{\partial \phi} - \frac{2m\dot{\rho}\dot{\phi}}{\rho} \quad (2.1.19)$$

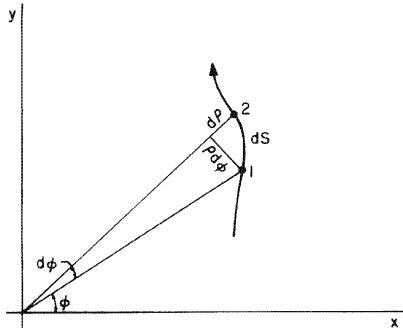


Figure 2.3. Points (1) and (2) are positions of the particle at times differing by Δt .

Notice the difference in form between Eqs. (2.1.16) and (2.1.17) on the one hand and Eqs. (2.1.18) and (2.1.19) on the other.

In the Lagrangian scheme one has only to recompute \mathcal{L} in polar coordinates. From Fig. 2.3 it is clear that the distance traveled by the particle in time Δt is

$$dS = [(d\rho)^2 + (\rho d\phi)^2]^{1/2}$$

so that the magnitude of velocity is

$$v = \frac{dS}{dt} = [(\dot{\rho})^2 + \rho^2(\dot{\phi})^2]^{1/2}$$

and

$$\mathcal{L} = \frac{1}{2}m(\dot{\rho}^2 + \rho^2\dot{\phi}^2) - V(\rho, \phi) \quad (2.1.20)$$

(Notice that in these coordinates T involves not just the velocities $\dot{\rho}$ and $\dot{\phi}$ but also the coordinate ρ . This does not happen in Cartesian coordinates.) The equations of motion generated by this \mathcal{L} are

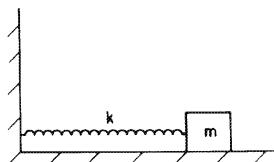
$$\frac{d}{dt}(m\dot{\rho}) = -\frac{\partial V}{\partial \rho} + m\rho\dot{\phi}^2 \quad (2.1.21)$$

$$\frac{d}{dt}(m\rho^2\dot{\phi}) = -\frac{\partial V}{\partial \phi} \quad (2.1.22)$$

which are the same as Eqs. (2.1.18) and (2.1.19). In Eq. (2.1.22) the canonical momentum $p_\phi = m\rho^2\dot{\phi}$ is the angular momentum and the generalized force $-\partial V/\partial \phi$ is the torque, both along the z axis. Notice how easily the centrifugal and Coriolis forces came out.

Finally, if $V(\rho, \phi) = V(\rho)$, the conservation of p_ϕ is obvious in Eq. (2.1.22). The conservation of p_ϕ follows from Eq. (2.1.19) only after some manipulations and is practically invisible in Eqs. (2.1.16) and (2.1.17). Both the conserved quantity and its conservation law arise naturally in the Lagrangian scheme. \square

*Exercise 2.1.1.** Consider the following system, called a *harmonic oscillator*. The block has a mass m and lies on a frictionless surface. The spring has a force constant k .



Write the Lagrangian and get the equations of motion.

*Exercise 2.1.2.** Do the same for the coupled-mass problem discussed at the end of Section 1.8. Compare the equations of motion with Eqs. (1.8.24) and (1.8.25).

*Exercise 2.1.3.** A particle of mass m moves in three dimensions under a potential $V(r, \theta, \phi) = V(r)$. Write its \mathcal{L} and find the equations of motion.

2.2. The Electromagnetic Lagrangian‡

Recall that the force on a charge q due to an electric field \mathbf{E} and magnetic field \mathbf{B} is given by

$$\mathbf{F} = q \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right) \quad (2.2.1)$$

where $\mathbf{v} = \dot{\mathbf{r}}$ is the velocity of the particle. Since the force is velocity-dependent, we must analyze the problem afresh, not relying on the preceding discussion, which was restricted to velocity-independent forces.

Now it turns out that if we use

$$\mathcal{L}_{e.m.} = \frac{1}{2} m \mathbf{v} \cdot \mathbf{v} - q\phi + \frac{q}{c} \mathbf{v} \cdot \mathbf{A} \quad (2.2.2)$$

we get the correct electromagnetic force laws. In Eq. (2.2.2) c is the velocity of light, while ϕ and \mathbf{A} are the scalar and vector potentials related to \mathbf{E} and \mathbf{B} via

$$\mathbf{E} = -\nabla\phi - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} \quad (2.2.3)$$

and

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (2.2.4)$$

‡ See Section 18.4 for a review of classical electromagnetism.

The Euler–Lagrange equations corresponding to $\mathcal{L}_{e.m}$ are

$$\frac{d}{dt} \left(m\dot{x}_i + \frac{q}{c} A_i \right) = -q \frac{\partial \phi}{\partial x_i} + \frac{q}{c} \frac{\partial(\mathbf{v} \cdot \mathbf{A})}{\partial x_i}, \quad i=1, 2, 3 \quad (2.2.5)$$

Combining the three equations above into a single vector equation we get

$$\frac{d}{dt} \left(m\mathbf{v} + \frac{q\mathbf{A}}{c} \right) = -q\nabla\phi + \frac{q}{c}\nabla(\mathbf{v} \cdot \mathbf{A}) \quad (2.2.6)$$

The canonical momentum is

$$\mathbf{p} = m\mathbf{v} + \frac{q\mathbf{A}}{c} \quad (2.2.7)$$

Rewriting Eq. (2.2.6), we get

$$\frac{d}{dt} (m\mathbf{v}) = -q\nabla\phi + \frac{q}{c} \left[-\frac{d\mathbf{A}}{dt} + \nabla(\mathbf{v} \cdot \mathbf{A}) \right] \quad (2.2.8)$$

Now, the total derivative $d\mathbf{A}/dt$ has two parts: an explicit time dependence $\partial\mathbf{A}/\partial t$, plus an implicit one $(\mathbf{v} \cdot \nabla)\mathbf{A}$ which represents the fact that a spatial variation in \mathbf{A} will appear as a temporal variation to the moving particle. Now Eq. (2.2.8) becomes

$$\frac{d}{dt} (m\mathbf{v}) = -q\nabla\phi - \frac{q}{c} \frac{\partial\mathbf{A}}{\partial t} + \frac{q}{c} [\nabla(\mathbf{v} \cdot \mathbf{A}) - (\mathbf{v} \cdot \nabla)\mathbf{A}] \quad (2.2.9)$$

which is identical to Eq. (2.2.1) by virtue of the identity

$$\mathbf{v} \times (\nabla \times \mathbf{A}) = \nabla(\mathbf{v} \cdot \mathbf{A}) - (\mathbf{v} \cdot \nabla)\mathbf{A}$$

Notice that $\mathcal{L}_{e.m}$ is not of the form $T - V$, for the quantity $U = q\phi - (q/c)\mathbf{v} \cdot \mathbf{A}$ (sometimes called the *generalized potential*) cannot be interpreted as the potential energy of the charged particle. First of all, the force due to a time-dependent electromagnetic field is not generally conservative and does not admit a path-independent work function to play the role of a potential. Even in the special cases when the force is conservative, only $q\phi$ can be interpreted as the electrical potential energy. The $[-q(\mathbf{v} \cdot \mathbf{A})/c]$ term is not a magnetic potential energy, since the magnetic force $\mathbf{F}_B = q(\mathbf{v} \times \mathbf{B})/c$ never does any work, being always perpendicular to the velocity. To accommodate forces such as the electro-magnetic, we must, therefore, redefine \mathcal{L} to be that function $\mathcal{L}(q, \dot{q}, t)$ which, when fed into the Euler–Lagrange equations, reproduces the correct dynamics. The rule $\mathcal{L} = T - V$ becomes just a useful mnemonic for the case of conservative forces.

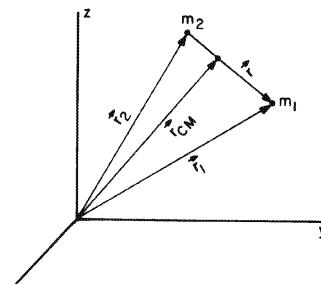


Figure 2.4. The relation between \mathbf{r}_1 , \mathbf{r}_2 and \mathbf{r}_{CM} , \mathbf{r} .

2.3. The Two-Body Problem

We discuss here a class of problems that plays a central role in classical physics: that of two masses m_1 and m_2 exerting equal and opposite forces on each other. Since the particles are responding to each other and nothing external, it follows that the potential between them depends only on the *relative coordinate* $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ and not the individual positions \mathbf{r}_1 and \mathbf{r}_2 . But $V(\mathbf{r}_1, \mathbf{r}_2) = V(\mathbf{r}_1 - \mathbf{r}_2)$ means in turn that there are three cyclic coordinates, for V depends on only three variables rather than the possible six. (In Cartesian coordinates, since T is a function only of velocities, a coordinate missing in V is also cyclic.) The corresponding conserved momenta will of course be the three components of the total momentum, which are conserved in the absence of external forces. To bring out these features, it is better to trade \mathbf{r}_1 and \mathbf{r}_2 in favor of

$$\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2 \quad (2.3.1)$$

and

$$\mathbf{r}_{CM} = \frac{m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2}{m_1 + m_2} \quad (2.3.2)$$

where \mathbf{r}_{CM} is called the *center-of-mass (CM) coordinate*. One can invert Eqs. (2.3.1) and (2.3.2) to get (see Fig. 2.4)

$$\mathbf{r}_1 = \mathbf{r}_{CM} + \frac{m_2 \mathbf{r}}{m_1 + m_2} \quad (2.3.3)$$

$$\mathbf{r}_2 = \mathbf{r}_{CM} - \frac{m_1 \mathbf{r}}{m_1 + m_2} \quad (2.3.4)$$

If one rewrites the Lagrangian

$$\mathcal{L} = \frac{1}{2} m_1 |\dot{\mathbf{r}}_1|^2 + \frac{1}{2} m_2 |\dot{\mathbf{r}}_2|^2 - V(\mathbf{r}_1 - \mathbf{r}_2) \quad (2.3.5)$$

in terms of \mathbf{r}_{CM} and \mathbf{r} , one gets

$$\mathcal{L} = \frac{1}{2} (m_1 + m_2) |\dot{\mathbf{r}}_{CM}|^2 + \frac{1}{2} \frac{m_1 m_2}{m_1 + m_2} |\dot{\mathbf{r}}|^2 - V(\mathbf{r}) \quad (2.3.6)$$

The main features of Eq. (2.3.6) are the following.

(1) The problem of two mutually interacting particles has been transformed to that of two fictitious particles that do not interact with each other. In other words, the equations of motion for \mathbf{r} do not involve \mathbf{r}_{CM} and vice versa, because $\mathcal{L}(\mathbf{r}, \dot{\mathbf{r}}; \mathbf{r}_{\text{CM}}, \dot{\mathbf{r}}_{\text{CM}}) = \mathcal{L}(\mathbf{r}, \dot{\mathbf{r}}) + \mathcal{L}(\mathbf{r}_{\text{CM}}, \dot{\mathbf{r}}_{\text{CM}})$.

(2) The first fictitious particle is the CM, of mass $M = m_1 + m_2$. Since \mathbf{r}_{CM} is a cyclic variable, the momentum $\mathbf{p}_{\text{CM}} = M\dot{\mathbf{r}}_{\text{CM}}$ (which is just the total momentum) is conserved as expected. Since the motion of the CM is uninteresting one usually ignores it. One clear way to do this is to go to the CM frame in which $\dot{\mathbf{r}}_{\text{CM}} = 0$, so that the CM is completely eliminated in the Lagrangian.

(3) The second fictitious particle has mass $\mu = m_1 m_2 / (m_1 + m_2)$ (called the *reduced mass*), momentum $\mathbf{p} = \mu\dot{\mathbf{r}}$ and moves under a potential $V(\mathbf{r})$. One has just to solve this one-body problem. If one chooses, one may easily return to the coordinates \mathbf{r}_1 and \mathbf{r}_2 at the end, using Eqs. (2.3.1) and (2.3.2).

*Exercise 2.3.1.** Derive Eq. (2.3.6) from (2.3.5) by changing variables.

2.4. How Smart Is a Particle?

The Lagrangian formalism seems to ascribe to a particle a tremendous amount of foresight: a particle at (x_i, t_i) destined for (x_f, t_f) manages to calculate ahead of time the action for every possible path linking these points, and takes the one with the least action. But this, of course, is an illusion. The particle need not know its entire trajectory ahead of time, it needs only to obey the Euler–Lagrange equations at each instant in time to minimize the action. This in turn means just following Newton’s law, which is to say, the particle has to sample the potential in its immediate vicinity and accelerate in the direction of greatest change.

Our esteem for the particle will sink further when we learn quantum mechanics. We will discover that far from following any kind of strategy, the particle, in a sense, goes from (x_i, t_i) to (x_f, t_f) along all possible paths, giving equal weight to each! How it is that despite this, classical particles do seem to follow $x_{\text{cl}}(t)$ is an interesting question that will be answered when we come to the path integral formalism of quantum mechanics.

2.5. The Hamiltonian Formalism

In the Lagrangian formalism, the independent variables are the coordinates q_i and velocities \dot{q}_i . The momenta are derived quantities defined by

$$p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \quad (2.5.1)$$

In the Hamiltonian formalism one exchanges the roles of \dot{q} and p : one replaces the Lagrangian $\mathcal{L}(q, \dot{q})$ [‡] by a Hamiltonian $\mathcal{H}(q, p)$ which generates the equations of motion, and \dot{q} becomes a derived quantity,

$$\dot{q}_i = \frac{\partial \mathcal{H}}{\partial p_i} \quad (2.5.2)$$

thereby completing the role reversal of the \dot{q} 's and the p 's.

There exists a standard procedure for effecting such a change, called a *Legendre transformation*, which is illustrated by the following simple example. Suppose we have a function $f(x)$ with

$$u(x) = \frac{df}{dx} \quad (2.5.3)$$

Let it be possible to invert $u(x)$ to get $x(u)$. [For example if $u(x) = x^3$, $x(u) = u^{1/3}$, etc.] If we define a function

$$g(u) = x(u)u - f(x(u)) \quad (2.5.4)$$

then

$$\frac{dg}{du} = \frac{dx}{du} \cdot u + x(u) - \frac{df}{dx} \cdot \frac{dx}{du} = x(u) \quad (2.5.5)$$

That is to say, in going from f to g (or vice versa) we exchange the roles of x and u . One calls Eq. (2.5.4) a *Legendre transformation* and f and g *Legendre transforms* of each other.

More generally, if $f = f(x_1, x_2, \dots, x_n)$, one can eliminate a subset $\{x_i, i=1 \text{ to } j\}$ in favor of the partial derivatives $u_i = \partial f / \partial x_i$ by the transformation

$$g(u_1, \dots, u_j, x_{j+1}, \dots, x_n) = \sum_{i=1}^j u_i x_i - f(x_1, \dots, x_n) \quad (2.5.6)$$

It is understood in the right-hand side of Eq. (2.5.6) that all the x_i 's to be eliminated have been rewritten as functions of the allowed variables in g . It can be easily verified that

$$\frac{\partial g}{\partial u_i} = x_i \quad (2.5.7)$$

where in taking the above partial derivative, one keeps all the other variables in g constant.

[‡] We will often refer to q_1, \dots, q_n as q and p_1, \dots, p_n as p .

Table 2.1. Comparison of the Lagrangian and Hamiltonian Formalisms

Lagrangian formalism	Hamiltonian formalism
(1) The state of a system with n degrees of freedom is described by n coordinates (q_1, \dots, q_n) and n velocities ($\dot{q}_1, \dots, \dot{q}_n$), or in a more compact notation by (q, \dot{q}) .	(1) The state of a system with n degrees of freedom is described by n coordinates and n momenta ($q_1, \dots, q_n; p_1, \dots, p_n$) or, more succinctly, by (q, p) .
(2) The state of the system may be represented by a point moving with a definite velocity in an n -dimensional configuration space.	(2) The state of the system may be represented by a point in a $2n$ -dimensional phase space, with coordinates $(q_1, \dots, q_n; p_1, \dots, p_n)$.
(3) The n coordinates evolve according to n second-order equations.	(3) The $2n$ coordinates and momenta obey $2n$ first-order equations.
(4) For a given \mathcal{L} , several trajectories may pass through a given point in configuration space depending on \dot{q} .	(4) For a given \mathcal{H} only one trajectory passes through a given point in phase space.

Applying these methods to the problem in question, we define

$$\mathcal{H}(q, p) = \sum_{i=1}^n p_i \dot{q}_i - \mathcal{L}(q, \dot{q}) \quad (2.5.8)$$

where the \dot{q} 's are to be written as functions of q 's and p 's. This inversion is generally easy since \mathcal{L} is a polynomial of rank 2 in \dot{q} , and $p_i = \partial \mathcal{L} / \partial \dot{q}_i$ is a polynomial of rank 1 in the \dot{q} 's, e.g., Eq. (2.2.7). Consider now

$$\frac{\partial \mathcal{H}}{\partial p_i} = \frac{\partial}{\partial p_i} \left(\sum_j p_j \dot{q}_j - \mathcal{L} \right) \quad (2.5.9)$$

$$\begin{aligned} &= \dot{q}_i + \sum_j p_j \frac{\partial \dot{q}_j}{\partial p_i} - \sum_j \frac{\partial \mathcal{L}}{\partial \dot{q}_j} \frac{\partial \dot{q}_j}{\partial p_i} \\ &= \dot{q}_i \quad \left(\text{since } p_j = \frac{\partial \mathcal{L}}{\partial \dot{q}_j} \right) \end{aligned} \quad (2.5.10)$$

[There are no $(\partial \mathcal{L} / \partial \dot{q}_j)(\partial \dot{q}_j / \partial p_i)$ terms since q is held constant in $\partial \mathcal{H} / \partial p_i$; that is, q and p are independent variables.] Similarly,

$$\frac{\partial \mathcal{H}}{\partial q_i} = \sum_j p_j \frac{\partial \dot{q}_j}{\partial q_i} - \frac{\partial \mathcal{L}}{\partial q_i} - \sum_j \frac{\partial \mathcal{L}}{\partial \dot{q}_j} \frac{\partial \dot{q}_j}{\partial q_i} = -\frac{\partial \mathcal{L}}{\partial q_i} \quad (2.5.11)$$

We now feed in the dynamics by replacing $(\partial \mathcal{L} / \partial \dot{q}_i)$ by \dot{p}_i , and obtain *Hamilton's canonical equations*:

$$\frac{\partial \mathcal{H}}{\partial p_i} = \dot{q}_i, \quad -\frac{\partial \mathcal{H}}{\partial q_i} = \dot{p}_i \quad (2.5.12)$$

Note that we have altogether $2n$ first-order equations (in time) for a system with n degrees of freedom. Given the initial-value data, $(q_i(0), p_i(0))$, $i = 1, \dots, n$, we can integrate the equations to get $(q_i(t), p_i(t))$.

Table 2.1 provides a comparison of the Lagrangian and Hamiltonian formalisms.

Now, just as \mathcal{L} may be interpreted as $T - V$ if the force is conservative, so there exists a simple interpretation for \mathcal{H} in this case. Consider the sum $\sum_i p_i \dot{q}_i$. Let us use Cartesian coordinates, in terms of which

$$T = \sum_{i=1}^n \frac{1}{2} m_i \dot{x}_i^2$$

$$p_i = \frac{\partial \mathcal{L}}{\partial \dot{x}_i} = \frac{\partial T}{\partial \dot{x}_i} = m_i \dot{x}_i$$

and

$$\sum_{i=1}^n p_i \dot{x}_i = \sum_{i=1}^n m_i \dot{x}_i^2 = 2T \quad (2.5.13)$$

so that

$$\mathcal{H} = \sum_i p_i \dot{x}_i - \mathcal{L} = T + V \quad (2.5.14)$$

the total energy. Notice that although we used Cartesian coordinates along the way, the resulting equation (2.5.14) is a relation among scalars and thus coordinate independent.

Exercise 2.5.1. Show that if $T = \sum_i \sum_j T_{ij}(q) \dot{q}_i \dot{q}_j$, where \dot{q} 's are generalized velocities, $\sum_i p_i \dot{q}_i = 2T$.

The Hamiltonian method is illustrated by the simple example of a harmonic oscillator, for which

$$\mathcal{L} = \frac{1}{2} m \dot{x}^2 - \frac{1}{2} k x^2$$

The canonical momentum is

$$p = \frac{\partial \mathcal{L}}{\partial \dot{x}} = m \dot{x}$$

It is easy to invert this relation to obtain \dot{x} as a function of p :

$$\dot{x} = p/m$$

and obtain

$$\begin{aligned}\mathcal{H}(x, p) &= T + V = \frac{1}{2}m[\dot{x}(p)]^2 + \frac{1}{2}kx^2 \\ &= \frac{p^2}{2m} + \frac{1}{2}kx^2\end{aligned}\quad (2.5.15)$$

The equations of motion are

$$\frac{\partial \mathcal{H}}{\partial p} = \dot{q} \rightarrow \frac{p}{m} = \dot{x} \quad (2.5.16)$$

$$-\frac{\partial \mathcal{H}}{\partial q} = \dot{p} \rightarrow -kx = \dot{p} \quad (2.5.17)$$

These equations can be integrated in time, given the initial q and p . If, however, we want the familiar second-order equation, we differentiate Eq. (2.5.16) with respect to time, and feed it into Eq. (2.5.17) to get

$$m\ddot{x} + kx = 0$$

Exercise 2.5.2. Using the conservation of energy, show that the trajectories in phase space for the oscillator are ellipses of the form $(x/a)^2 + (p/b)^2 = 1$, where $a^2 = 2E/k$ and $b^2 = 2mE$.

Exercise 2.5.3. Solve Exercise 2.1.2 using the Hamiltonian formalism.

*Exercise 2.5.4.** Show that \mathcal{H} corresponding to \mathcal{L} in Eq. (2.3.6) is $\mathcal{H} = |\mathbf{p}_{CM}|^2/2M + |\mathbf{p}|^2/2\mu + V(\mathbf{r})$, where M is the total mass, μ is the reduced mass, \mathbf{p}_{CM} and \mathbf{p} are the momenta conjugate to \mathbf{r}_{CM} and \mathbf{r} , respectively.

2.6. The Electromagnetic Force in the Hamiltonian Scheme

The passage from \mathcal{L}_{em} to its Legendre transform \mathcal{H}_{em} is not sensitive in any way to the velocity-dependent nature of the force. If \mathcal{L}_{em} generated the correct force laws, so will \mathcal{H}_{em} , the dynamical content of the schemes being identical. In contrast, the velocity independence of the force was assumed in showing that the numerical value of \mathcal{H} is $T + V$, the total energy. Let us therefore repeat the analysis for the electromagnetic case. As

$$\mathcal{L}_{em} = \frac{1}{2}m\mathbf{v} \cdot \mathbf{v} - q\phi + \frac{q}{c}\mathbf{v} \cdot \mathbf{A}$$

and[‡]

[‡] Note that in this discussion, q is the charge and not the coordinate. The (Cartesian) coordinate \mathbf{r} is hidden in the functions $\mathbf{A}(\mathbf{r}, t)$ and $\phi(\mathbf{r}, t)$.

we have

$$\begin{aligned}\mathcal{H}_{e.m} &= \mathbf{p} \cdot \mathbf{v} - \mathcal{L}_{e.m} \\ &= m\mathbf{v} \cdot \mathbf{v} + q \frac{\mathbf{v} \cdot \mathbf{A}}{c} - \frac{1}{2} m\mathbf{v} \cdot \mathbf{v} + q\phi - \frac{q\mathbf{v} \cdot \mathbf{A}}{c} \\ &= \frac{1}{2}m\mathbf{v} \cdot \mathbf{v} + q\phi = T + q\phi\end{aligned}\quad (2.6.1)$$

Now, there is something very disturbing about Eq. (2.6.1): the vector potential \mathbf{A} seems to have dropped out along the way. How is $\mathcal{H}_{e.m}$ to generate the correct dynamics without knowing what \mathbf{A} is? The answer is, of course, the \mathcal{H} is more than just $T + q\phi$; it is $T + q\phi$ written in terms of the correct variables, in particular, in terms of \mathbf{p} and not \mathbf{v} . Making the change of variables, we get

$$\mathcal{H}_{e.m} = \frac{|(\mathbf{p} - q\mathbf{A}/c)|^2}{2m} + q\phi \quad (2.6.2)$$

with the vector potential very much in the picture.

2.7. Cyclic Coordinates, Poisson Brackets, and Canonical Transformations

Cyclic coordinates are defined here just as in the Lagrangian case and have the same significance: if a coordinate q_i is missing in \mathcal{H} , then

$$\dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q_i} = 0 \quad (2.7.1)$$

Now, there will be other quantities, such as the energy, that may be conserved in addition to the canonical momenta.[§] There exists a nice method of characterizing these in the Hamiltonian formalism. Let $\omega(p, q)$ be some function of the state variables, *with no explicit dependence on t*. Its time variation is given by

$$\begin{aligned}\frac{d\omega}{dt} &= \sum_i \left(\frac{\partial \omega}{\partial q_i} \dot{q}_i + \frac{\partial \omega}{\partial p_i} \dot{p}_i \right) \\ &= \sum_i \left(\frac{\partial \omega}{\partial q_i} \frac{\partial \mathcal{H}}{\partial p_i} - \frac{\partial \omega}{\partial p_i} \frac{\partial \mathcal{H}}{\partial q_i} \right) \\ &\equiv \{\omega, \mathcal{H}\}\end{aligned}\quad (2.7.2)$$

[§] Another example is the conservation of $l_z = xp_y - yp_x$ when $V(x, y) = V(x^2 + y^2)$. There are no cyclic coordinates here. Of course, if we work in polar coordinates, $V(\rho, \phi) = V(\rho)$, and $p_\phi = m\rho^2 \dot{\phi} = l_z$ is conserved because it is the momentum conjugate to the cyclic coordinate ϕ .

where we have defined the Poisson bracket (PB) between two variables $\omega(p, q)$ and $\lambda(p, q)$ to be

$$\{\omega, \lambda\} \equiv \sum_i \left(\frac{\partial \omega}{\partial q_i} \frac{\partial \lambda}{\partial p_i} - \frac{\partial \omega}{\partial p_i} \frac{\partial \lambda}{\partial q_i} \right) \quad (2.7.3)$$

It follows from Eq. (2.7.2) that any variable whose PB with \mathcal{H} vanishes is constant in time, i.e., conserved. In particular \mathcal{H} itself is a constant of motion (identified as the total energy) if it has no explicit t dependence.

*Exercise 2.7.1.** Show that

$$\{\omega, \lambda\} = -\{\lambda, \omega\}$$

$$\{\omega, \lambda + \sigma\} = \{\omega, \lambda\} + \{\omega, \sigma\}$$

$$\{\omega, \lambda\sigma\} = \{\omega, \lambda\}\sigma + \lambda\{\omega, \sigma\}$$

Note the similarity between the above and Eqs. (1.5.10) and (1.5.11) for commutators.

Of fundamental importance are the PB between the q 's and the p 's. Observe that

$$\{q_i, q_j\} = \{p_i, p_j\} = 0 \quad (2.7.4a)$$

$$\{q_i, p_j\} = \delta_{ij} \quad (2.7.4b)$$

since (q_1, \dots, p_n) are independent variables ($\partial q_i / \partial q_j = \delta_{ij}$, $\partial q_i / \partial p_k = 0$, etc.). Hamilton's equations may be written in terms of PB as

$$\dot{q}_i = \{q_i, \mathcal{H}\} \quad (2.7.5a)$$

$$\dot{p}_i = \{p_i, \mathcal{H}\} \quad (2.7.5b)$$

by setting $\omega = q_i$ or p_i in Eq. (2.7.2).

*Exercise 2.7.2.** (i) Verify Eqs. (2.7.4) and (2.7.5). (ii) Consider a problem in two dimensions given by $\mathcal{H} = p_x^2 + p_y^2 + ax^2 + by^2$. Argue that if $a = b$, $\{l_z, \mathcal{H}\}$ must vanish. Verify by explicit computation.

Canonical Transformations

We have seen that the Euler–Lagrange equations are form invariant under an arbitrary‡ change of coordinates in configuration space

$$q_i \rightarrow \bar{q}_i(q_1, \dots, q_n), \quad i = 1, \dots, n \quad (2.7.6a)$$

‡ We assume the transformation is invertible, so we may write q in terms of \bar{q} : $q = q(\bar{q})$. The transformation may also depend on time explicitly [$\bar{q} = \bar{q}(q, t)$], but we do not consider such cases.

or more succinctly

$$q \rightarrow \bar{q}(q) \quad (2.7.6b)$$

The response of the velocities to this transformation follows from Eq. (2.7.6a):

$$\dot{q}_i = \dot{\bar{q}}_i = \frac{d\bar{q}_i}{dt} = \sum_j \left(\frac{\partial \bar{q}_i}{\partial q_j} \right) \dot{q}_j \quad (2.7.7)$$

The response of the canonical momenta may be found by rewriting \mathcal{L} in terms of $(\bar{q}, \dot{\bar{q}})$ and taking the derivative with respect to $\dot{\bar{q}}$:

$$\bar{p}_i = \frac{\partial \mathcal{L}(\bar{q}, \dot{\bar{q}})}{\partial \dot{\bar{q}}_i} \quad (2.7.8)$$

The result is (Exercise 2.7.8):

$$\bar{p}_i = \sum_j \left(\frac{\partial q_j}{\partial \bar{q}_i} \right) p_j \quad (2.7.9)$$

Notice that although \mathcal{L} enters Eq. (2.7.8), it drops out in Eq. (2.7.9), which connects \bar{p} to the old variables. This is as it should be, for we expect that the response of the momenta to a coordinate transformation (say, a rotation) is a purely kinematical question.

A word of explanation about $\mathcal{L}(\bar{q}, \dot{\bar{q}})$. By $\mathcal{L}(\bar{q}, \dot{\bar{q}})$ we mean the Lagrangian (say $T - V$, for definiteness) written in terms of \bar{q} and $\dot{\bar{q}}$. Thus the numerical value of the Lagrangian is unchanged under $(q, \dot{q}) \rightarrow (\bar{q}, \dot{\bar{q}})$; for (q, \dot{q}) and $(\bar{q}, \dot{\bar{q}})$ refer to the *same physical state*. The functional form of the Lagrangian, however, *does* change and so we should really be using two different symbols $\mathcal{L}(q, \dot{q})$ and $\bar{\mathcal{L}}(\bar{q}, \dot{\bar{q}})$. Nonetheless we follow the convention of denoting a given dynamical variable, such as the Lagrangian, by a fixed symbol in all coordinate systems.

The invariance of the Euler–Lagrange equations under $(q, \dot{q}) \rightarrow (\bar{q}, \dot{\bar{q}})$ implies the invariance of Hamilton’s equation under $(q, p) \rightarrow (\bar{q}, \bar{p})$, i.e., (\bar{q}, \bar{p}) obey

$$\dot{\bar{q}}_i = \partial \mathcal{H} / \partial \bar{p}_i, \quad \dot{\bar{p}}_i = -(\partial \mathcal{H} / \partial \bar{q}_i) \quad (2.7.10)$$

where $\mathcal{H} = \mathcal{H}(\bar{q}, \bar{p})$ is the Hamiltonian written in terms of \bar{q} and \bar{p} . The proof is simple: we start with $\mathcal{L}(\bar{q}, \dot{\bar{q}})$, perform a Legendre transform, and use the fact that \bar{q} obeys Euler–Lagrange equations.

The transformation

$$q_i \rightarrow \bar{q}_i(q_1, \dots, q_n), \quad \bar{p}_i = \sum_j \left(\frac{\partial q_j}{\partial \bar{q}_i} \right) p_j \quad (2.7.11)$$

is called a *point transformation*. If we view the Hamiltonian formalism as something derived from the Lagrangian scheme, which is formulated in n -dimensional configuration space, this is the most general (time-independent) transformation which preserves the form of Hamilton's equations (that we can think of). On the other hand, if we view the Hamiltonian formalism in its own right, the backdrop is the $2n$ -dimensional phase space. In this space, the point transformation is unnecessarily restrictive. One can contemplate a more general transformation of phase space coordinates:

$$\begin{aligned} q &\rightarrow \bar{q}(q, p) \\ p &\rightarrow \bar{p}(q, p) \end{aligned} \quad (2.7.12)$$

Although all sets of $2n$ independent coordinates (\bar{q}, \bar{p}) are formally adequate for describing the state of the system, not all of them will preserve the canonical form of Hamilton's equations. (This is like saying that although Newton's laws may be written in terms of any complete set of coordinates, the simple form $m\ddot{q}_i = -\partial V/\partial q_i$ is valid only if the q_i are Cartesian). If, however, (\bar{q}, \bar{p}) obey the canonical equations (2.7.10), we say that they are *canonical coordinates* and that Eq. (2.7.12) defines a *canonical transformation*. Any set of coordinates (q_1, \dots, q_n) , and the corresponding momenta generated in the Lagrangian formalism ($p_i = \partial \mathcal{L}/\partial \dot{q}_i$), are canonical coordinates. Given one set, (q, p) , we can get another, (\bar{q}, \bar{p}) , by the point transformation, which is a special case of the canonical transformation. This does not, however, exhaust the possibilities. Let us now ask the following question. *Given a new set of coordinates $(\bar{q}(q, p), \bar{p}(q, p))$, how can we tell if they are canonical [assuming (q, p) are]?* Now it is true for any $\omega(q, p)$ that

$$\dot{\omega} = \{\omega, \mathcal{H}\} = \sum_i \left(\frac{\partial \omega}{\partial q_i} \frac{\partial \mathcal{H}}{\partial p_i} - \frac{\partial \omega}{\partial p_i} \frac{\partial \mathcal{H}}{\partial q_i} \right) \quad (2.7.13)$$

Applying this to $\bar{q}_j(q, p)$ we find

$$\dot{\bar{q}}_j = \sum_i \left(\frac{\partial \bar{q}_j}{\partial q_i} \frac{\partial \mathcal{H}}{\partial p_i} - \frac{\partial \bar{q}_j}{\partial p_i} \frac{\partial \mathcal{H}}{\partial q_i} \right) \quad (2.7.14)$$

If we view \mathcal{H} as a function of (\bar{q}, \bar{p}) and use the chain rule, we get

$$\frac{\partial \mathcal{H}(q, p)}{\partial p_i} = \frac{\partial \mathcal{H}(\bar{q}, \bar{p})}{\partial p_i} = \sum_k \left(\frac{\partial \mathcal{H}}{\partial \bar{q}_k} \frac{\partial \bar{q}_k}{\partial p_i} + \frac{\partial \mathcal{H}}{\partial \bar{p}_k} \frac{\partial \bar{p}_k}{\partial p_i} \right) \quad (2.7.15a)$$

and

$$\frac{\partial \mathcal{H}(q, p)}{\partial q_i} = \frac{\partial \mathcal{H}(\bar{q}, \bar{p})}{\partial q_i} = \sum_k \left(\frac{\partial \mathcal{H}}{\partial \bar{q}_k} \frac{\partial \bar{q}_k}{\partial q_i} + \frac{\partial \mathcal{H}}{\partial \bar{p}_k} \frac{\partial \bar{p}_k}{\partial q_i} \right) \quad (2.7.15b)$$

Feeding all this into Eq. (2.7.14) we find, upon regrouping terms,

$$\dot{\bar{q}}_j = \sum_k \left(\frac{\partial \mathcal{H}}{\partial \bar{q}_k} \{ \bar{q}_j, \bar{q}_k \} + \frac{\partial \mathcal{H}}{\partial \bar{p}_k} \{ \bar{q}_j, \bar{p}_k \} \right) \quad (2.7.16)$$

It can similarly be established that

$$\dot{\bar{p}}_j = \sum_k \left(\frac{\partial \mathcal{H}}{\partial \bar{q}_k} \{ \bar{p}_j, \bar{q}_k \} + \frac{\partial \mathcal{H}}{\partial \bar{p}_k} \{ \bar{p}_j, \bar{p}_k \} \right) \quad (2.7.17)$$

If Eqs. (2.7.16) and (2.7.17) are to reduce to the canonical equations (2.7.10) for any $\mathcal{H}(q, p)$, we must have

$$\begin{aligned} \{ \bar{q}_j, \bar{q}_k \} &= 0 = \{ \bar{p}_j, \bar{p}_k \} \\ \{ \bar{q}_j, \bar{p}_k \} &= \delta_{jk} \end{aligned} \quad (2.7.18)$$

These then are the conditions to be satisfied by the new variables if they are to be canonical. Notice that these constraints make no reference to the specific functional form of \mathcal{H} : the equations defining canonical variables are purely kinematical and true for any $\mathcal{H}(q, p)$.

Exercise 2.7.3. Fill in the missing steps leading to Eq. (2.7.18) starting from Eq. (2.7.14).

Exercise 2.7.4. Verify that the change to a rotated frame

$$\bar{x} = x \cos \theta - y \sin \theta$$

$$\bar{y} = x \sin \theta + y \cos \theta$$

$$\bar{p}_x = p_x \cos \theta - p_y \sin \theta$$

$$\bar{p}_y = p_x \sin \theta + p_y \cos \theta$$

is a canonical transformation.

Exercise 2.7.5. Show that the polar variables $\rho = (x^2 + y^2)^{1/2}$, $\phi = \tan^{-1}(y/x)$,

$$p_\rho = \hat{e}_\rho \cdot \mathbf{p} = \frac{x p_x + y p_y}{(x^2 + y^2)^{1/2}}, \quad p_\phi = x p_y - y p_x (= l_z)$$

are canonical. (\hat{e}_ρ is the unit vector in the radial direction.)

*Exercise 2.7.6.** Verify that the change from the variables $\mathbf{r}_1, \mathbf{r}_2, \mathbf{p}_1, \mathbf{p}_2$ to $\mathbf{r}_{\text{CM}}, \mathbf{p}_{\text{CM}}, \mathbf{r}$, and \mathbf{p} is a canonical transformation. (See Exercise 2.5.4).

Exercise 2.7.7. Verify that

$$\bar{q} = \ln(q^{-1} \sin p)$$

$$\bar{p} = q \cot p$$

is a canonical transformation.

Exercise 2.7.8. We would like to derive here Eq. (2.7.9), which gives the transformation of the momenta under a coordinate transformation in configuration space:

$$q_i \rightarrow \bar{q}_i(q_1, \dots, q_n)$$

(1) Argue that if we invert the above equation to get $q = q(\bar{q})$, we can derive the following counterpart of Eq. (2.7.7):

$$\dot{q}_i = \sum_j \frac{\partial q_i}{\partial \dot{\bar{q}}_j} \dot{\bar{q}}_j$$

(2) Show from the above that

$$\left(\frac{\partial \dot{q}_i}{\partial \dot{\bar{q}}_j} \right)_q = \frac{\partial q_i}{\partial \bar{q}_j}$$

(3) Now calculate

$$\bar{p}_i = \left[\frac{\partial \mathcal{L}(\bar{q}, \dot{\bar{q}})}{\partial \dot{\bar{q}}_i} \right]_q = \left[\frac{\partial \mathcal{L}(q, \dot{q})}{\partial \dot{q}_i} \right]_q$$

Use the chain rule and the fact that $q = q(\bar{q})$ and not $q(\bar{q}, \dot{\bar{q}})$ to derive Eq. (2.7.9).

(4) Verify, by calculating the PB in Eq. (2.7.18), that the point transformation is canonical.

If (q, p) and (\bar{q}, \bar{p}) are both canonical, we must give them both the same status, for Hamilton's equations have the same appearance when expressed in terms of either set. Now, we have defined the PB of two variables ω and σ in terms of (q, p) as

$$\{\omega, \sigma\} = \sum_i \left(\frac{\partial \omega}{\partial q_i} \frac{\partial \sigma}{\partial p_i} - \frac{\partial \omega}{\partial p_i} \frac{\partial \sigma}{\partial q_i} \right) \equiv \{\omega, \sigma\}_{q,p}$$

Should we not also define a PB, $\{\omega, \sigma\}_{\bar{q}, \bar{p}}$ for every canonical pair (\bar{q}, \bar{p}) ? Fortunately it turns out that *the PB are invariant under canonical transformations*:

$$\{\omega, \sigma\}_{q,p} = \{\omega, \sigma\}_{\bar{q}, \bar{p}} \quad (2.7.19)$$

(It is understood that ω and σ are written as functions of \bar{q} and \bar{p} on the right-hand side.)

Exercise 2.7.9. Verify Eq. (2.7.19) by direct computation. Use the chain rule to go from q, p derivatives to \bar{q}, \bar{p} derivatives. Collect terms that represent PB of the latter.

Besides the proof by direct computation (as per Exercise 2.7.9 above) there is an alternate way to establish Eq. (2.7.19).

Consider first $\sigma = \mathcal{H}$. We know that since (q, p) obey canonical equations,

$$\dot{\omega} = \{\omega, \mathcal{H}\}_{q,p}$$

But then (\bar{q}, \bar{p}) also obey canonical equations, so

$$\dot{\omega} = \{\omega, \mathcal{H}\}_{\bar{q},\bar{p}}$$

Now ω is some physical quantity such as the kinetic energy or the component of angular momentum in some fixed direction, so its rate of change is independent of the phase space coordinates used, i.e., $\dot{\omega}$ is $\dot{\omega}$, whether $\omega = \omega(q, p)$ or $\omega(\bar{q}, \bar{p})$. So

$$\{\omega, \mathcal{H}\}_{q,p} = \{\omega, \mathcal{H}\}_{\bar{q},\bar{p}} \quad (2.7.20)$$

Having proved the result for what seems to be the special case $\sigma = \mathcal{H}$, we now pull the following trick. Note that nowhere in the derivation did we have to assume that \mathcal{H} was any particular function of q and p . In fact, Hamiltonian dynamics, as a consistent mathematical scheme, places no restriction on \mathcal{H} . It is the physical requirement that the time evolution generated by \mathcal{H} coincide with what is *actually* observed, that restricts \mathcal{H} to be $T + V$. Thus \mathcal{H} could have been any function at all in the preceding argument and in the result Eq. (2.7.20) (which is just a relation among partial derivatives.) If we understand that \mathcal{H} is not $T + V$ in *this argument* but an arbitrary function, call it σ , we get the desired result.

Active Transformations

So far, we have viewed the transformation

$$\bar{q} = \bar{q}(q, p)$$

$$\bar{p} = \bar{p}(q, p)$$

as passive: both (q, p) and (\bar{q}, \bar{p}) refer to the same point in phase space described in two different coordinate systems. Under the transformation $(q, p) \rightarrow (\bar{q}, \bar{p})$, the numerical values of all dynamical variables are unchanged (for we are talking about the same physical state), but their functional form is changed. For instance, under a change from Cartesian to spherical coordinates, $\omega(x, y, z) = x^2 + y^2 + z^2 \rightarrow \omega(r, \theta, \phi) = r^2$. As mentioned earlier, we use the same symbol for a given variable even if its functional dependence on the coordinates changes when we change coordinates.

Consider now a restricted class of transformations, called *regular transformations*, which preserve the range of the variables: (q, p) and (\bar{q}, \bar{p}) have the same range. A change from one Cartesian coordinate to a translated or rotated one is

regular (each variable goes from $-\infty$ to $+\infty$ before and after), whereas a change to spherical coordinates (where some coordinates are nonnegative, some are bounded by 2π , etc.) is not.

A regular transformation $(q, p) \rightarrow (\bar{q}, \bar{p})$ permits an alternate interpretation: instead of viewing (\bar{q}, \bar{p}) as the same phase space point in a new coordinate system, we may view it as a new point in the same coordinate system. This corresponds to an active transformation which changes the state of the system. Under this change, the numerical value of any dynamical variable $\omega(q, p)$ will generally change: $\omega(q, p) \neq \omega(\bar{q}, \bar{p})$, though its functional dependence will not: $\omega(\bar{q}, \bar{p})$ is the same function $\omega(q, p)$ evaluated at the new point ($q = \bar{q}$, $p = \bar{p}$).

We say that ω is *invariant* under the regular transformation $(q, p) \rightarrow (\bar{q}, \bar{p})$ if

$$\omega(q, p) = \omega(\bar{q}, \bar{p}) \quad (2.7.21)$$

(This equation has content only if we are talking about the active transformations, for it is true for any ω under a passive transformation.)

Whether we view the transformation $(q, p) \rightarrow (\bar{q}, \bar{p})$ as active or passive, it is called canonical if (\bar{q}, \bar{p}) obey Eq. (2.7.18). As we shall see, only regular *canonical* transformations are physically interesting.

2.8. Symmetries and Their Consequences

Let us begin our discussion by examining what the word “symmetry” means in daily usage. We say that a sphere is a very symmetric object because it looks the same when seen from many directions. Or, equivalently, a sphere looks the same before and after it is subjected to a rotation around *any* axis passing through its center. A cylinder has symmetry too, but not as much: the rotation must be performed around its axis. Generally then, the symmetry of an object implies its invariance under some transformations, which in our examples are rotations.

A symmetry can be discrete or continuous, as illustrated by the example of a hexagon and a circle. While the rotation angles that leave a hexagon unchanged form a discrete set, namely, multiples of 60° , the corresponding set for a circle is a continuum. We may characterize the continuous symmetry of the circle in another way. Consider the *identity transformation*, which does nothing, i.e., rotates by 0° in our example. This leaves both the circle and the hexagon invariant. Consider next an *infinitesimal transformation*, which is infinitesimally “close” to the identity; in our example this is a rotation by an infinitesimal angle ε . The infinitesimal rotation leaves the circle invariant but not the hexagon. The circle is thus characterized by its invariance under infinitesimal rotations. Given this property, its invariance under finite rotations follows, for any finite rotation may be viewed as a sequence of infinitesimal rotations (each of which leaves it invariant).

It is also possible to think of functions of some variables as being symmetric in the sense that if one changes the values of the variables in a certain way, the value of the function is invariant. Consider for example

$$f(x, y) = x^2 + y^2$$

If we make the following change

$$\begin{aligned} x \rightarrow \bar{x} &= x \cos \theta - y \sin \theta \\ y \rightarrow \bar{y} &= x \sin \theta + y \cos \theta \end{aligned} \quad (2.8.1)$$

in the arguments, we find that f is invariant. We say that f is symmetric under the above transformation. In the terminology introduced earlier, the transformation in question is continuous: its infinitesimal version is

$$\begin{aligned} x \rightarrow \bar{x} &= x \cos \varepsilon - y \sin \varepsilon = x - y\varepsilon \\ y \rightarrow \bar{y} &= x \sin \varepsilon + y \cos \varepsilon = x\varepsilon + y \quad (\text{to order } \varepsilon) \end{aligned} \quad (2.8.2)$$

Consider now the function $\mathcal{H}(q, p)$. There are two important dynamical consequences that follow from its invariance under *regular canonical* transformations.

I. If \mathcal{H} is invariant under the following *infinitesimal* transformation (which you may verify is canonical, Exercise 2.8.2),

$$\begin{aligned} q_i \rightarrow \bar{q}_i &= q_i + \varepsilon \frac{\partial g}{\partial p_i} \equiv q_i + \delta q_i \\ p_i \rightarrow \bar{p}_i &= p_i - \varepsilon \frac{\partial g}{\partial q_i} \equiv p_i + \delta p_i \end{aligned} \quad (2.8.3)$$

where $g(q, p)$ is any dynamical variable, *then* g is *conserved*, i.e., a constant of motion. One calls g the *generator of the transformation*.

II. If \mathcal{H} is invariant under the regular, canonical, but not necessarily infinitesimal, transformation $(q, p) \rightarrow (\bar{q}, \bar{p})$, and if $(q(t), p(t))$ is a solution to the equations of motion, so is the transformed (translated, rotated, etc.) trajectory, $(\bar{q}(t), \bar{p}(t))$.

Let us now analyze these two consequences.

Consequence I. Let us first verify that g is indeed conserved if \mathcal{H} is invariant under the transformation it generates. Working to first order in ε , if we equate the change in \mathcal{H} under the change of its arguments to zero, we get

$$\delta \mathcal{H} = \sum_i \frac{\partial \mathcal{H}}{\partial q_i} \left(\varepsilon \frac{\partial g}{\partial p_i} \right) + \frac{\partial \mathcal{H}}{\partial p_i} \left(-\varepsilon \frac{\partial g}{\partial q_i} \right) = \varepsilon \{ \mathcal{H}, g \} = 0 \quad (2.8.4)$$

But according to Eq. (2.7.2),

$$\{g, \mathcal{H}\} = 0 \rightarrow g \text{ is conserved} \quad (2.8.5)$$

(More generally, the response of any variable ω to the transformation is

$$\delta \omega = \varepsilon \{ \omega, g \} \quad (2.8.6)$$

Note that δp and δq in Eq. (2.8.3) may also be written as PBs.) Consider as an example, a particle in one dimension and the case $g=p$. From Eq. (2.8.3),

$$\begin{aligned}\delta x &= \varepsilon \frac{\partial p}{\partial p} = \varepsilon \\ \delta p &= -\varepsilon \frac{\partial p}{\partial x} = 0\end{aligned}\tag{2.8.7}$$

which we recognize to be an infinitesimal translation. Thus the linear momentum p is the generator of spatial translations and is conserved in a translationally invariant problem. The physics behind this result is clear. Since p is unchanged in a translation, so is $T=p^2/2m$. Consequently $V(x+\varepsilon)=V(x)$. But if the potential doesn't vary from point to point, there is no force and p is conserved.

Next consider an example from two dimensions with $g=l_z=xp_y-yp_x$. Here,

$$\begin{aligned}\delta x &= -y\varepsilon \left(= \varepsilon \frac{\partial l_z}{\partial p_x} \right) \\ \delta y &= x\varepsilon \left(= \varepsilon \frac{\partial l_z}{\partial p_y} \right) \\ \delta p_x &= -p_y\varepsilon \left(= -\varepsilon \frac{\partial l_z}{\partial x} \right) \\ \delta p_y &= p_x\varepsilon \left(= -\varepsilon \frac{\partial l_z}{\partial y} \right)\end{aligned}\tag{2.8.8}$$

which we recognize to be an infinitesimal rotation around the z axis, [Eq. (2.8.2)]. Thus the angular momentum around the z axis is the generator of rotations around that axis, and is conserved if \mathcal{H} is invariant under rotations of the state around that axis. The relation between the symmetry and the conservation law may be understood in the following familiar terms. Under the rotation of the coordinates and the momenta, $|\mathbf{p}|$ doesn't change and so neither does $T=|\mathbf{p}|^2/2m$. Consequently, V is a constant as we go along any circle centered at the origin. This in turn means that there is no force in the tangential direction and so no torque around the z axis. The conservation of l_z then follows.

Exercise 2.8.1. Show that $p=p_1+p_2$, the total momentum, is the generator of infinitesimal translations for a two-particle system.

*Exercise 2.8.2.** Verify that the infinitesimal transformation generated by any dynamical variable g is a canonical transformation. (Hint: Work, as usual, to first order in ε .)

Exercise 2.8.3. Consider

$$\mathcal{H} = \frac{p_x^2 + p_y^2}{2m} + \frac{1}{2}m\omega^2(x^2 + y^2)$$

whose invariance under the rotation of the coordinates *and* momenta leads to the conservation of l_z . But \mathcal{H} is also invariant under the rotation of *just the coordinates*. Verify that this is a *noncanonical* transformation. Convince yourself that in this case it is not possible to write $\delta\mathcal{H}$ as $\varepsilon\{\mathcal{H}, g\}$ for any g , i.e., that no conservation law follows.

*Exercise 2.8.4.** Consider $\mathcal{H} = \frac{1}{2}p^2 + \frac{1}{2}x^2$, which is invariant under infinitesimal rotations in *phase space* (the x - p plane). Find the generator of this transformation (after verifying that it is canonical). (You could have guessed the answer based on Exercise 2.5.2.).

The preceding analysis yields, as a by-product, a way to generate infinitesimal canonical transformations. We take any function $g(q, p)$ and obtain the transformation given by Eq. (2.8.6). (Recall that although we defined a canonical transformation earlier, until now we had no means of generating one.) Given an infinitesimal canonical transformation, we can get a finite one by “integrating” it. The following examples should convince you that this is possible. Consider the transformation generated by $g = \mathcal{H}$. We have

$$\begin{aligned}\delta q_i &= \varepsilon\{q_i, \mathcal{H}\} \\ \delta p_i &= \varepsilon\{p_i, \mathcal{H}\}\end{aligned}\tag{2.8.9}$$

But we know from the equations of motion that $\dot{q}_i = \{q_i, \mathcal{H}\}$ etc. So

$$\begin{aligned}\delta q_i &= \varepsilon\dot{q}_i \\ \delta p_i &= \varepsilon\dot{p}_i\end{aligned}\tag{2.8.10}$$

Thus the new point in phase space $(\bar{q}, \bar{p}) = (q + \delta q, p + \delta p)$ obtained by this canonical transformation of (q, p) is just the point to which (q, p) would move in an infinitesimal time interval ε . In other words, the motion of points in phase space under the time evolution generated by \mathcal{H} is an active canonical transformation. Now, you know that by integrating the equations of motion, we can find (\bar{q}, \bar{p}) at any future time, i.e., get the finite canonical transformation. Consider now a general case of $g \neq \mathcal{H}$. We still have

$$\begin{aligned}\delta q_i &= \varepsilon\{q_i, g\} \\ \delta p_i &= \varepsilon\{p_i, g\}\end{aligned}\tag{2.8.11}$$

Mathematically, these equations are identical to Eq. (2.8.9), with g playing the role of the Hamiltonian. Clearly there should be no problem integrating these equations for the evolution of the phase space points under the “fake” Hamiltonian g , and fake “time” ε . Let us consider for instance the case $g = l_z$ which has units erg sec and the corresponding fake time $\varepsilon = \delta\theta$, an angle. The transformation of the coordinates is

$$\begin{aligned}\delta x &= \varepsilon\{x, l_z\} = -\varepsilon y \equiv (-\delta\theta)y \\ \delta y &= (\delta\theta)x\end{aligned}\tag{2.8.12}$$

The fake equations of motion are

$$\frac{dx}{d\theta} = -y, \quad \frac{dy}{d\theta} = x \quad (2.8.13)$$

Differentiating first with respect to θ , and using the second, we get

$$\frac{d^2x}{d\theta^2} + x = 0$$

and likewise,

$$\frac{d^2y}{d\theta^2} + y = 0$$

So

$$x = A \cos \theta + B \sin \theta$$

$$y = C \sin \theta + D \cos \theta$$

We find the constants from the “initial” ($\theta = 0$) coordinates and “velocities”: $A = x_0$, $D = y_0$, $B = (\partial x / \partial \theta)_0 = -y_0$, $C = (\partial y / \partial \theta)_0 = x_0$. Reverting to the standard notation in which (x, y) , rather than (x_0, y_0) , labels the initial point and (\bar{x}, \bar{y}) , rather than (x, y) , denotes the transformed one, we may write the finite canonical transformation (a finite rotation) as

$$\begin{aligned} \bar{x} &= x \cos \theta - y \sin \theta \\ \bar{y} &= x \sin \theta + y \cos \theta \end{aligned} \quad (2.8.14)$$

Similar equations may be derived for \bar{p}_x and \bar{p}_y in terms of p_x and p_y .

Although a wide class of canonical transformations is now open to us, there are many that aren’t. For instance, $(q, p) \rightarrow (-q, -p)$ is a discrete canonical transformation that has no infinitesimal version. There are also the transformations that are not regular, such as the change from Cartesian to spherical coordinates, which have neither infinitesimal forms, nor an active interpretation. We do not consider ways of generating these.[‡]

Consequence II. Let us understand the content of this result through an example before turning to the proof. Consider a two-particle system whose Hamiltonian is invariant under the translation of the entire system, i.e., both particles. Let an observer S_A prepare, at $t = 0$, a state $(x_1^0, x_2^0; p_1^0, p_2^0)$ which evolves as $(x_1(t), x_2(t); p_1(t), p_2(t))$ for some time and ends up in the state $(x_1^T, x_2^T; p_1^T, p_2^T)$ at time T . Let

[‡] For an excellent and lucid treatment of this question and many other topics in advanced classical mechanics, see H. Goldstein, *Classical Mechanics*, Addison-Wesley, Reading, Massachusetts (1950); E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective*, Wiley, New York (1974).

us call the final state the outcome of the experiment conducted by S_A . We are told that as a result of the translational invariance of \mathcal{H} , any other trajectory that is related to this by an arbitrary translation a is also a solution to the equations of motion. In this case, the initial state, for example, is $(x_1^0 + a, x_2^0 + a; p_1^0, p_2^0)$. *The final state and all intermediate states are likewise displaced by the same amount.* To an observer S_B , displaced relative to S_A by an amount a , the evolution of the second system will appear to be identical to what S_A saw in the first. Assuming for the sake of this argument that S_B had in fact prepared the second system, we may say that a given experiment and its translated version will give the same result (as seen by the observers who conducted them) if \mathcal{H} is translationally invariant.

The physical idea is the following. For the usual reasons, translational invariance of \mathcal{H} implies the invariance of $V(x_1, x_2)$. This in turn means that $V(x_1, x_2) = V(x_1 - x_2)$. Thus each particle cares only about where the other is relative to it, and not about where the system as a whole is in space. Consequently the outcome of the experiment is not affected by an overall translation.

Consequence II is just a generalization of this result to other canonical transformations that leave \mathcal{H} invariant. For instance, if \mathcal{H} is rotationally invariant, a given experiment and its rotated version will give the same result (according to the observers who conducted them).

Let us now turn to the proof of the general result.

Proof. Imagine a trajectory $(q(t), p(t))$ in phase space that satisfies the equations of motion. Let us associate with it an image trajectory, $(\bar{q}(t), \bar{p}(t))$, which is obtained by transforming each point (q, p) to the image point (\bar{q}, \bar{p}) by means of a regular canonical transformation. We ask if the image point moves according to Hamilton's equation of motion, i.e., if

$$\dot{\bar{q}}_j = \frac{\partial \mathcal{H}(\bar{q}, \bar{p})}{\partial \bar{p}_j}, \quad \dot{\bar{p}}_j = -\frac{\partial \mathcal{H}(\bar{q}, \bar{p})}{\partial \bar{q}_j} \quad (2.8.15)$$

if \mathcal{H} is invariant under the transformation $(q, p) \rightarrow (\bar{q}, \bar{p})$. Now $\dot{\bar{q}}_j(q, p)$, like any dynamical variable $\omega(q, p)$, obeys

$$\dot{\bar{q}}_j = \{\bar{q}_j, \mathcal{H}(q, p)\}_{q,p} \quad (2.8.16)$$

If $(q, p) \rightarrow (\bar{q}, \bar{p})$ were a *passive* canonical transformation, we could write, since the PB are invariant under such a transformation,

$$\dot{\bar{q}}_j = \{\bar{q}_j, \mathcal{H}(q, p)\}_{q,p} = \{\bar{q}_j, \mathcal{H}(\bar{q}, \bar{p})\}_{\bar{q},\bar{p}} = \frac{\partial \mathcal{H}(\bar{q}, \bar{p})}{\partial \bar{p}_j}$$

But it is an active transformation. However, *because of the symmetry of \mathcal{H}* , i.e., $\mathcal{H}(q, p) = \mathcal{H}(\bar{q}, \bar{p})$, we can go through the very same steps that led to Eq. (2.7.16) from Eq. (2.7.14) and prove the result. If you do not believe this, you may verify it

by explicit computation using $\mathcal{H}(q, p) = \mathcal{H}(\bar{q}, \bar{p})$. A similar argument shows that

$$\dot{\bar{p}}_j = -\frac{\partial \mathcal{H}(\bar{q}, \bar{p})}{\partial \bar{q}_j} \quad (2.8.17)$$

So the image point moves according to Hamilton's equations. Q.E.D.

Exercise 2.8.5. Why is it that a *noncanonical* transformation that leaves \mathcal{H} invariant does not map a solution into another? Or, in view of the discussions on consequence II, why is it that an experiment and its transformed version do not give the same result when the transformation that leaves \mathcal{H} invariant is not canonical? It is best to consider an example. Consider the potential given in Exercise 2.8.3. Suppose I release a particle at $(x=a, y=0)$ with $(p_x=b, p_y=0)$ and you release one in the transformed state in which $(x=0, y=a)$ and $(p_x=b, p_y=0)$, i.e., you rotate the coordinates but not the momenta. This is a noncanonical transformation that leaves \mathcal{H} invariant. Convince yourself that at later times the states of the two particles are not related by the same transformation. Try to understand what goes wrong in the general case.

As you go on and learn quantum mechanics, you will see that the symmetries of the Hamiltonian have similar consequences for the dynamics of the system.

A Useful Relation Between S and E

We now prove a result that will be invoked in Chapter 16:

$$\frac{\partial S_{\text{cl}}(x_f, t_f; x_i, t_i)}{\partial t_f} = -\mathcal{H}(t_f)$$

where $S_{\text{cl}}(x_f, t_f; x_i, t_i)$ is the action of the classical path from x_i, t_i to x_f, t_f and \mathcal{H} is the Hamiltonian at the upper end point. Since we shall be working with problems where energy is conserved we may write

$$\frac{\partial S_{\text{cl}}(x_f, t_f; x_i, t_i)}{\partial t_f} = -E \quad (2.8.18)$$

where E is the conserved energy, constant on the whole trajectory.

At first sight you may think that since

$$S_{\text{cl}} = \int_{t_i}^{t_f} \mathcal{L} dt$$

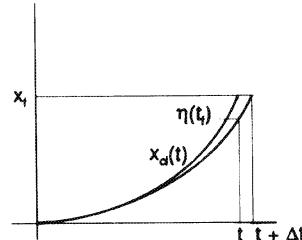


Figure 2.5. The upper trajectory takes time t while the lower takes $t + \Delta t$.

the right side must equal \mathcal{L} and not $-E$. The explanation requires Fig. 2.5 wherein we have set $x_i = t_i = 0$ for convenience.

The derivative we are computing is governed by the change in action of the *classical path* due to a change in travel by Δt holding the end points x_i and x_f fixed. From the figure it is clear that now the particle takes a different classical trajectory

$$x(t) = x_{\text{cl}}(t) + \eta(t) \quad \text{with} \quad \eta(0) = 0.$$

so that the total change in action comes from the difference in paths between $t=0$ and $t=t_f$ as well as the entire action due to the extra travel between t_f and $t_f + \Delta t$. Only the latter is given $\mathcal{L}\Delta t$. The correct answer is then

$$\begin{aligned} \delta S_{\text{cl}} &= \int_0^{t_f} \left[\frac{\partial \mathcal{L}}{\partial x} \eta(t) + \frac{\partial \mathcal{L}}{\partial \dot{x}} \dot{\eta}(t) \right] dt + \mathcal{L}(t_f) \Delta t \\ &= \int_0^{t_f} \left(-\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{x}} + \frac{\partial \mathcal{L}}{\partial x} \right)_{x_{\text{cl}}} \eta(t) dt + \int_0^{t_f} \frac{d}{dt} \left[\frac{\partial \mathcal{L}}{\partial \dot{x}} \eta(t) \right] dt + \mathcal{L}(t_f) \Delta t \\ &= 0 + \left. \frac{\partial \mathcal{L}}{\partial \dot{x}} \eta(t) \right|_{t_f} + \mathcal{L}(t_f) \Delta t. \end{aligned}$$

It is clear from the figure that $\eta(t_f) = -\dot{\eta}(t_f) \Delta t$ so that

$$\delta S = \left[-\frac{\partial \mathcal{L}}{\partial \dot{x}} \dot{x} + \mathcal{L} \right]_{t_f} \Delta t = -\mathcal{H}(t_f) \Delta t$$

from which the result follows.

Exercise 2.8.6. Show that $\partial S_{\text{cl}} / \partial x_f = p(t_f)$.

Exercise 2.8.7. Consider the harmonic oscillator, for which the general solution is

$$x(t) = A \cos \omega t + B \sin \omega t.$$

106

CHAPTER 2

Express the energy in terms of A and B and note that it does not depend on time. Now choose A and B such that $x(0)=x_1$ and $x(T)=x_2$. Write down the energy in terms of x_1 , x_2 , and T . Show that the action for the trajectory connecting x_1 and x_2 is

$$S_{\text{cl}}(x_1, x_2, T) = \frac{m\omega}{2 \sin \omega T} [(x_1^2 + x_2^2) \cos \omega T - 2x_1 x_2].$$

Verify that $\partial S_{\text{cl}} / \partial T = -E$.

3

All Is Not Well with Classical Mechanics

It was mentioned in the Prelude that as we keep expanding our domain of observations we must constantly check to see if the existing laws of physics continue to explain the new phenomena, and that, if they do not, we must try to find new laws that do. In this chapter you will get acquainted with experiments that betray the inadequacy of the classical scheme. The experiments to be described were never performed exactly as described here, but they contain the essential features of the actual experiments that were performed (in the first quarter of this century) with none of their inessential complications.

3.1. Particles and Waves in Classical Physics

There exist in classical physics two distinct entities: particles and waves. We have studied the particles in some detail in the last chapter and may summarize their essential features as follows. Particles are localized bundles of energy and momentum. They are described at any instant by the state parameters q and \dot{q} (or q and p). These parameters evolve in time according to some equations of motion. Given the initial values $q(t_i)$ and $\dot{q}(t_i)$ at time t_i , the trajectory $q(t)$ may be deduced for all future times from the equations of motion. A wave, in contrast, is a disturbance spread over space. It is described by a wave function $\psi(\mathbf{r}, t)$ which characterizes the disturbance at the point \mathbf{r} at time t .

In the case of sound waves, ψ is the excess air pressure above the normal, while in the case of electromagnetic waves, ψ can be any component of the electric field vector \mathbf{E} . The analogs of q and \dot{q} for a wave are ψ and $\dot{\psi}$ at each point \mathbf{r} , assuming ψ obeys a second-order wave equation in time, such as

$$\nabla^2 \psi = \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2}$$

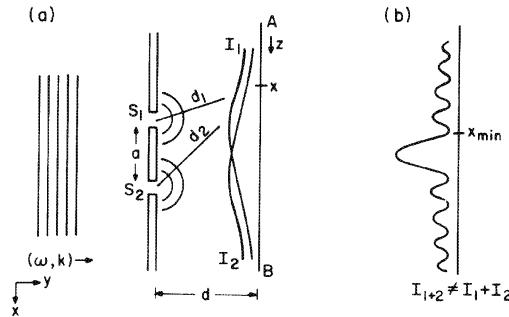


Figure 3.1. (a) When a wave $\psi = e^{i(ky - \omega t)}$ is incident on the screen with either slit S_1 or S_2 open, the intensity patterns I_1 and I_2 , respectively, are measured by the row of detectors on AB . (b) With both slits open, the pattern I_{1+2} is observed. Note that $I_{1+2} \neq I_1 + I_2$. This is called interference.

which describes waves propagating at the speed of light, c . Given $\psi(\mathbf{r}, 0)$ and $\dot{\psi}(\mathbf{r}, 0)$ one can get the wave function $\psi(\mathbf{r}, t)$ for all future times by solving the wave equation.

Of special interest to us are waves that are periodic in space and time, called *plane waves*. In one dimension, the plane wave may be written as

$$\psi(x, t) = A \exp\left[i\left(\frac{2\pi}{\lambda}x - \frac{2\pi}{T}t\right)\right] \equiv A \exp[i\phi] \quad (3.1.1)$$

At some given time t , the wave is periodic in space with a period λ , called its *wavelength*, and likewise at a given point x , it is periodic in time, repeating itself every T seconds, T being called the *time period*. We will often use, instead of λ and T , the related quantities $k = 2\pi/\lambda$ called the *wave number* and $\omega = 2\pi/T$ called the (*angular*) *frequency*. In terms of the phase ϕ in Eq. (3.1.1), k measures the phase change per unit length at any fixed time t , while ω measures the phase change per unit time at any fixed point x . This wave travels at a speed $v = \omega/k$. To check this claim, note that if we start out at a point where $\phi = 0$ and move along x at a rate $x = (\omega/k)t$, ϕ remains zero. The overall scale A up front is called the *amplitude*. For any wave, the intensity is defined to be $I = |\psi|^2$. For a plane wave this is a constant equal to $|A|^2$. If ψ describes an electromagnetic wave, the intensity is a measure of the energy and momentum carried by the wave. [Since the electromagnetic field is real, only the real part of ψ describes it. However, time averages of the energy and momentum flow are still proportional to the intensity (as defined above) in the case of plane waves.]

Plane waves in three dimension are written as

$$\psi(\mathbf{r}, t) = A e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}, \quad \omega = |\mathbf{k}|v \quad (3.1.2)$$

where each component k_i gives the phase changes per unit length along the i th axis. One calls \mathbf{k} the *wave vector*.[‡]

3.2. An Experiment with Waves and Particles (Classical)

Waves exhibit a phenomenon called *interference*, which is peculiar to them and is not exhibited by particles described by classical mechanics. This phenomenon is illustrated by the following experiment (Fig. 3.1a). Let a wave $\psi = A e^{i(ky - \omega t)}$ be

[‡] Unfortunately we also use \mathbf{k} to denote the unit vector along the z axis. It should be clear from the context what it stands for.

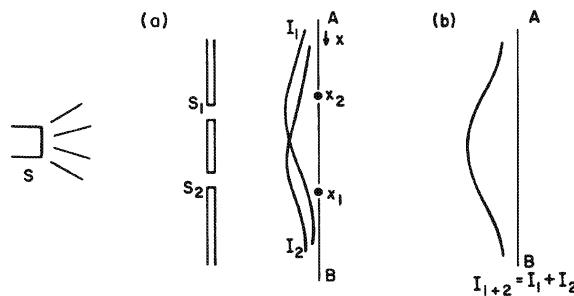


Figure 3.2. (a) Intensity pattern when S_1 or S_2 is open, due to a beam of incident particles. (b) The pattern with both slits open according to classical mechanics ($I_{1+2} = I_1 + I_2$).

incident normally on a screen with slits S_1 and S_2 , which are a distance a apart. At a distance d parallel to it is a row of detectors that measures the intensity as a function of the position x measured along AB .

If we first keep only S_1 open, the incident wave will come out of S_1 and propagate radially outward. One may think of S_1 as the virtual source of this wave ψ_1 , which has the same frequency and wavelength as the incident wave. The intensity pattern $I_1 = |\psi_1|^2$ is registered by the detectors. Similarly if S_2 is open instead of S_1 , the wave ψ_2 produces the pattern $I_2 = |\psi_2|^2$. In both cases the arrival of energy at the detectors is a smooth function of x and t .

Now if both S_1 and S_2 are opened, both waves ψ_1 and ψ_2 are present and produce an intensity pattern $I_{1+2} = |\psi_1 + \psi_2|^2$.

The interesting thing is that $I_{1+2} \neq I_1 + I_2$, but rather the interference pattern shown in Fig. 3.1b. The ups and downs are due to the fact that the waves ψ_1 and ψ_2 have to travel different distances d_1 and d_2 to arrive at some given x (see Fig. 3.1a) and thus are not always in step. In particular, the maxima correspond to the case $d_2 - d_1 = n\lambda$ (n is an integer), when the waves arrive exactly in step, and the minima correspond to the case $d_2 - d_1 = (2n+1)\lambda/2$, when the waves are exactly out of step. In terms of the phases ϕ_1 and ϕ_2 , $\phi_2(x) - \phi_1(x) = 2n\pi$ at a maximum and $\phi_2(x) - \phi_1(x) = (2n+1)\pi$ at a minimum. One can easily show that the spacing Δx between two adjacent maxima is $\Delta x = \lambda d/a$.

The feature to take special note of is that if x_{\min} is an interference minimum, there is more energy flowing into x_{\min} with just one slit open than with both. In other words, the opening of an extra slit can actually reduce the energy flow into x_{\min} .

Consider next the experiment with particles (Fig. 3.2a). The source of the incident plane waves is replaced by a source of particles that shoots them toward the screen with varying directions but fixed energy. Let the line AB be filled with an array of particle detectors. Let us define the intensity $I(x)$ to be the number of particles arriving per second at any given x . The patterns with S_1 or S_2 open are shown in (Fig. 3.2a). These look very much like the corresponding patterns for the wave. The only difference will be that the particles arrive not continuously, but in a staccato fashion, each particle triggering a counter at some single point x at the time of arrival. Although this fact may be obscured if the beam is dense, it can be easily detected as the incident flux is reduced.

What if both S_1 and S_2 are opened? Classical mechanics has an unambiguous prediction: $I_{1+2} = I_1 + I_2$. The reasoning is as follows: each particle travels along a definite trajectory that passes via S_1 or S_2 to the destination x . To a particle headed

for S_1 , it is immaterial whether S_2 is open or closed. Being localized in space it has no way of even knowing if S_2 is open or closed, and thus cannot respond to it in any way. Thus the number coming via S_1 to x is independent of whether S_2 is open or not and vice versa. It follows that $I_{1+2} = I_1 + I_2$ (Fig. 3.2b).

The following objection may be raised: although particles heading for S_1 are not aware that S_2 is open, they certainly can be deflected by those coming out of S_2 , if, for instance, the former are heading for x_1 and the latter for x_2 (see Fig. 3.1a).

This objection can be silenced by sending in one particle at a time. A given particle will of course not produce a pattern like I_1 or I_2 by itself, it will go to some point x . If, however, we make a histogram, the envelope of this histogram, after many counts, will define the smooth functions I_1 , I_2 , and I_{1+2} . Now the conclusion $I_{1+2} = I_1 + I_2$ is inevitable.

This is what classical physics predicts particles and waves will do in the double-slit experiment.

3.3. The Double-Slit Experiment with Light

Consider now what happens when we perform the following experiment to check the classical physics notion that light is an electromagnetic wave phenomenon.

We set up the double slit as in Fig. 3.1a, with a row of light-sensitive meters along AB and send a beam $\psi = A e^{i(ky - \omega t)}$ in a direction perpendicular to the screen. (Strictly speaking, the electromagnetic wave must be characterized by giving the orientation of the \mathbf{E} and \mathbf{B} vectors in addition to ω and k . However, for a plane wave, \mathbf{B} is uniquely fixed by \mathbf{E} . If we further assume \mathbf{E} is polarized perpendicular to the page, this polarization is unaffected by the double slit. We can therefore suppress the explicit reference to this constant vector and represent the field as a scalar function ψ .) We find that with the slits open one at a time we get patterns I_1 and I_2 , and with both slits open we get the interference pattern I_{1+2} as in Figs. 3.1a and 3.1b. (The interference pattern is of course what convinced classical physicists that light was a wave phenomenon.) The energy arrives at the detectors smoothly and continuously as befitting a wave.

Say we repeat the experiment with a change that is expected (in classical physics) to produce no qualitative effects. We start with S_1 open and cut down the intensity. A very strange thing happens. We find that the energy is not arriving continuously, but in sudden bursts, a burst here, a burst there, etc. We now cut down the intensity further so that only one detector gets activated at a given time and there is enough of a gap, say a millisecond, between counts. As each burst occurs at some x , we record it and plot a histogram. With enough data, the envelope of the histogram becomes, of course, the pattern I_1 . We have made an important discovery: light energy is not continuous—it comes in bundles. This discrete nature is obscured in intense beams, for the bundles come in so fast and all over the line AB , that the energy flow seems continuous in space and time.

We pursue our study of these bundles, called photons, in some detail and find the following properties:

1. Each bundle carries the same energy E .
2. Each bundle carries the same momentum p .

3. $E=pc$. From the famous equation $E^2=p^2c^2+m^2c^4$, we deduce that these bundles are particles of zero mass.

4. If we vary the frequency of the light source we discover that

$$E=\hbar\omega \quad (3.3.1)$$

$$p=\hbar k \quad (3.3.2)$$

where $\hbar=h/2\pi$ is a constant. The constant h is called *Planck's constant*, and has the dimensions of erg sec, which is the same as that of action and angular momentum. Its value is

$$\frac{h}{2\pi}=\hbar\simeq 10^{-27} \text{ erg sec} \quad (3.3.3)$$

For those interested in history, the actual experiment that revealed the granular nature of light is called the *photoelectric effect*. The correct explanation of this experiment, in terms of photons, was given by Einstein in 1905.

That light is made of particles will, of course, surprise classical physicists but will not imply the end of classical physics, for physicists are used to the idea that phenomena that seem continuous at first sight may in reality be discrete. They will cheerfully plunge into the study of the dynamics of the photons, trying to find the equations of motion for its trajectory and so on. What really undermines classical physics is the fact that if we now open both slits, still keeping the intensity so low that only one photon is in the experimental region at a given time, and watch the histogram take shape, we won't find that I_{1+2} equals $I_1 + I_2$ as would be expected of particles, but is instead an interference pattern characteristic of wave number k . This result completely rules out the possibility that photons move in well-defined trajectories like the particles of classical mechanics—for if this were true, a photon going in via S_1 should be insensitive to whether S_2 is open or not (and vice versa), and the result $I_{1+2}=I_1 + I_2$ is inescapable! To say this another way, consider a point x_{\min} which is an interference minimum. More photons arrive here with either S_1 or S_2 open than with both open. If photons followed definite trajectories, it is incomprehensible how opening an extra pathway can *reduce* the number coming to x_{\min} . Since we are doing the experiment with one photon at a time, one cannot even raise the improbable hypothesis that photons coming out of S_1 collide with those coming out of S_2 to modify (miraculously) the smooth pattern $I_1 + I_2$ into the wiggly interference pattern.

From these facts Born drew the following conclusion: with *each* photon is associated a wave ψ , called the *probability amplitude* or simply *amplitude*, whose modulus squared $|\psi(x)|^2$ gives the probability of finding the particle at x . [Strictly speaking, we must not refer to $|\psi(x)|^2$ as the probability for a given x , but rather as the probability density at x since x is a continuous variable. These subtleties can, however, wait.] The entire experiment may be understood in terms of this hypothesis as follows. Every incoming photon of energy E and momentum p has a wave function ψ associated with it, which is a plane wave with $\omega=E/\hbar$ and $k=p/\hbar$. This wave interferes with itself and forms the oscillating pattern $|\psi(x)|^2$ along AB , which gives

the probability that the given photon will arrive at x . A given photon of course arrives at some definite x and does not reveal the probability distribution. If, however, we wait till several photons, all described by the same ψ , have arrived, the number at any x will become proportional to the probability function $|\psi(x)|^2$. Likewise, if an intense (macroscopic) monochromatic beam is incident, many photons, all described by the same wave and hence the same probability distribution, arrive at the same time and all along the line AB . The intensity distribution then assumes the shape of the probability distribution right away and the energy flow seems continuous and in agreement with the predictions of classical electromagnetic theory.

The main point to note, besides the probability interpretation, is that a wave is associated not with a beam of photons, but with *each* photon. If the beam is monochromatic, every photon is given by the same ψ and the same probability distribution. A large ensemble of such photons will reproduce the phenomena expected of a classical electromagnetic wave ψ and the probabilistic aspect will be hidden.

3.4. Matter Waves (de Broglie Waves)

That light, which one thought was a pure wave phenomenon, should consist of photons, prompted de Broglie to conjecture that entities like the electron, generally believed to be particles, should exhibit wavelike behavior. More specifically, he conjectured, in analogy with photons, that particles of momentum p will produce an interference pattern corresponding to a wave number $k = p/\hbar$ in the double-slit experiment. This prediction was verified for electrons by Davisson and Germer, shortly thereafter. It is now widely accepted that all particles are described by probability amplitudes $\psi(x)$, and that the assumption that they move in definite trajectories is ruled out by experiment.

But what about common sense, which says that billiard balls and baseballs travel along definite trajectories? How did classical mechanics survive for three centuries? The answer is that the wave nature of matter is not apparent for macroscopic phenomena since \hbar is so small. The precise meaning of this explanation will become clear only after we fully master quantum mechanics. Nonetheless, the following example should be instructive. Suppose we do the double-slit experiment with pellets of mass 1 g, moving at 1 cm/sec. The wavelength associated with these particles is

$$\lambda = \frac{2\pi}{k} = \frac{\hbar}{p} \simeq 10^{-26} \text{ cm}$$

which is 10^{-13} times smaller than the radius of the proton! For any reasonable values of the parameters a and d (see Fig. 3.1b), the interference pattern would be so dense in x that our instruments will only measure the smooth average, which will obey $I_{1+2} = I_1 + I_2$ as predicted classically.

3.5. Conclusions

The main objective of this chapter was to expose the inadequacy of classical physics in explaining certain phenomena and, incidentally, to get a glimpse of what

the new (quantum) physics ought to look like. We found that entities such as the electron are particles in the classical sense in that when detected they seem to carry all their energy, momentum, charge, etc. in localized form; and at the same time they are not particlelike in that assuming they move along definite trajectories leads to conflict with experiment. It appears that each particle has associated with it a wave function $\psi(x, t)$, such that $|\psi(x, t)|^2$ gives the probability of finding it at a point x at time t . This is called *wave-particle duality*.

The dynamics of the particle is then the dynamics of this function $\psi(x, t)$ or, if we think of functions as vectors in an infinite-dimensional space, of the ket $|\psi(t)\rangle$. In the next chapter the postulates of quantum theory will define the dynamics in terms of $|\psi(t)\rangle$. The postulates, which specify what sort of information is contained in $|\psi(t)\rangle$ and how $|\psi(t)\rangle$ evolves with time, summarize the results of the double-slit experiment and *many others not mentioned here*. The double-slit experiment was described here to expose the inadequacy of classical physics and not to summarize the entire body of experimental results from which all the postulates could be inferred. Fortunately, the double-slit experiment contains most of the central features of the theory, so that when the postulates are encountered in the next chapter, they will appear highly plausible.

4

The Postulates—a General Discussion

Having acquired the necessary mathematical training and physical motivation, you are now ready to get acquainted with the postulates of quantum mechanics. In this chapter the postulates will be stated and discussed in broad terms to bring out the essential features of quantum theory. The subsequent chapters will simply be applications of these postulates to the solution of a variety of physically interesting problems. Despite your preparation you may still find the postulates somewhat abstract and mystifying on this first encounter. These feelings will, however, disappear after you have worked with the subject for some time.

4.1. The Postulates[‡]

The following are the postulates of nonrelativistic quantum mechanics. We consider first a system with one degree of freedom, namely, a single particle in one space dimension. The straightforward generalization to more particles and higher dimensions will be discussed towards the end of the chapter. In what follows, the quantum postulates are accompanied by their classical counterparts (in the Hamiltonian formalism) to provide some perspective.

- | Classical Mechanics | Quantum Mechanics |
|---|---|
| I. The state of a particle at any given time is specified by the two variables $x(t)$ and $p(t)$, i.e., as a point in a two-dimensional phase space. | I. The state of the particle is represented by a vector $ \psi(t)\rangle$ in a Hilbert space. |
| II. Every dynamical variable ω is a function of x and p : $\omega = \omega(x, p)$. | II. The independent variables x and p of classical mechanics are represented |

[‡] Recall the discussion in the Preface regarding the sense in which the word is used here.

by Hermitian operators X and P with the following matrix elements in the eigenbasis of X [‡]

$$\begin{aligned}\langle x|X|x'\rangle &= x\delta(x-x') \\ \langle x|P|x'\rangle &= -i\hbar\delta'(x-x')\end{aligned}$$

The operators corresponding to dependent variables $\omega(x, p)$ are given Hermitian operators

$$\Omega(X, P) = \omega(x \rightarrow X, p \rightarrow P)\S$$

III. If the particle is in a state given by x and p , the measurement^{||} of the variable ω will yield a value $\omega(x, p)$. The state will remain unaffected.

III. If the particle is in a state $|\psi\rangle$, measurement^{||} of the variable (corresponding to) Ω will yield one of the eigenvalues ω with probability $P(\omega) \propto |\langle \omega | \psi \rangle|^2$. The state of the system will change from $|\psi\rangle$ to $|\omega\rangle$ as a result of the measurement.

IV. The state variables change with time according to Hamilton's equations:

$$\dot{x} = \frac{\partial \mathcal{H}}{\partial p}$$

$$\dot{p} = -\frac{\partial \mathcal{H}}{\partial x}$$

IV. The state vector $|\psi(t)\rangle$ obeys the *Schrödinger equation*

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = H|\psi(t)\rangle$$

where $H(X, P) = \mathcal{H}(x \rightarrow X, p \rightarrow P)$ is the quantum Hamiltonian operator and \mathcal{H} is the Hamiltonian for the corresponding classical problem.

4.2. Discussion of Postulates I–III

The postulates (of classical and quantum mechanics) fall naturally into two sets: the first three, which tell us how the system is depicted at a given time, and the last, which specifies how this picture changes with time. We will confine our attention to the first three postulates in this section, leaving the fourth for the next.

The first postulate states that a particle is described by a ket $|\psi\rangle$ in a Hilbert space which, you will recall, contains *proper vectors* normalizable to unity as well as

[‡] Note that the X operator is the same one discussed at length in Section 1.10. Likewise $P = \hbar K$, where K was also discussed therein. You may wish to go over that section now to refresh your memory.

[§] By this we mean that Ω is the same function of X and P as ω is of x and p .

^{||} That is, in an ideal experiment consistent with the theory. It is assumed you are familiar with the ideal classical measurement which can determine the state of the system without disturbing it in any way. A discussion of ideal quantum measurements follows.

improper vectors, normalizable only to the Dirac delta functions.[‡] Now, a ket in such a space has in general an infinite number of components in a given basis. One wonders why a particle, which had only two independent degrees of freedom, x and p , in classical mechanics, now needs to be specified by an infinite number of variables. What do these variables tell us about the particle? To understand this we must go on to the next two postulates, which answer exactly this question. For the present let us note that the double-slit experiment has already hinted to us that a particle such as the electron needs to be described by a wave function $\psi(x)$. We have seen in Section 1.10 that a function $f(x)$ may be viewed as a ket $|f\rangle$ in a Hilbert space. The ket $|\psi\rangle$ of quantum mechanics is none other than the vector representing the probability amplitude $\psi(x)$ introduced in the double-slit experiment.

When we say that $|\psi\rangle$ is an element of a vector space we mean that if $|\psi\rangle$ and $|\psi'\rangle$ represent possible states of a particle so does $\alpha|\psi\rangle + \beta|\psi'\rangle$. This is called the *principle of superposition*. The principle by itself is not so new: we know in classical physics, for example, that if $f(x)$ and $g(x)$ [with $f(0)=f(L)=g(0)=g(L)=0$] are two possible displacements of a string, so is the superposition $\alpha f(x) + \beta g(x)$. What is new is the interpretation of the superposed state $\alpha|\psi\rangle + \beta|\psi'\rangle$. In the case of the string, the state $\alpha f + \beta g$ has very different attributes from the states f and g : it will look different, have a different amount of stored elastic energy, and so on. In quantum theory, on the other hand, the state $\alpha|\psi\rangle + \beta|\psi'\rangle$ will, loosely speaking, have attributes that sometimes resemble that of $|\psi\rangle$ and at other times those of $|\psi'\rangle$. There is, however, no need to speak loosely, since we have postulates II and III to tell us exactly how the state vector $|\psi\rangle$ is to be interpreted in quantum theory. Let us find out.

In classical mechanics when a state (x, p) is given, one can say that any dynamical variable ω has a value $\omega(x, p)$, in the sense that if the variable is measured the result $\omega(x, p)$ will obtain. What is the analogous statement one can make in quantum mechanics given that the particle is in a state $|\psi\rangle$? The answer is provided by Postulates II and III, in terms of the following steps:

Step 1. Construct the corresponding quantum operator $\Omega = \omega(x \rightarrow X, p \rightarrow P)$, where X and P are the operators defined in postulate II.

Step 2. Find the orthonormal eigenvectors $|\omega_i\rangle$ and eigenvalues ω_i of Ω .

Step 3. Expand $|\psi\rangle$ in this basis:

$$|\psi\rangle = \sum_i |\omega_i\rangle \langle \omega_i | \psi \rangle$$

Step 4. The probability $P(\omega)$ that the result ω will obtain is proportional to the modulus squared of the projection of $|\psi\rangle$ along the eigenvector $|\omega\rangle$, that is $P(\omega) \propto |\langle \omega | \psi \rangle|^2$. In terms of the projection operator $\mathbb{P}_\omega = |\omega\rangle \langle \omega|$, $P(\omega) \propto |\langle \omega | \psi \rangle|^2 = \langle \psi | \omega \rangle \langle \omega | \psi \rangle = \langle \psi | \mathbb{P}_\omega | \psi \rangle = \langle \psi | \mathbb{P}_\omega \mathbb{P}_\omega | \psi \rangle = \langle \mathbb{P}_\omega \psi | \mathbb{P}_\omega \psi \rangle$.

There is a tremendous amount of information contained in these steps. Let us note, for the present, the following salient points.

[‡] The status of the two classes will be clarified later in this chapter.

(1) The theory makes only probabilistic predictions for the result of a measurement of Ω . Further, it assigns (relative) probabilities only for obtaining some eigenvalue ω of Ω . *Thus the only possible values of Ω are its eigenvalues.* Since postulate II demands that Ω be Hermitian, these eigenvalues are all real.

(2) Since we are told that $P(\omega_i) \propto |\langle \omega_i | \psi \rangle|^2$, the quantity $|\langle \omega_i | \psi \rangle|^2$ is only the relative probability. To get the absolute probability, we divide $|\langle \omega_i | \psi \rangle|^2$ by the sum of all relative probabilities:

$$P(\omega_i) = \frac{|\langle \omega_i | \psi \rangle|^2}{\sum_j |\langle \omega_j | \psi \rangle|^2} = \frac{|\langle \omega_i | \psi \rangle|^2}{\langle \psi | \psi \rangle} \quad (4.2.1)$$

It is clear that if we had started with a normalized state

$$|\psi'\rangle = \frac{|\psi\rangle}{\langle \psi | \psi \rangle^{1/2}}$$

we would have had

$$P(\omega_i) = |\langle \omega_i | \psi' \rangle|^2 \quad (4.2.2)$$

If $|\psi\rangle$ is a proper vector, such a rescaling is possible and will be assumed hereafter. The probability interpretation breaks down if $|\psi\rangle$ happens to be one of the improper vectors in the space, for in this case $\langle \psi | \psi \rangle = \delta(0)$ is the only sensible normalization. The status of such vectors will be explained in Example 4.2.2 below.

Note that the condition $\langle \psi | \psi \rangle = 1$ is a matter of convenience and not a physical restriction on the proper vectors. (In fact the set of all normalized vectors does not even form a vector space. If $|\psi\rangle$ and $|\psi'\rangle$ are normalized, then an arbitrary linear combination, $\alpha|\psi\rangle + \beta|\psi'\rangle$, is not.)

Note that the relative probability distributions corresponding to the states $|\psi\rangle$ and $\alpha|\psi\rangle$, when they are renormalized to unity, reduce to the same absolute probability distribution. Thus, corresponding to each physical state, there exists not one vector, but a *ray* or “direction” in Hilbert space. When we speak of the state of the particle, we usually mean the ket $|\psi\rangle$ with unit norm. Even with the condition $\langle \psi | \psi \rangle = 1$, we have the freedom to multiply the ket by a number of the form $e^{i\theta}$ without changing the physical state. This freedom will be exploited at times to make the components of $|\psi\rangle$ in some basis come out real.

(3) If $|\psi\rangle$ is an eigenstate $|\omega_i\rangle$, the measurement of Ω is guaranteed to yield the result ω_i . A particle in such a state may be said to have a value ω_i for Ω in the classical sense.

(4) When two states $|\omega_1\rangle$ and $|\omega_2\rangle$ are superposed to form a (normalized) state, such as

$$|\psi\rangle = \frac{\alpha|\omega_1\rangle + \beta|\omega_2\rangle}{(|\alpha|^2 + |\beta|^2)^{1/2}}$$

one gets the state, which upon measurement of Ω , can yield either ω_1 or ω_2 with probabilities $|\alpha|^2/(|\alpha|^2 + |\beta|^2)$ and $|\beta|^2/(|\alpha|^2 + |\beta|^2)$, respectively. This is the peculiar

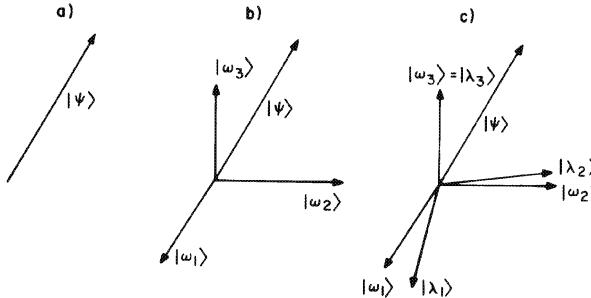


Figure 4.1. (a) The normalized ket in $\mathbb{V}^3(R)$ representing the state of the particle. (b) The Ω basis, $|\omega_1\rangle$, $|\omega_2\rangle$, and $|\omega_3\rangle$. (c) The Ω and the Λ bases. To get the statistical information on a variable, we find the eigenvectors of the corresponding operator $|\psi\rangle$ on that basis.

consequence of the superposition principle in quantum theory, referred to earlier. It has no analog in classical mechanics. For example, if a dynamical variable of the string in the state $\alpha f + \beta g$ is measured, one does not expect to get the value corresponding to f some of the time and that corresponding to g the rest of the time; instead, one expects a unique value generally distinct from both. Likewise, the functions f and αf (α real) describe two distinct configurations of the string and are not physically equivalent.

(5) When one wants information about another variable Λ , one repeats the whole process, finding the eigenvectors $|\lambda_i\rangle$ and the eigenvalues λ_i . Then

$$P(\lambda) = |\langle \lambda | \psi \rangle|^2$$

The bases of Ω and Λ will of course be different in general. In summary, we have a single ket $|\psi\rangle$ representing the state of the particle in Hilbert space, and it contains the statistical prediction for all observables. To extract this information for any observable, we must determine the eigenbasis of the corresponding operator and find the projection of $|\psi\rangle$ along all its eigenkets.

(6) As our interest switches from one variable Ω , to another, Λ , so does our interest go from the kets $|\omega\rangle$, to the kets $|\lambda\rangle$. There is, however, no need to change the basis each time. Suppose for example we are working in the Ω basis in which

$$|\psi\rangle = \sum_i |\omega_i\rangle \langle \omega_i | \psi \rangle$$

and $P(\omega_i) = |\langle \omega_i | \psi \rangle|^2$. If we want $P(\lambda_i)$ we take the operator Λ (which is some given matrix with elements $\Lambda_{ij} = \langle \omega_i | \Lambda | \omega_j \rangle$); find its eigenvectors $|\lambda_i\rangle$ (which are column vectors with components $\langle \omega_j | \lambda_i \rangle$), and take the inner product $\langle \lambda_i | \psi \rangle$ in this basis:

$$\langle \lambda_i | \psi \rangle = \sum_j \langle \lambda_i | \omega_j \rangle \langle \omega_j | \psi \rangle$$

Example 4.2.1. Consider the following example from a fictitious Hilbert space $\mathbb{V}^3(R)$ (Fig. 4.1). In Fig. 4.1a we have the normalized state $|\psi\rangle$, with no reference

to any basis. To get predictions on Ω , we find its eigenbasis and express the state vector $|\psi\rangle$ in terms of the orthonormal eigenvectors $|\omega_1\rangle$, $|\omega_2\rangle$, and $|\omega_3\rangle$ (Fig. 4.1b). Let us suppose

$$|\omega\rangle = \frac{1}{2}|\omega_1\rangle + \frac{1}{2}|\omega_2\rangle + \frac{1}{2^{1/2}}|\omega_3\rangle$$

This means that the values ω_1 , ω_2 , and ω_3 are expected with probabilities $\frac{1}{4}$, $\frac{1}{4}$, and $\frac{1}{2}$, respectively, and other values of ω are impossible. If instead $|\psi\rangle$ were some eigenvector, say $|\omega_1\rangle$, then the result ω_1 would obtain with unit probability. Only a particle in a state $|\psi\rangle=|\omega_i\rangle$ has a well-defined value of Ω in the classical sense. If we want $P(\lambda_i)$ we construct the basis $|\lambda_1\rangle$, $|\lambda_2\rangle$, and $|\lambda_3\rangle$, which can in general be distinct from the Ω basis. In our example (Fig. 4.1c) there is just one common eigenvector $|\omega_3\rangle=|\lambda_3\rangle$. \square

Returning to our main discussion, there are a few complications that could arise as one tries to carry out the steps 1–4. We discuss below the major ones and how they are to be surmounted.

Complication 1: The Recipe $\Omega=\omega(x\rightarrow X, p\rightarrow P)$ Is Ambiguous. If, for example, $\omega=xp$, we don't know if $\Omega=XP$ or PX since $xp=px$ classically. There is no universal recipe for resolving such ambiguities. In the present case, the rule is to use the symmetric sum: $\Omega=(XP+PX)/2$. Notice incidentally that symmetrization also renders Ω Hermitian. Symmetrization is the answer as long as Ω does not involve products of two or more powers of X with two or more powers of P . If it does, only experiment can decide the correct prescription. We will not encounter such cases in this book.

Complication 2: The Operator Ω Is Degenerate. Let us say $\omega_1=\omega_2=\omega$. What is $P(\omega)$ in this case? We select some orthonormal basis $|\omega, 1\rangle$ and $|\omega, 2\rangle$ in the eigenspace \mathbb{V}_ω with eigenvalue ω . Then

$$P(\omega)=|\langle\omega, 1|\psi\rangle|^2+|\langle\omega, 2|\psi\rangle|^2$$

which is the modulus squared of the projection of $|\psi\rangle$ in the degenerate eigenspace. This is the result we will get if we assume that ω_1 and ω_2 are infinitesimally distinct and ask for $P(\omega_1)$ or $P(\omega_2)$. In terms of the projection operator for the *eigenspace*,

$$\mathbb{P}_\omega=|\omega, 1\rangle\langle\omega, 1|+|\omega, 2\rangle\langle\omega, 2| \quad (4.2.3a)$$

we have

$$P(\omega)=\langle\psi|\mathbb{P}_\omega|\psi\rangle=\langle\mathbb{P}_\omega\psi|\mathbb{P}_\omega\psi\rangle \quad (4.2.3b)$$

In general, one can replace in Postulate III

$$P(\omega)\propto\langle\psi|\mathbb{P}_\omega|\psi\rangle$$

where \mathbb{P}_ω is the projection operator for the eigenspace with eigenvalue ω . Then postulate III as stated originally would become a special case in which there is no degeneracy and each eigenspace is simply an eigenvector.

In our example from $\mathbb{V}^3(R)$, if $\omega_1 = \omega_2 = \omega$ (Fig. 4.1b) then $P(\omega)$ is the square of the component of $|\psi\rangle$ in the “ x y ” plane.

Complication 3: The Eigenvalue Spectrum of Ω Is Continuous. In this case one expands $|\psi\rangle$ as

$$|\psi\rangle = \int |\omega\rangle \langle \omega| \psi\rangle d\omega$$

One expects that as ω varies continuously, so will $\langle \omega | \psi \rangle$, that is to say, one expects $\langle \omega | \psi \rangle$ to be a smooth function $\psi(\omega)$. To visualize this function one introduces an auxiliary one-dimensional space, called the ω space, the points in which are labeled by the coordinate ω . In this space $\psi(\omega)$ will be a smooth function of ω and is called the *wave function in the ω space*. We are merely doing the converse of what we did in Section 1.10 wherein we started with a function $f(x)$ and tried to interpret it as the components of an infinite-dimensional ket $|\psi\rangle$ in the $|x\rangle$ basis. As far as the state vector $|\psi\rangle$ is concerned, there is just one space, the Hilbert space, in which it resides. The ω space, the λ space, etc. are auxiliary manifolds introduced for the purpose of visualizing the components of the infinite-dimensional vector $|\psi\rangle$ in the Ω basis, the Λ basis, and so on. The wave function $\psi(\omega)$ is also called the *probability amplitude* for finding the particle with $\Omega = \omega$.

Can we interpret $|\langle \omega | \psi \rangle|^2$ as the probability for finding the particle with a value ω for Ω ? No. Since the number of possible values for ω is infinite and the total probability is unity, each single value of ω can be assigned only an infinitesimal probability. One interprets $P(\omega) = |\langle \omega | \psi \rangle|^2$ to be the *probability density* at ω , by which one means that $P(\omega) d\omega$ is the probability of obtaining a result between ω and $\omega + d\omega$. This definition meets the requirement that the total probability be unity, since

$$\begin{aligned} \int P(\omega) d\omega &= \int |\langle \omega | \psi \rangle|^2 d\omega = \int \langle \psi | \omega \rangle \langle \omega | \psi \rangle d\omega \\ &= \langle \psi | I | \psi \rangle = \langle \psi | \psi \rangle = 1 \end{aligned} \quad (4.2.4)$$

If $\langle \psi | \psi \rangle = \delta(0)$ is the only sensible normalization possible, the state cannot be normalized to unity and $P(\omega)$ must be interpreted as the *relative probability density*. We will discuss such improper states later.

An important example of a continuous spectrum is that of X , the operator corresponding to the position x . The wave function in the X basis (or the x space), $\psi(x)$, is usually referred to as just the wave function, since the X basis is almost always what one uses. In our discussions in the last chapter, $|\psi(x)|^2$ was referred to as the probability for finding the particle *at a given x* , rather than as the probability density, in order to avoid getting into details. Now the time has come to become precise!

Earlier on we were wondering why it was that a classical particle defined by just two numbers x and p now needs to be described by a ket which has an infinite number of components. The answer is now clear. A classical particle has, at any given time, a definite position. One simply has to give this value of x in specifying the state. A quantum particle, on the other hand, can take on any value of x upon measurement and one must give the relative probabilities for *all possible outcomes*. This is part of the information contained in $\psi(x) = \langle x | \psi \rangle$, the components of $|\psi\rangle$ in the X basis. Of course, in the case of the classical particle, one needs also to specify the momentum p as well. In quantum theory one again gives the odds for getting different values of momenta, but one doesn't need a new vector for specifying this; the same ket $|\psi\rangle$ when expanded in terms of the eigenkets $|p\rangle$ of the momentum operator P gives the odds through the wave function in p space, $\psi(p) = \langle p | \psi \rangle$.

Complication 4: The Quantum Variable Ω Has No Classical Counterpart. Even “point” particles such as the electron are now known to carry “spin,” which is an internal angular momentum, that is to say, angular momentum unrelated to their motion through space. Since such a degree of freedom is absent in classical mechanics, our postulates do not tell us which operator is to describe this variable in quantum theory. As we will see in Chapter 14, the solution is provided by a combination of intuition and semi-classical reasoning. It is worth bearing in mind that no matter how diligently the postulates are constructed, they must often be supplemented by intuition and classical ideas.

Having discussed the four-step program for extracting statistical information from the state vector, we continue with our study of what else the postulates of quantum theory tell us.

Collapse of the State Vector

We now examine another aspect of postulate III, namely, that the measurement of the variable Ω changes the state vector, which is in general some superposition of the form

$$|\psi\rangle = \sum_{\omega} |\omega\rangle \langle \omega | \psi \rangle$$

into the eigenstate $|\omega\rangle$ corresponding to the eigenvalue ω obtained in the measurement. This phenomenon is called the *collapse or reduction of the state vector*.

Let us first note that any definitive statement about the impact of the measurement process presupposes that the measurement process is of a definite kind. For example, the classical mechanics maxim that any dynamical variable can be measured without changing the state of the particle, assumes that the measurement is an ideal measurement (consistent with the classical scheme). But one *can* think up *nonideal* measurements which *do* change the state; imagine trying to locate a chandelier in a dark room by waving a broom till one makes contact. What makes Postulate III profound is that the measurement process referred to there is an *ideal quantum measurement*, which in a sense is the best one can do. We now illustrate the notion of an ideal quantum measurement and the content of this postulate by an example.

Consider a particle in a momentum eigenstate $|p\rangle$. The postulate tells us that if the momentum in this state is measured we are assured a result p , and that the state will be the same after the measurement (since $|\psi\rangle = |p\rangle$ is already an eigenstate of the operator P in question). One way to measure the momentum of the particle is by *Compton scattering*, in which a photon of definite momentum bounces off the particle.

Let us assume the particle is forced to move along the x -axis and that we send in a right-moving photon of energy $\hbar\omega$ that bounces off the particle and returns as a left-moving photon of energy $\hbar\omega'$. (How do we know what the photon energies are? We assume we have atoms that are known to emit and absorb photons of any given energy.) Using momentum and energy conservation:

$$cp' = cp + \hbar(\omega + \omega')$$

$$E' = E + \hbar(\omega - \omega')$$

it is now possible from this data to reconstruct the initial and final momenta of the particle:

$$cp = -\frac{(\hbar\omega + \hbar\omega')}{2} + \sqrt{1 + \frac{m^2 c^4}{\hbar^2 \omega \omega'} \frac{\hbar\omega - \hbar\omega'}{2}}$$

$$cp' = \frac{(\hbar\omega + \hbar\omega')}{2} + \sqrt{1 + \frac{m^2 c^4}{\hbar^2 \omega \omega'} \frac{\hbar\omega - \hbar\omega'}{2}}$$

Solving for ω' and p' in terms of ω and p , one readily sees that for any choice of p , if $\omega \rightarrow 0$, then so does ω' . Thus one can always make the change in momentum $p' - p$ arbitrarily small. Hereafter, when we speak of a momentum measurement, this is what we will mean. We will also assume that to each dynamical variable there exists a corresponding ideal measurement. We will discuss, for example, the ideal position measurement, which, when conducted on a particle in state $|x\rangle$, will give the result x with unit probability and leave the state vector unchanged.

Suppose now that we measure the *position* of a particle in a *momentum* eigenstate $|p\rangle$. Since $|p\rangle$ is a sum of position eigenkets $|x\rangle$,

$$|p\rangle = \int |x\rangle \langle x| p \rangle dx$$

the measurement will force the system into some state $|x\rangle$. Thus even the *ideal* position measurement will change the state which is not a position eigenstate. Why does a position measurement alter the state $|p\rangle$, while momentum measurement does not? The answer is that an ideal position measurement uses photons of infinitely high momentum (as we will see) while an ideal momentum measurement uses photons of infinitesimally low momentum (as we have seen).

This then is the big difference between classical and quantum mechanics: an ideal measurement of any variable ω in classical mechanics leaves any state invariant,

whereas the ideal measurement of Ω in quantum mechanics leaves only the eigenstates of Ω invariant.

The effect of measurement may be represented schematically as follows:

$$|\psi\rangle \xrightarrow[\Omega \text{ measured, } \omega \text{ obtained}]{\quad} = \frac{\mathbb{P}_\omega |\psi\rangle}{\langle \mathbb{P}_\omega \psi | \mathbb{P}_\omega \psi \rangle^{1/2}}$$

where \mathbb{P}_ω is the projection operator associated with $|\omega\rangle$, and the state after measurement has been normalized. If ω is degenerate,

$$|\psi\rangle \rightarrow \frac{\mathbb{P}_\omega |\psi\rangle}{\langle \mathbb{P}_\omega \psi | \mathbb{P}_\omega \psi \rangle^{1/2}}$$

where \mathbb{P}_ω is the projection operator for the eigenspace \mathbb{V}_ω . Special note should be taken of the following point: if the initial state $|\psi\rangle$ were unknown, and the measurement yielded a degenerate eigenvalue ω , we could not say what the state was after the measurement, except that it was some state in the eigenspace with eigenvalue ω . On the other hand, if the initial state $|\psi\rangle$ were known, and the measurement yielded a degenerate value ω , the state after measurement is known to be $\mathbb{P}_\omega |\psi\rangle$ (up to normalization). Consider our example from $\mathbb{V}^3(\mathbb{R})$ (Fig. 4.1b). Say we had $\omega_1 = \omega_2 = \omega$. Let us use an orthonormal basis $|\omega, 1\rangle, |\omega, 2\rangle, |\omega_3\rangle$, where, as usual, the extra labels 1 and 2 are needed to distinguish the basis vectors in the degenerate eigenspace. If in this basis we know, for example, that

$$|\psi\rangle = \frac{1}{2}|\omega, 1\rangle + \frac{1}{2}|\omega, 2\rangle + (\frac{1}{2})^{1/2}|\omega_3\rangle$$

and the measurement gives a value ω , the normalized state after measurement is known to us to be

$$|\psi\rangle = 2^{-1/2}(|\omega, 1\rangle + |\omega, 2\rangle)$$

If, on the other hand, the initial state were unknown and a measurement gave a result ω , we could only say

$$|\psi\rangle = \frac{\alpha|\omega, 1\rangle + \beta|\omega, 2\rangle}{(\alpha^2 + \beta^2)^{1/2}}$$

where α and β are arbitrary real numbers.

Note that although we do not know what α and β are from the measurement, they are not arbitrary. In other words, the system had a well-defined state vector $|\psi\rangle$ before the measurement, though we did not know $|\psi\rangle$, and has a well-defined state vector $\mathbb{P}_\omega |\psi\rangle$ after the measurement, although all we know is that it lies within a subspace \mathbb{V}_ω .

How to Test Quantum Theory

One of the outstanding features of classical mechanics is that it makes fully deterministic predictions. It may predict for example that a particle leaving $x=x_i$ with momentum p_i in some potential $V(x)$ will arrive 2 seconds later at $x=x_f$ with momentum $p=p_f$. To test the prediction we release the particle at $x=x_i$ with $p=p_i$ at $t=0$ and wait at $x=x_f$ and see if the particle arrives there with $p=p_f$ at $t=2$ seconds.

Quantum theory, on the other hand, makes statistical predictions about a particle in a state $|\psi\rangle$ and claims that this state evolves in time according to Schrödinger's equation. To test these predictions we must be able to

- (1) Create particles in a well-defined state $|\psi\rangle$.
- (2) Check the probabilistic predictions at any time.

The collapse of the state vector provides us with a good way of preparing definite states: we begin with a particle in an arbitrary state $|\psi\rangle$ and measure a variable Ω . If we get a nondegenerate eigenvalue ω , we have in our hands the state $|\omega\rangle$. (If ω is degenerate, further measurement is needed. We are not ready to discuss this problem.) Notice how in quantum theory, measurement, instead of telling us what the system was doing *before* the measurement, tells us what it is doing just *after* the measurement. (Of course it does tell us that the original state had some projection on the state $|\omega\rangle$ obtained after measurement. But this information is nothing compared to the complete specifications of the state just *after* measurement.)

Anyway, assume we have prepared a state $|\omega\rangle$. If we measure some variable Λ , immediately thereafter, so that the state could not have changed from $|\omega\rangle$, and if say,

$$|\omega\rangle = \frac{1}{3^{1/2}} |\lambda_1\rangle + \left(\frac{2}{3}\right)^{1/2} |\lambda_2\rangle + 0 \cdot (\text{others})$$

the theory predicts that λ_1 and λ_2 will obtain with probabilities $1/3$ and $2/3$, respectively. If our measurement gives a λ_i , $i \neq 1, 2$ (or worse still a $\lambda \neq$ any eigenvalue!) that is the end of the theory. So let us assume we get one of the allowed values, say λ_1 . This is consistent with the theory but does not fully corroborate it, since the odds for λ_1 could have been $1/30$ instead of $1/3$ and we could still get λ_1 . Therefore, we must repeat the experiment many times. But we cannot repeat the experiment with *this* particle, since after the measurement the state of the particle is $|\lambda_1\rangle$. We must start afresh with another particle in $|\omega\rangle$. For this purpose we require a *quantum ensemble*, which consists of a large number N of particles *all in the same state* $|\omega\rangle$. If a measurement of Λ is made on every one of these particles, approximately $N/3$ will yield a value λ_1 and end up in the state $|\lambda_1\rangle$ while approximately $2N/3$ will yield a value λ_2 and end up in a state $|\lambda_2\rangle$. For sufficiently large N , the deviations from the fractions $1/3$ and $2/3$ will be negligible. The chief difference between a classical ensemble, of the type one encounters in, say, classical statistical mechanics, and the quantum ensemble referred to above, is the following. If in a classical ensemble of N particles $N/3$ gave a result λ_1 and $2N/3$ a result λ_2 , one can think of the ensemble as having contained $N/3$ particles with $\lambda = \lambda_1$ and the others with $\lambda = \lambda_2$ *before* the

measurement. In a quantum ensemble, on the other hand, every particle is assumed to be in the same state $|\omega\rangle$ prior to measurement (i.e., every particle is potentially capable of yielding *either* result λ_1 or λ_2). Only after that measurement are a third of them forced into the state $|\lambda_1\rangle$ and the rest into $|\lambda_2\rangle$.

Once we have an ensemble, we can measure any other variable and test the expectations of quantum theory. We can also prepare an ensemble, let it evolve in time, and study it at a future time to see if the final state is what the Schrödinger equation tells us it should be.

Example 4.2.2. An example of an ensemble being used to test quantum theory was encountered in the double-slit experiment, say with photons. A given photon of momentum p and energy E was expected to hit the detectors with a probability density given by the oscillating function $|\psi(x)|^2$. One could repeat the experiment N times, sending one such photon at a time to see if the final number distribution indeed was given by $|\psi(x)|^2$. One could equally well send in a macroscopic, monochromatic beam of light of frequency $\omega = E/\hbar$ and wave number $k = p/\hbar$, which consists of a large number of photons of energy E and momentum p . If one makes the assumption (correct to a high degree) that the photons are noninteracting, sending in the beam is equivalent to experimenting with the ensemble. In this case the intensity pattern will take the shape of the probability density $|\psi(x)|^2$, the instant the beam is turned on. \square

Example 4.2.3. The following example is provided to illustrate the distinction between the probabilistic descriptions of systems in classical mechanics and in quantum mechanics.

We choose as our classical system a six-faced die for which the probabilities $P(n)$ of obtaining a number n have been empirically determined. As our quantum system we take a particle in a state

$$|\psi\rangle = \sum_{i=1}^6 C_i |\omega_i\rangle$$

Suppose we close our eyes, toss the die, and cover it with a mug. Its statistical description has many analogies with the quantum description of the state $|\psi\rangle$:

- (1) The state of the die is described by a probability function $P(n)$ before the mug is lifted.
- (2) The only possible values of n are 1, 2, 3, 4, 5, and 6.
- (3) If the mug is lifted, and some value—say $n=3$ —is obtained, the function $P(n)$ collapses to δ_{n3} .
- (4) If an ensemble of N such dice are thrown, $NP(n)$ of them will give the result n (as $N \rightarrow \infty$).

The corresponding statements for the particle in the state $|\psi\rangle$ are no doubt known to you. Let us now examine some of the key differences between the statistical descriptions in the two cases.

(1) It is possible, *at least in principle*, to predict exactly which face of the die will be on top, given the mass of the die, its position, orientation, velocity, and angular velocity at the time of release, the viscosity of air, the elasticity of the table top, and so on. The statistical description is, however, the only possibility in the quantum case, *even in principle*.

(2) If the result $n=3$ was obtained upon lifting the mug, it is consistent to assume that the die was in such a state *even prior to measurement*. In the quantum case, however, the state after measurement, say $|\omega_3\rangle$, is not the state before measurement, namely $|\psi\rangle$.

(3) If N such dice are tossed and covered with N mugs, there will be $NP(1)$ dice with $n=1$, $NP(2)$ dice with $n=2$, etc. in the ensemble *before and after the measurement*. In contrast, the quantum ensemble corresponding to $|\psi\rangle$ will contain N particles all of which are in the same state $|\psi\rangle$ (that is, each can yield any of the values $\omega_1, \dots, \omega_6$) before the measurement, and $NP(\omega_i)$ particles in $|\omega_i\rangle$ *after* the measurement. Only the ensemble before the measurement represents the state $|\psi\rangle$. The ensemble after measurement is a mixture of six ensembles representing the states $|\omega_1\rangle, \dots, |\omega_6\rangle$.[‡] \square

Having seen the utility of the ensemble concept in quantum theory, we now define and discuss the two statistical variables that characterize an ensemble.

Expectation Value

Given a large ensemble of N particles in a state $|\psi\rangle$, quantum theory allows us to predict what fraction will yield a value ω if the variable Ω is measured. This prediction, however, involves solving the eigenvalue problem of the operator Ω . If one is not interested in such detailed information on the state (or the corresponding ensemble) one can calculate instead an average over the ensemble, called the *expectation value*, $\langle\Omega\rangle$. The expectation value is just the mean value defined in statistics:

$$\begin{aligned}\langle\Omega\rangle &= \sum_i P(\omega_i)\omega_i = \sum_i |\langle\omega_i|\psi\rangle|^2\omega_i \\ &= \sum_i \langle\psi|\omega_i\rangle\langle\omega_i|\psi\rangle\omega_i\end{aligned}\tag{4.2.5}$$

But for the factors ω_i multiplying each projection operator $|\omega_i\rangle\langle\omega_i|$, we could have used $\sum_i |\omega_i\rangle\langle\omega_i| = I$. To get around this, note that $\omega_i|\omega_i\rangle = \Omega|\omega_i\rangle$. Feeding this in and continuing, we get

$$\langle\Omega\rangle = \sum_i \langle\psi|\Omega|\omega_i\rangle\langle\omega_i|\psi\rangle$$

Now we can use $\sum_i |\omega_i\rangle\langle\omega_i| = I$ to get

$$\langle\Omega\rangle = \langle\psi|\Omega|\psi\rangle\tag{4.2.6}$$

[‡] This is an example of a *mixed* ensemble. These will be discussed in the digression on density matrices, which follows in a while.

There are a few points to note in connection with this formula.

- (1) To calculate $\langle \Omega \rangle$, one need only be given the state vector and the operator Ω (say as a column vector and a matrix, respectively, in some basis). There is no need to find the eigenvectors or eigenvalues of Ω .
- (2) If the particle is in an eigenstate of Ω , that is $\Omega|\psi\rangle = \omega|\psi\rangle$, then $\langle \Omega \rangle = \omega$.
- (3) By the average value of Ω we mean the average over the ensemble. A given particle will of course yield only one of the eigenvalues upon measurement. The mean value will generally be an inaccessible value for a single measurement unless it accidentally equals an eigenvalue. [A familiar example of this phenomenon is that of the mean number of children per couple, which may be 2.12, although the number in a given family is restricted to be an integer.]

The Uncertainty

In any situation described probabilistically, another useful quantity to specify besides the mean is the *standard deviation*, which measures the average fluctuation around the mean. It is defined as

$$\Delta\Omega = \langle (\Omega - \langle \Omega \rangle)^2 \rangle^{1/2} \quad (4.2.7)$$

and often called the root-mean-squared deviation. In quantum mechanics, it is referred to as the *uncertainty in Ω* . If Ω has a discrete spectrum

$$(\Delta\Omega)^2 = \sum_i P(\omega_i)(\omega_i - \langle \Omega \rangle)^2 \quad (4.2.8)$$

and if it has a continuous spectrum,

$$(\Delta\Omega)^2 = \int P(\omega)(\omega - \langle \Omega \rangle)^2 d\omega \quad (4.2.9)$$

Notice that $\Delta\Omega$, just like $\langle \Omega \rangle$, is also calculable given just the state and the operator, for Eq. (4.2.7) means just

$$\Delta\Omega = [\langle \psi | (\Omega - \langle \Omega \rangle)^2 | \psi \rangle]^{1/2} \quad (4.2.10)$$

Usually the expectation value and the uncertainty provide us with a fairly good description of the state. For example, if we are given that a particle has $\langle X \rangle = a$ and $\Delta X = \Delta$, we know that the particle is likely to be spotted near $x = a$, with deviations of order Δ .

So far, we have concentrated on the measurement of a single variable at a time. We now turn our attention to the measurement of more than one variable at a time. (Since no two independent measurements can really be performed at the same time, we really mean the measurement of two or more dynamical variables in rapid succession.)

Exercise 4.2.1 (Very Important). Consider the following operators on a Hilbert space $\mathbb{V}^3(C)$:

$$L_x = \frac{1}{2^{1/2}} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad L_y = \frac{1}{2^{1/2}} \begin{bmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{bmatrix}, \quad L_z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

- (1) What are the possible values one can obtain if L_z is measured?
- (2) Take the state in which $L_z=1$. In this state what are $\langle L_x \rangle$, $\langle L_x^2 \rangle$, and ΔL_x ?
- (3) Find the normalized eigenstates and the eigenvalues of L_x in the L_z basis.
- (4) If the particle is in the state with $L_z=-1$, and L_x is measured, what are the possible outcomes and their probabilities?
- (5) Consider the state

$$|\psi\rangle = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2^{1/2} \end{bmatrix}$$

in the L_z basis. If L_z^2 is measured in this state and a result +1 is obtained, what is the state after the measurement? How probable was this result? If L_z is measured immediately afterwards, what are the outcomes and respective probabilities?

(6) A particle is in a state for which the probabilities are $P(L_z=1)=1/4$, $P(L_z=0)=1/2$, and $P(L_z=-1)=1/4$. Convince yourself that the most general, normalized state with this property is

$$|\psi\rangle = \frac{e^{i\delta_1}}{2} |L_z=1\rangle + \frac{e^{i\delta_2}}{2^{1/2}} |L_z=0\rangle + \frac{e^{i\delta_3}}{2} |L_z=-1\rangle$$

It was stated earlier on that if $|\psi\rangle$ is a normalized state then the state $e^{i\theta} |\psi\rangle$ is a physically equivalent normalized state. Does this mean that the factors $e^{i\delta_i}$ multiplying the L_z eigenstates are irrelevant? [Calculate for example $P(L_x=0)$.]

Compatible and Incompatible Variables

A striking feature of quantum theory is that given a particle in a state $|\psi\rangle$, one cannot say in general that the particle has a definite value for a given dynamical variable Ω : a measurement can yield any eigenvalue ω for which $\langle \omega | \psi \rangle$ is not zero. The exceptions are the states $|\omega\rangle$. A particle in one of these states can be said, as in classical mechanics, to have a value ω for Ω , since a measurement is assured to give this result. To produce such states we need only take an arbitrary state $|\psi\rangle$ and measure Ω . The measurement process acts as a filter that lets through just one component of $|\psi\rangle$, along some $|\omega\rangle$. The probability that this will happen is $P(\omega) = |\langle \omega | \psi \rangle|^2$.

We now wish to extend these ideas to more than one variable. We consider first the question of two operators. The extension to more than two will be

straightforward. We ask:

Question 1. Is there some multiple filtering process by which we can take an ensemble of particles in some state $|\psi\rangle$ and produce a state with well-defined values ω and λ for two variables Ω and Λ ?

Question 2. What is the probability that the filtering will give such a state if we start with the state $|\psi\rangle$?

To answer these questions, let us try to devise a multiple filtering scheme. Let us first measure Ω on the ensemble described by $|\psi\rangle$ and take the particles that yield a result ω . These are in a state that has a well-defined value for Ω . We *immediately* measure Λ and pick those particles that give a result λ . Do we have now an ensemble that is in a state with $\Omega=\omega$ and $\Lambda=\lambda$? Not generally. The reason is clear. After the first measurement, we had the system in the state $|\omega\rangle$, which assured a result ω for Ω , but nothing definite for Λ (since $|\omega\rangle$ need not be an eigenstate of Λ). Upon performing the second measurement, the state was converted to

$$|\psi'\rangle = |\lambda\rangle$$

and we are now assured a result λ for Λ , but nothing definite for Ω (since $|\lambda\rangle$ need not be an eigenstate of Ω).

In other words, the second filtering generally alters the state produced by the first. This change is just the collapse of the state vector $|\omega\rangle = \sum |\lambda\rangle \langle \lambda| \omega\rangle$ into the eigenstate $|\lambda\rangle$.

An exception occurs when the state produced after the first measurement is unaffected by the second. This in turn requires that $|\omega\rangle$ also be an eigenstate of Λ . The answer to the first question above is then in the affirmative only for the simultaneous eigenstates $|\omega\lambda\rangle$. The means for producing them are just as described above. These kets satisfy the equations

$$\Omega|\omega\lambda\rangle = \omega|\omega\lambda\rangle \quad (4.2.11)$$

$$\Lambda|\omega\lambda\rangle = \lambda|\omega\lambda\rangle \quad (4.2.12)$$

The question that arises naturally is: When will two operators admit simultaneous eigenkets? A necessary (but not sufficient) condition is obtained by operating Eq. (4.2.12) with Ω , Eq. (4.2.11) with Λ , and taking the difference:

$$(\Omega\Lambda - \Lambda\Omega)|\omega\lambda\rangle = 0 \quad (4.2.13)$$

Thus $[\Omega, \Lambda]$ must have eigenkets with zero eigenvalue if simultaneous eigenkets are to exist. A pair of operators Ω and Λ will fall into one of the three classes:

- A. Compatible: $[\Omega, \Lambda] = 0$
- B. Incompatible: $[\Omega, \Lambda] = \text{something that obviously has no zero eigenvalue}$
- C. Others

Class A. If two operators commute, we know a complete basis of simultaneous eigenkets can be found. Each element $|\omega\lambda\rangle$ of this basis has well-defined values for Ω and Λ .

Class B. The most famous example of this class is provided by the position and momentum operators X and P , which obey the *canonical commutation rule*

$$[X, P] = i\hbar \quad (4.2.14)$$

Evidently we cannot ever have $i\hbar|\psi\rangle = 0|\psi\rangle$ for any nontrivial $|\psi\rangle$. This means there doesn't exist even a single ket for which both X and P are well defined. Any attempt to filter X is ruined by a subsequent filtering for P and vice versa. This is the origin of the famous *Heisenberg uncertainty principle*, which will be developed as we go along.

Class C. In this case there are *some* states that are simultaneous eigenkets. There is nothing very interesting we can say about this case except to emphasize that even if two operators don't commute, one can still find a few common eigenkets, though not a full basis. (Why?)

Let us now turn to the second question of the probability of obtaining a state $|\omega\lambda\rangle$ upon measurement of Ω and Λ in a state $|\psi\rangle$. We will consider just case A; the question doesn't arise for case B, and case C is not very interesting. (You should be able to tackle case C yourself after seeing the other two cases.)

Case A. Let us first assume there is no degeneracy. Thus, to a given eigenvalue λ , there is just one ket and this must be a simultaneous eigenket $|\omega\lambda\rangle$. Suppose we measured Ω first. We get ω with a probability $P(\omega) = |\langle\omega\lambda|\psi\rangle|^2$. After the measurement, the particle is in a state $|\omega\lambda\rangle$. The measurement of Λ is certain to yield the result λ . The probability for obtaining ω for Ω and λ for Λ is just the product of the two probabilities

$$P(\omega, \lambda) = |\langle\omega\lambda|\psi\rangle|^2 \cdot 1 = |\langle\omega\lambda|\psi\rangle|^2$$

Notice that if Λ were measured first and Ω next, the probability is the same for getting the results λ and ω . Thus if we expand $|\psi\rangle$ in the complete common eigenbasis as

$$|\psi\rangle = \sum |\omega\lambda\rangle \langle\omega\lambda|\psi\rangle \quad (4.2.15a)$$

then

$$P(\omega, \lambda) = |\langle\omega\lambda|\psi\rangle|^2 = P(\lambda, \omega) \quad (4.2.15b)$$

The reason for calling Ω and Λ compatible if $[\Omega, \Lambda] = 0$ is that the measurement of one variable followed by the other doesn't alter the *eigenvalue* obtained in the first measurement and we have in the end a state with a well-defined value for both observables. Note the emphasis on the invariance of the *eigenvalue* under the second measurement. In the non-degenerate case, this implies the invariance of the state vector as well. In the degenerate case, the state vector can change due to the second

measurement, though the eigenvalue will not, as the following example will show. Consider two operators Λ and Ω on $\mathbb{V}^3(R)$. Let $|\omega_3\lambda_3\rangle$ be one common eigenvector. Let $\lambda_1 = \lambda_2 = \lambda$. Let $\omega_1 \neq \omega_2$ be the eigenvalues of Ω in this degenerate space. Let us use as a basis $|\omega_1\lambda\rangle$, $|\omega_2\lambda\rangle$, and $|\omega_3\lambda_3\rangle$. Consider a normalized state

$$|\psi\rangle = \alpha|\omega_3\lambda_3\rangle + \beta|\omega_1\lambda\rangle + \gamma|\omega_2\lambda\rangle \quad (4.2.16)$$

Let us say we measure Ω first and get ω_3 . The state becomes $|\omega_3\lambda_3\rangle$ and the subsequent measurement of Λ is assured to give a value λ_3 and to leave the state alone. Thus $P(\omega_3, \lambda_3) = |\langle\omega_3\lambda_3|\psi\rangle|^2 = \alpha^2$. Evidently $P(\omega_3, \lambda_3) = P(\lambda_3, \omega_3)$.

Suppose that the measurement of Ω gave a value ω_1 . The resulting state is $|\omega_1\lambda\rangle$ and the probability for this outcome is $|\langle\omega_1\lambda|\psi\rangle|^2$. The subsequent measurement of Λ will leave the state alone and yield the result λ with unit probability. Thus $P(\omega_1, \lambda)$ is the product of the probabilities:

$$P(\omega_1, \lambda) = |\langle\omega_1\lambda|\psi\rangle|^2 \cdot 1 = |\langle\omega_1\lambda|\psi\rangle|^2 = \beta^2 \quad (4.2.17)$$

Let us now imagine the measurements carried out in reverse order. Let the result of the measurement be λ . The state $|\psi'\rangle$ after measurement is the projection of $|\psi\rangle$ in the degenerate λ eigenspace:

$$|\psi'\rangle = \frac{\mathbb{P}_\lambda|\psi\rangle}{|\langle\mathbb{P}_\lambda|\psi|\mathbb{P}_\lambda|\psi\rangle|^{1/2}} = \frac{\beta|\omega_1\lambda\rangle + \gamma|\omega_2\lambda\rangle}{(\beta^2 + \gamma^2)^{1/2}} \quad (4.2.18)$$

where, in the expression above, the projected state has been normalized. The probability for this outcome is $P(\lambda) = \beta^2 + \gamma^2$, the square of the projection of $|\psi\rangle$ in the eigenspace. If Ω is measured now, both results ω_1 and ω_2 are possible. The probability for obtaining ω_1 is $|\langle\omega_1\lambda|\psi'\rangle|^2 = \beta^2/(\beta^2 + \gamma^2)$. Thus, the probability for the result $\Lambda = \lambda$, $\Omega = \omega_1$, is the product of the probabilities:

$$P(\lambda, \omega_1) = (\beta^2 + \gamma^2) \cdot \frac{\beta^2}{\beta^2 + \gamma^2} = \beta^2 = P(\omega_1, \lambda) \quad (4.2.19)$$

Thus $P(\omega_1, \lambda) = P(\lambda, \omega_1)$ independent of the degeneracy. *But this time the state suffered a change due to the second measurement* (unless by accident $|\psi'\rangle$ has no component along $|\omega_2\lambda\rangle$). Thus compatibility generally implies the invariance under the second measurement of the *eigenvalue* measured in the first. Therefore, the state can only be said to remain in the same eigenspace after the second measurement. If the first eigenvalue is non-degenerate, the eigenspace is one dimensional and the state vector itself remains invariant.

In our earlier discussion on how to produce well-defined states $|\psi\rangle$ for testing quantum theory, it was observed that the measurement process could itself be used as a preparation mechanism: if the measurement of Ω on an arbitrary, unknown initial state given a result ω , we are sure we have the state $|\psi\rangle = |\omega\rangle$. But this presumes ω is not a degenerate eigenvalue. If it is degenerate, we cannot nail down the state, except to within an eigenspace. It was therefore suggested that we stick to variables with a nondegenerate spectrum. We can now lift that restriction. Let us

say a degenerate eigenvalue ω for the variable Ω was obtained. We have then some vector in the ω eigenspace. We now measure another compatible variable Λ . If we get a result λ , we have a definite state $|\omega\lambda\rangle$, unless the value (ω, λ) itself is degenerate. We must then measure a third variable Γ compatible with Ω and Λ and so on. Ultimately we will get a state that is unique, given all the simultaneous eigenvalues: $|\omega, \lambda, \gamma, \dots\rangle$. It is presumed that such a set of compatible observables, called a *complete set of commuting observables*, exists. To prepare a state for studying quantum theory then, we take an arbitrary initial state and filter it by a sequence of compatible measurements till it is down to a unique, known vector. Any nondegenerate operator, all by itself, is a “complete set.”

Incidentally, even if the operators Ω and Λ are incompatible, we can specify the probability $P(\omega, \lambda)$ that the measurement of Ω followed by that of Λ on a state $|\psi\rangle$ will give the results ω and λ , respectively. However, the following should be noted:

(1) $P(\omega, \lambda) \neq P(\lambda, \omega)$ in general.

(2) The probability $P(\omega, \lambda)$ is not the probability for producing a final state that has well-defined values ω and λ for Ω and Λ . (Such a state doesn't exist by the definition of incompatibility.) The state produced by the two measurements is just the eigenstate of the second operator with the measured eigenvalue.

The Density Matrix—a Digression‡

So far we have considered ensembles of N systems all in the same state $|\psi\rangle$. They are hard to come by in practice. More common are ensembles of N systems, n_i ($i = 1, 2, \dots, k$) of which are in the state $|i\rangle$. (We restrict ourselves to the case where $|i\rangle$ is an element of an orthonormal basis.) Thus the ensemble is described by k kets $|1\rangle, |2\rangle, \dots, |k\rangle$, and k occupancy numbers n_1, \dots, n_k . A convenient way to assemble all this information is in the form of the *density matrix* (which is really an operator that becomes a matrix in some basis):

$$\rho = \sum_i p_i |i\rangle \langle i| \quad (4.2.20)$$

where $p_i = n_i/N$ is the probability that a system picked randomly out of the ensemble is in the state $|i\rangle$. The ensembles we have dealt with so far are said to be *pure*; they correspond to all $p_i = 0$ except one. A general ensemble is *mixed*.

Consider now the ensemble average of Ω . It is

$$\langle \bar{\Omega} \rangle = \sum_i p_i \langle i | \Omega | i \rangle \quad (4.2.21)$$

The bar on $\langle \bar{\Omega} \rangle$ reminds us that two kinds of averaging have been carried out: a quantum average $\langle i | \Omega | i \rangle$ for each system in $|i\rangle$ and a classical average over the

‡ This digression may be omitted or postponed without loss of continuity.

systems in different states $|i\rangle$. Observe that

$$\begin{aligned}\text{Tr}(\Omega\rho) &= \sum_j \langle j|\Omega\rho|j\rangle \\ &= \sum_j \sum_i \langle j|\Omega|i\rangle \langle i|j\rangle p_i = \sum_i \sum_j \langle i|j\rangle \langle j|\Omega|i\rangle p_i \\ &= \sum_i \langle i|\Omega|i\rangle p_i \\ &= \langle \bar{\Omega} \rangle\end{aligned}\tag{4.2.22}$$

The density matrix contains all the statistical information about the ensemble. Suppose we want, not $\langle \bar{\Omega} \rangle$, but instead $P(\omega)$, the probability of obtaining a particular value ω . We first note that, for a pure ensemble,

$$P(\omega) = |\langle \omega | \psi \rangle|^2 = \langle \psi | \omega \rangle \langle \omega | \psi \rangle = \langle \psi | \mathbb{P}_\omega | \psi \rangle = \langle \mathbb{P}_\omega \rangle$$

which combined with Eq. (4.2.22) tells us that

$$\overline{P(\omega)} = \text{Tr}(\mathbb{P}_\omega \rho)$$

The following results may be easily established:

- (1) $\rho^\dagger = \rho$
- (2) $\text{Tr } \rho = 1$
- (3) $\rho^2 = \rho$ for a pure ensemble
- (4) $\rho = (1/k)I$ for an ensemble uniformly distributed over k states
- (5) $\text{Tr } \rho^2 \leq 1$ (equality holds for a pure ensemble)

You are urged to convince yourself of these relations.

Example 4.2.4. To gain more familiarity with quantum theory let us consider an infinite-dimensional ket $|\psi\rangle$ expanded in the basis $|x\rangle$ of the position operator X :

$$|\psi\rangle = \int_{-\infty}^{\infty} |x\rangle \langle x| \psi \rangle dx = \int_{-\infty}^{\infty} |x\rangle \psi(x) dx$$

We call $\psi(x)$ the wave function (in the X basis). Let us assume $\psi(x)$ is a Gaussian, that is, $\psi(x) = A \exp[-(x-a)^2/2\Delta^2]$ (Fig. 4.2a). We now try to extract information about this state by using the postulates. Let us begin by normalizing the state:

$$\begin{aligned}1 &= \langle \psi | \psi \rangle = \int_{-\infty}^{\infty} \langle \psi | x \rangle \langle x | \psi \rangle dx = \int_{-\infty}^{\infty} |\psi(x)|^2 dx \\ &= \int_{-\infty}^{\infty} A^2 e^{-(x-a)^2/\Delta^2} dx = A^2 (\pi \Delta^2)^{1/2} \quad (\text{see Appendix A.2})\end{aligned}$$

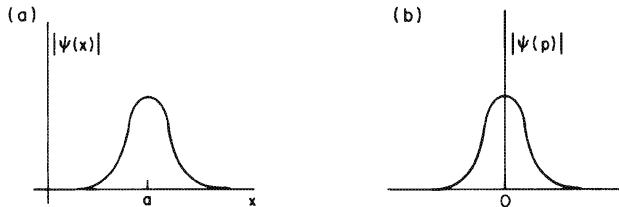


Figure 4.2. (a) The modulus of the wave function, $|\langle x|\psi \rangle| = |\psi(x)|$. (b) The modulus of the wave function, $|\langle p|\psi \rangle| = |\psi(p)|$.

So the normalized state is

$$\psi(x) = \frac{1}{(\pi\Delta^2)^{1/4}} e^{-(x-a)^2/2\Delta^2}$$

The probability for finding the particle between x and $x+dx$ is

$$P(x) dx = |\psi(x)|^2 dx = \frac{1}{(\pi\Delta^2)^{1/2}} e^{-(x-a)^2/\Delta^2} dx$$

which looks very much like Fig. 4.2a. Thus the particle is most likely to be found around $x=a$, and chances of finding it away from this point drop rapidly beyond a distance Δ . We can quantify these statements by calculating the expectation value and uncertainty for X . Let us do so.

Now, the operator X defined in postulate II is the same one we discussed at length in Section 1.10. Its action in the X basis is simply to multiply by x , i.e., if

$$\langle x | \psi \rangle = \psi(x)$$

then,

$$\begin{aligned} \langle x | X | \psi \rangle &= \int_{-\infty}^{\infty} \langle x | X | x' \rangle \langle x' | \psi \rangle dx' = \int_{-\infty}^{\infty} x \delta(x-x') \psi(x') dx' \\ &= x \psi(x) \end{aligned}$$

Using this result, the mean or expectation value of X is

$$\begin{aligned} \langle X \rangle &= \langle \psi | X | \psi \rangle = \int_{-\infty}^{\infty} \langle \psi | x \rangle \langle x | X | \psi \rangle dx \\ &= \int_{-\infty}^{\infty} \psi^*(x) x \psi(x) dx \\ &= \frac{1}{(\pi\Delta^2)^{1/2}} \int_{-\infty}^{\infty} e^{-(x-a)^2/\Delta^2} x dx \end{aligned}$$

If we define $y = x - a$,

$$\begin{aligned}\langle X \rangle &= \frac{1}{(\pi\Delta^2)^{1/2}} \int_{-\infty}^{\infty} (y + a) e^{-y^2/\Delta^2} dy \\ &= a\end{aligned}$$

We should have anticipated this result of course, since the probability density is symmetrically distributed around $x = a$.

Next, we calculate the fluctuations around $\langle X \rangle = a$, i.e., the uncertainty

$$\begin{aligned}\Delta X &= [\langle \psi | (X - \langle X \rangle)^2 | \psi \rangle]^{1/2} \\ &= [\langle \psi | X^2 - 2X\langle X \rangle + \langle X \rangle^2 | \psi \rangle]^{1/2} \\ &= [\langle \psi | X^2 - \langle X \rangle^2 | \psi \rangle]^{1/2} \quad (\text{since } \langle \psi | X | \psi \rangle = \langle X \rangle) \\ &= [\langle X^2 \rangle - \langle X \rangle^2]^{1/2} \\ &= [\langle X^2 \rangle - a^2]^{1/2}\end{aligned}$$

Now

$$\begin{aligned}\langle X^2 \rangle &= \frac{1}{(\pi\Delta^2)^{1/2}} \int_{-\infty}^{\infty} e^{-(x-a)^2/2\Delta^2} \cdot x^2 \cdot e^{-(x-a)^2/2\Delta^2} dx \\ &= \frac{1}{(\pi\Delta^2)^{1/2}} \int_{-\infty}^{\infty} e^{-y^2/\Delta^2} (y^2 + 2ya + a^2) dy = \frac{\Delta^2}{2} + 0 + a^2\end{aligned}$$

So

$$\Delta X = \frac{\Delta}{2^{1/2}}$$

So much for the information on the variable X . Suppose we next want to know the probability distribution for different values of another dynamical variable, say the momentum P .

- (1) First we must construct the operator P in this basis.
- (2) Then we must find its eigenvalues p , and eigenvectors $|p\rangle$.
- (3) Finally, we must take the inner product $\langle p|\psi\rangle$.
- (4) If p is discrete, $|\langle p_i|\psi\rangle|^2 = P(p_i)$, and if p is continuous, $|\langle p|\psi\rangle|^2 = P(p)$, the probability density.

Now, the P operator is just the K operator discussed in Section 1.10 multiplied by \hbar and has the action of $-i\hbar d/dx$ in the X basis, for if

$$\langle x|\psi\rangle = \psi(x)$$

$$\begin{aligned}
 \langle x|P|\psi\rangle &:= \int_{-\infty}^{\infty} \langle x|P|x'\rangle \langle x'|\psi\rangle dx' \\
 &= \int_{-\infty}^{\infty} [-i\hbar\delta'(x-x')] \psi(x') dx' \quad (\text{Postulate II}) \\
 &= -i\hbar \frac{d\psi}{dx}
 \end{aligned}$$

Thus, if we project the eigenvalue equation

$$P|p\rangle = p|p\rangle$$

onto the X basis, we get

$$\langle x|P|p\rangle = p\langle x|p\rangle$$

or

$$-i\hbar \frac{d\psi_p(x)}{dx} = p\psi_p(x)$$

where $\psi_p(x) = \langle x|p\rangle$. The solutions, normalized to the Dirac delta function[‡] are (from Section 1.10)

$$\psi_p(x) = \frac{1}{(2\pi\hbar)^{1/2}} e^{ipx/\hbar}$$

Now we can compute

$$\begin{aligned}
 \langle p|\psi\rangle &= \int \langle p|x\rangle \langle x|\psi\rangle dx = \int \psi_p^*(x) \psi(x) dx \\
 &= \int_{-\infty}^{\infty} \frac{e^{-ipx/\hbar}}{(2\pi\hbar)^{1/2}} \frac{e^{-(x-a)^2/2\Delta^2}}{(\pi\Delta^2)^{1/4}} dx = \left(\frac{\Delta^2}{\pi\hbar^2}\right)^{1/4} e^{-ipa/\hbar} e^{-p^2\Delta^2/2\hbar^2}
 \end{aligned}$$

The modulus of $\psi(p)$ is a Gaussian (Fig. 4.2b) of width $\hbar/2^{1/2}\Delta$. It follows that $\langle P\rangle = 0$, and $\Delta P = \hbar/2^{1/2}\Delta$. Since $\Delta X = \Delta/2^{1/2}$; we get the relation

$$\Delta X \cdot \Delta P = \hbar/2$$

[‡] Here we want $\langle p|p'\rangle = \delta(p-p') = \delta(k-k')/\hbar$, where $p = \hbar k$. This explains the $(2\pi\hbar)^{-1/2}$ normalization factor.

The Gaussian happens to saturate the lower bound of the uncertainty relation (to be formally derived in chapter 9):

$$\Delta X \cdot \Delta P \geq \hbar/2$$

The uncertainty relation is a consequence of the general fact that anything narrow in one space is wide in the transform space and vice versa. So if you are a 110-lb weakling and are taunted by a 600-lb bully, just ask him to step into momentum space! \square

This is a good place to point out that the plane waves $e^{ipx/\hbar}$ (and *all improper vectors*, i.e., vectors that can't be normalized to unity but only to the Dirac delta function) are introduced into the formalism as purely mathematical entities. Our inability to normalize them to unity translates into our inability to associate with them a sensible absolute probability distribution, so essential to the physical interpretation of the wave function. In the present case we have a particle whose relative probability density is uniform in all of space. Thus the absolute probability of finding it in any finite volume, even as big as our solar system, is zero. Since any particle that we are likely to be interested in will definitely be known to exist in some finite volume of such large dimensions, it is clear that no physically interesting state will be given by a plane wave. But, since the plane waves are eigenfunctions of P , does it mean that states of well-defined momentum do not exist? Yes, in the strict sense. However, there do exist states that are both normalizable to unity (i.e., correspond to *proper* vectors) and come arbitrarily close to having a precise momentum. For example, a wave function that behaves as $e^{ip_0x/\hbar}$ over a large region of space and tapers off to zero beyond, will be normalizable to unity and will have a Fourier transform so sharply peaked at $p = p_0$ that momentum measurements will only give results practically indistinguishable from p_0 . Thus there is no conflict between the fact that plane waves are unphysical, while states of well-defined momentum exist, for "well defined" never means "mathematically exact," but only "exact to any measurable accuracy." Thus a particle coming out of some accelerator with some advertised momentum, say $500 \text{ GeV}/c$, is in a proper normalizable state (since it is known to be located in our laboratory) and not in a plane wave state corresponding to $|p = 500 \text{ GeV}/c\rangle$.

But despite all this, we will continue to use the eigenkets $|p\rangle$ as basis vectors and to speak of a particle being in the state $|p\rangle$, because these vectors are so much more convenient to handle mathematically than the proper vectors. It should, however, be borne in mind that when we say a particle is (coming out of the accelerator) in a state $|p_0\rangle$, it is really in a proper state with a momentum space wave function so sharply peaked at $p = p_0$ that it may be replaced by a delta function $\delta(p - p_0)$.

The other set of improper kets we will use in the same spirit are the position eigenkets $|x\rangle$, which also form a convenient basis. Again, when we speak of a particle being in a state $|x_0\rangle$ we shall mean that its wave function is so sharply peaked at $x = x_0$ that it may be treated as a delta function to a good accuracy.[‡]

[‡] Thus, by the physical Hilbert space, we mean the space of interest to physicists, not one whose elements all correspond to physically realizable states.

Occasionally, the replacement of a proper wave function by its improper counterpart turns out to be a poor approximation. Here is an example from Chapter 19: Consider the probability that a particle coming out of an accelerator with a nearly exact momentum scatters off a target and enters a detector placed far away, and not in the initial direction. Intuition says that the answer must be zero if the target is absent. This reasonable condition is violated if we approximate the initial state of the particle by a plane wave (which is nonzero everywhere). So we proceed as follows. In the vicinity of the target, we use the plane wave to approximate the initial wave function, for the two are indistinguishable over the (finite and small) range of influence of the target. At the detector, however, we go back to the proper wave (which has tapered off) to represent the initial state.

*Exercise 4.2.2.** Show that for a real wave function $\psi(x)$, the expectation value of momentum $\langle P \rangle = 0$. (Hint: Show that the probabilities for the momenta $\pm p$ are equal.) Generalize this result to the case $\psi = c\psi_r$, where ψ_r is real and c an arbitrary (real or complex) constant. (Recall that $|\psi\rangle$ and $a|\psi\rangle$ are physically equivalent.)

*Exercise 4.2.3.** Show that if $\psi(x)$ has mean momentum $\langle P \rangle$, $e^{ip_0x/\hbar}\psi(x)$ has mean momentum $\langle P \rangle + p_0$.

Example 4.2.5. The collapse of the state vector and the uncertainty principle play a vital role in explaining the following extension of the double slit experiment. Suppose I say, “I don’t believe that a given particle (let us say an electron) doesn’t really go through one slit or the other. So I will set up a light source in between the slits to the right of the screen. Each passing electron will be exposed by the beam and I note which slit it comes out of. Then I note where it arrives on the screen. I make a table of how many electrons arrive at each x and which slit they came from. Now there is no escape from the conclusion that the number arriving at a given x is the sum of the numbers arriving via S_1 and S_2 . So much for quantum theory and its interference pattern!”

But the point of course is that quantum theory no longer predicts an interference pattern! The theory says that if an electron of definite momentum p is involved, the corresponding wave function is a wave with a well-defined wave number $k = p/\hbar$, which interferes with itself and produces a nice interference pattern. This prediction is valid only as long as the state of the electron is what we say it is. But this state is necessarily altered by the light source, which upon measuring the position of the electron (as being next to S_1 , say) changes its wave function from something that was extended in space to something localized near S_1 . Once the state is changed, the old prediction of interference is no longer valid.

Now, once in a while some electrons will get to the detectors without being detected by the light source. We note where these arrive, but cannot classify them as coming via S_1 or S_2 . When the distribution of just these electrons is plotted; sure enough we get the interference pattern. We had better, for quantum theory predicts it, the state not having been tampered with in these cases.

The above experiment can also be used to demystify to some extent the collapse of the wave function under measurement. Why is it that even the ideal measurement produces unavoidable changes in the state? The answer, as we shall see, has to do with the fact that \hbar is not zero.

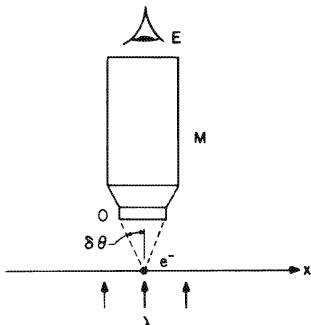


Figure 4.3. Light of frequency λ bounces off the electron, enters the objective O of the microscope, and enters the eye E of the observer.

Consider the schematic set up in Fig. 4.3. Light of wavelength λ illuminates an electron (e^-), enters the objective (O) of a microscope (M) and reaches our eye (E). If $\delta\theta$ is the opening angle of the cone of light entering the objective after interacting with the electron, classical optics limits the accuracy of the position measurement by an uncertainty

$$\Delta X \cong \lambda / \sin \delta\theta$$

Both classically and quantum mechanically, we can reduce ΔX to 0 by reducing λ to zero.[‡] In the latter description however, the improved accuracy in the position measurement is at the expense of producing an increased uncertainty in the x component (p_x) of the electron momentum. The reason is that light of wavelength λ is not a continuous wave whose impact on the electron momentum may be arbitrarily reduced by a reduction of its amplitude, but rather a flux of photons of momentum $p = 2\pi\hbar/\lambda$. As λ decreases, the collisions between the electron and the photons become increasingly violent. This in itself would not lead to an uncertainty in the electron momentum, were it not for the fact that the x component of the photons entering the objective can range from 0 to $p \sin \delta\theta = 2\pi\hbar \sin \delta\theta / \lambda$. Since at least one photon must reach our eyes after bouncing off the electron for us to see it, there is a minimum uncertainty in the recoil momentum of the electron given by

$$\Delta P_x \cong \frac{2\pi\hbar}{\lambda} \sin \delta\theta$$

Consequently, we have at the end of our measurement an electron whose position and momenta are uncertain by ΔX and ΔP_x such that

$$\Delta X \cdot \Delta P_x \cong 2\pi\hbar \cong \hbar$$

[The symbols ΔX and ΔP_x are not precisely the quantities defined in Eq. (4.2.7) but are of the same order of magnitude.] This is the famous *uncertainty principle*. There is no way around it. If we soften the blow of each photon by increasing λ or narrowing the objective to better constrain the final photon momentum, we lose in resolution.

[‡] This would be the ideal position measurement.

More elaborate schemes, which determine the recoil of the microscope, are equally futile. Note that if \hbar were 0, we could have ΔX and ΔP_x simultaneously 0. Physically, it means that we can increase our position resolution without increasing the punch carried by the photons. Of course \hbar is not zero and we can't make it zero in any experiment. But what we can do is to use bigger and bigger objects for our experiment so that in the scale of these objects \hbar appears to be negligible. We then regain classical mechanics. The position of a billiard ball can be determined very well by shining light on it, but this light hardly affects its momentum. This is why one imagines in classical mechanics that momentum and position can be well defined simultaneously. \square

Generalization to More Degrees of Freedom

Our discussion so far has been restricted to a system with one degree of freedom—namely, a single particle in one dimension. We now extend our domain to a system with N degrees of freedom. The only modification is in postulate II, which now reads as follows.

Postulate II. Corresponding to the N Cartesian coordinates x_1, \dots, x_N describing the classical system, there exist in quantum theory N mutually commuting operators X_1, \dots, X_N . In the simultaneous eigenbasis $|x_1, x_2, \dots, x_N\rangle$ of these operators, called the *coordinate basis* and normalized as

$$\langle x_1, x_2, \dots, x_N | x'_1, x'_2, \dots, x'_N \rangle = \delta(x_1 - x'_1) \dots \delta(x_N - x'_N)$$

(the product of delta functions vanishes unless all the arguments vanish) we have the following correspondence:

$$|\psi\rangle \rightarrow \langle x_1, \dots, x_N | \psi \rangle = \psi(x_1, \dots, x_N)$$

$$X_i |\psi\rangle \rightarrow \langle x_1, \dots, x_N | X_i | \psi \rangle = x_i \psi(x_1, \dots, x_N)$$

$$P_i |\psi\rangle \rightarrow \langle x_1, \dots, x_N | P_i | \psi \rangle = -i\hbar \frac{\partial}{\partial x_i} \psi(x_1, \dots, x_N)$$

P_i being the momentum operator corresponding to the classical momentum p_i . Dependent dynamical variables $\omega(x_i, p_j)$ are represented by operators $\Omega = \omega(x_i \rightarrow X_i, p_j \rightarrow P_j)$.

The other postulates remain the same. For example $|\psi(x_1, \dots, x_N)|^2 \times dx_1 \dots dx_N$ is the probability that the particle coordinates lie between x_1, x_2, \dots, x_N and $x_1 + dx_1, x_2 + dx_2, \dots, x_N + dx_N$.

This postulate is stated in terms of Cartesian coordinates since only in terms of these can one express the operator assignments in the simple form $X_i \rightarrow x_i$, $P_i \rightarrow -i\hbar \partial/\partial x_i$. Once the substitutions have been made and the desired equations obtained in the coordinate basis, one can perform any desired change of variable before solving them. Suppose, for example, that we want to find the eigenvalues and

$$\omega = \frac{p_1^2 + p_2^2 + p_3^2}{2m} + x_1^2 + x_2^2 + x_3^2 \quad (4.2.24)$$

where x_1 , x_2 , and x_3 are the three Cartesian coordinates and p_i the corresponding momenta of a particle of mass m in three dimensions. Since the coordinates are usually called x , y , and z , let us follow this popular notation and rewrite Eq. (4.2.24) as

$$\omega = \frac{p_x^2 + p_y^2 + p_z^2}{2m} + x^2 + y^2 + z^2 \quad (4.2.25)$$

To solve the equation

$$\Omega|\omega\rangle = \omega|\omega\rangle$$

with

$$\Omega = \frac{P_x^2 + P_y^2 + P_z^2}{2m} + X^2 + Y^2 + Z^2$$

we make the substitution

$$|\omega\rangle \rightarrow \psi_\omega(x, y, z)$$

$$X \rightarrow x, \quad P_x \rightarrow -i\hbar \frac{\partial}{\partial x}$$

etc. and get

$$\left[\frac{-\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + x^2 + y^2 + z^2 \right] \psi_\omega(x, y, z) = \omega \psi_\omega(x, y, z) \quad (4.2.26)$$

Once we have obtained this differential equation, we can switch to any other set of coordinates. In the present case the spherical coordinates r , θ , and ϕ recommend themselves. Since

$$\begin{aligned} & \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \\ & \equiv \nabla^2 \equiv \frac{1}{r^2} \left[\frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right] \end{aligned}$$

$$\frac{-\hbar^2}{2m} \left[\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi_\omega}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \psi_\omega}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi_\omega}{\partial \phi^2} \right] + r^2 \psi_\omega = \omega \psi_\omega \quad (4.2.27)$$

What if we wanted to go directly from ω in spherical coordinates

$$\omega = \frac{1}{2m} \left(p_r^2 + \frac{p_\theta^2}{r^2} + \frac{p_\phi^2}{r^2 \sin^2 \theta} \right) + r^2$$

to Eq. (4.2.27)? It is clear upon inspection that there exists no simple rule [such as $p_r \rightarrow (-i\hbar \partial/\partial r)$] for replacing the classical momenta by differential operators in r , θ , and ϕ which generates Eq. (4.2.27) starting from the ω above. There does exist a complicated procedure for quantizing in non-Cartesian coordinates, but we will not discuss it, since the recipe eventually reproduces what the Cartesian recipe (which seems to work[‡]) yields so readily.

There are further generalizations, namely, to relativistic quantum mechanics and to quantum mechanics of systems in which particles are created and destroyed (so that the number of degrees of freedom changes!). Except for a brief discussion of these toward the end of the program, we will not address these matters.

4.3. The Schrödinger Equation (Dotting Your *i*'s and Crossing Your *#*'s)

Having discussed in some detail the state at a given time, we now turn our attention to postulate IV, which specifies the change of this state with time. According to this postulate, the state obeys the Schrödinger equation

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = H|\psi(t)\rangle \quad (4.3.1)$$

Our discussion of this equation is divided into three sections:

- (1) Setting up the equation
- (2) General approach to its solution
- (3) Choosing a basis for solving the equation

Setting Up the Schrödinger Equation

To set up the Schrödinger equation one must simply make the substitution $\mathcal{H}(x \rightarrow X, p \rightarrow P)$, where \mathcal{H} is the classical Hamiltonian for the same problem. Thus,

[‡] In the sense that in cases where comparison with experiment is possible, as in say the hydrogen spectrum, there is agreement.

if we are describing a harmonic oscillator, which is classically described by the Hamiltonian

$$\mathcal{H} = \frac{p^2}{2m} + \frac{1}{2} m\omega^2 x^2 \quad (4.3.2)$$

the Hamiltonian operator in quantum mechanics is

$$H = \frac{P^2}{2m} + \frac{1}{2} m\omega^2 X^2 \quad (4.3.3)$$

In three dimensions, the Hamiltonian operator for the quantum oscillator is likewise

$$H = \frac{P_x^2 + P_y^2 + P_z^2}{2m} + \frac{1}{2} m\omega^2 (X^2 + Y^2 + Z^2) \quad (4.3.4)$$

assuming the force constant is the same in all directions.

If the particle in one dimension is subject to a constant force f , then

$$\mathcal{H} = \frac{p^2}{2m} - fx$$

and

$$H = \frac{P^2}{2m} - fX \quad (4.3.5)$$

For a particle of charge q in an electromagnetic field in three dimensions,

$$\mathcal{H} = \frac{|\mathbf{p} - (q/c)\mathbf{A}(\mathbf{r}, t)|^2}{2m} + q\phi(\mathbf{r}, t) \quad (4.3.6)$$

In constructing the corresponding quantum Hamiltonian operator, we must use the symmetrized form

$$H = \frac{1}{2m} \left(\mathbf{P} \cdot \mathbf{P} - \frac{q}{c} \mathbf{P} \cdot \mathbf{A} - \frac{q}{c} \mathbf{A} \cdot \mathbf{P} + \frac{q^2}{c^2} \mathbf{A} \cdot \mathbf{A} \right) + q\phi \quad (4.3.7)$$

since \mathbf{P} does not commute with \mathbf{A} , which is a function of X , Y , and Z .

In this manner one can construct the Hamiltonian H for any problem with a classical counterpart. Problems involving spin have no classical counterparts and some improvisation is called for. We will discuss this question when we study spin in some detail in Chapter 14.

General Approach to the Solution

Let us first assume that H has no explicit t dependence. In this case the equation

$$i\hbar|\dot{\psi}\rangle = H|\psi\rangle$$

is analogous to equations discussed in Chapter 1

$$|\ddot{x}\rangle = \Omega|x\rangle$$

and

$$|\ddot{\psi}\rangle = -K^2|\psi\rangle$$

describing the coupled masses and the vibrating string, respectively. Our approach will once again be to find the eigenvectors and eigenvalues of H and to construct the propagator $U(t)$ in terms of these. Once we have $U(t)$, we can write

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle$$

There is no need to make assumptions about $|\psi(0)\rangle$ here, since it is determined by Eq. (4.3.1):

$$|\dot{\psi}(0)\rangle = \frac{-i}{\hbar}H|\psi(0)\rangle$$

In other words, Schrödinger's equation is first order in time, and the specification of $|\psi\rangle$ at $t=0$ is a sufficient initial-value datum.

Let us now construct an explicit expression for $U(t)$ in terms of $|E\rangle$, the normalized eigenkets of H with eigenvalues E which obey

$$H|E\rangle = E|E\rangle \quad (4.3.8)$$

This is called the *time-independent Schrödinger equation*. Assume that we have solved it and found the kets $|E\rangle$. If we expand $|\psi\rangle$ as

$$|\psi(t)\rangle = \sum |E\rangle \langle E|\psi(t)\rangle \equiv \sum a_E(t)|E\rangle \quad (4.3.9)$$

the equation for $a_E(t)$ follows if we act on both sides with $(i\hbar\partial/\partial t - H)$:

$$0 = (i\hbar\partial/\partial t - H)|\psi(t)\rangle = \sum (i\hbar\dot{a}_E - Ea_E)|E\rangle \Rightarrow i\hbar\dot{a}_E = Ea_E \quad (4.3.10)$$

where we have used the linear independence of the kets $|E\rangle$. The solution to Eq. (4.3.10) is

$$a_E(t) = a_E(0) e^{-iEt/\hbar} \quad (4.3.11a)$$

or

$$\langle E | \psi(t) \rangle = \langle E | \psi(0) \rangle e^{-iEt/\hbar} \quad (4.3.11b)$$

so that

$$|\psi(t)\rangle = \sum_E |E\rangle \langle E | \psi(0) \rangle e^{-iEt/\hbar} \quad (4.3.12)$$

We can now extract $U(t)$:

$$U(t) = \sum_E |E\rangle \langle E| e^{-iEt/\hbar} \quad (4.3.13)$$

We have been assuming that the energy spectrum is discrete and nondegenerate. If E is degenerate, one must first introduce an extra label α (usually the eigenvalue of a compatible observable) to specify the states. In this case

$$U(t) = \sum_{\alpha} \sum_E |E, \alpha\rangle \langle E, \alpha| e^{-iEt/\hbar}$$

If E is continuous, the sum must be replaced by an integral. The normal modes

$$|E(t)\rangle = |E\rangle e^{-iEt/\hbar}$$

are also called *stationary states* for the following reason: the probability distribution $P(\omega)$ for any variable Ω is time-independent in such a state:

$$\begin{aligned} P(\omega, t) &= |\langle \omega | \psi(t) \rangle|^2 \\ &= |\langle \omega | E(t) \rangle|^2 \\ &= |\langle \omega | E \rangle e^{-iEt/\hbar}|^2 \\ &= |\langle \omega | E \rangle|^2 \\ &= P(\omega, 0) \end{aligned}$$

There exists another expression for $U(t)$ besides the sum, Eq. (4.3.13), and that is

$$U(t) = e^{-iHt/\hbar} \quad (4.3.14)$$

If this exponential series converges (and it sometimes does not), this form of $U(t)$ can be very useful. (Convince yourself that $|\psi(t)\rangle = e^{-iHt/\hbar} |\psi(0)\rangle$ satisfies Schrödinger's equation.)

Since H (the energy operator) is Hermitian, it follows that $U(t)$ is unitary. We may therefore think of the time evolution of a ket $|\psi(t)\rangle$ as a "rotation" in Hilbert

space. One immediate consequence is that the norm $\langle \psi(t) | \psi(t) \rangle$ is invariant:

$$\langle \psi(t) | \psi(t) \rangle = \langle \psi(0) | U^\dagger(t) U(t) | \psi(0) \rangle = \langle \psi(0) | \psi(0) \rangle \quad (4.3.15)$$

so that a state, once normalized, stays normalized. There are other consequences of the fact that the time evolution may be viewed as a rotation. For example, one can abandon the fixed basis we have been using, and adopt one that also rotates at the same rate as the state vectors. In such a basis the vectors would appear frozen, but the operators, which were constant matrices in the fixed basis, would now appear to be time dependent. Any physical entity, such as a matrix element, would, however, come out the same as before since $\langle \phi | \Omega | \psi \rangle$, which is the dot product of $\langle \phi |$ and $|\Omega \psi \rangle$, is invariant under rotations. This view of quantum mechanics is called the *Heisenberg picture*, while the one we have been using is called the *Schrödinger picture*. Infinitely many pictures are possible, each labeled by how the basis is rotating. So if you think you were born too late to make a contribution to quantum theory fear not, for you can invent your own picture. We will take up the study of various pictures in Chapter 18.

Let us now consider the case $H = H(t)$. We no longer look for normal modes, since the operator in question is changing with time. There exists no fixed strategy for solving such problems. In the course of our study we will encounter a time-dependent problem involving spin which can be solved exactly. We will also study a systematic approximation scheme for solving problems with

$$H(t) = H^0 + H^1(t)$$

where H^0 is a large time-independent piece and $H^1(t)$ is a small time-dependent piece.

What is the propagator $U(t)$ in the time-dependent case? In other words, how is $U(t)$ in $|\psi(t)\rangle = U(t)|\psi(0)\rangle$ related to $H(t)$? To find out, we divide the interval $(0-t)$ into N pieces of width $\Delta = t/N$, where N is very large and Δ is very small. By integrating the Schrödinger equation over the first interval, we can write *to first order in Δ*

$$\begin{aligned} |\psi(\Delta)\rangle &= |\psi(0)\rangle + \Delta \frac{d|\psi\rangle}{dt} \Big|_0 \\ &= |\psi(0)\rangle - \frac{i\Delta}{\hbar} H(0)|\psi(0)\rangle \\ &= \left[1 - \frac{i\Delta}{\hbar} H(0) \right] |\psi(0)\rangle \end{aligned}$$

which, to this order

$$= \exp \left[\frac{-i\Delta}{\hbar} H(0) \right] |\psi(0)\rangle$$

[One may wonder whether in the interval $0 - \Delta$, one must use $H(0)$ or $H(\Delta)$ or $H(\Delta/2)$ and so on. The difference between these possibilities is of order Δ and hence irrelevant, since there is already one power of Δ in front of H .] Inchng forth in steps of Δ , we get

$$|\psi(t)\rangle = \prod_{n=0}^{N-1} e^{-i\Delta H(n\Delta)/\hbar} |\psi(0)\rangle$$

We cannot simply add the exponents to get, in the $N \rightarrow \infty$ limit,

$$U(t) = \exp \left[-(i/\hbar) \int_0^t H(t') dt' \right]$$

since

$$[H(t_1), H(t_2)] \neq 0$$

in general. For example, if

$$H(t) = X^2 \cos^2 \omega t + P^2 \sin^2 \omega t$$

then

$$H(0) = X^2$$

and

$$H(\pi/2\omega) = P^2$$

and

$$[H(0), H(\pi/2\omega)] \neq 0$$

It is common to use the symbol, called the *time-ordered integral*

$$T \left\{ \exp \left[-(i/\hbar) \int_0^t H(t') dt' \right] \right\} = \lim_{N \rightarrow \infty} \prod_{n=0}^{N-1} \exp[-(i/\hbar)H(n\Delta)\Delta]$$

in such problems. We will not make much use of this form of $U(t)$. But notice that being a product of unitary operators, $U(t)$ is unitary, and time evolution continues to be a “rotation” whether or not H is time independent.

Whether or not H depends on time, the propagator satisfies the following conditions:

$$\begin{aligned} U(t_3, t_2)U(t_2, t_1) &= U(t_3, t_1) \\ U^\dagger(t_2, t_1) &= U^{-1}(t_2, t_1) = U(t_1, t_2) \end{aligned} \quad (4.3.16)$$

It is intuitively clear that these equations are correct. You can easily prove them by applying the U 's to some arbitrary state and using the fact that U is unitary and $U(t, t) = I$.

Choosing a Basis for Solving Schrödinger's Equation

Barring a few exceptions, the Schrödinger equation is always solved in a particular basis. Although all bases are equal mathematically, some are more equal than others. First of all, since $H = H(X, P)$ the X and P bases recommend themselves, for in going to one of them the corresponding operator is rendered diagonal. Thus one can go to the X basis in which $X \rightarrow x$ and $P \rightarrow -i\hbar d/dx$ or to the P basis in which $P \rightarrow p$ and $X \rightarrow i\hbar d/dp$. The choice between the two depends on the Hamiltonian. Assuming it is of the form (in one dimension)

$$H = T + V = \frac{P^2}{2m} + V(X) \quad (4.3.17)$$

the choice is dictated by $V(X)$. Since $V(X)$ is usually a more complicated function of X than T is of P , one prefers the X basis. Thus if

$$H = \frac{P^2}{2m} + \frac{1}{\cosh^2 X} \quad (4.3.18)$$

the equation

$$H|E\rangle = E|E\rangle$$

becomes in the X basis the second-order equation

$$\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + \frac{1}{\cosh^2 x} \right) \psi_E(x) = E\psi_E(x) \quad (4.3.19)$$

which can be solved. Had one gone to the P basis, one would have ended up with the equation

$$\left[\frac{p^2}{2m} + \frac{1}{\cosh^2(i\hbar d/dp)} \right] \psi_E(p) = E\psi_E(p) \quad (4.3.20)$$

which is quite frightening.

A problem where the P basis is preferred is that of a particle in a constant force field f , for which

$$H = \frac{P^2}{2m} - fX \quad (4.3.21)$$

In the P basis one gets a first-order differential equation

$$\left(\frac{P^2}{2m} - i\hbar f \frac{d}{dp} \right) \psi_E(p) = E\psi_E(p) \quad (4.3.22)$$

whereas in the X basis one gets the second-order equation

$$\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} - fx \right) \psi_E(x) = E\psi_E(x) \quad (4.3.23)$$

The harmonic oscillator can be solved with equal ease in either basis since H is quadratic in X and P . It turns out to be preferable to solve it in a third basis in which neither X nor P is diagonal! You must wait till Chapter 7 before you see how this happens.

There exists a built-in bias in favor of the X basis. This has to do with the fact that the x space is the space we live in. In other words, when we speak of the probability of obtaining a value between x and $x+dx$ if the variable X is measured, we mean simply the probability of finding the particle between x and $x+dx$ in our space. One may thus visualize $\psi(x)$ as a function in our space, whose modulus squared gives the probability density for finding a particle near x . Such a picture is useful in thinking about the double-slit experiment or the electronic states in a hydrogen atom.

But like all pictures, it has its limits. First of all it must be borne in mind that even though $\psi(x)$ can be visualized as a wave in our space, it is not a real wave, like the electromagnetic wave, which carries energy, momentum, etc. To understand this point, consider a particle in three dimensions. The function $\psi(x, y, z)$ can be visualized as a wave in our space. But, if we consider next a two-particle system, $\psi(x_1, y_1, z_1, x_2, y_2, z_2)$ is a function in a six-dimensional *configuration space* and cannot be visualized in our space.

Thus the case of the single particle is really an exception: there is only one position operator and the space of its eigenvalues *happens to coincide* with the space in which we live and in which the drama of physics takes place.

This brings us to the end of our general discussion of the postulates. We now turn to the application of quantum theory to various physical problems. For pedagogical reasons, we will restrict ourselves to problems of a single particle in one dimension in the next few chapters.

5

Simple Problems in One Dimension

Now that the postulates have been stated and explained, it is all over but for the applications. We begin with the simplest class of problems—concerning a single particle in one dimension. Although these one-dimensional problems are somewhat artificial, they contain most of the features of three-dimensional quantum mechanics but little of its complexity. One problem we will not discuss in this chapter is that of the harmonic oscillator. This problem is so important that a separate chapter has been devoted to its study.

5.1. The Free Particle

The simplest problem in this family is of course that of the free particle. The Schrödinger equation is

$$i\hbar|\psi\rangle = H|\psi\rangle = \frac{P^2}{2m}|\psi\rangle \quad (5.1.1)$$

The normal modes or stationary states are solutions of the form

$$|\psi\rangle = |E\rangle e^{-iEt/\hbar} \quad (5.1.2)$$

Feeding this into Eq. (5.1.1), we get the time-independent Schrödinger equation for $|E\rangle$:

$$H|E\rangle = \frac{P^2}{2m}|E\rangle = E|E\rangle \quad (5.1.3)$$

This problem can be solved without going to any basis. First note that any eigenstate

of P is also an eigenstate of P^2 . So we feed the trial solution $|p\rangle$ into Eq. (5.1.3) and find

$$\frac{P^2}{2m}|p\rangle = E|p\rangle$$

or

$$\left(\frac{p^2}{2m} - E\right)|p\rangle = 0 \quad (5.1.4)$$

Since $|p\rangle$ is not a null vector, we find that the allowed values of p are

$$p = \pm(2mE)^{1/2} \quad (5.1.5)$$

In other words, there are two orthogonal eigenstates for each eigenvalue E :

$$|E, +\rangle = |p = (2mE)^{1/2}\rangle \quad (5.1.6)$$

$$|E, -\rangle = |p = -(2mE)^{1/2}\rangle \quad (5.1.7)$$

Thus, we find that to the eigenvalue E there corresponds a degenerate two-dimensional eigenspace, spanned by the above vectors. Physically this means that a particle of energy E can be moving to the right or to the left with momentum $|p| = (2mE)^{1/2}$. Now, you might say, “This is exactly what happens in classical mechanics. So what’s new?” What is new is the fact that the state

$$|E\rangle = \beta|p = (2mE)^{1/2}\rangle + \gamma|p = -(2mE)^{1/2}\rangle \quad (5.1.8)$$

is also an eigenstate of energy E and represents a *single* particle of energy E that can be caught moving either to the right or to the left with momentum $(2mE)^{1/2}$!

To construct the complete orthonormal eigenbasis of H , we must pick from each degenerate eigenspace any two orthonormal vectors. The obvious choice is given by the kets $|E, +\rangle$ and $|E, -\rangle$ themselves. In terms of the ideas discussed in the past, we are using the eigenvalue of a compatible variable P as an extra label within the space degenerate with respect to energy. Since P is a nondegenerate operator, the label p by itself is adequate. In other words, there is no need to call the state $|p, E=P^2/2m\rangle$, since the value of $E=E(p)$ follows, given p . We shall therefore drop this redundant label.

The propagator is then

$$\begin{aligned} U(t) &= \int_{-\infty}^{\infty} |p\rangle \langle p| e^{-iE(p)t/\hbar} dp \\ &= \int_{-\infty}^{\infty} |p\rangle \langle p| e^{-ip^2t/2m\hbar} dp \end{aligned} \quad (5.1.9)$$

Exercise 5.1.1. Show that Eq. (5.1.9) may be rewritten as an integral over E and a sum over the \pm index as

$$U(t) = \sum_{\alpha=\pm} \int_0^\infty \left[\frac{m}{(2mE)^{1/2}} \right] |E, \alpha\rangle \langle E, \alpha| e^{-iEt/\hbar} dE$$

*Exercise 5.1.2.** By solving the eigenvalue equation (5.1.3) in the X basis, regain Eq. (5.1.8), i.e., show that the general solution of energy E is

$$\psi_E(x) = \beta \frac{\exp[i(2mE)^{1/2}x/\hbar]}{(2\pi\hbar)^{1/2}} + \gamma \frac{\exp[-i(2mE)^{1/2}x/\hbar]}{(2\pi\hbar)^{1/2}}$$

[The factor $(2\pi\hbar)^{-1/2}$ is arbitrary and may be absorbed into β and γ .] Though $\psi_E(x)$ will satisfy the equation even if $E < 0$, are these functions in the Hilbert space?

The propagator $U(t)$ can be evaluated explicitly in the X basis. We start with the matrix element

$$\begin{aligned} U(x, t; x') &\equiv \langle x| U(t)| x' \rangle = \int_{-\infty}^{\infty} \langle x| p \rangle \langle p| x' \rangle e^{-ip^2t/2m\hbar} dp \\ &= \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} e^{ip(x-x')/\hbar} \cdot e^{-ip^2t/2m\hbar} dp \\ &= \left(\frac{m}{2\pi\hbar it} \right)^{1/2} e^{im(x-x')^2/2\hbar t} \end{aligned} \quad (5.1.10)$$

using the result from Appendix A.2 on Gaussian integrals. In terms of this propagator, any initial-value problem can be solved, since

$$\psi(x, t) = \int U(x, t; x') \psi(x', 0) dx' \quad (5.1.11)$$

Had we chosen the initial time to be t' rather than zero, we would have gotten

$$\psi(x, t) = \int U(x, t; x', t') \psi(x', t') dx' \quad (5.1.12)$$

where $U(x, t; x', t') = \langle x| U(t-t')| x' \rangle$, since U depends only on the time interval $t-t'$ and not the absolute values of t and t' . [Had there been a time-dependent potential such as $V(t) = V_0 e^{-\alpha t^2}$ in H , we could have told what absolute time it was by looking at $V(t)$. In the absence of anything defining an absolute time in the problem, only time differences have physical significance.] Whenever we set $t'=0$, we will resort to our old convention and write $U(x, t; x', 0)$ as simply $U(x, t; x')$.

A nice physical interpretation may be given to $U(x, t; x', t')$ by considering a special case of Eq. (5.1.12). Suppose we started off with a particle localized at

$x' = x'_0$, that is, with $\psi(x', t') = \delta(x' - x'_0)$. Then

$$\psi(x, t) = U(x, t; x'_0, t') \quad (5.1.13)$$

In other words, the propagator (in the X basis) is the amplitude that a particle starting out at the space-time point (x'_0, t') ends with at the space-time point (x, t) . [It can obviously be given such an interpretation in any basis: $\langle \omega | U(t, t') | \omega' \rangle$ is the amplitude that a particle in the state $|\omega'\rangle$ at t' ends up with in the state $|\omega\rangle$ at t .] Equation (5.1.12) then tells us that the total amplitude for the particle's arrival at (x, t) is the sum of the contributions from all points x' with a weight proportional to the initial amplitude $\psi(x', t')$ that the particle was at x' at time t' . One also refers to $U(x, t; x'_0, t')$ as the "fate" of the delta function $\psi(x', t') = \delta(x' - x'_0)$.

Time Evolution of the Gaussian Packet

There is an unwritten law which says that the derivation of the free-particle propagator be followed by its application to the Gaussian packet. Let us follow this tradition.

Consider as the initial wave function the wave packet

$$\psi(x', 0) = e^{ip_0 x'/\hbar} \frac{e^{-x'^2/2\Delta^2}}{(\pi\Delta^2)^{1/4}} \quad (5.1.14)$$

This packet has mean position $\langle X \rangle = 0$, with an uncertainty $\Delta X = \Delta/2^{1/2}$, and mean momentum p_0 with uncertainty $\hbar/2^{1/2}\Delta$. By combining Eqs. (5.1.10) and (5.1.12) we get

$$\begin{aligned} \psi(x, t) &= \left[\pi^{1/2} \left(\Delta + \frac{i\hbar t}{m\Delta} \right) \right]^{-1/2} \cdot \exp \left[\frac{-(x - p_0 t/m)^2}{2\Delta^2(1 + i\hbar t/m\Delta^2)} \right] \\ &\times \exp \left[\frac{ip_0}{\hbar} \left(x - \frac{p_0 t}{2m} \right) \right] \end{aligned} \quad (5.1.15)$$

The corresponding probability density is

$$P(x, t) = \frac{1}{\pi^{1/2} (\Delta^2 + \hbar^2 t^2/m^2 \Delta^2)^{1/2}} \exp \left\{ \frac{-[x - (p_0/m)t]^2}{\Delta^2 + \hbar^2 t^2/m^2 \Delta^2} \right\} \quad (5.1.16)$$

The main features of this result are as follows:

- (1) The mean position of the particles is

$$\langle X \rangle = \frac{p_0 t}{m} = \frac{\langle P \rangle t}{m}$$

In other words, the classical relation $x = (p/m)t$ now holds between average quantities. This is just one of the consequences of the *Ehrenfest theorem* which states that the classical equations obeyed by dynamical variables will have counterparts in quantum mechanics as relations among expectation values. The theorem will be proved in the next chapter.

(2) The width of the packet grows as follows:

$$\Delta X(t) = \frac{\Delta(t)}{2^{1/2}} = \frac{\Delta}{2^{1/2}} \left(1 + \frac{\hbar^2 t^2}{m^2 \Delta^4} \right)^{1/2} \quad (5.1.17)$$

The increasing uncertainty in position is a reflection of the fact that any uncertainty in the initial velocity (that is to say, the momentum) will be reflected with passing time as a growing uncertainty in position. In the present case, since $\Delta V(0) = \Delta P(0)/m = \hbar/2^{1/2} m \Delta$, the uncertainty in X grows approximately as $\Delta X \simeq \hbar t / 2^{1/2} m \Delta$ which agrees with Eq. (5.1.17) for large times. Although we are able to understand the spreading of the wave packet in classical terms, the fact that the initial spread $\Delta V(0)$ is *unavoidable* (given that we wish to specify the position to an accuracy Δ) is a purely quantum mechanical feature.

If the particle in question were macroscopic, say of mass 1 g, and we wished to fix its initial position to within a proton width, which is approximately 10^{-13} cm, the uncertainty in velocity would be

$$\Delta V(0) \simeq \frac{\hbar}{2^{1/2} m \Delta} \simeq 10^{-14} \text{ cm/sec}$$

It would be over 300,000 years before the uncertainty $\Delta(t)$ grew to 1 millimeter! We may therefore treat a macroscopic particle classically for any reasonable length of time. This and similar questions will be taken up in greater detail in the next chapter.

Exercise 5.1.3 (Another Way to Do the Gaussian Problem). We have seen that there exists another formula for $U(t)$, namely, $U(t) = e^{-iHt/\hbar}$. For a free particle this becomes

$$U(t) = \exp \left[\frac{i}{\hbar} \left(\frac{\hbar^2 t}{2m} \frac{d^2}{dx^2} \right) \right] = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{i\hbar t}{2m} \right)^n \frac{d^{2n}}{dx^{2n}} \quad (5.1.18)$$

Consider the initial state in Eq. (5.1.14) with $p_0 = 0$, and set $\Delta = 1$, $t' = 0$:

$$\psi(x, 0) = \frac{e^{-x^2/2}}{(\pi)^{1/4}}$$

Find $\psi(x, t)$ using Eq. (5.1.18) above and compare with Eq. (5.1.15).

Hints: (1) Write $\psi(x, 0)$ as a power series:

$$\psi(x, 0) = (\pi)^{-1/4} \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{n!(2)^n}$$

(2) Find the action of a few terms

$$1, \quad \left(\frac{i\hbar t}{2m}\right)\frac{d^2}{dx^2}, \quad \frac{1}{2!}\left(\frac{i\hbar t}{2m}\frac{d^2}{dx^2}\right)^2$$

etc., on this power series.

(3) Collect terms with the same power of x .

(4) Look for the following series expansion in the coefficient of x^{2n} :

$$\left(1 + \frac{i\hbar t}{m}\right)^{-n+1/2} = 1 - (n+1/2)\left(\frac{i\hbar t}{m}\right) + \frac{(n+1/2)(n+3/2)}{2!}\left(\frac{i\hbar t}{m}\right)^2 + \dots$$

(5) Juggle around till you get the answer.

Exercise 5.1.4: A Famous Counterexample. Consider the wave function

$$\begin{aligned}\psi(x, 0) &= \sin\left(\frac{\pi x}{L}\right), & |x| \leq L/2 \\ &= 0, & |x| > L/2\end{aligned}$$

It is clear that when this function is differentiated any number of times we get another function confined to the interval $|x| \leq L/2$. Consequently the action of

$$U(t) = \exp\left[\frac{i}{\hbar}\left(\frac{\hbar^2 t}{2m}\right)\frac{d^2}{dx^2}\right]$$

on this function is to give a function confined to $|x| \leq L/2$. What about the spreading of the wave packet?

[Answer: Consider the derivatives at the boundary. We have here an example where the (exponential) operator power series doesn't converge. Notice that the convergence of an operator power series depends not just on the operator but also on the operand. So there is no paradox: if the function dies abruptly as above, so that there seems to be a paradox, the derivatives are singular at the boundary, while if it falls off continuously, the function will definitely leak out given enough time, no matter how rapid the falloff.]

Some General Features of Energy Eigenfunctions

Consider now the energy eigenfunctions in some potential $V(x)$. These obey

$$\psi'' = -\frac{2m(E-V)}{\hbar^2} \psi$$

where each prime denotes a spatial derivative. Let us ask what the continuity of $V(x)$ implies. Let us start at some point x_0 where ψ and ψ' have the values $\psi(x_0)$ and $\psi'(x_0)$. If we pretend that x is a time variable and that ψ is a particle coordinate, the problem of finding ψ everywhere else is like finding the trajectory of a particle (for all times past and future) given its position and velocity at some time and its acceleration as a function of its position and time. It is clear that if we integrate

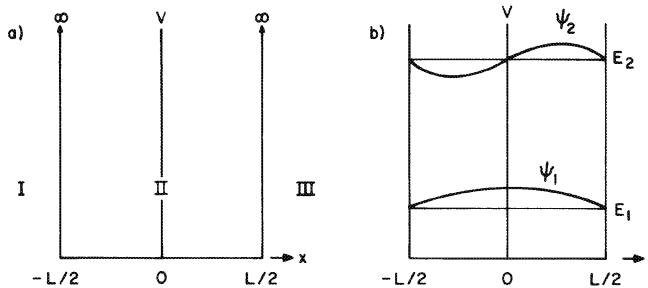


Figure 5.1. (a) The box potential. (b) The first two levels and wave functions in the box.

these equations we will get continuous $\psi'(x)$ and $\psi(x)$. This is the typical situation. There are, however, some problems where, for mathematical simplicity, we consider potentials that change abruptly at some point. This means that ψ'' jumps abruptly there. However, ψ' will still be continuous, for the area under a function is continuous even if the function jumps a bit. What if the change in V is infinitely large? It means that ψ'' is also infinitely large. This in turn means that ψ' can change abruptly as we cross this point, for the area under ψ'' can be finite over an infinitesimal region that surrounds this point. But whether or not ψ' is continuous, ψ , which is the area under it, will be continuous.[‡]

Let us turn our attention to some specific cases.

5.2. The Particle in a Box

We now consider our first problem with a potential, albeit a rather artificial one:

$$\begin{aligned} V(x) &= 0, & |x| < L/2 \\ &= \infty, & |x| \geq L/2 \end{aligned} \quad (5.2.1)$$

This potential (Fig. 5.1a) is called the box since there is an infinite potential barrier in the way of a particle that tries to leave the region $|x| < L/2$. The eigenvalue equation in the X basis (which is the only viable choice) is

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2} (E - V) \psi = 0 \quad (5.2.2)$$

We begin by partitioning space into three regions I, II, and III (Fig. 5.1a). The solution ψ is called ψ_I , ψ_{II} , and ψ_{III} in regions I, II, and III, respectively.

Consider first region III, in which $V = \infty$. It is convenient to first consider the case where V is not infinite but equal to some V_0 which is greater than E . Now

[‡] We are assuming that the jump in ψ' is finite. This will be true even in the artificial potentials we will encounter. But can you think of a potential for which this is not true? (Think delta.)

Eq. (5.2.2) becomes

$$\frac{d^2\psi_{\text{III}}}{dx^2} - \frac{2m(V_0 - E)}{\hbar^2} \psi_{\text{III}} = 0 \quad (5.2.3)$$

which is solved by

$$\psi_{\text{III}} = A e^{-\kappa x} + B e^{\kappa x} \quad (5.2.4)$$

where $\kappa = [2m(V_0 - E)/\hbar^2]^{1/2}$.

Although A and B are arbitrary coefficients from a mathematical standpoint, we must set $B = 0$ on physical grounds since $B e^{\kappa x}$ blows up exponentially as $x \rightarrow \infty$ and such functions are not members of our Hilbert space. If we now let $V \rightarrow \infty$, we see that

$$\psi_{\text{III}} \equiv 0$$

It can similarly be shown that $\psi_I \equiv 0$. In region II, since $V = 0$, the solutions are exactly those of a free particle:

$$\psi_{\text{II}} = A \exp[i(2mE/\hbar^2)^{1/2}x] + B \exp[-i(2mE/\hbar^2)^{1/2}x] \quad (5.2.5)$$

$$= A e^{ikx} + B e^{-ikx}, \quad k = (2mE/\hbar^2)^{1/2} \quad (5.2.6)$$

It therefore appears that the energy eigenvalues are once again continuous as in the free-particle case. This is not so, for $\psi_{\text{II}}(x) = \psi$ only in region II and not in all of space. We must require that ψ_{II} goes continuously into its counterparts ψ_I and ψ_{III} as we cross over to regions I and III, respectively. In other words we require that

$$\psi_I(-L/2) = \psi_{\text{II}}(-L/2) = 0 \quad (5.2.7)$$

$$\psi_{\text{III}}(+L/2) = \psi_{\text{II}}(+L/2) = 0 \quad (5.2.8)$$

(We make no such continuity demands on ψ' at the walls of the box since V jumps to infinity there.) These constraints applied to Eq. (5.2.6) take the form

$$A e^{-ikL/2} + B e^{ikL/2} = 0 \quad (5.2.9a)$$

$$A e^{ikL/2} + B e^{-ikL/2} = 0 \quad (5.2.9b)$$

or in matrix form

$$\begin{bmatrix} e^{-ikL/2} & e^{ikL/2} \\ e^{ikL/2} & e^{-ikL/2} \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (5.2.10)$$

Such an equation has nontrivial solutions only if the determinant vanishes:

159

$$e^{-ikL} - e^{ikL} = -2i \sin(kL) = 0 \quad (5.2.11)$$

SIMPLE
PROBLEMS IN
ONE DIMENSION

that is, only if

$$k = \frac{n\pi}{L}, \quad n = 0, \pm 1, \pm 2, \dots \quad (5.2.12)$$

To find the corresponding eigenfunctions, we go to Eqs. (5.2.9a) and (5.2.9b). Since only one of them is independent, we study just Eq. (5.2.9a), which says

$$A e^{-in\pi/2} + B e^{in\pi/2} = 0 \quad (5.2.13)$$

Multiplying by $e^{in\pi/2}$, we get

$$A = -e^{in\pi} B \quad (5.2.14)$$

Since $e^{in\pi} = (-1)^n$, Eq. (5.2.6) generates two families of solutions (normalized to unity):

$$\psi_n(x) = \left(\frac{2}{L}\right)^{1/2} \sin\left(\frac{n\pi x}{L}\right), \quad n \text{ even} \quad (5.2.15)$$

$$= \left(\frac{2}{L}\right)^{1/2} \cos\left(\frac{n\pi x}{L}\right), \quad n \text{ odd} \quad (5.2.16)$$

Notice that the case $n=0$ is uninteresting since $\psi_0 \equiv 0$. Further, since $\psi_n = \psi_{-n}$ for n odd and $\psi_n = -\psi_{-n}$ for n even, and since eigenfunctions differing by an overall factor are not considered distinct, we may restrict ourselves to positive nonzero n . In summary, we have

$$\psi_n = \left(\frac{2}{L}\right)^{1/2} \cos\left(\frac{n\pi x}{L}\right), \quad n = 1, 3, 5, 7, \dots \quad (5.2.17a)$$

$$= \left(\frac{2}{L}\right)^{1/2} \sin\left(\frac{n\pi x}{L}\right), \quad n = 2, 4, 6, \dots \quad (5.2.17b)$$

and from Eqs. (5.2.6) and (5.2.12),

$$E_n = \frac{\hbar^2 k_n^2}{2m} = \frac{\hbar^2 \pi^2 n^2}{2m L^2} \quad (5.2.17c)$$

[It is tacitly understood in Eqs. (5.2.17a) and (5.2.17b) that $|x| < L/2$.]

We have here our first encounter with the quantization of a dynamical variable. Both the variables considered so far, X and P , had a continuous spectrum of eigenvalues from $-\infty$ to $+\infty$, which coincided with the allowed values in classical mechanics. In fact, so did the spectrum of the Hamiltonian in the free-particle case. The particle in the box is the simplest example of a situation that will be encountered again and again, wherein Schrödinger's equation, combined with appropriate boundary conditions, leads to the quantization of energy. These solutions are also examples of *bound states*, namely, states in which a potential prevents a particle from escaping to infinity. Bound states are thus characterized by

$$\psi(x) \xrightarrow[|x| \rightarrow \infty]{} 0$$

Bound states appear in quantum mechanics exactly where we expect them classically, namely, in situations where $V(\pm\infty)$ is greater than E .

The energy levels of bound states are always quantized. Let us gain some insight into how this happens. In the problem of the particle in a box, quantization resulted from the requirement that ψ_{II} completed an integral number of half-cycles within the box so that it smoothly joined its counterparts ψ_I and ψ_{III} which vanished identically. Consider next a particle bound by a finite well, i.e., by a potential that jumps from 0 to V_0 at $|x|=L/2$. We have already seen [Eq. (5.2.4)] that in the classically forbidden region ($E < V_0$, $|x| \geq L/2$) ψ is a sum of rising and falling exponentials (as $|x| \rightarrow \infty$) and that we must choose the coefficient of the rising exponential to be zero to get an admissible solution. In the classically allowed region ($|x| \leq L/2$) ψ is a sum of a sine and cosine. Since V is everywhere finite, we demand that ψ and ψ' be continuous at $x = \pm L/2$. Thus we impose four conditions on ψ , which has only three free parameters. (It may seem that there are four—the coefficients of the two falling exponentials, the sine, and the cosine. However, the overall scale of ψ is irrelevant both in the eigenvalue equation and the continuity conditions, these being *linear* in ψ and ψ' . Thus if say, ψ' does not satisfy the continuity condition at $x=L/2$, an overall rescaling of ψ and ψ' will not help.) Clearly, the continuity conditions cannot be fulfilled except possibly at certain special energies. (See Exercise 5.2.6 for details). This is the origin of energy quantization here.

Consider now a general potential $V(x)$ which tends to limits V_{\pm} as $x \rightarrow \pm\infty$ and which binds a particle of energy E (less than both V_{\pm}). We argue once again that we have one more constraint than we have parameters, as follows. Let us divide space into tiny intervals such that in each interval $V(x)$ is essentially constant. As $x \rightarrow \pm\infty$, these intervals can be made longer and longer since V is stabilizing at its asymptotic values V_{\pm} . The right- and leftmost intervals can be made infinitely wide, since by assumption V has a definite limit as $x \rightarrow \pm\infty$. Now in all the finite intervals, ψ has two parameters: these will be the coefficients of the sine/cosine if $E > V$ or growing/falling exponential if $E < V$. (The rising exponential is not disallowed, since it doesn't blow up within the finite intervals.) Only in the left- and rightmost intervals does ψ have just one parameter, for in these infinite intervals, the growing exponential can blow up. All these parameters are constrained by the continuity of ψ and ψ' at each interface between adjacent regions. To see that we have one more constraint than we have parameters, observe that every extra interval brings with it two free parameters and one new interface, i.e., two new constraints. Thus as we go from

three intervals in the finite well to the infinite number of intervals in the arbitrary potential, the constraints are always one more than the free parameters. Thus only at special energies can we expect an allowed solution.

[Later we will study the oscillator *potential*, $V = \frac{1}{2}m\omega^2x^2$, which grows without limit as $|x| \rightarrow \infty$. How do we understand energy quantization here? Clearly, any *allowed* ψ will vanish even more rapidly than before as $|x| \rightarrow \infty$, since $V - E$, instead of being a constant, grows quadratically, so that the particle is “even more forbidden than before” from escaping to infinity. If E is an allowed energy,[‡] we expect ψ to fall off rapidly as we cross the classical turning points $x_0 = \pm(2E/m\omega^2)^{1/2}$. To a particle in such a state, it shouldn’t matter if we flatten out the potential to some constant at distances much greater than $|x_0|$, i.e., the *allowed* levels and eigenfunctions must be the same in the two potentials which differ only in a region that the particle is so strongly inhibited from going to. Since the flattened-out potential has the asymptotic behavior we discussed earlier, we can understand energy quantization as we did before.]

Let us restate the origin of energy quantization in another way. Consider the search for acceptable energy eigenfunctions, taking the finite well as an example. If we start with some arbitrary values $\psi(x_0)$ and $\psi'(x_0)$, at some point x_0 to the right of the well, we can integrate Schrödinger’s equation numerically. (Recall the analogy with the problem of finding the trajectory of a particle given its initial position and velocity and the force on it.) As we integrate out to $x \rightarrow \infty$, ψ will surely blow up since ψ_{III} contains a growing exponential. Since $\psi(x_0)$ merely fixes the overall scale, we vary $\psi'(x_0)$ until the growing exponential is killed. [Since we can solve the problem analytically in region III, we can even say what the desired value of $\psi'(x_0)$ is: it is given by $\psi'(x_0) = -\kappa\psi(x_0)$. Verify, starting with Eq. (5.2.4), that this implies $B = 0$.] We are now out of the fix as $x \rightarrow \infty$, but we are committed to whatever comes out as we integrate to the left of x_0 . We will find that ψ grows exponentially till we reach the well, whereupon it will oscillate. When we cross the well, ψ will again start to grow exponentially, for ψ_1 also contains a growing exponential in general. Thus there will be no acceptable solution at some randomly chosen energy. It can, however, happen that for certain values of energy, ψ will be exponentially damped in both regions I and III. [At any point x'_0 in region I, there is a ratio $\psi'(x'_0)/\psi(x'_0)$ for which only the damped exponential survives. The ψ we get integrating from region III will not generally have this feature. At special energies, however, this can happen.] These are the allowed energies and the corresponding functions are the allowed eigenfunctions. Having found them, we can choose $\psi(x_0)$ such that they are normalized to unity. For a nice numerical analysis of this problem see the book by Eisberg and Resnick.[§]

It is clear how these arguments generalize to a particle bound by some arbitrary potential: if we try to keep ψ exponentially damped as $x \rightarrow -\infty$, it blows up as $x \rightarrow \infty$ (and vice versa), except at some special energies. It is also clear why there is no quantization of energy for unbound states: since the particle is classically allowed at infinity, ψ oscillates there and so we have two more parameters, one from each end (why?), and so two solutions (normalizable to $\delta(0)$) at *any* energy.

[‡] We are not assuming E is quantized.

[§] R. Eisberg and R. Resnick, *Quantum Physics of Atoms, Molecules, Solids, Nuclei and Particles*, Wiley, New York (1974). See Section 5.7 and Appendix F.

Let us now return to the problem of the particle in a box and discuss the fact that the lowest energy is not zero (as it would be classically, corresponding to the particle at rest inside the well) but $\hbar^2\pi^2/2mL^2$. The reason behind it is the uncertainty principle, which prevents the particle, whose position (and hence ΔX) is bounded by $|x| \leq L/2$, from having a well-defined momentum of zero. This in turn leads to a lower bound on the energy, which we derive as follows. We begin with[‡]

$$H = \frac{P^2}{2m} \quad (5.2.18)$$

so that

$$\langle H \rangle = \frac{\langle P^2 \rangle}{2m} \quad (5.2.19)$$

Now $\langle P \rangle = 0$ in any bound state for the following reason. Since a bound state is a stationary state, $\langle P \rangle$ is time independent. If this $\langle P \rangle \neq 0$, the particle must (in the average sense) drift either to the right or to the left and eventually escape to infinity, which cannot happen in a bound state.

Consequently we may rewrite Eq. (5.2.19) as

$$\langle H \rangle = \frac{\langle (P - \langle P \rangle)^2 \rangle}{2m} = \frac{(\Delta P)^2}{2m}$$

If we now use the uncertainty relation

$$\Delta P \cdot \Delta X \geq \hbar/2$$

we find

$$\langle H \rangle \geq \frac{\hbar^2}{8m(\Delta X)^2}$$

Since the variable x is constrained by $-L/2 \leq x \leq L/2$, its standard deviation ΔX cannot exceed $L/2$. Consequently

$$\langle H \rangle \geq \hbar^2/2mL^2$$

In an energy eigenstate, $\langle H \rangle = E$ so that

$$E \geq \hbar^2/2mL^2 \quad (5.2.20)$$

The actual ground-state energy E_1 happens to be π^2 times as large as the lower

[‡] We are suppressing the infinite potential due to the walls of the box. Instead we will restrict x to the range $|x| \leq L/2$.

bound. The uncertainty principle is often used in this fashion to provide a quick order-of-magnitude estimate for the ground-state energy.

If we denote by $|n\rangle$ the abstract ket corresponding to $\psi_n(x)$, we can write the propagator as

$$U(t) = \sum_{n=1}^{\infty} |n\rangle \langle n| \exp\left[-\frac{i}{\hbar}\left(\frac{\hbar^2\pi^2n^2}{2mL^2}\right)t\right] \quad (5.2.21)$$

The matrix elements of $U(t)$ in the X basis are then

$$\begin{aligned} \langle x| U(t)|x'\rangle &= U(x, t; x') \\ &= \sum_{n=1}^{\infty} \psi_n(x)\psi_n^*(x') \exp\left[-\frac{i}{\hbar}\left(\frac{\hbar^2\pi^2n^2}{2mL^2}\right)t\right] \end{aligned} \quad (5.2.22)$$

Unlike in the free-particle case, there exists no simple closed expression for this sum.

*Exercise 5.2.1.** A particle is in the ground state of a box of length L . Suddenly the box expands (symmetrically) to twice its size, leaving the wave function undisturbed. Show that the probability of finding the particle in the ground state of the new box is $(8/3\pi)^2$.

*Exercise 5.2.2.** (a) Show that for any normalized $|\psi\rangle$, $\langle\psi|H|\psi\rangle \geq E_0$, where E_0 is the lowest-energy eigenvalue. (Hint: Expand $|\psi\rangle$ in the eigenbasis of H .)

(b) Prove the following theorem: Every attractive potential in one dimension has at least one bound state. Hint: Since V is attractive, if we define $V(\infty)=0$, it follows that $V(x)=-|V(x)|$ for all x . To show that there exists a bound state with $E < 0$, consider

$$\psi_\alpha(x) = \left(\frac{\alpha}{\pi}\right)^{1/4} e^{-\alpha x^2/2}$$

and calculate

$$E(\alpha) = \langle\psi_\alpha|H|\psi_\alpha\rangle, \quad H = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} - |V(x)|$$

Show that $E(\alpha)$ can be made negative by a suitable choice of α . The desired result follows from the application of the theorem proved above.

*Exercise 5.2.3.** Consider $V(x) = -aV_0\delta(x)$. Show that it admits a bound state of energy $E = -ma^2V_0^2/2\hbar^2$. Are there any other bound states? Hint: Solve Schrödinger's equation outside the potential for $E < 0$, and keep only the solution that has the right behavior at infinity and is continuous at $x=0$. Draw the wave function and see how there is a cusp, or a discontinuous change of slope at $x=0$. Calculate the change in slope and equate it to

$$\int_{-\varepsilon}^{+\varepsilon} \left(\frac{d^2\psi}{dx^2}\right) dx$$

(where ε is infinitesimal) determined from Schrödinger's equation.

Exercise 5.2.4. Consider a particle of mass m in the state $|n\rangle$ of a box of length L . Find the force $F = -\partial E/\partial L$ encountered when the walls are slowly pushed in, assuming the particle remains in the n th state of the box as its size changes. Consider a classical particle of energy E_n in this box. Find its velocity, the frequency of collision on a given wall, the momentum transfer per collision, and hence the average force. Compare it to $-\partial E/\partial L$ computed above.

*Exercise 5.2.5.** If the box extends from $x=0$ to L (instead of $-L/2$ to $L/2$) show that $\psi_n(x) = (2/L)^{1/2} \sin(n\pi x/L)$, $n=1, 2, \dots, \infty$ and $E_n = \hbar^2\pi^2 n^2 / 2mL^2$.

Exercise 5.2.6. Square Well Potential.* Consider a particle in a square well potential:

$$V(x) = \begin{cases} 0, & |x| \leq a \\ V_0, & |x| \geq a \end{cases}$$

Since when $V_0 \rightarrow \infty$, we have a box, let us guess what the lowering of the walls does to the states. First of all, all the bound states (which alone we are interested in), will have $E \leq V_0$. Second, the wave functions of the low-lying levels will look like those of the particle in a box, with the obvious difference that ψ will not vanish at the walls but instead spill out with an exponential tail. The eigenfunctions will still be even, odd, even, etc.

- (1) Show that the even solutions have energies that satisfy the transcendental equation

$$k \tan ka = \kappa \quad (5.2.23)$$

while the odd ones will have energies that satisfy

$$k \cot ka = -\kappa \quad (5.2.24)$$

where k and $i\kappa$ are the real and complex wave numbers inside and outside the well, respectively. Note that k and κ are related by

$$k^2 + \kappa^2 = 2mV_0/\hbar^2 \quad (5.2.25)$$

Verify that as V_0 tends to ∞ , we regain the levels in the box.

(2) Equations (5.2.23) and (5.2.24) must be solved graphically. In the $(\alpha = ka, \beta = \kappa a)$ plane, imagine a circle that obeys Eq. (5.2.25). The bound states are then given by the intersection of the curve $\alpha \tan \alpha = \beta$ or $\alpha \cot \alpha = -\beta$ with the circle. (Remember α and β are positive.)

(3) Show that there is always one even solution and that there is no odd solution unless $V_0 \geq \hbar^2\pi^2/8ma^2$. What is E when V_0 just meets this requirement? Note that the general result from Exercise 5.2.2b holds.

5.3. The Continuity Equation for Probability

We interrupt our discussion of one-dimensional problems to get acquainted with two concepts that will be used in the subsequent discussions, namely, those of the *probability current density* and the *continuity equation* it satisfies. Since the probability current concept will also be used in three-dimensional problems, we discuss here a particle in three dimensions.

As a prelude to our study of the continuity equation in quantum mechanics, let us recall the analogous equation from electromagnetism. We know in this case that the total charge in the universe is a constant, that is

$$Q(t) = \text{const, independent of time } t \quad (5.3.1)$$

This is an example of a global conservation law, for it refers to the total charge in the universe. But charge is also conserved locally, a fact usually expressed in the form of the continuity equation

$$\frac{\partial \rho(\mathbf{r}, t)}{\partial t} = -\nabla \cdot \mathbf{j} \quad (5.3.2)$$

where ρ and \mathbf{j} are the charge and current densities, respectively. By integrating this equation over a volume V bounded by a surface S_V we get, upon invoking Gauss's law,

$$\frac{d}{dt} \int_V \rho(\mathbf{r}, t) d^3\mathbf{r} = - \int_V \nabla \cdot \mathbf{j} d^3\mathbf{r} = - \int_{S_V} \mathbf{j} \cdot d\mathbf{S} \quad (5.3.3)$$

This equation states that any decrease in charge in the volume V is accounted for by the flow of charge out of it, that is to say, charge is not created or destroyed in any volume.

The continuity equation forbids certain processes that obey global conservation, such as the sudden disappearance of charge from one region of space and its immediate reappearance in another.

In quantum mechanics the quantity that is globally conserved is the total probability for finding the particle anywhere in the universe. We get this result by expressing the invariance of the norm in the coordinate basis: since

$$\langle \psi(t) | \psi(t) \rangle = \langle \psi(0) | U^\dagger(t) U(t) | \psi(0) \rangle = \langle \psi(0) | \psi(0) \rangle$$

then

$$\begin{aligned} \text{const} &= \langle \psi(t) | \psi(t) \rangle = \iiint \langle \psi(t) | x, y, z \rangle \langle x, y, z | \psi(t) \rangle dx dy dz \dagger \\ &= \iiint \langle \psi(t) | \mathbf{r} \rangle \langle \mathbf{r} | \psi(t) \rangle d^3\mathbf{r} \\ &= \iiint \psi^*(\mathbf{r}, t) \psi(\mathbf{r}, t) d^3\mathbf{r} \\ &= \iiint P(\mathbf{r}, t) d^3\mathbf{r} \end{aligned} \quad (5.3.4)$$

\dagger The range of integration will frequently be suppressed when obvious.

This global conservation law is the analog of Eq. (5.3.1). To get the analog of Eq. (5.3.2), we turn to the Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi + V\psi \quad (5.3.5)$$

and its conjugate

$$-i\hbar \frac{\partial \psi^*}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi^* + V\psi^* \quad (5.3.6)$$

Note that V has to be real if H is to be Hermitian. Multiplying the first of these equations by ψ^* , the second by ψ , and taking the difference, we get

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} (\psi^* \psi) &= -\frac{\hbar^2}{2m} (\psi^* \nabla^2 \psi - \psi \nabla^2 \psi^*) \\ \frac{\partial P}{\partial t} &= -\frac{\hbar}{2mi} \nabla \cdot (\psi^* \nabla \psi - \psi \nabla \psi^*) \\ \frac{\partial P}{\partial t} &= -\nabla \cdot \mathbf{j} \end{aligned} \quad (5.3.7)$$

where

$$\mathbf{j} = \frac{\hbar}{2mi} (\psi^* \nabla \psi - \psi \nabla \psi^*) \quad (5.3.8)$$

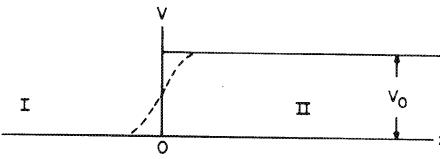
is the *probability current density*, that is to say, the probability flow per unit time per unit area perpendicular to \mathbf{j} . To regain the global conservation law, we integrate Eq. (5.3.7) over all space:

$$\frac{d}{dt} \int P(\mathbf{r}, t) d^3\mathbf{r} = - \int_{S_\infty} \mathbf{j} \cdot d\mathbf{S} \quad (5.3.9)$$

where S_∞ is the sphere at infinity. For (typical) wave functions which are normalizable to unity, $r^{3/2}\psi \rightarrow 0$ as $r \rightarrow \infty$ in order that $\int \psi^* \psi r^2 dr d\Omega$ is bounded, and the surface integral of \mathbf{j} on S_∞ vanishes. The case of momentum eigenfunctions that do not vanish on S_∞ is considered in one of the following exercises.

Exercise 5.3.1. Consider the case where $V = V_r - iV_i$, where the imaginary part V_i is a constant. Is the Hamiltonian Hermitian? Go through the derivation of the continuity equation and show that the total probability for finding the particle decreases exponentially as $e^{-2V_i t/\hbar}$. Such complex potentials are used to describe processes in which particles are absorbed by a sink.

Figure 5.2. The single-step potential. The dotted line shows a more realistic potential idealized by the step, which is mathematically convenient. The total energy E and potential energy V are measured along the y axis.



Exercise 5.3.2. Convince yourself that if $\psi = c\tilde{\psi}$, where c is constant (real or complex) and $\tilde{\psi}$ is real, the corresponding \mathbf{j} vanishes.

Exercise 5.3.3. Consider

$$\psi_p = \left(\frac{1}{2\pi\hbar} \right)^{3/2} e^{i(\mathbf{p}\cdot\mathbf{r})/\hbar}$$

Find \mathbf{j} and P and compare the relation between them to the electromagnetic equation $\mathbf{j} = \rho\mathbf{v}$, \mathbf{v} being the velocity. Since ρ and \mathbf{j} are constant, note that the continuity Eq. (5.3.7) is trivially satisfied.

*Exercise 5.3.4.** Consider $\psi = A e^{ipx/\hbar} + B e^{-ipx/\hbar}$ in one dimension. Show that $j = (|A|^2 - |B|^2)p/m$. The absence of cross terms between the right- and left-moving pieces in ψ allows us to associate the two parts of j with corresponding parts of ψ .

Ensemble Interpretation of \mathbf{j}

Recall that $\mathbf{j} \cdot d\mathbf{S}$ is the rate at which probability flows past the area $d\mathbf{S}$. If we consider an ensemble of N particles all in some state $\psi(\mathbf{r}, t)$, then $N\mathbf{j} \cdot d\mathbf{S}$ particles will trigger a particle detector of area $d\mathbf{S}$ per second, assuming that N tends to infinity and that \mathbf{j} is the current associated with $\psi(\mathbf{r}, t)$.

5.4. The Single-Step Potential: A Problem in Scattering‡

Consider the step potential (Fig. 5.2)

$$\begin{aligned} V(x) &= 0 & x < 0 & \text{(region I)} \\ &= V_0 & x > 0 & \text{(region II)} \end{aligned} \tag{5.4.1}$$

Such an abrupt change in potential is rather unrealistic but mathematically convenient. A more realistic transition is shown by dotted lines in the figure.

Imagine now that a classical particle of energy E is shot in from the left (region I) toward the step. One expects that if $E > V_0$, the particle would climb the barrier and travel on to region II, while if $E < V_0$, it would get reflected. We now compare this classical situation with its quantum counterpart.

‡ This rather difficult section may be postponed till the reader has gone through Chapter 7 and gained more experience with the subject. It is for the reader or the instructor to decide which way to go.

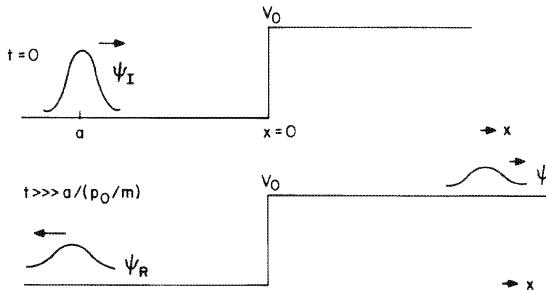


Figure 5.3. A schematic description of the wave function long before and long after it hits the step. The area under $|\psi_I|^2$ is unity. The areas under $|\psi_R|^2$ and $|\psi_T|^2$, respectively, are the probabilities for reflection and transmission.

First of all, we must consider an initial state that is compatible with quantum principles. We replace the incident particle possessing a well-defined trajectory with a wave packet.[‡] Though the detailed wave function will be seen to be irrelevant in the limit we will consider, we start with a Gaussian, which is easy to handle analytically[§]:

$$\psi_I(x, 0) = \psi_I(x) = (\pi\Delta^2)^{-1/4} e^{ik_0(x+a)} e^{-(x+a)^2/2\Delta^2} \quad (5.4.2)$$

This packet has a mean momentum $p_0 = \hbar k_0$, a mean position $\langle X \rangle = -a$ (which we take to be far away from the step), with uncertainties

$$\Delta X = \frac{\Delta}{2^{1/2}}, \quad \Delta P = \frac{\hbar}{2^{1/2}\Delta}$$

We shall be interested in the case of large Δ , where the particle has essentially well-defined momentum $\hbar k_0$ and energy $E_0 \simeq \hbar^2 k_0^2 / 2m$. We first consider the case $E_0 > V_0$.

After a time $t \simeq a[p_0/m]^{-1}$, the packet will hit the step and in general break into two packets: ψ_R , the reflected packet, and ψ_T , the transmitted packet (Fig. 5.3). The area under $|\psi_R|^2$ at large t is the probability of finding the particle in region I in the distant future, that is to say, the probability of reflection. Likewise the area under $|\psi_T|^2$ at large t is the probability of transmission. Our problem is to calculate the *reflection coefficient*

$$R = \int |\psi_R|^2 dx, \quad t \rightarrow \infty \quad (5.4.3)$$

and *transmission coefficient*

$$T = \int |\psi_T|^2 dx, \quad t \rightarrow \infty \quad (5.4.4)$$

Generally R and T will depend on the detailed shape of the initial wave function. If, however, we go to the limit in which the initial momentum is well defined (i.e.,

[‡] A wave packet is any wave function with reasonably well-defined position and momentum.

[§] This is just the wave packet in Eq. (5.1.14), displaced by an amount $-a$.

when the Gaussian in x space has infinite width), we expect the answer to depend only on the initial energy, it being the only characteristic of the state. In the following analysis we will assume that $\Delta X = \Delta/2^{1/2}$ is large and that the wave function in k space is very sharply peaked near k_0 .

We follow the standard procedure for finding the fate of the incident wave packet, ψ_I :

Step 1: Solve for the *normalized* eigenfunction of the step potential Hamiltonian, $\psi_E(x)$.

Step 2: Find the projection $a(E) = \langle \psi_E | \psi_I \rangle$.

Step 3: Append to each coefficient $a(E)$ a time dependence $e^{-iEt/\hbar}$ and get $\psi(x, t)$ at any future time.

Step 4: Identify ψ_R and ψ_T in $\psi(x, t \rightarrow \infty)$ and determine R and T using Eqs. (5.4.3) and (5.4.4).

Step 1. In region I, as $V=0$, the (*unnormalized*) solution is the familiar one:

$$\psi_E(x) = A e^{ik_1 x} + B e^{-ik_1 x}, \quad k_1 = \left(\frac{2mE}{\hbar^2} \right)^{1/2} \quad (5.4.5)$$

In region II, we simply replace E by $E - V_0$ [see Eq. (5.2.2)],

$$\psi_E(x) = C e^{ik_2 x} + D e^{-ik_2 x}, \quad k_2 = \left[\frac{2m(E - V_0)}{\hbar^2} \right]^{1/2} \quad (5.4.6)$$

(We consider only $E > V_0$; the eigenfunction with $E < V_0$ will be orthogonal to ψ_I as will be shown on the next two pages.) Of interest to us are eigenfunctions with $D = 0$, since we want only a transmitted (right-going) wave in region II, and incident plus reflected waves in region I. If we now impose the continuity of ψ and its derivative at $x=0$; we get

$$A + B = C \quad (5.4.7)$$

$$ik_1(A - B) = ik_2C \quad (5.4.8)$$

In anticipation of future use, we solve these equations to express B and C in terms of A :

$$B = \left(\frac{k_1 - k_2}{k_1 + k_2} \right) A = \left(\frac{E^{1/2} - (E - V_0)^{1/2}}{E^{1/2} + (E - V_0)^{1/2}} \right) A \quad (5.4.9)$$

$$C = \left(\frac{2k_1}{k_1 + k_2} \right) A = \left(\frac{2E^{1/2}}{E^{1/2} + (E - V_0)^{1/2}} \right) A \quad (5.4.10)$$

Note that if $V_0=0$, $B=0$ and $C=A$ as expected. The solution with energy E is then

$$\psi_E(x) = A \left[\left(e^{ik_1 x} + \frac{B}{A} e^{-ik_1 x} \right) \theta(-x) + \frac{C}{A} e^{ik_2 x} \theta(x) \right] \quad (5.4.11)$$

where

$$\begin{aligned} \theta(x) &= 1 && \text{if } x > 0 \\ &= 0 && \text{if } x < 0 \end{aligned}$$

Since to each E there is a unique $k_1 = +(\Delta^2/\hbar^2)^{1/2}$, we can label the eigenstates by k_1 instead of E . Eliminating k_2 in favor of k_1 , we get

$$\begin{aligned} \psi_{k_1}(x) &= A \left[\left(\exp(ik_1 x) + \frac{B}{A} \exp(-ik_1 x) \right) \theta(-x) \right. \\ &\quad \left. + \frac{C}{A} \exp[i(k_1^2 - 2mV_0/\hbar^2)^{1/2}x] \theta(x) \right] \end{aligned} \quad (5.4.12)$$

Although the overall scale factor A is generally arbitrary (and the physics depends only on B/A and C/A), here we must choose $A=(2\pi)^{-1/2}$ because ψ_k has to be properly normalized in the four-step procedure outlined above. We shall verify shortly that $A=(2\pi)^{-1/2}$ is the correct normalization factor.

Step 2. Consider next

$$\begin{aligned} a(k_1) &= \langle \psi_{k_1} | \psi_I \rangle \\ &= \frac{1}{(2\pi)^{1/2}} \left\{ \int_{-\infty}^{\infty} \left[e^{-ik_1 x} + \left(\frac{B}{A} \right)^* e^{ik_1 x} \right] \theta(-x) \psi_I(x) dx \right. \\ &\quad \left. + \int_{-\infty}^{\infty} \left(\frac{C}{A} \right)^* e^{-ik_2 x} \theta(x) \psi_I(x) dx \right\} \end{aligned} \quad (5.4.13)$$

The second integral vanishes (to an excellent approximation) since $\psi_I(x)$ is nonvanishing far to the left of $x=0$, while $\theta(x)$ is nonvanishing only for $x>0$. Similarly the second piece of the first integral also vanishes since ψ_I in k space is peaked around $k=+k_0$ and is orthogonal to (left-going) negative momentum states. [We can ignore the $\theta(-x)$ factor in Eq. (5.4.13) since it equals 1 where $\psi_I(x)\neq 0$.] So

$$\begin{aligned} a(k_1) &= \left(\frac{1}{2\pi} \right)^{1/2} \int_{-\infty}^{\infty} e^{-ik_1 x} \psi_I(x) dx \\ &= \left(\frac{\Delta^2}{\pi} \right)^{1/4} e^{-(k_1 - k_0)^2 \Delta^2/2} e^{ik_1 a} \end{aligned} \quad (5.4.14)$$

is just the Fourier transform of ψ_I . Notice that for large Δ , $a(k_1)$ is very sharply peaked at $k_1 = k_0$. This justifies our neglect of eigenfunctions with $E < V_0$, for these correspond to k_1 not near k_0 .

Step 3. The wave function at any future time t is

$$\psi(x, t) = \int_{-\infty}^{\infty} a(k_1) e^{-iE(k_1)t/\hbar} \psi_{k_1}(x) dk_1 \quad (5.4.15)$$

$$\begin{aligned} &= \left(\frac{\Delta^2}{4\pi^3} \right)^{1/4} \int_{-\infty}^{\infty} \exp\left(\frac{-i\hbar k_1^2 t}{2m}\right) \cdot \exp\left[\frac{-(k_1 - k_0)^2 \Delta^2}{2}\right] \exp(ik_1 a) \\ &\times \left\{ e^{ik_1 x} \theta(-x) + \left(\frac{B}{A} \right) e^{-ik_1 x} \theta(-x) \right. \\ &\left. + \left(\frac{C}{A} \right) \exp[i(k_1^2 - 2mV_0/\hbar^2)^{1/2} x] \theta(x) \right\} dk_1 \end{aligned} \quad (5.4.16)$$

You can convince yourself that if we set $t = 0$ above we regain $\psi_I(x)$, which corroborates our choice $A = (2\pi)^{-1/2}$.

Step 4. Consider the first of the three terms. If $\theta(-x)$ were absent, we would be propagating the original Gaussian. After replacing x by $x + a$ in Eq. (5.1.15), and inserting the $\theta(-x)$ factor, the first term of $\psi(x, t)$ is

$$\begin{aligned} &\theta(-x) \pi^{-1/4} \left(\Delta + \frac{i\hbar t}{m} \right)^{-1/2} \exp\left[\frac{-(x + a - \hbar k_0 t / m)^2}{2\Delta^2(1 + i\hbar t / m\Delta^2)} \right] \\ &\times \exp\left[ik_0 \left(x + a - \frac{\hbar k_0 t}{2m} \right) \right] \equiv \theta(-x) G(-a, k_0, t) \end{aligned} \quad (5.4.17)$$

Since the Gaussian $G(-a, k_0, t)$ is centered at $x = -a + \hbar k_0 t / m \simeq \hbar k_0 t / m$ as $t \rightarrow \infty$, and $\theta(-x)$ vanishes for $x > 0$, the product θG vanishes. Thus the initial packet has disappeared and in its place are the reflected and transmitted packets given by the next two terms. In the middle term if we replace B/A , which is a function of k_1 , by its value $(B/A)_0$ at $k_1 = k_0$ (because $a(k_1)$ is very sharply peaked at $k_1 = k_0$) and pull it out of the integral, changing the dummy variable from k_1 to $-k_1$, it is easy to see that apart from the factor $(B/A)_0 \theta(-x)$ up front, the middle term represents the free propagation of a normalized Gaussian packet that was originally peaked at $x = +a$ and began drifting to the left with mean momentum $-\hbar k_0$. Thus

$$\psi_R = \theta(-x) G(a, -k_0, t) (B/A)_0 \quad (5.4.18)$$

As $t \rightarrow \infty$, we can set $\theta(-x)$ equal to 1, since G is centered at $x = a - \hbar k_0 t / m \simeq -\hbar k_0 t / m$. Since the Gaussian G has unit norm, we get from Eqs. (5.4.3) and (5.4.9),

$$R = \int |\psi_R|^2 dx = \frac{|B|^2}{|A|_0} = \frac{\left|E_0^{1/2} - (E_0 - V_0)^{1/2}\right|^2}{\left|E_0^{1/2} + (E_0 - V_0)^{1/2}\right|^2}$$

where

$$E_0 = \frac{\hbar^2 k_0^2}{2m} \quad (5.4.19)$$

This formula is exact only when the incident packet has a well-defined energy E_0 , that is to say, when the width of the incident Gaussian tends to infinity. But it is an excellent approximation for *any* wave packet that is narrowly peaked in momentum space.

To find T , we can try to evaluate the third piece. But there is no need to do so, since we know that

$$R + T = 1 \quad (5.4.20)$$

which follows from the global conservation of probability. It then follows that

$$T = 1 - R = \frac{4E_0^{1/2}(E_0 - V_0)^{1/2}}{[E_0^{1/2} + (E_0 - V_0)^{1/2}]^2} = \frac{|C|^2}{|A|_0} \frac{(E_0 - V_0)^{1/2}}{E_0^{1/2}} \quad (5.4.21)$$

By inspecting Eqs. (5.4.19) and (5.4.21) we see that both R and T are readily expressed in terms of the ratios $(B/A)_0$ and $(C/A)_0$ and a kinematical factor, $(E_0 - V_0)^{1/2}/E_0^{1/2}$. Is there some way by which we can directly get to Eqs. (5.4.19) and (5.4.21), which describe the dynamic phenomenon of scattering, from Eqs. (5.4.9) and (5.4.10), which describe the static solution to Schrödinger's equation? Yes.

Consider the unnormalized eigenstate

$$\begin{aligned} \psi_{k_0}(x) &= [A_0 \exp(ik_0 x) + B_0 \exp(-ik_0 x)]\theta(-x) \\ &\quad + C_0 \exp\left[i\left(k_0^2 - \frac{2mV_0}{\hbar^2}\right)^{1/2} x\right]\theta(x) \end{aligned} \quad (5.4.22)$$

The incoming plane wave $A e^{ik_0 x}$ has a probability current associated with it equal to

$$j_I = |A_0|^2 \frac{\hbar k_0}{m} \quad (5.4.23)$$

while the currents associated with the reflected and transmitted pieces are

$$j_R = |B_0|^2 \frac{\hbar k_0}{m} \quad (5.4.24)$$

and

$$j_T = |C_0|^2 \frac{\hbar(k_0^2 - 2mV_0/\hbar^2)^{1/2}}{m} \quad (5.4.25)$$

(Recall Exercise 5.3.4, which provides the justification for viewing the two parts of the j in region I as being due to the incident and reflected wave functions.) In terms of these currents

$$R = \frac{j_R}{j_I} = \frac{|B_0|^2}{|A_0|^2} \quad (5.4.26)$$

and

$$T = \frac{j_T}{j_I} = \frac{|C_0|^2}{|A_0|^2} \frac{(k_0^2 - 2mV_0/\hbar^2)^{1/2}}{k_0} = \frac{|C_0|^2}{|A_0|^2} \frac{(E_0 - V_0)^{1/2}}{E_0^{1/2}} \quad (5.4.27)$$

Let us now enquire as to why it is that R and T are calculable in these two ways. Recall that R and T were exact only for the incident packet whose momentum was well defined and equal to $\hbar k_0$. From Eq. (5.4.2) we see that this involves taking the width of the Gaussian to infinity. As the incident Gaussian gets wider and wider (we ignore now the $\Delta^{-1/2}$ factor up front and the normalization) the following things happen:

- (1) It becomes impossible to say when it hits the step, for it has spread out to be a right-going plane wave in region I.
- (2) The reflected packet also gets infinitely wide and coexists with the incident one, as a left-going plane wave.
- (3) The transmitted packet becomes a plane wave with wave number $(k_0^2 - 2mV_0/\hbar^2)^{1/2}$ in region II.

In other words, the dynamic picture of an incident packet hitting the step and disintegrating into two becomes the steady-state process described by the eigenfunction Eq. (5.4.22). We cannot, however, find R and T by calculating areas under $|\psi_T|^2$ and $|\psi_R|^2$ since all the areas are infinite, the wave packets having been transformed into plane waves. We find instead that the ratios of the probability currents associated with the incident, reflected, and transmitted waves give us R and T . The equivalence between the wave packet and static descriptions that we were able to demonstrate in this simple case happens to be valid for any potential. When we come to scattering in three dimensions, we will assume that the equivalence of the two approaches holds.

Exercise 5.4.1 (Quite Hard). Evaluate the third piece in Eq. (5.4.16) and compare the resulting T with Eq. (5.4.21). [Hint: Expand the factor $(k_1^2 - 2mV_0/\hbar^2)^{1/2}$ near $k_1 = k_0$, keeping just the first derivative in the Taylor series.]

Before we go on to examine some of the novel features of the reflection and transmission coefficients, let us ask how they are used in practice. Consider a general problem with some $V(x)$, which tends to constants V_+ and V_- as $x \rightarrow \pm \infty$. For simplicity we take $V_{\pm} = 0$. Imagine an accelerator located to the far left ($x \rightarrow -\infty$) which shoots out a beam of nearly monoenergetic particles with $\langle P \rangle = \hbar k_0$ toward the potential. The question one asks in practice is what fraction of the particles will get transmitted and what fraction will get reflected to $x = -\infty$, respectively. In general, the question cannot be answered because we know only the mean momenta of the particles and not their individual wave functions. But the preceding analysis shows that *as long as the wave packets are localized sharply in momentum space, the reflection and transmission probabilities (R and T) depend only on the mean momentum and not the detailed shape of the wave functions*. So the answer to the question raised above is that a fraction $R(k_0)$ will get reflected and a fraction $T(k_0) = 1 - R(k_0)$ will get transmitted. To find R and T we solve for the time-independent eigenfunctions of $H = T + V$ with energy eigenvalue $E_0 = \hbar^2 k_0^2 / 2m$, and asymptotic behavior

$$\psi_{k_0}(x) \begin{cases} \longrightarrow A e^{ik_0 x} + B e^{-ik_0 x} \\ \xrightarrow{x \rightarrow -\infty} \\ \longrightarrow C e^{ik_0 x} \\ \xrightarrow{x \rightarrow \infty} \end{cases}$$

and obtain from it $R = |B/A|^2$ and $T = |C/A|^2$. Solutions with this asymptotic behavior (namely, free-particle behavior) will always exist provided V vanishes rapidly enough as $|x| \rightarrow \infty$. [Later we will see that this means $|xV(x)| \rightarrow 0$ as $|x| \rightarrow \infty$.] The general solution will also contain a piece $D \exp(-ik_0 x)$ as $x \rightarrow \infty$, but we set $D = 0$ here, for if a $\exp(ik_0 x)$ is to be identified with the incident wave, it must only produce a right-moving transmitted wave $C e^{ik_0 x}$ as $x \rightarrow \infty$.

Let us turn to Eqs. (5.4.19) and (5.4.21) for R and T . These contain many nonclassical features. First of all we find that an incident particle with $E_0 > V_0$ gets reflected some of the time. It can also be shown that a particle with $E_0 > V_0$ incident from the right will also get reflected some of the time, contrary to classical expectations.

Consider next the case $E_0 < V_0$. Classically one expects the particle to be reflected at $x = 0$, and never to get to region II. This is not so quantum mechanically. In region II, the solution to

$$\frac{d^2 \psi_{II}}{dx^2} + \frac{2m}{\hbar^2} (E_0 - V_0) \psi_{II} = 0$$

with $E_0 < V_0$ is

$$\psi_{II}(x) = C e^{-\kappa x}, \quad \kappa = \left(\frac{2m|E_0 - V_0|}{\hbar^2} \right)^{1/2} \quad (5.4.28)$$

(The growing exponential $e^{\kappa x}$ does not belong to the physical Hilbert space.) Thus there is a finite probability for finding the particle in the region where its kinetic energy $E_0 - V_0$ is negative. There is, however, no steady flow of probability current into region II, since $\psi_{II}(x) = C\tilde{\psi}$, where $\tilde{\psi}$ is real. This is also corroborated by the fact the reflection coefficient in this case is

$$R = \left| \frac{(E_0)^{1/2} - (E_0 - V_0)^{1/2}}{(E_0)^{1/2} + (E_0 - V_0)^{1/2}} \right|^2 = \left| \frac{k_0 - i\kappa}{k_0 + i\kappa} \right|^2 = 1 \quad (5.4.29)$$

The fact that the particle can penetrate into the classically forbidden region leads to an interesting quantum phenomenon called *tunneling*. Consider a modification of Fig. 5.2, in which $V = V_0$ only between $x = 0$ and L (region II) and is once again zero beyond $x = L$ (region III). If now a plane wave is incident on this barrier from the left with $E < V_0$, there is an exponentially small probability for the particle to get to region III. Once a particle gets to region III, it is free once more and described by a plane wave. An example of tunneling is that of α particles trapped in the nuclei by a barrier. Every once in a while an α particle manages to penetrate the barrier and come out. The rate for this process can be calculated given V_0 and L .

Exercise 5.4.2. (a)* Calculate R and T for scattering of a potential $V(x) = V_0\delta(x)$. (b) Do the same for the case $V=0$ for $|x| > a$ and $V=V_0$ for $|x| < a$. Assume that the energy is positive but less than V_0 .

Exercise 5.4.3. Consider a particle subject to a constant force f in one dimension. Solve for the propagator in momentum space and get

$$U(p, t; p', 0) = \delta(p - p' - ft) e^{i(p'^3 - p^3)/6m\hbar f} \quad (5.4.30)$$

Transform back to coordinate space and obtain

$$U(x, t; x', 0) = \left(\frac{m}{2\pi\hbar^2 t} \right)^{1/2} \exp \left\{ \frac{i}{\hbar} \left[\frac{m(x-x')^2}{2t} + \frac{1}{2} ft(x+x') - \frac{f^2 t^3}{24m} \right] \right\} \quad (5.4.31)$$

[Hint: Normalize $\psi_E(p)$ such that $\langle E|E' \rangle = \delta(E-E')$. Note that E is not restricted to be positive.]

5.5. The Double-Slit Experiment

Having learned so much quantum mechanics, it now behooves us to go back and understand the double-slit experiment (Fig. 3.1). Let us label by I and II the regions to the left and right of the screen. The incident particle, which must really be represented by a wave packet, we approximate by a plane wave of wave number $k = p/\hbar$. The impermeable screen we treat as a region with $V = \infty$, and hence the region of vanishing ψ . Standard wave theory (which we can borrow from classical electromagnetism) tells us what happens in region II: the two slits act as sources of radially outgoing waves of the same wavelength. These two waves interfere on the

line AB and produce the interference pattern. We now return to quantum mechanics and interpret the intensity $|\psi|^2$ as the probability density for finding the particle.

5.6. Some Theorems

Theorem 15. There is no degeneracy in one-dimensional bound states.

Proof. Let ψ_1 and ψ_2 be two solutions with the same eigenvalue E :

$$\frac{-\hbar^2}{2m} \frac{d^2\psi_1}{dx^2} + V\psi_1 = E\psi_1 \quad (5.6.1)$$

$$\frac{-\hbar^2}{2m} \frac{d^2\psi_2}{dx^2} + V\psi_2 = E\psi_2 \quad (5.6.2)$$

Multiply the first by ψ_2 , the second by ψ_1 and subtract, to get

$$\psi_1 \frac{d^2\psi_2}{dx^2} - \psi_2 \frac{d^2\psi_1}{dx^2} = 0$$

or

$$\frac{d}{dx} \left(\psi_1 \frac{d\psi_2}{dx} - \psi_2 \frac{d\psi_1}{dx} \right) = 0$$

so that

$$\psi_1 \frac{d\psi_2}{dx} - \psi_2 \frac{d\psi_1}{dx} = c \quad (5.6.3)$$

To find the constant c , go to $|x| \rightarrow \infty$, where ψ_1 and ψ_2 vanish, since they describe bound states by assumption.[‡] It follows that $c=0$. So

$$\frac{1}{\psi_1} d\psi_1 = \frac{1}{\psi_2} d\psi_2$$

$$\log \psi_1 = \log \psi_2 + d \quad (d \text{ is a constant})$$

$$\psi_1 = e^d \psi_2 \quad (5.6.4)$$

[‡] The theorem holds even if ψ vanishes at either $+\infty$ or $-\infty$. In a bound state it vanishes at both ends. But one can think of situations where the potential confines the wave function at one end but not the other.

Thus the two eigenfunctions differ only by a scale factor and represent the same state. Q.E.D.

What about the free-particle case, where to every energy there are two degenerate solutions with $p = \pm(2mE/\hbar^2)^{1/2}$? The theorem doesn't apply here since $\psi_p(x)$ does not vanish at spatial infinity. [Calculate c in Eq. (5.6.3).]

Theorem 16. The eigenfunctions of H can always be chosen pure real in the coordinate basis.

Proof. If

$$\left[\frac{-\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] \psi_n = E_n \psi_n$$

then by conjugation

$$\left[\frac{-\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] \psi_n^* = E_n \psi_n^*$$

Thus ψ_n and ψ_n^* are eigenfunctions with the same eigenvalue. It follows that the real and imaginary parts of ψ_n ,

$$\psi_r = \frac{\psi_n + \psi_n^*}{2}$$

and

$$\psi_i = \frac{\psi_n - \psi_n^*}{2i}$$

are also eigenfunctions with energy E . Q.E.D.

The theorem holds in higher dimensions as well for *Hamiltonians of the above form*, which in addition to being Hermitian, are *real*. Note, however, that while Hermiticity is preserved under a unitary change of basis, reality is not.

If the problem involves a magnetic field, the Hamiltonian is no longer real in the coordinate basis, as is clear from Eq. (4.3.7). In this case the eigenfunctions cannot be generally chosen real. This question will be explored further at the end of Chapter 11.

Returning to one dimension, due to nondegeneracy of bound states, we must have

$$\psi_i = c\psi_r, \quad c, \text{ a constant}$$

Consequently,

$$\psi = \psi_r + i\psi_i = (1 + ic)\psi_r = \tilde{c}\psi_r$$

Since the overall scale \tilde{c} is irrelevant, we can ignore it, i.e., work with real eigenfunctions with no loss of generality.

This brings us to the end of our study of one-dimensional problems, except for the harmonic oscillator, which is the subject of Chapter 7.

6

The Classical Limit

It is intuitively clear that when quantum mechanics is applied to a macroscopic system it should reproduce the results of classical mechanics, very much the way that relativistic dynamics, when applied to slowly moving ($v/c \ll 1$) objects, reproduces Newtonian dynamics. In this chapter we examine how classical mechanics is regained from quantum mechanics in the appropriate domain. When we speak of regaining classical mechanics, we refer to the numerical aspects. Qualitatively we know that the deterministic world of classical mechanics does not exist. Once we have bitten the quantum apple, our loss of innocence is permanent.

We commence by examining the time evolution of the expectation values. We find

$$\begin{aligned} \frac{d}{dt} \langle \Omega \rangle &= \frac{d}{dt} \langle \psi | \Omega | \psi \rangle \\ &= \langle \dot{\psi} | \Omega | \psi \rangle + \langle \psi | \Omega | \dot{\psi} \rangle + \langle \psi | \dot{\Omega} | \psi \rangle^{\ddagger} \end{aligned} \quad (6.1)$$

In what follows we will assume that Ω has no explicit time dependence. We will therefore drop the third term $\langle \psi | \dot{\Omega} | \psi \rangle$. From the Schrödinger equation, we get

$$|\dot{\psi}\rangle = \frac{-i}{\hbar} H |\psi\rangle$$

and from its adjoint,

$$\langle \dot{\psi} | = \frac{i}{\hbar} \langle \psi | H$$

[†] If you are uncomfortable differentiating bras and kets, work in a basis and convince yourself that this step is correct.

Feeding these into Eq. (6.1) we get the relation

$$\begin{aligned}\frac{d}{dt} \langle \Omega \rangle &= \left(\frac{-i}{\hbar} \right) \langle \psi | [\Omega, H] | \psi \rangle \\ &= \left(\frac{-i}{\hbar} \right) \langle [\Omega, H] \rangle\end{aligned}\quad (6.2)$$

which is called *Ehrenfest's theorem*.

Notice the structural similarity between this equation and the corresponding one from classical mechanics:

$$\frac{d\omega}{dt} = \{\omega, \mathcal{H}\} \quad (6.3)$$

We continue our investigation to see how exactly the two mechanics are related. Let us, for simplicity, discuss a particle in one dimension. If we consider $\Omega = X$ we get

$$\langle \dot{X} \rangle = \left(\frac{-i}{\hbar} \right) \langle [X, H] \rangle \quad (6.4)$$

If we assume

$$H = \frac{P^2}{2m} + V(X)$$

then

$$\langle \dot{X} \rangle = \left(\frac{-i}{\hbar} \right) \langle [X, P^2/2m] \rangle$$

Now

$$\begin{aligned}[X, P^2] &= P[X, P] + [X, P]P \quad [\text{from Eq. (1.5.10)}] \\ &= 2i\hbar P\end{aligned}$$

so that

$$\langle \dot{X} \rangle = \frac{\langle P \rangle}{m} \quad (6.5)$$

The relation $\dot{x} = p/m$ of classical mechanics now appears as a relation among the mean values. We can convert Eq. (6.5) to a more suggestive form by writing

$$\frac{P}{m} = \frac{\partial H}{\partial P}$$

where $\partial H/\partial P$ is a formal derivative of H with respect to P , calculated by pretending that H , P , and X are just c numbers. The rule for finding such derivatives is just as in calculus, as long as the function being differentiated has a power series, as in this case. We now get, in the place of Eq. (6.5),

$$\langle \dot{X} \rangle = \left\langle \frac{\partial H}{\partial P} \right\rangle \quad (6.6)$$

Consider next

$$\begin{aligned} \langle \dot{P} \rangle &= \frac{1}{i\hbar} \langle [P, H] \rangle \\ &= \frac{1}{i\hbar} \langle [P, V(X)] \rangle \end{aligned}$$

To find $[P, V(X)]$ we go to the X basis, in which

$$P \rightarrow -i\hbar \frac{d}{dx} \quad \text{and} \quad V(X) \rightarrow V(x)$$

and for any $\psi(x)$,

$$\left[-i\hbar \frac{d}{dx}, V(x) \right] \psi(x) = -i\hbar \frac{dV}{dx} \psi(x)$$

We conclude that in the abstract,

$$[P, V(X)] = -i\hbar \frac{dV}{dX} \quad (6.7)$$

where dV/dX is again a formal derivative. Since $dV/dX = \partial H/\partial X$, we get

$$\langle \dot{P} \rangle = \left\langle -\frac{\partial H}{\partial X} \right\rangle \quad (6.8)$$

The *similarity* between Eqs. (6.6) and (6.8) and Hamilton's equations is rather striking. We would like to see how the quantum equations reduce to Hamilton's equations when applied to a macroscopic particle (of mass 1 g, say).

First of all, it is clear that we must consider an initial state that resembles the states of classical mechanics, i.e., states with well-defined position and momentum. Although simultaneous eigenstates of X and P do not exist, there do exist states which we can think of as approximate eigenstates of both X and P . In these states, labeled $|x_0 p_0 \Delta\rangle$, $\langle X \rangle = x_0$ and $\langle P \rangle = p_0$, with uncertainties $\Delta X = \Delta$ and $\Delta P \simeq \hbar/\Delta$, both of which are small in the *macroscopic scale*. A concrete example of such a state is

$$|x_0 p_0 \Delta\rangle \rightarrow \Psi_{x_0, p_0, \Delta} = \left(\frac{1}{\pi \Delta^2} \right)^{1/4} e^{ip_0 x / \hbar} e^{-(x - x_0)^2 / 2\Delta^2} \quad (6.9)$$

If we choose $\Delta \simeq 10^{-13}$ cm, say, which is the size of a proton, $\Delta P \simeq 10^{-14}$ g cm/sec. For a particle of mass 1 g, this implies $\Delta V \simeq 10^{-14}$ cm/sec, an uncertainty far below the experimentally detectable range. In the classical scale, such a state can be said to have well-defined values for X and P , namely, x_0 and p_0 , since the uncertainties (fluctuations) around these values are truly negligible. If we let such a state evolve with time, the mean values $x_0(t)$ and $p_0(t)$ will follow Hamilton's equations, once again with negligible deviations. We establish this result as follows.

Consider Eqs. (6.6) and (6.8) which govern the evolution of $\langle X \rangle = x_0$ and $\langle P \rangle = p_0$. These would reduce to Hamilton's equations if we could replace the mean values of the functions on the right-hand side by the functions of the mean values:

$$\dot{x}_0 = \langle \dot{X} \rangle = \left\langle \frac{\partial H(X, P)}{\partial P} \right\rangle \simeq \frac{\partial H}{\partial P} \Big|_{(X=x_0, P=p_0)} = \frac{\partial \mathcal{H}(x_0, p_0)}{\partial p_0} \quad (6.10)$$

and

$$\dot{p}_0 = \langle \dot{P} \rangle = - \left\langle \frac{\partial H}{\partial X} \right\rangle \simeq - \frac{\partial H}{\partial X} \Big|_{(X=x_0, P=p_0)} = - \frac{\partial \mathcal{H}(x_0, p_0)}{\partial x_0} \quad (6.11)$$

If we consider some function of X and P , we will find in the same approximation

$$\langle \Omega(X, P) \rangle \simeq \Omega(x_0, p_0) = \omega(x_0, p_0) \quad (6.12)$$

Thus we regain classical physics as a good approximation whenever it is a good approximation to replace the mean of the functions $\partial H/\partial P$, $-\partial H/\partial X$, and $\Omega(X, P)$ by the functions of the mean. This in turn requires that the fluctuations about the mean have to be small. (The result is exact if there are no fluctuations.) Take as a concrete example Eqs. (6.10) and (6.11). There is no approximation involved in the first equation since $\langle \partial H/\partial P \rangle$ is just $\langle P/m \rangle = p_0/m$. In the second one, we need to approximate $\langle \partial H/\partial X \rangle = \langle dV/dX \rangle = \langle V'(X) \rangle$ by $V'(X=x_0)$. To see when this is a good approximation, let us expand V' in a Taylor series around x_0 . Here it is convenient to work in the coordinate basis where $V(X) = V(x)$. The series is

$$V'(x) = V'(x_0) + (x - x_0) V''(x_0) + \frac{1}{2}(x - x_0)^2 V'''(x_0) + \dots$$

Let us now take the mean of both sides. The first term on the right-hand side, which alone we keep in our approximation, corresponds to the classical force at x_0 , and thus reproduces Newton's second law. The second vanishes in all cases, since the mean of $x - x_0$ does. The succeeding terms, which are corrections to the classical approximation, represent the fact that unlike the classical particle, which responds only to the force $F = -V'$ at x_0 , the quantum particle responds to the force at neighboring points as well. (Note, incidentally, that these terms are zero if the potential is at the most quadratic in the variable x .) Each of these terms is a product of two factors, one of which measures the size or nonlocality of the wave packet and the other, the variation of the force with x . (See the third term for example.) At an intuitive level, we may say that these terms are negligible if the force varies very little over the "size" of the wave packet. (There is no unique definition of "size." The uncertainty is one measure. We see above that the uncertainty squared has to be much smaller than the inverse of the second derivative of the force.) In the present case, where the size of the packet is of the order of 10^{-13} cm, it is clear that the classical approximation is good for any potential that varies appreciably only over macroscopic scales.

There is one apparent problem: although we may start the system out in a state with $\Delta \approx 10^{-13}$ cm, which is certainly a very small uncertainty, we know that with passing time the wave packet will spread. The uncertainty in the particle's position will inevitably become macroscopic. True. But recall the arguments of Section 5.1. We saw that the spreading of the wave packet can be attributed to the fact that any initial uncertainty in velocity, however small, will eventually manifest itself as a giant uncertainty in position. But in the present case ($\Delta V \approx 10^{-14}$ cm/sec) it would take 300,000 years before the packet is even a millimeter across! (It is here that we invoke the fact that the particle is macroscopic: but for this, a small ΔP would not imply a small ΔV .) The problem is thus of academic interest only; and besides, it exists in classical mechanics as well, since the perfect measurement of velocity is merely an idealization.

There remains yet another question. We saw that for a macroscopic particle prepared in a state $|x_0 p_0 \Delta\rangle$, the time evolution of x_0 and p_0 will be in accordance with Hamilton's equations. Question: While it is true that a particle in such a conveniently prepared state obeys classical mechanics, are these the only states one encounters in classical mechanics? What if the initial position of the macroscopic particle is fixed to an accuracy of 10^{-27} cm? Doesn't its velocity now have uncertainties that are classically detectable? Yes. But such states do not occur in practice. The classical physicist talks about making exact position measurements, but never does so in practice. This is clear from the fact that he uses light of a finite frequency to locate the particle's positions, while only light of infinite frequency has perfect resolution. For example light in the visible spectrum has a wavelength of $\lambda \approx 10^{-5}$ cm and thus the minimum ΔX is $\approx 10^{-5}$ cm. If one really went towards the classical ideal and used photons of decreasing wavelength, one would soon find that the momentum of the macroscopic particle is affected by the act of measuring its position. For example, by the time one gets to a wavelength of 10^{-27} cm, each photon would carry a momentum of approximately 1 g cm/sec and one would see *macroscopic* objects recoiling under their impact.

In summary then, a typical macroscopic particle, described classically as possessing a well-defined value of x and p , is in reality an approximate eigenstate $|x_0 p_0 \Delta\rangle$,

where Δ is at least 10^{-5} cm if visible light is used to locate the particle. The quantum equations for the time evolution of these approximate eigenvalues x_0 and p_0 reduce to Hamilton's equations, up to truly negligible uncertainties. The same goes for any other dynamical variable dependent on x and p .

We conclude this chapter by repeating an earlier observation to underscore its importance. Ehrenfest's theorem does not tell us that, in general, the expectation values of quantum operators evolve as do their classical counterparts. In particular, $\langle X \rangle = x_0$ and $\langle P \rangle = p_0$ do not obey Hamilton's equations in all problems. For them to obey Hamilton's equations, we must be able to replace the mean values (expectation values) of the functions $\partial H / \partial P$ and $\partial H / \partial X$ of X and P by the corresponding functions of the mean values $\langle X \rangle = x_0$ and $\langle P \rangle = p_0$. For Hamiltonians that are at the most quadratic in X and P , this replacement can be done with no error for all wave functions. In the general case, such a replacement is a poor approximation unless the fluctuations about the means x_0 and p_0 are small. Even in those cases where x_0 and p_0 obey classical equations, the expectation value of some dependent variable $\Omega(X, P)$ need not, unless we can replace $\langle \Omega(X, P) \rangle$ by $\Omega(\langle X \rangle, \langle P \rangle) = \omega(x_0, p_0)$.

Example 6.1. Consider $\langle \Omega(X) \rangle$, where $\Omega = X^2$, in a state given by $\psi(x) = A \exp[-(x - a)^2 / 2\Delta^2]$. Is $\langle \Omega(X) \rangle = \Omega(\langle X \rangle)$? No, for the difference between the two is $\langle X^2 \rangle - \langle X \rangle^2 = (\Delta X)^2 \neq 0$.

The Harmonic Oscillator

7.1. Why Study the Harmonic Oscillator?

In this section I will put the harmonic oscillator in its place—on a pedestal. Not only is it a system that can be exactly solved (in classical and quantum theory) and a superb pedagogical tool (which will be repeatedly exploited in this text), but it is also a system of great physical relevance. As will be shown below, any system fluctuating by small amounts near a configuration of stable equilibrium may be described either by an oscillator or by a collection of decoupled harmonic oscillators. Since the dynamics of a collection of noninteracting oscillators is no more complicated than that of a single oscillator (apart from the obvious N -fold increase in degrees of freedom), in addressing the problem of the oscillator we are actually confronting the general problem of small oscillations near equilibrium of an arbitrary system.

A concrete example of a single harmonic oscillator is a mass m coupled to a spring of force constant k . For small deformations x , the spring will exert the force given by Hooke's law, $F = -kx$, (k being its force constant) and produce a potential $V = \frac{1}{2}kx^2$. The Hamiltonian for this system is

$$\mathcal{H} = T + V = \frac{p^2}{2m} + \frac{1}{2}m\omega^2x^2 \quad (7.1.1)$$

where $\omega = (k/m)^{1/2}$ is the classical frequency of oscillation. Any Hamiltonian of the above form, quadratic in the coordinate and momentum, will be called the *harmonic oscillator Hamiltonian*. Now, the mass-spring system is just one among the following family of systems described by the oscillator Hamiltonian. Consider a particle moving in a potential $V(x)$. If the particle is placed at one of its minima x_0 , it will remain there in a state of stable, static equilibrium. (A maximum, which is a point of unstable static equilibrium, will not interest us here.) Consider now the dynamics of this particle as it fluctuates by small amounts near $x = x_0$. The potential it experiences may be expanded in a Taylor series:

$$V(x) = V(x_0) + \frac{dV}{dx}\bigg|_{x_0}(x - x_0) + \frac{1}{2!} \frac{d^2V}{dx^2}\bigg|_{x_0}(x - x_0)^2 + \dots \quad (7.1.2)$$

Now, the constant piece $V(x_0)$ is of no physical consequence and may be dropped. [In other words, we may choose $V(x_0)$ as the arbitrary reference point for measuring the potential.] The second term in the series also vanishes since x_0 is a minimum of $V(x)$, or equivalently, since at a point of static equilibrium, the force, $-dV/dx$, vanishes. If we now shift our origin of coordinates to x_0 Eq. (7.1.2) reads

$$V(x) = \frac{1}{2!} \left. \frac{d^2 V}{dx^2} \right|_0 x^2 + \frac{1}{3!} \left. \frac{d^3 V}{dx^3} \right|_0 x^3 + \dots \quad (7.1.3)$$

For *small* oscillations, we may neglect all but the leading term and arrive at the potential (or Hamiltonian) in Eq. (7.1.1), d^2V/dx^2 being identified with $k=m\omega^2$. (By definition, x is small if the neglected terms in the Taylor series are small compared to the leading term, which alone is retained. In the case of the mass-spring system, x is small as long as Hooke's law is a good approximation.)

As an example of a system described by a collection of independent oscillators, consider the coupled-mass system from Example 1.8.6. (It might help to refresh your memory by going back and reviewing this problem.) The Hamiltonian for this system is

$$\begin{aligned} \mathcal{H} &= \frac{p_1^2}{2m} + \frac{p_2^2}{2m} + \frac{1}{2} m\omega^2 [x_1^2 + x_2^2 + (x_1 - x_2)^2] \\ &= \mathcal{H}_1 + \mathcal{H}_2 + \frac{1}{2} m\omega^2 (x_1 - x_2)^2 \end{aligned} \quad (7.1.4)$$

Now this \mathcal{H} is not of the promised form, since the oscillators corresponding to \mathcal{H}_1 and \mathcal{H}_2 (associated with the coordinates x_1 and x_2) are coupled by the $(x_1 - x_2)^2$ term. But we already know of an alternate description of this system in which it can be viewed as two *decoupled* oscillators. The trick is of course the introduction of normal coordinates. We exchange x_1 and x_2 for

$$x_{\text{I}} = \frac{x_1 + x_2}{2^{1/2}} \quad (7.1.5a)$$

and

$$x_{\text{II}} = \frac{x_1 - x_2}{2^{1/2}} \quad (7.1.5b)$$

By differentiating these equations with respect to time, we get similar ones for the velocities, and hence the momenta. In terms of the normal coordinates (and the corresponding momenta),

$$\mathcal{H} = \mathcal{H}_{\text{I}} + \mathcal{H}_{\text{II}} = \frac{p_{\text{I}}^2}{2m} + \frac{1}{2} m\omega^2 x_{\text{I}}^2 + \frac{p_{\text{II}}^2}{2m} + \frac{3}{2} m\omega^2 x_{\text{II}}^2 \quad (7.1.6)$$

Thus the problem of the two coupled masses reduces to that of two uncoupled oscillators of frequencies $\omega_{\text{I}} = \omega = (k/m)^{1/2}$ and $\omega_{\text{II}} = 3^{1/2}\omega = (3k/m)^{1/2}$.

Let us rewrite Eq. (7.1.4) as

$$\mathcal{H} = \frac{1}{2m} \sum_{i=1}^2 \sum_{j=1}^2 p_i \delta_{ij} p_j + \frac{1}{2} \sum_{i=1}^2 \sum_{j=1}^2 x_i V_{ij} x_j \quad (7.1.7)$$

where V_{ij} are elements of a real symmetric (Hermitian) matrix V with the following values:

$$V_{11} = V_{22} = 2m\omega^2, \quad V_{12} = V_{21} = -m\omega^2 \quad (7.1.8)$$

In switching to the normal coordinates x_1 and x_{II} (and p_1 and p_{II}), we are going to a basis that diagonalizes V and reduces the potential energy to a sum of decoupled terms, one for each normal mode. The kinetic energy piece remains decoupled in both bases.

Now, just as the mass-spring system was just a representative element of a family of systems described by the oscillator Hamiltonian, the coupled-mass system is also a special case of a family that can be described by a collection of coupled harmonic oscillators. Consider a system with N Cartesian degrees of freedom $x_1 \dots x_N$, with a potential energy function $V(x_1, \dots, x_N)$. Near an equilibrium point (chosen as the origin), the expansion of V , in analogy with Eq. (7.1.3), is

$$V(x_1 \dots x_N) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \left. \frac{\partial^2 V}{\partial x_i \partial x_j} \right|_0 x_i x_j + \dots \quad (7.1.9)$$

For small oscillations, the Hamiltonian is

$$\mathcal{H} = \sum_{i=1}^N \sum_{j=1}^N \frac{p_i \delta_{ij} p_j}{2m} + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N x_i V_{ij} x_j \quad (7.1.10)$$

where

$$V_{ij} = \left. \frac{\partial^2 V}{\partial x_i \partial x_j} \right|_0 = \left. \frac{\partial^2 V}{\partial x_j \partial x_i} \right|_0 = V_{ji} \quad (7.1.11)$$

are the elements of a *Hermitian* matrix V . (We are assuming for simplicity that the masses associated with all N degrees of freedom are equal.) From the mathematical theory of Chapter 1, we know that there exists a new basis (i.e., a new set of coordinates x_1, x_{II}, \dots) which will diagonalize V and reduce \mathcal{H} to a sum of N decoupled oscillator Hamiltonians, one for each normal mode. Thus the general problem of small fluctuations near equilibrium of an arbitrary system reduces to the study of a single harmonic oscillator.

This section concludes with a brief description of two important systems which are described by a collection of independent oscillators. The first is a crystal (in three dimensions), the atoms in which jiggle about their mean positions on the lattice. The second is the electromagnetic field in free space. A crystal with N_0 atoms (assumed to be point particles) has $3N_0$ degrees of freedom, these being the displacements from

equilibrium points on the lattice. For small oscillations, the Hamiltonian will be quadratic in the coordinates (and of course the momenta). Hence there will exist $3N_0$ normal coordinates and their conjugate momenta, in terms of which \mathcal{H} will be a decoupled sum over oscillator Hamiltonians. What are the corresponding normal modes? Recall that in the case of two coupled masses, the normal modes corresponded to collective motions of the entire system, with the two masses in step in one case, and exactly out of step in the other. Likewise, in the present case, the motion is collective in the normal modes, and corresponds to plane waves traveling across the lattice. For a given wavevector \mathbf{k} , the atoms can vibrate parallel to \mathbf{k} (*longitudinal polarization*) or in any one of the two independent directions perpendicular to \mathbf{k} (*transverse polarization*). Most books on solid state physics will tell you why there are only N_0 possible values for \mathbf{k} . (This must of course be so, for with three polarizations at each \mathbf{k} , we will have exactly $3N_0$ normal modes.) The modes, labeled (\mathbf{k}, λ) , where λ is the polarization index ($\lambda = 1, 2, 3$), form a complete basis for expanding any state of the system. The coefficients of the expansion, $a(\mathbf{k}, \lambda)$, are the normal coordinates. The normal frequencies are labeled $\omega(\mathbf{k}, \lambda)$.†

In the case of the electromagnetic field, the coordinate is the potential $\mathbf{A}(\mathbf{r}, t)$ at each point in space. [$\dot{\mathbf{A}}(\mathbf{r}, t)$ is the “velocity” corresponding to the coordinate $\mathbf{A}(\mathbf{r}, t)$.] The normal modes are once again plane waves but with two differences: there is no restriction on \mathbf{k} , but the polarization has to be transverse. The quantum theory of the field will be discussed at length in Chapter 18.

7.2. Review of the Classical Oscillator

The equations of motion for the oscillator are, from Eq. (7.1.1),

$$\dot{x} = \frac{\partial \mathcal{H}}{\partial p} = \frac{p}{m} \quad (7.2.1)$$

$$\dot{p} = -\frac{\partial \mathcal{H}}{\partial x} = -m\omega^2 x \quad (7.2.2)$$

By eliminating \dot{p} , we arrive at the familiar equation

$$\ddot{x} + \omega^2 x = 0$$

with the solution

$$x(t) = A \cos \omega t + B \sin \omega t = x_0 \cos(\omega t + \phi) \quad (7.2.3)$$

where x_0 is the amplitude and ϕ the phase of oscillator. The conserved energy associated with the oscillator is

$$E = T + V = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}m\omega^2 x^2 = \frac{1}{2}m\omega^2 x_0^2 \quad (7.2.4)$$

† To draw a parallel with the two-mass system, (\mathbf{k}, λ) is like I or II, $a(\mathbf{k}, \lambda)$ is like x_I or x_{II} and $\omega(\mathbf{k}, \lambda)$ is like $(k/m)^{1/2}$ or $(3k/m)^{1/2}$.

Since x_0 is a continuous variable, so is the energy of the classical oscillator. The lowest value for E is zero, and corresponds to the particle remaining at rest at the origin.

By solving for \dot{x} in terms of E and x from Eq. (7.2.4) we obtain

$$\dot{x} = (2E/m - \omega^2 x^2)^{1/2} = \omega(x_0^2 - x^2)^{1/2} \quad (7.2.5)$$

which says that the particle starts from rest at a turning point ($x = \pm x_0$), picks up speed till it reaches the origin, and slows down to rest by the time it reaches the other turning point.

You are reminded of these classical results, so that you may readily compare and contrast them with their quantum counterparts.

7.3. Quantization of the Oscillator (Coordinate Basis)

We now consider the quantum oscillator, that is to say, a particle whose state vector $|\psi\rangle$ obeys the Schrödinger equation

$$i\hbar \frac{d}{dt} |\psi\rangle = H|\psi\rangle$$

with

$$H = \mathcal{H}(x \rightarrow X, p \rightarrow P) = \frac{P^2}{2m} + \frac{1}{2} m\omega^2 X^2$$

As observed repeatedly in the past, the complete dynamics is contained in the propagator $U(t)$, which in turn may be expressed in terms of the eigenvectors and eigenvalues of H . In this section and the next, we will solve the eigenvalue problem in the X basis and the H basis, respectively. In Section 7.5 the passage from the H basis to the X basis will be discussed. The solution in the P basis, trivially related to the solution in the X basis in this case, will be discussed in an exercise.

With an eye on what is to follow, let us first establish that the eigenvalues of H cannot be negative. For any $|\psi\rangle$,

$$\begin{aligned} \langle H \rangle &= \frac{1}{2m} \langle \psi | P^2 | \psi \rangle + \frac{1}{2} m\omega^2 \langle \psi | X^2 | \psi \rangle \\ &= \frac{1}{2m} \langle \psi | P^\dagger P | \psi \rangle + \frac{1}{2} m\omega^2 \langle \psi | X^\dagger X | \psi \rangle \\ &= \frac{1}{2m} \langle P\psi | P\psi \rangle + \frac{1}{2} m\omega^2 \langle X\psi | X\psi \rangle \geq 0 \end{aligned}$$

since the norms of the states $|P\psi\rangle$ and $|X\psi\rangle$ cannot be negative. If we now set $|\psi\rangle$ equal to any eigenstate of H , we get the desired result.

Armed with the above result, we are now ready to attack the problem in the X basis.

We begin by projecting the eigenvalue equation,

$$\left(\frac{P^2}{2m} + \frac{1}{2} m\omega^2 X^2 \right) |E\rangle = E|E\rangle \quad (7.3.1)$$

onto the X basis, using the usual substitutions

$$\begin{aligned} X &\rightarrow x \\ P &\rightarrow -i\hbar \frac{d}{dx} \\ |E\rangle &\rightarrow \psi_E(x) \end{aligned}$$

and obtain

$$\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + \frac{1}{2} m\omega^2 x^2 \right) \psi = E\psi \quad (7.3.2)$$

(The argument of ψ and the subscript E are implicit.)

We can rearrange this equation to the form

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2} \left(E - \frac{1}{2} m\omega^2 x^2 \right) \psi = 0 \quad (7.3.3)$$

We wish to find all solutions to this equation that lie in the physical Hilbert space (of functions normalizable to unity or the Dirac delta function). Follow the approach closely—it will be invoked often in the future.

The first step is to write Eq. (7.3.3) in terms of dimensionless variables. We look for a new variable y which is dimensionless and related to x by

$$x = by \quad (7.3.4)$$

where b is a scale factor with units of length. Although any length b (say the radius of the solar system) will generate a dimensionless variable y , the idea is to choose the natural length scale generated by the equation itself. By feeding Eq. (7.3.4) into Eq. (7.3.3), we arrive at

$$\frac{d^2\psi}{dy^2} + \frac{2mEb^2}{\hbar^2} \psi - \frac{m^2\omega^2b^4}{\hbar^2} y^2 \psi = 0 \quad (7.3.5)$$

The last terms suggests that we choose

$$b = \left(\frac{\hbar}{m\omega} \right)^{1/2} \quad (7.3.6)$$

Let us also define a dimensionless variable ε corresponding to E :

$$\varepsilon = \frac{mEb^2}{\hbar^2} = \frac{E}{\hbar\omega} \quad (7.3.7)$$

(We may equally well choose $\varepsilon = 2mEb^2/\hbar^2$. Constants of order unity are not uniquely suggested by the equation. In the present case, our choice of ε is in anticipation of the results.) In terms of the dimensionless variables, Eq. (7.3.5) becomes

$$\psi'' + (2\varepsilon - y^2)\psi = 0 \quad (7.3.8)$$

where the prime denotes differentiation with respect to y .

Not only do dimensionless variables lead to a more compact equation, they also provide the natural scales for the problem. By measuring x and E in units of $(\hbar/m\omega)^{1/2}$ and $\hbar\omega$, which are scales generated intrinsically by the parameters entering the problem, we develop a feeling for what the words "small" and "large" mean: for example the displacement of the oscillator is large if y is large. If we insist on using the same units for all problems ranging from the atomic physics to cosmology, we will not only be dealing with extremely large or extremely small numbers, we will also have no feeling for the size of quantities in the relevant scale. (A distance of 10^{-20} parsecs, small on the cosmic scale, is enormous if one is dealing with an atomic system.)

The next step is to examine Eq. (7.3.8) at limiting values of y to learn about the solution in these limits. In the limit $y \rightarrow \infty$, we may neglect the $2\varepsilon\psi$ term and obtain

$$\psi'' - y^2\psi = 0 \quad (7.3.9)$$

The solution to this equation *in the same limit* is

$$\psi = Ay^m e^{\pm y^2/2}$$

for

$$\begin{aligned} \psi'' &= Ay^{m+2} \cdot e^{\pm y^2/2} \left[1 \pm \frac{2m+1}{y^2} + \frac{m(m-1)}{y^4} \right] \\ &\xrightarrow[y \rightarrow \infty]{} Ay^{m+2} e^{\pm y^2/2} = y^2\psi \end{aligned}$$

where we have dropped all but the leading power in y as $y \rightarrow \infty$. Of the two possibilities $y^m e^{\pm y^2/2}$, we pick $y^m e^{-y^2/2}$, for the other possibility is not a part of the physical Hilbert space since it grows exponentially as $y \rightarrow \infty$.

Consider next the $y \rightarrow 0$ limit. Equation (7.3.8) becomes, upon dropping the $y^2\psi$ term,

$$\psi'' + 2\varepsilon\psi = 0$$

which has the solution

$$\psi = A \cos[\sqrt{2\varepsilon}y] + B \sin[\sqrt{2\varepsilon}y]$$

Since we have dropped the y^2 term in the equation as being too small, consistency demands that we expand the cosine and sine and drop terms of order y^2 and beyond. We then get

$$\psi \xrightarrow[y \rightarrow 0]{} A + cy + O(y^2)$$

where c is a new constant [$=B(2\varepsilon)^{1/2}$].

We therefore infer that ψ is of the form

$$\psi(y) = u(y) e^{-y^2/2} \quad (7.3.10)$$

where u approaches $A + cy$ (plus higher powers) as $y \rightarrow 0$, and y^m (plus lower powers) as $y \rightarrow \infty$. To determine $u(y)$ completely, we feed the above *ansatz* into Eq. (7.3.8) and obtain

$$u'' - 2yu' + (2\varepsilon - 1)u = 0 \quad (7.3.11)$$

This equation has the desired features (to be discussed in Exercise 7.3.1) that indicate that a power-series solution is possible, i.e., if we assume

$$u(y) = \sum_{n=0}^{\infty} C_n y^n \quad (7.3.12)$$

the equation will determine the coefficients. [The series begins with $n=0$, and not some negative n , since we know that as $y \rightarrow 0$, $u \rightarrow A + cy + O(y^2)$.] Feeding this series into Eq. (7.3.11) we find

$$\sum_{n=0}^{\infty} C_n [n(n-1)y^{n-2} - 2ny^n + (2\varepsilon - 1)y^n] = 0 \quad (7.3.13)$$

Consider the first of three pieces in the above series:

$$\sum_{n=0}^{\infty} C_n n(n-1)y^{n-2}$$

Due to the $n(n-1)$ factor, this series also equals

$$\sum_{n=2}^{\infty} C_n n(n-1) y^{n-2}$$

In terms of a new variable $m=n-2$ the series becomes

$$\sum_{m=0}^{\infty} C_{m+2}(m+2)(m+1)y^m \equiv \sum_{n=0}^{\infty} C_{n+2}(n+2)(n+1)y^n$$

since m is a dummy variable. Feeding this equivalent series back into Eq. (7.3.13) we get

$$\sum_{n=0}^{\infty} y^n [C_{n+2}(n+2)(n+1) + C_n(2\varepsilon - 1 - 2n)] = 0 \quad (7.3.14)$$

Since the functions y^n are linearly independent (you cannot express y^n as a linear combination of other powers of y) each coefficient in the linear relation above must vanish. We thus find

$$C_{n+2} = C_n \frac{(2n+1-2\varepsilon)}{(n+2)(n+1)} \quad (7.3.15)$$

Thus for any C_0 and C_1 , the *recursion relation* above generates C_2, C_4, C_6, \dots and C_3, C_5, C_7, \dots . The function $u(y)$ is given by

$$\begin{aligned} u(y) &= C_0 \left[1 + \frac{(1-2\varepsilon)y^2}{(0+2)(0+1)} + \frac{(1-2\varepsilon)}{(0+2)(0+1)} \frac{(4+1-2\varepsilon)}{(2+2)(2+1)} y^4 + \dots \right] \\ &\quad + C_1 \left[y + \frac{(2+1-2\varepsilon)y^3}{(1+2)(1+1)} + \frac{(2+1-2\varepsilon)}{(1+2)(1+1)} \frac{(6+1-2\varepsilon)}{(3+2)(3+1)} y^5 + \dots \right] \end{aligned} \quad (7.3.16)$$

where C_0 and C_1 are arbitrary.

It appears as if the energy of the quantum oscillator is arbitrary, since ε has not been constrained in any way. But we know something is wrong, since we saw at the outset that the oscillator eigenvalues are nonnegative. The first sign of sickness in our solution, Eq. (7.3.16), is that $u(y)$ does not behave like y^m as $y \rightarrow \infty$ (as deduced at the outset) since it contains arbitrarily high powers of y . There is only one explanation. We have seen that as $y \rightarrow \infty$, there are just two possibilities

$$\psi(y) \xrightarrow[y \rightarrow \infty]{} y^m e^{\pm y^2/2}$$

If we write $\psi(y) = u(y) e^{-y^2/2}$, then the two possibilities for $u(y)$ are

$$u(y) \xrightarrow[y \rightarrow \infty]{} y^m \quad \text{or} \quad y^m e^{y^2}$$

Clearly $u(y)$ in Eq. (7.3.16), which is not bounded by any finite power of y as $y \rightarrow \infty$, corresponds to the latter case. We may explicitly verify this as follows.

Consider the power series for $u(y)$ as $y \rightarrow \infty$. Just as the series is controlled by C_0 (the coefficient of the lowest power of y) as $y \rightarrow 0$, it is governed by its coefficients $C_{n \rightarrow \infty}$ as $y \rightarrow \infty$. The growth of the series is characterized by the ratio [see Eq. (7.3.15)]

$$\frac{C_{n+2}}{C_n} \xrightarrow{n \rightarrow \infty} \frac{2}{n} \quad (7.3.17)$$

Compare this to the growth of $y^m e^{y^2}$. Since

$$y^m e^{y^2} = \sum_{k=0}^{\infty} \frac{y^{2k+m}}{k!}$$

C_n = coefficient of $y^n = 1/k!$; with $n = 2k + m$ or $k = (n - m)/2$. Likewise

$$C_{n+2} = \frac{1}{[(n+2-m)/2]!}$$

so

$$\frac{C_{n+2}}{C_n} \xrightarrow{n \rightarrow \infty} \frac{[(n-m)/2]!}{[(n+2-m)/2]!} = \frac{1}{(n-m+2)/2} \sim \frac{2}{n}$$

In other words, $u(y)$ in Eq. (7.3.16) grows as $y^m e^{y^2}$, so that $\psi(y) \simeq y^m e^{y^2} e^{-y^2/2} \simeq y^m e^{+y^2/2}$, which is the rejected solution raising its ugly head! Our predicament is now reversed: from finding that every ε is allowed, we are now led to conclude that no ε is allowed. Fortunately there is a way out. If ε is one of the special values

$$\varepsilon_n = \frac{2n+1}{2}, \quad n = 0, 1, 2, \dots \quad (7.3.18)$$

the coefficient C_{n+2} (and others dependent on it) vanish. If we choose $C_1 = 0$ when n is even (or $C_0 = 0$ when n is odd) we have a finite polynomial of order n which satisfies the differential equation and behaves as y^n as $y \rightarrow \infty$:

$$\psi(y) = u(y) e^{-y^2/2} = \left\{ \begin{array}{l} C_0 + C_2 y^2 + C_4 y^4 + \dots + C_n y^n \\ C_1 y + C_3 y^3 + C_5 y^5 + \dots + C_n y^n \end{array} \right\} e^{-y^2/2} \quad (7.3.19)$$

Equation (7.3.18) tells us that energy is quantized: the only allowed values for $E = \varepsilon \hbar \omega$ (i.e., values that yield solutions in the physical Hilbert space) are

$$E_n = (n + \frac{1}{2}) \hbar \omega, \quad n = 0, 1, 2, \dots \quad (7.3.20)$$

For each value of n , Eq. (7.3.15) determines the corresponding polynomials of n th order, called *Hermite polynomials*, $H_n(y)$:

$$\begin{aligned} H_0(y) &= 1 \\ H_1(y) &= 2y \\ H_2(y) &= -2(1 - 2y^2) \\ H_3(y) &= -12(y - \frac{2}{3}y^3) \\ H_4(y) &= 12(1 - 4y^2 + \frac{4}{3}y^4) \end{aligned} \quad (7.3.21)$$

The arbitrary initial coefficients C_0 and C_1 in H_n are chosen according to a standard convention. The normalized solutions are then

$$\begin{aligned} \psi_E(x) &\equiv \psi_{(n+1/2)\hbar\omega}(x) \equiv \psi_n(x) \\ &= \left(\frac{m\omega}{\pi\hbar 2^{2n}(n!)^2} \right)^{1/4} \exp\left(-\frac{m\omega x^2}{2\hbar}\right) H_n\left[\left(\frac{m\omega}{\hbar}\right)^{1/2} x\right] \end{aligned} \quad (7.3.22)$$

The derivation of the normalization constant

$$A_n = \left[\frac{m\omega}{\pi\hbar 2^{2n}(n!)^2} \right]^{1/4} \quad (7.3.23)$$

is rather tedious and will not be discussed here in view of a shortcut to be discussed in the next section.

The following recursion relations among Hermite polynomials are very useful:

$$H'_n(y) = 2nH_{n-1} \quad (7.3.24)$$

$$H_{n+1}(y) = 2yH_n - 2nH_{n-1} \quad (7.3.25)$$

as is the integral

$$\int_{-\infty}^{\infty} H_n(y) H_{n'}(y) e^{-y^2} dy = \delta_{nn'} (\pi^{1/2} 2^n n!) \quad (7.3.26)$$

which is just the orthonormality condition of the eigenfunctions $\psi_n(x)$ and $\psi_{n'}(x)$ written in terms of $y = (m\omega/\hbar)^{1/2}x$.

We can now express the propagator as

$$\begin{aligned} U(x, t; x', t') &= \sum_{n=0}^{\infty} A_n \exp\left(-\frac{m\omega}{2\hbar} x^2\right) H_n(x) A_n \exp\left(-\frac{m\omega}{2\hbar} x'^2\right) \\ &\times H_n(x') \exp[-i(n+1/2)\omega(t-t')] \end{aligned} \quad (7.3.27)$$

Evaluation of this sum is a highly formidable task. We will not attempt it here since we will find an extremely simple way for calculating U in Chapter 8, devoted to the path integral formalism. The result happens to be

$$U(x, t; x', t') = \left(\frac{m\omega}{2\pi i\hbar \sin \omega T} \right)^{1/2} \exp \left[\frac{i m \omega}{\hbar} \frac{(x^2 + x'^2) \cos \omega T - 2xx'}{2 \sin \omega T} \right] \quad (7.3.28)$$

where $T = t - t'$.

This concludes the solution of the eigenvalue problem. Before analyzing our results let us recapitulate our strategy.

- Step 1. Introduce dimensionless variables natural to the problem.
- Step 2. Extract the asymptotic ($y \rightarrow \infty, y \rightarrow 0$) behavior of ψ .
- Step 3. Write ψ as a product of the asymptotic form and an unknown function u . The function u will usually be easier to find than ψ .
- Step 4. Try a power series to see if it will yield a recursion relation of the form Eq. (7.3.15).

*Exercise 7.3.1.** Consider the question why we tried a power-series solution for Eq. (7.3.11) but not Eq. (7.3.8). By feeding in a series into the latter, verify that a three-term recursion relation between C_{n+2} , C_n , and C_{n-2} obtains, from which the solution does not follow so readily. The problem is that ψ'' has two powers of y less than $2\varepsilon\psi$, while the $-y^2$ piece has two more powers of y . In Eq. (7.3.11) on the other hand, of the three pieces u'' , $-2yu'$, and $(2\varepsilon - 1)u$, the last two have the same powers of y .

Exercise 7.3.2. Verify that $H_3(y)$ and $H_4(y)$ obey the recursion relation, Eq. (7.3.15).

Exercise 7.3.3. If $\psi(x)$ is even and $\phi(x)$ is odd under $x \rightarrow -x$, show that

$$\int_{-\infty}^{\infty} \psi(x)\phi(x) dx = 0$$

Use this to show that $\psi_2(x)$ and $\psi_1(x)$ are orthogonal. Using the values of Gaussian integrals in Appendix A.2 verify that $\psi_2(x)$ and $\psi_0(x)$ are orthogonal.

Exercise 7.3.4. Using Eqs. (7.3.23)–(7.3.25), show that

$$\langle n' | X | n \rangle = \left(\frac{\hbar}{2m\omega} \right)^{1/2} [\delta_{n',n+1}(n+1)^{1/2} + \delta_{n',n-1}n^{1/2}]$$

$$\langle n' | P | n \rangle = \left(\frac{m\omega\hbar}{2} \right)^{1/2} i[\delta_{n',n+1}(n+1)^{1/2} - \delta_{n',n-1}n^{1/2}]$$

*Exercise 7.3.5.** Using the symmetry arguments from Exercise 7.3.3 show that $\langle n | X | n \rangle = \langle n | P | n \rangle = 0$ and thus that $\langle X^2 \rangle = (\Delta X)^2$ and $\langle P^2 \rangle = (\Delta P)^2$ in these states. Show that $\langle 1 | X^2 | 1 \rangle = 3\hbar/2m\omega$ and $\langle 1 | P^2 | 1 \rangle = \frac{3}{2}m\omega\hbar$. Show that $\psi_0(x)$ saturates the uncertainty bound $\Delta X \cdot \Delta P \geq \hbar/2$.

*Exercise 7.3.6.** Consider a particle in a potential

$$\begin{aligned} V(x) &= \frac{1}{2}m\omega^2x^2, & x > 0 \\ &= \infty, & x \leq 0 \end{aligned}$$

What are the boundary conditions on the wave functions now? Find the eigenvalues and eigenfunctions.

We now discuss the eigenvalues and eigenfunctions of the oscillator. The following are the main features:

(1) The energy is quantized. In contrast to the classical oscillator whose energy is continuous, the quantum oscillator has a discrete set of levels given by Eq. (7.3.20). Note that the quantization emerges only after we supplement Schrödinger's equation with the requirement that ψ be an element of the physical Hilbert space. In this case it meant the imposition of the boundary condition $\psi(|x| \rightarrow \infty) \rightarrow 0$ [as opposed to $\psi(|x| \rightarrow \infty) \rightarrow \infty$, which is what obtained for all but the special values of E].

Why does the classical oscillator seem to have a continuum of energy values? The answer has to do with the relative sizes of the energy gap and the total energy of the classical oscillator. Consider, for example, a mass of 2 g, oscillating at a frequency of 1 rad/sec, with an amplitude of 1 cm. Its energy is

$$E = \frac{1}{2}m\omega^2x_0^2 = 1 \text{ erg}$$

Compare this to the gap between allowed energies:

$$\Delta E = \hbar\omega \simeq 10^{-27} \text{ erg}$$

At the macroscopic level, it is practically impossible to distinguish between a system whose energy is continuous and one whose allowed energy levels are spaced 10^{-27} erg apart. Stated differently, the *quantum number* associated with this oscillator is

$$n = \frac{E}{\hbar\omega} - \frac{1}{2} \simeq 10^{27}$$

while the difference in n between adjacent levels is unity. We have here a special case of the *correspondence principle*, which states that as the quantum number tends to infinity, we regain the classical picture. (We know vaguely that when a system is big, it may be described classically. The correspondence principle tells us that the quantum number is a good measure of bigness.)

(2) The levels are spaced uniformly. The fact that the oscillator energy levels go up in steps of $\hbar\omega$ allows one to construct the following picture. We pretend that associated with an oscillator of classical frequency ω there exist fictitious particles called *quanta* each endowed with energy $\hbar\omega$. We view the $n\hbar\omega$ piece in the energy formula Eq. (7.3.20) as the energy of n such quanta. In other words, we forget about the mass and spring and think in terms of the quanta. When the quantum number n goes up (or down) by Δn , we say that Δn quanta have been created (or destroyed).

Although it seems like a matter of semantics, thinking of the oscillator in terms of these quanta has proven very useful.

In the case of the crystal, there are $3N_0$ oscillators, labeled by the $3N_0$ values of (\mathbf{k}, λ) , with frequencies $\omega(\mathbf{k}, \lambda)$. The quantum state of the crystal is specified by giving the number of quanta, called *phonons*, at each (\mathbf{k}, λ) . For a crystal whose Hamiltonian is exactly given by a sum of oscillator pieces, the introduction of the phonon concept is indeed a matter of semantics. If, however, we consider deviations from this, say to take into account nonleading terms in the Taylor expansion of the potential, or the interaction between the crystal and some external probe such as an electron shot at it, the phonon concept proves very useful. (The two effects mentioned above may be seen as phonon-phonon interactions and phonon-electron interactions, respectively.)

Similarly, the interaction of the electromagnetic field with matter may be viewed as the interaction between light quanta or *photons* and matter, which is discussed in Chapter 18.

(3) The lowest possible energy is $\hbar\omega/2$ and not 0. Unlike the classical oscillator, which can be in a state of zero energy (with $x=p=0$) the quantum oscillator has a minimum energy of $\hbar\omega/2$. This energy, called the *zero-point energy*, is a reflection of the fact that the simultaneous eigenstate $|x=0, p=0\rangle$ is precluded by the canonical commutation relation $[X, P]=i\hbar$. This result is common to all oscillators, whether they describe a mechanical system or a normal mode of the electromagnetic field, since all these problems are mathematically identical and differ only in what the coordinate and its conjugate momentum represent. Thus, a crystal has an energy $\frac{1}{2}\hbar\omega(\mathbf{k}, \lambda)$ in each mode (\mathbf{k}, λ) even when phonons are absent, and the electromagnetic field has an energy $\frac{1}{2}\hbar\omega(\mathbf{k}, \lambda)$ in each mode of frequency ω even when photons are absent. (The zero-point fluctuation of the field has measurable consequences, which will be discussed in Chapter 18.)

In the following discussion let us restrict ourselves to the mechanical oscillator and examine more closely the zero-point energy. We saw that it is the absence of the state $|x=0, p=0\rangle$ that is responsible for this energy. Such a state, with $\Delta X=\Delta P=0$, is forbidden by the uncertainty principle. Let us therefore try to find a state that is quantum mechanically allowed and comes as close as possible (in terms of its energy) to the classical state $x=p=0$. If we choose a wave function $\psi(x)$ that is sharply peaked near $x=0$ to minimize the mean potential energy $\langle \frac{1}{2}m\omega^2 X^2 \rangle$, the wave function in P space spreads out and the mean kinetic energy $\langle P^2/2m \rangle$ grows. The converse happens if we pick a momentum space wave function sharply peaked near $p=0$. What we need then is a compromise $\psi_{\min}(x)$ that minimizes the *total* mean energy without violating the uncertainty principle. Let us now begin our quest for $\psi_{\min}(x)$. We start with a normalized trial state $|\psi\rangle$ and consider

$$\langle \psi | H | \psi \rangle = \langle H \rangle = \frac{\langle P^2 \rangle}{2m} + \frac{1}{2} m\omega^2 \langle X^2 \rangle \quad (7.3.29)$$

Now

$$(\Delta P)^2 = \langle P^2 \rangle - \langle P \rangle^2 \quad (7.3.30)$$

and

$$(\Delta X)^2 = \langle X^2 \rangle - \langle X \rangle^2 \quad (7.3.31)$$

so that

$$\langle H \rangle = \frac{(\Delta P)^2 + \langle P \rangle^2}{2m} + \frac{1}{2} m\omega^2 [(\Delta X)^2 + \langle X \rangle^2] \quad (7.3.32)$$

The first obvious step in minimizing $\langle H \rangle$ is to restrict ourselves to states with $\langle X \rangle = \langle P \rangle = 0$. (Since $\langle X \rangle$ and $\langle P \rangle$ are independent of each other and of $(\Delta X)^2$ and $(\Delta P)^2$, such a choice is always possible.) For these states (from which we must pick the winner)

$$\langle H \rangle = \frac{(\Delta P)^2}{2m} + \frac{1}{2} m\omega^2 (\Delta X)^2 \quad (7.3.33)$$

Now we use the uncertainty relation

$$\Delta X \cdot \Delta P \geq \hbar/2 \quad (7.3.34)$$

where the equality sign holds only for a Gaussian, as will be shown in Section 9.3.
We get

$$\langle H \rangle \geq \frac{\hbar^2}{8m(\Delta X)^2} + \frac{1}{2} m\omega^2 (\Delta X)^2 \quad (7.3.35)$$

We minimize $\langle H \rangle$ by choosing a Gaussian wave function, for which

$$\langle H \rangle_{\text{Gaussian}} = \frac{\hbar^2}{8m(\Delta X)^2} + \frac{1}{2} m\omega^2 (\Delta X)^2 \quad (7.3.36)$$

What we have found is that the mean energy associated with the trial wave function is sensitive only to the corresponding ΔX and that, of all functions with the same ΔX , the Gaussian has the lowest energy. Finally we choose, from the family of Gaussians, the one with the ΔX that minimizes $\langle H \rangle_{\text{Gaussian}}$. By requiring

$$\frac{\partial \langle H \rangle_{\text{Gaussian}}}{\partial (\Delta X)^2} = 0 = \frac{-\hbar^2}{8m(\Delta X)^4} + \frac{1}{2} m\omega^2 \quad (7.3.37)$$

we obtain

$$(\Delta X)^2 = \hbar/2m\omega \quad (7.3.38)$$

and

$$\langle H \rangle_{\min} = \hbar\omega/2 \quad (7.3.39)$$

Thus, by systematically hunting in Hilbert space, we have found that the following normalized function has the lowest mean energy:

$$\psi_{\min}(x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} \exp\left(-\frac{m\omega x^2}{2\hbar}\right), \quad \langle H \rangle_{\min} = \frac{\hbar\omega}{2} \quad (7.3.40)$$

If we apply the above result

$$\langle \psi_{\min} | H | \psi_{\min} \rangle \leq \langle \psi | H | \psi \rangle \quad (\text{for all } |\psi\rangle)$$

to $|\psi\rangle = |\psi_0\rangle$ = ground-state vector, we get

$$\langle \psi_{\min} | H | \psi_{\min} \rangle \leq \langle \psi_0 | H | \psi_0 \rangle = E_0 \quad (7.3.41)$$

Now compare this with the result of Exercise 5.2.2:

$$E_0 = \langle \psi_0 | H | \psi_0 \rangle \leq \langle \psi | H | \psi \rangle \quad \text{for all } |\psi\rangle$$

If we set $|\psi\rangle = |\psi_{\min}\rangle$ we get

$$E_0 = \langle \psi_0 | H | \psi_0 \rangle \leq \langle \psi_{\min} | H | \psi_{\min} \rangle \quad (7.3.42)$$

It follows from Eq. (7.3.41) and (7.3.42) that

$$E_0 = \langle \psi_0 | H | \psi_0 \rangle = \langle \psi_{\min} | H | \psi_{\min} \rangle = \frac{\hbar\omega}{2} \quad (7.3.43)$$

Also, since there was only one state, $|\psi_{\min}\rangle$, with energy $\hbar\omega/2$, it follows that

$$|\psi_0\rangle = |\psi_{\min}\rangle \quad (7.3.44)$$

We have thus managed to find the oscillator ground-state energy and state vector without solving the Schrödinger equation.

It would be a serious pedagogical omission if it were not emphasized at this juncture that the uncertainty relation has been unusually successful in the above context. Our ability here to obtain all the information about the ground state using the uncertainty relation is a consequence of the special form of the oscillator Hamiltonian [which allowed us to write $\langle H \rangle$ in terms of $(\Delta X)^2$ and $(\Delta P)^2$] and the fact that its ground-state wave function is a Gaussian (which has a privileged role with respect to the uncertainty relation). In more typical instances, the use of the uncertainty relation will have to be accompanied by some hand-waving [before $\langle H \rangle$ can be approximated by a function of $(\Delta X)^2$ and $(\Delta P)^2$] and then too will yield only an estimate for the ground-state energy. As for the wave function, we can only get an estimate for ΔX , the spread associated with it.

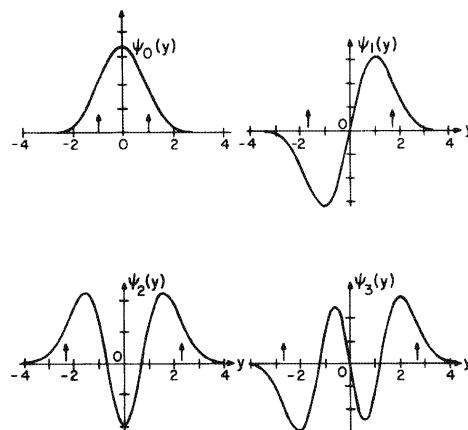


Figure 7.1. Normalized eigenfunctions for $n=0, 1, 2$, and 3 . The small arrows at $|y|=(2n+1)^{1/2}$ stand for the classical turning points. Recall that $y=(m\omega/\hbar)^{1/2}x$.

(4) The solutions (Fig. 7.1) $\psi_n(x)$ contain only even or odd powers of x , depending on whether n is even or odd. Consequently the eigenfunctions are even or odd:

$$\begin{aligned}\psi_n(-x) &= \psi_n(x), & n \text{ even} \\ &= -\psi_n(x), & n \text{ odd}\end{aligned}$$

In Chapter 11 on symmetries it will be shown that the eigenfunctions had to have this property.

(5) The wave function does not vanish beyond the classical turning points, but dies out exponentially as $x \rightarrow \infty$. [Verify that the classical turning points are given by $y_0=\pm(2n+1)^{1/2}$.] Notice, however, that when n is large (Fig. 7.2) the excursions outside the turning points are small compared to the classical amplitude. This exponentially damped amplitude in the classically forbidden region was previously encountered in Chapter 5 when we studied tunneling.

(6) The probability distribution $P(x)$ is very different from the classical case. The position of a given classical oscillator is of course exactly known. But we could ask the following probabilistic question: if I suddenly walk into a room containing the oscillator, where am I likely to catch it? If the velocity of the oscillator at a point x is $v(x)$, the time it spends near the x , and hence the probability of our catching it there during a random spot check, varies inversely with $v(x)$:

$$P_{\text{cl}}(x) \propto \frac{1}{v(x)} = \frac{1}{\omega(x_0^2 - x^2)^{1/2}} \quad (7.3.45)$$

which is peaked near $\pm x_0$ and has a minimum at the origin. In the quantum case, for the ground state in particular, $|\psi(x)|^2$ seems to go just the other way (Fig. 7.1). There is no contradiction here, for quantum mechanics *is* expected to differ from classical mechanics. The correspondence principle, however, tells us that for large n

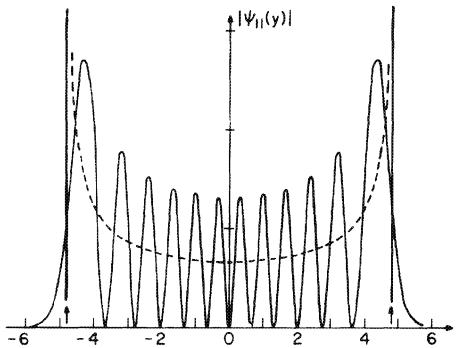


Figure 7.2. Probability density in the state $n=11$. The broken curve gives the classical probability distribution in a state with the same energy.

the two must become indistinguishable. From Fig. 7.2, which shows the situations at $n=11$, we can see how the classical limit is reached: the quantum distribution $P(x)=|\psi(x)|^2$ wiggles so rapidly (in a scale set by the classical amplitude) that only its mean can be detected at these scales, and this agrees with $P_{cl}(x)$. We are reminded here of the double-slit experiment performed with macroscopic particles: there is a dense interference pattern, whose mean is measured in practice and agrees with the classical probability curve.

A remark that was made in more general terms in Chapter 6: the classical oscillator that we often refer to, is a figment lodged in our imagination and doesn't exist. In other words, all oscillators, including the 2-g mass and spring system, are ultimately governed by the laws of quantum mechanics, and thus have discrete energies, can shoot past the "classical" turning points, and have a zero-point energy of $\frac{1}{2}\hbar\omega$ even while they play dead. Note however that what I am calling nonexistent is an oscillator that *actually* has the properties attributed to it in classical mechanics, and not one that *seems* to have them when examined at the macroscopic level.

Exercise 7.3.7. The Oscillator in Momentum Space.* By setting up an eigenvalue equation for the oscillator in the P basis and comparing it to Eq. (7.3.2), show that the momentum space eigenfunctions may be obtained from the ones in coordinate space through the substitution $x \rightarrow p$, $m\omega \rightarrow 1/m\omega$. Thus, for example,

$$\psi_0(p) = \left(\frac{1}{m\pi\hbar\omega} \right)^{1/4} e^{-p^2/2m\hbar\omega}$$

There are several other pairs, such as ΔX and ΔP in the state $|n\rangle$, which are related by the substitution $m\omega \rightarrow 1/m\omega$. You may wish to watch out for them. (Refer back to Exercise 7.3.5.)

7.4. The Oscillator in the Energy Basis

Let us orient ourselves by recalling how the eigenvalue equation

$$\left(\frac{P^2}{2m} + \frac{1}{2} m\omega^2 X^2 \right) |E\rangle = E |E\rangle \quad (7.4.1)$$

was solved in the coordinate basis: (1) We made the assignments $X \rightarrow x$, $P \rightarrow -i\hbar d/dx$. (2) We solved for the components $\langle x | E \rangle = \psi_E(x)$ and the eigenvalues.

To solve the problem in the momentum basis, we first compute the X and P operators in this basis, given their form in the coordinate basis. For instance,

$$\begin{aligned} \langle p' | X | p \rangle &= \iint \underbrace{\langle p' | x \rangle}_{\frac{e^{-ip'x/\hbar}}{(2\pi\hbar)^{1/2}}} \underbrace{\langle x | X | x' \rangle}_{x\delta(x-x')} \underbrace{\langle x' | p \rangle}_{\frac{e^{ip'x'/\hbar}}{(2\pi\hbar)^{1/2}}} dx dx' \\ &= -i\hbar \delta'(p - p') \end{aligned}$$

We then find P and $H(X, P)$ in this basis. The eigenvalue equation, (7.4.1), will then become a differential equation that we will proceed to solve.

Now suppose that we want to work in the energy basis. We must first find the eigenfunctions of H , i.e., $\langle x | E \rangle$, so that we can carry out the change of basis. But finding $\langle x | E \rangle = \psi_E(x)$ amounts to solving the full eigenvalue problem in the coordinate basis. Once we have done this, there is not much point in setting up the problem in the E basis.

But there is a clever way due to Dirac, which allows us to work in the energy basis without having to know ahead of time the operators X and P in this basis. All we will need is the commutation relation

$$[X, P] = i\hbar I = i\hbar \quad (7.4.2)$$

which follows from $X \rightarrow x$, $P \rightarrow -i\hbar d/dx$, but is basis independent. The next few steps will seem rather mysterious and will not fit into any of the familiar schemes discussed so far. You must be patient till they begin to pay off.

Let us first introduce the operator

$$a = \left(\frac{m\omega}{2\hbar} \right)^{1/2} X + i \left(\frac{1}{2m\omega\hbar} \right)^{1/2} P \quad (7.4.3)$$

and its adjoint

$$a^\dagger = \left(\frac{m\omega}{2\hbar} \right)^{1/2} X - i \left(\frac{1}{2m\omega\hbar} \right)^{1/2} P \quad (7.4.4)$$

(Note that $m\omega \rightarrow 1/m\omega$ as $X \leftrightarrow P$.) They satisfy the commutation relation (which you should verify)

$$[a, a^\dagger] = 1 \quad (7.4.5)$$

Note next that the Hermitian operator $a^\dagger a$ is simply related to H :

$$\begin{aligned} a^\dagger a &= \frac{m\omega}{2\hbar} X^2 + \frac{1}{2m\omega\hbar} P^2 + \frac{i}{2\hbar} [X, P] \\ &= \frac{H}{\hbar\omega} - \frac{1}{2} \end{aligned}$$

so that

$$H = (a^\dagger a + 1/2)\hbar\omega \quad (7.4.6)$$

[This method is often called the “method of factorization” since we are expressing $H = P^2 + X^2$ (ignoring constants) as a product of $(X + iP) = a$ and $(X - iP) = a^\dagger$. The extra $\hbar\omega/2$ in Eq. (7.4.6) comes from the non-commutative nature of X and P .]

Let us next define an operator \hat{H} ,

$$\hat{H} = \frac{H}{\hbar\omega} = (a^\dagger a + 1/2) \quad (7.4.7)$$

whose eigenvalues ε measure energy in units of $\hbar\omega$. We wish to solve the eigenvalue equation for \hat{H} :

$$\hat{H}|\varepsilon\rangle = \varepsilon|\varepsilon\rangle \quad (7.4.8)$$

where ε is the energy measured in units of $\hbar\omega$. Two relations we will use shortly are

$$[a, \hat{H}] = [a, a^\dagger a + 1/2] = [a, a^\dagger a] = a \quad (7.4.9)$$

and

$$[a^\dagger, \hat{H}] = -a^\dagger \quad (7.4.10)$$

The utility of a and a^\dagger stems from the fact that given an eigenstate of \hat{H} , they generate others. Consider

$$\begin{aligned} \hat{H}a|\varepsilon\rangle &= (a\hat{H} - [a, \hat{H}])|\varepsilon\rangle \\ &= (a\hat{H} - a)|\varepsilon\rangle \\ &= (\varepsilon - 1)a|\varepsilon\rangle \end{aligned} \quad (7.4.11)$$

We infer from Eq. (7.4.11) that $a|\varepsilon\rangle$ is an eigenstate with eigenvalue $\varepsilon - 1$, i.e.,

$$a|\varepsilon\rangle = C_\varepsilon |\varepsilon - 1\rangle \quad (7.4.12)$$

where C_ε is a constant, and $|\varepsilon - 1\rangle$ and $|\varepsilon\rangle$ are normalized eigenkets.[‡]

Similarly we see that

$$\begin{aligned} \hat{H}a^\dagger|\varepsilon\rangle &= (a^\dagger\hat{H} - [a^\dagger, H])|\varepsilon\rangle \\ &= (a^\dagger\hat{H} + a^\dagger)|\varepsilon\rangle \\ &= (\varepsilon + 1)a^\dagger|\varepsilon\rangle \end{aligned} \quad (7.4.13)$$

so that

$$a^\dagger|\varepsilon\rangle = C_{\varepsilon+1}|\varepsilon+1\rangle \quad (7.4.14)$$

One refers to a and a^\dagger as *lowering and raising operators* for obvious reasons. They are also called *destruction and creation operators* since they destroy or create quanta of energy $\hbar\omega$.

We are thus led to conclude that if ε is an eigenvalue of \hat{H} , so are $\varepsilon + 1, \varepsilon + 2, \varepsilon + 3, \dots, \varepsilon + \infty$; and $\varepsilon - 1, \dots, \varepsilon - \infty$. The latter conclusion is in conflict with the result that the eigenvalues of H are nonnegative. So, it must be that the downward chain breaks at some point: there must be a state $|\varepsilon_0\rangle$ that cannot be lowered further:

$$a|\varepsilon_0\rangle = 0 \quad (7.4.15)$$

Operating with a^\dagger , we get

$$a^\dagger a|\varepsilon_0\rangle = 0$$

or

$$(\hat{H} - 1/2)|\varepsilon_0\rangle = 0 \quad [\text{from Eq. (7.4.7)}]$$

or

$$\hat{H}|\varepsilon_0\rangle = \frac{1}{2}|\varepsilon_0\rangle$$

or

$$\varepsilon_0 = \frac{1}{2} \quad (7.4.16)$$

[‡] We are using the fact that there is no degeneracy in one dimension.

We may, however, raise the state $|\varepsilon_0\rangle$ indefinitely by the repeated application of a^\dagger . We thus find that the oscillator has a sequence of levels given by

$$\varepsilon_n = (n + 1/2), \quad n = 0, 1, 2, \dots$$

or

$$E_n = (n + 1/2)\hbar\omega, \quad n = 0, 1, 2, \dots \quad (7.4.17)$$

Are these the only levels? If there were another family, it too would have to have a ground state $|\varepsilon'_0\rangle$ such that

$$a|\varepsilon'_0\rangle = 0$$

or

$$a^\dagger a|\varepsilon'_0\rangle = 0$$

or

$$\hat{H}|\varepsilon'_0\rangle = \frac{1}{2}|\varepsilon'_0\rangle \quad (7.4.18)$$

But we know that there is no degeneracy in one dimension (Theorem 15). Consequently it follows from Eqs. (7.4.16) and (7.4.18) that $|\varepsilon_0\rangle$ and $|\varepsilon'_0\rangle$ represent the same state. The same goes for the families built from $|\varepsilon_0\rangle$ and $|\varepsilon'_0\rangle$ by the repeated action of a^\dagger .

We now calculate the constants C_ε and $C_{\varepsilon+1}$ appearing in Eqs. (7.4.12) and (7.4.14). Since $\varepsilon = n + 1/2$, let us label the kets by the integer n . We want to determine the constant C_n appearing in the equation

$$a|n\rangle = C_n|n-1\rangle \quad (7.4.19a)$$

Consider the adjoint of this equation

$$\langle n|a^\dagger = \langle n-1|C_n^* \quad (7.4.19b)$$

By combining these equations we arrive at

$$\begin{aligned} \langle n|a^\dagger a|n\rangle &= \langle n-1|n-1\rangle C_n^* C_n \\ \langle n|\hat{H} - \frac{1}{2}|n\rangle &= C_n^* C_n \quad (\text{since } |n-1\rangle \text{ is normalized}) \\ \langle n|n|n\rangle &= |C_n|^2 \quad (\text{since } \hat{H}|n\rangle = (n + 1/2)|n\rangle) \\ |C_n|^2 &= n \\ C_n &= (n)^{1/2} e^{i\phi} \quad (\phi \text{ is arbitrary}) \end{aligned} \quad (7.4.20)$$

It is conventional to choose ϕ as zero. So we have

$$a|n\rangle = n^{1/2}|n-1\rangle \quad (7.4.21)$$

It can similarly be shown (by you) that

$$a^\dagger|n\rangle = (n+1)^{1/2}|n+1\rangle \quad (7.4.22)$$

[Note that in Eqs. (7.4.21) and (7.4.22) the larger of the n 's labeling the two kets appears under the square root.] By combining these two equations we find

$$a^\dagger a|n\rangle = a^\dagger n^{1/2}|n-1\rangle = n^{1/2}n^{1/2}|n\rangle = n|n\rangle \quad (7.4.23)$$

In terms of

$$N = a^\dagger a \quad (7.4.24)$$

called the *number operator* (since it counts the quanta)

$$\hat{H} = N + \frac{1}{2} \quad (7.4.25)$$

Equations (7.4.21) and (7.4.22) are very important. They allow us to compute the matrix elements of all operators in the $|n\rangle$ basis. First consider a and a^\dagger themselves:

$$\langle n'|a|n\rangle = n^{1/2}\langle n'|n-1\rangle = n^{1/2}\delta_{n,n-1} \quad (7.4.26)$$

$$\langle n'|a^\dagger|n\rangle = (n+1)^{1/2}\langle n'|n+1\rangle = (n+1)^{1/2}\delta_{n,n+1} \quad (7.4.27)$$

To find the matrix elements of X and P , we invert Eqs. (7.4.3) and (7.4.4) to obtain

$$X = \left(\frac{\hbar}{2m\omega} \right)^{1/2} (a + a^\dagger) \quad (7.4.28)$$

$$P = i \left(\frac{m\omega\hbar}{2} \right)^{1/2} (a^\dagger - a) \quad (7.4.29)$$

and then use Eqs. (7.4.26) and (7.4.27). The details are left as an exercise. The two basic matrices in this energy basis are

$$a^\dagger \leftrightarrow \begin{matrix} n=0 & n=1 & n=2 & \dots \\ \begin{matrix} n=0 \\ n=1 \\ \cdot \\ \cdot \end{matrix} & \begin{bmatrix} 0 & 0 & 0 & \dots \\ 1^{1/2} & 0 & 0 & \\ 0 & 2^{1/2} & 0 & \\ 0 & 0 & 3^{1/2} & \\ \vdots & & & \end{bmatrix} & \end{matrix} \quad (7.4.30)$$

$$a \leftrightarrow \begin{bmatrix} 0 & 1^{1/2} & 0 & 0 & \dots \\ 0 & 0 & 2^{1/2} & 0 \\ 0 & 0 & 0 & 3^{1/2} \\ \vdots & & & \end{bmatrix} \quad (7.4.31)$$

Both matrices can be constructed either from Eqs. (7.4.26) and (7.4.27) or Eqs. (7.4.21) and (7.4.22) combined with our mnemonic involving images of the transformed vectors $a^\dagger|n\rangle$ and $a|n\rangle$. We get the matrices representing X and P by turning to Eqs. (7.4.28) and (7.4.29):

$$X \leftrightarrow \left(\frac{\hbar}{2m\omega} \right)^{1/2} \begin{bmatrix} 0 & 1^{1/2} & 0 & 0 & \dots \\ 1^{1/2} & 0 & 2^{1/2} & 0 \\ 0 & 2^{1/2} & 0 & 3^{1/2} \\ 0 & 0 & 3^{1/2} & 0 \\ \vdots & & & \end{bmatrix} \quad (7.4.32)$$

$$P \leftrightarrow i \left(\frac{m\omega\hbar}{2} \right)^{1/2} \begin{bmatrix} 0 & -1^{1/2} & 0 & 0 & \dots \\ 1^{1/2} & 0 & -2^{1/2} & 0 \\ 0 & 2^{1/2} & 0 & -3^{1/2} \\ 0 & 0 & 3^{1/2} & 0 \\ \vdots & & & \end{bmatrix} \quad (7.4.33)$$

The Hamiltonian is of course diagonal in its own basis:

$$H \leftrightarrow \hbar\omega \begin{bmatrix} 1/2 & 0 & 0 & 0 & \dots \\ 0 & 3/2 & 0 & 0 \\ 0 & 0 & 5/2 & 0 \\ \vdots & & & \end{bmatrix} \quad (7.4.34)$$

Equation (7.4.22) also allows us to express all normalized eigenvectors $|n\rangle$ in terms of the ground state $|0\rangle$:

$$|n\rangle = \frac{a^\dagger}{n^{1/2}} |n-1\rangle = \frac{a^\dagger}{n^{1/2}} \frac{a^\dagger}{(n-1)^{1/2}} |n-2\rangle \cdots = \frac{(a^\dagger)^n}{(n!)^{1/2}} |0\rangle \quad (7.4.35)$$

The a and a^\dagger operators greatly facilitate the calculation of the matrix of elements of other operators between oscillator eigenstates. Consider, for example, $\langle 3|X^3|2\rangle$. In

the X basis one would have to carry out the following integral:

$$\begin{aligned}\langle 3|X^3|2\rangle &= \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} \left(\frac{1}{2^3 3!} \cdot \frac{1}{2^2 2!}\right)^{1/2} \int_{-\infty}^{\infty} \left\{ \exp\left(-\frac{m\omega x^2}{2\hbar}\right) \right. \\ &\quad \times H_3\left[\left(\frac{m\omega}{\hbar}\right)^{1/2} x\right] x^3 \exp\left(-\frac{m\omega x^2}{2\hbar}\right) H_2\left[\left(\frac{m\omega}{\hbar}\right)^{1/2} x\right] \left. \right\} dx\end{aligned}$$

whereas in the $|n\rangle$ basis

$$\begin{aligned}\langle 3|X^3|2\rangle &= \left(\frac{\hbar}{2m\omega}\right)^{3/2} \langle 3|(a+a^\dagger)^3|2\rangle \\ &= \left(\frac{\hbar}{2m\omega}\right)^{3/2} \langle 3|(a^3 + a^2 a^\dagger + a a^\dagger a + a a^\dagger a^\dagger \\ &\quad + a^\dagger a a + a^\dagger a a^\dagger + a^\dagger a^\dagger a + a^\dagger a^\dagger a^\dagger)|2\rangle\end{aligned}$$

Since a lowers n by one unit and a^\dagger raises it by one unit and we want to go up by one unit from $n=2$ to $n=3$, the only nonzero contribution comes from $a^\dagger a^\dagger a$, $a a^\dagger a^\dagger$, and $a^\dagger a a^\dagger$. Now

$$\begin{aligned}a^\dagger a^\dagger a|2\rangle &= 2^{1/2} a^\dagger a^\dagger |1\rangle = 2^{1/2} 2^{1/2} a^\dagger |2\rangle = 2^{1/2} 2^{1/2} 3^{1/2} |3\rangle \\ a a^\dagger a^\dagger |2\rangle &= 3^{1/2} a a^\dagger |3\rangle = 3^{1/2} 4^{1/2} a |4\rangle = 3^{1/2} 4^{1/2} 4^{1/2} |3\rangle \\ a^\dagger a a^\dagger |2\rangle &= 3^{1/2} a^\dagger a |3\rangle = 3^{1/2} N |3\rangle = 3^{1/2} 3 |3\rangle\end{aligned}$$

so that

$$\langle 3|X^3|2\rangle = \left(\frac{\hbar}{2m\omega}\right)^{3/2} [2(3^{1/2}) + 4(3^{1/2}) + 3(3^{1/2})]$$

What if we want not some matrix element of X , but the probability of finding the particle in $|n\rangle$ at position x ? We can of course fall back on Postulate III, which tells us to find the eigenvectors $|x\rangle$ of the matrix X [Eq. (7.4.32)] and evaluate the inner product $\langle x|n\rangle$. A more practical way will be developed in the next section.

Consider a remarkable feature of the above solution to the eigenvalue problem of H . Usually we work in the X basis and set up the eigenvalue problem (as a differential equation) by invoking Postulate II, which gives the action of X and P in the X basis ($X \rightarrow x$, $P \rightarrow -i\hbar d/dx$). In some cases (the linear potential problem), the P basis recommends itself, and then we use the Fourier-transformed version of Postulate II, namely, $X \rightarrow i\hbar d/dp$, $P \rightarrow p$. In the present case we could not transform this operator assignment to the energy eigenbasis, for to do so we first had to solve for the energy eigenfunctions in the X basis, which was begging the question. Instead we used just the commutation relation $[X, P] = i\hbar$, which follows from Postulate II, but is true in all bases, in particular the energy basis. Since we obtained the complete

solution given just this information, it would appear that the essence of Postulate II is just the commutator. This in fact is the case. In other words, we may trade our present Postulate II for a more general version:

Postulate II. The independent variables x and p of classical mechanics now become Hermitian operators X and P defined by the canonical commutator $[X, P] = i\hbar$. Dependent variables $\omega(x, p)$ are given by operators $\Omega = \omega(x \rightarrow X, p \rightarrow P)$.

To regain our old version, we go to the X basis. Clearly in its own basis $X \rightarrow x$. We must then pick P such that $[X, P] = i\hbar$. If we make the conventional choice $P = -i\hbar d/dx$, we meet this requirement and arrive at Postulate II as stated earlier. But the present version of Postulate II allows us some latitude in the choice of P , for we can add to $-i\hbar d/dx$ any function of x without altering the commutator: the assignment

$$X \xrightarrow[X \text{ basis}]{} x \quad (7.4.36a)$$

$$P \xrightarrow[X \text{ basis}]{} -i\hbar \frac{d}{dx} + f(x) \quad (7.4.36b)$$

is equally satisfactory. Now, it is not at all obvious that in every problem (and not just the harmonic oscillator) the same physics will obtain if we make this our starting point. For example if we project the eigenvalue equation

$$P|p\rangle = p|p\rangle \quad (7.4.37a)$$

onto the X basis, we now get

$$\left[-i\hbar \frac{d}{dx} + f(x) \right] \psi_p(x) = p \psi_p(x) \quad (7.4.37b)$$

from which it follows that $\psi_p(x)$ is no longer a plane wave $\propto e^{ipx/\hbar}$. How can the physics be the same as before? The answer is that the wave function is never measured directly. What we do measure are probabilities $|\langle \omega | \psi \rangle|^2$ for obtaining some result ω when Ω is measured, squares of matrix elements $|\langle \psi_1 | \Omega | \psi_2 \rangle|^2$, or the eigenvalue spectrum of operators such as the Hamiltonian. In one of the exercises that follows, you will be guided toward the proof that these measurable quantities are in fact left invariant under the change to the nontraditional operator assignment Eq. (7.4.36).

Dirac emphasized the close connection between the commutation rule

$$[X, P] = i\hbar$$

of the quantum operators and the Poisson brackets (PB) of their classical counterparts

$$\{x, p\} = 1$$