

Notes on Quantum Mechanics

Daniel V. Schroeder

Weber State University

January 2022



Copyright ©2015–2022, Daniel V. Schroeder.

Permission is granted to download and print this book for personal use only. Further distribution, in electronic or printed form, is prohibited without written permission from the author. All other rights reserved. Please check <http://physics.weber.edu/schroeder/quantum/> periodically for updates.

Contents

Preface	v
1 Wave mechanics	1
1.1 Einstein and de Broglie	1
1.2 Wavefunctions	5
1.3 Time evolution	15
1.4 The infinite square well	19
1.5 The Schrödinger equations	33
2 Bound states in one dimension	42
2.1 Qualitative solutions of the TISE	42
2.2 The shooting method	48
2.3 The matrix method	56
2.4 The harmonic oscillator	63
2.5 Multiple wells	69
3 Unbound states in one dimension	78
3.1 Momentum space	78
3.2 Wavepackets	82
3.3 Numerically solving the TDSE	88
3.4 Scattering from a barrier	93
3.5 Summary of one-dimensional wave mechanics	104
4 Beyond one dimension	107
4.1 Multiple spatial dimensions	107
4.2 Numerical methods for multiple dimensions	119
4.3 Multiple particles	126
4.4 Particles with internal structure	139
5 Working with operators	148
5.1 More about operators	148
5.2 Ladder operators for the harmonic oscillator	151
5.3 Compatible and incompatible observables	155
5.4 The principles of quantum mechanics	158

6 Central forces	161
6.1 Spherical coordinates	161
6.2 Angular momentum	164
6.3 Spherical harmonics	170
6.4 The radial equation	174
6.5 The hydrogen atom	179
6.6 The helium atom	186
7 Spin	192
7.1 Spin 1/2	193
7.2 Spins in magnetic fields	201
7.3 Photon polarization	205
7.4 Addition of angular momenta	207
A The mathematics of quantum mechanics	214
A.1 Probability and statistics	214
A.2 The Dirac delta function	218
A.3 Complex numbers	220
A.4 Vector spaces	225
A.5 Matrices	229
B Mathematica tutorial	235
B.1 Basic plots and basic syntax	235
B.2 Lists, tables, and animation	240
B.3 Complex numbers, functions, and plots	242
B.4 Plotting functions of two variables	243
B.5 Calculus and other math	248
B.6 Loops and simulations	254

Preface

There are two reasons to study quantum mechanics: it's *cool*, and it's *useful*.

Quantum mechanics is *cool* because it's far more bizarre than any idea ever dreamed up by science fiction writers. Particles behave like waves! Measurement outcomes are random! Energies can have only certain discrete values! Events in separated places have spooky correlations! There's a sense in which nobody truly "understands" quantum mechanics—and yet there's a definite logic to it, encoded in some of the most beautiful mathematics ever devised.

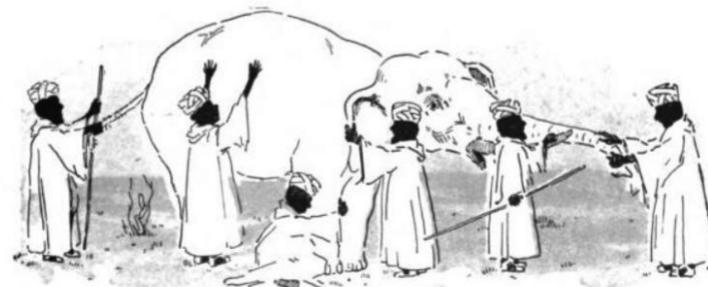
Meanwhile, quantum mechanics is *useful* because it governs the structure of all matter and interactions. It tells us why atoms are stable, why radioactive nuclei decay, and why carbon dioxide is a greenhouse gas. Quantum mechanics underlies all of chemistry, all of materials science, and all of subatomic physics. We can use quantum theory to design microscopes, lasers, solar cells, nuclear reactors, and secure methods of sending encrypted information. Scientists around the world are constantly discovering new applications of quantum mechanics.

My goal in this book is to introduce you to quantum mechanics in a way that honors *both* its coolness and its usefulness. Even if your initial interest is in only one or the other of these aspects, I hope you will find that each enhances the other. That has certainly been the case historically, as practical laboratory investigations gradually forced researchers to accept radical revisions to their world views—and then curiosity about quantum weirdness led to commercially promising applications of quantum information technology.

At the same time, I fear that this book might end up conveying *neither* the coolness nor the usefulness of quantum mechanics. That's because quantum mechanics is not an easy subject to learn, or to write about. How could it be easy, when it encompasses so much, while totally defying our common sense? I will need to ask you to work hard, and I have tried to do my part, working hard to prepare the pages that follow.

The best metaphor I know for the difficulty of learning quantum mechanics is the ancient Indian parable of the blind men trying to apprehend an elephant.¹ One of them feels a tusk and decides the elephant is like a spear. Another feels a leg and

¹This parable is apparently very old, and there are many versions. Probably the most famous English-language version is the delightful 1872 poem by John Godfrey Saxe, which you can find at https://en.wikisource.org/wiki/The_poems_of_John_Godfrey_Saxe/The_Blind_Men_and_the_Elephant.



An illustration of the parable of the blind men and the elephant, from C. M. Stebbins and M. H. Coolidge, *Golden Treasury Readers: Primer* (American Book Co., New York, 1909).

decides it is like a tree. Still others feel the trunk, the side, an ear, and the tail, deciding in turn that the elephant is like a snake, a wall, a fan, or a rope. They're all partially correct, but none of them can picture the whole elephant.

Similarly, quantum mechanics can appear to be many different things, depending on how we approach it—and because none of us experiences the quantum world directly through our senses we are all initially blind. But we have to start somewhere, and so I am faced with a choice: Which part of the quantum “elephant” should I take you to first?

My choice is the traditional one: We will begin by studying the quantum mechanics of a single, structureless particle moving in one dimension. This approach will bring us to some real-world applications fairly soon, and it lets you use your intuition for classical wave behavior to understand interference patterns and energy quantization. Mathematically, this part of quantum mechanics looks like a lot of ordinary and partial differential equations and their solutions. But as you go along, you'll also get some glimpses of matrix methods, linear algebra, and even two-state “qubit” systems. All of these aspects of quantum mechanics will reappear later in the book, and all can seem rather disconnected from wave mechanics. Part of my job is to help you see the connections.

Starting in Chapter 4, you'll see how to generalize the principles of quantum mechanics to multiple dimensions, multiple particles, and particles with internal structure such as spin. A big idea in all three cases is that we can't always describe these multiple degrees of freedom independently of each other, one at a time. In the case of multiple particles, we say that most of the allowed quantum states are “entangled.”

The later chapters of the book cover (or will cover) further applications (especially to atoms) and further exploration of quantum mysteries (especially entangled states). Unfortunately, in this introductory book I have had to leave out many advanced topics, both cool and useful, including path integrals, most aspects of relativistic systems, and the more intricate applications to chemistry, materials science, and quantum computing. I have also left out nearly all of the history of the

subject. If you want to see the elephant from all angles you will need to continue your studies.

One feature that sets this book apart is a rather heavy emphasis on computer-based numerical methods. These methods are more broadly useful, are usually easier to learn, and often provide deeper insight than some of the traditional pencil-and-paper tricks. Computer graphics can also help us visualize much of the mathematics—and I hope you'll like the pretty color representations of complex-valued functions. Awkwardly, though, using the computer as a tool means that I have had to choose a particular software platform. My choice is Wolfram Mathematica, which I find especially convenient for plotting formulas and doing short, interactive calculations. Although you could easily translate *some* of this book's computational exercises to other computing environments, in other cases the translation is rather difficult and not something that most students should attempt at the same time as learning quantum mechanics. For readers who are not already proficient in using Mathematica, I have provided an introductory tutorial in Appendix B.

You will also need some physics and math background. It will be helpful, though not absolutely necessary, if you are already somewhat familiar with the ideas of quantum physics, at the level of a standard “modern physics” textbook (or the modern physics chapters in an introductory physics textbook). You should also be comfortable with the rest of what's covered in a year-long introductory physics course, and with calculus, differential equations, and linear algebra.

As of this writing, this book is still unfinished and quite unsuitable for formal publication. It needs more problems, more applications, more advanced topics, and an index. Still, I am posting this draft for others to use in the hope that it is better than nothing, offering a somewhat fresh approach to one of the most profound and important subjects in all of human knowledge. I would welcome feedback sent to dschroeder@weber.edu.

Acknowledgments

It is a pleasure to acknowledge the teachers and authors who have most influenced my understanding of quantum mechanics and, hence, this draft textbook: Mike Casper, Daniel Gillespie, David Griffiths, Frank Marsiglio, Thomas Moore, Ramamurti Shankar, David Sipfle, and Daniel Styer. I apologize to the others who are omitted from this list due to my poor memory. I am also grateful to countless students for their enthusiasm and their questions, always pushing me to rethink how to teach quantum mechanics.

Daniel V. Schroeder
Ogden, Utah
January 2022

Chapter 1

Wave mechanics

1.1 Einstein and de Broglie

Quantum mechanics began with two deceptively simple formulas:

$$E = hf \quad \text{and} \quad p = \frac{h}{\lambda}. \quad (1.1)$$

I'll refer to these as the Einstein and de Broglie relations, respectively, after the two physicists most responsible for introducing them, in the early 20th century. You may already be familiar with both of these equations, but let me summarize where they came from and what they tell us.

The Einstein relation, $E = hf$, says that a particle's energy E is proportional to its frequency f . In 1905 Albert Einstein proposed this relation for light, introducing the radical idea that light comes in discrete lumps, now called photons, each with an energy determined by the light's frequency. So a blue photon (high frequency) has more energy than a red photon (low frequency), while a gamma-ray photon has far more energy than either, and a radio-wave photon has much less. The first direct experimental evidence for the Einstein relation came from the photoelectric effect, in which high-frequency light, aimed at a metal surface, ejects electrons with more energy than low-frequency light. (Each electron, it turns out, absorbs the energy of just one photon from the light.) If you plot the energy vs. the frequency (see Problem 1.4) you get a straight line, whose slope is the constant of proportionality,

$$h = 6.63 \times 10^{-34} \text{ J s} = 4.14 \times 10^{-15} \text{ eV s}, \quad (1.2)$$

called *Planck's constant*.

But the Einstein relation doesn't just apply to particles of light; it applies equally well to electrons, protons, quarks, neutrinos, and (as far as we know) baseballs. Defining what we mean by the "frequency" of one of these particles is trickier than for light, so in introductory courses we usually stick to photons—and warn you not to apply $E = hf$ to anything else unless you know what you're doing. In Section 1.3 I'll show you how to apply the Einstein relation to any particle at all.

The de Broglie relation, $p = h/\lambda$, says that a particle's momentum p is inversely proportional to its wavelength λ . For photons, this relation is a straightforward consequence of $E = hf$ (since a light wave has $p = E/c$ and $f = c/\lambda$, where c is the speed of light). But in 1924 Louis de Broglie proposed that *every* particle has a wavelength that's inversely proportional to its momentum, with the same universal constant of proportionality, h . The wavelength of a pitched baseball (large p) is far too tiny to measure, but for low-mass particles such as electrons (small p), it's not hard to measure the wavelength in a diffraction experiment—as Clinton Davisson and Lester Germer did, by accident, soon after de Broglie's proposal. More recent experiments have measured the wavelengths of all sorts of subatomic particles, as well as entire atoms and molecules.

Diffraction experiments with *particles*, however, are extremely odd, because each particle can land on the detector in only one place. Figure 1.1 shows a sequence of actual photographs from a “two-slit” interference experiment performed with electrons, with the beam current increasing from left to right. At low beam currents, you can see the distinct blips (dots) left by individual electrons on the detection screen, in apparently random locations. At higher currents, the familiar maxima and minima of the interference pattern emerge, allowing us to determine the wavelength λ from the size of the pattern. Thus, it appears that λ (together with the experimental geometry) determines the *probability* of an electron arriving in any particular location. Randomness and probabilities seem to be inherent in the de Broglie relation. Interference experiments with photons yield similar results: random blips at low intensity, with the wavelength-dependent pattern emerging at higher intensity.¹ Explaining how a particle can be localized under some conditions but spread-out and wavelike under other conditions is quite a challenge, which I'll begin to address in the next section.

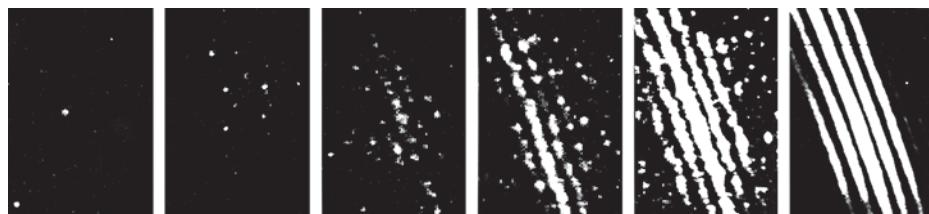


Figure 1.1: A sequence of photographs of an electron interference pattern, with the beam current increasing from left to right. Instead of a pair of slits, this experiment actually used a positively charged wire, perpendicular to the electron beam, allowing the electrons to pass the wire on either side and then interfere before hitting the detection screen. P. G. Merli, G. F. Missiroli, and G. Pozzi, “On the statistical aspect of electron interference phenomena,” Am. J. Phys. **44**(3), 306–307 (1976), <http://dx.doi.org/10.1119/1.10184>.

¹See, for example, R. S. Aspden, M. J. Padgett, and G. C. Spalding, “Video recording true single-photon double-slit interference,” Am. J. Phys. **84**(9), 671–677 (2016), <http://dx.doi.org/10.1119/1.4955173>.

The similarity of the Einstein and de Broglie relations becomes more apparent if we express the former in terms of the period of the wave, $T = 1/f$:

$$E = \frac{h}{T} \quad \text{and} \quad p = \frac{h}{\lambda}. \quad (1.3)$$

In other words, energy is to time (period) as momentum is to space (wavelength). Alternatively, we can write these relations in terms of the angular frequency, $\omega = 2\pi f = 2\pi/T$, and the angular wavenumber, $k = 2\pi/\lambda$. For convenience, we usually absorb the factors of 2π into Planck's *reduced constant*,

$$\hbar = \frac{h}{2\pi} = 1.05 \times 10^{-34} \text{ Js} = 6.58 \times 10^{-16} \text{ eV s} \quad (1.4)$$

(pronounced "h bar"), so we have simply

$$E = \hbar\omega \quad \text{and} \quad p = \hbar k. \quad (1.5)$$

Versions 1.1, 1.3, and 1.5 express exactly the same physics, so feel free to use whichever seems most convenient in any given situation.

Problem 1.1. In problems involving photons the combination hc appears frequently. Calculate the value of this combination in units of eV-nm, and make a note of the result so you can use it to simplify the arithmetic in future calculations.

Problem 1.2. Estimate (to within a factor of 2) the number of photons given off by a 10-watt LED light bulb in one second. How would your answer change for a 60-watt incandescent bulb, whose light looks the same to the eye?

Problem 1.3. Use the Einstein relation to explain why it is unlikely that the radiation from mobile phones could cause cancer.

Problem 1.4. In the introductory physics labs at Weber State University, students study the photoelectric effect by measuring (with a high-impedance voltmeter) the voltage that builds up across a phototube when the cathode is illuminated by five different spectral lines from a mercury light source. Here is some actual data that I obtained using this equipment:

Color	Wavelength (nm)	Voltage (V)
Yellow	578	0.707
Green	546	0.812
Violet	436	1.424
Violet	405	1.638
Near UV	365	1.913

Use this data (or your own, if you've done the experiment yourself) to plot a graph of voltage vs. frequency. From this graph, make your best estimate of Planck's constant. Make your reasoning clear, and express your answer both in Js and eVs.

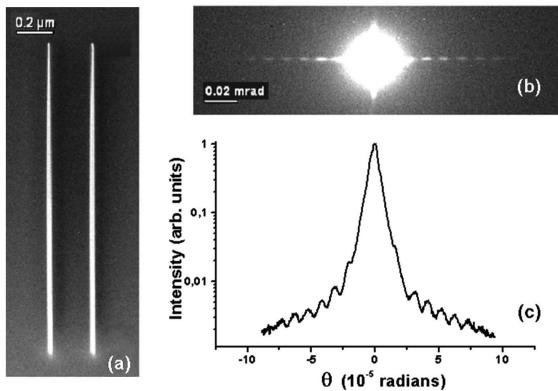


Figure 1.2: A double slit (a), interference pattern (b), and intensity plot (c), from S. Frabboni, G. C. Gazzadi, and G. Pozzi, “Young’s double-slit interference experiment with electrons,” Am. J. Phys. **75**(11), 1053–1055 (2007), <http://dx.doi.org/10.1119/1.2757621>.

Problem 1.5. Calculate the approximate de Broglie wavelength of each of the following: (a) a microwave photon, with an energy of 10^{-5} eV; (b) an electron inside the magnetron that creates the microwaves, which has been accelerated through a voltage difference of 5 kV; (c) a rubidium-87 atom in a magneto-optical trap, with a kinetic energy of 10^{-11} eV; (d) a proton at the Large Hadron Collider with a kinetic energy of 6.5 TeV; (e) a gnat with a mass of 0.2 mg, flying at a speed of 0.5 m/s. (Hint: Some of these particles are nonrelativistic, with speed v much less than c and kinetic energy K much less than mc^2 , so that $K \approx \frac{1}{2}mv^2 = p^2/(2m)$. Others are ultra-relativistic, moving at or near the speed of light, with energy much greater than mc^2 so that $E \approx pc$. Be sure to choose the correct energy-momentum relation in each case, and to justify your choice.)

Problem 1.6. Figure 1.2 shows a recent electron two-slit interference experiment. At left (a) is an image of the two slits, fabricated in a 500-nm thick silicon nitride membrane using a focused ion beam. The beam of a transmission electron microscope was then aimed at the slits, and the resulting interference pattern imaged (b) with a CCD camera. Panel (c) shows the intensity of the pattern in the CCD image. Please ignore the central bright spot, which was due to the beam traveling directly through the partially transparent membrane.

- Measure the slit spacing and the angle between interference fringes directly from the images (using the scales shown), and use these numbers, together with what you remember about two-slit interference, to determine the wavelength of the electrons.
- Each electron in this beam had a kinetic energy of 200 keV. Is your measured wavelength consistent with the de Broglie relation? (Hint: 200 keV is barely small enough compared to the electron’s rest energy of 511 keV that it’s reasonable to use the nonrelativistic energy-momentum relation $E^2 = (mc^2)^2 + (pc)^2$, where E is the sum of the electron’s kinetic energy and rest energy.)

1.2 Wavefunctions

To create a mathematical theory of the wave properties of particles, we introduce the concept of a *wavefunction*: a function of space that encodes the current state of a system.

For now, let's assume that the “system” consists of a single particle living in a one-dimensional universe. (I'll generalize to more complicated systems in Chapter 4.) Then, if this were a *classical* particle, the “state” of the system would consist of just two numbers: its position x and its momentum p_x (or velocity v_x , which you can easily calculate from the momentum). (You can think of these two numbers as the initial conditions that you would need in order to predict the particle's full trajectory using Newton's second law.) For a *quantum* particle, the “state” instead consists of the *wavefunction* $\psi(x)$, a whole infinity of numbers (one for each x). Quantum states are vastly more complicated, and interesting, than classical states.

An example wavefunction

For example, if we draw the x axis across the two-slit interference pattern in Figure 1.1 (and ignore the other two dimensions of space), then the wavefunction of each particle, just before it hits the detection screen, might look something like what I've drawn in Figure 1.3(a). This wavefunction has five “bumps,” corresponding to the five bright lines in the interference pattern. The dark lines in the pattern are at the locations where the wavefunction is zero. More precisely, the brightness of the interference pattern is proportional to the *square* of the wavefunction, in analogy to the way the energy in a mechanical or electromagnetic wave is proportional to the square of the wave amplitude. Figure 1.3(b) shows the square of our five-bump wavefunction. The height of this graph at any x is then proportional to the probability of finding the particle at x , when the subsequent interaction with the detection screen “measures” the particle's position. After many such measurements are made on identically prepared particles, the five-line interference pattern emerges.

Computing probabilities

Because x is a continuous variable, the actual probability of finding a particle at any *exact* value of x is infinitesimal. To get around this awkwardness, we can instead ask for the probability of finding a particle *between* two values of x . We obtain this probability by calculating the corresponding area under the graph of $|\psi|^2$. For a generic wavefunction, you can visualize the calculation as shown in Figure 1.4. Written as an equation, the rule for calculating probabilities is to integrate:

$$\left(\begin{array}{l} \text{Probability of finding} \\ \text{particle between } a \text{ and } b \end{array} \right) = \int_a^b |\psi(x)|^2 dx. \quad (1.6)$$

This formula is called the *Born rule*, after German physicist Max Born (1882–1970), one of the founders of quantum mechanics. The function $|\psi(x)|^2$ is called the

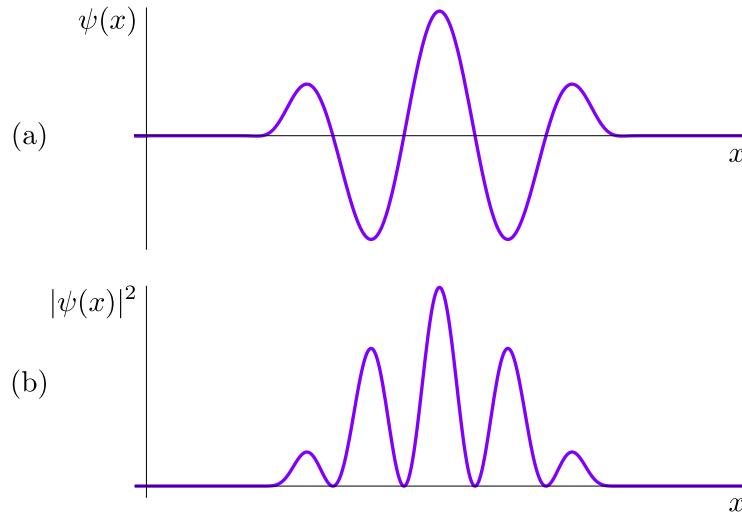


Figure 1.3: An example wavefunction (a) and the corresponding probability density (b).

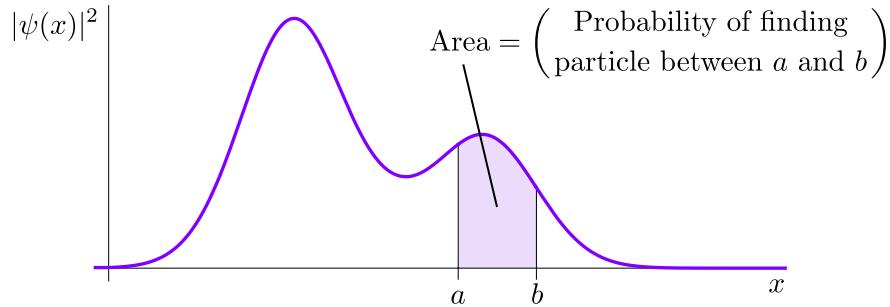


Figure 1.4: The area under any interval of the graph of $|\psi(x)|^2$ is the probability of finding the particle within that interval.

probability density, and I like to think of it as a function *whose purpose in life is to be integrated*. In order for the rule to work, however, we must impose the condition that the total probability of finding the particle *somewhere* equals exactly 100%:

$$\int_{-\infty}^{\infty} |\psi(x)|^2 dx = 1. \quad (1.7)$$

Any function ψ that satisfies this condition is said to be *normalized*.

Problem 1.7. For the probability density plotted in Figure 1.4, make a rough graphical estimate of the probability of finding the particle between a and b .

Problem 1.8. What are the units (dimensions) of ψ ? Explain your answer.

Problem 1.9. Suppose that the wavefunction of a particle in one dimension has the form

$$\psi(x) = \begin{cases} Ax(1-x)e^{-x} & \text{for } x > 0, \\ 0 & \text{for } x < 0, \end{cases}$$

where A is a normalization constant and x is measured in some suitable distance unit. (A wavefunction similar to this one will come up in Chapter 6, but don't worry for now about how we might prepare a particle in this wavefunction.)

- (a) Find the normalization constant A . (I suggest doing the integral with Mathematica; see Appendix B.5 for guidance if necessary.)
- (b) Plot graphs of $\psi(x)$ and $|\psi(x)|^2$, showing the full range of x values where ψ isn't negligibly small. (Again I suggest using Mathematica; see Appendix B.1 for guidance if necessary.)
- (c) Suppose now that you measure this particle's position. Make a rough graphical estimate of the probability of finding it at $x < 1$. Explain your reasoning.
- (d) Now do a careful calculation to check your estimate in part (c).
- (e) At (or near) what value of x are you *most* likely to find this particle? Justify your answer.

Computing averages

Sometimes, instead of calculating the probability of finding a particle at (or near) a particular location, we want to know its *average* position. For example, the average x for the five-bump wavefunction drawn above would be in the middle of the central bump, while the average x for the function drawn in Figure 1.4 would be somewhere to the right of the highest peak but to the left of $x = a$. I'll use the symbol $\langle x \rangle$ to denote the average value of x .

To figure out the general formula for $\langle x \rangle$, it's helpful to first imagine that the values of x are discrete, separated by intervals of dx , so that the probability of any particular x_i is $|\psi(x_i)|^2 dx$. Then the average x is just the sum of all the possible values x_i , weighted by their probabilities:

$$\langle x \rangle = \sum_i x_i |\psi(x_i)|^2 dx \longrightarrow \int_{-\infty}^{\infty} x |\psi(x)|^2 dx, \quad (1.8)$$

where in the last expression I've taken the continuum limit.

You can use a similar formula to calculate the average value of any *function* of x :

$$\langle f(x) \rangle = \int_{-\infty}^{\infty} f(x) |\psi(x)|^2 dx. \quad (1.9)$$

A common example would be to compute the average value of x^2 , which lets you then calculate the *standard deviation* of x , denoted σ_x :

$$\sigma_x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}. \quad (1.10)$$

This is a rough measure of the “width” or “spread” of a wavefunction.²

Although we commonly say that these average values apply to a single particle whose wavefunction is ψ , it’s important to remember that if you actually *measure* the value of x , you’ll get some random value that may or may not be near the average, $\langle x \rangle$. It’s therefore impossible to determine $\langle x \rangle$ (or $\langle x^2 \rangle$ or σ_x) with a single measurement. To determine these quantities in the laboratory you must first prepare *many* identical particles in the same wavefunction ψ , then measure all their positions and calculate the averages.

Problem 1.10. Find the average value of x for the wavefunction of Problem 1.9. Show graphically that your answer is reasonable, with a brief explanation.

Localized wavefunctions

Any normalized function of x is an allowed wavefunction, which you could use to describe a quantum particle under suitable conditions. By a *function*, I mean any single-valued graph that you can draw on a set of axes—regardless of whether you know how to write down a formula for it.

Sometimes a quantum particle is localized within a very small region of space. To describe such a particle we would use a “spiky” wavefunction that’s zero everywhere except in that narrow region, as shown in Figure 1.5. You could use a variety of formulas for such a function, but often the exact shape doesn’t matter, and it’s convenient to take the limit where the spike becomes infinitely narrow and infinitely tall to become a *Dirac delta function*,³ denoted $\delta(x - a)$. Borrowing a word from German, we say that a delta function is an *eigenfunction* (which could be translated “characteristic” or “particular” function) of position, meaning that it’s a function for which the particle’s position is precisely defined. A delta function is *not* a properly normalized wavefunction, however.

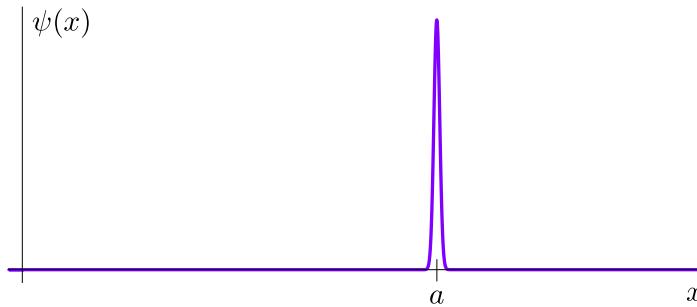


Figure 1.5: A sharp, narrow wavefunction, representing a particle whose position is approximately well defined.

²For a review of probability and statistics, including the concept of a standard deviation, see Appendix A.1.

³If you’re not already familiar with the Dirac delta function, please see Appendix A.2.

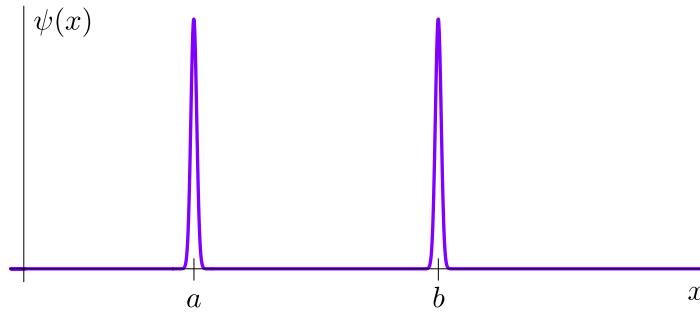


Figure 1.6: A wavefunction with two narrow spikes and a gap in between. If you measure the position of such a particle you are equally likely to find it near a or b .

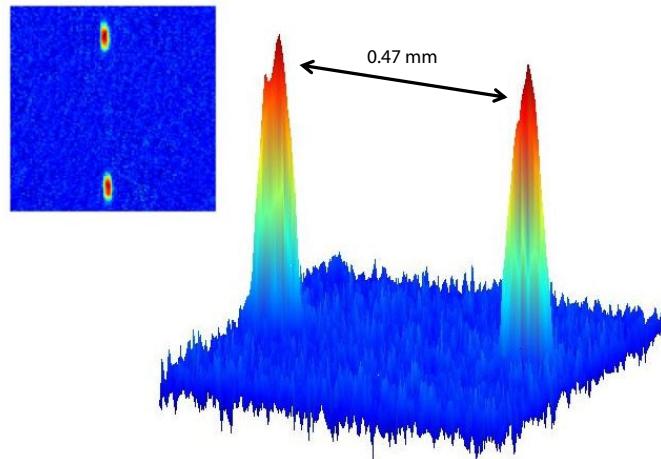


Figure 1.7: Distribution of measured positions for a collection of about 10,000 identically prepared rubidium atoms, each with a wavefunction consisting of two narrow peaks, separated by about half a millimeter. From J. H. Burke, “Improvements and Applications of a Guided-Wave Bose-Einstein Condensate Interferometer,” PhD thesis, University of Virginia, 2010, http://galileo.phys.virginia.edu/research/groups/sackett/publications/Burke_PhD.pdf.

A wavefunction with a single narrow spike is in some ways like a classical particle. But an equally valid wavefunction could have *two* narrow spikes, separated by a gap, as shown in Figure 1.6. This wavefunction has no classical counterpart, because it describes a particle that is (apparently) half in one location and half in another—and unambiguously *not* half way in-between the two locations! We sometimes refer to such a wavefunction as a *cat state*, in honor of Schrödinger’s cat, a hypothetical creature that is (supposedly) in a state in which it is half alive and half dead.

However, just because a wavefunction is mathematically *allowed* doesn’t mean it’s easy to physically *prepare* a particle to have that wavefunction—at least in such a way that you can then verify it in the laboratory. Putting a particle into

a two-peak “cat state” is pretty easy if the particle is a photon (just aim it at a screen with a double slit, or at a semi-reflective mirror that causes half to reflect and half to pass through), and becomes progressively more difficult for heavier particles. Figure 1.7 shows the result of a recent state-of-the-art experiment in which a collection of about 10,000 identically prepared rubidium atoms were each split between two distinct locations separated by about half a millimeter. Scaling up such an experiment from rubidium atoms to cats is, as far as anyone knows, science fiction.

Momentum eigenfunctions

Completely different from a spiky, localized wavefunction would be one that describes a particle with well-defined *momentum*. For this purpose, according to the de Broglie relation, we should use a periodic function with wavelength $\lambda = h/p$. The simplest periodic function would be a sine or a cosine, as shown in Figure 1.8. A long wavelength would correspond to a small momentum, and a short wavelength would correspond to a large momentum. A true sinusoidal function would repeat forever in both directions, and therefore wouldn’t be normalizable, so we should remember in the back of our minds that the oscillations must eventually die out in both directions. As we’ll see later, this means that no normalized wavefunction has a *perfectly* well-defined momentum; a perfectly well-defined momentum is an idealized but unphysical limiting case, just as perfectly well-defined position, described by a delta function, is an idealized limiting case.

But besides the issue of normalization, our sinusoidal wavefunction has two other awkward features. First, the square of this wavefunction alternates back and forth between some maximum value and zero, so it describes a particle whose probability density varies periodically with location. There’s no reason a particle can’t have such a wavefunction, but surely there must be a way to describe a particle with well-defined momentum whose probability density is more uniform. Second, our sinusoidal wavefunction doesn’t seem to encode the *direction* of the particle’s momentum; it would apparently look the same whether p_x is positive or negative

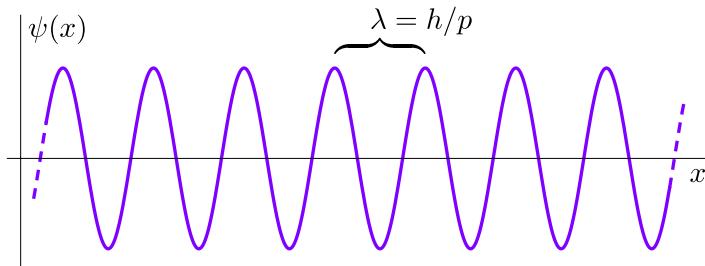


Figure 1.8: A sinusoidal function with well-defined wavelength and therefore a well-defined momentum magnitude, $|p| = h/\lambda$. However, the direction of the momentum is not well defined.

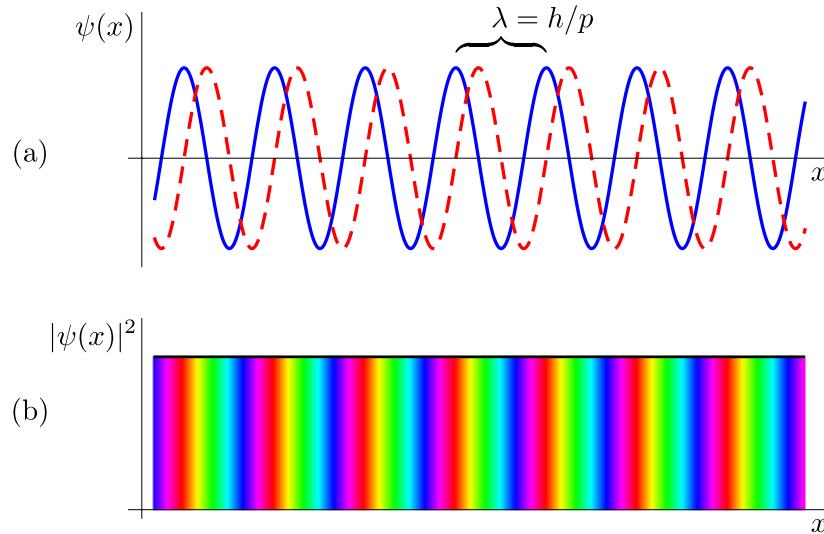


Figure 1.9: A wavefunction with a definite value of the momentum p_x , which in this case is positive. (a) Here the two components are plotted as solid and dashed lines, respectively, with the correct quarter-cycle phase offset to represent a positive-momentum eigenfunction. (b) Another visual representation of the same function, with the squared magnitude (which is constant) plotted vertically and the space below the plot filled in with colors to indicate the signs and relative sizes of the two components. More precisely, the color hue indicates the phase angle in the complex plane, as explained in the following subsection.

(since λ can only be positive).

The way to fix both of these defects is to give the wavefunction a second component at each location in space—that is, to make it a *pair* of functions instead of just one. Then, to describe a particle with well-defined p_x and a uniform probability density, we use two sinusoidal functions that are out of phase by a quarter cycle ($\pi/2$, or 90°), so that one component is large in magnitude where the other is zero and vice-versa. And to distinguish left-moving from right-moving particles, we associate a quarter-cycle phase difference in one direction with positive p_x , and a quarter-cycle phase difference in the other direction with negative p_x . Figure 1.9(a) shows a plot of both components, where I've used a solid line for the first component and a dashed line for the second component. By convention, we associate this particular phase relationship (second component a quarter-cycle to the *right* of the first component) with a *right*-moving wavefunction, that is, with *positive* p_x . To draw the analogous *left*-moving wavefunction, you would shift either component left or right by a half cycle, or, equivalently, flip either component upside-down.

Complex notation

One way to write a formula for this two-component wavefunction would be to just list the two components in order:

$$\psi(x) = \{A \cos(kx + \phi), A \sin(kx + \phi)\}, \quad (1.11)$$

where A is the wave amplitude, ϕ is a phase constant that determines where the cycles start, and $k = 2\pi/\lambda$. Notice that the two components share the same A , ϕ , and k , and that the sum of the squares of the components is a constant, A^2 . But we can write this formula much more concisely using *complex numbers*.⁴ By convention, we associate the two components of ψ with the so-called “real” and “imaginary” parts of a single complex-valued function:

$$\psi(x) = A \cos(kx + \phi) + iA \sin(kx + \phi). \quad (1.12)$$

You may now recognize that $\cos \theta + i \sin \theta$ is the *definition* of the complex exponential function, $e^{i\theta}$, so we can write simply

$$\psi(x) = A e^{i(kx+\phi)}. \quad (1.13)$$

Moreover, if we factor out $e^{i\phi}$ from the exponential, we recognize that this factor is just a complex constant, and so without loss of generality we can absorb it into the amplitude A :

$$\psi(x) = A e^{ikx}, \quad (1.14)$$

where A is now complex.

Again, this is the formula for a wavefunction with definite momentum, also called a momentum eigenfunction. Specifically, the momentum value corresponding to this function is

$$p_x = \frac{h}{\lambda} = \frac{hk}{2\pi} = \hbar k. \quad (1.15)$$

Moreover, as I said above, equation 1.14, with positive k , describes a particle with positive p_x . But if we allow k to be negative, then equations 1.14 and 1.15 actually work for negative p_x as well: the “imaginary” component of ψ is then a quarter cycle to the left of the “real” part, as desired. In essence, we have now implemented the de Broglie relation in the precise language of wavefunctions, by identifying the periodic function $Ae^{ip_xx/\hbar}$ as the momentum eigenfunction corresponding to momentum value p_x .

Momentum eigenfunctions aren’t the only wavefunctions that are complex. Virtually all wavefunctions are complex, with two separate components that (in general) needn’t be related to each other in any particular way:

$$\psi(x) = \text{Re } \psi(x) + i \text{Im } \psi(x), \quad (1.16)$$

⁴For a review of the properties of complex numbers, see Appendix A.3.

where Re and Im denote the “real” and “imaginary” parts, respectively. I started with examples of real-valued functions (for which $\text{Im } \psi = 0$) only for simplicity. But I need to go back and explain more carefully how to “square” a complex-valued wavefunction, so you’ll know how to compute probabilities. The probability density, $|\psi|^2$, is actually the square *modulus* of ψ , defined as

$$|\psi|^2 = \psi^* \psi = (\text{Re } \psi - i \text{Im } \psi)(\text{Re } \psi + i \text{Im } \psi) = (\text{Re } \psi)^2 + (\text{Im } \psi)^2. \quad (1.17)$$

Here $*$ indicates the complex conjugate, in which the “real” part is the same and the “imaginary” part has opposite sign. If you think of ψ (at a given x) as a two-component vector, then $|\psi|^2$ is just the squared magnitude of that vector (which is always a real, nonnegative number). Notice again that $|\psi|^2$ is constant for the momentum eigenstate of equation 1.14.

To visualize a complex-valued wavefunction you can always just plot its “real” and “imaginary” parts separately, as I did in Figure 1.9(a). A useful (and pretty) alternative, though, is to plot $|\psi|^2$ (or sometimes just $|\psi|$) on the vertical axis and then fill in the area under the plot with color hues to indicate the relative sizes and signs of the “real” and “imaginary” parts. The scheme that I’ll use in this book assigns red to a number that’s “real” and positive, light green to “imaginary” and positive, cyan to “real” and negative, purple to “imaginary” and negative, and intermediate hues to numbers that have both “real” and “imaginary” parts. More precisely, if you imagine plotting each function value in the complex plane and then describing its location in polar coordinates, then I’m mapping the angular coordinate, which mathematicians call the *argument* of the number but physicists simply call the *phase*, to the hue as normally defined in modern computer graphics systems.⁵ Figure 1.9(b) shows a phase-as-color plot of the same momentum eigenfunction as in Figure 1.9(a). Figure 1.10 shows a more generic function plotted in both ways.

Incidentally, I’ve been putting scare quotes around the words “real” and “imaginary” because neither part of the wavefunction is any more real or imaginary than the other, in the everyday meanings of these words. Both contribute equally to the probability density. Whether wavefunctions are actually real physical entities or merely imaginary human constructs is another question entirely, which physicists and philosophers have debated since 1926.

In any case, given an arbitrary complex-valued wavefunction $\psi(x)$, you can compute $|\psi(x)|^2$ and then check whether it’s normalized according to equation 1.7. If it is, then you can integrate it over any desired interval (according to equation 1.6) to obtain the probability of finding the particle within that interval.

There is a similar formula, which I’ll give you in Chapter 3, for finding the probabilities of various outcomes of *momentum* measurements.

Problem 1.11. Sketch a portion (a few cycles) of the wavefunction of a particle in one dimension with a well-defined momentum that is negative. Sketch the real and

⁵ Appendix B.3 shows how to produce plots of this type using Mathematica.

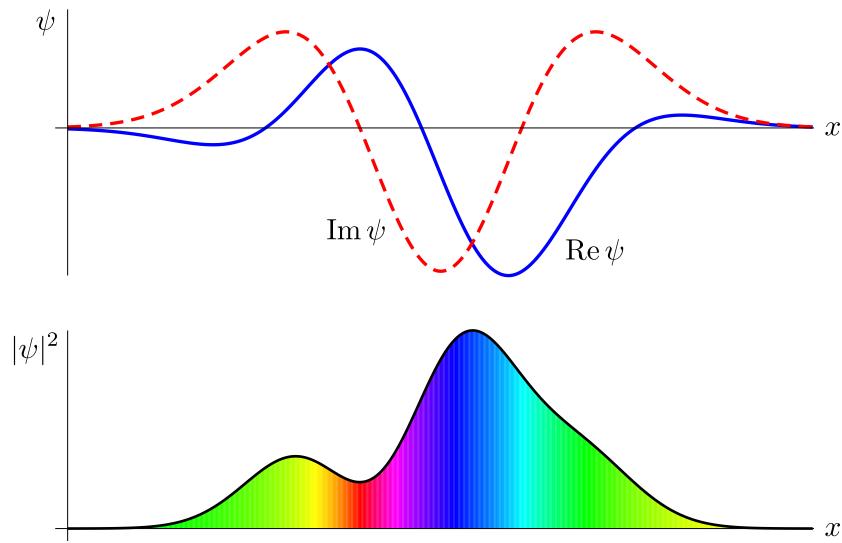


Figure 1.10: An arbitrary complex wavefunction can be plotted either as separate real and imaginary parts (top) or as a probability density, filled underneath with color hues to indicate the complex phase (bottom).

imaginary parts of the wavefunction on the same graph, using different colors or dashing patterns. Then, underneath this graph, use colored pencils to make a density-phase plot of the same wavefunction, representing the phases by colors as in Figures 1.9 and 1.10. Be sure that the various colors line up correctly with the places where the wavefunction is pure real and positive, pure imaginary and positive, pure real and negative, and so on.

Problem 1.12. Prove that you can multiply a wavefunction by $e^{i\phi}$, for any real number ϕ , without affecting the probabilities of any position measurement outcomes. We often call such a multiplicative factor a “pure constant phase.” What are some interesting special cases of pure constant phase factors? (As we’ll see later, the same is true for measurements of other quantities. Therefore a wavefunction’s overall phase factor is completely unphysical and you can choose it to suit your convenience.)

The effect of a measurement

And now we come to the most important fact about wavefunctions and measurements. Immediately *after* you perform a measurement, the measured quantity (for example, position or momentum) will be well defined, and so the particle’s wavefunction must be the corresponding eigenfunction. If you measured the particle’s position and found it at $x = a$, then the particle’s wavefunction is now the eigenfunction $\delta(x - a)$ (at least approximately, and never mind the normalization). Or, if you measured the particle’s momentum and found it to have $p = p_a$, then the particle’s wavefunction is now the eigenfunction $e^{ip_a x/\hbar}$ (at least approximately, and never mind the normalization). Either way, we say that the measurement process

causes the wavefunction to *collapse*, in a random way, into whatever eigenfunction corresponds to the measurement outcome.

But if a measurement causes the wavefunction to “collapse,” then what exactly constitutes a measurement? Amazingly, quantum mechanics provides no clean answer to this question. Fortunately, though, we won’t need a clean answer in order to apply quantum mechanics to laboratory experiments. The messy answer is that a measurement always involves some sort of interaction between the quantum system you’re studying and the apparatus you’re using to study it. Usually this apparatus is much larger than the system, so you can think of it as a classical (that is, non-quantum) object. Fundamentally, of course, quantum mechanics should apply to everything in your laboratory, including your measurement apparatus. For that matter, quantum mechanics should also apply to you and to everything else in the universe. How, then, should we think about quantum measurements fundamentally? Does the universe have a wavefunction and if so, does it ever “collapse”? These questions are fun to contemplate, and there is still no consensus on the answers.

Problem 1.13. A particle in a one-dimensional universe has the wavefunction $\psi(x) = Ae^{-\alpha x^2}$, where $\alpha = 1.2 \times 10^{18}$ in SI units.

- (a) What are the units of α ? (Explain briefly.)
- (b) What is the value of the normalization constant A , both exactly in terms of α and π , and as an actual number, with units? (Hint: To carry out the integral, change to a dimensionless variable y so the exponent in the integrand is simply $-y^2$. If the integral isn’t familiar to you, do it with Mathematica. If this hint seems unnecessary, please set aside your skepticism and follow it anyway, because it will make the next part of the problem easier.)
- (c) Suppose that you now measure the position of this particle. What is the probability of finding it between $x = 0$ and $x = +0.5$ nm? (Hint: Make the same change of variables as in part (b), simplifying the expression algebraically as much as you can before attempting to carry out the integral. Then do the integral numerically with Mathematica. As a reality check, make a graphical estimate of the answer.)
- (d) Suppose that when you measure the position, you find the particle at $x = -0.4$ nm. Describe the wavefunction of the particle immediately after this measurement.

1.3 Time evolution

So far I’ve been discussing wavefunctions only as functions of space, at a single instant in time. What happens to a wavefunction as time passes?

The Einstein relation, $E = hf$, gives us a clue: If the wavefunction represents a state in which the particle has energy E , then it will *oscillate* in time at frequency $f = E/h$ (or angular frequency $\omega = E/\hbar$).

Most wavefunctions do *not* represent particles with well-defined energy, so this clue doesn’t tell us directly how they change over time. But for a *free* particle, with no potential energy (and therefore acted upon by no forces), the energy is entirely kinetic. In this case, the momentum eigenfunctions Ae^{ikx} , described in the previous

section, must also be energy eigenfunctions, because knowing the momentum means you can calculate the energy.

And how exactly do we take the wavefunction Ae^{ikx} and make it oscillate in time at a given frequency? The answer is simple and elegant: we multiply it by $e^{-i\omega t}$, to obtain

$$\psi(x, t) = Ae^{ikx}e^{-i\omega t} = Ae^{i(kx - \omega t)}. \quad (1.18)$$

This is the general formula for the full time-dependent wavefunction of a free particle with well-defined momentum, and it takes the familiar form of a *traveling wave*: a function only of the combination $kx - \omega t$. The individual “ripples” in the wave move at the so-called phase velocity,

$$v_{\text{phase}} = \frac{\omega}{k} = \frac{E/\hbar}{p_x/\hbar} = \frac{E}{p_x}. \quad (1.19)$$

The time-dependent factor $e^{-i\omega t}$ will come up repeatedly in this book, so let me give it a name: the *wiggle factor*.⁶ The wiggle factor must be a complex exponential (and not merely a sine or a cosine) in order to get the simple traveling-wave form in equation 1.18 (that is, a function only of the combination $kx - \omega t$), and also because a sine or cosine would change the wavefunction’s normalization as time passes, violating probability conservation. The minus sign in the exponent of the wiggle factor is required in order for a positive- k traveling wave to travel to the *right*.⁷ You can visualize the wiggle factor as a unit-length “phaser” in the complex plane, rotating clockwise⁸ with angular velocity $\omega = E/\hbar$, as shown in Figure 1.11. The higher the energy, the faster the rotation.

Equations 1.18 and 1.19 work for both nonrelativistic and relativistic particles—the only difference being the formula that relates the momentum p_x to the (kinetic) energy E . The simplest case is actually for an ultrarelativistic (massless) particle such as the photon, for which $E = |p_x|c$ and therefore $v_{\text{phase}} = \pm c$, as you probably would have guessed. For a nonrelativistic particle, on the other hand, the result may surprise you. In this case, if the particle’s mass is m , then $E = p_x^2/2m$ and we obtain

$$v_{\text{phase}} = \frac{p_x^2/2m}{p_x} = \frac{p_x}{2m} = \frac{v_x}{2}, \quad (1.20)$$

where v_x is the particle’s velocity. The phase velocity of the wavefunction is only *half* the velocity of the particle! This may seem like nonsense, but it’s not, because you don’t measure a particle’s velocity by measuring the phase velocity of the ripples within its wavefunction.

To better visualize the time dependence of the momentum eigenfunctions, please take some time now to run the Sinusoidal Wave animation at <http://physics>.

⁶I learned this name from one of my college professors, Mike Casper.

⁷However, the plus sign in e^{+ikx} was a mere convention, so we *could* have instead chosen a minus sign there and a plus sign in the wiggle factor. Nobody actually does it that way, so please stick to the standard sign convention.

⁸When E includes potential energy, as we will consider soon, it can sometimes be negative. Then ω is also negative and the phasor rotation is counter-clockwise.

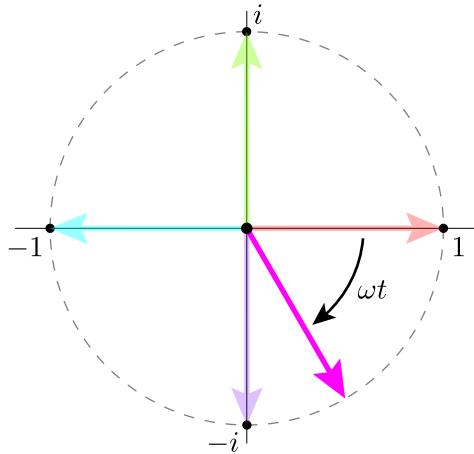


Figure 1.11: The wiggle factor $e^{-i\omega t}$ is a complex number of unit magnitude that rotates clockwise at angular velocity ω in the complex plane. (If ω is negative then the rotation is counter-clockwise.)

[weber.edu/schroeder/software/SinusoidalWave.html](http://physics.weber.edu/schroeder/software/SinusoidalWave.html). Notice how the wavelength, frequency, and phase velocity all change as you adjust the momentum slider. Also, as a review of the previous section, notice the difference in the phase relationships between wavefunctions with positive and negative momentum. To see how the phase velocity can be only half the particle velocity, take a look next at the Wavepackets simulation, <http://physics.weber.edu/schroeder/software/Wavepackets.html>.

We have now implemented both the Einstein and de Broglie relations, more precisely, in terms of the sinusoidal behavior of wavefunctions:

- When p_x is well defined, $\psi \propto e^{ikx}$ with $k = p_x/\hbar$ (de Broglie).
- When E is well defined, $\psi \propto e^{-i\omega t}$ with $\omega = E/\hbar$ (Einstein).

Although I've motivated the second statement by applying it to a free particle, it turns out to be completely general, even for systems with potential as well as kinetic energy.

For a single particle in one dimension, we will write the total energy as

$$E = K + V(x), \quad (1.21)$$

where the kinetic energy K is a function of momentum, and the potential energy $V(x)$ can be practically any function of position. When $V = 0$, the energy eigenfunctions are the same as the momentum eigenfunctions (that is, e^{ikx}), as described above. But those functions describe particles whose positions—and hence their potential energies—are not well defined at all. Conversely, the spiky position eigenfunctions describe particles with well-defined potential energy but undefined momentum, and hence undefined kinetic energy. You might therefore think that when

$V \neq 0$ there would be *no* functions for which $E = K + V$ is well defined, but in fact there always is a collection of such energy eigenfunctions—they’re just different from both the position eigenfunctions and the momentum eigenfunctions. Finding the energy eigenfunctions for a given $V(x)$ is not always easy, and in any case it must be done separately for each new function $V(x)$. But energy eigenfunctions are important, so a good deal of this book will be devoted to the process of finding them.

For now, though, my main point about the energy eigenfunctions is that they always evolve in time according to the wiggle factor $e^{-i\omega t}$. So, for example, if the wavefunction at $t = 0$ is $\psi(x, 0)$ and if $\psi(x, 0)$ is an energy eigenfunction with energy E , then

$$\psi(x, t) = \psi(x, 0) e^{-iEt/\hbar} \quad (\text{for an energy eigenfunction}). \quad (1.22)$$

This is just the general version of equation 1.18, which was specific to the case of a free particle.

What if $\psi(x, 0)$ is not an energy eigenfunction? The next simplest case is when it is a mixture of just two energy eigenfunctions:

$$\psi(x, 0) = c_1\psi_1(x) + c_2\psi_2(x), \quad (1.23)$$

where ψ_1 and ψ_2 are energy eigenfunctions and c_1 and c_2 are constants. Then, as you might guess, the principle of superposition applies: Each contribution to the total wavefunction simply oscillates at its own frequency, so

$$\psi(x, t) = c_1\psi_1(x) e^{-iE_1t/\hbar} + c_2\psi_2(x) e^{-iE_2t/\hbar}, \quad (1.24)$$

where E_1 and E_2 are the energies corresponding to the eigenfunctions ψ_1 and ψ_2 .

Equation 1.24 generalizes in the natural way to mixtures of three or more energy eigenfunctions. Moreover, we’ll see later that *any* wavefunction can be expressed as a linear combination of some number of energy eigenfunctions—although that number is sometimes infinite, so we might need to deal with an infinite sum. In the language of linear algebra, we say that the collection of energy eigenfunctions (for any potential energy $V(x)$) always forms a complete basis, from which all other wavefunctions can be built.

Problem 1.14. Prove in general that if the state of a quantum system is an energy eigenfunction, then the probability density, $|\psi|^2$, is independent of time. For this reason, energy eigenfunctions are also called *stationary states*.

Problem 1.15. For the mixture of two energy eigenfunctions expressed in equation 1.24, calculate $|\psi(t)|^2$, multiplying out the binomial and writing each term in a way that explicitly shows that it is real. For simplicity you may assume that $\psi_1(x)$ and $\psi_2(x)$ are pure real. (Hint: In the cross-terms you will find it helpful to write the constants in magnitude-phase form, e.g., $c_1 = |c_1|e^{i\phi_1}$, where ϕ_1 is a real number.) Discuss the result in some detail. How would you describe the time dependence of $|\psi|^2$?

Problem 1.16. Without writing out any lengthy formulas, discuss the time dependence of a mixture of *three* energy eigenstates: $\psi(t = 0) = c_1\psi_1 + c_2\psi_2 + c_3\psi_3$. At what frequencies will the time-dependent parts of $|\psi(t)|$ oscillate? (Be sure to work the previous problem before working this one.)

Problem 1.17. In many situations the energy E of a system is ambiguous, with an arbitrary additive constant. For instance, we measure potential energies with respect to an arbitrary reference location, and we usually neglect a particle's rest energy mc^2 in nonrelativistic situations. Discuss how an additive constant affects the wiggle factor and, therefore, the time dependence of definite-energy wavefunctions in quantum mechanics. Use the result of Problem 1.15 to explain why changing the additive constant will *not* affect the probability density.

1.4 The infinite square well

In the previous section I emphasized the *free* particle, for which $V(x) = 0$, because its energy eigenfunctions are so simple: they're the same as the momentum eigenfunctions, e^{ikx} . The next-simplest example is a particle that's free within a certain limited region (say from $x = 0$ to $x = a$), but absolutely confined to that region by powerful forces on each side. In the most idealized model, this system would have zero potential energy inside the region and infinite potential energy outside:

$$V(x) = \begin{cases} 0 & \text{for } 0 < x < a, \\ \infty & \text{elsewhere.} \end{cases} \quad (1.25)$$

This potential energy function is called an infinite square well or a one-dimensional “box.” Figure 1.12 shows one way to visualize $V(x)$. Of course, any confining forces in the real world would be neither infinitely strong nor infinitely abrupt. But this model is still useful for understanding a variety of real-world potential wells (e.g., electrons in long organic molecules, or in fabricated semiconductor layers), and provides a good starting point for understanding the general properties of quantum particles that are trapped within a finite region.

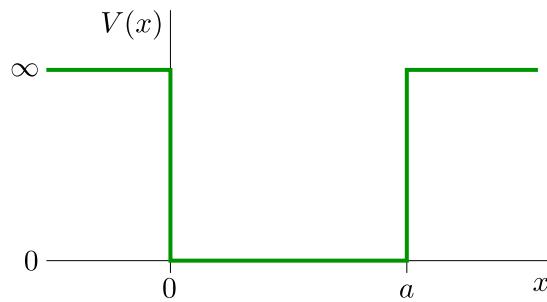


Figure 1.12: The infinite-square-well potential energy function, zero within a finite interval and infinite elsewhere.

The infinite potential energy outside the box means that there is zero probability of ever finding the particle there, so all of the allowed wavefunctions for this system are exactly zero at $x < 0$ and $x > a$. Inside the box the wavefunction can have any shape at all, so long as it is normalized. The wavefunction could be localized in a region much smaller than a , or spread over the full width of the box, and as always it can have both real and imaginary parts.

Energy eigenfunctions

The most important wavefunctions, however, are the energy eigenfunctions, whose shapes and formulas are not hard to guess. Inside the box there is no potential energy, so definite momentum would imply definite energy, just as for a free particle. The difference here is that the wavefunction must go to zero at each end, and the momentum eigenfunctions, e^{ikx} , don't go to zero anywhere. The way out of this difficulty is to realize that for a given kinetic energy there are actually *two* possible momenta, one positive and one negative. If we superimpose the two corresponding wavefunctions we get a standing wave, analogous to a standing wave on a string, with the same wavelength (and hence the same $|p_x|$ and the same kinetic energy) as the traveling waves, but with nodes, where the wavefunction goes to zero. In our case the correct superposition is proportional to

$$e^{ikx} - e^{-ikx} \propto \sin(kx), \quad (1.26)$$

which is automatically zero at $x = 0$, at the left end of the box. To make the wavefunction go to zero at the right end as well, we must restrict the values of k so that some integer number of half-wavelength “bumps” fits in the box:

$$\lambda = 2a, a, \frac{2a}{3}, \frac{2a}{4}, \dots = \frac{2a}{n}, \quad (1.27)$$

where n can be any positive integer (and is equal to the number of bumps). Figure 1.13 shows the first three wavefunctions. The corresponding k values are just $2\pi/\lambda = n\pi/a$, so the corresponding momentum magnitudes are

$$|p_x| = \hbar k = \frac{\hbar}{\lambda} = \frac{hn}{2a}. \quad (1.28)$$

The formula for the wavefunctions themselves is therefore

$$\psi_n(x) = \begin{cases} A \sin\left(\frac{n\pi x}{a}\right) & \text{for } 0 < x < a, \\ 0 & \text{elsewhere,} \end{cases} \quad (1.29)$$

where the normalization constant A equals $\sqrt{2/a}$ (as you can easily show), times an arbitrary unit-modulus phase factor $e^{i\phi}$, which we might as well take to be 1. Again, these are eigenfunctions of energy but *not* momentum; inside the box they are mixtures of positive-momentum and negative-momentum eigenstates, in equal parts.

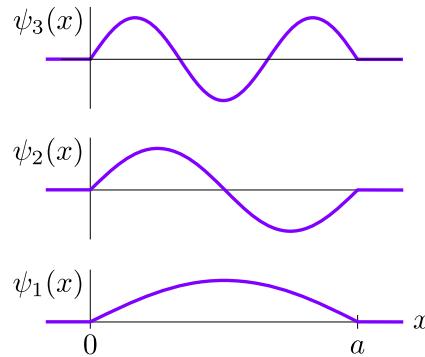


Figure 1.13: The first three definite-energy wavefunctions for a particle in an infinite square well.

Problem 1.18. Prove that the normalization constant A in equation 1.29 is $\sqrt{2/a}$.

Problem 1.19. Suppose that a particle is in the $n = 2$ energy eigenstate of an infinite square well of width a . (a) Sketch the probability density $|\psi(x)|^2$, labeling the axes in terms of a . (b) Suppose you now measure the position of this particle. Estimate the probability (to the nearest half percent or less) of finding it within each of the following three intervals: $0.25a$ to $0.26a$; $0.50a$ to $0.51a$; and $0.62a$ to $0.63a$.

Quantized energies

To find the energy values that correspond to these eigenfunctions, we need to know the formula for kinetic energy in terms of momentum. If the particle is nonrelativistic, we have simply

$$E = K = \frac{p_x^2}{2m} = \frac{\hbar^2 n^2}{8ma^2}, \quad (1.30)$$

where the first equals sign holds only inside the box, but the result must be valid for the wavefunction as a whole. Thus, we have a series of *discrete* allowed energy values, with intermediate values not allowed. We say that the energy is *quantized*; this is where the “quantum” in quantum mechanics comes from.

Figure 1.14 shows an energy level diagram (that is, a plot of energy vs. nothing) for this system. Such a diagram is sometimes called a *quantum ladder*, and in this case the rungs get farther and farther apart as you go up. The lowest energy level is called the *ground state*, while the rest are called *excited states*. Figure 1.15 combines the energy level diagram with plots of the potential energy and the definite-energy wavefunctions.

Let’s pause briefly to think about the *reason* why the energy is quantized: In order to fit within the box, a sinusoidal wavefunction must have an integer number of bumps. Partial bumps are not allowed, because the wavefunction must go to zero at both ends of the box (and a discontinuity at either end of the wavefunction would be equivalent to an infinitely short wavelength at that point, and hence an

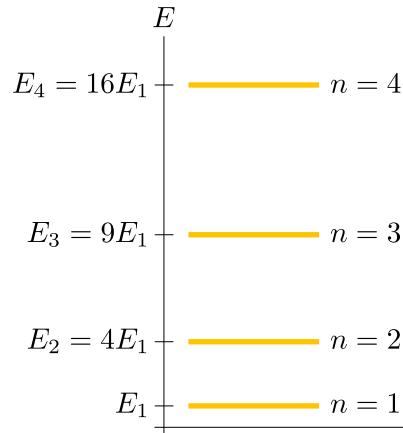


Figure 1.14: Energy level diagram or “quantum ladder” for the one-dimensional infinite square well. The energies are in units of $E_1 = h^2/8ma^2$.

infinitely high energy—subverting our goal of finding wavefunctions for which the energy is well defined and finite).

Notice also that the energy of the ground state is nonzero; in fact it is $h^2/8ma^2$. This is the lowest possible energy for a (nonrelativistic) particle trapped inside an infinite square well of width a . It is nonzero because the wavefunction must have at least one full bump inside the box, and therefore the longest possible wavelength is $2a$. If you squeeze the particle into a smaller box (that is, reduce the value of a), the maximum wavelength decreases and therefore the ground-state energy increases. This inverse relation between size and energy is quite general in quantum mechanics, and is ultimately responsible for the stability of matter: If there were no energy cost to squeezing electrons into smaller and smaller spaces, atoms would collapse under the attractive electrostatic forces between electrons and nuclei.

I wish I could now show you an *experimental* energy level diagram for a real particle trapped in a real potential that approximates an infinite square well. Perhaps there’s a good example out there somewhere, but so far I’ve been unable to find it. As I mentioned above, the one-dimensional infinite square well can be used to model electrons in certain organic molecules, or in semiconductor layers. If you shine light on these systems, the wavelengths absorbed will correspond to electron transitions from one rung on the quantum ladder up to a higher rung. Typically, however, the data show only a single transition—not enough to decode the full structure of the energy level diagram or even to show that the levels get farther apart as you go up. Many of the experiments verify the dependence on the well width a in equation 1.30, but even extracting this information involves quite a few real-world complications that I’d rather not spend time on now.

So please be patient. The real reason to study the one-dimensional infinite square well is not because of its direct, real-world applications, but because it serves as a foundation and starting point for understanding other quantum systems.

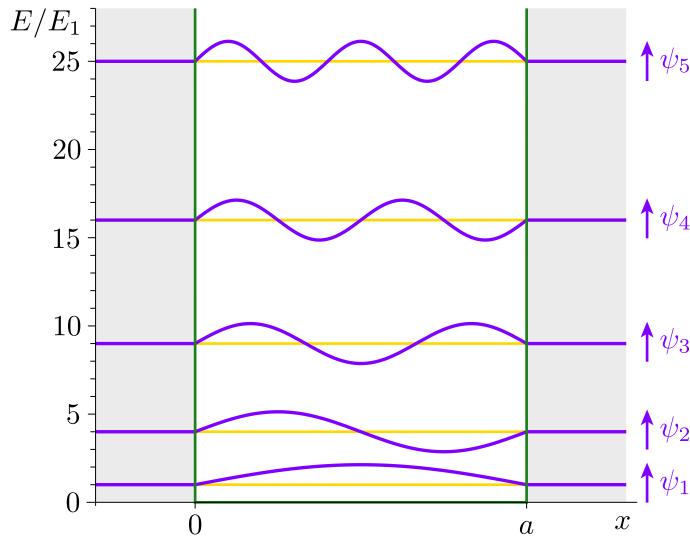


Figure 1.15: Combined plot of the lowest five energy levels and definite-energy wavefunctions for the infinite square well. The energies are in units of the ground-state energy E_1 . This picture is confusing because each wavefunction is plotted on a separate horizontal axis, using a vertical scale that is unrelated to the energy scale (since wavefunction values are not measured in energy units!). Nevertheless, plots of this type concisely show a lot of information at once, so I will use them from time to time.

Problem 1.20. Calculate the ground-state energy for an electron trapped in a one-dimensional box of width (a) 1 cm; (b) 0.1 nm; (c) 1 fm. Express your answers in electron-volts, and compare them to other energies that might be relevant to systems of these sizes. (Hint: When is the electron nonrelativistic? When is it relativistic? In the relativistic case you'll need to use the relativistic energy-momentum relation instead of equation 1.30.)

Time dependence

If a particle in a box starts out in one of the energy eigenfunctions $\psi_n(x)$, what happens as time passes? I told you the answer—for energy eigenfunctions in general—in Section 1.3: Just multiply ψ_n by the usual time-dependent wiggle factor to obtain

$$\psi(x, t) = \psi_n(x) e^{-iE_n t/\hbar}. \quad (1.31)$$

In other words, the magnitude of the wavefunction remains unchanged at every x , but the phase oscillates clockwise in the complex plane, at a rate proportional to the energy E_n . Because the probability density does not change with time, we often refer to energy eigenfunctions as *stationary states* (as mentioned in Problem 1.14).

Things get more interesting if we start out in a mixture of two or more stationary states. Again, the general prescription was given in Section 1.3: If $\psi(t = 0)$ is a sum of stationary-state terms, then $\psi(t)$ is given by the same formula with the corresponding wiggle factor, $e^{-iE_n t/\hbar}$, multiplying each term (see equation 1.24).

That is, each term oscillates sinusoidally at a frequency proportional to its energy. The Quantum Square Well simulation, posted at <http://physics.weber.edu/schroeder/software/SquareWell.html>, animates the time dependence of an arbitrary mixture of the eight lowest-energy eigenfunctions. Please spend some time playing with that simulation now.

But what if a system starts in a wavefunction that *isn't* a mixture of energy eigenstates? Well, it turns out that no such wavefunctions exist! To see how this works (and why), we need to investigate some of the mathematical properties of the energy eigenfunctions.

Problem 1.21. A nonrelativistic particle in an infinite square well that starts out in the mixture

$$\psi(t=0) = \frac{1}{\sqrt{2}}(\psi_1(x) + \psi_2(x)), \quad (1.32)$$

where ψ_1 and ψ_2 are the usual energy eigenfunctions for this system. Write down an expression for the wavefunction $\psi(t)$ at future times, and use Mathematica to plot $\psi(t)$ at several different times—enough to show the pattern. As in Figures 1.9 and 1.10, plot $|\psi|^2$ vertically and color the space underneath according to the phase of ψ .⁹ Be careful to use the same vertical scale in every plot. Use units in which $a = 1$, and choose a similarly convenient unit of time (perhaps so that $E_1/\hbar = 1$). Describe the time evolution of the probability density in words, with reference to the more general behavior that you found in Problem 1.15. Also comment on how the sequence of phases (color hues) relates to what the wavefunction is doing.

Completeness and orthogonality

The claim that I just made can be stated as a mathematical theorem: *Any* function $\psi(x)$, within the interval from 0 through a , can be written as a linear combination of the energy eigenfunctions, for some set of coefficients c_n :

$$\psi(x) = \sum_{n=1}^{\infty} c_n \psi_n(x) = \sqrt{\frac{2}{a}} \sum_{n=1}^{\infty} c_n \sin\left(\frac{n\pi x}{a}\right). \quad (1.33)$$

This theorem is sometimes called *Dirichlet's theorem*, and you may recognize this formula as the *Fourier series* for $\psi(x)$. I won't try to prove the theorem, but you'll probably be convinced of it if you spend twenty minutes playing with the Wave Builder web app at <http://physics.weber.edu/schroeder/software/WaveBuilder.html>. Notice that in general the sum requires an infinite number of terms, although it's common in physics to stop after a finite number of terms as an approximation (since we never know the wavefunction *exactly* anyway). If the function $\psi(x)$ is complex, then so are at least some of the constants c_n . The formal name for this property of the ψ_n functions is *completeness*: they form a *complete set*, also called a *basis*, from which any other wavefunction can be built as a linear combination.

⁹See Appendix B.3 for how to make these kinds of plots in Mathematica.

You might wonder, though, how to *find* all the c_n values for a given $\psi(x)$. Before I tell you how, I need to mention another mathematical property of the energy eigenfunctions: they are *orthogonal*, in the sense that

$$\int_0^a \psi_m(x) \psi_n(x) dx = 0 \quad \text{for } m \neq n. \quad (1.34)$$

I like to think of this integral as a kind of infinite-dimensional dot-product, where we multiply the “components” of ψ_m and ψ_n together at each x , then add them all up. In fact, we define the *inner product* of any two one-dimensional functions $\alpha(x)$ and $\beta(x)$ as

$$\langle \alpha | \beta \rangle = \int_{-\infty}^{\infty} \alpha^*(x) \beta(x) dx, \quad (1.35)$$

where the complex conjugation is needed to ensure that any normalized function’s inner product with itself equals 1 (but is unnecessary in the present context where all of the our ψ_n functions are real). Problems 1.25 through 1.27 demonstrate that this inner product is zero for any two distinct sine-wave energy eigenfunctions of the infinite square well.

We usually combine equation 1.34 with the normalization condition to write

$$\int_0^a \psi_m(x) \psi_n(x) dx = \delta_{mn}, \quad (1.36)$$

where δ_{mn} is an abbreviation called the *Kronecker delta symbol*, defined as

$$\delta_{mn} = \begin{cases} 1 & \text{if } m = n, \\ 0 & \text{if } m \neq n. \end{cases} \quad (1.37)$$

The verbal way of expressing equation 1.36 is then to say that the energy eigenfunctions are *orthonormal*. This property is not unique to the infinite square well; it actually applies to any quantum system whose energy levels are discrete.

I’m now ready to tell you how to find the coefficients c_n in equation 1.33. Textbook author David Griffiths¹⁰ calls the method by the apt name *Fourier’s trick*: Multiply equation 1.33 by an arbitrary $\psi_m(x)$, and integrate from 0 to a :

$$\int_0^a \psi_m(x) \psi(x) dx = \sum_{n=1}^{\infty} c_n \int_0^a \psi_m(x) \psi_n(x) dx = \sum_{n=1}^{\infty} c_n \delta_{mn} = c_m. \quad (1.38)$$

In mathematical language, we’re taking the inner product of our function $\psi(x)$ with one of the “basis” functions $\psi_m(x)$, to project out the “component” of $\psi(x)$ along this basis function. After moving the integral inside the sum, we obtain the same integral as in equation 1.36, and then the Kronecker delta kills every term in the sum except one, leaving us with a formula for c_m . The formula for c_m involves

¹⁰David J. Griffiths and Darrell F. Schroeter, *Introduction to Quantum Mechanics*, third edition (Cambridge University Press, 2018).

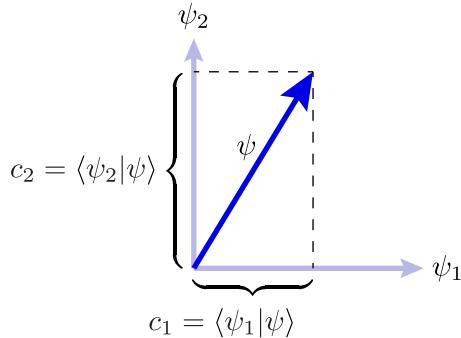


Figure 1.16: In an abstract two-dimensional space, an arbitrary vector ψ can be expanded in terms of two orthonormal basis vectors ψ_1 and ψ_2 . The components of ψ in this basis are $\langle \psi_1 | \psi \rangle$ and $\langle \psi_2 | \psi \rangle$, respectively.

an integral, and you might feel daunted by the prospect of having to carry out an infinite number of such integrals to find the whole list of c_m values. But even if this process would be tedious, at least it's straightforward—and often we can get computers to do the tedious work for us.

The term “inner product” is supposed to remind you of an ordinary dot product of vectors. In fact, everything we’re doing here is analogous to the more familiar example of expanding an arbitrary vector in terms of an orthonormal set of unit-length basis vectors. Figure 1.16 shows how to visualize the process for ordinary vectors in two dimensions. In the analysis above, the energy eigenfunctions ψ_n are our basis “vectors,” which must be orthogonal to each other and normalized to unit length (equation 1.36). Because these basis functions form a complete set, any other wavefunction ψ can be expanded in terms of its components c_n in this basis. And to find any particular component c_n , you just “dot” the corresponding basis function ψ_n into the function ψ .

Of course my little two-dimensional picture leaves out two important complications: (1) we’re actually working in an *infinite*-dimensional vector space,¹¹ with an infinite number of mutually orthogonal basis functions ψ_n ; and (2) our components c_n can be *complex* numbers. I can’t really visualize either of these complications, but I still find my little two-dimensional picture to be far better than no picture at all.

Problem 1.22. After playing with the Wave Builder web app (<http://physics.weber.edu/schroeder/software/WaveBuilder.html>) for a while, set it on “Challenge round 1” and match all four of the target functions. Try not to use too many hints. To keep a record of your success, make a screen capture when you see the message “Target 4 of 4 matched!”

Problem 1.23. Repeat the previous problem for “Challenge round 2.”

Problem 1.24. Repeat Problem 1.22 for “Challenge round 3” (which is much harder than rounds 1 and 2).

¹¹For a systematic overview of the properties of vector spaces, see Appendix A.4.

Problem 1.25. Sketch the first two energy eigenfunctions for the infinite square well, then use your sketch to explain (without any calculation) why these functions must be orthogonal to each other. Can you generalize your argument to other pairs of energy eigenfunctions? Explain.

Problem 1.26. Carry out the integral to prove that the infinite square well energy eigenfunctions $\psi_1(x)$ and $\psi_3(x)$ are orthogonal. Do the integral “by hand,” rewriting the integrand using a product-to-sum trig identity. Also use a computer to plot the integrand. Does it *look* like the total area under the graph is zero? Explain.

Problem 1.27. Prove in general that the infinite square well eigenfunctions $\psi_m(x)$ and $\psi_n(x)$ are orthogonal for $m \neq n$. Do the integral “by hand,” using a product-to-sum trig identity. Why does the proof break down when $m = n$?

Problem 1.28. Prove the following properties of the inner product (defined in equation 1.35), where the α s and β s are functions and c is any complex number:

- (a) $\langle \beta | \alpha \rangle = \langle \alpha | \beta \rangle^*$,
- (b) $\langle \alpha | \beta_1 + \beta_2 \rangle = \langle \alpha | \beta_1 \rangle + \langle \alpha | \beta_2 \rangle$,
- (c) $\langle \alpha_1 + \alpha_2 | \beta \rangle = \langle \alpha_1 | \beta \rangle + \langle \alpha_2 | \beta \rangle$,
- (d) $\langle \alpha | c\beta \rangle = c\langle \alpha | \beta \rangle$,
- (e) $\langle c\alpha | \beta \rangle = c^*\langle \alpha | \beta \rangle$.

Aside from the notation and the complex conjugations in identities (a) and (e), these properties should be familiar to you from ordinary dot products of real vectors in three-dimensional space..

An example initial state

Now let’s work through an example that applies the ideas of the previous two subsections.

Suppose that a nonrelativistic particle, trapped in an infinite square well, initially has the rather localized wavefunction shown in Figure 1.17. I’ve arbitrarily centered this wavefunction at $a/4$, and arbitrarily made it several times narrower than the well. For simplicity I’ve made it purely real. *Question:* How will this wavefunction change as time passes?

The formula I used to plot Figure 1.17 is

$$\psi(x) = A \exp \left[-\left(\frac{x - 0.25}{0.05} \right)^2 \right], \quad (1.39)$$

where all distances are measured as multiples of the well width a (so in these units, $a = 1$). This function is what we call a *Gaussian bell curve*: basically e^{-x^2} , shifted so the center is at $x = 0.25$ and squeezed so the half-width, from the center to where the value is smaller by a factor of e , is 0.05. I chose the Gaussian shape because it’s smooth and cuts off rapidly enough to be essentially zero at either edge of the box. You can check (Problem 1.29) that the normalization constant A is slightly less than 4.0.

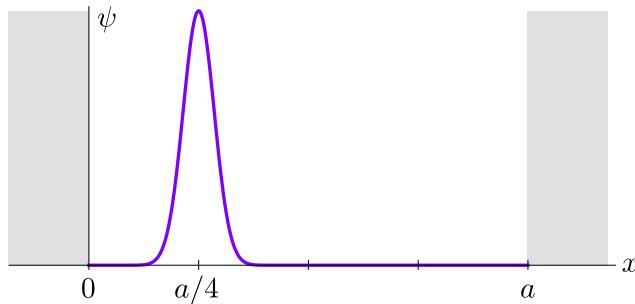


Figure 1.17: A Gaussian initial state for a particle in an infinite square well.

By the completeness property, this wavefunction must be some mixture of the energy eigenfunctions:

$$\psi(x) = \sum_{n=1}^{\infty} c_n \psi_n(x), \quad (1.40)$$

for some set of coefficients c_n , where $\psi_n(x)$ is the n -bump sine wave (equation 1.29). To find the n th coefficient, we “dot” our wavefunction into the n th basis function, as written in equation 1.38. Here is some Mathematica code that defines the initial $\psi(x)$ function, defines the $\psi_n(x)$ basis functions, and then computes the first 30 expansion coefficients:

```
initialPsi[x_] := 3.99471*Exp[-((x-0.25)/0.05)^2];
basis[n_, x_] := Sqrt[2]*Sin[n*Pi*x];
c = Table[
  NIntegrate[initialPsi[x]*basis[n, x], {x, 0, 1}],
  {n, 1, 30}]
```

Please, if at all possible, fire up your Mathematica system and run this code now. You should find that it produces¹² the following list:

```
{ 0.351845, 0.48846, 0.334903, 2.3348 × 10-14, -0.303429, -0.400962,
  -0.261675, 4.5866 × 10-14, 0.214801, 0.270178, 0.167833, 6.7859 × 10-14,
  -0.124821, -0.149442, -0.088362, 8.8199 × 10-14, 0.059541, 0.067853,
  0.038188, 1.0712 × 10-13, -0.023314, -0.025289, -0.013548, 1.2419 × 10-13,
  0.007494, 0.007737, 0.003945, 1.3905 × 10-13, -0.001977, -0.001943 }
```

These coefficients are the *components* of $\psi(x)$ in the “directions” of the first 30 sine-wave basis functions. To get a better feel for them, we can plot them with the instruction `ListPlot[c, Filling->Axis]`, as shown in Figure 1.18. The values oscillate as a function of n , but the more important pattern is that they generally

¹²For brevity I’ve rounded the numbers off a little more than Mathematica does by default. Mathematica also displays some warning messages about the near-zero results possibly being inaccurate. You can just ignore these messages, or eliminate them with an option like `AccuracyGoal->8` in the `NIntegrate` function.

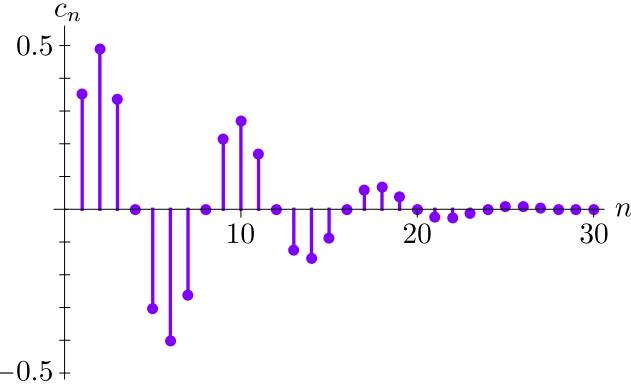


Figure 1.18: Plot of the first 30 expansion coefficients of the wavefunction shown in Figure 1.17, in the infinite square well energy eigenfunction basis.

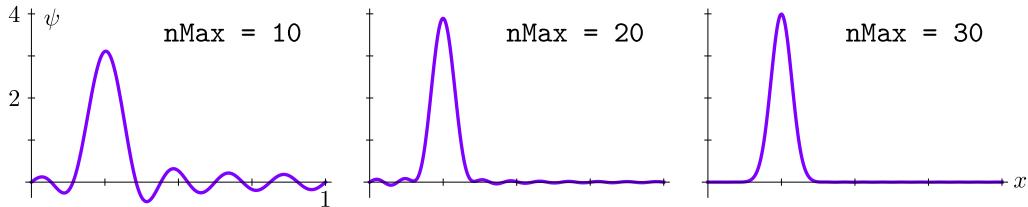


Figure 1.19: Plots of the reassembled $\psi(x)$, computed from equation 1.40, keeping the first 10, 20, and 30 terms.

decrease as n increases, and are pretty negligible once n reaches 30. (I chose to stop at 30 after some trial and error with other upper limits.) As a further check that the expansion has converged, we can reassemble $\psi(x)$ from its components (using equation 1.40 and plot the result:

```
Plot[Sum[c[[n]]*basis[n, x], {n, 1, nMax}], {x, 0, 1}]
```

Figure 1.19 shows the output of this instruction for n_{Max} values of 10, 20, and 30.

Now that we have our expansion coefficients, we can obtain the wavefunction at any future time by inserting the appropriate wiggle factor into each term of the expansion (see equation 1.24):

$$\psi(x, t) = \sum_n c_n \psi_n(x) e^{-i E_n t / \hbar}. \quad (1.41)$$

In this case $E_n = \hbar^2 n^2 / (8ma^2)$, so the wiggle factor frequencies are proportional to n^2 . To simplify this formula we can just absorb all the constants into our unit of time, so the wiggle factor becomes $e^{-in^2 t}$ (see Problem 1.29 for details). Then the Mathematica code to define a function for $\psi(x, t)$ is:

```
psi[x_, t_] := Sum[c[[n]]*basis[n, x]*Exp[-I n^2 t], {n, 1, 30}]
```

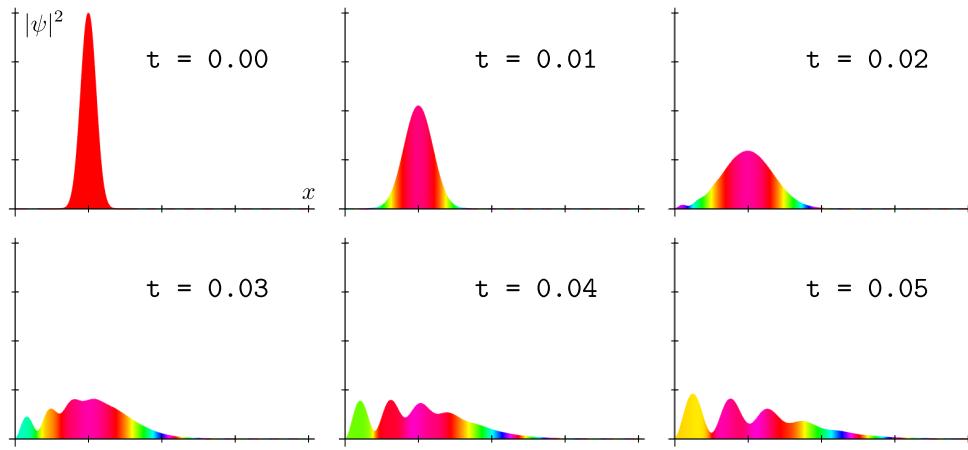


Figure 1.20: Time evolution of the initial wavefunction shown in Figure 1.17, with times measured in units of \hbar/E_1 .

What does this function look like? For $t > 0$ it's complex, so let's plot $|\psi|^2$ on the vertical axis and use color hues to represent the phases (as in Figures 1.9 and 1.10, using Mathematica as described in Appendix B.3):

```
Plot[Abs[psi[x, t]]^2, {x, 0, 1}, PlotRange -> {0, 16},
    PlotPoints -> 300, Filling -> Axis,
    ColorFunction -> Function[x, Hue[Arg[psi[x, t]]/(2Pi)]],
    ColorFunctionScaling -> False]
```

Figure 1.20 shows a sequence of plots produced in this way for t values up to 0.05. The initially narrow Gaussian spreads horizontally, as its tails acquire a positive imaginary part (orange to yellow to green) and its center acquires a negative imaginary part (magenta). Soon the left tail hits the edge of the box, and the reflections produce interference fringes in the probability distribution. Problem 1.29 further explores the time evolution of this wavefunction. The general point, though, is that once we know the initial wavefunction $\psi(x)$ and the energy eigenfunctions $\psi_n(x)$, we can use wiggle factors and superposition to predict the wavefunction at any future time.

Problem 1.29. As you use Mathematica to work through the calculations and reproduce the plots in this subsection, fill in the following details:

- Find the normalization constant A in equation 1.39.
- Explain why $c_n \approx 0$ whenever n is a multiple of 4.
- Suppose that the quantum particle is an electron and $a = 1.0$ nm. Calculate E_1 in electron-volts, then use this value to find the unit of time used in Figure 1.20 (and in the corresponding code). Express this time unit in femtoseconds. About how long does the wavefunction take to spread out enough to fill the “box”?
- Explore the time evolution of this wavefunction over longer time scales. What patterns do you notice? Can you explain them?

Problem 1.30. A nonrelativistic particle in an infinite square well has the initial ($t = 0$) wavefunction

$$\psi(x) = A(x - 0.5) \exp\left[-\left(\frac{x - 0.5}{0.1}\right)^2\right],$$

in units where the well width a equals 1.

- (a) Plot this function.
- (b) Find the normalization constant A .
- (c) Find the expansion coefficients c_n in the basis of energy eigenfunctions, and plot these coefficients as in Figure 1.18. Explain why some of the c_n are zero (although `NIntegrate` returns very small nonzero values).
- (d) Plot the approximate reassembled wavefunction as in Figure 1.19, keeping a few different numbers of terms in the sum and showing how many are needed for the plot to visually match the original $\psi(x)$. How many terms *are* needed to visually match the original $\psi(x)$?
- (e) Calculate and plot the wavefunction at several future times (using color hues to represent the phases as in Figure 1.20), to illustrate how it evolves. Discuss the appearance of these plots in some detail, explaining as many features as you can. How much time passes before the behavior repeats?

Problem 1.31. The method used in this subsection is similar to the way we predict the motion of a taut string with fixed ends. Consider, for instance, a guitar string stretched between $x = 0$ and $x = 1$ (in suitable units), whose transverse displacement is y . At $t = 0$ you pluck the string so it has the initial triangular shape

$$y(t=0) = \begin{cases} x & \text{for } 0 < x < 0.2, \\ 0.25(1-x) & \text{for } 0.2 < x < 1, \end{cases}$$

where I'm assuming that y is measured in much smaller units than x .

- (a) Use Mathematica to plot this function. Then calculate its first 50 expansion coefficients in the basis of normalized sine waves, analogous to the expansion in equation 1.40. Make plots analogous to Figures 1.18 and 1.19, and discuss them briefly.
- (b) Instead of oscillating with a complex “wiggle factor” $e^{-i\omega_n t}$, a sinusoidal normal mode of a taut string that is initially at rest oscillates according to a cosine function: $\cos(\omega_n t)$. The other difference from the nonrelativistic quantum system is that instead of ω_n being a quadratic function of n , for the string it is a linear function (with a constant of proportionality that depends on the string's mass and tension). Otherwise, the procedure for predicting $y(x, t)$ is the same as for a quantum particle in a box: Just insert the oscillating time-dependent factor into each term in the expansion. So define a Mathematica function to calculate $y(x, t)$, and plot this function for enough different times to show the pattern in the behavior.

Probabilities for energy measurements

If a particle in an infinite square well starts out in an energy eigenfunction and then you measure its position, you know how to calculate the probabilities of the various possible outcomes: according to the Born rule (equation 1.6), you square the

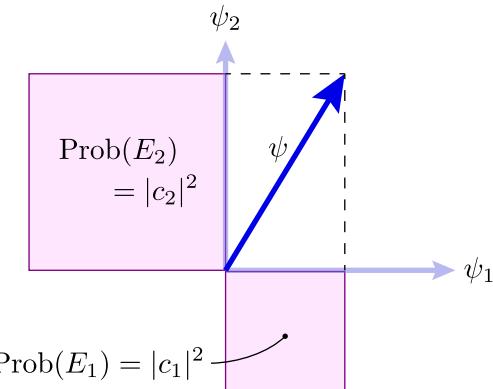


Figure 1.21: When we measure the energy of a system that's not in an energy eigenstate, the probability of obtaining any particular outcome E_n is the squared modulus of the component c_n of the system's state ψ in the direction of the eigenvector ψ_n that corresponds to E_n . The normalization condition on ψ and the Pythagorean theorem guarantee that the sum of all the energy outcome probabilities equals 1.

wavefunction to get the position probability density. For an energy eigenfunction this probability density is spread over the width of the well, and varies sinusoidally between 0 at the nodes and $2/a$ at the antinodes.

But what if, on the other hand, the particle starts out in some arbitrary wavefunction and you try to measure its *energy*? If the wavefunction is an energy eigenfunction ψ_n , then of course you will obtain the corresponding E_n value with 100% certainty. If the wavefunction is a mixture of two energy eigenfunctions, such as

$$\psi(x) = c_1\psi_1(x) + c_2\psi_2(x), \quad (1.42)$$

then presumably you could obtain either E_1 or E_2 as your result, with probabilities that somehow depend on the sizes of the coefficients c_1 and c_2 . But the probabilities can't be *equal* to c_1 and c_2 , because these coefficients could be negative or even complex; in fact the probabilities are equal to the *square moduli* of these coefficients, $|c_1|^2$ and $|c_2|^2$, respectively. To see why this rule makes sense, let's check the normalization of $\psi(x)$:

$$1 = \int_0^a |\psi(x)|^2 dx = \int_0^a [|c_1|^2\psi_1^2 + |c_2|^2\psi_2^2 + (c_1^*c_2 + c_2^*c_1)\psi_1\psi_2] dx. \quad (1.43)$$

Integrating the three terms separately, we find that the first two give simply $|c_1|^2$ and $|c_2|^2$, because ψ_1 and ψ_2 themselves are normalized, while the third term integrates to zero, because ψ_1 and ψ_2 are orthogonal. Therefore the squared coefficients must obey the relation $|c_1|^2 + |c_2|^2 = 1$, just as you would expect for probabilities (see Figure 1.21).

The formulas in the previous paragraph generalize straightforwardly to superpositions of three or more energy eigenfunctions. If you measure the energy for a

particle with *any* (normalized) wavefunction $\psi(x)$, the probability of obtaining the result E_n is the square modulus of the expansion coefficient c_n , when the wavefunction is expanded in terms of energy eigenfunctions as in equation 1.33. Therefore, according to the result of Fourier’s trick (equation 1.38), the probability is

$$\left(\begin{array}{l} \text{Probability of finding} \\ \text{particle with energy } E_n \end{array} \right) = |c_n|^2 = \left| \int_0^a \psi_n(x) \psi(x) dx \right|^2. \quad (1.44)$$

Moreover, normalization of $\psi(x)$ implies that the sum of all these probabilities equals 1:

$$\sum_{n=1}^{\infty} |c_n|^2 = 1. \quad (1.45)$$

Equation 1.44 is another version of the Born rule; to distinguish it from equation 1.6, I’ll call it the *Born rule for energy*. (Unlike the Born rule for position, equation 1.6, there’s no need here to integrate over a range of outcomes because the allowed energy values are discrete.)

Finally, I should mention that *after* your measurement, now that you have a definite energy E_n for the particle, the wavefunction will have “collapsed” to $\psi_n(x)$.

Problem 1.32. Suppose that at $t = 0$ you were to measure the energy of the particle described in the previous subsection, whose wavefunction is written in equation 1.39 and plotted in Figure 1.17.

- (a) What is the probability of finding this particle to have energy E_1 ?
- (b) What is the most likely outcome of this measurement, as a multiple of E_1 ? What is its probability?
- (c) Check that the total probability for all outcomes up to E_{30} is approximately equal to 1.
- (d) Suppose now that you wait for some time to pass before making the energy measurement. How do the energy measurement probabilities change with the passage of time? Explain carefully.

Problem 1.33. For the initial wavefunction described in Problem 1.30, list all the energies that you might obtain with probability 1% or greater, if you were to measure this particle’s energy. Express these energies as multiples of E_1 , the ground-state energy of a particle in this infinite well. Also list their respective probabilities. Finally, calculate the (approximate) average expected outcome of this energy measurement, being sure to make your reasoning clear.

1.5 The Schrödinger equations

The energy eigenfunctions for the infinite square well are simple and easy to guess—but they’re the exception, not the rule. For a more general (and more realistic) potential energy function $V(x)$ the energy eigenfunctions will be more complicated, and simply guessing them isn’t an option. We need a more systematic method. Unfortunately, there’s no general formula for the energy eigenfunctions themselves. But there *is* a general way of writing down the differential equation of which these functions are the solutions.

The time-independent Schrödinger equation

To see where this differential equation comes from, consider again the sinusoidal waves that are energy eigenfunctions for a free particle, or inside an infinite square well, or (it turns out) in any region where $V(x) = 0$. Whether the sinusoidal function is e^{ikx} or $\sin(kx)$ or $\cos(kx)$, its second derivative equals $-k^2$ times itself:

$$\frac{d^2}{dx^2}\psi(x) = -k^2\psi(x) \quad \text{for energy eigenfunctions where } V(x) = 0. \quad (1.46)$$

Here the wavenumber k is, as usual, p/\hbar , and we don't have to worry about whether p is positive or negative because it's squared.

From here on, unless I say otherwise, I will assume that the particle in question is nonrelativistic. Then $p^2 = 2mK$, where K is the kinetic energy, so

$$\frac{d^2}{dx^2}\psi(x) = -\frac{2mK}{\hbar^2}\psi(x). \quad (1.47)$$

Of course, for a free particle, we could just as well write E (total energy) instead of K (kinetic energy). But now think about what should happen where $V(x) \neq 0$. The second derivative of the wavefunction, or its curvature, is what determines its wavelength: the distance that it takes the wave to swing up and down to make a full cycle. The wavelength determines the momentum, which determines the *kinetic* energy, so adding in some potential energy, while keeping K fixed, shouldn't affect the second derivative or the relationship expressed in equation 1.47. Therefore, even where $V(x)$ is nonzero, we can substitute $E - V(x)$ for K in equation 1.47 to obtain

$$\frac{d^2}{dx^2}\psi(x) = -\frac{2m}{\hbar^2}(E - V(x))\psi(x) \quad \text{for any energy eigenfunction.} \quad (1.48)$$

Comparing to equation 1.46, we see that the quantity $\sqrt{2m(E - V(x))/\hbar}$ in this equation is playing the role of a position-dependent wavenumber, $k(x)$. Dividing this wavenumber into 2π gives us a corresponding *local wavelength*,

$$\lambda(x) = \frac{2\pi}{k(x)} = \frac{h}{\sqrt{2m(E - V(x))}}. \quad (1.49)$$

But the denominator of this expression is just the momentum that a classical particle of energy E would have if it were at position x ; we sometimes refer to this quantity as the *local momentum*. Figure 1.22 shows an example of the concept of local wavelength.

Equation 1.48 is the famous *time-independent Schrödinger equation*, or TISE for short. It says that the local wavelength and local momentum of an energy eigenfunction must satisfy the de Broglie relation at every x . In practical terms, the TISE is the differential equation that we must solve in order to find the energy eigenfunctions for an arbitrary potential energy $V(x)$. It's really a *different* differential equation for every new $V(x)$, and its solutions are different as well, so solving

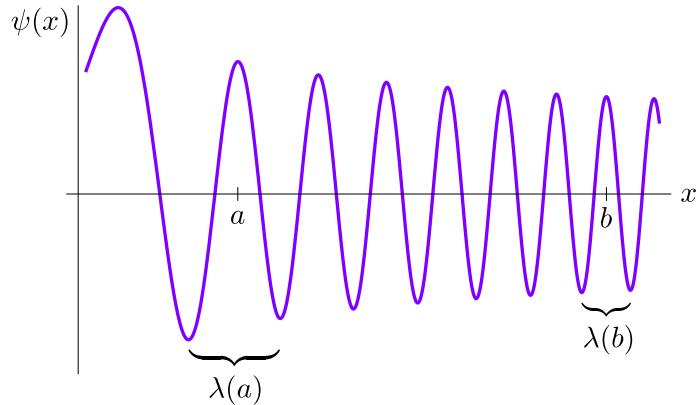


Figure 1.22: A wavefunction whose local wavelength decreases from left to right. Although it's easy in this case to estimate the local wavelength by measuring a full cycle of the wave, the precise definition of $\lambda(x)$ requires only a local comparison of ψ to its second derivative, to determine the local value of k^2 (see equation 1.46) and hence $\lambda = 2\pi/k$.

the TISE for the myriad possible $V(x)$ functions can be a never-ending task. We'll solve it repeatedly, in particular cases, in the chapters that follow. First, though, I'd like to digress into a more abstract interpretation of the TISE.

Problem 1.34. Assuming that the wavefunction shown in Figure 1.22 is an energy eigenfunction for some potential energy $V(x)$, does $V(x)$ increase or decrease from left to right? Explain carefully.

Problem 1.35. Consider a wavefunction $\psi(x)$ whose behavior in a small region around $x = 0$ is described by the function $\psi = 1 - x^2$ (where x is in some suitable distance unit). What is this function's local wavelength $\lambda(x)$ at $x = 0$? Illustrate your result with a sketch or a computer-drawn plot showing both $\psi(x)$ and the cosine function whose actual wavelength is $\lambda(0)$ so that it "fits" $\psi(x)$ at the origin.

Problem 1.36. Suppose $V(x)$ is a constant (independent of x) throughout some interval. Find the general solution to the TISE in this interval, in terms of E , treating the cases $E > V$ and $E < V$ separately. (In each case there will be a two-parameter family of solutions, because the TISE is a second-order differential equation.)

Problem 1.37. Usually, when we work with the time-independent Schrödinger equation (TISE), we assume a given $V(x)$ and try to solve for the eigenfunctions $\psi(x)$. But it's much easier to work backwards. Suppose, for example, that the Gaussian function $e^{-\alpha x^2}$ (where α is a constant) is a solution of the TISE for some energy E . What can you conclude about the potential energy function $V(x)$? If we assume $V = 0$ at $x = 0$, what are $V(x)$ and E ?

Operators and eigenvalues

With a simple algebraic rearrangement, we can instead write the TISE as

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x) + V(x)\psi(x) = E\psi(x). \quad (1.50)$$

I'll now “factor” out the $\psi(x)$ on the left-hand side to obtain

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] \psi(x) = E\psi(x), \quad (1.51)$$

where the quantity in brackets is what we call an *operator*, which acts on the function $\psi(x)$. In general, an operator is basically a “machine” that converts a function into some other function. Examples of operators would include d/dx (“take the derivative”), x (“multiply the function by x ”), and $-\hbar^2/2m$ (“multiply the function by the constant $-\hbar^2/2m$ ”). The operator in the brackets of equation 1.51 says, “Take the function’s second derivative, multiply the result by the constant $-\hbar^2/2m$, then add that onto what you get when you multiply the function by $V(x)$.” This multi-step operator is called the *Hamiltonian* operator and is abbreviated \hat{H} :

$$\hat{H} = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x). \quad (1.52)$$

(The “hat” symbol $\hat{}$ is to emphasize that it’s an operator.) Using this handy abbreviation, we can write the TISE as simply

$$\hat{H}\psi = E\psi, \quad (1.53)$$

where here it is understood that ψ is a function of x .

Equation 1.53 has the form of an *eigenvalue* equation: It says that the solutions to the TISE are precisely those functions that, when acted upon by the Hamiltonian operator, are left unchanged except for multiplication by a constant E (the so-called *eigenvalue*). This is why those solutions are called energy *eigenfunctions*.

You’ve probably heard of eigenvalues before, in the context of matrices and *eigenvectors*. In the next chapter I’ll show you how to rewrite equation 1.53 in vector-matrix form—although the matrix that represents \hat{H} and the vectors that represent ψ turn out to be infinite-dimensional. For now, just rest assured that the “eigen” terminology is entirely consistent with what you learned in linear algebra.

To provide more context for these ideas, let me digress even further to discuss some other operators and their eigenfunctions. The *momentum operator* is $\hat{p}_x = -i\hbar d/dx$, because the momentum eigenfunctions, $e^{ip_xx/\hbar}$, are eigenfunctions of this operator with eigenvalues p_x :

$$\left[-i\hbar \frac{d}{dx} \right] \psi(x) = p_x \psi(x), \quad \text{for } \psi(x) = e^{ip_xx/\hbar}. \quad (1.54)$$

From the momentum operator we can construct the *kinetic energy operator* using the formula $\hat{p}_x^2/2m$; that is, apply the momentum operator twice, then divide by the constant $2m$. The result is just the first term in the Hamiltonian operator, equation 1.52.

The *position operator*, meanwhile, is simply $\hat{x} = x$, that is, “multiply by a factor of x .” The eigenfunctions of this operator are Dirac delta functions, because the eigenvalue equation

$$\hat{x}\psi(x) = x_0\psi(x) \quad (1.55)$$

Physical quantity	Operator	Eigenvalue symbol	Corresponding eigenfunction
Position	$\hat{x} = x$	x_0	$\delta(x - x_0)$
Momentum	$\hat{p}_x = -i\hbar \frac{d}{dx}$	p_x	$e^{ip_x x/\hbar}$
Energy	$\hat{H} = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x)$	E	Depends on $V(x)$

Table 1.1: The most important operators, and their corresponding eigenvalues and eigenfunctions, for a quantum system consisting of a single, structureless particle in one dimension.

(where x_0 is a constant) is satisfied by the delta function $\delta(x - x_0)$. (The equation must be satisfied for *all* x , but it is: check it separately for $x = x_0$ and $x \neq x_0$.) The operator for any *function* of x is simply (to multiply by) that function; in particular, the *potential energy operator* is simply (to multiply by) $V(x)$.

Equation 1.52, therefore, simply says that the Hamiltonian operator is the sum of the kinetic energy and potential energy operators: it is the *total energy* operator, and that's why its eigenvalues are the energy values E that correspond to its eigenfunctions. (The reason it's called the Hamiltonian and not just the energy operator has to do with the Hamiltonian formulation of classical mechanics, which you may or may not have studied in the past.¹³)

Table 1.1 summarizes the notation and formulas associated with the position, momentum, and Hamiltonian operators, and their eigenvalues and eigenfunctions.

Problem 1.38. Prove that all three of the operators in Table 1.1 are *linear*, that is, that $\hat{A}(\psi_1 + \psi_2) = \hat{A}\psi_1 + \hat{A}\psi_2$ and $\hat{A}(k\psi) = k\hat{A}\psi$, where \hat{A} can be \hat{x} , \hat{p} , or \hat{H} , for any wavefunctions ψ , ψ_1 , and ψ_2 , and any constant k . All of the operators that we use in quantum mechanics are linear.

Properties of the energy eigenfunctions

In Section 1.4 I claimed that the energy eigenfunctions for the infinite square well are mutually orthogonal, so that if they are also normalized we can write the inner product of any two of them as

$$\langle \psi_m | \psi_n \rangle = \delta_{mn}, \quad (1.56)$$

where δ_{mn} is the Kronecker delta symbol (1 if $m = n$ and 0 otherwise). I further claimed that these functions form a *complete* set, so that any other wavefunction

¹³The Hamiltonian is named after Irish mathematician William Rowan Hamilton, whose story is told in a brilliant video on the YouTube A Capella Science channel, <https://www.youtube.com/watch?v=SZXHoWwBcDc>.

$\psi(x)$ can be written as a linear combination of them:

$$\psi(x) = \sum_n c_n \psi_n(x), \quad (1.57)$$

for some set of complex coefficients c_n . Finally, I used “Fourier’s trick” to show how to calculate these coefficients as inner products:

$$c_n = \langle \psi_n | \psi \rangle. \quad (1.58)$$

These results are not unique to the infinite square well: the energy eigenfunctions for *any* potential energy function $V(x)$ must be both orthogonal and complete. There are some subtle complications when $V(x)$ does not grow to infinity in both directions, in which case the “unbound” eigenfunctions are not quite normalizable and the corresponding eigenvalues vary continuously instead of being quantized. I’d prefer to avoid these complications for now, so let’s assume that $V(x)$ *does* grow to infinity in both directions, causing all wavefunctions to be well behaved, dying out to zero on both sides. Then the energy eigenfunctions will form a discrete (but infinite) set, which we can number in order of increasing energy according to how many nodes they have. In general they won’t be simple sine waves, but they will still obey equations 1.56, 1.57, and 1.58, as long as we define the inner product of two functions as an integral over all x ,

$$\langle \psi_\alpha | \psi_\beta \rangle = \int_{-\infty}^{\infty} \psi_\alpha^*(x) \psi_\beta(x) dx, \quad (1.59)$$

instead of cutting it off at 0 and a as I did in Section 1.4.

Furthermore, and also as in Section 1.4, the coefficients c_n that appear in equations 1.57 and 1.58 have an important physical interpretation: According to the Born rule for energy, the square modulus of c_n is the *probability* of obtaining the corresponding energy E_n if you measure the energy of a particle whose wavefunction is ψ :

$$\left(\begin{array}{l} \text{Probability of finding} \\ \text{particle with energy } E_n \end{array} \right) = |c_n|^2 = |\langle \psi_n | \psi \rangle|^2. \quad (1.60)$$

Again, I’ve written these formulas (1.56, 1.57, 1.58, and 1.60) for the case in which the energy eigenvalues are discrete rather than continuous. When the eigenvalues are continuous the same concepts apply, but it’s trickier to express the concepts with mathematical precision. For that matter, these concepts also apply to other observable quantities besides energy—most notably position and momentum, whose eigenvalues are always continuous. For position, though, I’ve already written down the Born rule in equation 1.6. I’ll show how to express orthogonality and completeness for the momentum eigenfunctions, and write down the Born rule for momentum, in Section 3.1.

Problem 1.39. Suppose that you prepare a million identical systems in the same state ψ , and then measure all of their energies. The predicted *average* outcome would be the sum of all possible outcomes, weighted by their probabilities:

$$\langle E \rangle = \sum_n E_n |\langle \psi_n | \psi \rangle|^2, \quad (1.61)$$

where ψ_n is the eigenfunction corresponding to E_n (and again I'm assuming that the E_n values are discrete). Show that you can also, if you wish, compute the average energy as the inner product

$$\langle E \rangle = \langle \psi | \hat{H} \psi \rangle, \quad (1.62)$$

where \hat{H} is the Hamiltonian operator. (Hint: Work backwards, and use the completeness property to write ψ in terms of energy eigenfunctions.) Can you write a similar formula for the predicted average outcome of a position measurement?

Problem 1.40. In Problem 1.12 you showed that multiplying a wavefunction by a constant phase factor has no effect on position measurement probabilities. Prove that the same is true of energy measurement probabilities.

The time-dependent Schrödinger equation

Among all the observable quantities, energy is special because it is closely tied to time evolution. Recall from Section 1.3 that an energy eigenstate evolves in time with a simple wiggle factor,

$$\psi(t) = \psi(0)e^{-iEt/\hbar} \quad (\text{for an energy eigenfunction}), \quad (1.63)$$

implementing the Einstein relation $E = \hbar\omega$. If we act on this function with the Hamiltonian operator \hat{H} , the wiggle factor simply factors out (since it doesn't depend on x) and we obtain the same factor of E as in equation 1.53:

$$\hat{H}\psi(t) = \hat{H}\psi(0)e^{-iEt/\hbar} = (\hat{H}\psi(0))e^{-iEt/\hbar} = E\psi(0)e^{-iEt/\hbar} = E\psi(t). \quad (1.64)$$

But now notice that we can obtain almost the same expression by instead taking the time derivative:

$$\frac{\partial}{\partial t}\psi(t) = \frac{\partial}{\partial t}(\psi(0)e^{-iEt/\hbar}) = (-iE/\hbar)\psi(0)e^{-iEt/\hbar} = (-iE/\hbar)\psi(t). \quad (1.65)$$

(It's a partial derivative because $\psi(t)$ is also a function of x .) Multiplying both sides by $+i\hbar$ and comparing to the previous equation, we have

$$\hat{H}\psi(t) = i\hbar \frac{\partial \psi}{\partial t}. \quad (1.66)$$

This result would be of little use if it applied only to energy eigenfunctions, for which we already have the more explicit formula 1.63. But amazingly, equation 1.66 applies to *any* wavefunction, because any wavefunction can be written as a sum of

energy eigenfunctions, each of which oscillates in time according to its own wiggle factor. The explicit expansion formula is

$$\psi(t) = \sum_n c_n \psi_n e^{-iE_n t/\hbar}, \quad (1.67)$$

for some collection of complex coefficients c_n , where the sum runs over all energy eigenfunctions ψ_n (and the sum should be replaced by an integral when the eigenvalues E_n are continuous instead of discrete). The calculations of equations 1.64 and 1.65 work the same as before, for each term in this sum. Each term has its own E_n , but everything still matches up, with the same extra factor of $-i/\hbar$, resulting in equation 1.66.

Equation 1.66 is a first-order differential equation that tells us how *any* wavefunction evolves in time. Although I've already written out its explicit solution in equation 1.67, that solution can be cumbersome because it entails an infinite sum and requires that we first find all the energy eigenfunctions and eigenvalues. Sometimes it's simpler just to work with the differential equation directly—not that that's always easy either.

If we write out the Hamiltonian operator explicitly, then equation 1.66 takes the form

$$\left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \right] \psi(x, t) = i\hbar \frac{\partial \psi}{\partial t}. \quad (1.68)$$

Notice that this is a *partial* differential equation, in which ψ is a function of both x and t . Given any initial wavefunction $\psi(x, 0)$, we can in principle use this equation to predict the wavefunction at all future times. Thus, this equation plays the same logical role in quantum mechanics that Newton's second law, $F = m(d^2x/dt^2)$, plays in classical mechanics.

Equation 1.68 (or 1.66) is called the *time-dependent Schrödinger equation*, or the TDSE for short. As you can see, it looks an awful lot like the time-*independent* Schrödinger equation (equation 1.48 or 1.53), because both involve the same Hamiltonian operator. But the logical roles of the two Schrödinger equations are *completely different*: The TISE is the equation that we must solve to find the energy eigenfunctions and their corresponding eigenvalues, while the TDSE is the equation that governs the time evolution of any wavefunction whatsoever.

Problem 1.41. True or false: Every wavefunction must satisfy the TISE. True or false: Every wavefunction must satisfy the TDSE. Explain your answers carefully.

Problem 1.42. A normalized wavefunction had better stay normalized as time passes, right? Therefore the time derivative of the normalization integral should be zero:

$$\frac{d}{dt} \int_{-\infty}^{\infty} \psi^* \psi dx = 0. \quad (1.69)$$

Prove this result. (Hints: Move the derivative inside the integral, where it becomes a partial derivative, because ψ depends on both x and t . Then apply the product rule, and use the TDSE to eliminate the time derivative in each term, being careful with

complex conjugation and minus signs. The potential energy terms should then cancel. In the remaining terms, integrate by parts so instead of a second derivative you have a product of two first derivatives. Why are the boundary terms zero?)

Chapter 2

Bound states in one dimension

Our goal in this chapter will be to solve the time-independent Schrödinger equation (TISE) for a variety of one-dimensional potential functions $V(x)$, focusing on situations in which the particle is trapped or “bound” within a limited region of space. In these situations the energy levels will be discrete, so we want to find these allowed energies and the corresponding energy eigenfunctions. Some of these solutions will have important real-world applications, for example, to molecular vibrations or to electrons in semiconductors.

We’ve already worked out the energies and eigenfunctions for the infinite square well, so a logical next step might be to pick a slightly more complicated $V(x)$ —perhaps a square well with a finite depth, or the smooth potential of a simple harmonic oscillator—and work out its solutions. Instead, though, I’d like to take a more general approach, focusing first on the qualitative features of the energy eigenfunctions, and then showing you how to use numerical methods to solve the TISE for any potential function whatsoever.

2.1 Qualitative solutions of the TISE

Consider a generic potential energy function $V(x)$, such as that shown in Figure 2.1. In the figure I’ve also drawn a horizontal line, at an arbitrary level, to indicate the total energy E of a hypothetical particle that’s subject to this potential.

If this were a *classical* particle, we could immediately determine the qualitative features of the motion from this diagram. We would first note that the particle’s kinetic energy,

$$K = E - V(x), \tag{2.1}$$

cannot be negative, so any x where E is less than $V(x)$ is *classically forbidden*: the particle simply cannot be there if its total energy is only E . The locations where E is greater than $V(x)$ are similarly called *classically allowed*. Within a classically allowed region, the particle will be moving faster where $V(x)$ is smaller (large K) and slower where $V(x)$ is larger (small K). The boundaries between the classically allowed and forbidden regions are called *classical turning points*: a particle at one

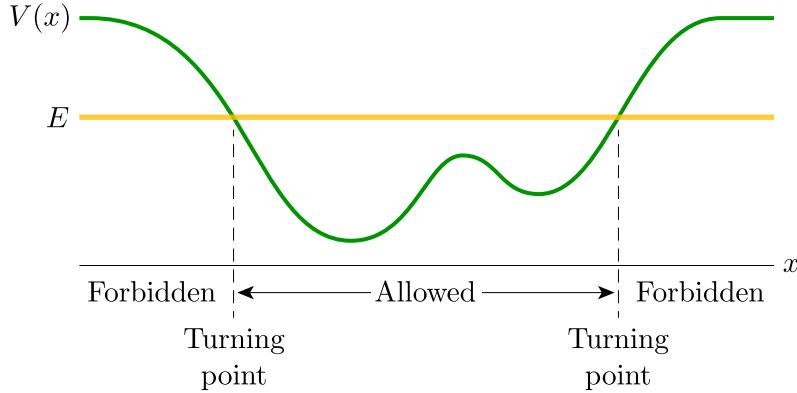


Figure 2.1: An arbitrary potential energy well and hypothetical energy level of a particle trapped in the well.

of these locations must have just come to rest ($K = 0$), after moving in from the allowed region and before moving back into the allowed region. Remember that the *force* on the particle is minus the slope of the potential energy: $F_x = -dV/dx$.

The distinction between allowed and forbidden regions is also critical in quantum mechanics, although we'll see that the word "forbidden" is somewhat of a misnomer. We're interested in wavefunctions with definite energy, so let's look at the TISE, which these functions must obey:

$$\frac{d^2\psi}{dx^2} = -\frac{2m}{\hbar^2}(E - V(x))\psi(x). \quad (2.2)$$

First note that there are no i 's in this equation, so we can always find solutions that are purely real; I'll take ψ to be real for the rest of this chapter. Now focus on the factor $(E - V(x))$ in the TISE. This factor is positive in an allowed region but negative in a forbidden region, and this sign change dramatically affects the shape of the solution $\psi(x)$. In an allowed region, ψ and its second derivative have *opposite* signs, while in a forbidden region, ψ and its second derivative have the *same* sign. The second derivative indicates the *curvature* of the graph of $\psi(x)$: concave-up or concave-down. Thus, $\psi(x)$ curves *toward* the x axis in an allowed region and *away from* the x axis in a forbidden region, as shown in Figure 2.2.

Now let's imagine drawing a sketch of $\psi(x)$, for some given energy. There's no obvious way to get started, so for now I'll just arbitrarily say that we can start with some positive ψ value and some positive slope ($d\psi/dx$), at some particular x value that lies in a classically allowed region (see the Figure 2.3). Then, by the TISE, the graph must be concave-down at this point, so as we move the pencil to the right we should curve downward through a local maximum and then continue downward until the graph crosses the x axis, at which point its curvature will have decreased to zero. Continuing to the right, the graph goes below the x axis but now begins curving upwards, so it then reaches a local minimum and heads back up, crossing the x axis again and finishing a complete "cycle."

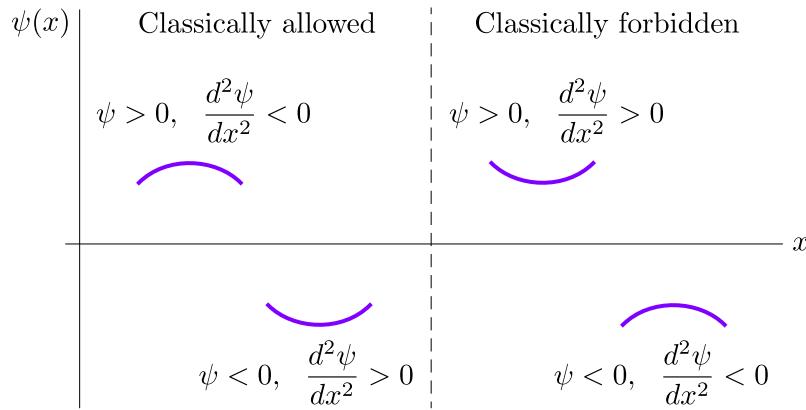


Figure 2.2: Four different cases of the relationship between the sign of ψ and the sign of its second derivative. Not shown are the borderline cases where $d^2\psi/dx^2 = 0$ (an “inflection point”), either at a classical turning point ($E = V(x)$) or where ψ crosses the x axis.

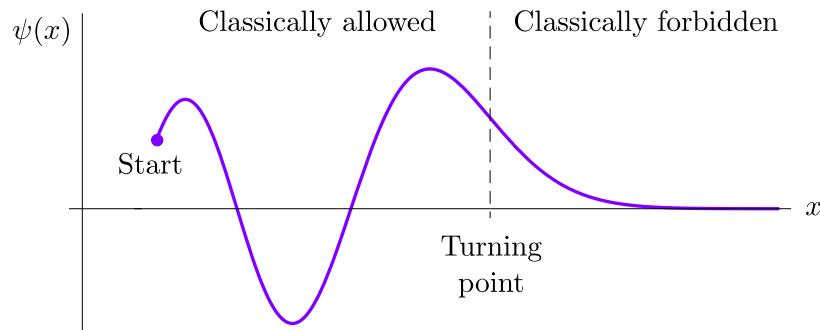


Figure 2.3: A partial sketch of a qualitatively accurate wavefunction (see text for details).

If the potential energy $V(x)$ were *constant* throughout this range of x values, then this oscillation cycle would be a perfect sine wave (see Problem 1.36). In general $V(x)$ will not be constant, so the “wavelength” and “amplitude” of $\psi(x)$ will vary from place to place. You can use the TISE to figure out exactly how they vary,¹ but it’s easier to go back to more basic principles. The de Broglie relation tells us that wavelength increases as the kinetic energy ($p^2/2m$) decreases, or as the potential energy increases. That is, the “waves” will be more spread-out horizontally in locations where $V(x)$ is larger. And the amplitude (that is, the vertical size) of the oscillations will be larger in these locations as well, in agreement with our classical intuition that a particle is more likely to be found in places where it’s moving more slowly. Looking at Figure 2.3, you can see that I drew oscillations that grow larger (both horizontally and vertically) from left to right, so I’ve assumed

¹A. P. French and E. F. Taylor, “Qualitative Plots of Bound State Wave Functions,” Am. J. Phys. **39**(8), 961–962 (1971), <http://dx.doi.org/10.1119/1.1986336>.

that $V(x)$ increases from left to right in the classically allowed region.

Eventually, however, as we move along in x we'll usually come to a classical turning point. There the curvature of the graph goes to zero and on the other side the curvature changes sign. In Figure 2.3 this happens to occur where ψ is positive, so the curvature changes from concave-down, in the classically allowed region on the left, to concave-up, in the classically forbidden region on the right. Often (though not always) the forbidden region extends indefinitely to the right, and in that case the *only* consistent behavior for $\psi(x)$ is to die out asymptotically to zero. (The TISE would also allow ψ to grow to infinity, but such a wavefunction wouldn't be normalizable.) In the special case where $V(x)$ is *constant* throughout a classically forbidden region that extends indefinitely to the right, $\psi(x)$ must be an exponential function of the form $e^{-\kappa x}$, where κ depends on the difference between V and E (again see Problem 1.36). More specifically, a large value of $V - E$ implies a large κ , which makes the wavefunction die out more abruptly. And if there is a classically forbidden region on the left, the wavefunction will die out asymptotically over there as well, taking the form $e^{+\kappa x}$ in the special case where $V(x)$ is constant.

While it is intuitively pleasing that wavefunctions tend to die out exponentially in classically forbidden regions, you may be wondering how it's possible for the wavefunction in these regions to be nonzero at all. How can there be a nonzero probability of finding a particle in a location where its potential energy is greater than its total energy? This behavior is one of the many wonders of quantum mechanics, and is a consequence of the TISE. And there's no actual paradox, because the process of measuring a particle's position will change its energy, or, more precisely, put the particle into a state for which its energy isn't even well defined.

In summary, the solutions to the TISE have the following properties:

- In a classically allowed region, the solutions oscillate. Both the amplitude and the wavelength of the oscillations will be smaller where $E - V(x)$ is large.
- In a classically forbidden region, the solutions have exponential-like behavior, dying out as the distance from a classically allowed region increases. The wavefunction dies out more abruptly where $V(x) - E$ is large.

Figure 2.4 shows an example to illustrate all of these properties, using a piecewise-constant potential function with four different levels.

Now let's think more generally about what an entire solution of the TISE should look like, over the whole range of x . Often, as illustrated in Figures 2.1 and 2.4, the classically allowed region is finite so the particle is trapped, with a classically forbidden region to either side. Then, in the classically allowed region, the wavefunction must go through some integer number of half-oscillations or "bumps," separated by zero-crossings or "nodes." The number of bumps (or nodes) must be an integer, so only certain (average) wavelengths are possible, and this implies that the allowed energies are quantized. The one-bump (zero-node) wavefunction will have the longest (average) wavelength and therefore the lowest energy; the two-bump (one-node) wavefunction will have the second-longest (average) wavelength

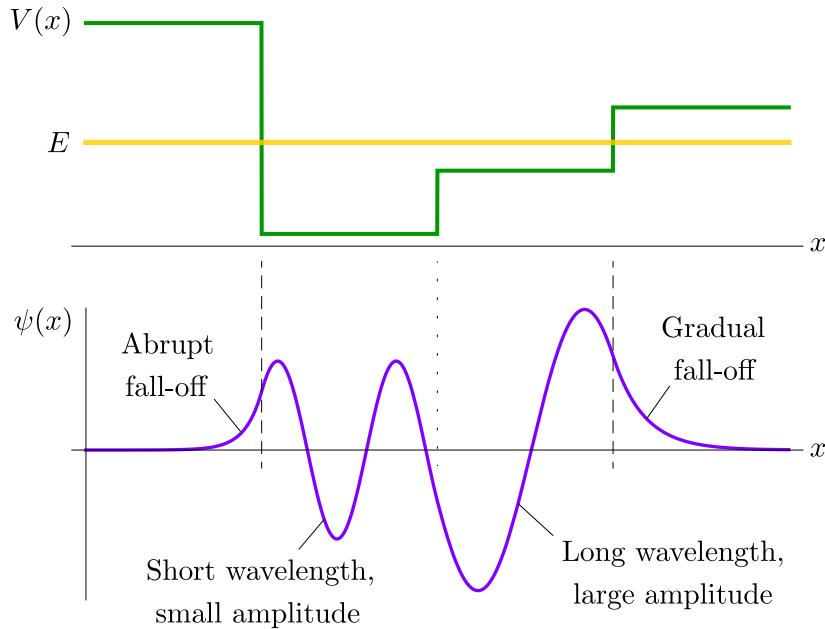


Figure 2.4: An example to illustrate the qualitative properties of a bound-state wavefunction in both classically allowed and classically forbidden zones, and in zones with larger and smaller values of $|E - V(x)|$. (I should confess that I actually used a computer to draw this wavefunction, just to make sure it's accurate. The numerical method that I used is described in the following section.)

and therefore the second-lowest energy; and so on. To determine the numerical values of these energies, of course, requires doing some actual calculations. Given an energy and the corresponding number of bumps or nodes, however, you should now be able to sketch a qualitatively accurate graph of the wavefunction without doing any calculations at all.

On the other hand, if the classically allowed region extends indefinitely to one side or both, then the wavefunction will go through an infinite number of oscillations. These oscillations can vary continuously in size, so the energies will not be quantized. In general, the energies of a quantum particle are quantized when it is “trapped” but not when it is “free.”

Problem 2.1. Notice that the wavefunction in Figure 2.4 is continuous and smooth (that is, its first derivative is continuous), even though $V(x)$ is discontinuous. Prove that as long as $V(x)$ is finite, *every* solution to the TISE must be continuous and smooth. (On the other hand, notice that the infinite square well solutions in Section 1.4 have sharp corners at the ends of the well.)

Problem 2.2. For a given potential well and a given energy, a one-dimensional bound-state wavefunction has a given local wavelength at every point and therefore a given overall shape. In other words, the solution to the TISE (for a given E and $V(x)$) is unique, aside from an uninteresting multiplicative constant. Prove this result more

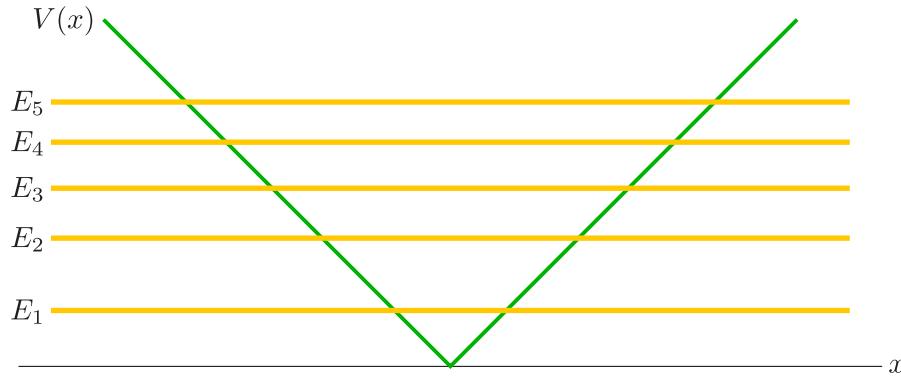


Figure 2.5: A V-shaped potential energy function and its five lowest energy levels (Problem 2.4).

formally, by imagining that there are *two* solutions, $\psi_1(x)$ and $\psi_2(x)$, with the same energy, and showing that they must be proportional to each other. (Hint: Write out the TISE separately for ψ_1 and ψ_2 , then combine these two equations to show that the quantity $\psi_1 d\psi_2/dx - \psi_2 d\psi_1/dx$ (called the *Wronskian*) is a constant. Use the asymptotic behavior of bound states to argue that this constant must be zero, then use this fact to prove that ψ_2 must be a multiple of ψ_1 .) Notice that this proof breaks down if the solutions are unbound, oscillating endlessly in both directions. Can you give a specific example of two distinct unbound solutions to the TISE that have the same energy?

Problem 2.3. Prove that if $V(x)$ is an even function of x , so that $V(-x) = V(x)$, then every bound-state energy eigenfunction must either be even ($\psi(-x) = \psi(x)$) or odd ($\psi(-x) = -\psi(x)$). (Hint: First show that if $\psi(x)$ is an energy eigenfunction then so is $\psi(-x)$. Then use the result of the previous problem.)

Problem 2.4. Figure 2.5 shows a V-shaped potential energy function and its five lowest energy levels. Using the same horizontal scale, draw qualitatively accurate graphs of the five corresponding energy eigenfunctions.

Problem 2.5. Figure 2.6 shows an asymmetric potential energy function and its three lowest energy levels. Using the same horizontal scale, draw qualitatively accurate graphs of the three corresponding energy eigenfunctions.

Problem 2.6. Figure 2.7 shows a symmetric “double well” potential energy function and its ground-state energy. Notice that for this energy there are two classically allowed regions, separated by a classically forbidden region. (a) Using the same horizontal scale, draw a qualitatively accurate graph of the ground-state wavefunction. (Hint: As in simpler examples, the ground-state wavefunction has no nodes; but in this example you shouldn’t make any assumption about the number of “bumps,” because the definition of a bump is ambiguous.) (b) Argue that the energy of the first excited state will be only slightly higher than that of the ground state, and draw a qualitatively accurate graph of the first-excited-state wavefunction.

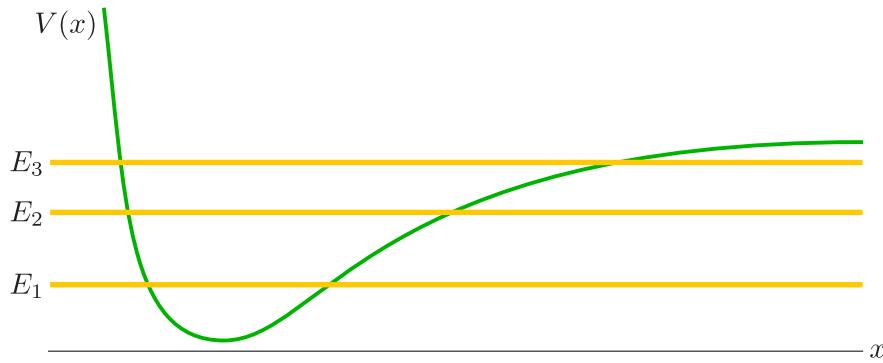


Figure 2.6: An asymmetric potential energy function and its three lowest energy levels (Problem 2.5).

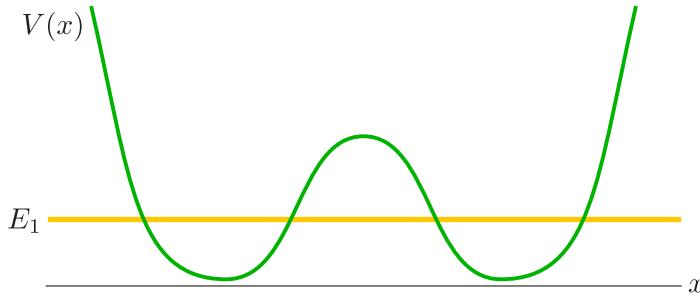


Figure 2.7: A “double well” potential and its lowest energy level (Problem 2.6).

2.2 The shooting method

Are you dissatisfied with the merely qualitative solutions of the previous section? Good! In this section I’ll make the same ideas quantitative, showing how to solve the time-independent Schrödinger equation (TISE) for any energy eigenvalue E and the corresponding eigenfunction $\psi(x)$. The solution will be in numerical form (a list or a plot of function values), because there is no analytical method that works for an arbitrary $V(x)$. As in the previous section, the idea is to work our way across the region a little at a time, using the TISE to adjust the wavefunction’s curvature.

Discretizing the TISE

For any given $V(x)$ and energy E , the TISE relates the value of the function $\psi(x)$ at each point to its second derivative at that point:

$$\frac{d^2\psi}{dx^2} = -\frac{2m}{\hbar^2}(E - V(x))\psi(x). \quad (2.3)$$

To solve this equation numerically, we need to replace the second derivative with a discrete approximation, expressed in terms of ordinary arithmetic instead of cal-

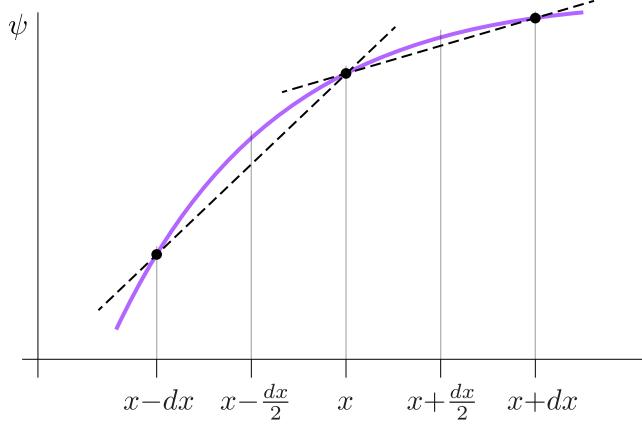


Figure 2.8: The slopes of the two dashed lines are good approximations to the derivative $d\psi/dx$ evaluated at $x + \frac{dx}{2}$ and $x - \frac{dx}{2}$, respectively. The difference of these slopes, divided by dx , therefore gives a good approximation to $d^2\psi/dx^2$ at the point x .

culus. I'll do this in two steps, one derivative at a time. First, because the second derivative is the derivative of the first derivative, we can write it approximately as

$$\frac{d^2\psi}{dx^2} \approx \frac{\left(\frac{d\psi}{dx} \text{ at } x + \frac{dx}{2}\right) - \left(\frac{d\psi}{dx} \text{ at } x - \frac{dx}{2}\right)}{dx}, \quad (2.4)$$

provided that the increment dx is sufficiently small. Note that I've used a symmetrical “centered difference” approximation, evaluating $d\psi/dx$ at a half step away from x in either direction, which is more accurate than stepping only in one direction, to $x + dx$ or $x - dx$. Now we do the same thing with each of the two first derivatives in the numerator:

$$\begin{aligned} \left(\frac{d\psi}{dx} \text{ at } x + \frac{dx}{2}\right) &\approx \frac{\psi(x+dx) - \psi(x)}{dx}, \\ \left(\frac{d\psi}{dx} \text{ at } x - \frac{dx}{2}\right) &\approx \frac{\psi(x) - \psi(x-dx)}{dx}. \end{aligned} \quad (2.5)$$

These approximations are illustrated in Figure 2.8. Plugging these expressions into equation 2.4, we obtain what I'll call the *second centered difference* approximation for $d^2\psi/dx^2$:

$$\frac{d^2\psi}{dx^2} \approx \frac{\psi(x+dx) + \psi(x-dx) - 2\psi(x)}{(dx)^2}. \quad (2.6)$$

As a check, note that this expression is positive if $\psi(x)$ is less than the average of ψ at the two neighboring points ($x+dx$ and $x-dx$), and negative if $\psi(x)$ is greater than the average of ψ at the two neighboring points.

With the approximation of equation 2.6 and a little rearrangement, the TISE becomes

$$\psi(x+dx) \approx 2\psi(x) - \frac{2m(dx)^2}{\hbar^2} (E - V(x)) \psi(x) - \psi(x-dx). \quad (2.7)$$

So if we know the values of ψ at two neighboring points, x and at $x-dx$, we can calculate its value at the next point to the right, $x+dx$. We can then move a step to the right and repeat the process to get ψ at the next point, and so on, eventually obtaining ψ values over the entire region of interest. The smaller the step dx , the more accurate these values will be. Plotting an accurate graph of the entire function $\psi(x)$ might require hundreds of calculation steps, but today's computers can carry out a billion calculations per second.

You may still be wondering how to find the two starting values $\psi(x)$ and $\psi(x-dx)$, as well as how to deal with our initial ignorance of the energy eigenvalue E . I'll address these issues in the context of a concrete example.

Problem 2.7. Check the accuracy of equation 2.6 for the function $\psi(x) = e^x$ at $x = 0$, for $dx = 1$, $dx = 0.1$, and $dx = 0.01$. What happens to the error when dx is reduced by a factor of 10?

Problem 2.8. Prove that the second centered difference approximation (equation 2.6) is *exact* if $\psi(x)$ is a quadratic polynomial ($ax^2 + bx + c$).

The finite square well

To illustrate the method, let me pick a specific $V(x)$: the *finite square well*, pictured in Figure 2.9 and defined as

$$V(x) = \begin{cases} 0 & \text{for } -a/2 < x < a/2, \\ V_0 & \text{elsewhere.} \end{cases} \quad (2.8)$$

This is the same potential as for the infinite square well, with ∞ replaced by V_0 ; I've shifted the well to center it at $x = 0$ because the resulting symmetry will slightly simplify the computer code and the description of the solutions. (This is actually an example that *can* be solved exactly, aside from the need to numerically solve a transcendental equation to match the wavefunction at the well boundary. But here I'll use it to illustrate the much more general method of numerically solving the TISE.)

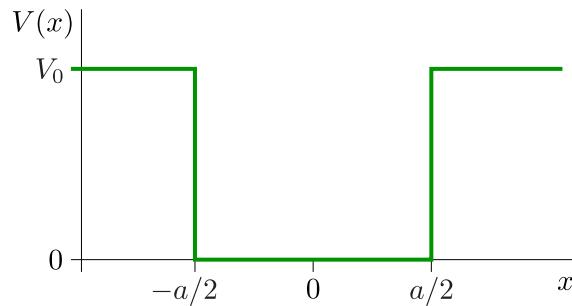


Figure 2.9: The finite square well potential, centered on $x = 0$ for convenience.

Problem 2.9. Use the principles of Section 2.1 to sketch qualitatively accurate graphs of the three lowest-energy eigenfunctions of the finite square well, assuming that all three energies are less than V_0 .

Natural units

Before typing a physics equation into a computer, you should almost always rewrite it in a system of units that is “natural” to the problem being solved. Doing so will free you from working with numbers that are awkwardly large or small, and from having to supply numerical values for parameters that turn out to be irrelevant to the mathematics. For example, in this problem the natural unit of distance is a , the width of the well, so I’ll set $a = 1$ in my computer code. I will also set $m/\hbar^2 = 1$; this combination has dimensions of $(\text{energy})^{-1}(\text{distance})^{-2}$, so setting it equal to 1 determines our unit of energy: all energies will now be measured in units of \hbar^2/ma^2 .

Note that after making these choices, we do *not* have the freedom to also set $V_0 = 1$. In other words, different V_0 values (in these units) represent different problems to solve, and we’ll have to choose a specific V_0 before solving the problem on a computer. And what would be some interesting V_0 values to choose? Well, recall that for an *infinite* square well, the energy eigenvalues are $\hbar^2 n^2 / 8ma^2$. Plugging in $\hbar = 2\pi\hbar$ and setting $\hbar^2/ma^2 = 1$, this becomes $(\pi^2/2)n^2 \approx 5n^2$, so the lowest energies in our units would be roughly 5, 20, 45, 80, and so on. The most interesting V_0 values should be in this range; values much less than 5 would tend not to trap the particle at all (though see Problem 2.12), while values much more than 100 start looking similar to infinity (at least for the low-energy states). I’ll use $V_0 = 50$.

Problem 2.10. Suppose we’re talking about an electron in a one-dimensional finite square well with a width of 10 nm. What is the size of one natural unit of energy, in electron-volts? If $V_0 = 50$ in natural units, what is V_0 in eV?

Mathematica code

Without further ado, here is some Mathematica code² for solving the TISE for a finite square well, using the units just described, with $V_0 = 50$:

```
v[x_] := If[Abs[x] < 0.5, 0, 50];
energy = 5;
xMax = 1.5;
dx = 0.001;
x = -xMax + dx;
lastPsi = 0.0;
thisPsi = 0.001*dx;
```

²If you aren’t already familiar with the basic syntax of Mathematica, including function definitions, lists, `While`, `AppendTo`, and `ListPlot`, see Appendix B.

```

psiList = {{-xMax, lastPsi}, {x, thisPsi}};
While[x < xMax,
  nextPsi = 2*thisPsi*(1 - dx^2*(energy - v[x])) - lastPsi;
  x += dx;
  AppendTo[psiList, {x, nextPsi}];
  lastPsi = thisPsi;
  thisPsi = nextPsi];
ListPlot[psiList, Joined -> True]

```

First I define the potential energy function $v[x_]$ (using lower case to avoid conflicts with built-in Mathematica functions, which always start with capital letters). Then I initialize three constants: an arbitrary initial guess for the energy (roughly equal to the ground state energy of an infinite well); the maximum x value that I'll ask the computer to look at (so the range of x values will be $-x_{\text{Max}}$ to x_{Max}); and the step size dx (which must be much less than 1, the natural unit of length—though you can get away with 0.01 or more when the potential energy doesn't have any discontinuities, as it does here). Next, to set up the step-by-step calculation, I initialize x to a value that's one step to the right of $-x_{\text{Max}}$, and initialize the variables `lastPsi` and `thisPsi` to represent $\psi(x-dx)$ and $\psi(x)$. My assumption here is that the wavefunction is negligibly small at $-x_{\text{Max}}$, so I can set `lastPsi` to zero and `thisPsi` to a tiny value. This tiny value is actually arbitrary, because it will merely determine the overall normalization of the wavefunction—but to keep things conceptually reasonable I've set the slope to a small value, coincidentally also equal to 0.001. The final initialization is to create a list, `psiList`, of the (x, ψ) pairs that have been determined so far.

With these initializations (and all the decisions they entail) out of the way, the calculation itself is fairly straightforward. I've used a `While` loop that will execute repeatedly as long as x is less than x_{Max} . During each iteration I set the variable `nextPsi` according to equation 2.7. Then I increment the value of x , append the new (x, ψ) pair to `psiList`, and get ready for the next step by renaming `thisPsi` to `lastPsi` and `nextPsi` to `thisPsi`. After the loop ends I pass the completed list to the `ListPlot` function so we can see what the solution looks like.

I strongly recommend that you type³ the preceding code into your Mathematica system and use it for your first few numerical solutions of the TISE. But once you understand what it's doing, you might as well use the following shortcut. Mathematica conveniently provides a numerical differential equation solving function called `NDSolve`, which accomplishes essentially the same thing as my code above but uses a more sophisticated algorithm that chooses the step size (equivalent to dx above) adaptively to try to attain a certain final accuracy. To use `NDSolve` you would keep the first three lines of my code above, but then replace the rest of it with the following:

³If you're working from an electronic copy of this book, please resist the urge to copy and paste my code. Typing each instruction with your own fingers will help you make the code your own, to develop a *working* knowledge of computational quantum mechanics.

```
solution = NDSolve[{psi''[x]==-2(energy-v[x])psi[x],
    psi[-xMax]==0, psi'[-xMax]==0.001}, psi, {x, -xMax, xMax}];
Plot[psi[x] /. solution, {x, -xMax, xMax}]
```

The `NDSolve` function requires a list (in curly braces) of the differential equation(s) and the boundary condition(s). Note that these equations are defined using double `==` signs, and that derivatives are denoted by primes (''). For a second-order differential equation, the required boundary conditions are the value and the first derivative of the function, both at the same point. After this list of equations, I supply the name of the function to solve for and a list consisting of the independent variable and its beginning and ending points. The result from `NDSolve` is stored in the variable `solution` as what Mathematica calls an interpolating function; the last line of code plots a graph of this function.

Results for the finite square well

If you execute either version of the code above (or the equivalent in some other software environment), you'll get a plot of a function that rises gradually from left to right, peaks a little to the left of $x = 0$, then falls gradually and crosses the x axis a little to the left of the edge of the well at $x = 0.5$. The function then becomes negative, but soon reaches $x = 0.5$ where it begins curving *away* from the horizontal axis, blowing up exponentially in the negative direction. This is *not* a normalizable wavefunction, and it teaches us a lesson: You can solve the TISE for *any* energy E , but not all energy values allow solutions that are normalizable.

The procedure, then, is to rerun the code with different E values until you get a

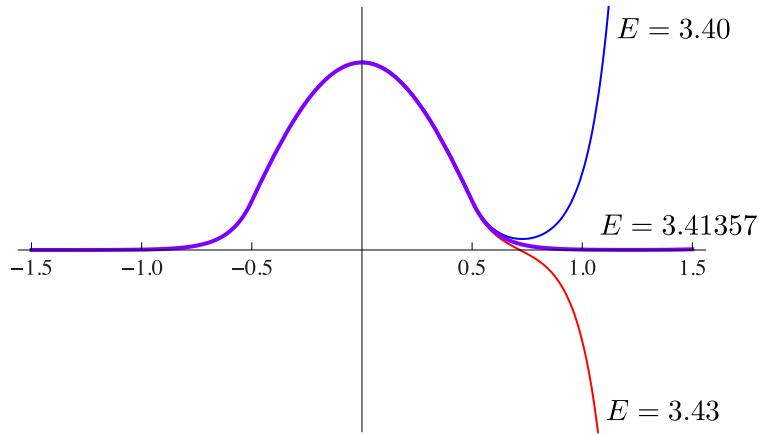


Figure 2.10: A combined plot showing the results of three numerical solutions to the TISE for the finite square well with $V_0 = 50$. Each numerical integration starts with a negligibly small function at the far left, as described in the text. For most E values the solution is not normalizable, blowing up to $+\infty$ or $-\infty$ in the classically forbidden region at the right. Fine-tuning the energy to $E \approx 3.41357$, however, produces a normalizable solution.

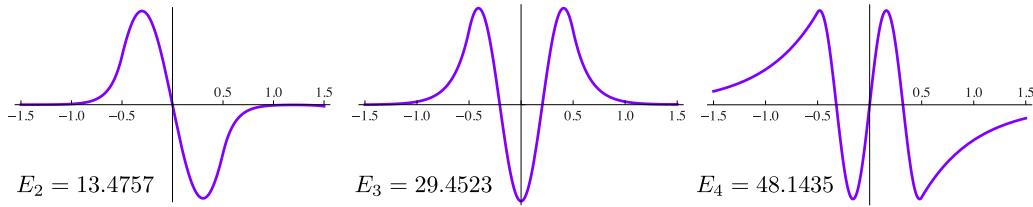


Figure 2.11: The first three excited states for a finite square well with $V_0 = 50$, obtained by the same numerical method as in Figure 2.10. The third excited-state wavefunction (with energy E_4) extends much farther into the classically forbidden regions, so I set `xMax` to 4.0, even though the plot extends only to ± 1.5 .

solution that “lies down flat” to the right of the potential energy well. It’s essentially a trial-and-error process, but with a little practice you can zero-in on an energy value that works, to several significant figures, in about 20 trials. Figure 2.10 shows three trials, one with the energy too low, one with the energy too high, and one with the energy just right to give the one-bump (ground-state) wavefunction. In the same way, I found the next three energies and wavefunctions, shown in Figure 2.11

Notice that all of these solutions are sinusoidal inside the well and exponential outside it. (Because the exponential fall-off is so gradual with the last of these wavefunctions, I increased `xMax` to 4.0 to get a consistent result—though I still cut off the plot at ± 1.5 .) Notice also that the energies are all significantly *less* than the corresponding infinite square well energies, $(\pi^2/2)n^2 \approx 4.93, 19.74, 44.41, 78.96$; that’s because the finite well lets part of the wavefunction “spill out” beyond the edges, allowing the wavelength inside to be longer for the same number of bumps. There are no further normalizable solutions with $E < 50$. For $E > 50$, the solutions are sinusoidal even outside the well, like the solutions for a free particle. Thus, this particular potential well admits exactly four bound-state solutions to the time-independent Schrödinger equation. Figure 2.12 shows all four bound-state wavefunctions and their energies, comparing to those of the infinite square well.

I haven’t labeled the vertical axes in Figures 2.10 and 2.11, because the vertical scales are determined by my arbitrary choice of $d\psi/dx = 0.001$ at the extreme left edge. To obtain normalized wavefunctions, we would have to compute $\int |\psi|^2 dx$ in each case, and divide ψ by the square root of the result.

There’s one more thing to notice about the four solutions pictured in Figures 2.10 and 2.11: Each of them is either an even or odd function of x . This will be true whenever the potential function $V(x)$ is an even function (see Problem 2.3), so for all such potentials there’s actually a better choice of boundary conditions: Instead of starting far out along the $-x$ axis, start at $x = 0$ and either set $\psi = 1$ and $d\psi/dx = 0$ to obtain the even functions, or set $\psi = 0$ and $d\psi/dx = 1$ to obtain the odd functions. (In both cases, the “1” is arbitrary; any other nonzero value will do.) Then both “tails” of the wavefunction will “wag” as you vary the energy, lying down flat when the energy is tuned to an eigenvalue. These boundary conditions avoid the awkwardness that arises when the starting point isn’t far enough to the

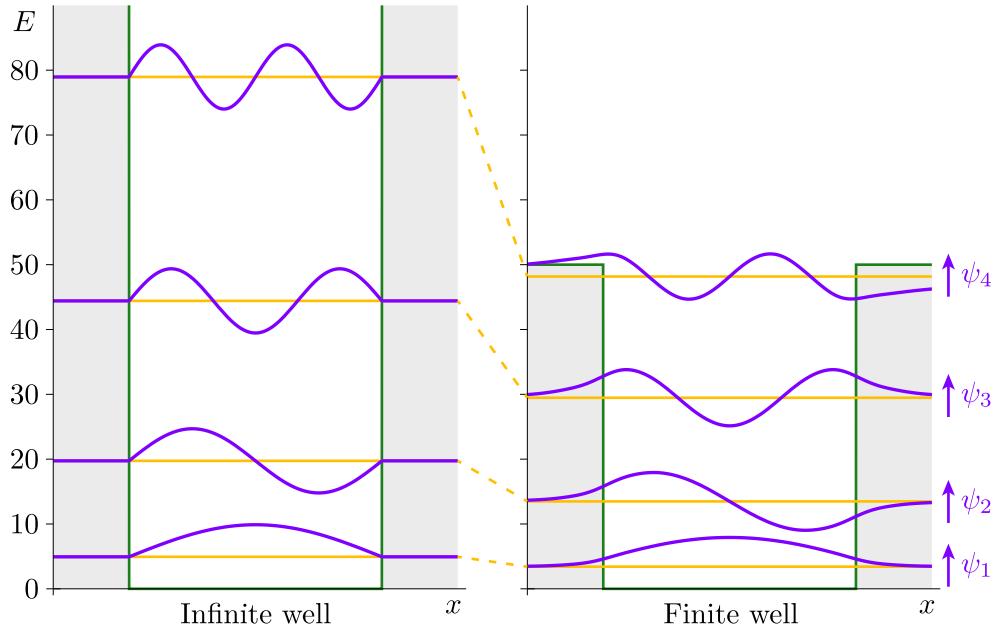


Figure 2.12: Comparison of the energy levels and eigenfunctions of infinite and finite square wells with the same width. All energies are measured in natural units of $\hbar^2/(ma^2)$, where m is the particle mass and a is the well width. The depth of the finite well is 50 in these units. Note that the vertical scale for the wavefunction graphs is unrelated to the energy scale. The energies in the finite well are lower than those in the infinite well because the finite well allows the wavefunctions to “spill out” into the classically forbidden regions, resulting in longer wavelengths.

left. In this example I used the more awkward boundary conditions because this method works even when $V(x)$ isn’t symmetric.

The algorithm that I’ve just described, in which we start at a point with a known boundary condition and adjust the energy until the other boundary condition is met, is called the *shooting method*, because it is reminiscent of repeatedly shooting projectiles while tuning the launch speed (or angle) to hit a fixed target. The shooting method is extremely accurate and computationally efficient, though it can be a bit tedious, finicky, and difficult to automate.

Problem 2.11. Run the explicit shooting-method code on page 51, to obtain the ground-state energy of the finite square well (with $V_0 = 50$) to four decimal places. Then increase the step size dx to 0.01 and repeat the process. By what percentage do the results differ? Still setting dx to 0.01, what happens if you modify the potential energy function definition to replace $<$ with $<=$? The problem, as you can now see, is that because both walls of the well fall on the step boundaries, the effective width of the well is slightly smaller than it should be when you use $<$, but slightly larger than it should be when you use $<=$. To fix the problem, shift the well to center it on an x value that is slightly (by less than dx) to one side of $x = 0$, and then repeat the shooting process one last time. Discuss the results. Also take a moment to show that

for an *infinite* square well, a 1% error in the well width results in a 2% error in the energy eigenvalues.

Problem 2.12. Use the shooting method to show that a finite square well with $V_0 = 5.0$ (and a width of 1 in natural units) has only one bound state. Then explore what happens as you further decrease V_0 to 2.0, 1.0, 0.5, 0.2, and 0.1. Find the bound-state energy (and plot the wavefunction) for each of these values, and discuss the results. (Hints: Use appropriate boundary conditions at $x = 0$, as described in the text. Note that you'll have to increase `xMax` more and more as V_0 decreases.)

Problem 2.13. Consider a particle subject to the potential energy function

$$V(x) = \begin{cases} \alpha x & \text{for } x > 0, \\ \infty & \text{for } x < 0, \end{cases}$$

where α is a positive constant.

- (a) First describe the motion of a *classical* particle subject to this potential energy function. Why is a quantum particle subject to this potential sometimes called a *quantum bouncer*?
- (b) For the quantum case, a natural system of units would set $\alpha = 1$ and $\hbar^2/m = 1$. In terms of these constants, what would then be the natural units of length and energy? Evaluate these natural length and energy units numerically for the case of an electron with $\alpha = 10$ eV/nm.
- (c) Carefully describe the boundary conditions that a solution to the TISE must obey, for a particle subject to this potential.
- (d) Use the shooting method to find the three lowest energy eigenvalues (to at least four significant figures) and plot the corresponding wavefunctions. Specify boundary conditions at $x = 0$, and be sure that your solutions also have the required behavior at large x . Use natural units, but also convert the energy eigenvalues to eV for the case of an electron with $\alpha = 10$ eV/nm.

2.3 The matrix method

Now let me describe a totally different numerical method for solving the TISE. This method is more abstract and takes more effort to understand than the shooting method, but the investment will pay off by providing a deeper perspective on the structure of quantum mechanics. Moreover, for many problems this method ends up being easier to use: instead of laboriously hunting for each individual bound-state solution, we can use the method of this section to find all of a system's bound states at once.⁴

Again, our goal is to solve the time-independent Schrödinger equation,

$$\hat{H}\psi = E\psi, \tag{2.9}$$

⁴Although this method dates back to the birth of quantum mechanics, I learned how practical it can be from a series of papers by F. Marsiglio and collaborators, beginning with “The harmonic oscillator in quantum mechanics: A third way,” Am. J. Phys. **77**(3), 253–258 (2009), <https://doi.org/10.1119/1.3042207>.

for the unknown eigenfunctions ψ and their corresponding eigenvalues E . The basic idea is to write ψ as a linear combination of some collection of orthonormal basis functions, which I'll call ϕ_n :

$$\psi = \sum_{n=1}^{\infty} c_n \phi_n. \quad (2.10)$$

Our goal is then to find the unknown coefficients c_n . The basis functions ϕ_n are *not* the solutions to the Schrödinger equation that we seek (if they were, the problem would already be solved!). They should instead be some collection of relatively simple functions that form an orthonormal basis. Here I will take the ϕ_n to be the sine functions that are energy eigenfunctions for an *infinite* square well that's wide enough to contain (to a good approximation—we hope!) all of the eigenfunctions ψ that we care about:

$$\phi_n(x) = \sqrt{\frac{2}{b}} \sin\left(\frac{n\pi x}{b}\right) \quad \text{for } 0 < x < b, \quad (2.11)$$

where b is the width of the (hypothetical) infinite well. (When I apply this method to the finite square well example in a moment, I will take b to be several times larger than a , and I'll center the finite well at $b/2$.)

Still working in general, we now insert the expansion for ψ (equation 2.10) into the TISE (equation 2.9), and move \hat{H} and E inside the sums to obtain

$$\sum_{n=1}^{\infty} \hat{H} c_n \phi_n = \sum_{n=1}^{\infty} E c_n \phi_n. \quad (2.12)$$

We can get rid of the sum on the right by using *Fourier's trick*: multiply on the left by ϕ_m^* (for some arbitrary index m) and integrate over x :

$$\int_{-\infty}^{\infty} \phi_m^* \sum_{n=1}^{\infty} \hat{H} c_n \phi_n dx = \int_{-\infty}^{\infty} \phi_m^* \sum_{n=1}^{\infty} E c_n \phi_n dx. \quad (2.13)$$

Now, on the right-hand side, we can move the integral inside the sum, and factor $E c_n$ out of the integral; the integral then gives simply δ_{mn} , because the ϕ_n functions are orthonormal, and therefore the only term that contributes to the sum is the one with $n = m$. The left-hand side isn't so simple, but we can at least move the integral inside the sum and factor the c_n out of the integral. Thus we obtain

$$\sum_{n=1}^{\infty} \left(\int_{-\infty}^{\infty} \phi_m^* \hat{H} \phi_n dx \right) c_n = E c_m. \quad (2.14)$$

The quantity inside the big parentheses is called the *mn matrix element* of the Hamiltonian operator,

$$H_{mn} = \text{"matrix element"} = \int_{-\infty}^{\infty} \phi_m^*(x) \hat{H} \phi_n(x) dx = \langle \phi_m | \hat{H} \phi_n \rangle, \quad (2.15)$$

and it's something that a computer can calculate straightforwardly. With this abbreviation, the TISE becomes

$$\sum_{n=1}^{\infty} H_{mn} c_n = E c_m, \quad (2.16)$$

which has the precise form of the eigenvalue equation for the *matrix*⁵ whose elements are H_{mn} :

$$\begin{pmatrix} H_{11} & H_{12} & \cdots \\ H_{21} & H_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \end{pmatrix} = E \begin{pmatrix} c_1 \\ c_2 \\ \vdots \end{pmatrix}. \quad (2.17)$$

Therefore the energy eigenvalues E are the eigenvalues of the H matrix, and the eigenfunctions $\psi(x)$ can be built from the basis functions $\phi_n(x)$ using the elements of the corresponding eigenvectors. This form of the TISE is useful because many mathematical computing environments include routines that can quickly find the eigenvalues and eigenvectors of remarkably large matrices. Of course, you first have to calculate all the matrix elements using equation 2.15, and that's often the most time-consuming part of the process.

Calculating the matrix elements is somewhat simplified if we break the Hamiltonian operator into two pieces:

$$\hat{H} = \hat{H}_0 + \Delta V(x). \quad (2.18)$$

Here \hat{H}_0 is the Hamiltonian whose eigenfunctions are $\psi_n(x)$, and $\Delta V(x)$ is whatever is left of our actual Hamiltonian. For our infinite-square-well basis, \hat{H}_0 is just the kinetic energy operator, $\Delta V(x)$ is simply the $V(x)$ for our actual potential energy function, and the integrals will run only from 0 to b because the basis functions are zero elsewhere. Thus, the matrix elements are

$$\begin{aligned} H_{mn} &= \int_0^b \phi_m^*(x) \hat{H}_0 \phi_n(x) dx + \int_0^b \phi_m^*(x) V(x) \phi_n(x) dx \\ &= E_n \delta_{mn} + \int_0^b \phi_m^*(x) V(x) \phi_n(x) dx, \end{aligned} \quad (2.19)$$

where in the first term I've used the fact that ϕ_n is an eigenfunction of H_0 with eigenvalue E_n , and the fact that the ϕ_n functions are orthonormal. Written out as a matrix, the first term would be simply a diagonal matrix whose entries are the eigenvalues of the infinite square well, $\pi^2 n^2 / 2b^2$ in natural units.

The finite square well

Now let me apply this method to the same example used in the previous section: a finite square well of width $a = 1$ and depth $V_0 = 50$ (in natural units). I'll center the

⁵If you need a review of matrices and matrix multiplication, see Appendix A.5.

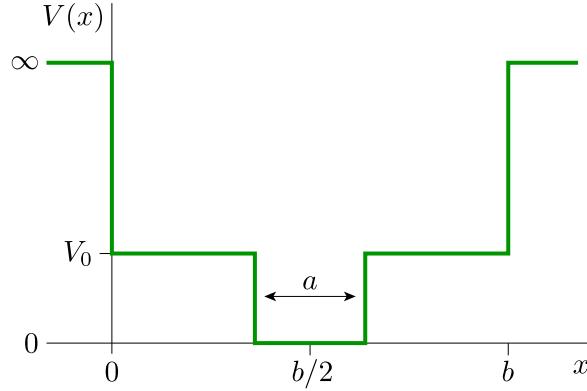


Figure 2.13: Potential energy function of a finite well of width a and depth V_0 centered within an infinite well of width b . The conventions used in this section are that the minimum x and minimum V are both zero.

finite well inside the infinite well, as shown in Figure 2.13. Here is some Mathematica code to find the energy eigenvalues E and the corresponding eigenvectors (c_1, c_2, \dots) for this system:

```
b = 4;
v[x_] := If[Abs[x - b/2] < 0.5, 0, 50];
nMax = 50;
basis[n_, x_] := Sqrt[2/b]*Sin[n*Pi*x/b];
hMatrix = Table[(n^2 Pi^2/(2b^2)) KroneckerDelta[m,n] +
  NIntegrate[basis[m,x]*v[x]*basis[n,x], {x, 0, b},
  AccuracyGoal->8], {m, 1, nMax}, {n, 1, nMax}];
```

In the first line I set $b = 4$, making the infinite well and its sine-wave basis functions four times as wide as the finite well of width $a = 1$ that we want to study. Then I define the potential energy function for the finite well, and next I set a cutoff `nMax` of 50, which will be the largest n value used throughout the calculations (and the dimension of the eigenvectors and matrices). The fourth line defines the sine-wave basis functions, properly normalized. Finally I build the H matrix using the `Table` function and formula 2.19. This calculation entails 50×50 numerical integrations, and is where nearly all of the computational time is spent; it takes about a minute to execute on my laptop computer (see Problem 2.17 for a clever way to speed up this step).

I've used a semicolon in the code to suppress the output of the Hamiltonian matrix, because it's cumbersome and you don't actually need to see it. Still, its structure is rather interesting, so let's take a quick look. To view all the matrix elements in decimal approximation, I suggest the following instruction:

```
PaddedForm[Chop[MatrixForm[hMatrix]], {3, 1}]
```

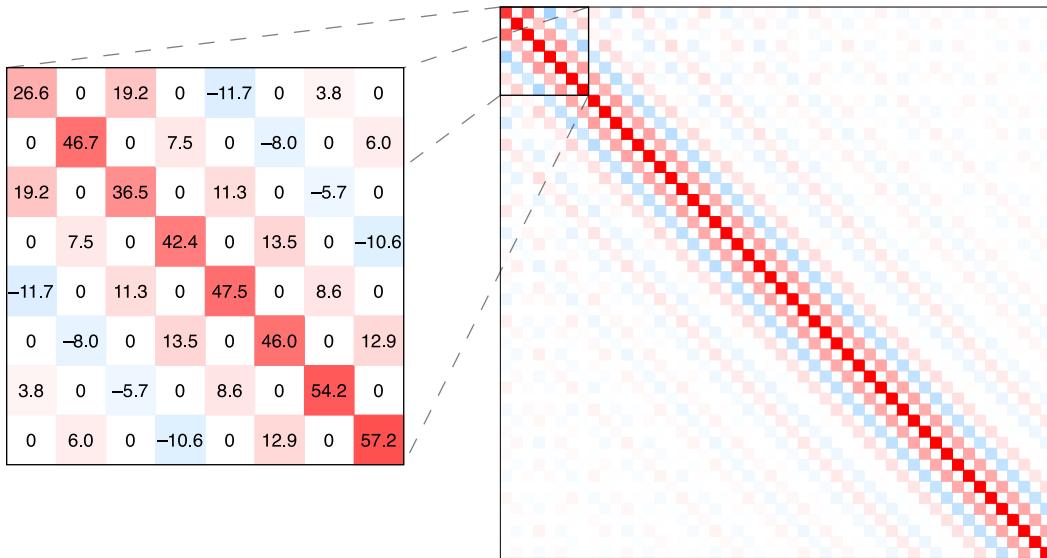


Figure 2.14: A graphical representation of the Hamiltonian matrix for the finite square well potential shown in Figure 2.13, using a basis of infinite-square-well eigenfunctions as described in the text. Shades of red represent positive values, shades of blue represent negative values, and white represents zero. The expanded image on the left shows numerical values for the upper-left 8×8 corner of the matrix.

Alternatively, you can use Mathematica's `MatrixPlot` function to display a pretty graphical representation of the matrix:

```
contrast = 0.01;
MatrixPlot[hMatrix*contrast + 0.5,
  ColorFunction -> Function[f, Blend[{Cyan, White, Red}, f]],
  ColorFunctionScaling -> False]
```

(The offset of 0.5 ensures that zero is represented by white. Choose the negative and positive colors to suit your taste, and adjust the `contrast` variable to emphasize different details.) Figure 2.14 shows both a graphical display and some selected numerical values. Notice the various patterns, which you can explore in Problem 2.15.

What we want, though, are the eigenvalues and eigenvectors of this matrix. To obtain them we can simply use Mathematica's `Eigensystem` function:

```
{eValues, eVectors} = Eigensystem[hMatrix];
```

Again I've used a semicolon to suppress the cumbersome output. But now the variable `eValues` contains a list of the eigenvalues, and if you type its name you get the following:

```
{811.179, 780.225, 749.343, 719.877, 690.352, 662.866, 634.565, 608.144,
581.992, 555.874, 531.607, 506.755, 483.011, 460.415, 437.107, 415.862,
```

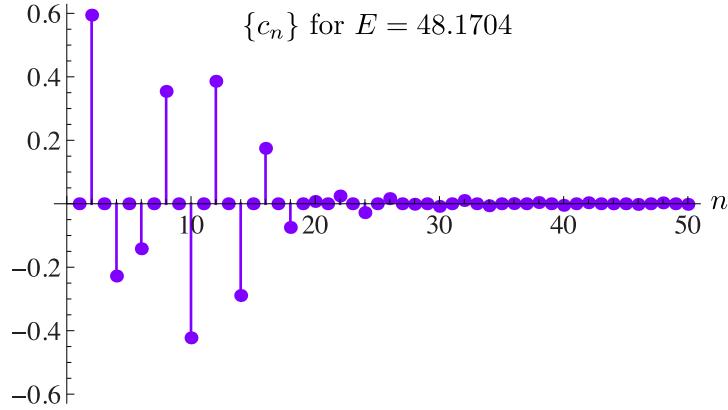


Figure 2.15: This plot shows the components c_n of one of the eigenvectors for the finite square well analyzed in this section. This particular eigenvector corresponds to eigenvalue $E = 48.1704$, just slightly below $V_0 = 50$ and therefore only weakly bound.

```
394.466, 373.352, 354.171, 333.967, 315.485, 297.698, 279.237, 263.283,
246.413, 230.559, 216.365, 200.724, 187.821, 174.449, 161.206, 150.544,
137.827, 128.003, 118.079, 107.628, 100.49, 90.6015, 84.3394, 77.3956,
70.5479, 66.78, 59.9372, 57.9902, 52.9032, 52.0802, 48.1704, 29.4697,
13.4824, 3.41566}
```

Mathematica inconveniently sorts the eigenvalues in descending order (and sorts the eigenvectors to correspond); although we could reverse the order with some list manipulation functions, I won't bother. The eigenvalues that we want are the last four, which are less than $V_0 = 50$ and therefore correspond to bound states of the finite well. Moreover, all four of these values agree to three significant figures with the more accurate values found by the shooting method in Section 2.2. The other 46 eigenvalues are artifacts of the fictitious infinite well; in fact the finite well allows *any* energy greater than V_0 .

We could also print out the components of the eigenvectors, but a picture is worth a thousand numerals. You can plot the components of any eigenvector using an instruction like `ListPlot[eVectors[[47]], PlotRange -> All]`, remembering again that the bound states are numbered 47 through 50. With some further tweaking for aesthetics, this instruction produces the plot shown in Figure 2.15. To see the actual function of x that corresponds to this eigenvector, we simply build it out of the basis functions according to equation 2.10:

```
Plot[Sum[eVectors[[47,n]]*basis[n,x], {n,1,nMax}], {x,0,b}]
```

This instruction (again with some tweaking for aesthetics) produces the plot shown in Figure 2.16. To the eye, this graph is identical to the one obtained using the shooting method, except near the end points where it is forced to zero more quickly by the walls of the fictitious infinite well. (We could avoid this problem by making the infinite well wider, but then we would also need to increase the maximum n value

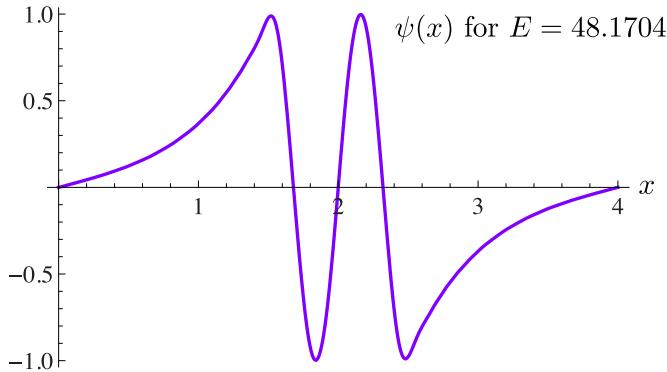


Figure 2.16: The eigenfunction whose vector components are plotted in Figure 2.15. Compare to the third plot in Figure 2.11.

to incorporate sufficiently short wavelengths.) Building the other three bound-state wavefunctions is no more difficult.

Finding the eigenvalues and eigenvectors of a matrix is often called *diagonalization*, because the eigenvectors can then be used as a new basis in which the matrix is diagonal, with entries equal to the eigenvalues. Solving the TISE is always equivalent to diagonalizing a matrix, and this method is practical whenever the infinite sums are well approximated by a reasonably small number of terms.

You may be wondering how the magical `Eigensystem` function actually works, finding all the eigenvalues and eigenvectors of a large matrix almost instantly. The details are proprietary and undoubtedly quite technical, but the basic idea of such routines is to successively rotate the basis vectors in various two-dimensional subspaces of the high-dimensional vector space, choosing the rotation angles to make the off-diagonal elements of the matrix smaller and smaller.⁶

Problem 2.14. Reproduce the calculation just outlined in the text, for a finite square well with $V_0 = 50$ in natural units, to obtain the list of eigenvalues and the plots shown in Figures 2.15 and 2.16. Also plot the eigenvector components and eigenfunctions corresponding to the three lowest eigenvalues. Explain why the odd-numbered eigenvector components are all zero for some eigenvectors, while the even-numbered components are all zero for others.

Problem 2.15. Explain the patterns in the Hamiltonian matrix shown in Figure 2.14. In particular: (a) Why is the matrix symmetric? (b) Why does the matrix have a “checkerboard” pattern, with half its elements equal to zero? (c) What is the pattern along the main diagonal? Can you find an approximate formula for H_{nn} that is valid for large n ? (d) Use a sketch of ϕ_1 , ϕ_3 , and V to explain why H_{13} is positive. (e) Use a

⁶See K. Randles, D. V. Schroeder, and B. R. Thomas, “Quantum matrix diagonalization visualized,” Am. J. Phys. **87**(11), 857–861 (2019), arXiv:1905.13269 [physics.ed-ph]. This paper includes Mathematica code to diagonalize the Hamiltonian matrix one step at a time, visualizing the results throughout the process. An interactive web app for the same purpose is posted at <http://physics.weber.edu/schroeder/software/QMatrixDiagVis.html>.

similar sketch to explain why H_{15} is negative. (f) Explain why elements far away from the main diagonal are very small.

Problem 2.16. For the quantum bouncer described in Problem 2.13, use the matrix diagonalization method to find the lowest 10 (or more) energy eigenvalues to reasonable accuracy. Use the infinite square well basis states, and think carefully about how wide you need to make the infinite well and how many basis states you need to include. How can you check the accuracy of your results? For the eigenvector that corresponds to the 10th eigenvalue, make a plot of the vector components (as a function of n) and also make a plot of the eigenfunction (as a function of x). Discuss a few of the notable features of these plots.

Problem 2.17. The time-consuming part of the matrix diagonalization code described in this section is not the `Eigensystem` function that diagonalizes the matrix, but rather the 2500 numerical integrations needed to construct the table of matrix elements. If you ever need to construct an even larger matrix in this way, you'll want to speed up the process. Fortunately, it isn't hard to do so. Using a product-to-sum trigonometric identity, show that all of the matrix elements can be written in terms of integrals of the form

$$\int_0^b V(x) \cos\left(\frac{N\pi x}{b}\right) dx,$$

where N is either $m + n$ or $m - n$. If the maximum value of n (or m) is n_{\max} , how many such integrals must you compute, in total? Rewrite the code to pre-compute these integrals and store their values in a one-dimensional list, then use this list to construct the two-dimensional table of matrix elements. Check that your code works correctly for $n_{\max} = 50$, then try it with larger values of n_{\max} . Comment on the speed improvement.

Problem 2.18. List some of the relative advantages and disadvantages of the matrix diagonalization method, compared to the shooting method of Section 2.2. Under what circumstances would you choose one or the other, and why?

2.4 The harmonic oscillator

It's time to study another example of solving the Schrödinger equation for a particular potential energy function $V(x)$. This example is the *harmonic oscillator*, for which $V(x)$ is quadratic:

$$V(x) = \frac{1}{2}k_s x^2 = \frac{1}{2}m\omega_c^2 x^2, \quad (2.20)$$

where k_s is some “spring constant” and $\omega_c = \sqrt{k_s/m}$ is the *classical oscillation frequency*, that is, the angular frequency of oscillation of a classical mass m attached to a rigid wall by a spring with constant k_s .

The quantum harmonic oscillator is important for two reasons.

First, it's a quantitatively useful model of almost anything small that wiggles, such as vibrating molecules and acoustic vibrations (“phonons”) in solids. The harmonic oscillator even serves as the basis for modeling the oscillations of the electromagnetic field and the other fundamental quantum fields of nature.

Second, the harmonic oscillator is another example of a one-dimensional quantum problem that can be solved exactly. Its detailed solutions will give us further insight into the behavior of quantum systems in general, helping us understand which features of the infinite square well are or aren't common to all trapped quantum particles. And although I won't do it in this book, you could also use the known harmonic oscillator energy eigenstates as an alternate "basis" for analyzing other quantum systems, as in the matrix diagonalization method described in the previous section.

Natural units

The full Hamiltonian for the harmonic oscillator is

$$\hat{H} = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + \frac{1}{2}m\omega_c^2 x^2. \quad (2.21)$$

Although this expression contains three constants (\hbar , m , and ω_c), they appear in only two different combinations. Without loss of generality we can choose units in which both of these combinations, \hbar^2/m and $m\omega_c^2$, are equal to 1; then the Hamiltonian becomes simply

$$\hat{H} = -\frac{1}{2} \frac{d^2}{dx^2} + \frac{1}{2}x^2. \quad (2.22)$$

To interpret these units, note that the combination

$$\sqrt{\frac{\hbar}{m\omega_c}} = \left(\frac{\hbar^2/m}{m\omega_c^2}\right)^{1/4} = 1 \quad (2.23)$$

has units of distance, while the combination

$$\hbar\omega_c = \sqrt{(\hbar^2/m)(m\omega_c^2)} = 1 \quad (2.24)$$

has units of energy. All distances and energies will therefore be measured in multiples of these quantities.

Problem 2.19. The fundamental oscillation frequency of a guitar's high-E string is 330 Hz. Suppose that we model this system as a simple harmonic oscillator with a mass of 0.25 g, oscillating with an amplitude of 1.0 mm. Find the amplitude and energy of this oscillator in natural (quantum) units.

Numerical solutions

It's completely straightforward to solve the time-independent Schrödinger equation, for the harmonic oscillator, using either of the numerical methods described in the previous two sections.

If you use the shooting method, you can exploit the fact that $V(x)$ is an even function and therefore assume that the solutions $\psi(x)$ are either even or odd, supplying boundary conditions $\psi(0) = 1$ and $\psi'(0) = 0$ for the even solutions and

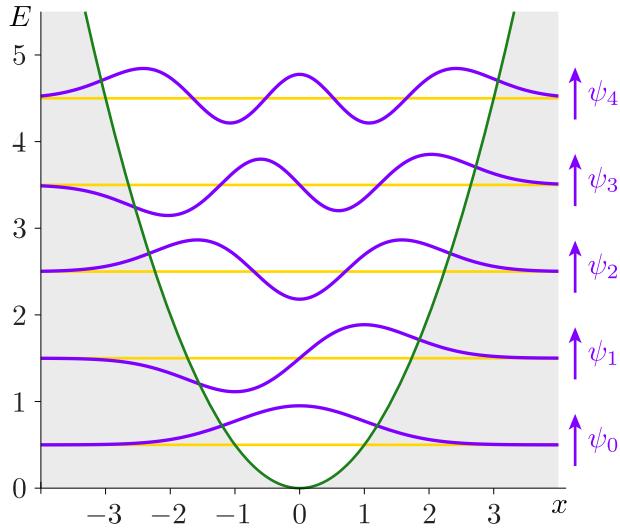


Figure 2.17: Combined plot of the harmonic oscillator potential energy (parabola), the first five energy levels (horizontal lines), and the five corresponding eigenfunctions. The energy and distance scales are labeled in natural units. Note that each wavefunction is plotted from a different horizontal axis, and that the vertical scale for the wavefunctions is unrelated to the energy scale. As expected, the bound-state wavefunctions oscillate within the classically allowed regions and decrease exponentially in the classically forbidden regions. Notice that the energy levels are evenly spaced, and that we number them starting from 0 instead of 1.

$\psi(0) = 0$ and $\psi'(0) = 1$ for the odd solutions. The wavefunction plots turn out as described in Section 2.1: oscillating in the classically allowed region, with increasing wavelength and amplitude as one moves outward, and exponentially decaying in the classically forbidden regions to either side. Unlike a generic $V(x)$, however, the harmonic oscillator potential yields an extremely simple set of energy eigenvalues: $1/2, 3/2, 5/2$, and so on, in natural units.

If instead you use the matrix diagonalization method, embedding the oscillator inside an infinite square well, you would center the oscillator inside the infinite well and choose the well width and number of basis functions to yield as many accurate eigenvalues and eigenfunctions as possible in whatever time you’re willing to wait for them. The answers for the eigenvalues are again $1/2, 3/2, 5/2$, and so on, and of course you can construct the associated wavefunctions out of the eigenvector components.

Figure 2.17 shows the lowest five energy levels superimposed on a graph of the potential energy, with the corresponding wavefunctions plotted using the energy levels as baselines, with same horizontal scale. Distances and energies are labeled in natural units. Notice that the energy rungs on this quantum ladder are *evenly spaced*, unlike the infinite square well for which they get farther apart as you go up. It’s conventional to number the harmonic oscillator energies and wavefunctions starting with 0 rather than 1, so the number indicates how many “units” of energy

the system has, relative to the ground state. This convention is a departure from the one that we use for essentially all other one-dimensional quantum systems.

The natural unit of energy is $\hbar\omega_c$, so in conventional units, the harmonic oscillator energy levels can be summarized in the formula

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega_c, \quad \text{for } n = 0, 1, 2, \dots \quad (2.25)$$

Problem 2.20. Use either the shooting method or the matrix diagonalization method (your choice!) to numerically find the five lowest energy eigenvalues and corresponding eigenfunctions of the harmonic oscillator.

Exact solutions

When you solve a problem numerically and get an unexpectedly simple answer, that's probably a clue that you could have solved the problem analytically. There are at least three approaches to analytically solving the TISE for the harmonic oscillator:

1. **Guess the answers.** Look at the ground-state wavefunction in Figure 2.17, and notice that it looks an awful lot like a Gaussian, e^{-ax^2} for some constant a . Plug this formula into the TISE and you'll see that it works as long as $a = 1/2$ and $E = 1/2$. One down. For the next solution, a look at the graph might lead you to the guess the formula xe^{-ax^2} , and if you plug this in you'll find that it works for the same $a = 1/2$, but with $E = 3/2$. That's two. At this point you might guess (correctly) that all the solutions are polynomial functions multiplied by the same Gaussian, $e^{-x^2/2}$. Each polynomial has only even or odd terms (to give the correct symmetry for the wavefunctions), and you can find the coefficients by requiring that the TISE be satisfied in each case. It gets laborious after the first few, but if you fiddle with the equations long enough you might notice some patterns and discover some general procedures for finding the coefficients.
2. **Power series.** This is the most traditional approach, and it's presented in all the traditional textbooks. By this method you can prove that the allowed energies are $n + 1/2$ for *any* nonnegative integer n , and that *all* of the associated wavefunctions are $e^{-x^2/2}$ times an n th-order polynomial. You end up with "recursion formulas" that let you calculate the coefficients of the polynomials in a straightforward way, but again it gets laborious to work out more than a handful of them.
3. **Ladder operators.** This is by far the most elegant method, although it's also the most abstract, and it's hard to see how anyone would have thought of it, and it's still laborious to work out more than a handful of the wavefunction formulas. I'll present this method in detail in Section 5.2, as we gear up to use a similar method to understand angular momentum in quantum mechanics. Feel free to look ahead if you're curious!

$$\begin{aligned}
H_0(x) &= 1 \\
H_1(x) &= 2x \\
H_2(x) &= 4x^2 - 2 \\
H_3(x) &= 8x^3 - 12x \\
H_4(x) &= 16x^4 - 48x^2 + 12 \\
H_5(x) &= 32x^5 - 160x^3 + 120x \\
H_6(x) &= 64x^6 - 480x^4 + 720x^2 - 120 \\
H_7(x) &= 128x^7 - 1344x^5 + 3360x^3 - 1680x
\end{aligned}$$

Table 2.1: The first eight Hermite polynomials. (You can make a table like this in Mathematica with just a single line: `TraditionalForm[TableForm[Table[HermiteH[n,x], {n,0,7}]]]`.)

Whatever the method used to obtain them, the harmonic oscillator energy eigenfunctions are n th-order polynomials multiplied by the Gaussian $e^{-x^2/2}$. There's no general formula for the polynomials themselves—just algorithms for calculating their coefficients. But they do have a name: they're called *Hermite polynomials*, abbreviated $H_n(x)$, and I've listed the first eight of them in Table 2.1. Notice the standard normalization convention⁷ of making the coefficient on x^n (the highest power that appears in H_n) equal to 2^n ; then all the other coefficients turn out to be integers. With the polynomials normalized in this way, $\psi_n(x)$ still has an n -dependent normalization coefficient that's not especially easy to work out, but at least it has a formula. The final formula for the normalized energy eigenfunctions is

$$\psi_n(x) = \frac{1}{\sqrt{2^n n! \sqrt{\pi}}} H_n(x) e^{-x^2/2}. \quad (2.26)$$

The Hermite polynomials are built into Mathematica as `HermiteH[n,x]`, so you can easily use that software to work with eigenfunctions up to $n = 100$ or more.

Like those of the infinite square well (and indeed, any other quantum system), the harmonic oscillator eigenfunctions are mutually orthogonal,

$$\int_{-\infty}^{\infty} \psi_m(x) \psi_n(x) dx = \delta_{mn}, \quad (2.27)$$

and they form a complete basis that you can use to expand any other wavefunction:

$$\psi(x) = \sum_{n=0}^{\infty} c_n \psi_n(x), \quad (2.28)$$

⁷I've read that mathematicians sometimes use a different normalization convention, so be careful if you ever look up Hermite polynomials in a source that's aimed at mathematicians rather than physicists.

for any wavefunction $\psi(x)$ and some set of complex coefficients $\{c_n\}$. I'll omit the proofs that go with these claims, but you can easily check some special cases (see Problems 2.21, 2.22, and 2.23).

Once you have the eigenfunctions and eigenvalues, and know that the eigenfunctions form an orthonormal basis, you can do all the usual things with them:

- Integrate $|\psi_n(x)|^2$ to calculate probabilities of finding the particle in various locations, when it's in a particular energy eigenfunction.
- Expand an arbitrary wavefunction in terms of energy eigenfunctions, to predict the probabilities of finding the particle with various energy values.
- Predict the time dependence of an arbitrary wavefunction, by expanding it in terms of energy eigenfunctions and inserting wiggle factors. (The Harmonic Oscillator web app at <http://physics.weber.edu/schroeder/software/HarmonicOscillator.html> can animate the behavior of any linear combination of ψ_0 through ψ_7 .)
- Use the harmonic oscillator eigenfunctions as basis functions for analyzing other one-dimensional quantum systems.

Problem 2.21. Using pencil and paper, check the analytical solutions to the quantum harmonic oscillator for $n = 0, 1$, and 2 . Also check the normalization and mutual orthogonality of these three eigenfunctions.

Problem 2.22. Set the Wave Builder app (<http://physics.weber.edu/schroeder/software/WaveBuilder.html>) to use the “Oscillator basis,” and play with it for a while to see what the basis functions look like, match a few target functions, and draw your own targets for the computer to solve. Then complete “Challenge round 4” (which uses the oscillator basis) and match all four of the target functions. Try not to use too many hints. To keep a record of your success, make a screen capture when you see the message “Target 4 of 4 matched!”

Problem 2.23. Use Mathematica to (a) print an explicit formula for the $n = 20$ energy eigenstate of a harmonic oscillator, $\psi_{20}(x)$; (b) plot a graph of $\psi_{20}(x)$ and a graph of the associated probability density; (c) check that $\psi_{20}(x)$ is normalized; (d) check that $\psi_{20}(x)$ is orthogonal to $\psi_{22}(x)$; and (e) check that $\psi_{20}(x)$ satisfies the time-independent Schrödinger equation. Be sure to add plenty of commentary to your work, to point out important and interesting details.

Problem 2.24. For a quantum harmonic oscillator in the ground state, calculate the probability of finding the particle outside the classically allowed region, if you were to measure its position. (Do the integral numerically.) Repeat the calculation for a few of the excited states, and discuss the results.

Problem 2.25. Consider a quantum harmonic oscillator that is initially in the following state (called a *coherent state*), built from a superposition of energy eigenstates:

$$\psi(t=0) = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} \psi_n(x), \quad (2.29)$$

where ψ_n is a normalized energy eigenstate and α is a dimensionless parameter that could be any complex number. Although it's interesting to explore the behavior of coherent states for different values of α , in this problem you may set $\alpha = 1$. You may also neglect the overall normalization factor, $e^{-|\alpha|^2/2}$.

- (a) Use a computer to make an accurate plot of this wavefunction. How many terms in the sum must you keep in order to get an accurate plot?
- (b) Write down a formula for the wavefunction at an arbitrary time t . Express the frequencies in the wiggle factors explicitly in terms of n .
- (c) Use a computer to make some plots to explore how this wavefunction changes with time. I suggest plotting the probability density, $|\psi(t)|^2$, although you may also wish to plot the real and imaginary parts of $\psi(t)$. Summarize the behavior in a sentence or two.

Problem 2.26. The vibrational frequency (ordinary frequency, not angular frequency) of a carbon dioxide molecule in its “flexing” mode is 2.0×10^{13} Hz. Assuming that these vibrations are harmonic, what are the energies (in electron-volts) of the three lowest vibrational energy levels for this mode? How much energy would a photon need to have, in order to excite the molecule up to the next-highest vibrational level when it is absorbed? What is the wavelength of such a photon? What part of the electromagnetic spectrum is it in?

2.5 Multiple wells

Very often, in molecules and solids, a quantum particle finds itself in a potential that has two or more local minima, separated by one or more barriers. For example, each electron in a hydrogen molecule (H_2) is attracted to both nuclei, so its potential is a “double well” with a barrier in between. Another example is the ammonia molecule, NH_3 , in which the three hydrogens form a triangle that defines a plane, and the nitrogen has two classically stable locations relative to them, one on either side of that plane (see Figure 2.18). Laboratory-fabricated double-well potentials are used to make light-emitting diodes, and are also being investigated for use as “gates” in quantum computing. Finally, a double-well potential gives us a starting point for understanding periodic potentials with more than two wells, such as the potential felt by a conduction electron in a crystalline metal.

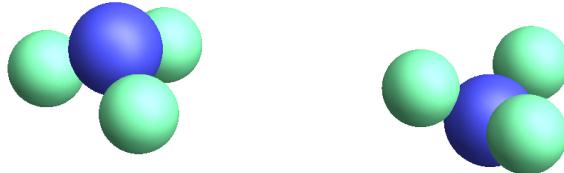


Figure 2.18: An ammonia molecule, NH_3 , has two classically stable configurations. If the three hydrogens define the horizontal plane, then the energy is minimized when the central nitrogen is either slightly above or below the plane, shown at left and right, respectively.

A double well

The most important features of double wells don't depend on the exact shape of the potential, so as a simple model I'll use a "rectangular" function as shown in Figure 2.19. Assuming that this potential has several bound states, with energies $E < V_0$, can you guess the shapes of the energy eigenfunctions and reach some qualitative conclusions about their energies? Take some time now to try to make some sketches and educated guesses (see also Problem 2.7).

To solve this problem quantitatively, it's most convenient to use the matrix diagonalization method. Here is some Mathematica code to define a rectangular potential with any number of wells:

```
nWells = 2; (* number of wells *)
bHeight = 50; (* barrier height *)
bWidth = 0.5; (* barrier width *)
b = nWells + (nWells+1)*bWidth; (* maximum x value *)
v[x_] := If[Mod[x-bWidth, 1+bWidth] < 1, 0, bHeight];
```

The first three lines define constants for the number of wells, the height of the barriers between them (V_0 in Figure 2.19), and the width of each barrier. The width a of each well is taken to be one unit of distance, and energies are measured in multiples of \hbar^2/ma^2 . The fourth line calculates the maximum value of x , which will be the width of the infinite well used to define the sine-wave basis functions; note that this width includes a "barrier" on either side, as well as the one(s) between the wells. The fifth line defines the potential energy function itself, using Mathematica's `Mod` function to make the pattern repeat over a distance of the well width (1) plus the barrier width.

The next steps are to define the infinite square well basis functions, calculate all the matrix elements, and diagonalize the Hamiltonian. The code to carry out these steps is essentially identical to that in Section 2.3, so I won't repeat it here. (I used

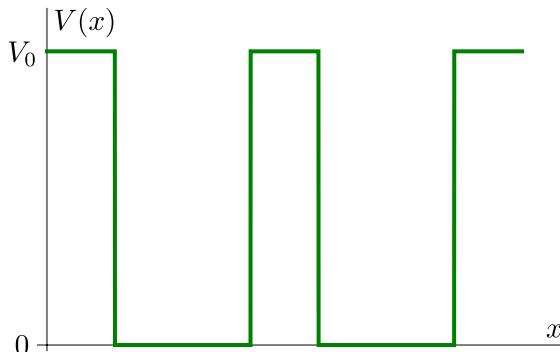


Figure 2.19: A rectangular double-well potential energy function. The calculation described in the text uses the width of a single well as the natural unit of distance, and takes the width of the central barrier to be $1/2$.

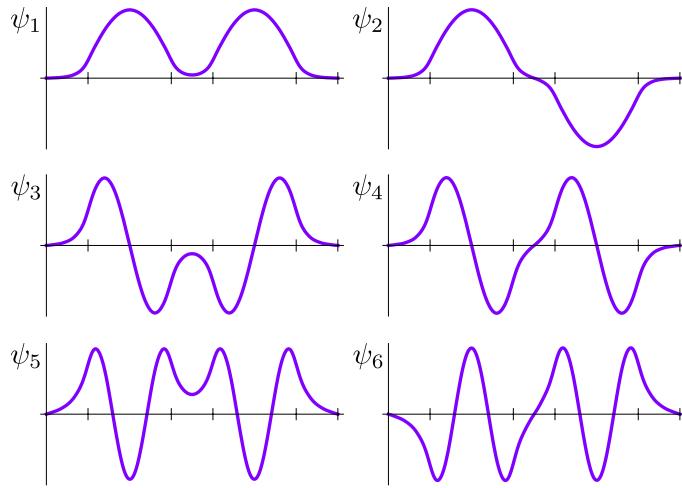


Figure 2.20: The lowest six energy eigenfunctions, numbered in order of increasing energy, for the rectangular double-well potential shown in Figure 2.19. The corresponding eigenvalues are listed in 2.30. Just as the energies are in closely spaced pairs, so also the eigenfunctions are in closely related pairs, differing mainly in the relative signs between the left and right halves and in how the two halves are joined.

the first 50 basis functions to build a 50×50 Hamiltonian matrix, but you can get pretty good results with only 25.) The result is that there are seven eigenvalues less than 50, corresponding to bound states, but I'll neglect the highest of these (47.9) because it's so weakly bound that the infinite square well embedding potential affects it significantly. The other six bound-state eigenvalues are approximately:

$$3.406, 3.424, 13.43, 13.53, 29.24, 29.71. \quad (2.30)$$

Notice that they come in closely spaced *pairs*, and that the values are quite close to those for a *single* square well with the same width and depth (3.414, 13.48, and 29.45, as we found in Section 2.2). The reason for this behavior becomes clearer if we look at the corresponding wavefunctions, shown in Figure 2.20. Each of these functions consists essentially of two side-by-side single-well eigenfunctions, modified slightly so they connect smoothly in the middle. Because the modifications are so small, the energies are almost the same as for the single well. Notice, though, that the total number of nodes for each eigenfunction is what you would expect for the system as a whole: no nodes for the ground state, one for the first excited state, and so on.

You could, of course, obtain these same eigenfunctions, along with the associated eigenvalues, using the shooting method.

In defiance of our classical intuition, every energy eigenstate puts *half* of the wavefunction on *each* side of the double-well potential. The two lowest states are what we sometimes call “cat states,” with two separated peaks. In principle, we could push the peaks as far apart as we like, by increasing the width of the central

potential barrier.

Once you have the eigenvalues and eigenfunctions, you can use them in the usual ways to examine probabilities and time evolution. As you can see, for each of these eigenfunctions the particle is equally likely to be found in either of the two wells. Wavefunctions that put the particle entirely in one well or the other do exist, but they are not energy eigenfunctions and therefore they change over time in a nontrivial way.

Problem 2.27. Use the matrix diagonalization method to reproduce the list of energy eigenvalues (2.30) for the rectangular double-well potential described above, and the corresponding eigenfunction plots shown in Figure 2.20. Then use the shooting method to check these results.

Problem 2.28. A simple formula for a smoothly varying double-well potential is $V(x) = x^4 - 10x^2$, where both x and V are understood to be measured in some natural system of units where $\hbar^2/m = 1$.

- (a) Plot this function and locate its two minima.
- (b) In preparation for finding the energy eigenvalues and eigenfunctions using the matrix diagonalization method, shift the potential energy function upward so its minimum value is zero, and shift it by 4 units in the $+x$ direction so that wavefunctions with energies less than the central barrier height will be entirely at positive x values (aside from negligible exponential tails). Write down the formula for this shifted potential function.
- (c) Use the matrix diagonalization method, with an infinite square well basis with width 8, to find the energy eigenvalues and eigenfunctions. Be sure to use enough basis functions to give a reasonably accurate picture of all the energy levels that are below the barrier height. Plot an energy level diagram for these levels. (Note that if the list of eigenvalues is called `eValues`, you can plot the diagram in Mathematica with an instruction like `Plot[eValues, {x, 0, 8}]`, and it's then easy to combine this plot with a plot of the potential energy function.) Discuss the patterns that occur in the set of energy eigenvalues.
- (d) Plot the eigenfunctions corresponding to the lowest six (or more) energy eigenvalues, and discuss the features of these eigenfunctions. (Note that you can use the Mathematica `Table` function to make several plots all at once.)

The two-state approximation

Imagine a double-well system in an environment where there is plenty of available energy (from collisions or radiation) to cause transitions from the ground state ψ_1 up to the first excited state ψ_2 , but not enough energy to excite any of the remaining, much higher-energy, states. Then only ψ_1 and ψ_2 will be reasonably probable, so we can neglect the higher-energy eigenstates to an excellent approximation.

Under these circumstances we call the double well a *two-state system*. However, a better name might be two-*eigenstate* system, or two-*basis*-state system, because the actual number of available states is infinite, consisting of all the linear superpositions

$$\psi = c_1\psi_1 + c_2\psi_2, \quad (2.31)$$

where c_1 and c_2 are complex numbers. In order for ψ to be normalized, these complex numbers must satisfy

$$1 = \langle \psi | \psi \rangle = \langle c_1 \psi_1 + c_2 \psi_2 | c_1 \psi_1 + c_2 \psi_2 \rangle = |c_1|^2 + |c_2|^2, \quad (2.32)$$

where in the last step I've used most of the properties of the inner product listed in Problem 1.28, along with the fact that ψ_1 and ψ_2 are orthonormal (see Problem 2.30). Two superpositions of particular interest are

$$\psi_L = \frac{1}{\sqrt{2}}(\psi_1 + \psi_2) \quad \text{and} \quad \psi_R = \frac{1}{\sqrt{2}}(\psi_1 - \psi_2), \quad (2.33)$$

which put the particle almost entirely inside the left well and almost entirely inside the right well, respectively (as you can see by mentally adding and subtracting the graphs in Figure 2.20).

One way to visualize these states is to plot the coefficients c_1 and c_2 in a two-dimensional plane, as shown in Figure 2.21. We can then think of each state as a unit-length vector in this abstract two-dimensional space, with c_1 and c_2 as its components. Notice that the vectors ψ_1 and ψ_2 are orthogonal to each other, as are ψ_L and ψ_R , while the latter pair make 45-degree angles with the former pair. Multiplying a state vector (wavefunction) by -1 has no effect on any physical predictions, so we can think of any one of these vectors as lying half-way between the two vectors of the other pair. Unit-length vectors lying at other angles on this diagram would represent states with other real values of c_1 and c_2 .

The diagram has a severe limitation, though, because it cannot show any of the equally valid states for which c_1 or c_2 is complex. For instance, there are two other states that make 45-degree angles with *all four* of the states shown in Figure 2.21:

$$\psi_+ = \frac{1}{\sqrt{2}}(\psi_1 + i\psi_2) \quad \text{and} \quad \psi_- = \frac{1}{\sqrt{2}}(\psi_1 - i\psi_2). \quad (2.34)$$

You're probably wondering what I even mean by the “angle” between two vectors when one of them has complex components. In this case it may be simpler to just say that ψ_+ , for example, can be written as a mixture of equal parts of the orthogonal vectors ψ_1 and ψ_2 , *or* as a mixture of equal parts of the orthogonal vectors ψ_L and ψ_R (so ψ_+ is another “cat state,” with half of the particle on each side of the double well). In general, though, we can define the angle between any two normalized state vectors as the inverse-cosine of the magnitude of their inner product (in analogy to the angle between ordinary vectors in physical space, aside from the fact that here the inner product can be complex). See Problem 2.31 for details.

Now let's think about how these states change over time. States ψ_1 and ψ_2 are energy eigenstates, so their time dependence is a simple wiggle factor: $e^{-iE_1 t/\hbar}$ or $e^{-iE_2 t/\hbar}$, where E_1 and E_2 are the respective energies. An overall phase factor of the form $e^{i\phi}$ has no effect on any physical predictions, so in a sense this time dependence is trivial and uninteresting—hence the name *stationary* states.

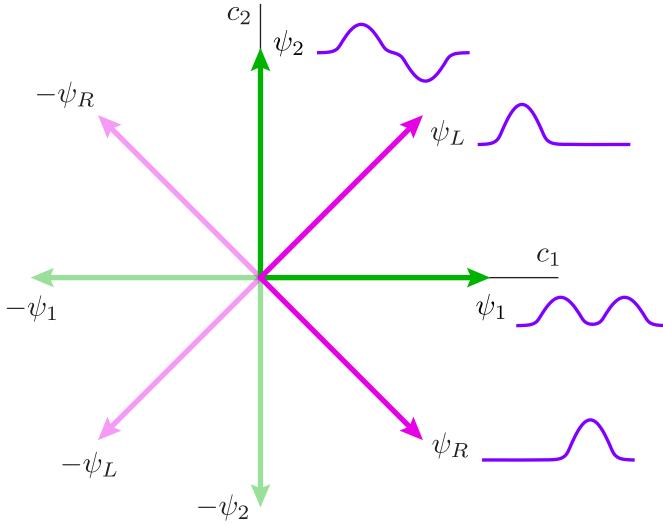


Figure 2.21: A visualization of the states ψ_1 , ψ_2 , ψ_L , and ψ_R , with the component c_1 plotted horizontally and c_2 plotted vertically. The wavefunction shapes are drawn near each of these four state vectors. An overall factor of -1 has no physical effect on a wavefunction, so the negated vectors represent the same four physical states. A normalized state vector has unit length on this diagram, but the diagram cannot show states whose components are complex.

But suppose instead that this system starts out in the state ψ_L , with the particle (almost) entirely in the left-hand potential well. Then, to predict the state at a later time, we expand this initial state in terms of stationary states and insert a different wiggle factor into each term (as in equation 1.24):

$$\psi(t) = \frac{1}{\sqrt{2}}(\psi_1 e^{-iE_1 t/\hbar} + \psi_2 e^{-iE_2 t/\hbar}). \quad (2.35)$$

To interpret this expression, it's convenient to factor out the first wiggle factor:

$$\psi(t) = e^{-iE_1 t/\hbar} \frac{1}{\sqrt{2}}(\psi_1 + \psi_2 e^{-i\Delta E t/\hbar}), \quad (2.36)$$

where $\Delta E = E_2 - E_1$ is the small difference in energy between ψ_1 and ψ_2 . The overall factor $e^{-iE_1 t/\hbar}$ now has no effect on physical predictions, so we can focus on the time dependence of the ψ_2 term, which oscillates clockwise in the complex plane at angular frequency $\Delta E/\hbar$. When $\Delta E t/\hbar = \pi/2$ this complex exponential equals $-i$, so the state is physically equivalent to ψ_- , with half of the probability on each side of the double well. After another such interval the complex exponential equals -1 , so the state is physically equivalent to ψ_R , with the particle (almost) entirely in the right-hand well. The oscillation then continues through a state that is equivalent to ψ_+ , and finally back to (a state equivalent to) ψ_L . Figure 2.22 shows the full oscillation cycle.

This back-and-forth motion between the two potential wells, right through the classically forbidden barrier, is an example of quantum-mechanical *tunneling*. The

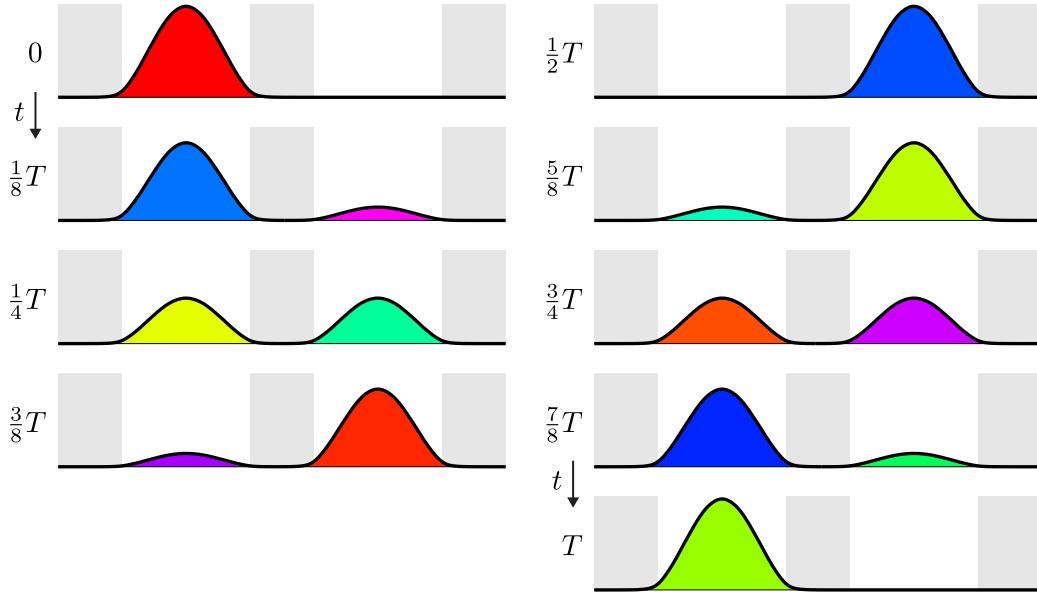


Figure 2.22: Sequence of frames showing a full oscillation period T of the wavefunction of equations 2.35 and 2.36, using the specific energies and eigenstates for a rectangular double well with depth 50 and a barrier width of 0.5 in natural units. In each frame the probability density $|\psi(x)|^2$ is plotted vertically, with color hues indicating the complex phase of the wavefunction. Note that these phases are oscillating very rapidly due to the overall wiggle factor in equation 2.36; the more important feature is that the phase in the right-hand well is a quarter-cycle ahead of that in the left-hand well during the first half-period of oscillation (left column), and a quarter-cycle behind during the second half-period (right column). Gray shading indicates the classically forbidden regions.

time required for the particle to tunnel through the barrier is half the oscillation period T :

$$\text{Tunneling time} = \frac{T}{2} = \frac{\pi\hbar}{\Delta E}. \quad (2.37)$$

Making the barrier between the wells higher and/or wider would dramatically reduce the energy splitting ΔE , and thus dramatically increase the tunneling time.

The problems below explore the features of this two-state system in more depth, and we will encounter other examples of two-state systems later in this book. Two-state systems are important not only because they are common in nature, but also because they illustrate many features of quantum behavior in the simplest possible way. Moreover, two-state systems are the foundation of the fast-growing field of quantum information science. In analogy to the way we can interpret a classical two-state system as a “binary digit” or *bit* (0 or 1), we often refer to a quantum two-state system as a “quantum bit” or *qubit*. In a classical bit, the two “states” are really just two arbitrary ranges of a property that actually has many possible values—such as the amount of charge on a transistor or the degree of magnetization of a small piece of material. In a qubit, on the other hand, the binary nature of

the system is intrinsic—but we can also put the system into arbitrary superposition states, of the form of equation 2.31, that are effectively mixtures of 0 and 1.

Problem 2.29. The difference in energy between the two lowest levels of an ammonia molecule is only 0.99×10^{-4} eV. If the molecule starts out in a mixture of these two levels that puts the nitrogen all on one side of the three hydrogens, at what frequency will it oscillate back and forth between the two minima of the potential energy function? How much time does it therefore take to “tunnel” through the potential barrier? If, in the oscillation process, it radiates an electromagnetic wave with the oscillation frequency, what is the wavelength? What is the energy of a photon with this wavelength? (This is the mechanism of the *ammonia maser*, an early predecessor of the laser.)

Problem 2.30. Fill in the missing steps in equation 2.32, carefully justifying every step.

Problem 2.31. Show that for two arbitrary states in the two-dimensional space,

$$\psi_c = c_1\psi_1 + c_2\psi_2 \quad \text{and} \quad \psi_d = d_1\psi_1 + d_2\psi_2, \quad (2.38)$$

the inner product is

$$\langle \psi_c | \psi_d \rangle = c_1^*d_1 + c_2^*d_2 = \begin{pmatrix} c_1^* & c_2^* \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}, \quad (2.39)$$

where the last expression is a convenient matrix version of the formula. Then use this formula to show that the wavefunctions ψ_+ and ψ_- defined in equation 2.34 are normalized and are orthogonal to each other. Finally, show that either ψ_+ or ψ_- makes a 45-degree angle with all four of the state vectors shown in Figure 2.21, where the angle θ between two normalized state vectors ψ_α and ψ_β is defined by $\cos \theta = |\langle \psi_\alpha | \psi_\beta \rangle|$.

Problem 2.32. Consider an arbitrary linear operator \hat{A} that acts in a two-dimensional state space, converting an arbitrary state ψ_c into some other state ψ_d :

$$\psi_d = \hat{A}\psi_c. \quad (2.40)$$

Show that if we write the states ψ_c and ψ_d as column vectors, using the ψ_1 - ψ_2 basis as in the previous problem, then \hat{A} becomes a 2×2 matrix whose elements are $A_{mn} = \langle \psi_m | \hat{A} \psi_n \rangle$, as in equation 2.15. (Hint: Start with equation 2.40 and expand each of the state vectors as in equation 2.38. Then take the inner product with ψ_1 or ψ_2 from the left.) What is the 2×2 matrix for the Hamiltonian operator? Can you find a matrix whose eigenvectors are ψ_L and ψ_R ? Can you find a matrix whose eigenvectors are ψ_+ and ψ_- ?

Problem 2.33. Although the matrix elements of the previous two problems are written using ψ_1 and ψ_2 as basis vectors, we can choose any orthogonal pair as our basis. Suppose, then, that we use ψ_L and ψ_R as basis vectors. What, then, are the components of ψ_1 , ψ_2 , ψ_+ , and ψ_- ? What is the Hamiltonian matrix? Starting with this rewritten Hamiltonian matrix, solve the characteristic equation to show that it has the same eigenvalues as before, then show that the eigenvectors are also as expected.

More than two wells

I've written the code on page 70 so you can add more potential wells simply by increasing the value of `nWells`. Can you guess what happens to the pattern in the energy eigenvalues when there are three, four, or more wells? Can you guess what the wavefunctions look like? (The detail that's probably hardest to guess is what happens when the number of wells is odd and the wavefunction needs to be an odd function, with a node in the middle. But this is just a detail.)

Rather than showing you more results here, let me just invite you to think about what should happen with more than two wells, and then to use Mathematica to check your guesses and derive some numbers and plots. You might also enjoy exploring the results from two "canned" simulations: Quantum Bound States⁸ from the PhET project at the University of Colorado, and the 1-d Quantum States Applet⁹ by Paul Falstad. Falstad's applet can even animate the time dependence of superposition states and tunneling through barriers.

When there are *many* potential wells in a periodic structure, the allowed energies come not in pairs or small groups but in quasi-continuous *bands*, each consisting of a very large number of closely spaced energies. The bands are separated by wide *gaps* in which no energies are allowed. This behavior underlies the dramatic differences between insulators, conductors, and semiconductors, as you can explore further in a course on solid state physics.

Problem 2.34. Use the matrix diagonalization method to find the bound-state energies and eigenfunctions for a four-well rectangular potential energy function with a well width of 1, a barrier width of 0.5, and barrier height of 50 in natural units. I suggest using the code on page 70 for defining the potential energy function (setting `nWells` to 4). Plot enough of the eigenfunctions to give a general idea of the patterns, and discuss the results in enough detail to show that you understand the patterns.

⁸<https://phet.colorado.edu/en/simulation/legacy/bound-states>

⁹<http://www.falstad.com/qm1d/>

Chapter 3

Unbound states in one dimension

The previous chapter was all about bound states, in which a particle is trapped inside a potential energy well and its energy can have only certain quantized values. Our main goal was to find those values, along with the associated energy eigenfunctions.

We now turn our attention to particles that are *not* trapped, so they have enough energy that the classically allowed region extends forever in one or both directions. In this case, as I explained at the end of Section 2.1, the energy eigenfunctions will go through an infinite number of spatial oscillations, whose wavelengths can vary continuously, so the energies are not quantized. We can pick any energy we like (at least above a certain threshold), and solve the TISE to find a corresponding eigenfunction.

But the energy eigenfunctions aren't necessarily the functions we want. For an untrapped particle these eigenfunctions extend over an infinite amount of space. They're technically not normalizable and, more importantly, they're unphysical: real particles are always at least *somewhat* localized. Also, in many cases, real untrapped particles are moving in approximately well-defined directions, at approximately well-defined speeds. So in Section 3.2 we will study wavefunctions that are approximately localized in both position and momentum. First, though, we need to take a careful look at how momentum factors into an arbitrary wavefunction.

3.1 Momentum space

I hope you're now comfortable with the idea of expanding an arbitrary wavefunction $\psi(x)$ in terms of basis functions:

$$\psi(x) = \sum_{n=1}^{\infty} c_n \phi_n(x), \quad (3.1)$$

where the ϕ_n can be any complete—but discrete—basis set. So far in this book, the ϕ_n have been energy eigenfunctions, associated with a $V(x)$ (like the infinite square well or the harmonic oscillator) that rises to infinity on both sides, trapping any particle and quantizing all energies.

What I want to do now is write the corresponding expansion in terms of *momentum* eigenfunctions, $e^{ipx/\hbar}$, where p can be any real number. (These functions are also energy eigenfunctions for a *free* particle, when $V(x) = 0$ everywhere.) If this collection of basis functions were discrete, we could express any other function $\psi(x)$ in terms of them by using a sum:

$$\psi(x) = \sum_p c_p e^{ipx/\hbar} \quad (\text{wrong}). \quad (3.2)$$

But because p is continuous, we need to replace the sum by an integral:

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \tilde{\psi}(p) e^{ipx/\hbar} dp. \quad (3.3)$$

In this formula I've also switched from the letter c to the symbol $\tilde{\psi}$ (which I pronounce “sigh twiddle”), and factored out a constant from $\tilde{\psi}$ for a reason that I'll explain in a moment. The important thing to notice is that instead of a discrete set of coefficients $\{c_n\}$, we now have a continuous function, $\tilde{\psi}(p)$, that encodes how much of each basis function $e^{ipx/\hbar}$ is incorporated into the wavefunction $\psi(x)$. This function, $\tilde{\psi}(p)$, has a name: the *momentum-space wavefunction*. For every (position-space) wavefunction $\psi(x)$, there must be a corresponding momentum-space wavefunction $\tilde{\psi}(p)$.

And how do we find $\tilde{\psi}(p)$ for a given $\psi(x)$? Again, think about the discrete case. There, to find a particular c_m , we would use *Fourier's trick*: multiply equation 3.1 by $\phi_m^*(x)$, integrate over x , and exploit the orthonormality of the basis functions to kill off every term in the sum except the one we want (see equation 1.38). Let's try the same trick here. Multiplying both sides of equation 3.3 by $e^{-ip'x/\hbar}$ (where p' is some arbitrary momentum value that in general is different from p) and integrating, we have

$$\begin{aligned} \int_{-\infty}^{\infty} dx \psi(x) e^{-ip'x/\hbar} &= \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp \tilde{\psi}(p) e^{ipx/\hbar} e^{-ip'x/\hbar} \\ &= \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} dp \tilde{\psi}(p) \int_{-\infty}^{\infty} dx e^{i(p-p')x/\hbar}, \end{aligned} \quad (3.4)$$

where in the second line I've interchanged the order of the integrals and moved $\tilde{\psi}(p)$ outside the x integral. (I've also written dx and dp next to their corresponding integral signs, so you can tell which integral goes with which variable.) The x integral is now an *inner product* of the two basis functions $e^{ipx/\hbar}$ and $e^{ip'x/\hbar}$. If these basis functions were a discrete and orthonormal set, this inner product would equal a Kronecker delta $\delta_{pp'}$. Here, where the set of basis functions is continuous,

we instead get a *Dirac* delta function, times a normalization constant:

$$\int_{-\infty}^{\infty} e^{i(p-p')x/\hbar} dx = 2\pi\hbar \delta(p - p'). \quad (3.5)$$

This lovely mathematical result may be unfamiliar to you, so think about it a moment: When $p \neq p'$, the integrand on the left oscillates and averages to zero, while the delta function on the right indeed equals zero. On the other hand, when $p = p'$, the integrand on the left is 1, so there's no cancelation and we get infinity—just as the delta function says. The factor of \hbar on the right-hand side comes from a change of variables; the more generic version of the formula would be simply

$$\int_{-\infty}^{\infty} e^{ikx} dx = 2\pi \delta(k). \quad (3.6)$$

The factor of 2π is not easy to guess, but I hope you'll accept it by the time we're through.

Plugging the orthonormality relation 3.5 into equation 3.4 and using the delta function to carry out the p integral, we obtain

$$\int_{-\infty}^{\infty} dx \psi(x) e^{-ip'x/\hbar} = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} dp \tilde{\psi}(p) 2\pi\hbar \delta(p - p') = \sqrt{2\pi\hbar} \tilde{\psi}(p'). \quad (3.7)$$

We can now rename $p' \rightarrow p$ to obtain our desired result,

$$\tilde{\psi}(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \psi(x) e^{-ipx/\hbar} dx. \quad (3.8)$$

You may recognize equations 3.3 and 3.8 as the formulas for a *Fourier transform* and *inverse Fourier transform*, respectively. (The factors of \hbar don't normally appear in math courses, but that's just a matter of using $p = \hbar k$ as our variable instead of k .) Mathematicians can give you a rigorous proof, without using delta functions, that either of these equations implies the other; that fact is called *Plancherel's theorem*. If nothing else, you should consult such a proof to see where the factor of 2π comes from. But as a physicist, I find it more intuitive to think in terms of “Fourier's trick,” which projects out the desired “component” of the “vector” $\psi(x)$, and to invoke the delta-function identity 3.5 (or 3.6) at the appropriate point in the calculation.

For a *free* particle, the momentum eigenfunctions $e^{ipx/\hbar}$ are also energy eigenfunctions, so equation 3.3 is just the expansion we need in order to insert wiggle factors and obtain the wavefunction as a function of time:

$$\psi(x, t) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \tilde{\psi}(p) e^{ipx/\hbar} e^{-iEt/\hbar} dp \quad (\text{free particle}). \quad (3.9)$$

Here E is a continuous variable that depends on p ; if the particle is nonrelativistic, then $E = p^2/2m$. So for a free particle, you can calculate the time dependence of any

initial wavefunction $\psi(x, 0)$ by first using equation 3.8 to find the momentum-space wavefunction and then plugging that into equation 3.9.

Unfortunately, carrying out Fourier-transform integrals with pencil and paper is feasible only for the simplest of wavefunctions. Professionals almost always rely on computers for this task, and fortunately, many computer software packages include powerful routines for “fast Fourier transforms” of numerical data. In Mathematica, the applicable functions are called `Fourier` and `InverseFourier`. Learning to use these functions would take a bit of time, however, so I’ve decided not to incorporate such calculations into this book (at least for now).

Problem 3.1. Derive equation 3.5 from equation 3.6.

Problem 3.2. Suppose $\psi(x) = e^{ip_0x/\hbar}$ (where p_0 is a constant). What is $\tilde{\psi}(p)$? Sketch both functions and discuss their physical interpretations.

Problem 3.3. Suppose $\psi(x) = \delta(x - x_0)$ (where x_0 is a constant). What is $\tilde{\psi}(p)$? Sketch both functions and discuss their physical interpretations.

Problem 3.4. Equation 3.3 expresses an arbitrary wavefunction $\psi(x)$ as a linear combination of the momentum eigenfunctions $e^{ipx/\hbar}$. What is the analogous equation that expresses $\psi(x)$ as a linear combination of the *position* eigenfunctions $\delta(x - x')$? (Here I’m using x' to label *which* position eigenfunction we want, just as the variable p in $e^{ipx/\hbar}$ labels *which* momentum eigenfunction we want.)

Probabilities and averages

Once you have the momentum-space wavefunction $\tilde{\psi}(p)$, you can use it to calculate momentum probabilities just as you would use $\psi(x)$ to calculate position probabilities:

$$\left(\begin{array}{l} \text{Probability of measuring} \\ \text{momentum between } p_a \text{ and } p_b \end{array} \right) = \int_{p_a}^{p_b} |\tilde{\psi}(p)|^2 dp. \quad (3.10)$$

This is the *Born rule for momentum*, analogous to equation 1.6 for position and equation 1.60 for energy. As Figure 3.1 shows, the probability of measuring a particle’s momentum to be within a certain range is the area under the graph of $|\tilde{\psi}(p)|^2$. Of course, this formula doesn’t make sense unless $\tilde{\psi}(p)$ is normalized, so that the integral from $-\infty$ to ∞ equals 1. But as you might guess, this will always be the case if you calculate $\tilde{\psi}(p)$ from a $\psi(x)$ that is itself normalized.

If you want to know the *average* momentum, you can get it from $\tilde{\psi}(p)$ in a way that’s exactly analogous to calculating the average position from $\psi(x)$ (see equation 1.8):

$$\langle p \rangle = \int_{-\infty}^{\infty} p |\tilde{\psi}(p)|^2 dp. \quad (3.11)$$

And, naturally, a similar formula works for any *function* of p . However, if average values are all you want, then there’s actually no need to calculate $\tilde{\psi}(p)$ at all. For example, you can get $\langle p \rangle$ directly from $\psi(x)$ by evaluating the integral

$$\langle p \rangle = \int_{-\infty}^{\infty} \psi^*(x) \left(-i\hbar \frac{d}{dx} \right) \psi(x) dx. \quad (3.12)$$

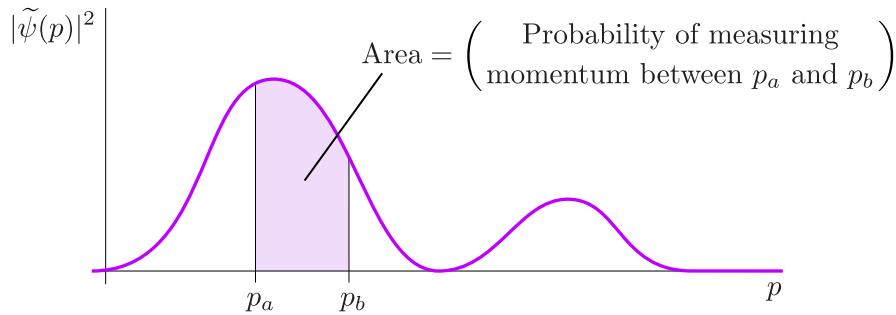


Figure 3.1: When you measure a particle’s momentum, the probability of obtaining a result within a given range equals the area under the graph of the square-modulus of the momentum-space wavefunction.

Notice that the quantity inside the big parentheses is the momentum operator. If instead it were the position operator (x), then this would just be the familiar formula for $\langle x \rangle$. To derive equation 3.12, just insert expansion 3.3 for both ψ and ψ^* on the right-hand side, being careful to use a different p variable in each. Then note that the x integral gives a delta function, which you can use to carry out one of the p integrals, leaving you with an expression that is precisely identical to the right-hand side of equation 3.11.

In fact, you can calculate the average value of *any* observable quantity using an expression of the form of equation 3.12, replacing the momentum operator with the operator of your choice. For powers of the momentum such as $\langle p^2 \rangle$, the proof is essentially the same as the proof of equation 3.12. For operators that involve both x and p , such as the Hamiltonian operator when $V(x)$ is nonzero, the proof is analogous but rests on the assumption that the operator has a complete set of mutually orthogonal eigenfunctions (as in Problem 1.39).

Problem 3.5. Derive equation 3.12, following the steps described in the text below it.

3.2 Wavepackets

We’re now ready to investigate the properties of wavefunctions that are *somewhat* localized in space, with *approximately* well-defined momentum, but with neither position nor momentum (nor any other observable quantity) *precisely* defined. Such a “compromise” wavefunction is called a *wavepacket*. It’s the closest we can come, in quantum mechanics, to making a particle behave classically.

The simplest version of a wavepacket consists of a momentum eigenfunction multiplied by an “envelope” function that’s large in some central region and dies out smoothly to either side. For mathematical convenience it’s easiest to take the envelope to be a Gaussian bell curve, so I’ll express the wavepacket as follows:

$$\psi(x) = A e^{-(x-x_0)^2/a^2} e^{ip_0 x/\hbar}. \quad (3.13)$$

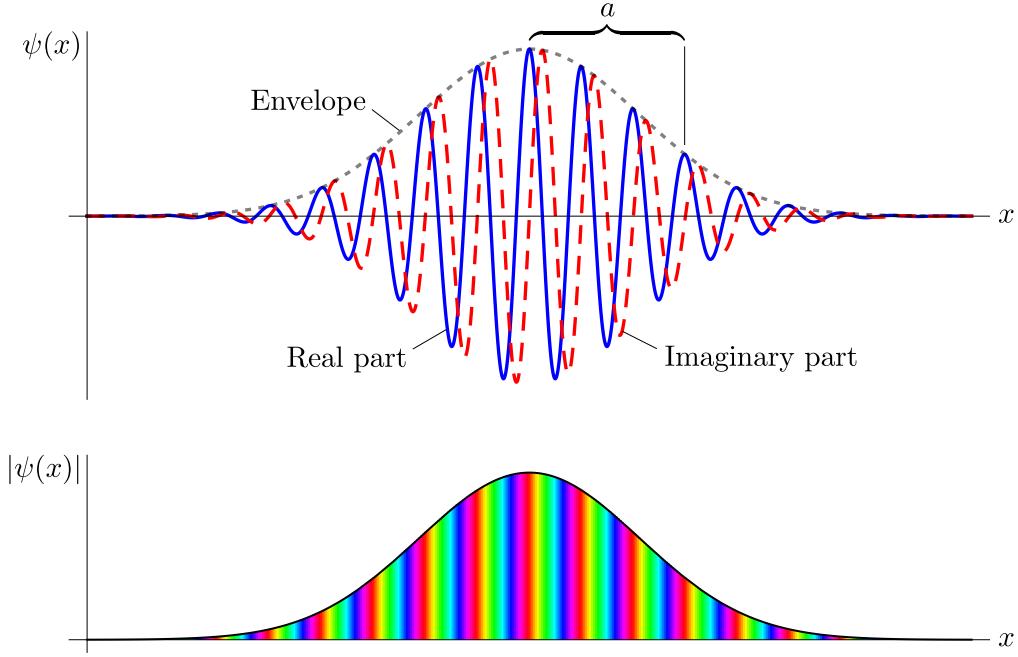


Figure 3.2: A Gaussian wavepacket (equation 3.13), plotted as real and imaginary parts (above) and as magnitude with phase as color hue (below). This particular wavepacket has $p_0 > 0$ and $a = 3\hbar/p_0$ (that is, positive nominal momentum, and width parameter equal to three wavelengths). The height at the peak is A and the horizontal displacement of the peak from the origin is x_0 .

Here I'm using the symbol p_0 for this wavefunction's nominal momentum value. The parameter a has units of length and is a rough measure of the width of the packet, while x_0 is the location of the wavepacket's peak amplitude. You can express the normalization constant A in terms of a (see Problem 3.8), but often it's handier to just write it as A .

What does this wavefunction look like? Figure 3.2 shows an example, but the details depend on the sign of p_0 and on how a compares to the oscillation wavelength, \hbar/p_0 .

But there's a reason I called p_0 the “nominal” momentum value. A wavepacket is *not* a momentum eigenfunction but rather a *mixture* of momentum eigenfunctions with a whole *range* of p values. To quantify exactly what mixture it is, we need to compute the momentum-space wavefunction, by taking the inverse Fourier transform (equation 3.8):

$$\tilde{\psi}(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} (A e^{-x^2/a^2} e^{ip_0 x/\hbar}) e^{-ipx/\hbar} dx. \quad (3.14)$$

Here I've set $x_0 = 0$ for simplicity. Be careful not to confuse p_0 , a constant parameter that defines our particular wavepacket, with p , the variable on which the momentum-space wavefunction depends. To carry out the integral, combine the three exponents,

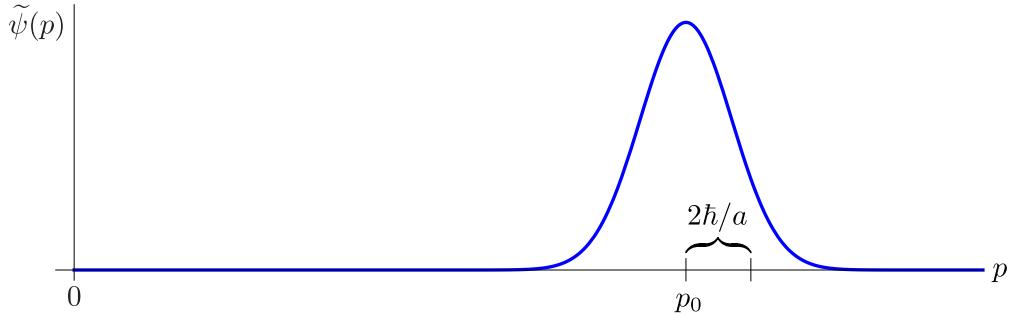


Figure 3.3: The momentum-space wavefunction corresponding to the wavepacket plotted in Figure 3.2, with $a = 3\hbar/p_0$. As in equation 3.16, the wavepacket is centered at $x_0 = 0$ for simplicity; this simplification makes the momentum-space wavefunction pure real.

complete the square, and use the basic Gaussian integration formula

$$\int_{-\infty}^{\infty} e^{-\alpha x^2} dx = \sqrt{\frac{\pi}{\alpha}}. \quad (3.15)$$

The result is

$$\tilde{\psi}(p) = B e^{-(a(p-p_0)/2\hbar)^2}, \quad (3.16)$$

where B is a normalization constant that you can express in terms of a if you wish. (Remarkably, $\tilde{\psi}(p)$ is pure real—but this doesn’t happen when x_0 is nonzero, as you’ll see in Problem 3.10.)

So the momentum-space wavefunction is also a Gaussian bell curve. It’s centered on p_0 as we would expect, with a width in momentum space of approximately $2\hbar/a$ (see Figure 3.3). This result means that if you were to measure the momentum of this Gaussian wavepacket, the most likely outcome would be p_0 but you would be reasonably likely to obtain any value in the range $p_0 \pm 2\hbar/a$. Notice that the width of $\tilde{\psi}(p)$ is inversely proportional to the width of the (position-space) wavefunction $\psi(x)$. So the more we try to localize a particle in space (by reducing the value of a), the more uncertainty we introduce into its momentum—and vice-versa. More precisely, if you calculate the standard deviations σ_x and σ_p for a Gaussian wavepacket you can show (see Problem 3.8) that their product is a constant, independent of a :

$$\sigma_x \sigma_p = \frac{\hbar}{2} \quad \text{for a Gaussian wavepacket.} \quad (3.17)$$

This is a special case of the famous *Heisenberg uncertainty principle*, which says more generally that there is *no* wavefunction for which this product of standard deviations is less than $\hbar/2$.

Problem 3.6. Use a computer to plot Gaussian wavepackets with $a = h/p_0$ and with $a = 10h/p_0$, both for $p_0 > 0$, and with $a = 3h/|p_0|$ for $p_0 < 0$. In each case, either plot the real and imaginary parts of ψ on the same graph, or plot $|\psi|$ and fill the area underneath with color hues determined by the complex phase, as in Figure 3.2.

Problem 3.7. To understand why a wavepacket of width a has a momentum-space width on the order of \hbar/a , imagine adding just two momentum eigenfunctions, $e^{ip_0x/\hbar}$ and $e^{ip_1x/\hbar}$ (rather than the actual continuous infinity of eigenfunctions that build $\psi(x)$ according to equation 3.3). These two waves interfere constructively at $x = 0$ to form the peak of this “packet,” and we want them to interfere destructively at around $x = a$ to make the “packet” die out over that distance. What, then, should be the difference between p_1 and p_0 ? Discuss your result briefly.

Problem 3.8. (a) Find the normalization constant A in equation 3.13, and use it to find the normalization constant B in equation 3.16. (b) Calculate $\langle x^2 \rangle$ and hence σ_x for a Gaussian wavepacket, in terms of the width parameter a used in equation 3.13. To derive the needed integration formula, differentiate both sides of equation 3.15 with respect to α . (c) Calculate σ_p for a Gaussian wavepacket, in terms of \hbar and a , and then check the uncertainty principle, equation 3.17.

Problem 3.9. Calculate the standard deviation σ_E of the energy, for a free Gaussian wavepacket with nominal momentum p_0 and width parameter a .

Problem 3.10. Calculate the momentum-space wavefunction of a Gaussian wavepacket (equation 3.13) *without* setting $x_0 = 0$ as I did in the text. How does the resulting function $\tilde{\psi}(p)$ change, as you vary x_0 ? Include a sketch (or a computer-drawn plot) as part of your answer.

Time evolution

Now suppose that this Gaussian wavepacket describes a nonrelativistic particle at time zero, and that the particle is *free*, with $V = 0$ everywhere. How will the wavepacket evolve as time passes? We could calculate the answer using equation 3.9, but it’s more enlightening to think about the features we expect the answer to have. (Another advantage of the reasoning that follows is that there is no need to assume that the shape of the envelope is precisely Gaussian, as long as we interpret a as a rough measure of its width.)

First of all, we expect the individual “ripples” within the wavepacket to move, on average, at the nominal phase velocity,

$$v_{\text{phase}} = \frac{\omega}{k} = \frac{E/\hbar}{p_0/\hbar} = \frac{p_0^2/2m}{p_0} = \frac{p_0}{2m}. \quad (3.18)$$

This is only *half* the nominal velocity of the particle, as we already saw in equation 1.20. But the packet itself moves not at the phase velocity but at the so-called *group velocity*, which (according to the standard formula for any type of wave) is

$$v_{\text{group}} = \frac{d\omega}{dk} = \frac{d}{dk} \left(\frac{\hbar k^2}{2m} \right)_{k=p_0/\hbar} = \frac{p_0}{m}. \quad (3.19)$$

So the packet as a whole moves at the nominal particle velocity, twice as fast as the ripples within it.

But the packet is actually a mixture of waves moving at a whole range of velocities, so as it moves along, some of these components move faster and others

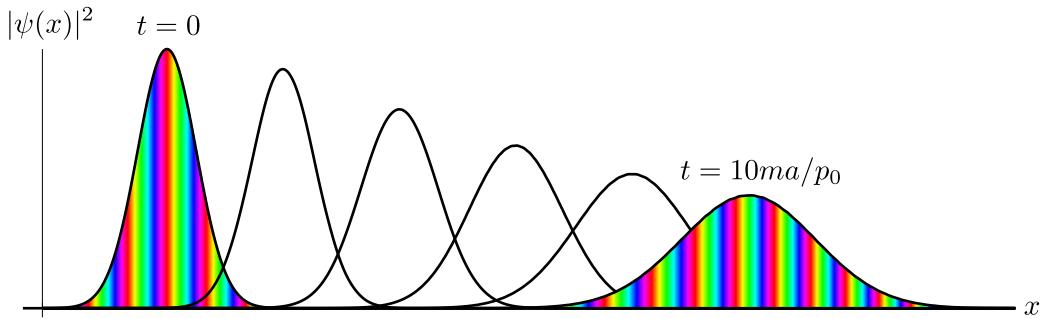


Figure 3.4: Time evolution of a free Gaussian wavepacket. The initial packet (left) has the form of equation 3.13 with $a = 1.5\hbar/p_0$. As time passes the packet moves to the right and spreads; the time interval between successive curves in this composite image is $2ma/p_0$. Look closely at the phase oscillations in the final packet (right) and notice that the wavelength is shorter in the leading portion than in the trailing portion, because the shorter-wavelength components move faster.

move more slowly. Soon this “dispersion” causes the packet to spread out, just as a large group of hikers naturally spreads out along the trail, with the faster ones in the lead and the slower ones behind. Although the wavelength of the oscillations within the packet is initially uniform, it will not remain uniform: as the packet moves and spreads, its leading edge will contain shorter-wavelength (higher-velocity) oscillations, while its trailing edge will contain longer-wavelength (lower-velocity) oscillations. Meanwhile the peak amplitude of the packet will decrease, to conserve probability as the width increases. Figure 3.4 illustrates all these effects for one particular choice of wavepacket parameters; you can explore a wide variety of parameters using the Quantum Wavepackets simulation at <http://physics.weber.edu/schroeder/software/Wavepackets.html>.

To quantify the spreading effect, recall from equation 3.16 that the faster components of the wavepacket have momentum values that are larger than average by about $2\hbar/a$. This means their velocity is larger than average by about $2\hbar/ma$, so at time t they will have traveled farther than average by about $2\hbar t/ma$. Once this quantity exceeds the initial width a , the spreading effect will dominate the wavepacket’s width:

$$\text{width} \sim \frac{2\hbar t}{ma} \quad \text{after } t \sim \frac{ma^2}{2\hbar}. \quad (3.20)$$

It’s interesting to compare the time scale for the spreading to become significant, $ma^2/2\hbar$, to the time scale for the wavepacket to move a distance equal to its width, ma/p_0 . These two time scales are roughly the same when the nominal wavelength, \hbar/p_0 , is equal to the packet width. When the wavelength is much less than the packet width, so that many waves fit inside, the packet moves many times its own width before it spreads significantly. On the other hand, if the wavelength is much longer than the packet width, the packet’s spreading will be more noticeable than its motion.

Naturally, all of these physical effects are predicted by the more formal mathematics. Problem 3.14 explores the spreading effect, and then Problem 3.15 adds in the motion. I've relegated these calculations to the problems because the calculations are somewhat laborious, and it's all too easy to get so bogged down in the algebra that the physics gets lost. But please do work the problems when you're ready, so you can see how the full formula for $\psi(x, t)$ backs up everything I've just described.

Alternatively, we can also explore wavepacket spreading numerically, as described in the following section (and as implemented in the Quantum Wavepackets simulation).

But whatever our tools—physical reasoning, analytical calculation, or numerical simulation—the behavior of a Gaussian wavepacket is a critically important example of quantum behavior. This example quantifies the tradeoff between states with well-defined position and states with well-defined momentum, showing exactly how close a particle can come to having both at the same time, and even showing how the passage of time amplifies the initial quantum fuzziness.

Problem 3.11. For the wavepacket shown in Figure 3.4, compare the characteristic spreading time, $ma^2/2\hbar$, to the time between the successive curves. Discuss.

Problem 3.12. Imagine that the wavepacket shown in Figure 3.4 has been undisturbed at times *before* $t = 0$, so the same principles (and equations) that determine its properties at positive times also apply at negative times. What would the packet have looked like at negative times? Would it have been narrower than at $t = 0$ or wider? Would the wavelength of its phase oscillations have been uniform, or larger in the leading portion, or larger in the trailing portion? Explain.

Problem 3.13. Suppose that we model a pitched baseball (mass 145 g, speed 40 m/s) as a Gaussian wavepacket with an initial width of 10^{-10} m (a little less than the width of an atom). Should the batter worry about this wavepacket spreading before the ball reaches the plate? Explain carefully, and also discuss some other possible values for the initial width.

Problem 3.14. Consider a *stationary* free Gaussian wavepacket, with $p_0 = 0$, centered at the origin (so $x_0 = 0$). Carry out the integral in equation 3.9 to find an explicit formula for $\psi(x, t)$. (Hints: Combine all the exponents and complete the square. Use the abbreviation $\tau = 2\hbar t/m a^2$; this is a dimensionless time parameter, equal to the time in units of the characteristic spreading time of equation 3.20. I also found it helpful to temporarily use the abbreviation α for the coefficient of the p^2 term in the exponent.) Your answer will initially be in a form that is hard to interpret, because it has factors of i in awkward places. To fix the exponent, multiply and divide by a complex expression to make the denominator real, thus moving the i to the numerator; then split the exponential into a real envelope and a pure phase factor. To fix the square root, write its argument in the form $b e^{i\phi}$ (where b and ϕ are real), then note that $\sqrt{e^{i\phi}} = e^{i\phi/2}$. Finally, having put your answer into a reasonable form, explain its properties in some detail.

Problem 3.15. Repeat Problem 3.14 for a moving Gaussian wavepacket, with $p_0 \neq 0$.

3.3 Numerically solving the TDSE

Even for the idealized example of a free Gaussian wavepacket, finding an explicit formula for the wavefunction at a future time, $\psi(x, t)$, is laborious. If the initial state is more complicated, or (more importantly) if the particle interacts with a nontrivial potential energy function $V(x)$, then chances are no such formula even exists. Fortunately, though, it's not hard to use a computer to predict the time evolution of an arbitrary initial wavefunction, subject to an arbitrary potential energy, by direct numerical integration of the time-dependent Schrödinger equation (TDSE).

First let me write the TDSE in natural units, setting $\hbar = m = 1$:

$$i \frac{\partial \psi}{\partial t} = -\frac{1}{2} \frac{\partial^2 \psi}{\partial x^2} + V(x)\psi. \quad (3.21)$$

We now discretize x into a list of evenly spaced values, using the second centered difference approximation (equation 2.6) for the second spatial derivative:

$$i \frac{\partial \psi}{\partial t} = -\frac{1}{2} \frac{\psi(x+dx) + \psi(x-dx) - 2\psi(x)}{(dx)^2} + V(x)\psi(x), \quad (3.22)$$

where dx is now the (non-infinitesimal) spacing between adjacent x values. To simplify this equation I will use units in which $dx = 1$. Then our unit of momentum will be \hbar/dx , our unit of velocity will be $\hbar/(m dx)$, and our unit of time will be $m(dx)^2/\hbar$. With this simplification and a bit of rearrangement, the TDSE becomes

$$i \frac{\partial \psi}{\partial t} = -\frac{1}{2} [\psi(x+1) + \psi(x-1)] + [1 + V(x)]\psi(x). \quad (3.23)$$

We still need to discretize time, replacing $\partial\psi/\partial t$ with a discrete approximation. The easiest way to do this would be to write

$$\frac{\partial \psi}{\partial t} \approx \frac{\psi(x, t+dt) - \psi(x, t)}{dt}, \quad (3.24)$$

with the understanding that each ψ on the right-hand side of equation 3.23 is evaluated at t (rather than at $t+dt$ or some other time). But this approximation is inaccurate because it has a forward bias, taking the future time $t+dt$ into account but not the past time $t-dt$ (see Figure 3.5). It is much more accurate to use the centered-difference approximation

$$\frac{\partial \psi}{\partial t} \approx \frac{\psi(x, t+dt) - \psi(x, t-dt)}{2 dt}, \quad (3.25)$$

which treats the near future and recent past symmetrically. Inserting this approximation into equation 3.23 and solving for $\psi(x, t+dt)$, we obtain our fully discretized version of the TDSE:

$$\psi(x, t+dt) = \psi(x, t-dt) + i dt [\psi(x+1, t) + \psi(x-1, t) - 2(1 + V(x))\psi(x, t)]. \quad (3.26)$$

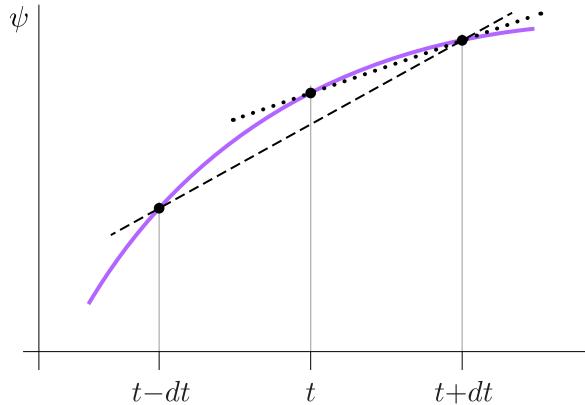


Figure 3.5: The slope of a centered secant line (dashed, equation 3.25) is a better approximation to the derivative of the actual function (solid curve) at t than is the slope of a forward-biased secant line (dotted, equation 3.24).

Equation 3.26 is the kernel of what I'll call the *centered-difference method* of solving the TDSE. The procedure is to loop over all x points on the discretized lattice, using this equation to calculate $\psi(x, t+dt)$, and then step forward in time and repeat the process to calculate ψ at the next time increment, and the next, and so on. There are just a few more details to fill in:

- **Choosing the time step.** The calculation will go faster if we use a larger value of dt , but it turns out that if we make dt too large, the process becomes unstable: small spurious inaccuracies grow exponentially with time until ψ becomes huge and nonsensical. The largest dt you can get away with depends on $V(x)$.¹ In our units this largest value is always less than 0.5, and in practice $dt = 0.45$ works well most of the time.
- **Boundary conditions.** Equation 3.26 won't work at either end of the spatial interval of our calculation, because it would then refer to x values that are beyond the ends by dx . The usual approach is therefore to set $\psi = 0$ at each end at all times, effectively embedding our system inside an infinite square well. We then use equation 3.26 only at the interior points.
- **Getting started.** Of course we must provide an initial wavefunction, $\psi(x, 0)$. But in order to use equation 3.26 for the first time, we also need to know $\psi(x, -dt)$, one step *before* time zero. A simple solution is to calculate it using

¹See P. B. Visscher, “A fast explicit algorithm for the time-dependent Schrödinger equation,” *Computers in Physics* **5** (6), 596–598 (1991); or H. Gould, J. Tobochnik, and W. Christian, *An Introduction to Computer Simulations Methods*, third edition (Pearson, San Francisco, 2007, electronic version available at <http://www.opensourcephysics.org/items/detail.cfm?ID=7375>), equation 16.34. I actually learned the algorithm described in this section from Thomas A. Moore (personal communication) in about 1982, but Visscher gets credit for the careful stability analysis.

equation 3.24 with a negative value of dt . Even though this approximation isn't very accurate, the harm is minimal because we're using it only once.

Now let me show an implementation of the centered-difference method using Mathematica. I'll start by defining the size of the spatial lattice, creating a table of potential energy values, and plotting the potential function:

```
xMax = 400;
v = Table[If[x < 200, 0, 0.06*((x - 200)/200)], {x, 1, xMax}];
ListPlot[v, Joined -> True]
```

Because Mathematica list indices start at 1 rather than 0, it's easiest to let x range from 1 to x_{Max} . My arbitrary potential energy function is 0 in the left half of the interval and linear in the right half, with a slope chosen so that my wavefunction (see below) will have enough energy to penetrate about half-way into the linear region.

Next I'll initialize the time variable, the time step, and the wavefunction itself:

```
t = 0;
dt = 0.45;
x0 = 100;
p0 = 0.25;
a = 30;
psi = Table[Exp[-(x-x0)^2/a^2] Exp[I p0 x], {x, 1, xMax}];
psiLast = psiNext = Table[0, {x, 1, xMax}];
For[x = 2, x < xMax, x++,
  psiLast[[x]] = psi[[x]]
  - (I*dt/2)*(psi[[x+1]] + psi[[x-1]] - 2(1+v[[x]])psi[[x]]]);
```

For the initial wavefunction I'm using a Gaussian of the form of equation 3.13, centered in the region where the potential energy is zero, with a positive momentum so it will move toward the linear-potential region. I chose the value of p_0 so the nominal wavelength, $2\pi/p_0$, would be significantly larger than the lattice spacing, $dx = 1$. After initializing the “current” wavefunction ψ , I've initialized the wavefunctions at the previous and next time steps (ψ_{Last} and ψ_{Next}) to zero—although here it would actually suffice just to initialize the endpoints. The `For` loop at the end of this code block uses equation 3.24 to properly initialize ψ_{Last} .

The initializations are now complete, but before we start calculating forward in time we need a plot to show what this wavefunction looks like. This part of the code will be pretty arcane, but I hope you'll like the result:

```
Dynamic[psiInterp = Interpolation[psi];
Plot[Abs[psiInterp[x]]^2, {x, 1, xMax},
  PlotRange -> {0, 1},
  Filling -> Axis,
```

```

ColorFunction -> Function[x, Hue[Arg[psiInterp[x]]/(2 Pi)]],
ColorFunctionScaling -> False,
PlotPoints -> 500,
Epilog -> Inset[Style[StringJoin["t = ",
ToString[Round[t]]], 20], {10, 0.9}, {Left, Top}]] )

```

The outer `Dynamic` function causes the plot to update itself whenever `psi` changes (as it soon will). Before creating the plot we create what Mathematica calls an interpolating function, so it can evaluate the function at any `x` value it wants (not just at integer values). The function we're plotting is the probability density, $|\psi|^2$, but we fill the area beneath it using a `ColorFunction` that assigns the hue based on the wavefunction's phase ("Arg"). The `Epilog` adds text to show the current value of the time variable, rounded to the nearest integer, in 20-point type, near the upper-left corner of the plot.

After all these preliminaries, the simulation code itself is refreshingly concise:

```

While[t < 2500,
  For[x = 2, x < xMax, x++,
    psiNext[[x]] = psiLast[[x]] +
      I*dt*(psi[[x+1]] + psi[[x-1]] - 2*(1+v[[x]])psi[[x]]);
    psiLast = psi;
    psi = psiNext;
    t += dt];

```

The `For` loop uses equation 3.26 to calculate the new wavefunction at each (interior) point. The rest of the code gets us ready for the next time step, by copying the current wavefunction into the space for the previous one, copying the next wavefunction into the space for the current one, and incrementing the time variable. All of this code repeats until the time variable reaches 2500, long enough for the wavepacket to move several hundred distance units at its nominal initial velocity of 0.25.

Figure 3.6 shows some of the results of running this code—but *please* don't take my word for it! Type the code in yourself and watch the full animation sequence. Check that your results agree with mine, then try changing the various parameters to better understand the effect of each. It's also worthwhile to try to explain what's happening qualitatively, as the wavepacket moves and spreads and slows and reverses. But I've chosen this particular example not because it has any fundamental importance, but rather because the results are sufficiently complicated that I could never have obtained them with pencil and paper. You now know how to predict the time evolution of *any* initial wavefunction, subject to *any* potential energy function!

The following problems ask you to modify the code to explore a variety of further scenarios. The next section takes a closer look at what happens when a quantum particle encounters a potential energy barrier.

Problem 3.16. The code on page 90 gives the initial wavepacket a nominal momentum of 0.25 in natural units. What is its nominal kinetic energy? If this were a *classical*

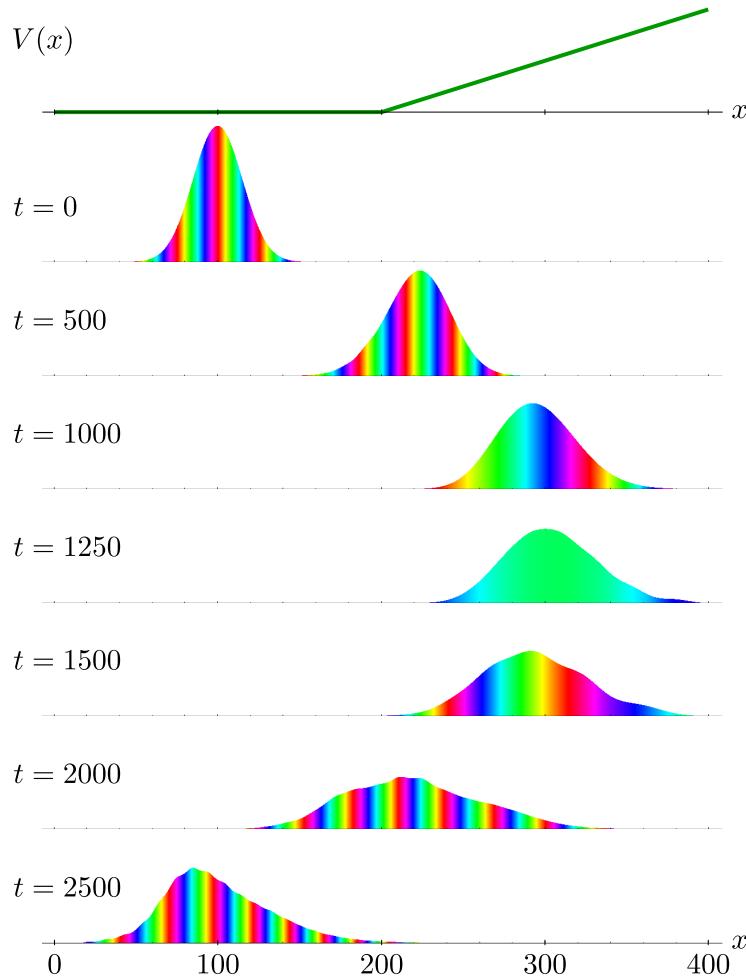


Figure 3.6: Selected animation frames generated by the numerical TDSE code presented in this section. The initial Gaussian wavepacket moves to the right and encounters a linear “ramp” potential beginning at $x = 200$. The wavepacket then slows down, stops (at about $t = 1250$), and reverses direction, as its shape spreads and becomes irregular.

particle with the same initial energy, how far would it penetrate into the region with the linear potential (that is, where would its turning point be)?

Problem 3.17. Explain the pattern of color hues (indicating the wavefunction phase) in each of the frames in Figure 3.6, relating the color pattern to the packet’s direction of motion, speed, and so on.

Problem 3.18. Run the code given in the text, and check that you can reproduce the results shown in Figure 3.6. Then run the same code using larger and smaller values of the time step dt , and describe the results.

Problem 3.19. Adapt the code in this section to reproduce the results shown in Figure 3.4, saving or printing a few representative animation frames as in the figure.

Explain carefully how you chose the initial wavepacket parameters, and discuss the wavepacket's behavior in some quantitative detail. Check that the wavelength within your final wavepacket varies from front to back as the figure shows.

Problem 3.20. Adapt the code in this section to study a spreading stationary Gaussian wavepacket, with $V = 0$, as in Problem 3.14. Start with a fairly narrow wavepacket that is centered in the simulation space, let it spread to fill most of the space, then save the resulting image and describe its qualitative features in some detail. Then repeat the simulation starting with a *pair* of narrow stationary wavepackets, separated by a gap that is a few times wider than their widths. Again save an image showing the results, and explain what is happening qualitatively.

Problem 3.21. Adapt the code in this section to study a moving Gaussian wavepacket with $V = 0$, and run it long enough to observe what happens when the wavepacket reaches the boundary of the simulation region. Describe and explain the results in as much detail as you can.

3.4 Scattering from a barrier

The example of the previous section (see Figure 3.6) becomes more interesting if we modify the potential function slightly, so it doesn't keep rising indefinitely as x

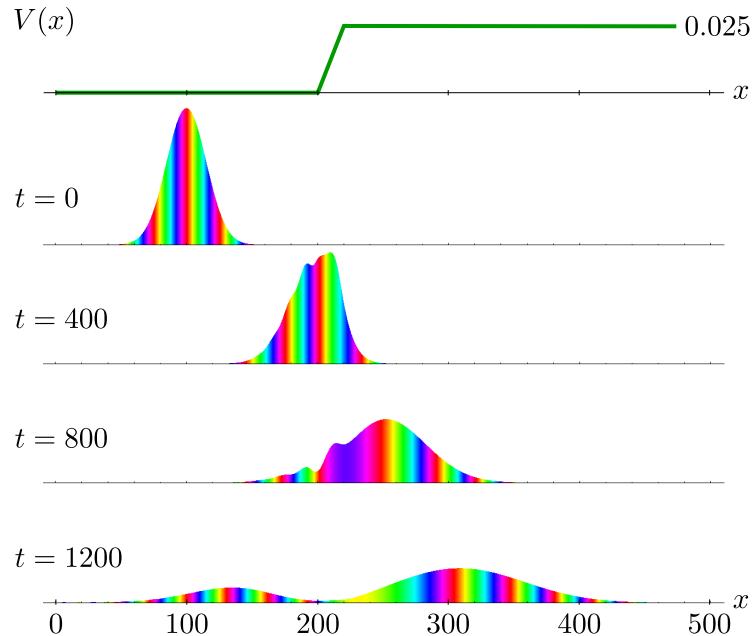


Figure 3.7: When a wavepacket hits a step potential, it can split into reflected and transmitted pieces. I produced these animation frames using the code in the previous section, with a modified potential function (and a larger value of `xMax`). The width of the sloped step is 20 units.

increases. Figure 3.7 shows a wavepacket hitting a sloped step potential that rises linearly to a maximum value of 0.025 in natural units (compared to 0.060 at the extreme right in Figure 3.6); I've also made the slope several times steeper than before, so the sloping portion has a width of just 20 units. The particle experiences a constant leftward force within this 20-unit interval, but no force on either side.

As before, this interaction causes intricate changes in the shape of the wavepacket. But now there is also a major qualitative difference in the outcome: the wavepacket *splits* into a reflected packet, which comes back to the left as before, and a transmitted packet, which continues to the right. We call this a *scattering* process, because the incoming particle can be sent outward in more than one direction. After it interacts with the potential barrier, the particle is in a so-called “cat state,” with two widely separated peaks.

We can calculate the relative probabilities of reflection and transmission by simply integrating the final $|\psi|^2$ over appropriate ranges. When I do this I get a reflection probability of approximately 24% and a transmission probability of approximately 76%. Throughout this section I'll refer to these probabilities as R and T , respectively.

To understand why the reflection and transmission probabilities have these values, recall from page 90 that the nominal momentum p_0 of the initial wavepacket is 0.25 in natural units. This means that its nominal kinetic energy, $\frac{1}{2}p_0^2$, is 0.03125, which is somewhat greater than the step height, 0.025. A *classical* particle with this momentum would have enough energy to overcome the leftward force and continue to the right (at reduced speed); its transmission probability would be 100%.

The main reason why the quantum wavepacket's transmission probability is less than 100% is because p_0 is merely its *nominal* momentum. As we saw in Section 3.2, the wavepacket is actually a mixture of a range of momentum states, with a width in momentum space that is inversely proportional to the width in position space. Here the position-space width is $a = 30$ in natural units, so the momentum-space width is on the order of 0.03. If we subtract this from the nominal momentum of 0.25 and then calculate the corresponding kinetic energy, we get only 0.024,

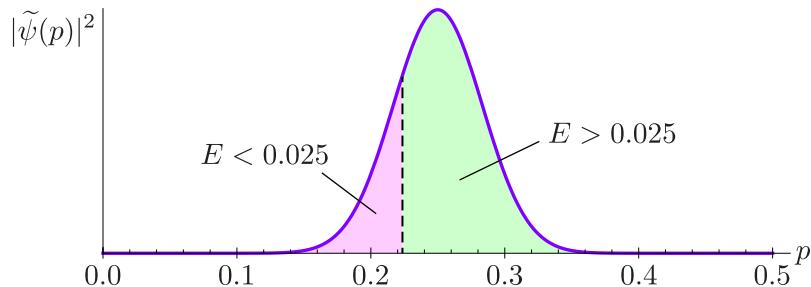


Figure 3.8: The momentum-space probability distribution of the initial wavepacket shown in Figure 3.7, calculated from equation 3.16. The portion to the right of the dashed line has enough energy to penetrate the potential barrier, but the portion to the left does not.

which is less than the step height. Figure 3.8 shows the initial wavepacket's entire momentum-space probability distribution, highlighting the momentum ranges that do and don't have enough energy to penetrate the force barrier. The barrier acts as a *filter*, allowing the high-energy (high-frequency) components of the packet through, and reflecting the low-energy (low-frequency) components. When I integrate $|\tilde{\psi}(p)|^2$ over the appropriate ranges I find that 78.6% of the wavepacket has more than the critical energy, while 21.4% has less.

Notice that this filtering effect doesn't quite account for all of the 24% reflection probability. That's because even when a quantum particle's energy is above the threshold for transmission, there is still a chance that it will reflect. In this respect, a quantum particle behaves less like a classical particle and more like a classical wave. For example, you already know that even when a light wave of a particular frequency is perfectly capable of penetrating a sheet of glass, a portion of it will instead reflect.

Problem 3.22. Modify the code in Section 3.3 to reproduce the scenario of Figure 3.7. Be sure to calculate the reflection and transmission probabilities. How would you expect the results to change if you increase the initial wavepacket width (a) from 30 to 50? After writing down your expectation, try it and see.

Problem 3.23. Draw a reasonably accurate sketch of the momentum-space probability distribution for the *final* ($t = 1200$) state shown in Figure 3.7. Show the probability distribution for the initial state (Figure 3.8) on the same sketch, for comparison.

Scattering probabilities for definite energies

To better understand the reflection/transmission process, it would be nice if we could focus on just one energy at a time, rather than simulating a narrow wavepacket that mixes a broad range of energies. Moreover, real-world wavepackets are often much wider in space, and therefore have much more precisely defined energies, than the narrow packet used in Figure 3.7.

But our brute-force method of numerically integrating the time-dependent Schrödinger equation is not well suited to studying wide wavepackets with precise energies. To simulate a wider wavepacket we would have to make room for it, extending our grid of x values in both directions, and we would have to wait longer for the entire wavepacket to interact with the barrier. The simulation would therefore take much longer to run, due to both the longer integration time and the larger number of x values to calculate during each time step.

Fortunately, there is a much better way. To determine the behavior of a definite-energy particle interacting with our potential barrier, we can simply solve the time-independent Schrödinger equation within the region of interest, using appropriate boundary conditions. The solution will represent a steady-state scenario or, equivalently, the limit of infinitely wide incoming and outgoing wavepackets. In the regions with constant V , to the left and right of where the force acts on the particle, the solution will consist of appropriate momentum eigenfunctions to represent

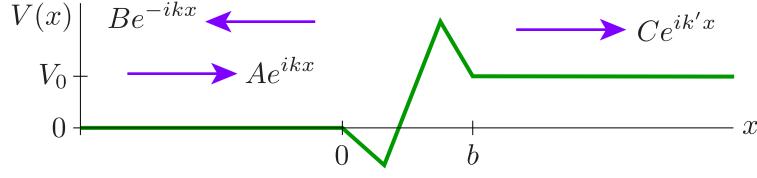


Figure 3.9: In a generic one-dimensional scattering situation, the potential energy takes an arbitrary form within a limited region (here 0 to b) but is constant on either side. The steady-state scattering solution to the TISE consists of incident plus reflected momentum eigenfunctions to the left of the region of the force, and a transmitted momentum eigenfunction to the right.

the incoming, reflected, and transmitted waves. The relative amplitudes of these eigenfunctions will determine the reflection and transmission probabilities.

To make all this more explicit, let me introduce some assumptions and some notation (see Figure 3.9). Let the nontrivial part of the potential energy function have a width of b , extending from $x = 0$ to $x = b$. I'll assume that the potential energy is constant (so there is no force) to either side of this region, and I'll put its (arbitrary) zero level on the left:

$$V(x) = \begin{cases} 0 & \text{for } x < 0, \\ V_0 & \text{for } x > b. \end{cases} \quad (3.27)$$

Then if the incident wave comes in from the left with energy E , we can write the wavefunction in each of the force-free regions as

$$\psi(x) = \begin{cases} Ae^{ikx} + Be^{-ikx} & \text{for } x < 0, \\ Ce^{ik'x} & \text{for } x > b, \end{cases} \quad (3.28)$$

where the angular wavenumbers are

$$k = \sqrt{2E} \quad \text{and} \quad k' = \sqrt{2(E - V_0)}, \quad (3.29)$$

in units where the particle mass and \hbar are equal to 1. The coefficients A , B , and C tell us the relative amplitudes of the incident, reflected, and transmitted waves, and can all be complex, to allow for phase shifts. But all that matters is the *relative* amplitudes and phases, so without loss of generality I will set C to a complex number of unit magnitude (and arbitrary phase, as you'll see below). Solving the TISE throughout the entire region will then tell us the corresponding values of A and B .

In principle, for a sufficiently simple potential function, we could try to solve the TISE analytically. But it's much easier, and much more generally doable, to solve it numerically. We can use essentially the same “shooting method” as in Section 2.2, except this time we're not trying to find an unknown energy eigenvalue; we can choose any energy we like, and shoot just once. We'll shoot from right to

left, because it's on the right that the wavefunction is completely known (given an arbitrary choice of C).

Without further ado, here is some Mathematica code to solve the TISE for these conditions. First I need to define the potential energy function:

```
v0 = 0.025; (* height of potential step *)
b = 20; (* width of sloped potential step *)
xMax = 100;
v[x_] := If[x < 0, 0, If[x > b, v0, x*v0/b]];
Plot[v[x], {x, -xMax, xMax}]
```

Here the constant `xMax` is the distance, to either side of $x = 0$, over which I will solve the TISE. For now I'm using the same potential function as in Figure 3.7 (aside from the horizontal shift to put $x = 0$ at the bottom of the "ramp"), but the rest of the code will work for essentially any potential in the middle region. To carry out the solution I'll use `NDSolve`:

```
energy = 0.03;
kp = Sqrt[2(energy-v0)];
solution = NDSolve[{psi'[x] == -2(energy-v[x])psi[x],
psi[xMax] == 1, psi'[xMax] == I*kp}, psi, {x, -xMax, xMax}];
```

Notice the boundary conditions in the last line: given that $\psi(x) = Ce^{ik'x}$ for any $x > b$, we must have $d\psi/dx = ik'Ce^{ik'x}$ in this region; I've arbitrarily set $\psi = 1$ at `xMax`, and this choice forces me to set $\psi = ik'$ at this same point. (In the code I've used `kp` for k' , and I've used equation 3.29 to calculate its value from the energy.)

Now that I have the solution, the first thing I want to do is plot it. I'll plot the probability density $|\psi|^2$, and use color hues to represent phases as usual:

```
Plot[Abs[psi[x]]^2 /. solution, {x, -xMax, xMax},
PlotRange -> {0, All},
Filling -> Axis,
ColorFunction -> Function[x, Hue[Arg[psi[x]]/.solution]/(2Pi)],
ColorFunctionScaling -> False,
PlotPoints -> 500]
```

Figure 3.10 shows the resulting plots for three different values of the wavefunction energy. There's a lot to notice in these plots! In the transmission region, to the right of $x = 20$, the numerical solution has a constant amplitude, with color hues indicating a rightward momentum flow and wavelengths that you can check are consistent with the formula $k' = \sqrt{2(E - V_0)}$. At the extreme right end, the red hue indicates that the wavefunction is pure real, in accord with my chosen boundary condition. To the left of $x = 0$, meanwhile, the probability density oscillates, showing interference between the incident and reflected waves. Even here, however, the color hues indicate that the net momentum is to the right, because the reflected amplitude

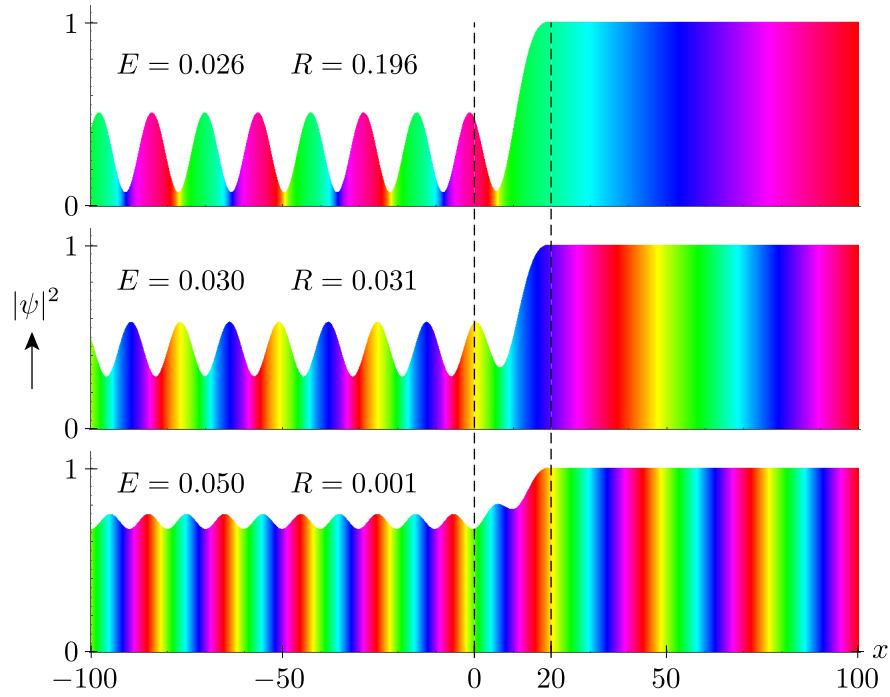


Figure 3.10: Numerical solutions of the TISE for three selected energies, for the same sloped step potential as in Figure 3.7 (aside from a horizontal shift). The dashed lines mark the ends of the sloped step. The boundary conditions are chosen so that the solution on the right is a pure right-going traveling wave. On the left we see interference between the incident and reflected waves. Larger interference fringes correspond to a larger reflection probability R , as described in the text. Here the distance units are arbitrary (inherited from the increment dx in Section 3.3), while the size of the energy unit is determined by the distance unit via $\hbar = m = 1$.

is always smaller than the incident amplitude. Notice that the wavelength of the interference fringes is exactly half the wavelength of the phase oscillations. Finally, and most importantly, the relative size of the interference fringes is large when the wavefunction energy is only slightly above V_0 , and small when E is substantially higher.

Think a little more about the size of the interference fringes. If there were no reflection at all, there would be only an incident wave in the left region so we would see a constant magnitude, with no fringes. On the other hand, if we had 100% reflection, the right- and left-moving waves in this region would have the same amplitude, so the fringes would be as large as possible, with total destructive interference at the minima. In general, the larger the relative size of the interference fringes, the greater the value of the reflection probability R .

Now let's quantify this idea. According to equation 3.28, the probability density

in the left region is

$$\begin{aligned} |\psi(x)|^2 &= (A^* e^{-ikx} + B^* e^{ikx})(A e^{ikx} + B e^{-ikx}) \\ &= |A|^2 + |B|^2 + AB^* e^{2ikx} + A^* B e^{-2ikx}. \end{aligned} \quad (3.30)$$

The cross-terms look a bit threatening, but we can tame them by writing A and B in amplitude-phase form:

$$A = |A|e^{i\alpha}, \quad B = |B|e^{i\beta}, \quad (3.31)$$

for some angles α and β (whose values won't concern us). Now with two lines of algebra you can obtain the much nicer formula

$$|\psi(x)|^2 = |A|^2 + |B|^2 + 2|A||B| \cos(2kx + \alpha - \beta), \quad (3.32)$$

which is explicitly real and shows exactly how the probability density oscillates in x with a wavelength of π/k . The high points of the oscillation are where the cosine equals $+1$, and we can take the square root at any of these points to obtain

$$|\psi|_{\max} = \sqrt{|A|^2 + |B|^2 + 2|A||B|} = |A| + |B|. \quad (3.33)$$

Similarly, the low points are where the cosine equals -1 , so

$$|\psi|_{\min} = \sqrt{|A|^2 + |B|^2 - 2|A||B|} = |A| - |B|. \quad (3.34)$$

It's easy to solve this pair of linear equations for $|A|$ and $|B|$ in terms of the maximum and minimum values of $|\psi|$, which we can obtain from our numerical solution. What we really want, though, is the reflection probability R , which is the ratio of the reflected intensity to the incident intensity. These intensities are proportional to the squared magnitudes of the respective amplitudes, just like any other quantum mechanical probabilities. The bottom line, then, is that the reflection probability is

$$R = \frac{|B|^2}{|A|^2} = \left(\frac{|\psi|_{\max} - |\psi|_{\min}}{|\psi|_{\max} + |\psi|_{\min}} \right)^2, \quad (3.35)$$

and because whatever isn't reflected must be transmitted, the transmission probability is $T = 1 - R$.

To use equation 3.35 we could just read the maximum and minimum $|\psi|^2$ values off of our plots, but it's more accurate to let the computer find them. The relevant Mathematica functions are `FindMaximum` and `FindMinimum`, and like other numerical analysis functions, they take some getting used to. Here's my code:

```
max = FindMaximum[Abs[psi[x]] /. solution,
                    {x, -0.8 xMax, -0.2 xMax}][[1]];
min = FindMinimum[Abs[psi[x]] /. solution,
                    {x, -0.8 xMax, -0.2 xMax}][[1]];
reflection = ((max - min)/(max + min))^2
```

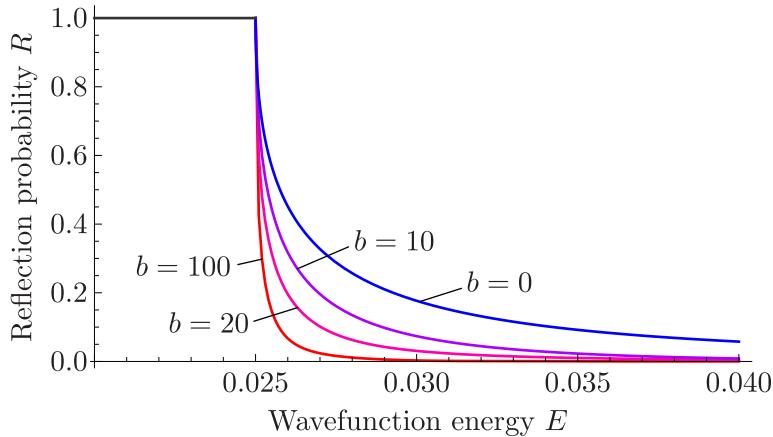


Figure 3.11: The reflection probability as a function of the wavefunction energy, for scattering of a quantum particle from a sloped step potential as in Figures 3.7 and 3.10. The height of the step is 0.025, and each curve is for a different value of the width b of the slope. I calculated the three curves for $b > 0$ numerically, but used equation 3.36 for the $b = 0$ curve.

For both the maximum and minimum, I told Mathematica to search an interval of x values that includes multiple fringes, but that stays away from the edges of the regions where the numerical solution is either undefined (beyond the far left) or no longer of the assumed form (right of $x = 0$). I appended “[[1]]” to extract the values of the maximum and minimum, throwing away the information about where these values occur (try each line without the [[1]] to see what I mean). The results of this calculation are written in Figure 3.10, for each of the three energies shown. I was a bit surprised by the small R values in relation to the sizes of the fringes, but that’s what you get from taking the square roots and such, according to equation 3.35.

Figure 3.11 shows more generally how the reflection coefficient falls off as the energy increases, at a rate that depends on the width b of the sloped potential step. When the slope is very shallow (large b), the quantum particle behaves almost as a classical particle, with essentially no reflection as long as its energy is just a little above the step height. At the other extreme, an abrupt step with $b = 0$ produces maximum reflection. This behavior is analogous to that of light, which reflects more strongly from an abrupt boundary than from a smoothly varying interface.

As you might guess, the $b = 0$ case is simple enough that there’s no need to calculate R numerically. In this case, as you can show in Problem 3.29, the reflection probability is given by the reasonably simple formula

$$R = \left(\frac{k - k'}{k + k'} \right)^2 = \left(\frac{1 - \sqrt{1 - V_0/E}}{1 + \sqrt{1 - V_0/E}} \right)^2. \quad (3.36)$$

Problem 3.24. Use a ruler (or the electronic equivalent) to measure the wavelength of each of the wavefunctions shown in Figure 3.10, both to the left of the step and to the right. (On the left, measure the wavelength of the phase oscillations, not the interference fringes.) Then calculate what each of these measured wavelengths should be, and check for agreement.

Problem 3.25. In the code on page 97 I set `xMax` to 100. Why do you think I chose that value? Would the code still work for a larger or a smaller value? Discuss.

Problem 3.26. Starting from equation 3.30, write out a complete derivation of equation 3.35, including the algebraic steps omitted in the text.

Problem 3.27. Use the code in the text to reproduce the three wavefunctions shown in Figure 3.10. Also try some intermediate energy values, and plot the reflection probability as a function of E . Check that your plot agrees with Figure 3.11.

Problem 3.28. All three of the wavefunction plots in Figure 3.10 are for values of E that are greater than the step height V_0 . What do you think would happen if E is less than V_0 ? Write down your expectation in some detail, then run the code for some $E < V_0$ and see if you're right. Does the solution have any features that you didn't expect?

Problem 3.29. Use the results that you proved in Problem 2.1 to derive equation 3.36.

Problem 3.30. Each of the curves in Figure 3.11 shows R as a function of E for fixed b , but we could do it the other way around. Sketch your best guess for what a plot of R vs. b for fixed E would look like, say for $V_0 = 0.025$ and $E = 0.030$. Then adapt the code to produce a quantitatively accurate version of this plot, being sure to include enough points to see the details. How accurate was your guess? Can you explain the interesting features of the plot?

Tunneling

There are many other potential shapes, besides the sloped step potential, that give rise to interesting scattering behavior. For instance, with a succession of two abrupt steps we can obtain constructive and destructive interference effects that are analogous to thin-film interference of light. I will save that scenario for Problem 3.35, and consider here the case of scattering from a barrier in which the potential energy is greater than the energy of the incident particle. The barrier is then a classically forbidden region, so a classical particle would have a 100% reflection probability. But if the barrier is sufficiently thin, with a classically allowed region on the other side, then a quantum particle has a chance to tunnel through it—in analogy to the tunneling between the two sides of a double well that we explored in Section 2.5.

Figure 3.12 shows an example. For simplicity I'm taking the barrier to be rectangular in shape: an inverted finite square well. More explicitly,

$$V(x) = \begin{cases} V_b & \text{for } 0 < x < b, \\ 0 & \text{elsewhere.} \end{cases} \quad (3.37)$$

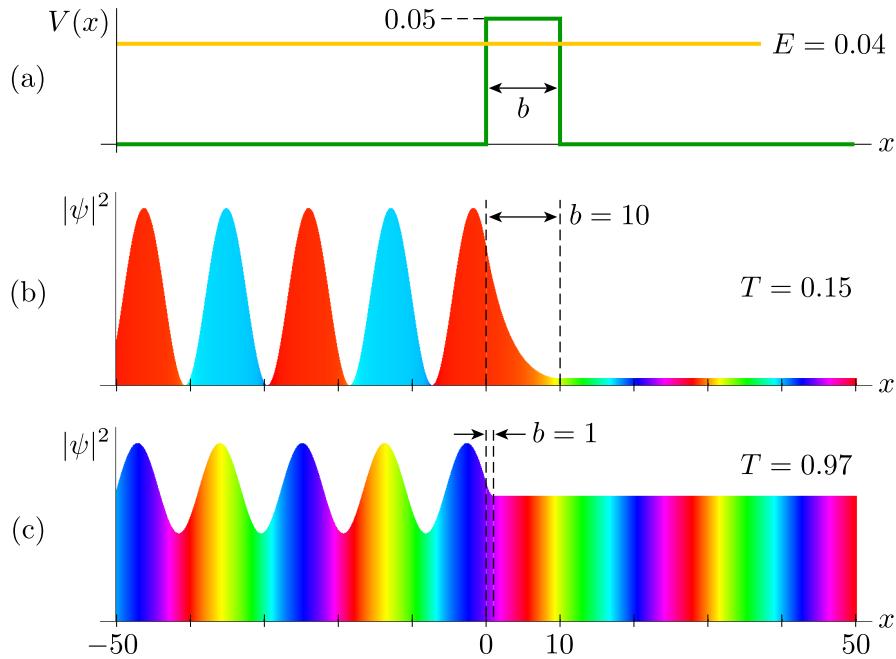


Figure 3.12: A rectangular potential barrier (a) with $V = 0.05$ inside, and scattering solutions of the TISE for this potential with $E = 0.04$, so that the inside of the barrier is a classically forbidden region. In (b) the barrier width is 10 units and the reflected wave is nearly as strong as the incident wave, producing a strong interference pattern on the left and a tunneling probability of only 15%. In (c) the barrier width is only 1 unit and the tunneling probability is 97%. The dashed vertical lines indicate the barrier edge locations. The distance unit is arbitrary (inherited from the increment dx in Section 3.3); the energy unit is determined by the distance unit via $\hbar = m = 1$.

The figure shows solutions to the TISE for $V_b = 0.05$, $E = 0.04$, and two values of b . As you can see, these barriers are narrow enough to result in significant tunneling probabilities: 15% for $b = 10$ and 97% for $b = 1$.

Producing wavefunction plots like those in Figure 3.12, and calculating the corresponding reflection and transmission probabilities R and T , is quite straightforward using a slightly modified version of the numerical TISE code that begins on page 97. It's equally easy to adapt that code to more realistic, non-rectangular barrier shapes. For the rectangular barrier, though, you can also solve the TISE analytically to get an explicit formula for the tunneling probability as a function of V_b , E , and b (see Problem 3.34).

Rather than diving straight into such a detailed calculation, I'd like to take a semi-quantitative approach and try to understand when the tunneling probability is large, and when it is small. Look closely at the wavefunction inside the barrier in Figure 3.12(b), and notice that it has the approximate shape of a decaying exponential function. The total amount of exponential decay determines the relative amplitudes of the wavefunction on either side, and hence the transmission probabil-

ity T . This decay amount is determined, in turn, by the distance b over which the decay occurs, along with the rate of decay, which depends on how far E is below V_b .

More precisely, recall from Problem 1.36 that the solution to the TISE inside the barrier is actually a sum of decaying and growing exponential functions:

$$\psi(x) = De^{-\kappa x} + Ge^{+\kappa x} \quad (\text{inside barrier}), \quad (3.38)$$

where

$$\kappa = \frac{\sqrt{2m(V_b - E)}}{\hbar} \quad (3.39)$$

is the forbidden-region counterpart to k (note the sign change inside the square root), and D and G are complex constants. The decaying and growing terms must actually be equal in magnitude at the right edge of the barrier, in order to match the horizontal slope of the pure right-moving traveling wave to the right (see Problem 3.32). But as we move leftward into the barrier from the right, the $e^{+\kappa x}$ term decreases exponentially, while the $e^{-\kappa x}$ term grows exponentially. For a sufficiently wide barrier, with $\kappa b \gg 1$, the $e^{+\kappa x}$ term soon becomes negligible and the $e^{-\kappa x}$ term then dominates. (Notice in Figure 3.12(b) that the phase of the wavefunction varies just inside the right edge of the barrier, where the $e^{+\kappa x}$ term is still significant, but is constant near the left edge, where this term is negligible.)

So let's assume that we're dealing with a wide barrier, with $\kappa b \gg 1$. Then the wavefunction amplitude decreases from left to right across the barrier by *roughly* a factor of $e^{-\kappa b}$, and the probability density decreases by the square of this factor. There's still the complication that the wavefunction on the left side of the barrier is a standing-wave superposition of incident and reflected waves, but the amplitude of this wave superposition at the barrier edge will ordinarily be within a factor of 2 or so of the amplitude of just the incident wave. Ignoring this factor for simplicity, we can conclude that the transmitted probability density is smaller than the incident probability density by a factor of $e^{-2\kappa b}$, that is,

$$T \sim e^{-2\kappa b}. \quad (3.40)$$

That's for a rectangular barrier of width b . For a barrier of arbitrary shape, we can divide the classically forbidden region into a succession of smaller intervals in which the potential is approximately constant. Then, if we apply the (admittedly fuzzy) logic of the previous paragraph to a single interval at location x with width dx , we can write

$$\psi(x + dx) \sim e^{-\kappa(x) dx} \psi(x), \quad (3.41)$$

where

$$\kappa(x) = \frac{\sqrt{2m(V(x) - E)}}{\hbar}. \quad (3.42)$$

Applying equation 3.41 to each interval in succession then gives

$$\begin{aligned} \psi(b) &\sim e^{-\kappa(x_1) dx} e^{-\kappa(x_2) dx} \dots e^{-\kappa(x_n) dx} \psi(0) \\ &= \exp\left(-\sum_i \kappa(x_i) dx\right) \psi(a), \end{aligned} \quad (3.43)$$

where x_i is the value of x in the i th interval, and I've assumed that the classically forbidden region extends from $x = a$ to $x = b$. Taking the square modulus of both sides and converting the sum to an integral, we find for the tunneling probability

$$T \sim \exp\left(-2 \int_a^b \kappa(x) dx\right). \quad (3.44)$$

Again, this formula is valid only when $T \ll 1$, and even then gives only a rough estimate.

Problem 3.31. Modify the code starting on page 97 to reproduce the wavefunction plots in Figure 3.12. Also check that you get the correct transmission probabilities.

Problem 3.32. Prove that for a tunneling solution to the TISE for a rectangular barrier, as shown in Figure 3.12, the magnitudes of the decaying and growing exponential contributions to the wavefunction (see equation 3.38) just inside the right edge of the barrier must be equal.

Problem 3.33. Use equation 3.40 to estimate the tunneling probability for the parameters used in Figure 3.12(b), and comment on the result.

Problem 3.34. [Derive explicit formula for tunneling through a rectangular barrier, and compare to approximate formula.]

Problem 3.35. [Explore scattering from a rectangular well or low barrier or double step, focusing on interference effects.]

3.5 Summary of one-dimensional wave mechanics

This chapter concludes our exclusive focus on a single, structureless, quantum particle moving in one dimension. Before we go on to consider a wider variety of systems, let's step back and summarize the logical principles of quantum mechanics in the one-dimensional case.

Principle 1: The states of the system correspond to wavefunctions.

By now you've worked with quite a variety of wavefunctions: position and momentum eigenfunctions, energy eigenfunctions for trapped and untrapped particles, wavepackets, and the complicated functions that result from wavepackets interacting with nontrivial potentials.

Wavefunctions are complex-valued, with both a magnitude and a phase at each point in space. But only *relative* phases matter, so we're always free to specify the phase of the wavefunction at one particular x value, or to multiply the entire wavefunction by a pure phase constant, $e^{i\phi}$. Some wavefunctions are simple enough that we can use this freedom to make them purely real.

As far as we know, *any* complex-valued function of x is an allowed wavefunction, so long as it is normalized. Sometimes we can get away with using unnormalized wavefunctions, when we care only about qualitative features or relative magnitudes.

And sometimes we can even get away with using wavefunctions that can't be normalized, such as delta functions and momentum eigenfunctions and the scattering solutions of Figures 3.10 and 3.12.

Although any normalized wavefunction is *allowed*, this doesn't mean that every normalized wavefunction is easy to realize in the laboratory. Localized wavepackets and harmonic oscillator eigenfunctions occur all the time in real-world situations, while "cat states" with two or more widely separated peaks require more specialized preparation.

Principle 2: Observable quantities correspond to operators.

I've already listed the important operators for one-dimensional quantum mechanics in Table 1.1: x for position, $-i\hbar(d/dx)$ for momentum, and the Hamiltonian operator,

$$\hat{H} = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x), \quad (3.45)$$

for total energy (kinetic plus potential). The most important property of these operators is that each has a complete set of eigenfunctions, for which the corresponding physical quantity is well defined and equal to the corresponding eigenvalue. Because the position eigenfunctions are different from the momentum eigenfunctions, it's impossible to put a particle into a single quantum state for which *both* position and momentum are well defined. The same is true for position and energy, and also for momentum and energy except in the trivial case where V is constant. However, each set of eigenfunctions forms a *complete* set, so we can express an arbitrary wavefunction as a mixture of position eigenfunctions (see Problem 3.4), or as a mixture of momentum eigenfunctions (see equation 3.3), or as a mixture of energy eigenfunctions (see equations 1.33 and 2.28).

Inconveniently for us, the position eigenfunctions (delta functions) and momentum eigenfunctions ($e^{ipx/\hbar}$) are not normalizable—but that doesn't stop us from expressing an arbitrary wavefunction in terms of them, or from using normalizable wavefunctions, such as very narrow spikes or very wide wavepackets, that are equivalent to the idealized eigenfunctions for all practical purposes.

Principle 3: Measurement probabilities are given by the Born rule.

Suppose that our particle is in state $\psi(x)$ and we measure an observable quantity whose operator is \hat{A} , with eigenvalues a_i and corresponding eigenfunctions $\alpha_i(x)$. Then the possible outcomes of the measurement are the various eigenvalues a_i , but in general we can predict only the *probability* of obtaining each of these outcomes. When the eigenvalues are discrete, the probability of obtaining a particular eigenvalue a_i equals the square modulus of the *component* of ψ along the corresponding eigenfunction α_i :

$$(\text{Probability of } a_i) = |\langle \alpha_i | \psi \rangle|^2 = \left| \int_{-\infty}^{\infty} \alpha_i^*(x) \psi(x) dx \right|^2. \quad (3.46)$$

This equation is called the *Born rule*. In the special case where $\psi(x)$ is one of the eigenstates α_i , the inner product equals 1, giving a probability of 100%, for obtaining that particular outcome a_i , while the probability of getting any other outcome is zero, because that corresponding eigenfunction will be orthogonal to ψ (inner product equal to zero).

When the eigenvalues are continuous, the probability of obtaining any one of them is infinitesimal, so we have to integrate this squared magnitude over some range of desired values as in equation 1.6 for position and equation 3.10 for momentum.

Principle 4: A measurement causes the state to collapse.

Whatever the result of a measurement, the measurement process causes the system's state to abruptly “collapse” to become whatever eigenstate corresponds to the value obtained. This means that if you immediately measure the same physical quantity again, you will always get the same result that you got the first time.

Principle 5: The TDSE governs time evolution.

As time passes, the wavefunction changes smoothly according to the time-dependent Schrödinger equation (TDSE),

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H}\psi, \quad (3.47)$$

where \hat{H} is the Hamiltonian operator in equation 3.45. In Section 3.3 you saw how to solve this equation numerically for any $V(x)$ and any initial wavefunction. Alternatively, if we can find the energy eigenfunctions, then we can expand any initial ψ in terms of them and write down the solution as in equation 1.24. In the special case in which the initial ψ is an energy eigenfunction, the TDSE tells us that its time dependence is a simple wiggle factor, $e^{-iEt/\hbar}$, and therefore the probability density does not change over time.

* * *

As I mentioned at the end of Section 1.2, Principles 3 and 4 awkwardly leave the word “measurement” undefined. This ambiguity rarely matters in practice, but it makes many of us uncomfortable. Some researchers have gone so far as to reject Principles 3 and 4 as parts of the fundamental theory, asserting (or at least hoping) that instead they somehow emerge as a consequence of the other three principles—when we generalize these principles to include not just the particle we’re measuring but also the apparatus that we’re using to measure it. Whether or not that idea has merit, we are now ready to turn our attention to more general quantum systems.

Chapter 4

Beyond one dimension

We're finally ready to generalize quantum mechanics beyond the system of a single, structureless particle in one dimension. Specifically, we will now consider:

- a particle moving in two- or three-dimensional space;
- systems of multiple particles; and
- particles with internal degrees of freedom (such as spin).

This chapter lays out the basic framework for treating this greater variety of systems, with examples that illustrate some of the new features they have. Later chapters will explore some of the most important applications in greater depth.

4.1 Multiple spatial dimensions

Let's now consider a quantum particle in two-dimensional or three-dimensional space. I'll mostly focus on the two-dimensional case, because the equations are less cumbersome and the pictures are easier to draw—but I'll occasionally add remarks about what happens in three dimensions.

Wavefunctions

In many ways, the generalization from one to two spatial dimensions is straightforward. The wavefunction ψ is now a complex-valued function of *two* variables, x and y , normalized so that

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy |\psi(x, y)|^2 = 1. \quad (4.1)$$

To compute the probability of finding the particle within some two-dimensional region, we would integrate $|\psi|^2$ only over that region.

If a particle is localized, then its wavefunction is nonzero only over a small range of x and y values (see Figure 4.1). In the idealized limit, such a function becomes

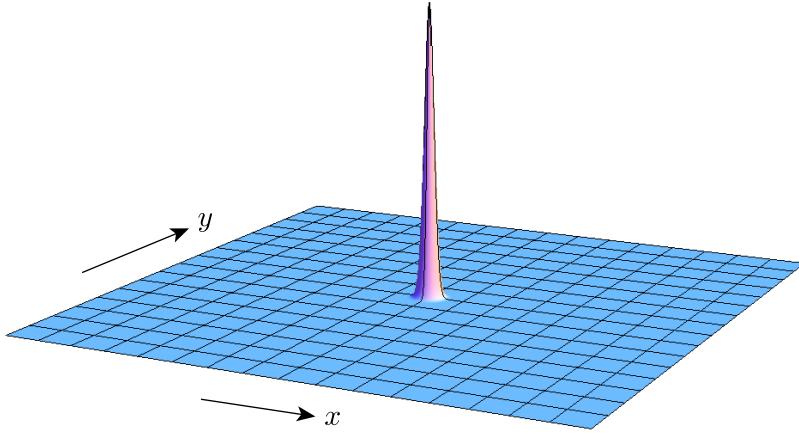


Figure 4.1: A localized wavefunction in two dimensions, shown as a “surface plot” with ψ (which in this case is real) plotted vertically. In the limiting case in which the width of the peak goes to zero and its height goes to infinity, this function becomes a position eigenfunction, $\delta^{(2)}(\vec{r} - \vec{r}_0)$.

a position eigenfunction, which is nonzero everywhere except at a single point, and can be expressed as a product of delta functions:

$$\delta(x - x_0)\delta(y - y_0) = \delta^{(2)}(\vec{r} - \vec{r}_0), \quad (4.2)$$

where $\vec{r} = (x, y)$ and the superscript (2) denotes a two-dimensional delta function. As in one dimension, however, a position eigenfunction is not normalizable.

An idealized momentum eigenfunction, on the other hand, would have the form

$$e^{ip_x x/\hbar} e^{ip_y y/\hbar} = e^{i\vec{p} \cdot \vec{r}/\hbar}, \quad (4.3)$$

where $\vec{p} = (p_x, p_y)$. This function varies only as \vec{r} changes in the \vec{p} direction, as shown in Figure 4.2(a). The wavefunction is constant along any line perpendicular to \vec{p} , so we call this a *linear wave*. In three dimensions the wavefunction would be constant along any plane perpendicular to \vec{p} , so we would call $e^{i\vec{p} \cdot \vec{r}/\hbar}$ a *plane wave*.

As an example of a more realistic wavefunction, Figure 4.2(b) shows a two-dimensional Gaussian wavepacket, whose formula would be something like

$$\psi(x, y) = e^{-(x^2 + y^2)/a^2} e^{i\vec{p} \cdot \vec{r}/\hbar}, \quad (4.4)$$

where for simplicity I've centered the wavepacket at the origin and neglected an overall normalization constant.

Figures 4.1 and 4.2 also illustrate two ways to visualize a two-dimensional wavefunction: either as a 3D surface plot with $|\psi|$ or $|\psi|^2$ on the vertical axis, or as a 2D density plot using saturation or brightness to represent $|\psi|$ or $|\psi|^2$. We can also use color hues to represent the wavefunction phases, as in earlier chapters. (Visualizing three-dimensional wavefunctions is quite a bit more difficult, but there's some cool software at <http://falstad.com> that can help.)

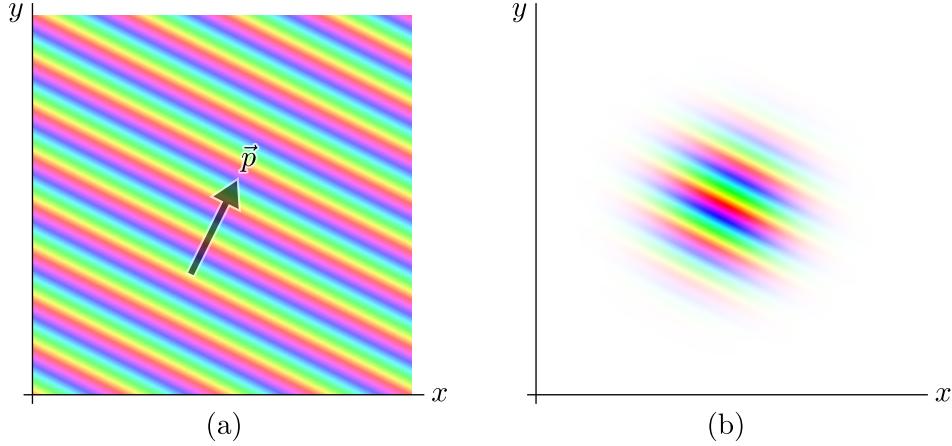


Figure 4.2: (a) A portion of a two-dimensional momentum eigenstate, $e^{i\vec{p}\cdot\vec{r}}$, with the direction of \vec{p} indicated. (b) A two-dimensional Gaussian wavepacket (equation 4.4) with the same average momentum as in (a). In these “density plots” the color hue indicates the phase of the wavefunction, while the saturation is proportional to the probability density $|\psi|^2$ (so white indicates a value of zero).

Probabilities

As I mentioned after equation 4.1, you can compute the probability of finding the particle in any two-dimensional region by integrating $|\psi|^2$ over that region. This is the two-dimensional version of the Born rule, equation 1.6. For a rectangular region, the explicit formula would be

$$\left(\begin{array}{c} \text{Probability of finding} \\ a < x < b \text{ and } c < y < d \end{array} \right) = \int_a^b dx \int_c^d dy |\psi(x, y)|^2, \quad (4.5)$$

as shown in Figure 4.3. Another important special case is where the region includes the full range of one coordinate but only a limited range of the other. For instance, if we measure x but not y , then the probability we care about is

$$\left(\begin{array}{c} \text{Probability of finding} \\ a < x < b \text{ and any } y \end{array} \right) = \int_a^b dx \int_{-\infty}^{\infty} dy |\psi(x, y)|^2 = \int_a^b dx P_x(x), \quad (4.6)$$

where the last expression defines the one-dimensional probability distribution $P_x(x)$:

$$P_x(x) = \int_{-\infty}^{\infty} |\psi(x, y)|^2 dy. \quad (4.7)$$

We can define a similar distribution $P_y(y)$ as the integral of $|\psi|^2$ over all x ; we would then integrate $P_y(y)$ over a limited range of y values to obtain the probability of finding the particle within that range, if we measure y but not x .

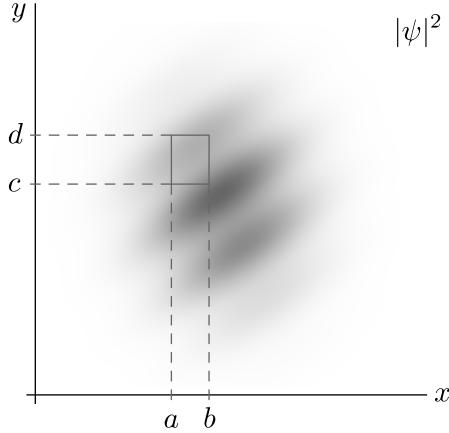


Figure 4.3: In two dimensions, we integrate $|\psi|^2$ over a two-dimensional region to obtain the probability of finding the particle within that region.

To obtain probabilities for momentum measurements, we use the two-dimensional inverse Fourier transform of ψ ,

$$\tilde{\psi}(p_x, p_y) = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \psi(x, y) e^{-ip\vec{r}/\hbar}, \quad (4.8)$$

in analogy to equation 3.8. The square modulus of $\tilde{\psi}$ is then the momentum-space probability density, which we can integrate over any desired region of momentum space to obtain the probability of finding \vec{p} in that region, in analogy to equation 3.10.

The Born rules for position and momentum require us to integrate over ranges of values, because the eigenvalues are continuous. In the simpler case where the eigenvalues are discrete, as for the energy of a trapped particle, no integration is needed. But to find the component of an arbitrary wavefunction ψ in the “direction” of particular eigenfunction α_i , we need to calculate the two-dimensional inner product,

$$\langle \alpha_i | \psi \rangle = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \alpha_i^*(x, y) \psi(x, y). \quad (4.9)$$

(Equation 4.8 is the special case in which $\alpha_i \propto e^{i\vec{p}\cdot\vec{r}/\hbar}$, and equation 4.1 simply says that the component of a normalized wavefunction in its own direction is 1.) If the eigenfunction α_i corresponds uniquely to eigenvalue a_i , then the probability of obtaining this measurement outcome (when we measure the quantity for which α_i is an eigenstate) is the square of the inner product, $|\langle \alpha_i | \psi \rangle|^2$.

In multiple dimensions, however, we will sometimes find that eigenvalues are *degenerate*:¹ two or more orthogonal eigenfunctions α_i can have the same eigenvalue a . In this case we need to sum over all these eigenfunctions to obtain the

¹Degeneracy also occurs in one dimension, but not as often. The most notable case is left- and right-moving traveling waves of the same wavelength, which are degenerate eigenfunctions of

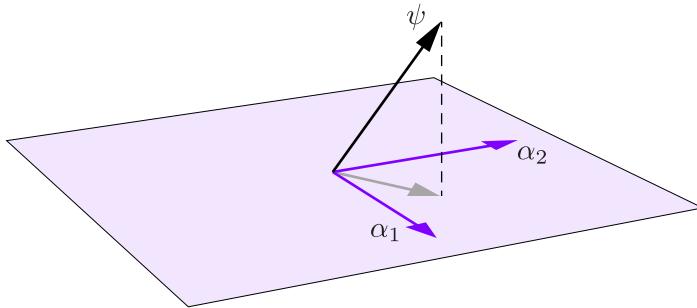


Figure 4.4: When two eigenvectors α_1 and α_2 have the same eigenvalue, they define a plane in which every other vector is also an eigenvector with the same eigenvalue. If the system's state vector is ψ , the probability of obtaining that eigenvalue as a measurement outcome is (according to equation 4.10) the squared length of the projection of ψ onto the plane. If the measurement yields this eigenvalue as a result, then the state collapses to the projected vector, renormalized so its length is 1. (This figure is drawn for vectors with real components. Complex vectors are much harder to visualize!)

probability we want:

$$(\text{Probability of } a) = \sum_i |\langle \alpha_i | \psi \rangle|^2, \quad (4.10)$$

where the sum runs over all i for which the eigenfunction α_i corresponds to eigenvalue a .

Degeneracy creates some further technical complications for us, because any linear combination of degenerate eigenfunctions is also an eigenfunction, with the same eigenvalue (see Problem 4.1). So, for instance, when we have two degenerate eigenfunctions, we really have a two-dimensional *eigenspace*, consisting of all the linear combinations that mix the original two functions in various proportions (see Figure 4.4 for a vector visualization of this space). The sum of squared components in equation 4.10 is equal (by the Pythagorean theorem) to the squared length of the *projection* of ψ onto the two-dimensional eigenspace. At least for the purpose of this particular observable quantity, the choice of basis functions within the eigenspace is arbitrary—so we can sometimes simplify calculations by aligning, say, α_1 with the projection of ψ , so that α_2 and ψ are orthogonal.

Degeneracy also creates an apparent ambiguity when the wavefunction collapses as a result of a measurement. If the measured result is degenerate, then which of the degenerate eigenfunctions should the wavefunction collapse to? Figure 4.4 again suggests the natural answer: ψ collapses to a vector (function) that lies along its projection in the degenerate subspace, so that its components in the α_1 and α_2 directions are in the same ratio to each other as before.

the Hamiltonian (energy) operator for a free particle. The energy levels of a *bound* particle in one dimension are never degenerate, as shown in Problem 2.2.

At the end of this section I'll give a specific example in which two-dimensional bound-state energies are degenerate. Meanwhile, we already saw an example of degeneracy in equation 4.6, where we imagined measuring the position of a particle along the x direction but not along y . The integral over y in that equation corresponds to the sum over degenerate eigenvectors in equation 4.10. Moreover, if we go ahead and measure x and obtain the result $x = a$, then the wavefunction collapses to a delta function in the x direction, multiplied by the original wavefunction evaluated at $x = a$:

$$\psi_{\text{new}}(x, y) \propto \psi_{\text{old}}(a, y) \delta(x - a). \quad (4.11)$$

This partially collapsed wavefunction is just the projection of the original wavefunction onto the degenerate subspace that corresponds to the measurement outcome.

Problem 4.1. Prove that if α_1 and α_2 are both eigenfunctions of some linear operator \hat{A} with the same eigenvalue a , then any linear combination of the form $c_1\alpha_1 + c_2\alpha_2$ is also an eigenfunction of \hat{A} with eigenvalue a .

Nonseparable wavefunctions

The example wavefunction formulas in equations 4.2, 4.3, and 4.4 all have the special property that they can be *faktored* into a function of x times a function of y :

$$\psi(x, y) = f(x)g(y) \quad (\text{usually not possible}). \quad (4.12)$$

We say that such functions are *separable* (in rectangular coordinates). However, this property is the exception rather than the rule. It's easy to write down wavefunction formulas, such as $e^{-x^2y^2/a^4}$ and $e^{ik\sqrt{x^2+y^2}}$, that are not separable.

One especially easy way to construct nonseparable wavefunctions is to combine two or more separable wavefunctions into a superposition. For example, consider the superposition of Gaussians

$$e^{-(x-a)^2} e^{-(y-b)^2} + e^{-(x-b)^2} e^{-(y-a)^2}, \quad (4.13)$$

where x , y , a , and b are understood to be measured as multiples of some suitable unit. Each of the two terms in this expression is separable, but the entire expression (which is an equally legitimate wavefunction) is not. If the difference $b - a$ is a few units, then this function looks something like Figure 4.5. This is just a two-dimensional “cat state,” similar to those in Figures 1.6 and 1.7, with the two peaks separated along a diagonal direction. Loosely speaking, we would say that the particle is half in one place and half in the other.

The nonseparability of most wavefunctions isn't a mere mathematical curiosity: it has the important *physical* consequence that measurements of the two coordinates are *correlated*. For instance, if you were to measure either the x or y coordinate of a particle in the cat state shown in Figure 4.5, you would have a 50/50 chance of obtaining a value either near a or near b . But if x is near a then y must be near b , and vice-versa. To put it more vividly, suppose you measure x first, and happen to

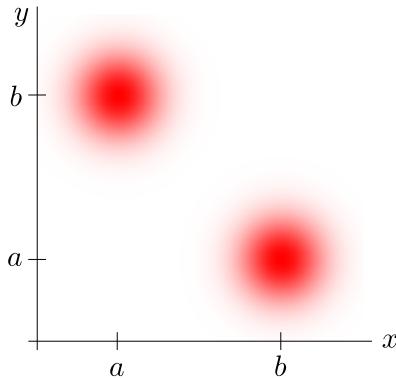


Figure 4.5: A two-peak “cat state” in two dimensions, consisting of Gaussian shapes centered at (a, b) and (b, a) . See equation 4.13.

obtain a value near a . This measurement collapses the wavefunction to lie entirely within the peak around $x = a$ and $y = b$, so you can now be certain that a subsequent measurement of y will yield a result near b . The measurement of x not only provides information you didn’t previously have about y ; it actually *changes* the probability distribution for the subsequent y measurement. (Problem 4.3 explores this scenario further.)

I should point out that sometimes a nonseparable wavefunction becomes separable if you switch to a different coordinate system. For example, the function $e^{ik\sqrt{x^2+y^2}}$ is separable in polar coordinates, while the wavefunction of Figure 4.5 can be factored if you rotate the coordinate axes by 45 degrees. Often the relevant coordinate system is dictated by some other outside condition, so a change of coordinates may or may not be appropriate. And if you make these wavefunctions just a little more complicated, they can’t be factored in any coordinate system.

Ironically, even though the vast majority of wavefunctions are *not* separable, we’ll spend most of our time working with wavefunctions that *are* separable. This is because they are mathematically simpler, and also because the separable wavefunctions are *complete*: any other wavefunction can always be expressed as a linear superposition of separable wavefunctions (as in equation 4.13). Just don’t let these separable wavefunctions give you the false impression that the quantum world is always so simple. In quantum mechanics you need to keep reminding yourself that your basis vectors are merely a basis, from which you can build arbitrary linear combinations to represent the unlimited variety of the world.

Problem 4.2. Show how to factor the two-dimensional Gaussian wavepacket function in equation 4.4, identifying the functions called $f(x)$ and $g(y)$ in equation 4.12.

Problem 4.3. Sketch the one-dimensional probability distributions $P_x(x)$ and $P_y(y)$ (as defined in equation 4.7) for the two-peaked wavefunction shown in Figure 4.5. Then sketch (as two-dimensional density plots, like Figure 4.5) two other wavefunctions, one separable and one not, that are different from this wavefunction yet have the same

$P_x(x)$ and $P_y(y)$. Describe in some detail what each of these wavefunctions predicts for the outcomes of sequential measurements of x and y . For instance, if you measure x and happen to obtain a result near a , what can you then predict about the outcome of a measurement of y ?

Problem 4.4. Make a color-phase density plot of each of the following two-dimensional wavefunctions, and in each case, explain how you can tell from the image whether or not the function is separable in rectangular (x, y) coordinates: (a) $e^{ik(x-y)}$; (b) $e^{ik(x-y)} + e^{-ik(x-y)}$; (c) $e^{ik\sqrt{x^2+y^2}}$; (d) $e^{ik^2(x^2+y^2)}$. (In all cases, take k to be a positive constant.)

Problem 4.5. Show both pictorially and algebraically that the wavefunction in equation 4.13 *is* separable if you rotate the coordinate axes by 45 degrees. Describe (and draw) at least two conceptually distinct ways in which you could modify this wavefunction so that it is not separable in any rotated coordinate system.

Operators

In two dimensions there are two independent position operators, $\hat{x} = x$ and $\hat{y} = y$, which we can think of as components of a two-dimensional vector position operator, $\hat{\vec{r}} = \vec{r}$. Similarly, there are two momentum operators,

$$\hat{p}_x = -i\hbar \frac{\partial}{\partial x} \quad \text{and} \quad \hat{p}_y = -i\hbar \frac{\partial}{\partial y}, \quad (4.14)$$

which we can think of as components of a two-dimensional vector momentum operator, $\hat{\vec{p}} = -i\hbar\nabla$. The sinusoidal linear waves in equation 4.3 are eigenfunctions of these operators, with the expected eigenvalues p_x and p_y .

To find the Hamiltonian operator for this two-dimensional system, we simply start with the classical expression for the system's total energy and substitute operators for the momentum components:

$$\hat{H} = \frac{1}{2m}(\hat{p}_x^2 + \hat{p}_y^2) + V(\hat{\vec{r}}) = -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + V(\vec{r}). \quad (4.15)$$

The differential operator inside the big parentheses is called the *Laplacian*, often abbreviated ∇^2 . With this Hamiltonian you can now write down either the TISE ($\hat{H}\psi = E\psi$) or the TDSE ($\hat{H}\psi = i\hbar\partial\psi/\partial t$) for this two-dimensional system.

Problem 4.6. Suppose that a particle in two dimensions is subject to a potential energy function of the form $V(x, y) = V_1(x) + V_2(y)$, where V_1 and V_2 are arbitrary functions of a single variable. Prove in this case that (a) you can solve the TISE by the method of separation of variables, to obtain a complete, separable set of energy eigenfunctions; and (b) if the particle's wavefunction is separable at $t = 0$, then it will remain separable as time passes.

The square infinite square well

Now let's take a first look at bound states in two dimensions. Just as we started with the one-dimensional infinite square well in Section 1.4, the natural place to

start here is with a *two*-dimensional infinite square well (or “box”). To keep things as simple as possible, I’ll assume that the well is literally a square:

$$V(x, y) = \begin{cases} 0 & \text{if } 0 < x < a \text{ and } 0 < y < a, \\ \infty & \text{elsewhere.} \end{cases} \quad (4.16)$$

This potential forces all wavefunctions to be zero outside the square box, and forces all wavefunctions to approach zero as x or y approaches 0 or a from inside. Otherwise, inside the box, any normalized complex-valued wavefunction is allowed.

As with any potential well, the most important wavefunctions are the energy eigenfunctions. To find them we could start with the two-dimensional TISE, but the answers are so easy to guess that there’s little point in being so formal. The energy eigenfunctions of a one-dimensional infinite square well are $\sqrt{2/a} \sin(n\pi x/a)$, so the natural guess for the two-dimensional case is a product of two such functions, that is,

$$\psi(x, y) = \frac{2}{a} \sin\left(\frac{n_x \pi x}{a}\right) \sin\left(\frac{n_y \pi y}{a}\right), \quad (4.17)$$

where n_x and n_y are separate quantum numbers, each of which can be any positive integer. (This formula applies only inside the box; again, every wavefunction must be zero outside.) Figure 4.6 shows nine of these eigenfunctions.

To find the corresponding energy eigenvalues, we can just use the two-dimensional formula for nonrelativistic kinetic energy (applied inside the well where $V = 0$), and then recognize that each term has the same form as the one-dimensional formula 1.30:

$$E = K = \frac{p_x^2}{2m} + \frac{p_y^2}{2m} = \frac{\hbar^2}{8ma^2}(n_x^2 + n_y^2). \quad (4.18)$$

Figure 4.7 shows a table and diagram of the low-lying energy levels. Note that most of these levels are doubly degenerate, because interchanging n_x and n_y leaves the energy unchanged.

Remember, though, that “doubly degenerate” actually means that there is an entire two-dimensional *space* of degenerate states. For instance, the degenerate states $\psi_{1,2}$ and $\psi_{2,1}$ (where the subscripts indicate n_x and n_y , respectively) each have energy $5\hbar^2/(8ma^2)$, but so does any linear combination of the form $c_1\psi_{1,2} + c_2\psi_{2,1}$, where c_1 and c_2 are complex numbers. Figure 4.8 shows a few of these infinitely many combinations.

A far greater variety of wavefunctions can be constructed as linear combinations of basis states that don’t necessarily have the same energy. In fact, the sine-wave products written in equation 4.17 form a *complete* basis, so *any* other wavefunction (that goes to zero at the four edges of the box) can be written in terms of them. They are also mutually orthogonal, by virtue of the orthogonality of the one-dimensional square well energy eigenfunctions.

Problem 4.7. Show by direct substitution that the wavefunction formula 4.17 solves the TISE, and verify the energy formula 4.18.

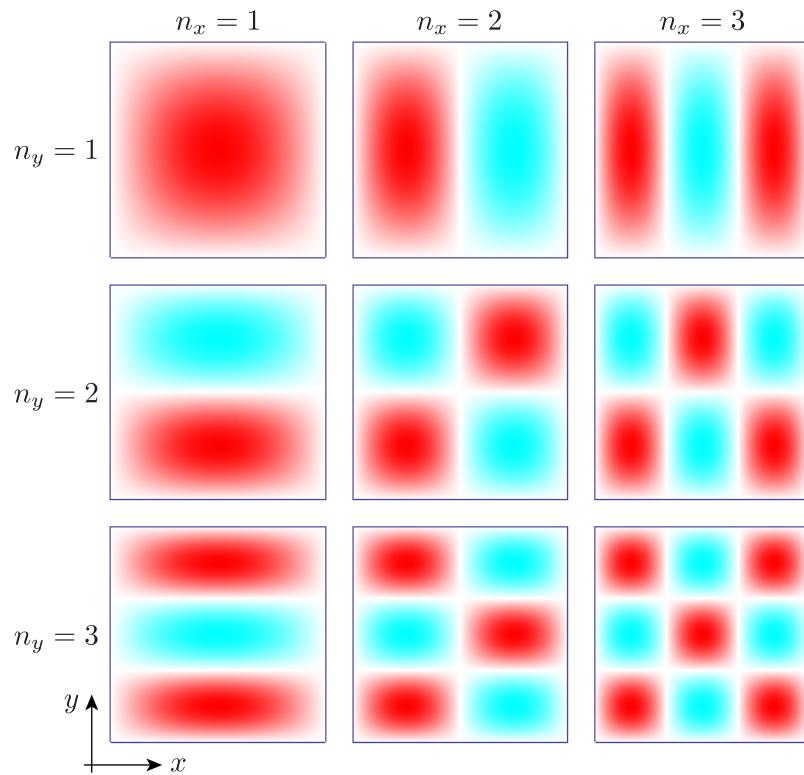


Figure 4.6: Separable energy eigenfunctions of the two-dimensional square infinite well (equation 4.17) with $n_x \leq 3$ and $n_y \leq 3$. In these density plots the saturation is proportional to $|\psi|$; red indicates positive values and cyan negative values, with zero represented by white.

Problem 4.8. Most of the energy levels shown in Figure 4.7 are doubly degenerate. What is the lowest level that is *triply* degenerate, and what are the corresponding quantum numbers n_x and n_y ? What is the lowest level with *four-fold* degeneracy, and what are the corresponding quantum numbers?

Problem 4.9. For the two-dimensional square infinite well, find an approximate formula for the number of independent energy eigenstates with energies less than E_{\max} , assuming that E_{\max} is sufficiently large so that this number is much larger than 1. (Hint: Represent the states as dots on a two-dimensional plot with n_x and n_y along the axes.) Then, working in the same approximation, find a formula for the number of independent eigenstates *per unit energy*; this quantity is called the *density of states*. You should find that the density of states for this system is a constant, independent of energy.²

Problem 4.10. Use a computer to reproduce each of the wavefunction plots in Figure 4.8. Then, using the computer if necessary, determine how the appearance of each

²The density of states is an important quantity in quantum statistical mechanics. See, for example, D. V. Schroeder, *An Introduction to Thermal Physics* (Addison-Wesley, San Francisco, 2000), chapter 7.

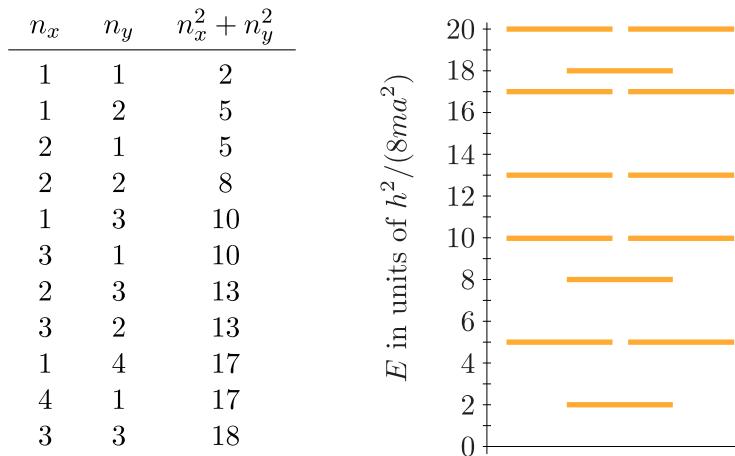


Figure 4.7: A list of some of the possible quantum number combinations for a two-dimensional square infinite well. The numbers in the third column are equal to the corresponding energies in units of $\hbar^2/(8ma^2)$, and are plotted vertically on the energy level diagram at right.

of these wavefunctions changes as time passes. Describe these changes carefully for each case.

Problem 4.11. Suppose that a particle in a two-dimensional square box is in the superposition state

$$\psi(x, y) \propto \psi_{2,3} + \frac{1}{2}\psi_{3,2},$$

where the subscripts indicate n_x and n_y , respectively, in equation 4.17.

- (a) Normalize this wavefunction (that is, find the constant of proportionality).
- (b) Use a computer to plot this wavefunction.
- (c) Imagine that you intend to measure the y coordinate of this particle. Calculate and plot the probability distribution $P_y(y)$ that shows the relative probability of obtaining each y value.
- (d) Suppose instead that before measuring y you first measure x , and you happen to obtain the result $x = a/4$. Calculate and plot the probability distribution $P_y(y)$ that now applies to your subsequent measurement of y , and compare to the answer to part (c).

Problem 4.12. Suppose that a particle in a two-dimensional square box is in the superposition state

$$\psi(x, y) \propto \psi_{2,3} + \frac{1}{2}\psi_{3,2} - \frac{2}{3}\psi_{1,4},$$

where the subscripts indicate n_x and n_y , respectively, in equation 4.17.

- (a) Normalize this wavefunction.
- (b) Suppose that you measure the energy of this particle. What values might you obtain, and what are their probabilities?
- (c) For each possible result that you might obtain in your energy measurement, what is the particle's normalized wavefunction immediately afterward?

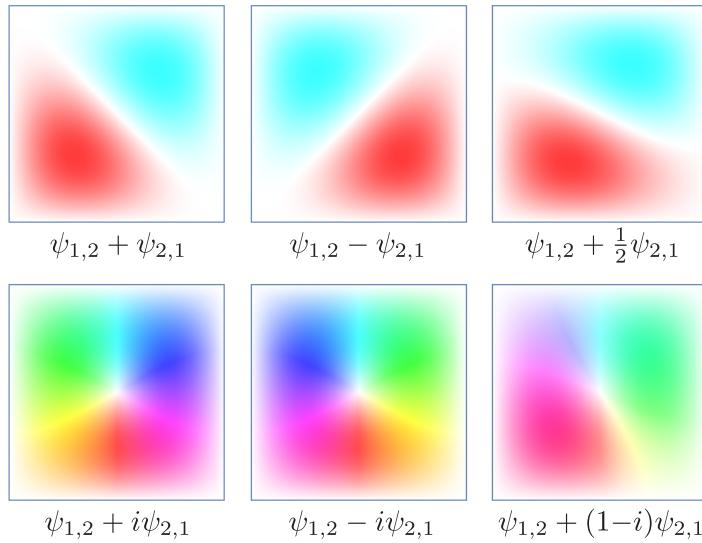


Figure 4.8: A selection of superpositions of two degenerate basis states for a particle in a square infinite square well. The two basis states, shown in Figure 4.6, are the first excited states $\psi_{1,2}$ and $\psi_{2,1}$, where the subscripts indicate n_x and n_y , respectively. Color hues represent complex phases as usual. (The expressions below the images are not normalized, and I've adjusted each image saturation independently.)

Problem 4.13. Consider a two-dimensional *rectangular* infinite square well, with width a in the x direction and width b in the y direction. What are the energy eigenfunctions for this system and their corresponding energies? Draw an energy level diagram for the case $a = 2b$ and another for the case $a = 10b$, in each case including enough states to show an interesting feature or two. Describe those features in some detail, comparing the two cases to each other and to the case $a = b$.

Problem 4.14. Consider a *three-dimensional* cube-shaped infinite square well, with width a in each dimension. Write down appropriate formulas for the energy eigenfunctions and eigenvalues, and draw an energy level diagram showing the low-lying levels and their degeneracies. What is the lowest level with a degeneracy other than 1, 3, or 6? (A “degeneracy” of 1 actually means that the level is not degenerate.)

Problem 4.15. Repeat Problem 4.9 for the case of a three-dimensional cube-shaped infinite square well. This time you should find that the density of states is *not* a constant function of energy, so sketch a graph of this function.

Problem 4.16. Consider a two-dimensional quantum harmonic oscillator, with the same natural frequency ω_c in both dimensions:

$$V(x, y) = \frac{1}{2}m\omega_c^2(x^2 + y^2).$$

Find a complete set of energy eigenfunctions for this system, and their corresponding eigenvalues. (Write the eigenfunctions in terms of Hermite polynomials, as in Section 2.4.) Draw an energy level diagram that shows degenerate basis states separately, with enough levels to make the pattern clear. Use a computer to plot a selection of the energy eigenfunctions, including some superpositions of the separable basis states.

4.2 Numerical methods for multiple dimensions

As you showed in Problem 4.6, some two-dimensional problems are *separable* into two one-dimensional problems. Whenever the two-dimensional potential energy function is simply a sum of one-dimensional potential energy functions,

$$V(x, y) = V_1(x) + V_2(y), \quad (4.19)$$

you can separate the TISE into independent equations for the x and y dimensions, and then solve these separately by the methods (analytical or numerical) of Chapters 2 and 3. If, in addition, the initial wavefunction $\psi(x, y)$ factors into a product of the form $f(x)g(y)$, then the wavefunction continues to have this separable form as time passes, with each factor evolving in time according to a one-dimensional TDSE.

Similar simplifications occur in three dimensions, and in Chapter 6 we'll see how to separate central-force problems in terms of spherical coordinates (or in two dimensions, polar coordinates).

The point I want to emphasize now is that whenever you *can* separate a multidimensional problem into two or more one-dimensional problems, you *should!* That's because one-dimensional problems are relatively easy, and multidimensional problems are inherently much more difficult. I don't know of *any* nonseparable multidimensional problems that can be solved analytically (probably there are a few contrived examples). So to solve nonseparable problems we're pretty much forced to use numerical methods—and these numerical solutions are much harder in multiple dimensions than in just one.

To get some idea of why numerical solutions are much harder in multiple dimensions, just consider the amount of computer memory required to store the wavefunction. For instance, if we discretize space into 100 grid locations in each dimension, then we can store a one-dimensional wavefunction as an array of 100 complex numbers. Storing a two-dimensional wavefunction at the same resolution requires $100^2 = 10,000$ complex numbers, and for a three-dimensional wavefunction this would become $100^3 = 1,000,000$ complex numbers! The time required for a computation grows in a similar way. More generally, the difficulty of a quantum problem grows *exponentially* with the number of dimensions.³ This is not true in classical mechanics, where the numerical difficulty of a problem is merely *proportional* to the number of dimensions.

Still, *difficult* doesn't mean *impossible*. Today's personal computers are sufficiently powerful that we can use them to solve a remarkable variety of quantum problems that are irreducibly multidimensional. This section describes several numerical methods for attacking these difficult problems.

³Richard Feynman famously made this point in “Simulating Physics with Computers,” *Intl. J. Theor. Phys.* **21**, 467–488 (1982), <https://doi.org/10.1007/BF02650179>.

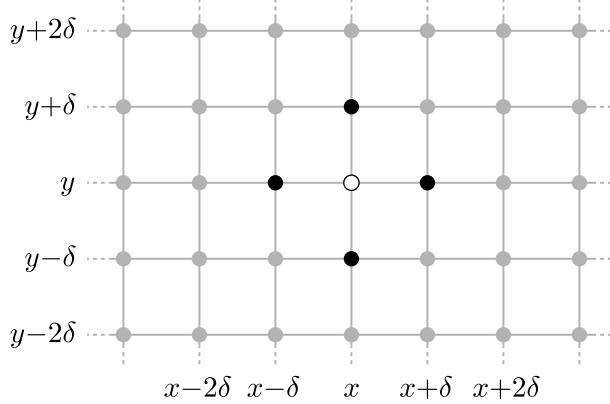


Figure 4.9: A square grid for numerically solving quantum problems in two dimensions. As shown in equation 4.21, the Laplacian operator ∇^2 compares $\psi(x, y)$ to the average value of ψ at the four nearest-neighbor points (highlighted in black).

Solving the TDSE

Let's start with time evolution. This is easy in principle, because the centered-difference method of solving the TDSE, described in Section 3.3, generalizes straightforwardly to multiple dimensions.

In two dimensions (to be specific), we can discretize space into a square grid, with grid spacing δ in both dimensions, as shown in Figure 4.9. Focusing on a particular grid location (x, y) , we approximate the second partial derivatives in the Hamiltonian operator (equation 4.15) using the second centered difference approximation (equation 2.6):

$$\begin{aligned}\frac{\partial^2 \psi}{\partial x^2} &\approx \frac{\psi(x+\delta, y) + \psi(x-\delta, y) - 2\psi(x, y)}{\delta^2}, \\ \frac{\partial^2 \psi}{\partial y^2} &\approx \frac{\psi(x, y+\delta) + \psi(x, y-\delta) - 2\psi(x, y)}{\delta^2}.\end{aligned}\tag{4.20}$$

Adding these equations gives us an approximation for the two-dimensional Laplacian:

$$\nabla^2 \psi \approx \frac{4}{\delta^2} (\bar{\psi}_{nn} - \psi(x, y)),\tag{4.21}$$

where $\bar{\psi}_{nn}$ is an abbreviation for the average of ψ at the four nearest-neighbor grid locations. Thus, the Laplacian operator tells us how different $\psi(x, y)$ is from the average value of ψ at the nearest-neighbor points. (If you've studied electrostatics in some detail, you may recall that the Laplacian of the electric potential in empty space is zero, implying that the potential at any point is *equal* to the average of its values at neighboring points.)

Combining approximation 4.21 for $\nabla^2 \psi$ with the centered difference approxima-

tion 3.25 for $\partial\psi/\partial t$, the two-dimensional TDSE becomes

$$i \frac{\psi(x, y, t+dt) - \psi(x, y, t-dt)}{2 dt} \approx -\frac{2}{\delta^2} (\bar{\psi}_{nn}(t) - \psi(x, y, t)) + V(x, y)\psi(x, y, t), \quad (4.22)$$

in units where $\hbar = m = 1$. Notice that everything on the right-hand side is evaluated at the “current” time t , while the left-hand side involves the future time $t+dt$ and the previous time $t-dt$. Solving for $\psi(x, y, t+dt)$ gives an explicit formula for the future value in terms of current and previous values, just as in the one-dimensional case described in Section 3.3.

Everything else about implementing the two-dimensional centered-difference algorithm is also analogous to the one-dimensional case, except that the correct choice of the time step dt isn’t obvious. It turns out that in two dimensions, dt/δ^2 must be somewhat less than 0.25, rather than 0.5 as in one dimension.

But I’m not going to show you any Mathematica code to solve the two-dimensional TDSE, because as of this writing, Mathematica on a personal computer is too slow to run such a simulation “while you wait.” Instead I’ve implemented the algorithm in JavaScript, which in today’s web browsers is many times faster. That simulation program is called Quantum Scattering in Two Dimensions, and is posted at <http://physics.weber.edu/schroeder/software/QuantumScattering2D.html>. The simulation uses a 400×400 grid, and displays the wavefunction as a color density plot as in the figures in the previous section. The user interface lets you choose from a variety of potential barriers and wells, and the initial state is a Gaussian wavepacket aimed toward the central region from the left.

Figure 4.10 shows some snapshots from three different runs of this simulation. I hope you can see from these examples that two-dimensional wavefunctions behave a lot like water waves, sound waves, or light waves. But please don’t settle for these still images! Run the simulation yourself to see the full animated processes, and a wider variety of scenarios. If you know just a little about JavaScript syntax (or Java or C, which are very similar), you can easily modify the simulation source code to use a potential function and initial state of your choice.

Problem 4.17. Generalize equations 4.21 and 4.22 to three spatial dimensions.

Problem 4.18. Use a ruler and protractor (or the electronic equivalent) to measure the hole spacing, the wavelength, and the approximate angles of the first-order interference maxima in the two-hole interference pattern shown in the middle-right image of Figure 4.10. Check that these quantities approximately satisfy the expected relationship. (Don’t expect perfect agreement, because the usual formula for two-slit interference is valid only very far away from the slits.)

Problem 4.19. Use the Quantum Scattering in Two Dimensions app to investigate single-hole diffraction. Set the hole size to 50 and the packet energy to 0.10, and use these values to predict the angle of the first minimum of the diffraction pattern. Run the simulation, use a protractor to measure the angle, and compare to your prediction. Then reduce the energy to 0.025 and repeat.

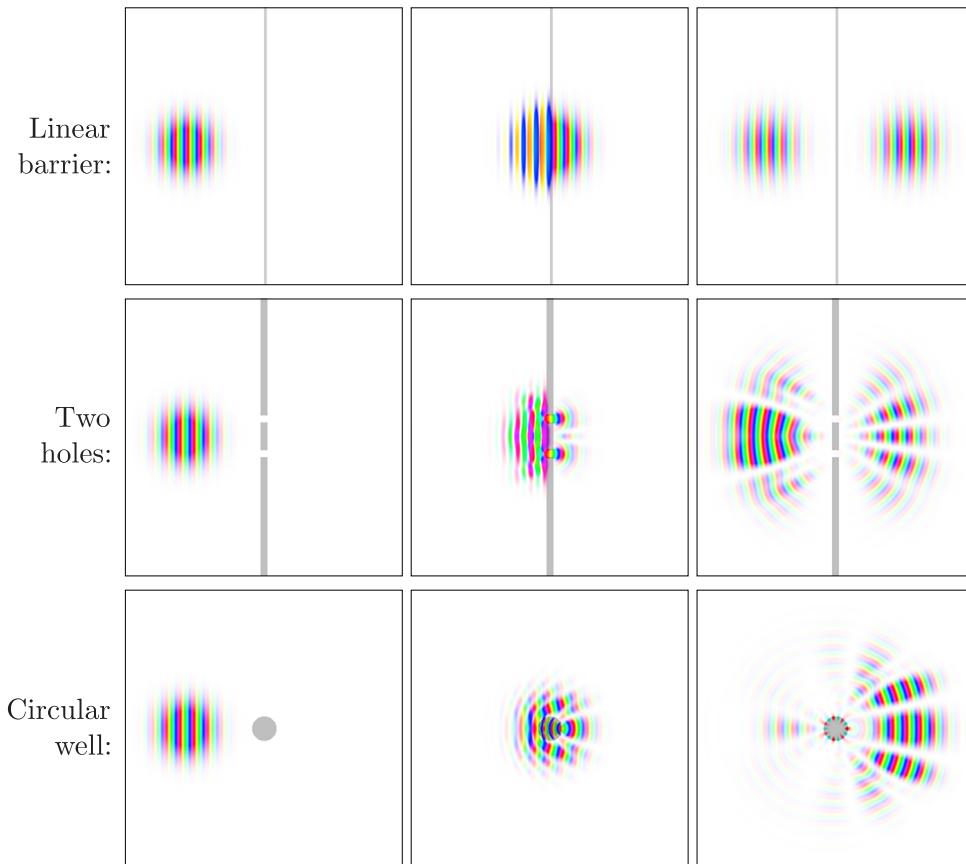


Figure 4.10: Snapshots from the Quantum Scattering in Two Dimensions simulation, showing a right-moving Gaussian wavepacket that encounters a potential barrier or well. Each row shows a sequence of three still images from a single simulation run (with the saturation adjusted independently for each image). Top row: A linear barrier whose potential is somewhat greater than the average wavepacket energy, but that is thin enough to give a large tunneling probability. Middle row: A linear barrier with a high enough potential to reflect the wavefunction, but with two holes, creating a two-source interference pattern on the right. Bottom row: A small circular well with a negative potential energy, which attracts the particle to produce an intricate interference pattern.

Solving the TISE on a grid

You might be wondering why I just explained how to numerically solve the two-dimensional time-*dependent* Schrödinger equation, before doing the same for the time-*independent* Schrödinger equation. After all, in one dimension it was easier to solve the latter—and it was much less computationally intensive, when we used the “shooting method” to integrate from one end of the region of interest across to the other.

The problem is that the shooting method does *not* generalize to more than one dimension! In multiple dimensions there are boundary conditions around all sides

of the region of interest, so there is no good place to shoot from, or to shoot to. We need a fundamentally different approach.

The approach that I'll describe in this subsection is to discretize space into a grid (as in the previous subsection) and start with a trial wavefunction on the grid that does *not* satisfy the TISE, but that *does* satisfy the boundary conditions. We'll specifically look at localized bound states on a two-dimensional square grid, for which the boundary conditions are simply that $\psi = 0$ along all four edges. In this case a reasonable starting function would be a product of one-bump sine waves (that is, the ground state of an infinite square well filling the region). We'll then gradually modify the trial wavefunction in a way that reduces its average energy. When this process converges, we'll have found the system's ground-state wavefunction. We can then use a slightly modified process to find the first excited state, then the second excited state, and so on.

There are a couple of different ways to accomplish this gradual modification of the trial wavefunction. One is an adaptation⁴ of the so-called *relaxation method* for solving Laplace's equation or Poisson's equation, which you may have studied in the context of electrostatics. Here I'll present an alternative method that is somewhat slower but more closely akin to the TDSE algorithm described in the previous subsection.

The trick is to evolve the trial wavefunction according to the partial differential equation

$$\frac{\partial \psi}{\partial \tau} = -\hat{H}\psi, \quad (4.23)$$

where \hat{H} is the usual Hamiltonian operator (equation 4.15 in two dimensions) and τ is a new variable that I'll call *fake time*.⁵ Some people call τ *imaginary time*, because this equation is identical to the TDSE except for the replacement $it \rightarrow \tau$. I claim that if we solve *this* differential equation, with our trial wavefunction as an “initial” condition, ψ will eventually converge to the system's ground-state wavefunction (a solution to the TISE).

To prove this claim, imagine expanding $\psi(\tau)$ (the solution to equation 4.23) in terms of the (unknown) energy eigenfunctions ψ_n :

$$\psi(\tau) = c_1(\tau)\psi_1 + c_2(\tau)\psi_2 + \dots \quad (4.24)$$

Such an expansion must exist, because the energy eigenfunctions always form a complete set. Plugging this expansion into equation 4.23 gives

$$\frac{\partial c_1}{\partial \tau}\psi_1 + \frac{\partial c_2}{\partial \tau}\psi_2 + \dots = -(E_1 c_1 \psi_1 + E_2 c_2 \psi_2 + \dots), \quad (4.25)$$

where each E_n is the energy eigenvalue that corresponds to ψ_n . But the ψ_n functions are mutually orthogonal (or in the case of degenerate eigenfunctions, we can specify

⁴D. V. Schroeder, “The variational-relaxation method for finding quantum bound states,” Am. J. Phys. **85**(9), 698–704 (2017), <https://doi.org/10.1119/1.4997165>.

⁵I learned both the term and the method from Steven E. Koonin and Dawn C. Meredith, *Computational Physics: Fortran Version* (Addison-Wesley, Reading, MA, 1990), Section 7.4.

that we're using an orthogonal basis), so we can project-out any one of them using Fourier's trick. That is, we take the inner product of both sides of equation 4.25 with a particular ψ_n , and this kills off all the rest of the terms in the equation, leaving us with simply $\partial c_n / \partial \tau = -E_n c_n$, whose solution is

$$c_n(\tau) = c_n(0) \cdot e^{-E_n \tau}. \quad (4.26)$$

In other words, the coefficients c_n decay at exponential rates that depend on their corresponding energies. The high-energy terms in the expansion 4.24 decay away most quickly, and the ground-state term decays most slowly. (If the ground-state energy is negative, then its coefficient grows the most quickly.) As $\tau \rightarrow \infty$, every term in the expansion becomes exponentially negligible in comparison to the ground state, so ψ becomes the ground state.

To actually carry out the fake time evolution, we need to discretize equation 4.23. For the right-hand side we can use exactly the same discretization as for the actual TDSE (equation 4.22 in two dimensions). To discretize the τ derivative, though, it turns out that we can't use the centered-difference approximation that we used for the actual TDSE, because in the fake-time case this would give a discrete equation that is unstable no matter how small the time step! (By *unstable* I mean that tiny round-off errors would grow exponentially over time, to give a huge and spurious “solution.”) But the fix is easy: We simply use the forward-time approximation in equation 3.24:

$$\frac{\partial \psi}{\partial \tau} \approx \frac{\psi(x, y, \tau + d\tau) - \psi(x, y, \tau)}{d\tau}. \quad (4.27)$$

This approximation was unsuitable for the actual TDSE because it doesn't conserve probability—but that's not a concern here, where we're intentionally making the wavefunction decay away over “time.”

With this approximation for the τ derivative, the discretized fake-time TDSE in two dimensions is

$$\frac{\psi(x, y, \tau + d\tau) - \psi(x, y, \tau)}{d\tau} \approx \frac{2}{\delta^2} (\bar{\psi}_{nn}(\tau) - \psi(x, y, \tau)) - V(x, y)\psi(x, y, \tau). \quad (4.28)$$

Notice that we can solve this equation for the “future” ψ entirely in terms of the “current” ψ ; there is no need to store the “previous” ψ , as when we used the centered-difference approximation for $\partial \psi / \partial t$.

In a future edition of this book I will provide an explicit example of carrying out the fake-time algorithm to find the ground state of a nonseparable two-dimensional quantum system. It's feasible to do this in Mathematica, because in many cases we can get good results with a grid size of only (say) 50×50 , much smaller than what is needed for a useful TDSE simulation. Meanwhile, please check out the Quantum Bound States in Two Dimensions web app at <http://physics.weber.edu/schroeder/software/BoundStates2D.html>, which can solve a variety of problems of this type. The web app actually uses the faster algorithm cited in footnote 4, but the fake-time algorithm would give the same results and would be similar to watch in action.

		$n_x \longrightarrow$			
		1	2	3	4
n_y	1	1	2	3	4
	2	5	6	7	8
	3	9	10	11	12
	4	13	14	15	16

Figure 4.11: An arbitrary method of ordering basis functions in the matrix representation of the Hamiltonian. Each table entry is a row number (and also a column number) in the Hamiltonian matrix, while n_x and n_y are the quantum numbers that we ordinarily use to label the basis functions, e.g., in equation 4.17 for the square infinite square well basis. In this example the basis is truncated after $n_x = n_y = 4$, yielding a 16×16 matrix; in practice the cutoff should be considerably higher.

I still need to explain how to find other TISE solutions besides the ground state. The method is to find the ground state first, then start over with a new trial function that has a node (for example, one of the two-bump first excited states of the square infinite square well). This time, during the fake-time evolution, we periodically project-out any ground-state component from the trial wavefunction. This ensures that the trial function remains orthogonal to the ground state, so the lowest-energy state for it to converge to will be the first excited state. Once we find that, we can start over and project-out both the ground state and the first excited state in order to find the second excited state, and so on. The process may sound a bit cumbersome, but as long as we need only a few low-lying energy eigenfunctions, it works remarkably well.

The matrix method

The other method that we used in Chapter 2 to solve the one-dimensional TISE was matrix diagonalization. Unlike the shooting method, the matrix method *does* generalize pretty straightforwardly to multiple dimensions. And it's usually the method of choice if you need to find more than a handful of low-lying energy eigenfunctions.

Setting up the Hamiltonian matrix for a two-dimensional system is just a tad cumbersome, because each basis function is naturally labeled with *two* quantum numbers (n_x and n_y , for the number of “bumps” in each dimension), and we need to assign each such pair to a unique row (and corresponding column) of the matrix. Figure 4.11 shows a way to do this that's as simple as any, for an unrealistically small cutoff of 4 for the maximum number of bumps in either dimension. In practice you would almost always need to go up to at least 10 bumps, and in that case the matrix would have 100 rows (and 100 columns). Diagonalizing a matrix of that size, or even larger, is no problem at all for Mathematica, but calculating all the matrix elements by brute force, as we did in Section 2.3, can take quite a while. So in practice you would want to use the trick of Problem 2.17, which speeds up the calculation enormously but unfortunately makes the code more complicated.

Problem 4.20. Referring to Figure 4.11, find a formula for the matrix row number in terms of the quantum numbers n_x and n_y . Do this first for a cutoff at n_x or n_y equal to 4, and then for a general cutoff n_{\max} . Then find the formulas for going backwards; that is, given a matrix row number N , find the formulas for n_x and n_y . Use the “floor” function, denoted $\lfloor x \rfloor$ or in Mathematica, `Floor[x]`, which rounds a real number down to the nearest integer.

Problem 4.21. List some of the apparent advantages and disadvantages of the matrix diagonalization method, compared to the fake-time-evolution method of the previous subsection. What is the practical limit on the accuracy in each case? For what types of potentials would you choose one method or the other? Which one seems easier to code?

4.3 Multiple particles

I'll get right to the point:

A system of two particles has only one wavefunction.

Read that sentence aloud. Repeatedly. It takes some getting used to.

And it gets worse: A system of three particles, or four, or 10^{23} , also has only one wavefunction.

But let's start with just two particles, and say they're moving in just one dimension (to keep things as simple as possible). Then if x_1 represents a possible position of the first particle and x_2 represents a possible position of the second particle, the wavefunction is a function of two variables: $\psi(x_1, x_2)$. We say that ψ lives in a two-dimensional *configuration space* (see Figure 4.12). In fact, the collection of all allowed wavefunctions for this system is mathematically the same as for a *single* particle in *two* dimensions: Just change (x, y) to (x_1, x_2) . (The preceding statement is strictly true only if the two particles are somehow distinguishable from each other. If they are identical in every way then some wavefunctions are not allowed, as I'll explain at the end of this section.)

How do we interpret $\psi(x_1, x_2)$? Well, if you integrate its square modulus over some range of x_1 values *and* some range of x_2 values, you should get the probability of finding the first particle in the first range *and* the second particle in the second range (if you measure both of their positions):

$$\left(\begin{array}{l} \text{Probability of finding} \\ a < x_1 < b \text{ and } c < x_2 < d \end{array} \right) = \int_a^b dx_1 \int_c^d dx_2 |\psi(x_1, x_2)|^2. \quad (4.29)$$

This two-particle Born rule is completely analogous to equation 4.5 for a single particle in two dimensions. And with the analogy in mind, it is straightforward to write formulas for the normalization condition, for the momentum-space wavefunction $\tilde{\psi}(p_1, p_2)$, and for a general inner product (see equations 4.1, 4.8, and 4.9).

Of particular interest in the two-particle case is the probability of finding one particle in a certain interval, regardless of the position of the other particle. To find

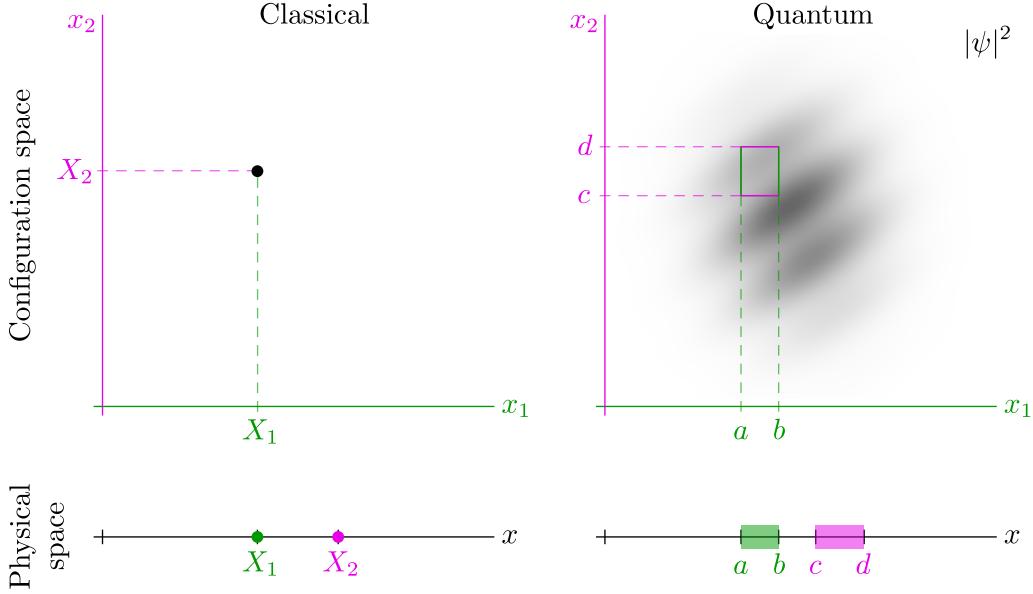


Figure 4.12: For a classical system of two particles in one dimension, a single point in configuration space (top left) tells us the locations of both particles in physical space (bottom left). For a quantum system of two particles in one dimension, integrating the squared magnitude of the wavefunction over any rectangular region in configuration space (top right) gives us the *probability* of finding the two particles within the corresponding intervals in physical space (bottom right).

this probability you need to integrate over *all* possible values of the other particle's coordinate:

$$\left(\text{Probability of finding } a < x_1 < b \text{ and any } x_2 \right) = \int_a^b dx_1 \int_{-\infty}^{\infty} dx_2 |\psi(x_1, x_2)|^2 = \int_a^b dx_1 P_1(x_1), \quad (4.30)$$

where the last expression defines the one-particle probability distribution

$$P_1(x_1) = \int_{-\infty}^{\infty} |\psi(x_1, x_2)|^2 dx_2, \quad (4.31)$$

in analogy with equations 4.6 and 4.7. (We can similarly define a function $P_2(x_2)$ by integrating $|\psi|^2$ over x_1 .) Moreover, if we go ahead and measure x_1 and obtain the result $x_1 = a$, then the wavefunction collapses to a delta function in the x_1 direction, multiplied by the original wavefunction evaluated at $x_1 = a$:

$$\psi_{\text{new}}(x_1, x_2) \propto \psi_{\text{old}}(a, x_2) \delta(x_1 - a), \quad (4.32)$$

precisely analogous to equation 4.11

We can also construct example wavefunctions that correspond to those in Section 4.1. If both particles have well-defined positions, then their wavefunction is just

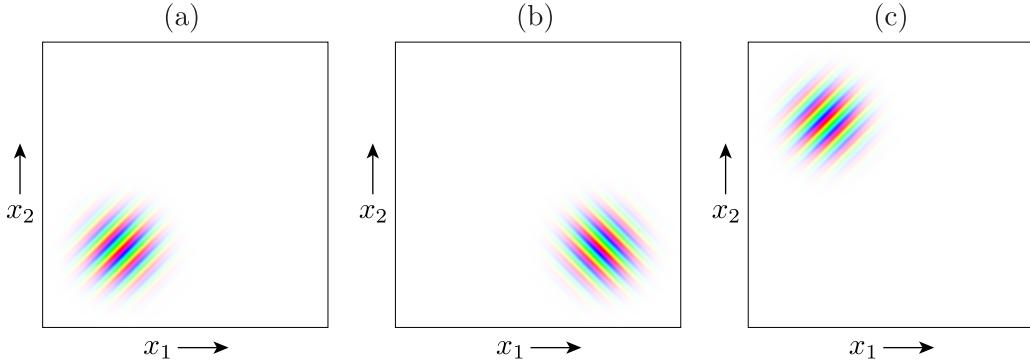


Figure 4.13: Some possible wavefunctions for a system of two particles in one dimension, with each particle in a Gaussian wavepacket state. See Problem 4.22.

a product of delta functions, $\delta(x_1 - a)\delta(x_2 - b)$. If both particles have well-defined momenta, their wavefunction is a product of complex exponentials, $e^{ip_1x_1/\hbar}e^{ip_2x_2/\hbar}$. And if both particles are in Gaussian wavepacket states, then their wavefunction is a product of a Gaussian wavepacket function of x_1 and a Gaussian wavepacket function of x_2 . To visualize these states we can use the very same illustrations as in Figures 4.1 and 4.2, again with the replacement $(x, y) \rightarrow (x_1, x_2)$. You just have to keep reminding yourself that the two directions on the plot correspond to the locations of the two particles along a single axis in physical space, rather than two different directions in space.

Problem 4.22. Figure 4.13 shows some two-particle wavefunctions for which each particle is in a Gaussian wavepacket state. The coordinates x_1 and x_2 are measured from the same origin and are plotted at the same scale. For each of the three plots, draw a sketch in one-dimensional physical space showing the approximate location of each of the two particles, with an arrow indicating each particle's direction of motion. Label the two particles “1” and “2” in your sketches.

Nonseparable wavefunctions

The example wavefunctions in Problem 4.22, and those described in the paragraph preceding it, all have the special property that they can be *faktored* into a function of x_1 times a function of x_2 :

$$\psi(x_1, x_2) = f(x_1)g(x_2) \quad (\text{usually not possible}). \quad (4.33)$$

We say that such functions are *separable*, and in such cases you can get away with saying that each particle has its own wavefunction. However, as in the case of a single particle in two spatial dimensions, this property is the exception rather than the rule. The vast majority of all possible two-particle wavefunctions are not separable.

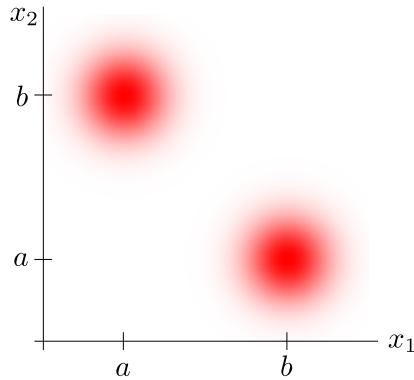


Figure 4.14: A two-peak “cat state” for a two-particle system, consisting of Gaussian shapes centered in configuration space at (a, b) and (b, a) . See equation 4.34.

One way to construct nonseparable wavefunctions is to combine two or more separable wavefunctions into a superposition. For example, consider the superposition of Gaussians

$$e^{-(x_1-a)^2} e^{-(x_2-b)^2} + e^{-(x_1-b)^2} e^{-(x_2-a)^2}, \quad (4.34)$$

where x_1 , x_2 , a , and b are understood to be measured as multiples of some suitable unit. Each of the two terms in this expression is separable, but the entire expression (which is an equally legitimate wavefunction) is not. If the difference $b - a$ is a few units, then this function looks something like Figure 4.14. This is a “cat state” in two-dimensional configuration space.

The nonseparability of most two-particle wavefunctions isn’t a mere mathematical curiosity: it has the important *physical* consequence that measurements performed on the two particles are *correlated*. For instance, if you were to measure the position of either of the particles in the cat state shown in Figure 4.14, you would have a 50/50 chance of obtaining a value either near a or near b . But if x_1 is near a then x_2 must be near b , and vice-versa. To put it more vividly, suppose you measure the position of particle 1 first, and happen to find it near a . This measurement collapses the wavefunction to lie entirely within the peak around $x_1 = a$ and $x_2 = b$, so you can now be certain that a subsequent measurement of the position of particle 2 will yield a result near b . The measurement of x_1 not only provides information you didn’t previously have about x_2 ; it actually *changes* the probability distribution for the subsequent x_2 measurement.

Whenever a two-particle wavefunction is not separable, we say that the two particles are *entangled*.⁶ Then it’s impossible to accurately describe the state of one particle without also referring to the other. Separate measurements performed

⁶Erwin Schrödinger coined this term in 1935, but it didn’t come into common use until the 1990s. See D. V. Schroeder, “Entanglement isn’t just for spin,” Am. J. Phys. **85**(11), 812–820 (2017), <https://doi.org/10.1119/1.5003808>. Significant portions of Sections 4.1 and 4.3 are based on this article.

on two entangled particles are correlated, so that a measurement of one alters the probability distribution for a measurement on the other.

Ironically, even though the vast majority of allowed multi-particle wavefunctions are entangled, we'll spend quite a bit of time working with wavefunctions that are separable, that is, not entangled. This is because separable wavefunctions are mathematically simpler, and also because they are *complete*: any other wavefunction can always be expressed as a linear superposition of separable wavefunctions (as in equation 4.34). Just don't let these separable wavefunctions give you the false impression that the quantum world is always so simple. In quantum mechanics you need to keep reminding yourself that your basis vectors are merely a basis, from which you can build arbitrary linear combinations to represent the unlimited variety of the world.

You may be wondering what physically *causes* particles to become entangled, or to remain unentangled, in the real world. To answer that question we need to take a look at two-particle quantum dynamics.

Problem 4.23. When you look at Figure 4.14, it's tempting to assume that each of the two peaks of the wavefunction represents one of the two particles. Why is this assumption wrong? What *does* each of the two peaks represent? Explain carefully.

The two-particle Hamiltonian

The operators for position and momentum in a two-particle system are exactly what you would expect: $\hat{x}_1 = x_1$, $\hat{p}_1 = -i\hbar\partial/\partial x_1$, and so on. So let's move right along and talk about the Hamiltonian operator, which governs the system's dynamics.

The Hamiltonian of a two-particle system is, as usual, the operator that corresponds to the system's total energy. So it consists of a kinetic energy term for each particle, plus an arbitrary potential energy term that can depend on both coordinates:

$$\hat{H} = -\frac{\hbar^2}{2m_1} \frac{\partial^2}{\partial x_1^2} - \frac{\hbar^2}{2m_2} \frac{\partial^2}{\partial x_2^2} + V(x_1, x_2), \quad (4.35)$$

where m_1 and m_2 are the masses of the two particles. The presence of two different masses makes this expression look a little more complicated than for a single particle in two spatial dimensions, but it doesn't add any actual mathematical complexity because the masses can (if we wish) be absorbed into the units that we use to measure x_1 and x_2 .

Moreover, we can usually simplify the two-particle potential energy function by writing it as a sum of three pieces:

$$V(x_1, x_2) = V_1(x_1) + V_2(x_2) + V_{12}(x_2 - x_1). \quad (4.36)$$

The first two terms on the right-hand side are from *external* forces acting on each of the two particles, while the third term is from the interaction *between* the particles. When all three terms are present, the Hamiltonian and the associated Schrödinger equations (time-dependent and -independent) are still pretty complicated. But when one or more terms is absent, things simplify.

First consider the important special case of *noninteracting* particles, for which $V_{12} = 0$. Then the results of Problem 4.6 (with the substitution $(x, y) \rightarrow (x_1, x_2)$) apply, and for most purposes we effectively have two single-particle systems that we can treat separately. More precisely, we know in this case that (a) there exists a complete set of separable solutions to the TISE; and (b) if the system starts out in a separable (unentangled) state, the TDSE predicts that it will remain in one.

The other important special case is when $V_1 = V_2 = 0$ in equation 4.36, meaning that there are no external forces acting on either particle. Then it is useful to rewrite the Hamiltonian in terms of center-of-mass and relative coordinates:

$$x_{\text{cm}} = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}, \quad x_r = x_2 - x_1. \quad (4.37)$$

In terms of these coordinates, the system effectively separates into a free “particle” of mass $m_1 + m_2$ whose coordinate is x_{cm} , and a second “particle” whose mass is the *reduced mass*, $m_1 m_2 / (m_1 + m_2)$, with coordinate x_r , subject to the interaction potential $V_{12}(x_r)$. These two fictitious “particles” will have a complete set of unentangled energy eigenstates, and will remain unentangled so long as they start out that way, even while the two actual particles interact and, typically, become entangled during the process.

Two particles in a box

Now recall from Section 4.1 the example of a single particle in a two-dimensional square infinite square well. With the substitution $(x, y) \rightarrow (x_1, x_2)$, we can translate that example to apply instead to a system of *two* particles trapped in a *one-dimensional* infinite square well (a one-dimensional “box”). To make the analogy precise we need to assume that the two particles interact only with the walls of the box and not with each other, so that V_{12} in equation 4.36 is zero. One way to make sure the particles don’t interact would be to put them in *separate* one-dimensional boxes, measuring x_1 and x_2 from the left edge of each of the respective boxes. Alternatively, the particles could be in the same box but of types that just don’t exert any significant forces on each other.

Assuming that the two particles don’t interact, the separable energy eigenfunctions are again products of sine waves (eigenfunctions for a one-dimensional box), as written in equation 4.17 and illustrated in Figure 4.6. If the two particles have the same mass (but are still distinguishable in some other way, for reasons I’ll explain later), then the energy levels are again given by equation 4.18, as illustrated in Figure 4.7. (If the two particles have different masses, then the energy level structure is more complicated, but still straightforward to work out.)

Furthermore, for two particles in a one-dimensional box we can again construct nonseparable superposition states such as those shown in Figure 4.8. These are states in which the two particles are entangled. Now, just because such states *exist* doesn’t necessarily mean they will actually *occur*. For instance, if the two particles are in separate boxes and don’t interact even indirectly, then there is no mechanism to put them into an entangled state.

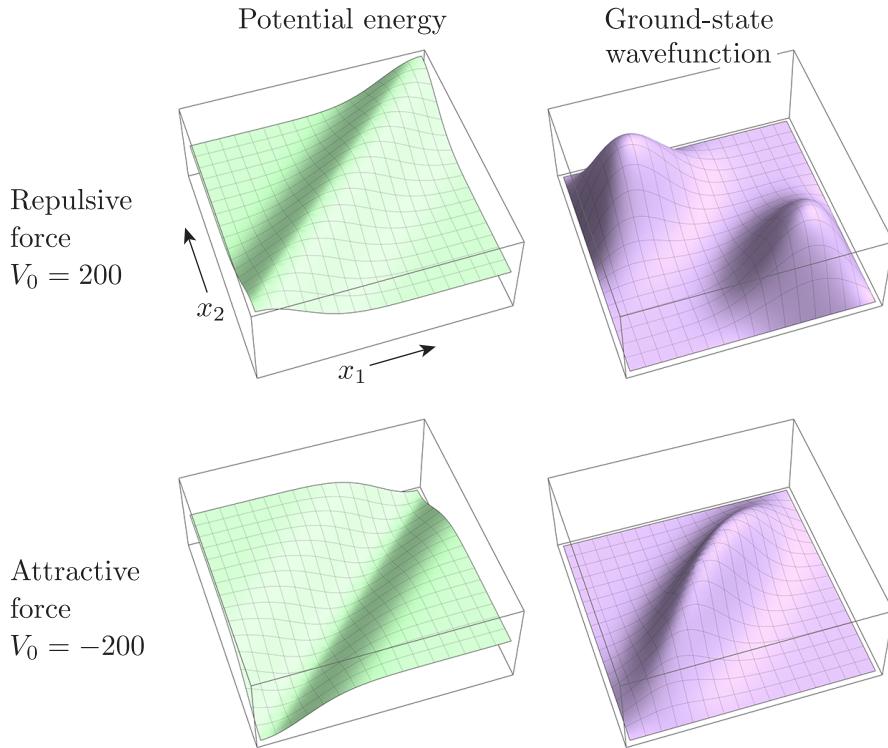


Figure 4.15: Potential energy functions (left) and ground-state wavefunctions (right), plotted in configuration space, for a system of two equal-mass particles in a one-dimensional box, interacting with each other via a Gaussian potential of the form of equation 4.38. The potential range b is equal to 1/4 the box width, and the potential strength V_0 is measured in units in which \hbar , the particle masses, and the box width are all equal to 1. (The infinite potential energy outside the box is not shown.) I calculated the wavefunctions using the fake-time method described in Section 4.2, but no calculations are needed to understand their qualitative shapes.

On the other hand, suppose that the two particles are in the same box, and that they do interact with each other. The exact form of the interaction isn't critical for the point I want to make, but to be specific, let's say the interaction potential is a smooth Gaussian,

$$V_{12}(x_2 - x_1) = V_0 e^{-(x_2 - x_1)^2/b^2}, \quad (4.38)$$

where V_0 and b are constants that determine the strength and the range of the interaction. The force between the two particles is repulsive if V_0 is positive and attractive if V_0 is negative. In configuration space, we can visualize this potential as a linear ridge (for $V_0 > 0$) or trough (for $V_0 < 0$) running along the main diagonal (see Figure 4.15).

In the repulsive case, the inter-particle potential divides the configuration-space box into a double well. The ground-state wavefunction therefore has two peaks, which become well separated for appropriate choices of V_0 and b , as shown in Fig-

ure 4.15) (top right). This two-particle wavefunction qualitatively resembles the entangled “cat state” shown in Figure 4.14. In physical space, it tells us that the particles will be found on opposite sides of the one-dimensional box, with no preference for either particle being on a particular side.

In the attractive case, the wavefunction gets pulled into the linear potential trough, as shown in Figure 4.15 (bottom right). This is also an entangled state, but this time it tells us that the two particles are close together instead of far apart.

I’ve shown only the ground-state wavefunctions in Figure 4.15, but in fact *all* of the energy eigenfunctions are entangled for this system of two trapped particles that interact with each other. The Entanglement in a Box web app, at <http://physics.weber.edu/schroeder/software/EntanglementInBox.html>, shows the four lowest-energy wavefunctions for this system, letting you adjust V_0 and b and watch how the wavefunctions depend on these parameters.

Although this example of two interacting particles in one dimension may seem artificially simple, it illustrates some of the same phenomena that arise in actual three-dimensional systems. For example, an ordinary hydrogen atom consists of two particles—a proton and an electron—that attract each other and are closely bound together even while the atom as a whole may be in less localized state. And in a helium atom, the two electrons repel each other even while both are confined by the stronger potential of the nucleus.

Problem 4.24. For the wavefunction shown in the bottom-right image of Figure 4.15, make a qualitatively accurate sketch of the single-particle probability densities $P_1(x_1)$ and $P_2(x_2)$. Then suppose that you measure x_1 and happen to obtain the result $x_1 = a/3$. Sketch the new probability density $P_2(x_2)$ that applies immediately after this measurement.

Problem 4.25. Sketch your best guess for what the first excited-state wavefunctions would look like for the systems shown in Figure 4.15, in both the repulsive and attractive cases. Then use the Entanglement in a Box web app to check your sketches.

Problem 4.26. If two equal-mass (but distinguishable) particles in a box do not interact, their second-lowest energy level is doubly degenerate, corresponding to either $\psi_{1,2}$ or $\psi_{2,1}$, or any linear combination of these, as illustrated in Figures 4.6 and 4.8 (with the substitution $(x, y) \rightarrow (x_1, x_2)$). Suppose, however, that the two particles interact weakly via a potential of the form of equation 4.38, with $|V_0|$ small and b less than the box width. Without doing any calculations, explain why this interaction will break the degeneracy. Determine which of the many mixtures of $\psi_{1,2}$ and $\psi_{2,1}$ are now the approximate energy eigenstates, and which of these eigenstates has the higher or lower energy. Treat the cases of repulsive and attractive interactions separately. (You can check your answers using the Entanglement in a Box web app.)

Two-particle scattering

It isn’t just energy eigenfunctions that become entangled when particles interact. The time-dependent Schrödinger equation also predicts that if two particles start out in a non-entangled state and then interact with each other, they will typically become entangled.

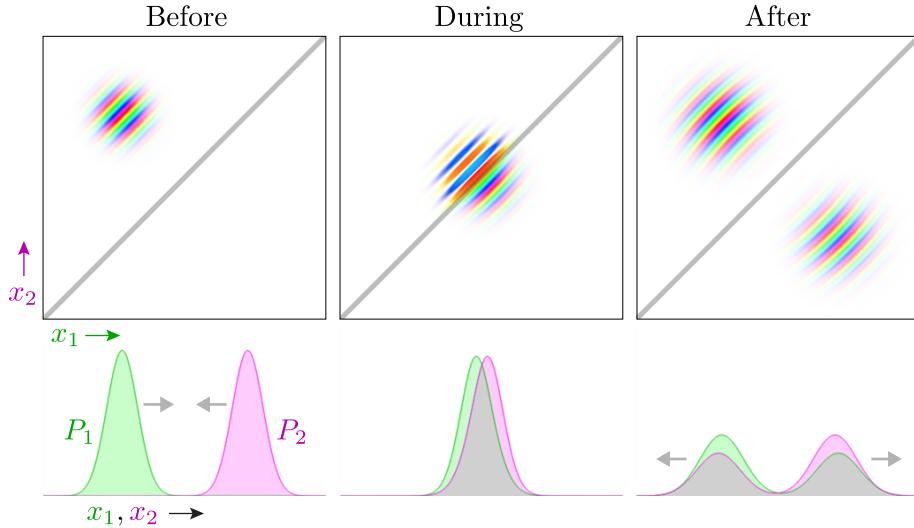


Figure 4.16: Scattering of two interacting particles in one dimension. The images at the top show the full wavefunction, plotted in configuration space, at three successive times. The gray diagonal band is the interaction potential, equation 4.39. Underneath are plotted the single-particle probability densities $P_1(x_1)$ (in green) and $P_2(x_2)$ (in magenta).

Figure 4.16 illustrates the simplest possible interaction of this type, in one spatial dimension. Each of the two particles is initially in a Gaussian wavepacket state, so their combined wavefunction (shown at the upper left) is separable. I've placed the two single-particle wavepackets at opposite sides of the region shown, and aimed them toward each other, with the same average speed. I've then assumed that the particles interact via a simple repulsive potential that is constant out to a certain distance and zero beyond that:

$$V(x_1, x_2) = \begin{cases} V_0 & \text{if } |x_2 - x_1| < b, \\ 0 & \text{otherwise.} \end{cases} \quad (4.39)$$

In configuration space this potential is a diagonal barrier, shown in gray in the figure. This scenario is mathematically equivalent to that in the top row of images in Figure 4.10—with the substitution $(x, y) \rightarrow (x_1, x_2)$ and a 45-degree rotation of the potential barrier and initial wavepacket orientation. I've adjusted the potential strength V_0 and range b to give a transmission probability close to 50% (actually a little less), so in the subsequent frames we see the expected interference pattern between the incident and reflected waves, and then the separation of the reflected and transmitted packets. The final state is entangled, similar to the two-peak cat state shown in Figure 4.14, but with the two peaks moving away from each other as time passes.

To create Figure 4.16 I used the Colliding Wavepackets web app at <http://physics.weber.edu/schroeder/software/CollidingPackets.html>, which simulates the time evolution of the two-particle wavefunction using the centered-difference

algorithm for the two-dimensional TDSE, as described in Section 4.2. By all means, run the app to see the full animation sequence, and adjust the parameters to vary the outcome of the scattering interaction.

As you can see, I didn't have to make the interaction between these two particles exotic or contrived in any way, in order to obtain an entangled final state. Entanglement is a generic outcome of interactions, so if we want to understand the quantum mechanics of particles that interact with each other, we will have to work with entangled states. Schrödinger himself emphasized this point when he coined the term "entangled" in 1935:

When two systems, of which we know the states by their respective representatives [i.e., wavefunctions], enter into temporary physical interaction due to known forces between them, and when after a time of mutual influence the systems separate again, then they can no longer be described in the same way as before, viz. by endowing each of them with a representative of its own. I would not call that *one* but rather *the* characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought. By the interaction the two representatives (or ψ -functions) have become entangled.⁷

Problem 4.27. Suppose that you wish to use a computer to simulate a system of n interacting particles in one dimension, using a spatial grid with a resolution of 100 grid locations (as I assumed in the discussion on page 119). If your computer uses eight bytes to store each complex number, what is the largest n for which it can store an arbitrary wavefunction for this system? (To answer this question you'll need to use a realistic value for the computer's total storage capacity; be sure to state this value.) Now suppose instead that you wish to simulate a single carbon atom, with six electrons in three-dimensional space, also at a resolution of 100 grid locations in each dimension. How much storage would you need for this system's wavefunction? Compare to the total capacity of all of Google's servers, which as of this writing is estimated to be on the order of 10 exabytes.

Identical particles

Quantum mechanics allows two (or more) particles to be *identical* in a sense that isn't really possible in classical mechanics. Consider what happens when two particles have all the same intrinsic properties (mass, electric charge, spin orientation, etc.), and in addition, their probability density is unchanged under the operation of interchanging its two arguments:

$$|\psi(x_2, x_1)|^2 = |\psi(x_1, x_2)|^2. \quad (4.40)$$

If this relation holds, then the two particles can't even be distinguished by their locations—or rather, by the probabilities of their being found at various locations—because the probability of finding particle 1 near $x = a$ and particle 2 near $x = b$

⁷E. Schrödinger, "Discussion of Probability Relations between Separated Systems," Math. Proc. Camb. Phil. Soc. **31**, 555–563 (1935), <https://doi.org/10.1017/S0305004100013554>.

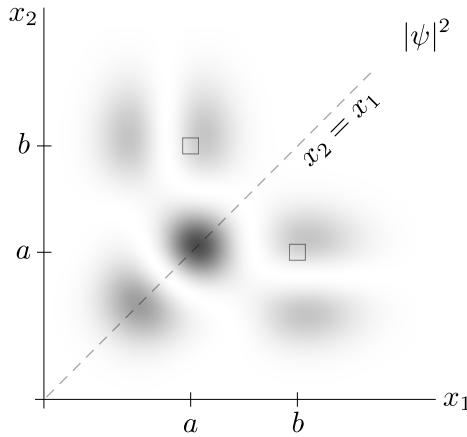


Figure 4.17: An example of a two-particle probability density $|\psi(x_1, x_2)|^2$ that is symmetric under the interchange of x_1 and x_2 , so its configuration space plot is symmetric under reflection through the diagonal line $x_2 = x_1$. This means we can't use the probability density function to distinguish the particles from each other; for instance, the probability of finding particle 1 near a and particle 2 near b (small upper square) is the same as the probability of finding particle 1 near b and particle 2 near a (small lower square).

is the same as that of finding particle 2 near $x = a$ and particle 1 near $x = b$, for any a and b (see Figure 4.17). This situation is fundamentally different from classical mechanics, in which you can always distinguish particles by their locations (if nothing else).

In terms of the wavefunction itself, there are actually two different ways⁸ that equation 4.40 can be realized:

$$\psi(x_2, x_1) = +\psi(x_1, x_2) \quad \text{or} \quad \psi(x_2, x_1) = -\psi(x_1, x_2). \quad (4.41)$$

In the first case the wavefunction is *symmetric* under interchanging x_1 and x_2 , while in the second case it is *antisymmetric*. Moreover, it's not hard to prove that if two particles are identical and either of equations 4.41 holds at one particular time, then the time-dependent Schrödinger equation will ensure that the same interchange relation holds at all other times. The symmetry or antisymmetry of the wavefunction for any two such particles is a permanent property of those particles, and it will never be meaningful to ask which one is which.

It seems that nature has taken advantage of both the symmetric and antisymmetric options, and filled the universe with particles that are truly identical to each other in one or the other of these two possible ways. Particles for which $\psi(x_2, x_1) = +\psi(x_1, x_2)$ are called *bosons* (after Satyendra Nath Bose, 1894–1974), and include photons, pions, helium-4 nuclei, and many other species of nuclei and atoms. Particles for which $\psi(x_2, x_1) = -\psi(x_1, x_2)$ are called *fermions* (after Enrico

⁸Equation 4.40 would also be satisfied if ψ changed by an arbitrary phase $e^{i\phi}$ when x_1 and x_2 are interchanged, but consistency requires that repeating the interchange operation restore the wavefunction to its original value, so ϕ can be only 0 or π .

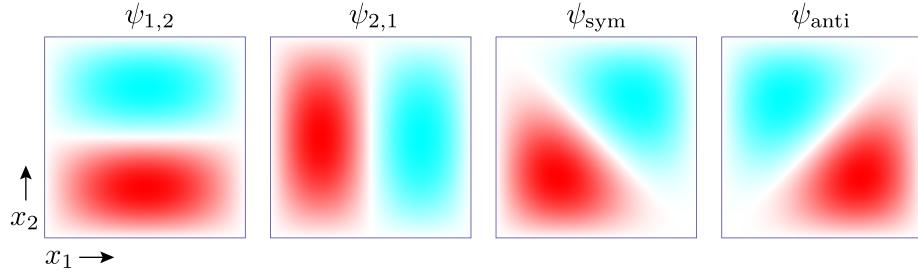


Figure 4.18: Four possible wavefunctions for a system of two particles in a one-dimensional box. In each of the four cases it is reasonable to say that one of the particles is in the single-particle state $\psi_1(x)$ and one of the particles is in the single-particle state $\psi_2(x)$. For states $\psi_{1,2}$ and $\psi_{2,1}$, we can also say *which* of the two particles is in each of these single-particle states, so the two particles are distinguishable. If the two particles are identical bosons, then only the symmetric state ψ_{sym} is allowed. If they are identical fermions, then only the antisymmetric state ψ_{anti} is allowed. For the states ψ_{sym} and ψ_{anti} it is not meaningful to ask which particle is in which single-particle state.

Fermi, 1901–1954), and include electrons, protons, neutrons, helium-3 nuclei, and many other species of nuclei and atoms. You can tell which particles are bosons and which are fermions by looking at their intrinsic spins, which I'll discuss further in Chapter 7.

Often, when describing a system of two identical particles, we start with a wavefunction $\psi_0(x_1, x_2)$ that's neither symmetric nor antisymmetric, and then we *symmetrize* or *antisymmetrize* it by adding or subtracting the same function with its arguments interchanged:

$$\begin{aligned}\psi_{\text{sym}}(x_1, x_2) &= A(\psi_0(x_1, x_2) + \psi_0(x_2, x_1)), \\ \psi_{\text{anti}}(x_1, x_2) &= B(\psi_0(x_1, x_2) - \psi_0(x_2, x_1)).\end{aligned}\quad (4.42)$$

Here A and B are normalization constants whose values depend on the function ψ_0 .

Figure 4.18 shows a simple example of these ideas, for a system of two particles in a one-dimensional box. If the particles are distinguishable, then two of the (many) possible system wavefunctions are

$$\begin{aligned}\psi_{1,2} &= \psi_1(x_1)\psi_2(x_2) = \frac{2}{a} \sin\left(\frac{\pi x_1}{a}\right) \sin\left(\frac{2\pi x_2}{a}\right), \\ \psi_{2,1} &= \psi_1(x_2)\psi_2(x_1) = \frac{2}{a} \sin\left(\frac{\pi x_2}{a}\right) \sin\left(\frac{2\pi x_1}{a}\right),\end{aligned}\quad (4.43)$$

where $\psi_1(x)$ and $\psi_2(x)$ are the usual single-particle energy eigenfunctions for a particle in a box of width a . Because $\psi_{1,2}$ and $\psi_{2,1}$ are separable, these system states separately describe the state of each particle. And for either $\psi_{1,2}$ or $\psi_{2,1}$ we can say *which* particle is in state $\psi_1(x)$ and *which* is in state $\psi_2(x)$. For instance, if particle 1 is a proton and particle 2 is an electron, then $\psi_{1,2}$ describes a state in

which the *proton* has the one-bump wavefunction $\psi_1(x)$ while the *electron* has the two-bump wavefunction $\psi_2(x)$.

If, on the other hand, the two particles are identical bosons, then neither $\psi_{1,2}$ nor $\psi_{2,1}$ is an allowed wavefunction because neither is symmetric under the interchange $x_1 \leftrightarrow x_2$. But the symmetrized combination $\psi_{\text{sym}} = A(\psi_{1,2} + \psi_{2,1})$, that is,

$$\psi_{\text{sym}} = A \left[\frac{2}{a} \sin\left(\frac{\pi x_1}{a}\right) \sin\left(\frac{2\pi x_2}{a}\right) + \frac{2}{a} \sin\left(\frac{\pi x_2}{a}\right) \sin\left(\frac{2\pi x_1}{a}\right) \right], \quad (4.44)$$

is allowed. Like $\psi_{1,2}$ and $\psi_{2,1}$, this wavefunction describes a state in which one of the particles is in state $\psi_1(x)$ and one of the particles is in state $\psi_2(x)$. The difference is that ψ_{sym} does not specify which of the particles is in which of these single-particle states.

Similarly, if the two particles are identical fermions, then neither $\psi_{1,2}$ nor $\psi_{2,1}$ nor ψ_{sym} is an allowed wavefunction because none of these is antisymmetric under the interchange $x_1 \leftrightarrow x_2$. But the antisymmetrized combination $\psi_{\text{sym}} = B(\psi_{1,2} - \psi_{2,1})$, that is,

$$\psi_{\text{anti}} = B \left[\frac{2}{a} \sin\left(\frac{\pi x_1}{a}\right) \sin\left(\frac{2\pi x_2}{a}\right) - \frac{2}{a} \sin\left(\frac{\pi x_2}{a}\right) \sin\left(\frac{2\pi x_1}{a}\right) \right], \quad (4.45)$$

is allowed. Again, this wavefunction describes a state in which one of the particles is in state $\psi_1(x)$ and one of the particles is in state $\psi_2(x)$, without specifying which of the particles is in which of these single-particle states.

An even simpler example would be if we try to put both particles in the *same* single-particle state:

$$\psi(x_1, x_2) = \psi_n(x_1)\psi_n(x_2), \quad (4.46)$$

where ψ_n is any single-particle wavefunction (not necessarily an energy eigenfunction). This two-particle wavefunction is already symmetric under $x_1 \leftrightarrow x_2$, so it could describe a system of two identical bosons (or a system of two distinguishable particles). But if we try to antisymmetrize it, by subtracting $\psi(x_2, x_1)$, we get zero, which is not a normalizable wavefunction. We can only conclude that it is *impossible* to put two identical fermions into the same single-particle state. This fact is called the *Pauli exclusion principle* (after Wolfgang Pauli, 1900–1958).

Problem 4.28. Work out the normalization constants A and B in equations 4.44 and 4.45. (Hint: We already know that the states $\psi_{1,2}$ and $\psi_{2,1}$ are normalized, so it's not necessary to carry out any integrals of sine functions.)

Problem 4.29. What are the energies of the four states depicted in Figure 4.18, assuming that these particles are nonrelativistic and do not interact with each other? (Express your answers in terms of the ground-state energy $E_1 = h^2/(8ma^2)$.) What can you say about the energies of the individual particles in each of the four cases?

Problem 4.30. Based on the four plots in Figure 4.18, discuss the relative probability of finding these two particles near each other, in each of the four cases. For which of these wavefunctions are the two particles closest together, on average? For which are they farthest apart, on average? How can you tell?

Problem 4.31. To quantify your answers to the previous problem, calculate the rms (root-mean-square) distance between the two particles for each of the four wavefunctions shown in Figure 4.18. To do this, you'll need to evaluate the integral

$$\langle (x_2 - x_1)^2 \rangle = \iint (x_2 - x_1)^2 |\psi(x_1, x_2)|^2 dx_1 dx_2, \quad (4.47)$$

which gives the average of the square of the distance, and then take the square root. (You can do these integrals with pencil and paper, but it's easier to use Mathematica. Numerical answers, expressed to two or three decimal places of precision, are fine.) Discuss briefly whether your answers are qualitatively reasonable.

Problem 4.32. Notice that the wavefunction ψ_{anti} in Figure 4.18 is zero along the diagonal line $x_2 = x_1$. What does this fact imply about the outcomes of position measurements made on the two particles? Does this property apply to all wavefunctions for systems of two identical fermions? Why or why not? Does this property ever apply to wavefunctions for systems of identical bosons, or for distinguishable particles? Explain your answers as clearly as you can.

Problem 4.33. Sketch four wavefunction density plots, analogous to those shown in Figure 4.18, for the case where one particle is in the single-particle state $\psi_1(x)$ while the other is in the *three-bump* single-particle state $\psi_3(x)$. Point out some of the notable features of your plots.

Problem 4.34. Consider a system of *three* particles in a one-dimensional box, with one particle in each of the single-particle states $\psi_1(x)$, $\psi_2(x)$, and $\psi_3(x)$. (a) If these particles are distinguishable, then one possible wavefunction for the system is $\psi_1(x_1)\psi_2(x_2)\psi_3(x_3)$. Write similar expressions for all the other such separable wavefunctions. (b) If these particles are identical bosons, then their wavefunction must be symmetric under any of the interchanges $x_1 \leftrightarrow x_2$, $x_2 \leftrightarrow x_3$, and $x_3 \leftrightarrow x_1$. What, then, is the fully symmetrized wavefunction for this system? (c) If these particles are identical fermions, then their wavefunction must be antisymmetric under any of the same interchanges. What, then, is the fully antisymmetrized wavefunction for this system?

4.4 Particles with internal structure

Our third and final generalization of quantum mechanics is to particles that have internal degrees of freedom. The quantum state of such a particle depends not only on variables that represent positions in space, but also on one or more internal variables.

The most familiar type of internal degree of freedom is *spin*, a particle's intrinsic angular momentum. But there are many other examples. Electrons within atoms and molecules can be internally excited above their ground states, and molecules can be in various rotational and vibrational states. Quarks and gluons each have several "color" states, and the quarks also have various "flavor" states (up, down, strange, charm, top, bottom), as do the leptons (electron, muon, tau, and three neutrino varieties). Even the familiar proton and neutron can sometimes be treated as two different states of a single "nucleon" system.

One common feature of all these internal properties is that we can enumerate the independent states in a discrete list. Although this list of independent states can sometimes be nominally infinite, the number of relevant states under any given conditions is usually finite and often small. In some of the most important cases there are just *two* independent states, for example, the two lowest-energy configurations of an ammonia molecule (as discussed in Section 2.5), or the spin-up and spin-down orientations of an electron along a given axis. (Any such two-state system is called a *qubit*.)

Because the number of independent internal states is usually small, it is convenient to represent these states as finite-dimensional vectors, and to represent the operators that act on them as matrices. I've already introduced this representation in Section 2.3 and Problems 2.31 and 2.32, but let me now summarize and generalize this formalism.

Matrix representation of states and operators

The idea is to pick a complete set of orthonormal states ϕ_n as our basis. We can then write the most general internal state as a linear combination:

$$\psi = c_1\phi_1 + c_2\phi_2 + \dots, \quad (4.48)$$

for some complex coefficients c_n that satisfy the normalization condition

$$\sum_n |c_n|^2 = 1 \quad (4.49)$$

(as in equation 1.45, except that here the sum is finite). If the list of basis states is understood, then to specify an arbitrary state we need only list the c_n values. For all practical purposes, then, this list of coefficients *is* the state of the system. We typically write the list as a column vector, so knowing the state of the system is equivalent to knowing the column vector of coefficients:

$$\psi \longleftrightarrow \begin{pmatrix} c_1 \\ c_2 \\ \vdots \end{pmatrix}. \quad (4.50)$$

When we express the state of a system in this column-vector form, we can also express all mathematical operations purely in terms of vector and matrix arithmetic. Addition, negation, subtraction, and scalar multiplication work the same as for any other types of vectors.

To compute inner products we need to be a little careful. For instance, if we express two system states ψ_c and ψ_d as

$$\psi_c = c_1\phi_1 + c_2\phi_2 + \dots, \quad \psi_d = d_1\phi_1 + d_2\phi_2 + \dots, \quad (4.51)$$

then their inner product is

$$\langle \psi_c | \psi_d \rangle = \sum_n c_n^* d_n, \quad (4.52)$$

with a complex conjugation on the components of the vector on the *left*, in analogy with equation 1.35 (see also Problem 2.31). We can also write this formula in matrix form, by transposing and conjugating the ψ_c column vector to make a row vector:

$$\langle \psi_c | \psi_d \rangle = (c_1^* \ c_2^* \ \dots) \begin{pmatrix} d_1 \\ d_2 \\ \vdots \end{pmatrix}. \quad (4.53)$$

Inner products are important in quantum mechanics because they appear in the *Born rule*, that is, in the formula for the probability of obtaining a particular measurement outcome. Suppose that the state ψ_d of equation 4.51 is the current system state, while ψ_c is an eigenstate, with eigenvalue λ , of some observable quantity that we measure. Then the probability of obtaining λ as our measurement outcome is the square magnitude of the inner product:

$$\left(\begin{array}{l} \text{Probability of obtaining} \\ \text{measurement result } \lambda \end{array} \right) = |\langle \psi_c | \psi_d \rangle|^2 = \left| (c_1^* \ c_2^* \ \dots) \begin{pmatrix} d_1 \\ d_2 \\ \vdots \end{pmatrix} \right|^2. \quad (4.54)$$

(If the eigenvalue λ is degenerate, then we need to sum this expression over all of its linearly independent eigenstates ψ_c .)

Furthermore, we can also express operator equations, including the Schrödinger equations, using matrix notation. Consider the action of an arbitrary linear operator \hat{A} acting on an arbitrary state ψ_c , resulting in some other state ψ_d :

$$\psi_d = \hat{A}\psi_c. \quad (4.55)$$

Inserting the expansions 4.51, this equation becomes

$$\sum_n d_n \phi_n = \sum_n c_n \hat{A} \phi_n, \quad (4.56)$$

where I've used the linearity of \hat{A} to move it inside the sum and past the constant c_n . Now take the inner product with an arbitrary basis state ϕ_m , from the left (Fourier's trick!):

$$\sum_n d_n \langle \phi_m | \phi_n \rangle = \sum_n \langle \phi_m | \hat{A} \phi_n \rangle c_n. \quad (4.57)$$

Here I've used the linearity of the inner product to distribute it over the sums and factor out the constants c_n and d_n . The inner product on the left is now δ_{mn} , which kills off all terms in the sum except the one with $n = m$, for which it is 1. The inner product on the right is what we call the *mn matrix element* of \hat{A} :

$$\text{"Matrix element"} = A_{mn} = \langle \phi_m | \hat{A} \phi_n \rangle. \quad (4.58)$$

We are therefore left with the equation

$$d_m = \sum_n A_{mn} c_n, \quad (4.59)$$

or in matrix notation,

$$\begin{pmatrix} d_1 \\ d_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & \cdots \\ A_{21} & A_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \end{pmatrix}. \quad (4.60)$$

This matrix equation is completely equivalent to the original operator equation 4.55. And any other equation involving operators, perhaps with additional terms, scalar multiplications, and/or time derivatives, can be similarly rewritten in matrix representation. In summary, when we associate each state ψ with a column vector, we also associate each operator \hat{A} with the appropriate matrix,

$$\hat{A} \longleftrightarrow \begin{pmatrix} A_{11} & A_{12} & \cdots \\ A_{21} & A_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}, \quad (4.61)$$

and the action of the operator becomes an ordinary matrix multiplication.

You may have noticed that when I introduced the symbols ψ and ϕ_n at the beginning of this section, I called them “states” rather than “wavefunctions,” and I wrote simply ψ and ϕ_n , not $\psi(x)$ and $\phi_n(x)$. Sometimes we can express these internal states as functions of some internal position variable(s); for instance, the vibrational states of a diatomic molecule can be represented as functions of x , where x is the intermolecular separation distance (see Section 2.4). But in other cases, including the spin of fundamental particles such as electrons and photons, there is no continuous internal variable (as far as we know) to play the role of x in a “wavefunction” $\psi(x)$. One nice thing about the column vector representation of quantum states is that it works equally well whether or not we also could represent the states as wavefunctions. But I should confess that in the cases when we don’t have actual wavefunctions, I haven’t actually derived all of the vector-matrix formulas written above. For instance, I motivated the inner product formula, equation 4.52, through an analogy to the inner product of wavefunctions in terms of an integral. In fact, if the basis states are functions of x , then you can derive equation 4.52 from that more familiar integral formula. But when there are no actual wavefunctions, equation 4.52 becomes a *definition*, which cannot be derived. Similarly, in deriving the matrix equation 4.60 I had to *assume* that the operator \hat{A} is linear, whereas we can derive this fact for the specific operators introduced in earlier chapters, written in terms of x and d/dx . The deep lesson here is that the underlying algebraic rules for inner products and operators are more fundamental than any explicit representation of those rules in terms of either calculus or matrix arithmetic.

I should also emphasize that all vector components (c_n, d_n) and matrix elements (A_{mn}) are relative to a particular orthonormal basis. We’re always free to choose a different orthonormal basis, in which case the components and matrix elements will be different. This situation is analogous to the arbitrariness of the choice of coordinate axes in three-dimensional space: rotating the axes changes the vector

components, but doesn't change the equations that relate various vectors to each other according to the laws of physics.

Problem 4.35. If you haven't already worked Problems 2.31 and 2.32, go back and work them now.

Problem 4.36. Consider a particle whose internal quantum states can be represented by vectors in just two dimensions. In this problem you will work with the following vectors and matrices:

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad \beta_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \psi = \begin{pmatrix} \sqrt{1/3} \\ \sqrt{2/3} \end{pmatrix}.$$

- (a) What are the eigenvalues and normalized eigenvectors of A ? (Call the vectors α_1 and α_2 .)
- (b) Show that β_1 is an eigenvector of B . Normalize it, and find the corresponding eigenvalue.
- (c) Find and normalize the other eigenvector (call it β_2) of B (hint: it's orthogonal to β_1), and find its corresponding eigenvalue.
- (d) Make a large plot of all of these vectors (α_1 , α_2 , β_1 , β_2 , and ψ) on a single set of axes, with their tails at the origin. As you work parts (e) through (g), illustrate the calculations on your plot.
- (e) Suppose this particle's current state is β_1 , and you measure the physical quantity whose corresponding matrix is A . What values might you obtain, and with what probabilities?
- (f) Suppose instead that this particle's current state is ψ , and you measure the physical quantity whose corresponding matrix is A . What values might you obtain, and with what probabilities?
- (g) Suppose again that this particle's current state is ψ , but that now you measure the physical quantity whose corresponding matrix is B . What values might you obtain, and with what probabilities?

Composite states

The column vectors of the previous subsection describe only the internal states of a particle. What if we also want to describe its motion through space? For instance, what if we want to describe both the motion of an electron and its spin?

Often it's sufficient to just specify the spatial state and the spin state separately. For instance, an electron in a hydrogen atom might have the spatial wavefunction $e^{-r/a}$ (where $r = \sqrt{x^2 + y^2 + z^2}$ and a is a constant) and the spin state $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ (conventionally indicating spin up along the z axis). More generally, there is an infinite-dimensional space of allowed spatial wavefunctions, and for each of these, there is a two-dimensional space of allowed spin states. (For a particle with a larger spin magnitude than the electron's, the spin states would live in a higher-dimensional space, as discussed in Chapter 7.)

But as you might guess by now, these separately-specified wavefunctions and spin states are really special cases. We must also allow for the possibility that the

electron is in a superposition of these separable states, and most superpositions are not separable. In fact, the most general superposition (to describe an electron's motion and spin) has the form

$$\psi = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \alpha(\vec{r}) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \beta(\vec{r}) = \begin{pmatrix} \alpha(\vec{r}) \\ \beta(\vec{r}) \end{pmatrix}, \quad (4.62)$$

where $\alpha(\vec{r})$ and $\beta(\vec{r})$ are arbitrary complex-valued functions that satisfy the normalization condition

$$\int (|\alpha(\vec{r})|^2 + |\beta(\vec{r})|^2) d^3r = 1. \quad (4.63)$$

When the functions $\alpha(\vec{r})$ and $\beta(\vec{r})$ are equal up to a constant factor, the overall state is separable and we can describe the spatial state and the spin separately. Otherwise, though, the electron's spin state is “entangled” with its spatial state. For instance, it's possible in principle to put an electron into a state that is a 50-50 mixture of spin-up in Uruguay and spin-down in Denmark. (To actually put an electron into a state with space and spin entangled, we would subject it to a spin-dependent force.)

I hope you can see the analogy between these composite motion-spin states and the two-particle states described in Section 4.3. In each case, the space of all allowed states consists of all possible separable states, constructed as products of the states of the subsystems, *together with all normalized linear combinations of these product states*. We often call this idea the *principle of superposition*: If ψ_1 and ψ_2 are both allowed states of some quantum system, then any normalized superposition of the form $c_1\psi_1 + c_2\psi_2$ is also an allowed state.⁹

I've been using spin and spatial motion as examples, but the principle of superposition applies to all of the degrees of freedom, spatial and internal, of any system of one or more quantum particles. To conclude this section, let me now show how to represent arbitrary states involving two different internal degrees of freedom, or any other two degrees of freedom for which we can represent the states using small column vectors.

To be specific, consider the internal states of a single quark, which has two independent spin states (just like an electron), and also has three independent “color” states. We can represent its spin state using a two-component column vector, and its color state using a three-component column vector:

$$\text{Spin state} \longleftrightarrow \begin{pmatrix} c_+ \\ c_- \end{pmatrix}, \quad \text{Color state} \longleftrightarrow \begin{pmatrix} c_r \\ c_g \\ c_b \end{pmatrix}. \quad (4.64)$$

Here + and – indicate the “spin-up” and “spin-down” components of the state, while r , g , and b represent the “red,” “green,” and “blue” components. (This

⁹This “principle of superposition” is more fundamental than the dynamical principle of superposition, mentioned in Section 1.3, which says that a superposition state evolves in time as if each of its component terms were evolving separately.

“color” terminology is entirely metaphorical, based on a rather imperfect analogy to the way our eyes perceive three primary colors.)

Each of the quark’s independent spin states can accompany each of its independent color states, so we can group these states in pairs to obtain *six* independent internal states for this system: $+r$, $+g$, $+b$, $-r$, $-g$, and $-b$. Every normalized superposition of these six basis states is allowed, so to represent an arbitrary internal state we can use a six-dimensional column vector. In the special case in which the state is separable, we can compute this column vector as

$$\begin{pmatrix} c_+ \\ c_- \end{pmatrix} \otimes \begin{pmatrix} c_r \\ c_g \\ c_b \end{pmatrix} = \begin{pmatrix} c_+c_r \\ c_+c_g \\ c_+c_b \\ c_-c_r \\ c_-c_g \\ c_-c_b \end{pmatrix}, \quad (4.65)$$

where the meaning of the symbol \otimes is defined by the right-hand side. This type of product is called a *tensor product*. (The ordering of the six components of the product vector is actually arbitrary, but it’s important that we be consistent, and the ordering I’ve used here is conventional. To remember the ordering, imagine “stretching” the first column vector to fill all six rows, then multiplying both the top half and the bottom half by the second column vector.)

For example, we would represent the state of a spin-down (along the z axis) red quark as

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad (4.66)$$

and we can similarly construct the five other basis states. We can also construct an infinite variety of other tensor-product states, such as

$$\begin{pmatrix} \sqrt{1/2} \\ -\sqrt{1/2} \end{pmatrix} \otimes \begin{pmatrix} \sqrt{1/3} \\ i\sqrt{2/3} \\ 0 \end{pmatrix} = \begin{pmatrix} \sqrt{1/6} \\ i\sqrt{1/3} \\ 0 \\ -\sqrt{1/6} \\ -i\sqrt{1/3} \\ 0 \end{pmatrix}. \quad (4.67)$$

This is a state in which a z -spin measurement would yield up or down with equal probability; a color measurement would yield red with probability $1/3$, green with probability $2/3$, and never blue; and, as with any tensor-product state, there would be no correlations between the spin outcomes and the color outcomes. (The factors of -1 and i don’t affect these particular probabilities, but they would affect probabilities for other possible measurements, such as spin along the x direction.)

But the vast majority of this system's six-component state vectors cannot be factored into tensor products. For instance,

$$\begin{pmatrix} \sqrt{1/3} \\ 0 \\ 0 \\ 0 \\ 0 \\ \sqrt{2/3} \end{pmatrix} \neq \begin{pmatrix} c_+ \\ c_- \end{pmatrix} \otimes \begin{pmatrix} c_r \\ c_g \\ c_b \end{pmatrix} \quad (\text{for any } c\text{'s}). \quad (4.68)$$

This state describes a quark that is either red and spin-up (with probability 1/3) or blue and spin-down (with probability 2/3). The spin and color are fully correlated, so if we prepare a quark in this state and then measure the color and get blue, we immediately know that the spin is down. The quark's color is entangled with its spin.

You might well be wondering how we would actually perform these spin and color measurements that I keep referring to. In Chapter 7 we'll study spin in detail, and I'll explain one way to measure the spin of a particle. Unfortunately, laboratory measurements of quark colors are indirect and beyond the scope of this book. But we could replace the color states in these examples with appropriately chosen states of orbital angular momentum (which is easier to measure), or with any other particle property whose measurement can yield just three possible outcomes.

In Chapter 7 I'll also show how to use tensor products of operator matrices to construct operators that act in the vector space of a composite, finite-dimensional system.

Problem 4.37. Evaluate each of the following tensor products:

$$(a) \begin{pmatrix} \sqrt{3/4} \\ -1/2 \end{pmatrix} \otimes \begin{pmatrix} \sqrt{1/2} \\ i\sqrt{1/2} \end{pmatrix}, \quad (b) \begin{pmatrix} 2 \\ -3 \end{pmatrix} \otimes \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \quad (c) \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \otimes \begin{pmatrix} 2 \\ -3 \end{pmatrix}.$$

Problem 4.38. Each of the six-component vectors below describes a quark's combined spin-plus-color state, using the basis of equation 4.65:

$$(a) \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad (b) \begin{pmatrix} \sqrt{1/20} \\ 0 \\ \sqrt{3/20} \\ -\sqrt{1/5} \\ 0 \\ -\sqrt{3/5} \end{pmatrix}, \quad (c) \begin{pmatrix} 0 \\ \sqrt{3/5} \\ -\sqrt{1/20} \\ 0 \\ \sqrt{1/5} \\ -\sqrt{3/20} \end{pmatrix}.$$

Interpret each of these vectors in terms of z -spin and color probabilities, as I did for the vectors in equations 4.66 through 4.68. Factor each vector into a tensor product of normalized spin and color vectors if possible; if this is not possible, prove it.

Problem 4.39. Prove that if the spin and color column vectors in equation 4.65 are separately normalized, then their tensor product, representing the overall (internal) state, is automatically normalized as well.

Problem 4.40. Evaluate the tensor product

$$\begin{pmatrix} \sqrt{1/2} \\ -\sqrt{1/2} \end{pmatrix} \otimes \begin{pmatrix} \sqrt{1/3} \\ i\sqrt{2/3} \end{pmatrix} \otimes \begin{pmatrix} 1/2 \\ \sqrt{3/4} \end{pmatrix}$$

for a three-qubit system, and check that you can carry out either of the two products first (so the tensor product is associative).

Problem 4.41. How large a column vector would you need to represent an arbitrary state of a system of four qubits? 10 qubits? 100 qubits? A *quantum computer* is one that uses entangled qubits (say 50 or 100 of them) for its calculations instead of ordinary classical bits. In principle, any quantum computer can be accurately simulated using a classical computer, by storing the N -qubit state as a list of complex numbers. Very roughly, how large can N be for such a simulation to remain feasible? Explain your reasoning.

Chapter 5

Working with operators

5.1 More about operators

According to the second principle of quantum mechanics (see Section 3.5), every observable quantity corresponds to a linear *operator* that acts in the appropriate space of functions (or vectors). So, for example, the most important operators for a structureless particle in one dimension are

$$x, \quad -i\hbar \frac{d}{dx}, \quad \text{and} \quad -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x), \quad (5.1)$$

which correspond to position, momentum, and energy, respectively. The eigenvalues of each such operator are the values that you can obtain when you measure the corresponding quantity, and the corresponding eigenvectors (or eigenfunctions) are the quantum states for which the measured quantity has each of those well-defined values. If you measure the quantity for some other quantum state, the Born rule tells us that the probability of getting any particular value is the square modulus of the component of the state vector along the direction of the corresponding eigenvector.

In order for this whole scheme to make sense, these eigenvalues must always be *real* numbers. In addition, the corresponding eigenvectors must form a complete, orthonormal basis (called an *eigenbasis*) for the space of state vectors (or functions), so that any other normalized vector in the space can be uniquely resolved into components whose square moduli (the probabilities) add up to 1. It's now time to investigate some of the mathematics of the operators that have these essential properties.

We say that an operator \hat{A} is *Hermitian* if, for all vectors ψ_1 and ψ_2 ,

$$\langle \hat{A}\psi_1 | \psi_2 \rangle = \langle \psi_1 | \hat{A}\psi_2 \rangle. \quad (5.2)$$

That is, a Hermitian operator can operate on *either* vector in an inner product with the same effect.

From this definition, it is easy to prove the following two theorems about Hermitian operators:

1. The eigenvalues of a Hermitian operator are real.
2. Eigenvectors of a Hermitian operator that have distinct eigenvalues are orthogonal to each other.

I suggest that you take a moment right now to try to prove each of these results, starting from the definition (5.2). You need to assume that the eigenvectors of your operator are normalizable, which is not the case when the eigenvalues are continuous (as they are for x and $-i\hbar d/dx$); the textbook by Griffiths has a nice discussion of how to handle continuous eigenvalues.

Moreover, we can also prove a theorem that is almost the converse of the two theorems above:

3. If a linear operator has a complete, orthonormal set of eigenvectors and a corresponding set of real eigenvalues, then it is Hermitian.

This theorem means that any operator that corresponds to an observable quantity in quantum mechanics must be Hermitian. Here is the proof.

Let's call the operator \hat{A} , its normalized eigenvectors α_n , and its eigenvalues a_n . Now consider the two arbitrary vectors ψ_1 and ψ_2 that appear in the inner product in equation 5.2. Because the set $\{\alpha_n\}$ is complete, we can expand each of these vectors in the α_n basis:

$$\psi_1 = \sum_n c_n \alpha_n, \quad \psi_2 = \sum_n d_n \alpha_n, \quad (5.3)$$

for some sets of complex coefficients $\{c_n\}$ and $\{d_n\}$. So let's insert these expansions into the left-hand side of equation 5.2:

$$\langle \hat{A}\psi_1 | \psi_2 \rangle = \left\langle \hat{A} \sum_n c_n \alpha_n \mid \sum_m d_m \alpha_m \right\rangle. \quad (5.4)$$

(Notice how I've used a different index, m , in the second sum, so I won't confuse it with the index in the first sum.) Because \hat{A} is linear, we can move it inside the first sum. And because the inner product obeys the algebraic rules that one would expect of a “complex dot-product,” we can move both of the sums, as well as the coefficients, outside of the inner product, picking up a $*$ on c_n , to obtain

$$\langle \hat{A}\psi_1 | \psi_2 \rangle = \sum_n \sum_m c_n^* d_m \langle \hat{A}\alpha_n | \alpha_m \rangle. \quad (5.5)$$

But α_n is an eigenvector of \hat{A} with real eigenvalue a_n , and the eigenvectors are orthonormal, so this becomes

$$\langle \hat{A}\psi_1 | \psi_2 \rangle = \sum_n \sum_m c_n^* d_m a_n \langle \alpha_n | \alpha_m \rangle = \sum_n \sum_m c_n^* d_m a_n \delta_{mn} = \sum_n c_n^* d_n a_n. \quad (5.6)$$

Through a completely analogous set of manipulations, you can show that the right-hand side of equation 5.2 reduces to exactly the same expression, and this completes the proof that \hat{A} is Hermitian.

Although only Hermitian operators can correspond to observable quantities in quantum mechanics, we do sometimes work with non-Hermitian operators. It's then useful to define the *adjoint*, \hat{A}^\dagger , of an operator \hat{A} to be the operator that has the same effect on the left side of an inner product that \hat{A} has on the right:

$$\langle \hat{A}^\dagger \psi_1 | \psi_2 \rangle = \langle \psi_1 | \hat{A} \psi_2 \rangle, \quad (5.7)$$

for any two vectors ψ_1 and ψ_2 . A Hermitian operator, then, is its own adjoint. A *unitary* operator is one whose adjoint is the same as its inverse: $\hat{U}^\dagger = \hat{U}^{-1}$, where $\hat{U}^{-1}\hat{U}$ is the identity operator.

Problem 5.1. Consider the rather trivial operator that consists of a simple constant c (which multiplies whatever vector it acts on). Under what condition is this operator Hermitian? Explain carefully.

Problem 5.2. Prove that, for the system consisting of a single particle in one dimension, the operators \hat{x} and $\hat{p} = -i\hbar d/dx$ are Hermitian.

Commuting and noncommuting operators

Often, in quantum mechanics, we will want to discuss two or more operators at the same time. For example, if two Hermitian operators \hat{A} and \hat{B} correspond to two observable quantities for a particular system, we might want to know whether these two quantities can be simultaneously well defined. As we shall later see, this question turns out to be closely related to the question of whether \hat{A} and \hat{B} *commute*, that is, whether their order matters when they operate, successively, on a vector:

$$\hat{A}\hat{B}\psi = \hat{A}(\hat{B}\psi) \stackrel{?}{=} \hat{B}(\hat{A}\psi) = \hat{B}\hat{A}\psi. \quad (5.8)$$

If $\hat{A}\hat{B}\psi = \hat{B}\hat{A}\psi$ for all ψ , then we say that operators \hat{A} and \hat{B} commute. And whether they commute or not, we define their *commutator*, denoted $[\hat{A}, \hat{B}]$, as the difference

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}, \quad (5.9)$$

where it is understood that both sides need to give the same result when acting on an arbitrary vector ψ . So if \hat{A} and \hat{B} commute, then their commutator is zero.

Problem 5.3. Evaluate the commutator $[\hat{x}, \hat{p}]$ of the position operator, \hat{x} , and the momentum operator, $\hat{p} = -i\hbar \partial/\partial x$.

Problem 5.4. Consider all the possible commutators of the position operators \hat{x} , \hat{y} , \hat{z} and the momentum operators \hat{p}_x , \hat{p}_y , and \hat{p}_z , in three dimensions. Which of these operators commute with each other and which don't? What are the commutators for the pairs that don't commute?

Problem 5.5. Prove that, for any three operators A , B , and C (I'm omitting the hats because they're getting cumbersome), (a) $[A + B, C] = [A, C] + [B, C]$, and (b) $[AB, C] = A[B, C] + [A, C]B$. (c) What are the analogous expansions of $[A, B + C]$ and $[A, BC]$?

Matrix representations

We've already seen in Sections 2.3 and 4.4 how to express operators in matrix form. With respect to a given set of basis vectors $\{\alpha_n\}$, any operator \hat{A} can be represented by the matrix whose elements are

$$A_{mn} = \langle \alpha_m | \hat{A} \alpha_n \rangle. \quad (5.10)$$

Representing operators as matrices in this way has the advantage of being concrete and vivid. For spin and other internal degrees of freedom for which the dimension of the vector space is finite and small, we normally express everything in matrix form from the start. But even when the dimension of the vector space is infinite, it's often useful to think of state vectors as columns of components, and to think of operators as matrices. Moreover, as we saw in Section 2.3, you can often get accurate results by keeping only a finite (and manageable) number of rows and columns.

From equation 5.10 you can prove the following further properties of operator-matrices:

- The adjoint matrix A^\dagger is the conjugate transpose, A^{T*} .
- A Hermitian matrix is its own conjugate transpose, so its real part is symmetric and its imaginary part is antisymmetric.
- The matrix for a Hermitian operator in its *own* eigenbasis is diagonal, with entries equal to its eigenvalues.

I suggest that you now spend a few minutes trying to prove each of these statements.

5.2 Ladder operators for the harmonic oscillator

Operators in quantum mechanics aren't merely a convenient way to keep track of eigenvalues (measurement outcomes) and eigenvectors (definite-value states). We can also use them to streamline calculations, stripping away unneeded calculus (or explicit matrix manipulations) and focusing on the essential algebra.

As an example, let's now go back to the one-dimensional harmonic oscillator, and use operator algebra to find the energy levels and associated eigenfunctions.

Recall that the harmonic oscillator Hamiltonian is

$$\hat{H} = \frac{1}{2m}\hat{p}^2 + \frac{1}{2}m\omega_c^2x^2, \quad (5.11)$$

where \hat{p} is the momentum operator, $\hat{p} = -i\hbar d/dx$. As in Section 2.4, it's easiest to use natural units in which we set the particle mass m , the classical oscillation frequency ω_c , and \hbar all equal to 1. Then the Hamiltonian is simply

$$\hat{H} = \frac{1}{2}\hat{p}^2 + \frac{1}{2}x^2, \quad (5.12)$$

with $\hat{p} = -id/dx$.

The basic trick, which I have no idea how to motivate, is to define two new operators that are linear combinations of x and \hat{p} :

$$\hat{a}_- = \frac{1}{\sqrt{2}}(x + i\hat{p}), \quad \hat{a}_+ = \frac{1}{\sqrt{2}}(x - i\hat{p}). \quad (5.13)$$

These are called the *lowering* and *raising* operators, respectively, for reasons that will soon become apparent. Unlike x and \hat{p} and all the other operators we've worked with so far, the lowering and raising operators are not Hermitian and do not represent any observable quantities. We're not especially concerned with their eigenvalues or eigenfunctions; instead we'll focus on how they mathematically convert one *energy* eigenfunction into another. (Note: I'm going to drop the cumbersome $\hat{\cdot}$ notation from here on. Also, most authors use the notation a and a^\dagger instead of a_- and a_+ , but the book by Griffiths uses the $-/+$ notation, and I too think it's more appropriate. In conventional units, by the way, we would insert factors of m , ω_c , and \hbar into these definitions as needed to make a_- and a_+ dimensionless. It's a good exercise to figure out exactly how to do this, especially if you're even a little uncomfortable with my use of natural units.)

The first thing to notice about a_- and a_+ is that if you multiply them together, you almost get the Hamiltonian:

$$\begin{aligned} a_- a_+ &= \frac{1}{2}(x^2 + p^2 - i(xp - px)) \\ &= H - \frac{i}{2}[x, p] \\ &= H + \frac{1}{2}, \end{aligned} \quad (5.14)$$

where I've used the basic commutator relation $[x, p] = i\hbar$ (with $\hbar = 1$) that you worked out in Problem 5.3. Similarly, multiplying them together in the other order gives $a_+ a_- = H - \frac{1}{2}$. We can therefore write the Hamiltonian as

$$H = a_- a_+ - \frac{1}{2} = a_+ a_- + \frac{1}{2}. \quad (5.15)$$

Now suppose that ψ is any energy eigenfunction, so that $H\psi = E\psi$ for some E . Then I claim that if we act on ψ with either the raising or the lowering operator, we get *another* energy eigenfunction. To prove this, I need to check what happens when the Hamiltonian operator acts on this new function. Here's what happens in the case of $a_+\psi$:

$$\begin{aligned} H(a_+\psi) &= (a_+ a_- + \frac{1}{2})(a_+\psi) \\ &= a_+(a_- a_+ + \frac{1}{2})\psi \\ &= a_+(H + 1)\psi \\ &= a_+(E + 1)\psi \\ &= (E + 1)(a_+\psi). \end{aligned} \quad (5.16)$$

So $a_+\psi$ is an energy eigenfunction, and its eigenvalue is exactly one unit greater than that of ψ itself. (Remember that our unit of energy is $\hbar\omega_c$.) This is why a_+ is called the raising operator: it mathematically *raises* any energy eigenstate ψ up the quantum ladder by one rung, to the next-highest energy eigenstate (see Figure 5.1). There's no limit to how many times we can apply the raising operator, so this proves that a quantum harmonic oscillator has an infinite ladder of energy eigenstates, with equally spaced levels separated in energy by $\hbar\omega_c$.

By a completely analogous calculation (see Problem 5.6) you can show that a_- lowers any energy eigenstate ψ by one rung down the quantum ladder:

$$H(a_-\psi) = (E - 1)(a_-\psi). \quad (5.17)$$

This is why a_- is called the lowering operator. This result would seem to imply that our infinite ladder of energy eigenstates continues *downward* in energy without limit—but that can't possibly be the case, because there can't be any states with negative energy when the potential energy is positive everywhere. The only way out is if, when a_- acts on the lowest-energy eigenstate (call it ψ_0), it gives zero:

$$a_-\psi_0 = 0. \quad (5.18)$$

Then equation 5.17 can still be true for ψ_0 even though there's no lower-energy state. And from this simple equation we immediately see that $H\psi_0 = \frac{1}{2}\psi_0$, so the ground-state energy is half a unit, or $\frac{1}{2}\hbar\omega_c$. The remaining energies are integer steps above this, so we've proved that the energies of a quantum harmonic oscillator are as claimed in Section 2.4:

$$E_n = (n + \frac{1}{2})\hbar\omega_c, \quad n = 1, 2, 3, \dots \quad (5.19)$$

(If you're reading carefully you may have noticed a couple of loopholes in the logic I just described. How do we know the quantum ladder goes up indefinitely, rather than ending with a state ψ_{\max} for which $a_+\psi_{\max} = 0$? And how do we know that the ladder that starts with ψ_0 includes *all* the states? Could there be others

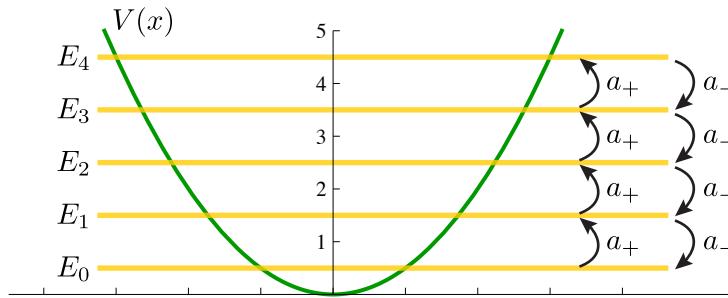


Figure 5.1: When the raising operator a_+ acts on a harmonic oscillator energy eigenstate, the result is the next eigenstate up on the quantum ladder. When the lowering operator a_- acts on an energy eigenstate, the result is the next eigenstate down on the ladder.

that we've missed, perhaps lying in between the ones we've found? These loopholes are actually pretty easy to close, and I'd encourage you to think about them, but only after you're comfortable with the rest of the calculations and reasoning described here.)

To find the energy eigenfunctions that correspond to the eigenvalues E_n , we still need to use calculus—but at least the method is intuitive and straightforward. To find the ground-state eigenfunction we can use equation 5.18, which becomes an ordinary differential equation for $\psi_0(x)$ when we express the lowering operator in terms of the explicit momentum operator, $p = -id/dx$. The solution to the differential equation is $\psi_0(x) = Ae^{-x^2/2}$, just as we saw in Section 2.4 (see Problem 5.7). Then, to find the first excited state, just apply the raising operator, also written in terms of $p = -id/dx$, to the ground state (see Problem 5.8). Keep applying the raising operator to work your way up the quantum ladder until the novelty wears off. As you might guess, it gets pretty tedious to work out more than the first few eigenfunctions by hand.

I hope you agree that the ladder-operator method is by far the most elegant way of solving the TISE for the harmonic oscillator. The bad news, though, is that no such elegant method exists for solving the TISE for other one-dimensional potential functions; the method worked here only because the Hamiltonian is quadratic in both p and x , allowing it to be factored, aside from an additive constant, into the product $a_- a_+$. So why spend time learning a method—however elegant—whose applicability is so restricted? There are two reasons.

The first reason is that harmonic oscillators really are ubiquitous in nature, from vibrating molecules to elastic solids to the electromagnetic field and other fundamental fields of elementary particle physics. In continuous vibrating systems we refer to the units of excitation energy as *particles*, such as photons in the case of the electromagnetic field, or as *quasiparticles*, such as phonons in the case of elastic vibrations. The study of these continuous vibrating quantum systems is called *quantum field theory*, and ladder operators are a fundamental tool of quantum field theorists. But I won't digress to explore quantum field theory in this book.

The second reason, though, is that ladder operators *will* come up again in this book in a somewhat different context: angular momentum. Instead of adding and removing energy, the ladder operators in that case will add and remove units of angular momentum (along the z axis). They will therefore be an extremely useful tool in our study of systems with spherical symmetry, especially atoms.

Problem 5.6. Show that if ψ is a harmonic oscillator energy eigenfunction with energy E , then when the Hamiltonian operator H acts on the state $(a_- \psi)$, the result is an energy eigenfunction with energy $E-1$ (in natural units).

Problem 5.7. Write equation 5.18 explicitly as a differential equation for $\psi_0(x)$, and solve it. Also work out the normalization constant for $\psi_0(x)$.

Problem 5.8. Starting with the harmonic oscillator ground-state wavefunction $\psi_0(x)$, apply the raising operator repeatedly to find the first three excited states, $\psi_1(x)$, $\psi_2(x)$, and $\psi_3(x)$. Feel free to use natural units, and don't worry about normalization.

Problem 5.9. The calculation in equation 5.16 shows that $a_+ \psi_n = c_+ \psi_{n+1}$ for some constant c_+ , and Problem 5.6 shows that $a_- \psi_n = c_- \psi_{n-1}$ for some constant c_- . In this problem you will determine the constants c_+ and c_- .

- (a) First prove that a_- and a_+ are adjoints of each other: $a_-^\dagger = a_+$ and $a_+^\dagger = a_-$. (You may assume that the operators x and p are Hermitian, as you proved in Problem 5.2.)
- (b) Now evaluate the inner product $\langle \psi_n | H \psi_n \rangle$ in three different ways: first by letting H act directly on ψ_n , then by expressing H in terms of $a_+ a_-$, and finally by expressing H in terms of $a_- a_+$. In each of the last two cases you can move the leftmost ladder operator to the other side of the inner product by converting it to its adjoint. Then compare the three results to determine c_+ and c_- . There is an undetermined phase factor $e^{i\phi}$, but you can set it to 1 (that is, take c_+ and c_- to be real and positive) without harm, because the phase factor doesn't affect the normalization of ψ_n . You should end up with the relations

$$a_- \psi_n = \sqrt{n} \psi_{n-1}, \quad a_+ \psi_n = \sqrt{n+1} \psi_{n+1}. \quad (5.20)$$

- (c) Look up (or recall) the explicit expressions for the normalized harmonic oscillator energy eigenstates $\psi_2(x)$ and $\psi_3(x)$, then check the results of part (b) by applying the raising operator, expressed in terms of d/dx , to $\psi_2(x)$, and the lowering operator, similarly expressed, to $\psi_3(x)$.

Problem 5.10. For the simple harmonic oscillator, using the energy eigenfunctions as a basis, write each of the following operators explicitly in matrix form, showing enough rows and columns of each matrix to make the patterns clear: H , a_- , a_+ , x , and p . (I suggest doing them in the order listed. Notice that you can express x and p as linear combinations of a_- and a_+ .)

5.3 Compatible and incompatible observables

Suppose you measure the position of a quantum particle, and then you measure its momentum, and then you measure its position again. Chances are that your second position measurement will be quite different from your first, because the intervening momentum measurement put the particle into a momentum eigenstate, spread out widely in space, effectively erasing all memory of the outcome of your first position measurement. We therefore say that position and momentum are *incompatible observables*.

An example of *compatible* observables would be momentum and kinetic energy: measuring one of these quantities will have no effect on subsequent measurements of the other. In general, two observables are compatible if you can measure one, then measure the other, then measure the first again, and be *guaranteed* of getting the same result in the final measurement that you got in the first one. In Chapter 6 we'll see that the magnitude of a particle's angular momentum is compatible with any one component of its angular momentum, and that both of these are compatible with energy whenever a particle is subject to a spherically symmetric potential energy function.

Mathematically, there are two ways to characterize whether observables are compatible (or incompatible). First, we can ask whether the operators for the two

observables possess a common eigenbasis (that is, a common set of eigenvectors that form a complete basis). Second, we can ask whether the operators themselves commute with each other. In summary, the following three statements are all equivalent:

1. A and B are compatible observables.
2. The A and B operators possess a common eigenbasis.
3. The A and B operators commute.

I won't try to write out a mathematician's proof of these equivalencies, but it's not hard to understand the basic ideas. I'll first explain why (1) and (2) are equivalent, then why (2) and (3) are equivalent.

Suppose, first, that A and B are compatible observables, and imagine measuring A first, then B , and then A again. The first measurement puts the system into one of the A eigenstates, and the second measurement (of B) must leave the system in an A eigenstate with the same eigenvalue, so that the final measurement is sure to yield the same outcome as the first. But this means that this A eigenstate is also a B eigenstate. In the simplest cases this final eigenstate will be the same one that the initial measurement of A put the system into. There's a complication, though, if the obtained value of A is degenerate, so that more than one linearly independent eigenvector of A has this eigenvalue. Then if B lacks this degeneracy, the B measurement will usually change the state to a *different* vector in the subspace of A -degenerate eigenstates. In any case, however, we can find *some* set of basis vectors that are eigenvectors of *both* A and B .

Conversely, if A and B have the same eigenvectors, then measuring A puts the system into an eigenstate of both, so a subsequent measurement of B doesn't change this state and therefore doesn't affect a subsequent measurement of A . Degeneracy again complicates this picture somewhat, but doesn't alter the conclusion that the two observables must be compatible.

Once we know that A and B have a common eigenbasis, we can see that they must commute by applying them in succession to an arbitrary vector ψ that is expanded in terms of this common eigenbasis. If the basis vectors are $\{\alpha_n\}$ and the expansion coefficients are $\{c_n\}$, then

$$AB\psi = AB \sum_n c_n \alpha_n = \sum_n c_n AB \alpha_n = \sum_n c_n a_n b_n \alpha_n, \quad (5.21)$$

where a_n and b_n are the associated eigenvalues of A and B , respectively. But applying A and B to ψ in the other order (that is, $BA\psi$) would give exactly the same expression, and therefore A and B commute.

Conversely, if we assume that A and B commute, then we can prove that if α is an eigenvector of A with eigenvalue a , then $B\alpha$ is also an eigenvector of A with the same eigenvalue:

$$A(B\alpha) = B(A\alpha) = B(a\alpha) = a(B\alpha). \quad (5.22)$$

If the eigenvalue a is nondegenerate, then this means that $B\alpha$ must be proportional to α itself, so α is also an eigenvector of B . In the degenerate case the vector $B\alpha$ could lie along some different direction in the subspace of degenerate eigenvectors of A , but there must always be a set of basis vectors in this subspace that are also eigenvectors of B .¹

Problem 5.11. Consider a quantum system with two observable quantities F and G , represented by the following operator matrices:

$$F = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad G = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (5.23)$$

- (a) Show that $[F, G] = 0$, and conclude that these two observables must be compatible.
- (b) What are the eigenvalues and eigenvectors of F ? (Be careful here, since one of the eigenvalues is degenerate.)
- (c) Which of the eigenvectors of F are also eigenvectors of G ? (A bit of trial and error should suffice to answer this question.)
- (d) Suppose that this system is initially in the state $\psi = (\sqrt{1/2}, 0, \sqrt{1/2})$. You then measure F , and obtain 2 as a result. What is the state vector after your measurement? (Hint: See Figure 4.4 and the accompanying discussion.)
- (e) Continuing the experiment described in part (d), suppose that you now measure G . What values might you obtain, and what will be the system's state vector after your measurement in each case?
- (f) Still continuing this experiment, suppose that you now measure F again. Will you obtain the same result (2) as before? Will this measurement cause any further change to the state vector? Explain.

The uncertainty principle

When two observables are incompatible, we might still wonder *how much* a measurement of one of them interferes with measuring the other. After all, in the classical limit there is no such thing as incompatibility: measurements don't affect the state of the system at all.

To quantify the degree of incompatibility, however, it's useful to rephrase the question. Instead of considering a succession of measurements (A then B then A again), let's ask whether we can find at least *some* states for which A and B are both *approximately* well defined. More precisely, imagine preparing a large number of identical systems in the same state, then measuring A for half of these systems and B for the other half. Each set of measurements will have some average value, $\langle A \rangle$ or $\langle B \rangle$, and will also have some amount of spread about the average, which we can characterize by the standard deviation, σ_A or σ_B . We say that a quantity is approximately well defined if its standard deviation is small.

¹For a proof of this statement, see F. Mandl, *Quantum Mechanics* (Wiley, Chichester, 1992), Section 3.1.

There is never any limit on how small we can make σ_A or σ_B : either could be zero, if the state in question is an eigenstate of that observable. And if A and B are compatible, then we can find simultaneous eigenstates for which *both* σ_A and σ_B are zero. But when A and B are incompatible, there is generally a limit on how small we can simultaneously make σ_A and σ_B . Specifically, one can prove that the product $\sigma_A\sigma_B$ must obey the inequality

$$\sigma_A\sigma_B \geq \left| \frac{1}{2i} \langle [A, B] \rangle \right|. \quad (5.24)$$

This is the most general version of the famous *uncertainty principle*. It says that there is *no* state for which the product of standard deviations is smaller than the right-hand side. Notice that the right-hand-side involves the average value of the commutator of the operators A and B ; for compatible observables, of course, this commutator would be zero so there would be no constraint on $\sigma_A\sigma_B$.

The proof of the generalized uncertainty principle is rather technical, so I won't present it here. You can find it in most traditional textbooks.²

For the special case of position and momentum (in one dimension), the commutator is simply $i\hbar$ (see Problem 5.3) so the generalized uncertainty principle reduces to

$$\sigma_x\sigma_p \geq \frac{\hbar}{2}. \quad (5.25)$$

This is the original uncertainty principle that we all know and love. For macroscopic objects, the smallness of \hbar implies that the uncertainty principle puts no practical constraint on precision with which position and momentum can be simultaneously well defined.

5.4 The principles of quantum mechanics

In Section 3.5 I laid out the five principles of quantum mechanics as applied to a single, structureless particle in one dimension. Now is a good time to pause and restate these principles for an arbitrary quantum system.

Principle 1: The states of a system correspond to vectors in a vector space.

In one-dimensional wave mechanics, these “vectors” are normalizable complex-valued functions of a single variable. The space of all such functions is infinite dimensional, since a complete basis for it has infinitely many basis functions. But the states of other quantum systems can live in even larger vector spaces, or in smaller ones. In many cases it suffices to use a finite-dimensional vector space, in which we can represent the vectors by simply listing their components.

²See, for example, Griffiths, Section 3.5.

In all cases, though, the vector space is complex, in the sense that the scalars, and the vector components, can be complex numbers.

If one vector is a scalar multiple of another, then those vectors represent the same physical state.

The essential property of a vector space is that any linear combination of two of its vectors is also a vector in the space. This means that in principle, we can put a composite system into an arbitrary entangled state.

Although any state vector is *allowed*, this doesn't mean that every state vector is easy to realize in the laboratory. Entangled states of a composite system sometimes arise naturally, but in other cases they require specialized preparation that may be difficult or even impractical.

Principle 2: Observable quantities correspond to Hermitian operators.

The eigenvalues of a Hermitian operator are real numbers, and these correspond to the possible outcomes when you measure the corresponding observable quantity.

The corresponding eigenvectors, in turn, correspond to states of the system for which this observable quantity is well defined. Eigenvectors with distinct eigenvalues are always orthogonal to each other.

If two observable quantities are compatible then their corresponding operators commute, and these operators have a common eigenbasis.

Principle 3: Measurement probabilities are given by the Born rule.

Suppose that a system's normalized state vector is ψ , and we measure an observable quantity whose operator is \hat{A} , with eigenvalues a_i and corresponding normalized eigenvectors α_i . Then the possible outcomes of the measurement are the various eigenvalues a_i , but in general we can predict only the *probability* of obtaining each of these outcomes. When the eigenvalues are discrete, the probability of obtaining a particular eigenvalue a_i equals the square of the *component* of ψ along the corresponding eigenvector α_i :

$$\text{(Probability of } a_i) = |\langle \alpha_i | \psi \rangle|^2. \quad (5.26)$$

This equation is called the *Born rule*. In the special case where ψ is one of the eigenstates α_i , the inner product equals 1, giving a probability of 100%, for obtaining that particular outcome a_i , while the probability of getting any other outcome is zero, because that corresponding eigenvector will be orthogonal to ψ (inner product equal to zero).

If two or more orthogonal eigenvectors are degenerate, with the same eigenvalue a_i , then on the right-hand side of equation 5.26 we must sum over all of these eigenvectors, as in equation 4.10.

When the eigenvalues are continuous, the probability of obtaining any one of them is infinitesimal, so we have to integrate this squared magnitude over some range of desired values as in equation 1.6 for position and equation 3.10 for momentum.

Principle 4: A measurement causes the state to collapse.

Whatever the result of a measurement, the measurement process causes the system's state to abruptly "collapse" to become whatever eigenstate corresponds to the value obtained. This means that if you immediately measure the same physical quantity again, you will always get the same result that you got the first time.

If the measured value is degenerate, with more than one orthogonal eigenstate, then the measurement causes the state to collapse to its projection in the degenerate subspace, as illustrated in Figure 4.4.

Principle 5: The TDSE governs time evolution.

As time passes, the state vector changes smoothly according to the time-dependent Schrödinger equation (TDSE),

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H}\psi, \quad (5.27)$$

where \hat{H} is the Hamiltonian operator, that is, the operator that corresponds to the system's total energy. If we can find the eigenfunctions of the Hamiltonian (by solving the time-independent Schrödinger equation, $\hat{H}\psi = E\psi$), then we can expand any initial ψ in terms of them and write down the solution to the TDSE by inserting the corresponding wiggle factors, as in equation 1.24. In the special case in which the initial ψ is an energy eigenfunction, the TDSE tells us that its time dependence is a simple wiggle factor, $e^{-iEt/\hbar}$, and therefore all measurement probabilities are independent of time.

Chapter 6

Central forces

6.1 Spherical coordinates

We now turn to the important special case of a single quantum particle in three dimensions, subject to a potential energy function V that is spherically symmetric—that is, V depends only on the distance r from some fixed center. It is natural in this case to use spherical coordinates (see Figure 6.1), in which case we can write

$$V(r, \theta, \phi) = V(r) \quad (\text{independent of } \theta \text{ and } \phi). \quad (6.1)$$

The most important example of a spherically symmetric potential is the attractive Coulomb potential, proportional to $-1/r$, between an atomic nucleus (which we'll treat as a fixed center of force) and an electron. But there are other interesting central potentials as well, and we can get pretty far in our analysis without assuming a specific formula for $V(r)$.

As in Chapter 2, our goal is to solve the time-independent Schrödinger equation to find the allowed energy levels and the associated wavefunctions. Once we

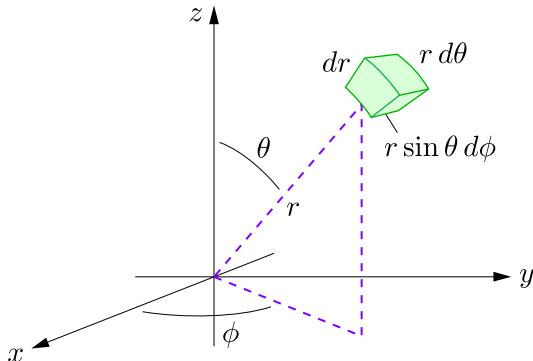


Figure 6.1: Definitions of the spherical coordinates r , θ , and ϕ . Note that θ is the polar angle. Also shown is an infinitesimal volume element whose dimensions are dr , $r d\theta$, and $r \sin \theta d\phi$.

have these, it is a straightforward exercise to combine these wavefunctions with their associated wiggle factors to build solutions to the time-dependent Schrödinger equation.

Even though V depends only on r , we'll need to consider wavefunctions that also depend on θ and ϕ . So let me pause a moment to summarize some basic facts about spherical coordinates.

Note from Figure 6.1 that θ is the polar angle, measured down from the z axis and ranging from 0 to π , while ϕ is the azimuthal angle, projected onto the xy plane, measured counter-clockwise when viewed from above, starting from the positive x axis, and ranging from 0 to 2π . This naming convention for the two angles is universal in physics, but differs from what you might find in some math textbooks.

If you need to convert from spherical to rectangular coordinates, the needed relations are

$$x = r \sin \theta \cos \phi, \quad y = r \sin \theta \sin \phi, \quad z = r \cos \theta, \quad (6.2)$$

as you can easily show using basic trigonometry. It's equally easy to derive the reverse relations for the spherical coordinates in terms of x , y , and z , but I won't bother to write those down.

When setting up integrals in spherical coordinates, you need to know that the infinitesimal volume element is $(dr)(r d\theta)(r \sin \theta d\phi)$, as shown in Figure 6.1 (note that $r \sin \theta$ is the projection of the length r onto the xy plane).

We can express vectors in spherical coordinates by writing them in terms of three orthogonal unit vectors \hat{r} , $\hat{\theta}$, and $\hat{\phi}$, shown in Figure 6.2. (In this chapter I will use the $\hat{}$ symbol only for unit vectors, never to denote an operator.) One way to write the relations between spherical and rectangular unit vectors is

$$\begin{aligned} \hat{r} &= \sin \theta \cos \phi \hat{i} + \sin \theta \sin \phi \hat{j} + \cos \theta \hat{k}, \\ \hat{\theta} &= \cos \theta \cos \phi \hat{i} + \cos \theta \sin \phi \hat{j} - \sin \theta \hat{k}, \\ \hat{\phi} &= -\sin \phi \hat{i} + \cos \phi \hat{j}, \end{aligned} \quad (6.3)$$

where \hat{i} , \hat{j} , and \hat{k} are the unit vectors in the x , y , and z directions, respectively.

In three dimensions there are three independent momentum operators, which we can express all at once using the gradient symbol:

$$\vec{p} = -i\hbar \left(\hat{i} \frac{\partial}{\partial x} + \hat{j} \frac{\partial}{\partial y} + \hat{k} \frac{\partial}{\partial z} \right) = -i\hbar \nabla. \quad (6.4)$$

To express this operator in spherical coordinates, we can write the gradient operator as

$$\nabla = \hat{r} \frac{\partial}{\partial r} + \hat{\theta} \frac{1}{r} \frac{\partial}{\partial \theta} + \hat{\phi} \frac{1}{r \sin \theta} \frac{\partial}{\partial \phi}. \quad (6.5)$$

Notice that the extra factors in the denominators are the same as those in the volume element.

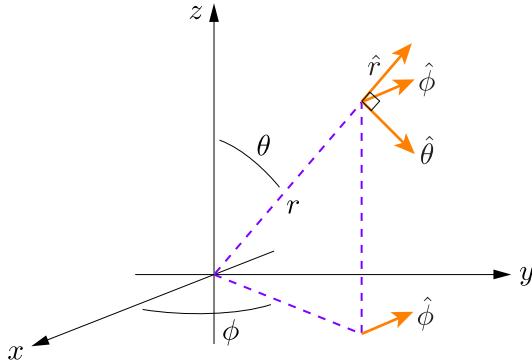


Figure 6.2: The unit vectors \hat{r} , $\hat{\theta}$, and $\hat{\phi}$ are each defined to point in the direction in which the corresponding coordinate is increasing. Note that \hat{r} points radially outward; $\hat{\theta}$ is perpendicular to \hat{r} in the plane defined by \hat{r} and the z axis; and $\hat{\phi}$ is in the xy plane, perpendicular to both \hat{r} and $\hat{\theta}$.

Finally, to write down the time-independent (or time-dependent) Schrödinger equation, we also need to express the Laplacian operator, ∇^2 , in spherical coordinates—and this is somewhat more difficult. First note that the naive guess is *wrong*:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \neq \frac{\partial^2}{\partial r^2} + \frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial \phi^2}. \quad (6.6)$$

(How can you tell?) The correct expression still has three terms: one with two r derivatives, one with two θ derivatives, and one with two ϕ derivatives. However, again there are various stray factors of r and $\sin \theta$. Here it is:

$$\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2}. \quad (6.7)$$

If you want to see a derivation of this expression, you can look in any textbook on mathematical methods in physics, or on classical electromagnetism.¹

The kinetic energy operator is the Laplacian times $-\hbar^2/2m$, so the TISE for a single particle subject to a central potential is

$$-\frac{\hbar^2}{2mr^2} \left[\frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) + \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} \right] + V(r)\psi = E\psi, \quad (6.8)$$

where ψ is a function of r , θ , and ϕ .

At this point, in order to solve the TISE, there are two ways to proceed. The more mathematical approach is to solve this partial differential equation by the method of separation of variables, looking for solutions $\psi(r, \theta, \phi)$ that factor into a function of r times a function of θ times a function of ϕ . (There will be other

¹See, for example, D. J. Griffiths, *Introduction to Electrodynamics*, 4th edition (Pearson, 2013, reprinted by Cambridge University Press, 2017).

solutions that don't factor in this way, but the separable solutions form a basis, so we can build the other solutions as linear combinations of them.) This method works fine, but it requires a fair amount of labor with little reward in the form of physical insight.²

The alternative is to realize that in any problem with spherical symmetry we expect the solutions to have a physical interpretation in terms of angular momentum. Recall that in classical mechanics, when a particle moves under the influence of a central potential $V(r)$, its angular momentum vector $\vec{L} = \vec{r} \times \vec{p}$ must be conserved. The quantum mechanical counterpart to this conservation law is more subtle, but we can ultimately find the separable solutions to the TISE (or more precisely, their angular dependence) by focusing first on the three components of the angular momentum operator, and by looking for their eigenfunctions and eigenvalues. This approach is just as laborious as the purely mathematical approach described above, but we'll gain much more physical understanding along the way.

Problem 6.1. Derive equations 6.2 for x , y , and z in terms of r , θ , and ϕ .

Problem 6.2. Draw a picture or pictures to show where the factors of r and $r \sin \theta$ in the spherical coordinates volume element come from.

Problem 6.3. Derive equations 6.3 for the unit vectors in spherical coordinates.

Problem 6.4. Give one obvious reason why the Laplacian operator in spherical coordinates can't be simply $\partial^2/\partial r^2 + \partial^2/\partial\theta^2 + \partial^2/\partial\phi^2$.

Problem 6.5. Prove that the three-dimensional TISE with a central potential $V(r)$ is separable in spherical coordinates.

6.2 Angular momentum

So we're looking for solutions to the TISE in spherical coordinates, for the case of a particle subject to a potential energy function V that depends only on r . We expect angular momentum to play an important role in this problem, because angular momentum would be conserved for a classical particle subject to such a potential energy function.

Let's start with the definition of angular momentum:

$$\vec{L} = \vec{r} \times \vec{p} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ x & y & z \\ p_x & p_y & p_z \end{vmatrix}, \quad (6.9)$$

where \hat{i} , \hat{j} , and \hat{k} are unit vectors in the x , y , and z directions. More explicitly, the three components of the \vec{L} vector are

$$L_x = yp_z - zp_y, \quad L_y = zp_x - xp_z, \quad L_z = xp_y - yp_x. \quad (6.10)$$

²See, for example, D. J. Griffiths, *Introduction to Quantum Mechanics*, Section 4.1.

In quantum mechanics we reinterpret every quantity in these definitions as an operator. Even so, we don't (yet) have to worry about the order in which we write the symbols, because y commutes with p_z , and so on.

But the three angular momentum operators do *not* commute with each other. As you can easily show (see Problem 6.6),

$$[L_x, L_y] = i\hbar L_z, \quad [L_y, L_z] = i\hbar L_x, \quad [L_z, L_x] = i\hbar L_y. \quad (6.11)$$

This means that the three angular momentum operators do not have a common eigenbasis: states with well-defined L_x , for instance, do not have well-defined L_y or L_z . When we look for angular momentum eigenfunctions, we'll need to specify which of the three components we want the eigenfunctions of. By convention, we ordinarily work with eigenfunctions of L_z .

There is, however, another related operator that commutes with all three angular momentum components. That operator is what we call L^2 , the sum of the squares of the three angular momentum components:

$$L^2 = L_x^2 + L_y^2 + L_z^2. \quad (6.12)$$

Any eigenfunction of this operator has an angular momentum vector whose *magnitude* is well defined, equal to the square root of the corresponding eigenvalue. (We can take the square root of the eigenvalue, but not of the operator itself!) The proof that L^2 commutes with L_z (and with L_x and L_y) is a purely algebraic calculation that makes use of equations 6.11 (see Problem 6.8). The result, in any case, is important enough to display:

$$[L^2, L_x] = [L^2, L_y] = [L^2, L_z] = 0. \quad (6.13)$$

The fact that L^2 commutes with L_z means that these two operators have a common eigenbasis, so our goal will be to find their common eigenstates and the corresponding eigenvalues. Again, these eigenstates will not have well-defined values of L_x or L_y . Although it's hard to imagine a vector with a well-defined magnitude and well-defined z component, but undefined x and y components, that's what quantum mechanics forces us to do.

The good news is that we can find all the eigenvalues that go with these eigenstates without doing any calculus—by using the raising and lowering operators

$$L_+ = L_x + iL_y, \quad L_- = L_x - iL_y. \quad (6.14)$$

These definitions should vaguely remind you of the ladder operators for the harmonic oscillator, equation 5.13. In this case, however, instead of adding and removing units of energy, L_+ and L_- will add and remove units of angular momentum—specifically, units of L_z . To prove this, it's helpful to first work out the commutators

$$[L_z, L_+] = \hbar L_+, \quad [L_z, L_-] = -\hbar L_- \quad (6.15)$$

(see Problem 6.9).

With these relations in hand, the proof goes as follows. Suppose that ψ is an eigenfunction of L_z , and, without loss of generality, suppose that its eigenvalue is $m\hbar$, where m is a unitless constant (nothing to do with mass). Then the function $L_+\psi$ is *also* an eigenfunction of L_z , because

$$\begin{aligned} L_z(L_+\psi) &= ([L_z, L_+] + L_+L_z)\psi \\ &= \hbar L_+\psi + L_+m\hbar\psi \\ &= (m+1)\hbar(L_+\psi). \end{aligned} \quad (6.16)$$

The key step here is in the first line, where I effectively interchanged L_z and L_+ , at the cost of an additional commutator term that was easy to evaluate. In conclusion, $L_+\psi$ is indeed an eigenfunction of L_z , and moreover, its eigenvalue is $(m+1)\hbar$, *greater* than the eigenvalue of ψ itself by one unit of Planck's constant.

In the same way (see Problem 6.10), you can show that

$$L_z(L_-\psi) = (m-1)\hbar(L_-\psi), \quad (6.17)$$

in other words, $L_-\psi$ is an eigenfunction of L_z with eigenvalue $(m-1)\hbar$, *less* than the eigenvalue of ψ itself by one unit of Planck's constant. We have discovered a quantum *ladder* of L_z eigenstates, equally spaced by units of \hbar , and we have raising and lowering operators L_+ and L_- that take us up a rung and down a rung, respectively.

Because L^2 commutes with L_z , we can also assume that ψ is an eigenfunction of L^2 (because L^2 and L_z must have a common eigenbasis). But L^2 also commutes with the raising and lowering operators, so it follows that $L_+\psi$ and $L_-\psi$ are also eigenfunctions of L^2 , with the *same* eigenvalue as ψ (see Problem 6.11). So the ladder operators change only the eigenvalue of L_z , not the eigenvalue of L^2 . There must be a *separate* ladder of states for each eigenvalue of L^2 . The rung spacing is the same on each ladder, but as we'll see, the ladders have different lengths.

Let's consider just one of these ladders, whose rungs all have the same L^2 eigenvalue—that is, the same angular momentum *magnitude*. The z component of a vector can't be bigger than the vector's magnitude, so there must be *highest* rung on the ladder, that is, a state ψ_{top} whose L_z eigenvalue is as high as possible for this particular ladder. Let's call this state's L_z eigenvalue $l\hbar$, where l is another unitless number:

$$L_z\psi_{\text{top}} = l\hbar\psi_{\text{top}}. \quad (6.18)$$

What happens if we use L_+ to try to raise this state? There can't be a higher- L_z state, so the only possible value of $L_+\psi_{\text{top}}$ is zero:

$$L_+\psi_{\text{top}} = 0. \quad (6.19)$$

Similarly, the z component of a vector can't be smaller than minus the vector's magnitude, so there must be a *lowest* possible rung on the ladder, that is, a state

ψ_{bot} whose L_z eigenvalue is as low as possible for this particular ladder. Let's call this state's L_z eigenvalue $l'\hbar$, where l' is another unitless number:

$$L_z\psi_{\text{bot}} = l'\hbar\psi_{\text{bot}}. \quad (6.20)$$

And what happens if we use L_- to try to lower this state? There can't be a lower- L_z state, so the only possible value of $L_-\psi_{\text{bot}}$ is zero:

$$L_-\psi_{\text{bot}} = 0. \quad (6.21)$$

I now want to prove that $l' = -l$, and also relate l to the eigenvalue of L^2 , but to do so I need two more tricky algebraic relations among the various L operators. To derive these relations we just need to fiddle around with products of the raising and lowering operators, for instance,

$$\begin{aligned} L_+L_- &= (L_x + iL_y)(L_x - iL_y) \\ &= L_x^2 + L_y^2 - i[L_x, L_y] \\ &= L^2 - L_z^2 - i(i\hbar L_z) \\ &= L^2 - L_z^2 + \hbar L_z. \end{aligned} \quad (6.22)$$

And if you put L_+ and L_- in the other order, you get (see Problem 6.12)

$$L_-L_+ = L^2 - L_z^2 - \hbar L_z. \quad (6.23)$$

Now if you apply this last equation to ψ_{top} , you get

$$0 = L^2\psi_{\text{top}} - l^2\hbar^2\psi_{\text{top}} - l\hbar^2\psi_{\text{top}}, \quad \text{or} \quad L^2\psi_{\text{top}} = l(l+1)\hbar^2\psi_{\text{top}}. \quad (6.24)$$

Similarly, if you apply equation 6.22 to ψ_{bot} , you get

$$0 = L^2\psi_{\text{bot}} - (l')^2\hbar^2\psi_{\text{bot}} + l'\hbar^2\psi_{\text{bot}}, \quad \text{or} \quad L^2\psi_{\text{bot}} = l'(l'-1)\hbar^2\psi_{\text{bot}}. \quad (6.25)$$

But ψ_{top} and ψ_{bot} are on the *same* quantum ladder, all of whose rungs have the same L^2 eigenvalue, so

$$l'(l'-1) = l(l+1). \quad (6.26)$$

This is a quadratic equation for l' in terms of l , so it has two solutions, which you can easily verify to be $l' = l+1$ and $l' = -l$. The first solution would be nonsense, since the bottom rung can't be higher than the top rung! Therefore it must be the case that $l' = -l$, that is, the ladder of L_z eigenvalues is symmetrical with respect to zero, descending just as far below as it ascends above $L_z = 0$.

One way to build a symmetrical ladder is to put the central rung at $m = 0$. Then every m value, including l , is an integer. In this case the ladder has an odd number of rungs. The other possibility is to put zero half-way between two rungs, so that every m value, including l , equals an integer plus $1/2$. In this case the ladder has an even number of rungs. In summary, the allowed l values are $0, 1/2, 1, 3/2$,

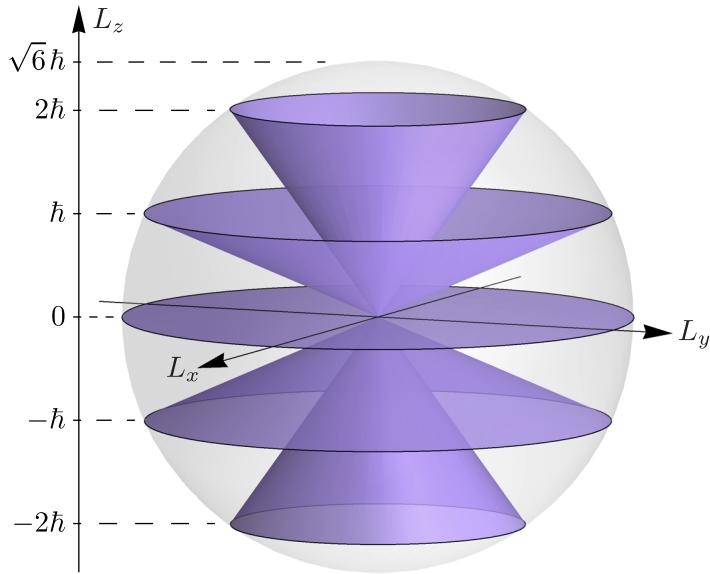


Figure 6.3: A way to visualize the allowed L_z eigenstates for $l = 2$. The magnitude of the angular momentum vector is well defined and equal to $\sqrt{6}\hbar$, the radius of the gray sphere in this three-dimensional space of \vec{L} vectors. For these states L_z is also well defined and equal to either $-2\hbar$, $-\hbar$, 0 , \hbar , or $2\hbar$. But L_x and L_y are *not* well defined, so we cannot visualize \vec{L} as a particular vector in this space; the best we can do is to picture a *cone* made of all the vectors with a given $|\vec{L}|$ and L_z .

2, and so on. Each of these l values has its own ladder, with m values running from $-l$ to $+l$ in integer steps.

For a given angular momentum ladder, the eigenvalue of L^2 is (according to equation 6.24) $l(l+1)\hbar^2$, and therefore the *magnitude* of the angular momentum vector is $\sqrt{l(l+1)}\hbar$. Again, we're working with states for which this magnitude is well defined, as is the z component of \vec{L} , but *not* the x and y components. Although I can't really visualize a "vector" with only one well-defined component, Figure 6.3 shows how you can visualize the set of all possible vectors with a given $|\vec{L}|$ and L_z as a cone. For a given l there are $2l+1$ different allowed m values, so you can visualize these states as a collection of $2l+1$ cones, as shown in the figure.

In the next section I'll show how to find the *eigenfunctions* that go with these angular momentum eigenvalues. The eigenfunctions are functions of the angular variables θ and ϕ , and the standard symbol for them is $Y_l^m(\theta, \phi)$, so we can write

$$L^2 Y_l^m = l(l+1)\hbar^2 Y_l^m \quad \text{and} \quad L_z Y_l^m = m\hbar Y_l^m. \quad (6.27)$$

Oddly, we'll find that half-integer values of m are inconsistent with the existence of such functions. But we *will* find a use for half-integer m values in the following chapter.

Problem 6.6. Derive equations 6.11 for the commutators of the three angular momentum component operators. (Hint: Use the definitions 6.10 and the basic commutator

$[x, p_x] = i\hbar$ to derive one of the angular momentum commutators, then cyclically permute the indexes to obtain the other two.)

Problem 6.7. Prove that the operators L_x , L_y , L_z , and L^2 are Hermitian.

Problem 6.8. Prove that $[L^2, L_z] = 0$, and hence argue that $[L^2, L_x] = [L^2, L_y] = 0$. (Hint: You may wish to use the result of Problem 5.5(b).)

Problem 6.9. Derive the commutator relations in equation 6.15 for $[L_z, L_+]$ and $[L_z, L_-]$.

Problem 6.10. Derive equation 6.17, proving that L_- is indeed a lowering operator.

Problem 6.11. Prove that $[L^2, L_+] = [L^2, L_-] = 0$, and then use these relations to prove that if ψ is an eigenfunction of L^2 , then $L_+\psi$ and $L_-\psi$ are also eigenfunctions of L^2 , with the same eigenvalue as ψ .

Problem 6.12. Derive equation 6.23 for the product L_-L_+ in terms of L^2 and L_z .

Problem 6.13. The operators L_+ and L_- convert one L_z eigenstate into another, but (like the harmonic oscillator ladder operators) they do *not* preserve the normalization of the states. Designating the eigenstates as Y_l^m , the correct relations are

$$\begin{aligned} L_+ Y_l^m &= \hbar \sqrt{l(l+1) - m(m+1)} Y_l^{m+1}, \\ L_- Y_l^m &= \hbar \sqrt{l(l+1) - m(m-1)} Y_l^{m-1}. \end{aligned} \quad (6.28)$$

Use equations 6.22 and 6.23 to derive these relations, referring to the analogous Problem 5.9 for additional guidance. Also check that these relations give the expected results when you apply L_+ to the $m = l$ state or L_- to the $m = -l$ state.

Problem 6.14. Draw accurately scaled and labeled cone diagrams, similar to Figure 6.3, for visualizing the allowed L_z eigenstates for $l = 1$ and $l = 3$.

Problem 6.15. Imagine a diatomic molecule made of two different atoms, such as CO or NO or HCl. In this problem you will explore the *rotational* states of this molecule, neglecting any translational motion, vibrational excitations, or electronic excitations. Then the configuration of the molecule can be described entirely by the *direction* of a vector drawn from the center of one atom to the center of the other, and this direction can be specified by the usual spherical coordinates θ and ϕ . In quantum mechanics, therefore, the wavefunction of this system is a function of these two angles (and no other variables). Please assume that the molecule's moment of inertia I , about its center of mass, is a given constant.

(a) Recall (or look up) the classical formula for the rotational kinetic energy of a rigid object in terms of its angular momentum and moment of inertia. Use this formula to write down the Hamiltonian operator of this system. (There is no potential energy.)

(b) Explain why the energy eigenfunctions of this system are the same as the angular momentum eigenfunctions. What are the corresponding energy eigenvalues? Please assume that only integer l values are allowed. Draw an energy level diagram, with a linear vertical scale, showing the lowest four energy levels and their degeneracies.

(c) For carbon monoxide (CO), the difference in energy between the rotational ground state and the first excited level is approximately 0.00048 eV. What frequency of electromagnetic radiation should you use to induce a transition from the ground state to

the first excited level? What frequency should you use to induce a transition from the first to the second excited level? What part of the electromagnetic spectrum are we talking about here?

- (d) Use this measured energy difference to find the approximate distance between the C and O nuclei.

6.3 Spherical harmonics

Now that we know the eigenvalues of L^2 and L_z , it's time to figure out the corresponding eigenfunctions. They're functions of the angles θ and ϕ (but not of r), so in order to find them, we need to express each of the angular momentum operators in terms of spherical coordinates.

Again we start with the definition $\vec{L} = \vec{r} \times \vec{p}$, but this time we write \vec{p} as a differential operator, $-i\hbar\nabla$, and use equation 6.5 to express the gradient in spherical coordinates:

$$\vec{L} = \vec{r} \times \vec{p} = -i\hbar \vec{r} \times \left(\hat{r} \frac{\partial}{\partial r} + \hat{\theta} \frac{1}{r} \frac{\partial}{\partial \theta} + \hat{\phi} \frac{1}{r \sin \theta} \frac{\partial}{\partial \phi} \right). \quad (6.29)$$

When we take the cross product the \hat{r} term drops out, and the factors of r cancel in the other two terms. Referring to Figure 6.2 and using the right-hand rule to get the signs right, we're left with

$$\vec{L} = -i\hbar \left(\hat{\phi} \frac{\partial}{\partial \theta} - \hat{\theta} \frac{1}{\sin \theta} \frac{\partial}{\partial \phi} \right). \quad (6.30)$$

This is a correct expression for \vec{L} in spherical coordinates, and it's remarkably simple, but it's not what we want! Instead, we want the *rectangular* components of \vec{L} , expressed in terms of θ and ϕ . So the next step is to express $\hat{\theta}$ and $\hat{\phi}$ in terms of the rectangular unit vectors, using equations 6.3. When we do this and gather up the terms, we obtain

$$L_x = i\hbar \left(\sin \phi \frac{\partial}{\partial \theta} + \cot \theta \cos \phi \frac{\partial}{\partial \phi} \right), \quad (6.31)$$

$$L_y = i\hbar \left(-\cos \phi \frac{\partial}{\partial \theta} + \cot \theta \sin \phi \frac{\partial}{\partial \phi} \right), \quad (6.32)$$

$$L_z = -i\hbar \frac{\partial}{\partial \phi}. \quad (6.33)$$

These are the formulas we need to find the angular momentum eigenfunctions.

We can put the L_z operator to immediate use. As in the previous section, I'll denote its eigenvalues as $m\hbar$ and its eigenfunctions as $Y(\theta, \phi)$. Then the L_z eigenvalue equation is

$$-i\hbar \frac{\partial Y}{\partial \phi} = m\hbar Y, \quad (6.34)$$

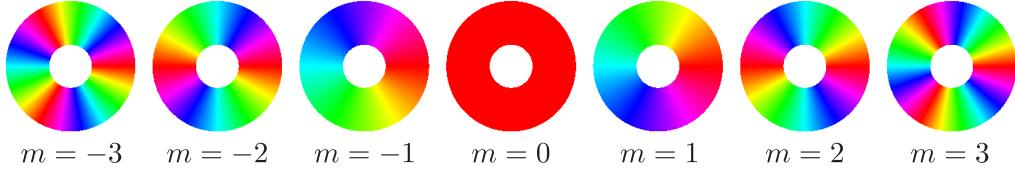


Figure 6.4: Representations of the ϕ dependence of the L_z eigenfunctions, equation 6.35, using color hues to indicate complex phases. In each image the x axis points to the right and the y axis points up, so we are looking “downward” from the positive z direction.

and the solution is

$$Y(\theta, \phi) = f(\theta) e^{im\phi}, \quad (6.35)$$

where $f(\theta)$ is an integration constant as far as ϕ is concerned, but still an undetermined function of θ . So the ϕ dependence of any L_z eigenfunction is simply a pure complex phase factor, which changes with ϕ at a rate proportional to the quantum number m . And this is why m must be an integer, as I mentioned at the end of the previous section. For $\phi = 0$ and $\phi = 2\pi$ are just two names for the same place, so Y must have the same value at both—but equation 6.35 has this property only if m is an integer. Figure 6.4 shows how we can visualize the ϕ dependence of the L_z eigenfunctions using color hues to represent phases.

We still need to find the θ dependence of the L_z eigenfunctions, specifically for those that are also eigenfunctions of L^2 . One way to do this would be to work out the L^2 operator in spherical coordinates and then solve its eigenvalue equation as a differential equation in terms of θ . A much easier approach, however, is to use ladder operators. From equations 6.31 and 6.32 it’s just a short step to obtain

$$L_+ = L_x + iL_y = \hbar e^{i\phi} \left(\frac{\partial}{\partial \theta} + i \cot \theta \frac{\partial}{\partial \phi} \right), \quad (6.36)$$

$$L_- = L_x - iL_y = \hbar e^{-i\phi} \left(-\frac{\partial}{\partial \theta} + i \cot \theta \frac{\partial}{\partial \phi} \right). \quad (6.37)$$

These operators must give zero when they act on the top and bottom rungs, respectively, of an L_z ladder. For the top rung, this condition says

$$0 = L_+ Y_l^l = \left(\frac{\partial}{\partial \theta} + i \cot \theta \frac{\partial}{\partial \phi} \right) Y_l^l(\theta, \phi). \quad (6.38)$$

If you plug in equation 6.35 for Y , with $m = l$, you get an ordinary differential equation for $f(\theta)$ that’s easy to solve (see Problem 6.16). The full solution for Y_l^l is

$$Y_l^l(\theta, \phi) \propto (\sin \theta)^l e^{il\phi}, \quad (6.39)$$

where I’ve omitted a normalization constant that I’ll discuss in a moment. Next, to find the eigenfunctions with $m = l - 1$ (one rung down from the top of the ladder), just apply the lowering operator L_- to this expression. Continue on down the ladder

in the same way—though at this point it's a lot easier to pick a particular l rather than trying to work them all out at once. Eventually you'll come to the bottom rung, $m = -l$, whose θ dependence also happens to be $(\sin \theta)^l$. Of course you could also have obtained this expression by *starting* at the bottom rung, solving the differential equation $L_- Y_l^{-l} = 0$.

The eigenfunctions $Y_l^m(\theta, \phi)$ are called *spherical harmonics*. They form a complete basis, in the sense that *any* function of θ and ϕ can be expressed as a linear combination of them. Table 6.1 shows all the spherical harmonics up to $l = 3$. You can easily work out others using the ladder operators if you like, or generate them in Mathematica with the instruction `SphericalHarmonicY[l,m,theta,phi]`. The θ dependence of the spherical harmonics can also be expressed in terms of so-called *associated Legendre functions* of $\cos \theta$, and you can look up those functions if you ever need them. The normalization constants of the spherical harmonics are defined by the arbitrary (but natural) condition that when you take the square modulus and integrate over all directions, you get 1:

$$\int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi |Y_l^m(\theta, \phi)|^2 = 1. \quad (6.40)$$

There is still an arbitrary overall phase factor in each Y_l^m function; we always take these phases to be real, but the most common convention inserts some surprising (though utterly unimportant) minus signs, as you can see in Table 6.1.

The best way to visualize the spherical harmonics is to draw them on the surfaces of spheres, as shown in Figure 6.5. Please spend some time looking at the patterns—especially how the number of nodes (black bands) varies as you move horizontally or vertically from one spherical harmonic to another. Can you predict the appearances of the $l = 4$ spherical harmonics?

Even though we've now found the eigenfunctions of the L^2 operator, I still haven't written down the operator itself. You can work out its formula in spherical coordinates either from the definition $L^2 = L_x^2 + L_y^2 + L_z^2$ and equations 6.31 through 6.33, or from one of the tricky formulas for L_+L_- and L_-L_+ , equations 6.22 and 6.23, along with equations 6.36 and 6.37. Either method takes about a page of algebra, and when the smoke clears, you obtain

$$L^2 = -\hbar^2 \left[\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right]. \quad (6.41)$$

I hope this expression looks vaguely familiar to you. In fact, it has exactly the same angular dependence as the angular terms in equation 6.7 for the Laplacian operator. This means we can write the kinetic energy operator, $-(\hbar^2/2m)\nabla^2$, in a pretty simple way in terms of L^2 . Adding on a central potential energy function gives us an extremely useful formula for the Hamiltonian:

$$H = -\frac{\hbar^2}{2mr^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{L^2}{2mr^2} + V(r). \quad (6.42)$$

m	$l = 0$	$l = 1$	$l = 2$	$l = 3$
3				$-\sqrt{\frac{35}{64\pi}} \sin^3 \theta e^{3i\phi}$
2		$\sqrt{\frac{15}{32\pi}} \sin^2 \theta e^{2i\phi}$	$\sqrt{\frac{105}{32\pi}} \cos \theta \sin^2 \theta e^{2i\phi}$	
1	$-\sqrt{\frac{3}{8\pi}} \sin \theta e^{i\phi}$	$-\sqrt{\frac{15}{8\pi}} \cos \theta \sin \theta e^{i\phi}$	$-\sqrt{\frac{21}{64\pi}} (5 \cos^2 \theta - 1) \sin \theta e^{i\phi}$	
0	$\sqrt{\frac{1}{4\pi}}$	$\sqrt{\frac{3}{4\pi}} \cos \theta$	$\sqrt{\frac{5}{16\pi}} (3 \cos^2 \theta - 1)$	$\sqrt{\frac{7}{16\pi}} (5 \cos^3 \theta - 3 \cos \theta)$
-1		$\sqrt{\frac{3}{8\pi}} \sin \theta e^{-i\phi}$	$\sqrt{\frac{15}{8\pi}} \cos \theta \sin \theta e^{-i\phi}$	$\sqrt{\frac{21}{64\pi}} (5 \cos^2 \theta - 1) \sin \theta e^{-i\phi}$
-2			$\sqrt{\frac{15}{32\pi}} \sin^2 \theta e^{-2i\phi}$	$\sqrt{\frac{105}{32\pi}} \cos \theta \sin^2 \theta e^{-2i\phi}$
-3				$\sqrt{\frac{35}{64\pi}} \sin^3 \theta e^{-3i\phi}$

Table 6.1: Formulas for the spherical harmonics Y_l^m (that is, the eigenfunctions of L^2 and L_z) for $l \leq 3$, including normalization constants and some conventional but unimportant minus signs. Each column is a “quantum ladder” for a given l value, so the ladder operators L_+ and L_- convert each function into the one above or below it, respectively (aside from a change in normalization, given by equations 6.28). Notice that the θ dependence of Y_l^{-m} is the same as that of Y_l^m .

Because L^2 depends on θ and ϕ but not on r , it is now easy to see that L^2 commutes with H (for central potentials), and therefore we know that L^2 and H have a common eigenbasis. Both H and L^2 also commute with L_z , so in fact we can find a common eigenbasis of all three of these operators. Specifying the eigenvalues of these three operators will be sufficient to determine the eigenfunctions uniquely (aside from an overall constant), so we refer to this set of operators as a *complete set of commuting observables*. We have already seen that the eigenfunctions of L^2 and L_z are the spherical harmonics, $Y_l^m(\theta, \phi)$. Next we turn to the radial portion of the Hamiltonian and its eigenfunctions.

Problem 6.16. Solve the differential equation 6.38 to obtain formula 6.39 for $Y_l^l(\theta, \phi)$. Then apply the lowering operator to obtain a general formula for $Y_l^{l-1}(\theta, \phi)$, and use Table 6.1 to check your results for $l = 1, 2$, and 3 . Don’t worry about normalization constants.

Problem 6.17. Derive the normalization constants for the $l = 0$ and $l = 1$ spherical harmonics.

Problem 6.18. Derive equation 6.41, using either of the methods described in the text just above it. (Hint: It’s helpful to write a test function, $f(\theta, \phi)$, to the right of each term, so you’ll remember to correctly use the product rule wherever a derivative acts on a nontrivial product.)

Problem 6.19. After studying Figure 6.5, sketch (using colored pencils or the like) the expected appearance of each of the $l = 4$ spherical harmonics. Then check your

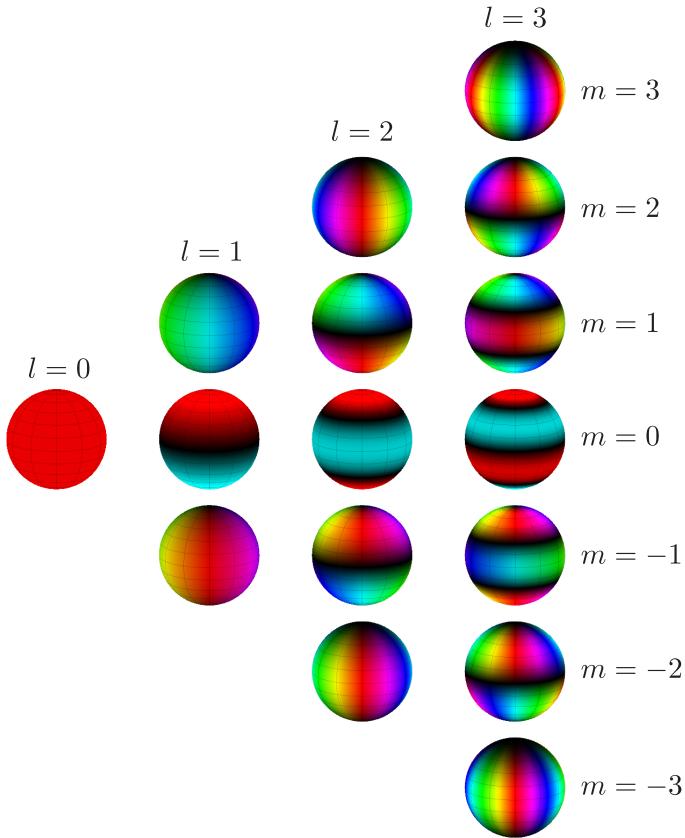


Figure 6.5: Spherical harmonics (that is, eigenfunctions of L^2 and L_z) for $l \leq 3$, shown as color-phase density plots on the surfaces of spheres. Unfortunately, only the front half of each sphere is visible. The z axis points up, the x axis points out of the page, and the y axis points to the right. Black indicates a function value of zero (that is, a node). The arrangement is the same as in Table 6.1.

sketches—especially the node locations—against the explicit formulas, obtained from Mathematica or another method of your choice.

Problem 6.20. Which spherical harmonics are nonzero on the z axis? What is the physical explanation for your answer?

6.4 The radial equation

Now that we know all about angular momentum, let's go back to the time-independent Schrödinger equation, in spherical coordinates, for a particle subject to a potential energy function that depends only on r . Writing the Hamiltonian in terms of L^2 as in equation 6.42, we obtain for the TISE

$$\left[-\frac{\hbar^2}{2mr^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{L^2}{2mr^2} + V(r) \right] \psi = E\psi. \quad (6.43)$$

As explained at the end of the previous section, we can find solutions to this equation that are also eigenfunctions of L^2 and L_z —and the angular dependence of these solutions is that of the spherical harmonics, $Y_l^m(\theta, \phi)$. More precisely, we can find solutions that factor into a spherical harmonic times a function of r alone. I'll call these r -dependent functions $R(r)$, so we can write

$$\psi(r, \theta, \phi) = R(r) Y_l^m(\theta, \phi). \quad (6.44)$$

Plugging this separable solution into the TISE, replacing L^2 with its known eigenvalue, and canceling the factor of Y_l^m , we obtain the *radial Schrödinger equation*,

$$\left[-\frac{\hbar^2}{2mr^2} \frac{d}{dr} \left(r^2 \frac{d}{dr} \right) + \frac{\hbar^2 l(l+1)}{2mr^2} + V(r) \right] R(r) = ER(r). \quad (6.45)$$

The radial Schrödinger equation simplifies somewhat if we make a change of variables from $R(r)$ to the function

$$u(r) = rR(r), \quad (6.46)$$

which is sometimes called the *reduced radial wavefunction*. Notice that

$$\begin{aligned} \frac{d}{dr} \left(r^2 \frac{d}{dr} \left(r^{-1} u(r) \right) \right) &= \frac{d}{dr} \left(r^2 \left(-r^{-2} u(r) + r^{-1} \frac{du}{dr} \right) \right) \\ &= \frac{d}{dr} \left(-u(r) + r \frac{du}{dr} \right) \\ &= -\frac{du}{dr} + \frac{du}{dr} + r \frac{d^2 u}{dr^2} \\ &= r \frac{d^2 u}{dr^2}. \end{aligned} \quad (6.47)$$

Plugging this simplification into the radial Schrödinger equation and multiplying through by r , we obtain the *reduced radial equation*,

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dr^2} + \frac{\hbar^2 l(l+1)}{2mr^2} + V(r) \right] u(r) = Eu(r). \quad (6.48)$$

The first term in brackets now looks just like the familiar one-dimensional kinetic energy operator. The second term, for any fixed l value, is a known function of r , which we can group with the potential energy to form the *effective potential energy* function,

$$V_{\text{eff}}(r) = \frac{\hbar^2 l(l+1)}{2mr^2} + V(r), \quad (6.49)$$

just as we often do in classical mechanics problems with central forces. Then the reduced radial equation becomes simply

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dr^2} + V_{\text{eff}}(r) \right] u(r) = Eu(r), \quad (6.50)$$

which is identical to the one-dimensional TISE but with x replaced by r , ψ replaced by u , and V replaced by V_{eff} .

Because of this mathematical similarity, we can analyze and solve the reduced radial equation using all of the techniques that we learned for the one-dimensional TISE in Chapter 2. We can sketch qualitative graphs of the solutions. In a few cases we can find exact solutions. And we can solve the equation numerically using the shooting method or the matrix diagonalization method. The only mathematical difference is that whereas x normally ranges from $-\infty$ to ∞ , negative r values are not allowed. In fact, $u(r)$ must equal zero at $r = 0$, as is almost obvious from the definition $u(r) = rR(r)$. (The only question that requires thought is whether $R(r)$ can ever be infinite at $r = 0$, and the answer is “no” in all the examples we’ll encounter.) There’s also a big difference when we interpret the solutions, because the full wavefunction is $u(r)/r$ times the applicable spherical harmonic. And we need to remember that V_{eff} depends on l , so we get to solve the reduced radial equation separately for $l = 0$, $l = 1$, $l = 2$, and so on.

The l -dependent term in V_{eff} is called the *centrifugal term* (as it is in classical mechanics). It’s zero for $l = 0$, but for $l > 0$ it creates an effective repulsive force that becomes infinite as $r \rightarrow 0$, pushing the particle out from the origin. (Of course this “force” is just a useful fiction that we invent so we can pretend that this is a one-dimensional problem; in three dimensions, we would instead say that assuming a nonzero angular momentum entails assuming that the particle avoids the origin.)

The spherical infinite square well

As a first example of solving the radial Schrödinger equation, consider the *spherical infinite square well*, with $V(r) = 0$ out to some radius a and $V = \infty$ beyond that. This potential is a somewhat reasonable model of the environment of a proton or neutron inside an atomic nucleus. In what follows I will use natural units in which $\hbar = m = a = 1$.

Inside the well, the effective potential energy is just the centrifugal term, $l(l + 1)/(2r^2)$. But we have an infinite and infinitely abrupt potential barrier at $r = 1$, and we can treat the boundary condition $u = 0$ at $r = 0$ as another such barrier. I’ve plotted $V_{\text{eff}}(r)$ for a few different l values in Figure 6.6(a).

The case $l = 0$ is mathematically equivalent to an ordinary one-dimensional infinite square well, so we can immediately write down the energy levels and reduced wavefunctions:

$$E_n = \frac{\pi^2 n^2}{2}, \quad u_n(r) \propto \sin(n\pi r), \quad n = 1, 2, 3, \dots \quad (\text{for } l = 0). \quad (6.51)$$

These energy levels are plotted in the leftmost column of Figure 6.6(b). Just remember that the *reduced* wavefunctions aren’t the same as the *actual* wavefunctions; to obtain the latter, we must divide by r and multiply by the appropriate spherical harmonic. For $l = 0$ the spherical harmonic is a mere constant, so the energy

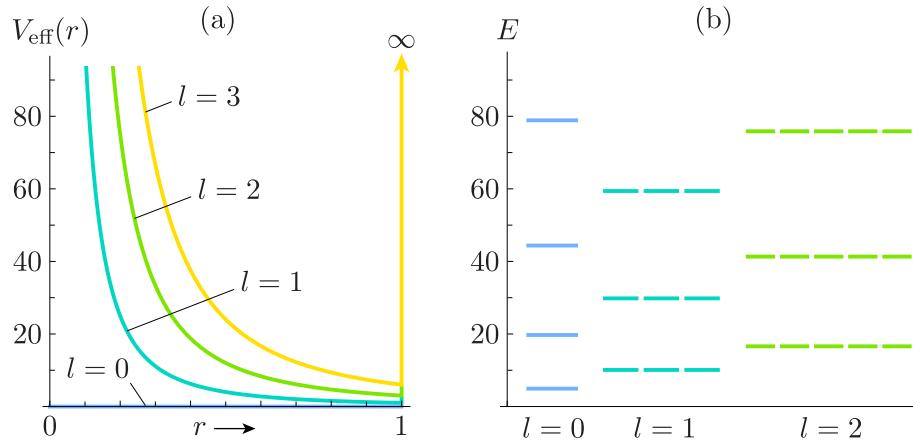


Figure 6.6: (a) The effective potential energy functions for the spherical infinite square well for $l = 0, 1, 2$, and 3 . Both axes use natural units in which $\hbar = m = a = 1$. (b) The resulting low-lying energy levels for $l = 0, 1$, and 2 , plotted on the same vertical scale. Note that the levels with nonzero l are degenerate, with $2l + 1$ possible values of the quantum number m . There are no degeneracies between different values of l .

eigenfunctions are

$$\psi_n(r, \theta, \phi) \propto \frac{\sin(n\pi r)}{r} \quad (\text{for } l = 0). \quad (6.52)$$

Unlike $u_n(r)$, these functions do *not* go to zero at the origin; in fact they reach their maximum values there.

For $l > 0$ the solutions are not so elementary, but they're still easy to find using the technology of Chapter 2. Qualitatively, we can see from Figure 6.6(a) that the centrifugal term will effectively push the wavefunctions away from the origin and up against the barrier at $r = 1$, reducing the wavelength and therefore increasing the energy, as l increases, for a wavefunction with a given number of radial bumps. Near the origin, any $l > 0$ state will have a classically forbidden region at small r values, where $u(r)$ will have exponential-like behavior (with a graph that curves away from, rather than toward, the r axis). In fact one can show that $u(r)$ goes to zero sufficiently rapidly as $r \rightarrow 0$ that even when we divide by r to obtain ψ , we find that the wavefunction goes to zero at the origin. A quantum particle with well-defined energy and nonzero angular momentum will *never* be found at the origin. This behavior is the quantum counterpart to the fact that a classical particle with a given energy and nonzero angular momentum, subject to a central force, cannot come closer to the origin than a certain minimum distance.

To find the $l > 0$ energies and wavefunctions quantitatively, we can use either the shooting method or the matrix diagonalization method. A few of the resulting energies are plotted in Figure 6.6(b). As expected, the energies increase as l increases. As indicated in the figure, the energy levels with nonzero l are degenerate, but there are no degeneracies between levels with different values of l .

In this and other spherically symmetric problems, it is customary among physicists (and chemists) to refer to the $l = 0$ states as *s waves*, the $l = 1$ states as *p waves*, the $l = 2$ states as *d waves*, and the $l = 3$ states as *f waves*. (Still higher- l states are labeled alphabetically starting with g.) The original meanings of these letters are of only historical interest, but we're stuck with them, so please learn them if you haven't already.

Problem 6.21. Make qualitatively accurate sketches of $u(r)$ for each of the three $l = 1$ levels shown in Figure 6.6. Use the same horizontal scale as for the graph of $V_{\text{eff}}(r)$.

Problem 6.22. Use colored pencils (or something similar) to make qualitatively accurate sketches of several of the wavefunctions for the spherical infinite square well. Start with the three lowest-energy $l = 0$ wavefunctions, then sketch the six lowest-energy $l = 1$ wavefunctions (three at each of the two lowest levels, corresponding to three different L_z eigenvalues). Rather than trying to make three-dimensional sketches, it is probably better to sketch two-dimensional slices, say in the xz plane, being sure to label the z direction and to make notes to indicate what changes in the third dimension.

Problem 6.23. Use either the shooting method or the matrix diagonalization method to determine all of the $l > 0$ energy levels shown in Figure 6.6(b), to at least four significant figures, and to plot the reduced wavefunctions.

Problem 6.24. Consider a particle of mass m confined by an attractive central potential in three dimensions that is directly proportional to the distance from the origin: $V = \alpha r$, for some constant α . Plot the effective potential energy functions for $l = 0$, 1, and 2, then use either the shooting method or the matrix diagonalization method to find the low-lying energies and wavefunctions for these l values. Use units in which $\hbar^2/m = 1$ and $\alpha = 1$. Plot or sketch an accurate energy-level diagram showing a few of the lowest levels for each of these l values (separating the different l values horizontally). Also plot both the reduced radial wavefunctions $u(r)$ and the actual radial wavefunctions $R(r)$ for a selection of the energy levels.

Problem 6.25. It is sometimes reasonable to model the attractive force between quarks (due to the so-called “strong interaction”) as constant, so the potential energy is linear. The light quarks that make up protons and neutrons are relativistic, but the c and b quarks are heavy enough for nonrelativistic quantum mechanics to apply to their bound states with reasonable accuracy. (Even though these quarks are unstable, they live long enough to form well-defined bound states.) Figure 6.7 shows a published energy level diagram for “charmonium,” the system of a c quark and c antiquark. In the diagram the letters S, P, and D correspond to $l = 0, 1$, and 2, respectively. The subscripts and superscripts refer to different spin states, which I’d like you to ignore for the purpose of this problem, so I’ve added broad gray lines to the diagram to guide your eye to “average” over the various spin states.

- (a) Compare the *qualitative* features of the charmonium energy level diagram to the energy levels that you found for a linear potential in the previous problem. Is the linear potential a reasonable model for this system? Explain in some detail.
- (b) Use the two lowest $l = 0$ levels to make a *quantitative* estimate of the c quark rest energy and the strength α of the (approximately) linear potential. Express your answers in GeV and in GeV/fm, respectively. Hints: First compare the spacing between these

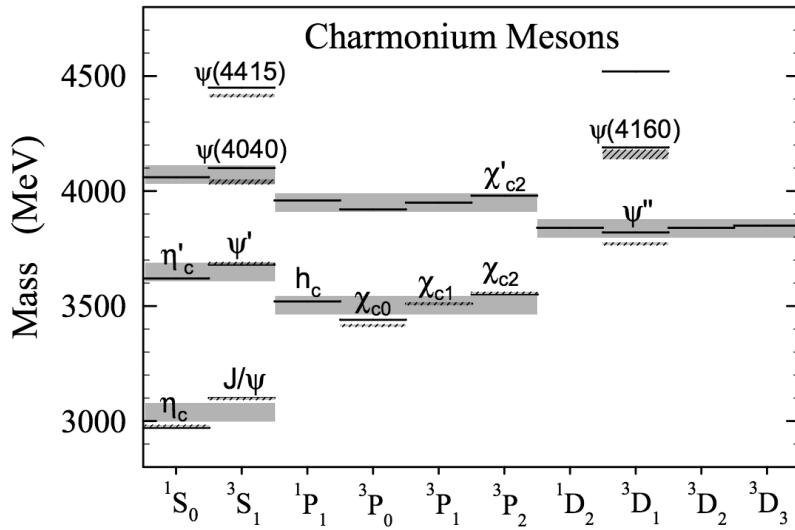


Figure 6.7: Measured energy levels of the “charmonium” system, consisting of a c quark and an anti- c quark. From S. Godfrey and S. L. Olsen, “The Exotic XYZ Charmonium-Like Mesons,” Ann. Rev. Nucl. Part. Sci. **58**, 51–71 (2008), arXiv:0801.3867 [hep-ph]. Note that although the vertical axis is labeled “mass,” it is actually energy, including the rest energies, mc^2 , of the quark and antiquark. The gray horizontal bands in the diagram are not from the original paper; I have drawn them to represent rough averages over the various spin states.

two levels to what you found for the two lowest $l = 0$ levels in the previous problem, and use this comparison to determine where zero on the charmonium energy scale “should” be; assume that any remaining energy comes from the rest energies of the two particles. Then determine what one unit of energy, in the previous problem, is in terms of \hbar , m , and α . Note that for a system of two moving particles, the mass m that enters the Schrödinger equation is actually the reduced mass, which in this case is *half* the actual mass of either of the particles (see the discussion on page 131). Because m is now known, you can use the values in the diagram to determine α .

6.5 The hydrogen atom

The most important example of a spherically symmetric potential energy is the *Coulomb potential energy*,

$$V(r) = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r}, \quad (6.53)$$

between two point charges q_1 and q_2 separated by a distance r . If one of the two charges is a heavy atomic nucleus and the other is a much lighter electron, then to a good approximation we can treat the nucleus as a fixed center of force and apply quantum mechanics only to the electron’s motion. (See page 131 for a brief discussion of why this approximation works.) In terms of the fundamental unit of

charge,

$$e = 1.602 \times 10^{-19} \text{ C}, \quad (6.54)$$

the electron's charge is $-e$ and the nuclear charge is Ze , where Z is the number of protons. For now we'll consider only the hydrogen atom, with $Z = 1$, so the potential energy is

$$V(r) = -\frac{e^2}{4\pi\epsilon_0 r} \frac{1}{r}. \quad (6.55)$$

Given this potential energy function, we can immediately write down the effective potential energy,

$$V_{\text{eff}}(r) = \frac{\hbar^2 l(l+1)}{2m_e r^2} - \frac{e^2}{4\pi\epsilon_0 r}, \quad (6.56)$$

where m_e is the electron's mass, and then use this V_{eff} in the (reduced) radial Schrödinger equation,

$$\left[-\frac{\hbar^2}{2m_e} \frac{d^2}{dr^2} + V_{\text{eff}}(r) \right] u(r) = Eu(r). \quad (6.57)$$

Atomic units

It looks like the TISE for the hydrogen atom involves four different constants: e , ϵ_0 , m_e , and \hbar . But they occur in only two different combinations,

$$\frac{e^2}{4\pi\epsilon_0} \quad \text{and} \quad \frac{\hbar^2}{m_e}, \quad (6.58)$$

and you can see from equation 6.56 that these combinations have dimensions of energy times distance and energy times distance squared, respectively. We can therefore divide the latter by the former to obtain a natural unit of distance,

$$a_0 = \frac{\hbar^2/m_e}{e^2/(4\pi\epsilon_0)} = 0.529 \times 10^{-10} \text{ m}, \quad (6.59)$$

called the *Bohr radius* (after Danish physicist Niels Bohr, 1885–1962). And then we can divide $e^2/(4\pi\epsilon_0)$ by a_0 to obtain a natural unit of energy,

$$E_h = \frac{e^2}{4\pi\epsilon_0 a_0} = \frac{\hbar^2}{m_e a_0^2} = \left(\frac{e^2}{4\pi\epsilon_0} \right)^2 \frac{m_e}{\hbar^2} = 27.2 \text{ eV}, \quad (6.60)$$

called the *Hartree energy* (after British physicist Douglas Hartree, 1897–1958, who made important contributions to theoretical and computational atomic physics).

These natural units are called *atomic units*, often abbreviated a.u. (not to be confused with AU for astronomical units!). Actually, the atomic unit system sets all four of the constants m_e , \hbar , e , and $1/(4\pi\epsilon_0)$ equal to 1.³

³Occasionally you might encounter a competing atomic unit system in which factors of 2 are absorbed into some of the units so the energy unit comes out half as large, 13.6 eV. That energy unit is called the *Rydberg energy*, and you can distinguish the two systems by saying *Hartree atomic units* or *Rydberg atomic units*. In this book I will use only Hartree atomic units.

With the understanding that all distances are measured in units of a_0 and all energies are measured in units of E_h , the radial Schrödinger equation becomes

$$\left[-\frac{1}{2} \frac{d^2}{dr^2} + \frac{l(l+1)}{2r^2} - \frac{1}{r} \right] u(r) = Eu(r). \quad (6.61)$$

Problem 6.26. Check the numerical values of the Bohr radius and Hartree energy, equations 6.59 and 6.60.

Problem 6.27. Find the correct expressions, and the numerical values in SI units, of the atomic units of time and velocity. Comment briefly on the results.

Qualitative solutions

Figure 6.8 shows the effective potential energy (equation 6.56) for $l = 0, 1$, and 2 . As you can see, the attractive Coulomb potential always dominates at large r , but when $l > 0$ the repulsive centrifugal term sends V_{eff} to $+\infty$ as $r \rightarrow 0$. The competition between the two terms results in a local minimum, which you can easily show to lie at $r = l(l+1)$, at which point $V_{\text{eff}} = -1/(2l(l+1))$ (check this on the graph!). Any solutions to the Schrödinger equation must have energies at least a little above this minimum, that is, above $-1/4$ for $l = 1$, above $-1/12$ for $l = 2$, and so on. Meanwhile, in order for the electron to be bound to the nucleus (that is, in order to have a hydrogen *atom* rather than an *ion*), its energy must be negative.

With these limits in mind, it's an instructive exercise to simply *guess* some energy levels, draw them on the graph to determine the associated classical turning points, and then sketch tentative graphs of the one-bump wavefunction, the two-bump wavefunction, and so on, just as you would do if this were a one-dimensional

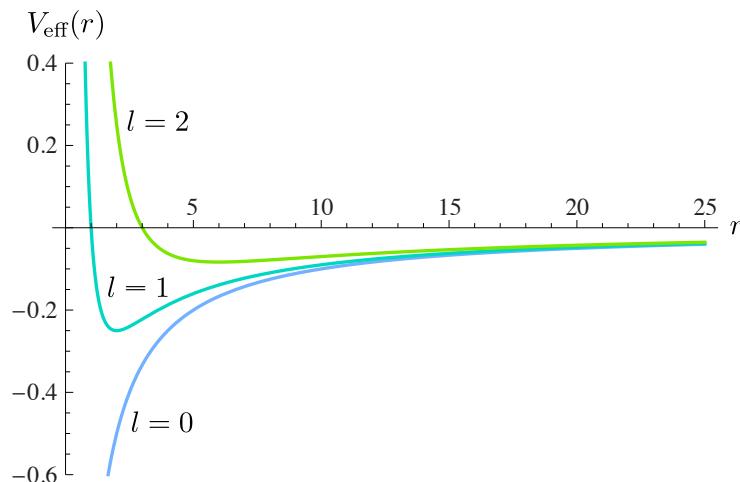


Figure 6.8: The effective potential energy for the electron in a hydrogen atom, for $l = 0, 1$, and 2 . Both distance and energy are measured in atomic units.

problem. There is a separate set of solutions for each l value, and these “wavefunctions” will actually be $u(r)$, which you have to divide by r to obtain $R(r)$. The widths of the “bumps” will grow as you go outward (away from the minimum of V_{eff}), and the centrifugal term pushes the wavefunctions farther and farther out as l increases. It’s not obvious what happens near the origin in the case $l = 0$, but it’s reasonable to guess that even then the *reduced* wavefunction must go to zero as $r \rightarrow 0$, and this guess turns out to be correct.

One other thing that isn’t obvious is how *many* bound states this system has for each l value. On one hand the potential wells for $l > 0$ don’t look especially deep, so you might guess that there aren’t very many bound states. But on the other hand, for values of E that are only slightly negative there is a great deal of horizontal space in which to fit plenty of bumps. It turns out that this second effect dominates, so the number of bound states is actually *infinite* for each l value, no matter how large.

Problem 6.28. Working from Figure 6.8, sketch a copy of the V_{eff} graph for $l = 1$, expanding the vertical scale somewhat. Guess three reasonable values for the three lowest energies for this l value, and show these three levels on the graph. Then, using the same horizontal scale, draw qualitatively reasonable sketches of the three corresponding reduced radial wavefunctions. Label the important features of your sketches.

Numerical solutions

The matrix diagonalization method isn’t well suited to the Coulomb potential, because the wavefunctions extend out to rather large r values, forcing you to use a wide “box” to enclose them, and for such a wide box you need to use a lot of sine waves (large `nMax`) to accurately fit the short-wavelength, small- r portions of the wavefunctions. You can still get a few of the lowest-energy states, but it’s computationally inefficient.

The shooting method, on the other hand, works just fine. For this purpose it’s a little easier to rewrite the radial equation with d^2u/dr^2 by itself on the left:

$$\frac{d^2u}{dr^2} = -2\left(E - \frac{l(l+1)}{2r^2} + \frac{1}{r}\right)u(r). \quad (6.62)$$

Because of the r ’s in the denominators, you need to start the integration a little away from the origin, say at $r_{\min} = 0.0001$. For $l = 0$ the functions $u(r)$ turn out to be linear near the origin, so it works well to use the boundary conditions $u(r_{\min}) = r_{\min}$ and $u'(r_{\min}) = 1$. For $l > 0$ the functions $u(r)$ die out more rapidly near the origin, so it’s best to set $u(r_{\min}) = 0$ and $u'(r_{\min}) = r_{\min}$ (or any other small value). You’ll find that you need to go out to surprisingly large maximum values of r .

You probably won’t be surprised to learn that the energy eigenvalues follow a simple pattern. For $l = 0$, still using atomic units, they are $-1/2$, $-1/8$, $-1/18$, $-1/32$, and so on, that is, $-1/(2n^2)$, where $n = 1, 2, 3, \dots$ is the number of bumps. Then an amazing (though perhaps familiar) thing happens for $l = 1$: The energy

values turn out to be exactly the same, except that the list omits $-1/2$ and instead starts at $-1/8$, so the one-bump $l = 1$ wavefunction is degenerate with the *two*-bump $l = 0$ wavefunction, while the two-bump $l = 1$ wavefunction is degenerate with the *three*-bump $l = 0$ wavefunction, and so on. And the list of energies for $l = 2$ is similarly degenerate with the others, but starting with $-1/18$. Please recall that there is *no* such degeneracy for the spherical infinite well. The degeneracy seems like a total coincidence! Of course it's not a coincidence, but the reason for it is rather difficult to understand so I won't go into it here.⁴

Because of this degeneracy, it's conventional to define the quantum number n so that the energy formula $-1/(2n^2)$ works even for $l > 0$. In conventional units,

$$E_n = -\frac{E_h}{2n^2} = -\frac{13.6 \text{ eV}}{n^2}. \quad (6.63)$$

This definition of n is confusing, because it means (for instance) that the wavefunction with $l = 1$ and $n = 2$ has only *one* bump (in the r direction). In general, the number of bumps in $u(r)$ equals $n - l$. Figure 6.9 shows an energy-level diagram with the values of l and n (but not the number of bumps) labeled. Of course it's also important to remember that for any given n and l values, there are still $2l + 1$ degenerate states with different values of the quantum number m (also called m_l when a plain m would be ambiguous). So in Figure 6.9, each $l = 1$ state is actually triply degenerate, each $l = 2$ state is five-fold degenerate, and so on.

Problem 6.29. Use the shooting method to verify the energies, and to plot the reduced radial wavefunctions, for the following states of a hydrogen atom: (a) the three lowest-energy states for $l = 0$; (b) the two lowest-energy states for $l = 1$; and (c) the lowest-energy state for $l = 2$. Use atomic units.

Analytic solutions

As with the one-dimensional harmonic oscillator, the existence of a simple formula for the energy eigenvalues is a sure sign that it must be possible to solve the TISE analytically. For the full analytic solution (via the power-series method) you can look in any traditional quantum mechanics textbook.⁵ Here I'll focus on the general form of the solutions and the specific formulas for a few of the simplest ones.

One key observation is that in the limit of large r , both $V(r)$ and the centrifugal term go to zero so the radial equation becomes simply

$$\frac{d^2u}{dr^2} \approx -2E u(r) = \frac{1}{n^2} u(r) \quad (\text{for } r \gg 1), \quad (6.64)$$

where in the last expression I've used the formula $E = -1/(2n^2)$ that we inferred from the pattern of the numerical solutions. The solutions to this differential equa-

⁴The “coincidental” degeneracy of the quantum Coulomb problem is related to the fact that bound orbits are closed in the classical Coulomb problem. If you'd like to learn more, look up the so-called Laplace-Runge-Lenz vector.

⁵For instance, Griffiths, Section 4.2.

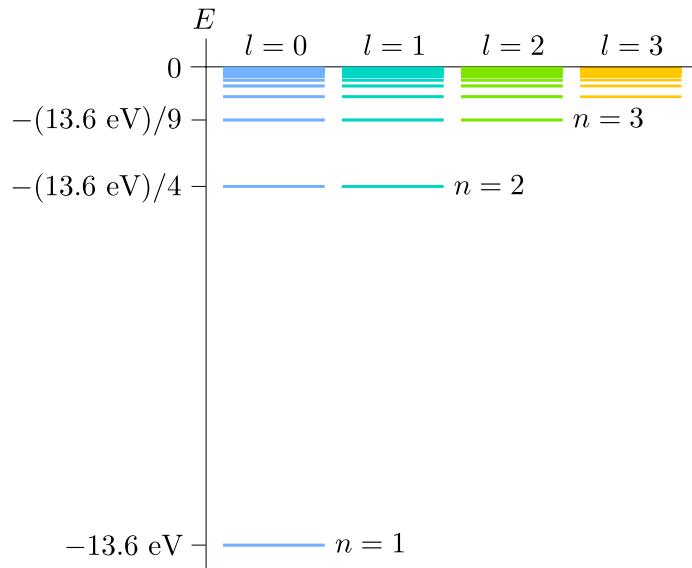


Figure 6.9: Energy levels of the hydrogen atom. Note the remarkable degeneracies between states with different l values. Also note that the $l > 0$ states have additional degeneracies, not shown here, due to the multiple allowed values of L_z .

tion are exponential functions, $e^{r/n}$ and $e^{-r/n}$, but the former isn't normalizable so we're left with the latter.

The (reduced) wavefunctions must also go to zero at $r = 0$, and most of them have multiple bumps and nodes. The simplest functions that have these properties are polynomial functions of r , so it's reasonable (and correct) to guess that the general formula for $u(r)$ is a polynomial in r times $e^{-r/n}$. There can be no constant terms in these polynomials, because they must go to zero at the origin.

The next thing to notice is that the $l = 0$ reduced wavefunctions are approximately linear near the origin (as you can see from the numerical solutions), so in these cases the polynomial must have a linear r^1 term. For the ground state this term is sufficient, while each additional node requires a term in the polynomial with the next-higher power of r .

For $l = 1$ the wavefunctions are concave-up near the origin, so a reasonable guess is that $u(r)$ begins with an r^2 term, again adding a term with the next-higher power of r for each additional node. For $l = 2$, the $u(r)$ functions begin with r^3 , and so on. This pattern is actually easier to remember for the *unreduced* radial wavefunctions, $R(r) = u(r)/r$, which begin with the power r^l .

Of course you can verify this pattern, at least for the one-bump wavefunctions, by simply plugging the formula into the radial Schrödinger equation and showing that it works. For the two-bump wavefunctions, you can make up a letter for the coefficient of the next polynomial term, plug in the formula, and solve for the value of the coefficient that works. Working out the polynomials for wavefunctions with

$$\begin{aligned}
R_{10} &= 2e^{-r} \\
R_{20} &= \frac{1}{\sqrt{2}} (1 - \frac{1}{2}r) e^{-r/2} \\
R_{21} &= \frac{1}{\sqrt{24}} r e^{-r/2} \\
R_{30} &= \frac{2}{\sqrt{27}} (1 - \frac{2}{3}r + \frac{2}{27}r^2) e^{-r/3} \\
R_{31} &= \frac{8}{27\sqrt{6}} (1 - \frac{1}{6}r) r e^{-r/3} \\
R_{32} &= \frac{4}{81\sqrt{30}} r^2 e^{-r/3} \\
R_{40} &= \frac{1}{4} (1 - \frac{3}{4}r + \frac{1}{8}r^2 - \frac{1}{192}r^3) e^{-r/4} \\
R_{41} &= \frac{\sqrt{5}}{16\sqrt{3}} (1 - \frac{1}{4}r + \frac{1}{80}r^2) r e^{-r/4} \\
R_{42} &= \frac{1}{64\sqrt{5}} (1 - \frac{1}{12}r) r^2 e^{-r/4} \\
R_{43} &= \frac{1}{768\sqrt{35}} r^3 e^{-r/4}
\end{aligned}$$

Table 6.2: Radial wavefunctions $R_{nl}(r)$ for the hydrogen atom, in atomic units. To translate these expressions into conventional units, replace r with r/a_0 wherever it appears, and multiply each expression by $a_0^{-3/2}$, where a_0 is the Bohr radius, equation 6.59.

more than two bumps is rather tedious, so at that point you're probably better off just plowing through the full power-series solution, which leads to a recursion relation for the polynomial coefficients.

These polynomials have names, by the way: After factoring out the overall powers of r , the remaining polynomials (suitably normalized) are called *associated Laguerre polynomials*. They're denoted in general by the letter L and labeled by a pair of integers, one written as a subscript and the other written as a superscript. But awkwardly, for our purposes, the polynomial that appears in $R_{nl}(r)$ is $L_{n-l-1}^{2l+1}(\frac{2r}{n})$. The full formula for $R_{nl}(r)$, normalized and everything, is⁶

$$R_{nl}(r) = \sqrt{\frac{(n-l-1)!}{(n+l)!}} \frac{2}{n^2} \left(\frac{2r}{n}\right)^l L_{n-l-1}^{2l+1}(\frac{2r}{n}) e^{-r/n}. \quad (6.65)$$

Traditionally one would use this formula, together with a table of the associated Laguerre polynomials, to obtain formulas for the hydrogen radial wavefunctions. Nowadays it's easier to use Mathematica, in which you can invoke the needed polynomials as `LaguerreL[n-1, 2l+1, 2r/n]`. I often find it most convenient, though, to simply work from an explicit table of the radial wavefunctions, so I've provided one (going up to $n = 4$) in Table 6.2.

⁶There are competing conventions for the normalization of the associated Laguerre polynomials, so the formula for the normalized R_{nl} may appear different in other textbooks. Here I'm following the normalization convention used by Mathematica.

You can find a marvelous web app for visualizing hydrogen wavefunctions at <http://falstad.com/qmatom/>.

Problem 6.30. Look up the exact formula for the $n = 3, l = 0$ hydrogen radial wavefunction in Table 6.2, and (a) plot it using atomic units. (b) Also plot the corresponding reduced radial wavefunction and check that this plot agrees (aside from normalization) with the one you made in Problem 6.29. (c) Show (either by hand or using Mathematica) that this exact function is normalized.

Problem 6.31. Look up the formula for the $n = 3, l = 1$ hydrogen radial wavefunction in Table 6.2, and then write down the corresponding reduced function $u(r)$. You may omit the overall normalization constant. Check explicitly, by hand, that this $u(r)$ satisfies the reduced radial Schrödinger equation.

Problem 6.32. Use Mathematica to produce density plots showing a slice, in the xz plane (with x horizontal and z vertical), through the probability density function for each of the $n = 2$ hydrogen wavefunctions (there are four of them). Note that you'll need to express θ in terms of x and z (and set $y = 0$ for your slice). Then use colored pencils (or something similar) to sketch three-dimensional “cloud” diagrams of each of these wavefunctions, using colors to represent phases as usual. It's difficult to draw the wavefunctions accurately in this way, but do the best you can. Take the z axis to point upward, with the viewer's location somewhat above the xy plane for perspective.

Problem 6.33. Repeat the previous problem for the $n = 3$ hydrogen wavefunctions.

6.6 The helium atom

On a scale of 0 to 10 in difficulty, the infinite square well is a 1. The one-dimensional harmonic oscillator is a 3. The hydrogen atom is an 8. And the helium atom is a 99.

But let's take a crack at it anyway, and see how far we can get.

I'll attack the problem by starting with the known solutions for the hydrogen atom, then introducing three changes, one at a time:

1. Increase the nuclear charge from e to $2e$.
2. Introduce a second electron that's also attracted to the nucleus, but neglect the repulsive force between the two electrons.
3. Finally, try to calculate the effects of the electron-electron repulsion.

Increasing the nuclear charge

For a single electron in the vicinity of a nucleus containing Z protons, the potential energy function becomes

$$V(r) = -\frac{Ze^2}{4\pi\epsilon_0 r} \frac{1}{r}. \quad (6.66)$$

Everything else in the Schrödinger equation is the same as for hydrogen, so there's no need to solve it from scratch; we can simply modify the hydrogen solutions by slipping in a factor of Z wherever e^2 appears. So the Bohr radius,

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{e^2m_e} = 0.529 \times 10^{-10} \text{ m}, \quad (6.67)$$

gets divided by Z , meaning that all the wavefunctions get pulled in closer to the nucleus by a factor of Z . Meanwhile the hydrogen energy levels,

$$E_n = -\frac{1}{2} \left(\frac{e^2}{4\pi\epsilon_0} \right)^2 \frac{m_e}{\hbar^2} \frac{1}{n^2} = -\frac{13.6 \text{ eV}}{n^2}, \quad (6.68)$$

get multiplied by a factor of Z^2 , meaning that all the levels become much more negative as Z increases.

We *could* simply modify our natural unit system to absorb these factors of Z , but nobody does that. We will continue to use atomic units, with distances in units of a_0 and energies in units of $E_h = 27.2 \text{ eV}$. In the formulas for the wavefunctions (see Table 6.2), we'll need to replace every a_0 with a_0/Z to transition from hydrogen to a heavier one-electron ion.

Helium, of course, has $Z = 2$, so for the He^+ ion, with only one electron, the wavefunctions are *half* as large as for hydrogen and the energies are *four* times as negative.

Adding a second electron

Now imagine that we start with a He^+ ion and bring in a second electron, but magically turn off the repulsive force between the two electrons. Then each of the two electrons behaves more or less independently, feeling the same force from the nucleus and having the same energy levels and definite-energy wavefunctions.

I say “more or less,” because electrons are identical fermions, so even if they don’t exert any forces on each other, their combined wavefunction still needs to be antisymmetric under the hypothetical operation of interchanging them with each other. However, this combined wavefunction also includes their spin states. As I’ll explain in Section 7.4, the spin state alone, for a system of two spin-1/2 particles, can be either symmetric or antisymmetric. This means that the *position*-dependent part of the wavefunction can be, respectively, either antisymmetric or symmetric under interchange. The bottom line is that the antisymmetrization requirement doesn’t affect the number of available spatial wavefunctions.

Moreover, as you’ll see in a moment, all we’re really looking for here is a set of orthonormal basis functions to use in our subsequent calculations. For this purpose we don’t even need to use symmetric or antisymmetric wavefunctions; we can simply multiply a wavefunction for electron 1 by a wavefunction for electron 2. Our basis, therefore, looks something like this:

$$\psi_{100}(\vec{r}_1)\psi_{100}(\vec{r}_2), \psi_{100}(\vec{r}_1)\psi_{200}(\vec{r}_2), \psi_{200}(\vec{r}_1)\psi_{100}(\vec{r}_2), \psi_{100}(\vec{r}_1)\psi_{210}(\vec{r}_2), \dots, \quad (6.69)$$

where the subscripts are the quantum numbers n , l , and m , respectively. (Remember that these ψ functions are all “squeezed” by a factor of 2 compared to those for hydrogen, as discussed above.) Still ignoring the repulsion between the two electrons, the energies of these wavefunctions are, in atomic units,

$$E_{n_1, n_2} = -\frac{2}{n_1^2} - \frac{2}{n_2^2} = -4, -2.5, -2.5, -2.5, \dots, \quad (6.70)$$

where n_1 and n_2 are the principal quantum numbers associated with \vec{r}_1 and \vec{r}_2 , respectively. Note that, as with hydrogen, the difference in an electron’s energy between $n = 1$ and $n = 2$ is greater than the difference between $n = 2$ and $n = \infty$; this means that if *both* electrons are in excited states, the atom has enough energy to ionize itself, ejecting one electron while the other drops to the ground state.

Electron-electron repulsion

The Hamiltonian for the hypothetical system I’ve just described, with two electrons but no repulsion between them, is (in atomic units)

$$H_0 = -\frac{1}{2}\nabla_1^2 - \frac{1}{2}\nabla_2^2 - \frac{2}{r_1} - \frac{2}{r_2}, \quad (6.71)$$

where ∇_1 is the ∇ operator with derivatives taken with respect to \mathbf{r}_1 , and similarly for ∇_2 . To put in the electrostatic repulsion between the two electrons, we add one more term to the Hamiltonian:

$$\Delta V = +\frac{1}{|\vec{r}_2 - \vec{r}_1|}. \quad (6.72)$$

This simple-looking addition makes it impossible to find exact formulas for the eigenfunctions and eigenvalues. My approach, therefore, will be to construct the Hamiltonian matrix (or at least a significant part of it) in the basis of “unperturbed” two-particle wavefunctions described above, and then diagonalize this matrix to obtain the new energy levels and their associated eigenfunctions.⁷

Notice that the basis functions (6.69) are already eigenfunctions of H_0 , with the eigenvalues listed in equation 6.70. This means that the H_0 matrix is diagonal, with entries equal to the H_0 eigenvalues. Our only task, then, is to evaluate the matrix elements of ΔV , so we can add this matrix onto H_0 .

A typical matrix element of ΔV looks like this:

$$\int d^3 r_1 \int d^3 r_2 \psi_{n_1 l_1 m_1}^*(\vec{r}_1) \psi_{n_2 l_2 m_2}^*(\vec{r}_2) \frac{1}{|\vec{r}_2 - \vec{r}_1|} \psi_{n_3 l_3 m_3}(\vec{r}_1) \psi_{n_4 l_4 m_4}(\vec{r}_2). \quad (6.73)$$

Yep, it’s a *six*-dimensional integral. Unfortunately, you can’t just type such an expression into Mathematica and tell it to `NIntegrate`. We’ll have to do at least some of the integrals by hand.

⁷The rest of this section is based on R. C. Massé and T. G. Walker, “Accurate energies of the He atom with undergraduate quantum mechanics,” Am. J. Phys. **83** (8), 730–732 (2015), <https://doi.org/10.1119/1.4921821>.

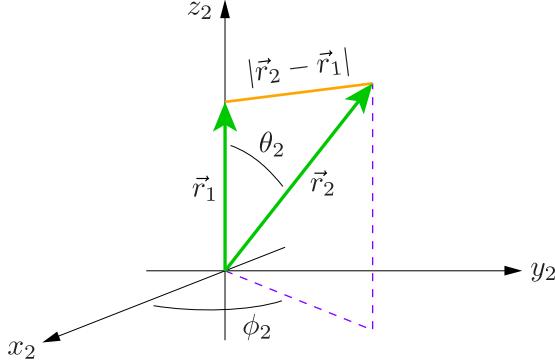


Figure 6.10: Coordinates for carrying out the integrals over ϕ_2 and θ_2 in equation 6.73.

I'll simplify the problem enormously by restricting my attention to the $l = 0$ matrix elements. These are sufficient to obtain the helium states that have zero orbital angular momentum, including the ground state, with pretty good accuracy. (There are sophisticated tricks, involving spherical harmonics and angular momentum algebra, for handling the higher- l states.) Each of the $l = 0$ wavefunctions consists of a radial wavefunction, $R_{nl}(r)$, multiplied by the trivial spherical harmonic $Y_0^0 = 1/\sqrt{4\pi}$. The only angular dependence in the integrand is then in ΔV itself, and this we can handle.

Taking \vec{r}_1 to be temporarily fixed, we orient the \vec{r}_2 coordinates with the z_2 axis pointing along \vec{r}_1 (see Figure 6.10). Then the integrand is independent of ϕ_2 , so the ϕ_2 integral gives a trivial factor of 2π . The θ_2 integral is nontrivial, but evaluates to a delightfully simple result:

$$\begin{aligned}
 \int_0^\pi \frac{\sin \theta_2}{|\vec{r}_2 - \vec{r}_1|} d\theta_2 &= \int_0^\pi \frac{\sin \theta_2}{\sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos \theta_2}} d\theta_2 \\
 &= \frac{1}{r_1 r_2} \left[\sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos \pi} - \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos 0} \right] \\
 &= \frac{1}{r_1 r_2} [(r_1 + r_2) - |r_1 - r_2|] \\
 &= \frac{2}{r_>},
 \end{aligned} \tag{6.74}$$

where $r_>$ is the greater of r_1 and r_2 . (The $\sin \theta_2$ in the numerator of the integrand comes from the measure of the integral in spherical coordinates.)

Now let's do some mental cleanup. We can combine the 2 in the numerator of equation 6.74 with the 2π from the ϕ_2 integral to cancel two of the factors of $1/\sqrt{4\pi}$ from the spherical harmonics. The θ_1 and ϕ_1 integrals give a trivial factor of 4π to cancel the other two spherical harmonics. The measures of the integrals in spherical coordinates also contain two factors of r_1 and two factors of r_2 , but we can absorb these into the radial wavefunctions to obtain the reduced radial

wavefunctions, $u_{nl}(r)$. Our generic $l_1 = l_2 = 0$ matrix element is therefore

$$\Delta V_{n_1 n_2 n_3 n_4} = \int dr_1 \int dr_2 u_{n_1 0}(r_1) u_{n_2 0}(r_2) \frac{1}{r_>} u_{n_3 0}(r_1) u_{n_4 0}(r_2). \quad (6.75)$$

We've reduced a six-dimensional integral to a two-dimensional integral. What's more, this two-dimensional integral can be done either analytically or numerically.

The paper cited in footnote 7 provides some Mathematica code to evaluate these matrix elements and then construct a truncated Hamiltonian matrix from just the following basis states (listing n_1 and n_2 , respectively, for each):

$$1,1 \quad 1,2 \quad 2,1 \quad 1,3 \quad 3,1 \quad 1,4 \quad 4,1 \quad (6.76)$$

(Notice that at least one of the electrons is always in the ground state, so none of these states have enough energy to self-ionize.) When we diagonalize this 7×7 matrix, we obtain for the three lowest eigenvalues, still in atomic units, -2.841 , -2.171 , and -2.137 . For comparison, the experimentally measured energies of the three lowest-energy states of helium are -2.903 , -2.175 , and -2.146 . So all of our values are a little too high, but we're off by only 2.1% for the ground state, and 0.2% and 0.4%, respectively, for the excited states.

The eigenvectors associated with these three lowest eigenvalues are (to three decimal places)

$$\begin{pmatrix} 0.954 \\ -0.198 \\ -0.198 \\ -0.068 \\ -0.068 \\ -0.041 \\ -0.041 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0.620 \\ -0.620 \\ -0.335 \\ 0.335 \\ -0.063 \\ 0.063 \end{pmatrix}, \quad \begin{pmatrix} 0.116 \\ 0.469 \\ 0.469 \\ -0.521 \\ -0.521 \\ -0.047 \\ -0.047 \end{pmatrix}, \quad (6.77)$$

where I've listed the components in the same order as the basis states (6.76). Notice that the first and third eigenstates are symmetric under interchange of the two particles, while the second is antisymmetric. Mentally squaring the vector components, we see that the ground state overlaps the $1,1$ basis state by about 90%, but includes about 4% each of the $1,2$ and $2,1$ states, plus smaller admixtures of the higher-energy basis states. Meanwhile, electron-electron repulsion splits the degeneracy of the two lowest excited states, with the antisymmetric state having a lower energy because it puts the two electrons farther apart on average. About 77% of this antisymmetric state comes from the $1,2$ and $2,1$ states, with most of the rest coming from $1,3$ and $3,1$. The symmetric excited state actually contains more of $1,3$ and $3,1$ than it does of $1,2$ and $2,1$. The *true* energy eigenstates would also contain small admixtures of $l > 0$ states (creating angular correlations between the electrons), and of positive-energy “continuum” states.

As we'll see in the following chapter, the symmetric and antisymmetric states of helium have different spin configurations. The ground state and the other spatially

symmetric states are called *parahelium*, while the spatially antisymmetric states are called *orthohelium*. You can find a diagram of the energy level structure in *Introduction to Quantum Mechanics* by Griffiths, or at <http://hyperphysics.phy-astr.gsu.edu/hbase/quantum/helium.html>.

Problem 6.34. In this problem you will try to get a feel for the approximate size of the electron-electron interaction energy in a helium atom. Please express your answers in atomic units.

- (a) Calculate the average value of r (that is, the average distance from the nucleus) for an electron in the ground state of a hydrogen atom.
- (a) Repeat part (a) for an electron in the ground state of a He^+ ion. (Don't start over; just modify your previous answer and justify the modification.)
- (a) Given your answer to part (b), what would be a reasonable guess for the average distance between the two electrons in the ground state of neutral helium? (Explain your answer.)
- (a) Use your answer to part (c) to estimate the average energy of repulsion between the two electrons. Compare to the actual difference between the ground-state energy of helium and what the energy would be without electron-electron repulsion.

Problem 6.35. Use either Table 6.2 or equation 6.65 to obtain explicit formulas for each of the seven two-particle helium basis states listed in 6.76. Be sure to reduce the horizontal scale by a factor of 2 to account for the doubled nuclear charge, and adjust the normalization constants accordingly. Check the normalizations. Make some plots (density, contour, or surface) of these functions, showing how they depend on r_1 and r_2 . Also plot the symmetric and antisymmetric linear combinations of the (1,2) and (2,1) wavefunctions. Finally, plot the combinations defined by each of the three eigenvectors listed in 6.77. Comment on the effects of inter-electron repulsion on these three lowest-energy wavefunctions.

Chapter 7

Spin

Chapter 6 treated the angular momentum associated with a particle's motion through space—often called *orbital* angular momentum. But a particle can also have internal angular momentum, called *spin*, which contributes just as legitimately to its total angular momentum. This chapter is all about spin.

To distinguish spin angular momentum from orbital angular momentum, we use the symbol S in place of L . So we have component operators S_x , S_y , and S_z , which we now *postulate* to obey the commutation relations

$$[S_x, S_y] = i\hbar S_z, \quad [S_y, S_z] = i\hbar S_x, \quad [S_z, S_x] = i\hbar S_y, \quad (7.1)$$

in analogy to equation 6.11. From these operators we then define a squared magnitude operator S^2 , along with raising and lowering operators S_+ and S_- , precisely in analogy to what we did with L in Section 6.2. Therefore it follows that *every* algebraic relation in Section 6.2 is still valid if we change every L to an S . We write the eigenvalues of S^2 as $s(s+1)\hbar^2$, replacing the symbol l with s for spin. We still write the eigenvalues of S_z as $m\hbar$, or as $m_s\hbar$ if there is a risk of confusion with other m symbols.

With spin, however, we can no longer argue that half-integer values of s are prohibited by the requirement that the wavefunction be single-valued (see page 171). That's because there *isn't* any spatial wavefunction that goes with spin angular momentum—or at least there needn't be. Sometimes we can look inside a particle and see how its spin is due to the relative motion of its constituents. But nobody has ever managed to look inside an electron or a neutrino or a quark, and we're pretty confident that these particles' spins are *not* due to internal motions of constituents, because they all have $s = 1/2$, which wouldn't be allowed for orbital angular momentum.

The allowed values of s , therefore, are $0, 1/2, 1, 3/2$, and so on, with the half-integer values fully allowed. The allowed values of m_s , according to the analysis of Section 6.2, run from $-s$ to s in steps of 1.

Even in those cases where spin *is* due to the motion of constituent particles, there's often no need to keep track of those particles' relative motions through

space. It's much easier to just treat spin states algebraically—and when there are no constituents, an algebraic treatment is our only choice. In this chapter I'll show how to express all of the algebra in terms of finite-dimensional vectors (called *spinors* in this case) and matrices.

Because the case $s = 1/2$ is so common among the fundamental particles, and because it is the simplest possible case (other than the trivial $s = 0$), with just two rungs on the S_z “ladder,” we will focus on this case in detail as we begin our study of spin.

7.1 Spin 1/2

If we know that a particle has $s = 1/2$ (“spin 1/2” for short), then the magnitude of its spin angular momentum vector is fixed at

$$|\vec{S}| = \sqrt{s(s+1)} \hbar = \sqrt{\frac{3}{4}} \hbar \approx 0.87 \hbar. \quad (7.2)$$

On the other hand, any particular component of its spin, such as the z component, has two possible values:

$$S_z = m\hbar = \frac{1}{2}\hbar \text{ or } -\frac{1}{2}\hbar. \quad (7.3)$$

Figure 7.1 shows a “cone diagram” for visualizing these two S_z eigenstates. As usual, these are states in which the x and y components of the angular momentum are undefined.

Spinors and spin matrices

The S_z eigenstates must form a complete basis, and since there are just two of them, we can represent *all* of the states of this system using a vector space with just two dimensions. I will use the symbol χ (Greek *chi*) for a generic two-dimensional

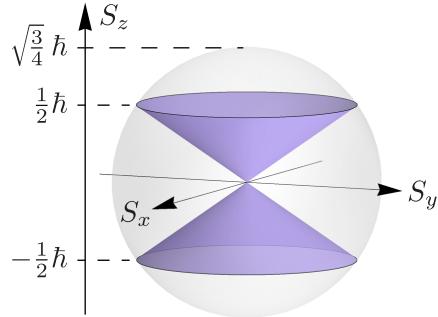


Figure 7.1: We can visualize each S_z eigenstate as a cone made of all vectors with a given $|\vec{S}|$ and S_z . For a particle with $s = 1/2$, these values are $|\vec{S}| = \sqrt{3/4} \hbar$ and $S_z = \pm(1/2)\hbar$. When a particle is in one of these states it does not have a well-defined value of S_x or S_y .

“spinor” that lives in this space. We conventionally stick with the S_z eigenstates as our basis, with the conventional assignments

$$\chi^+ = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \chi^- = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (7.4)$$

where + and – are abbreviations for the m values $+1/2$ and $-1/2$, and I’ve written them as superscripts for consistency with the notation Y_l^m . Using this basis, we can immediately write down the 2×2 matrices that represent S_z and S^2 , which are diagonal since they have these spinors as eigenspinors:

$$S_z = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad S^2 = \frac{3\hbar^2}{4} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (7.5)$$

(Check that these matrices give the correct results when they act on χ^+ and χ^- .)

The spin operators and eigenspinors for directions other than z are more interesting. To find them, we can start with the raising and lowering operators—and to find those, we can use equations 6.28, which I’ll rewrite here using spin notation:

$$\begin{aligned} S_+ \chi_s^m &= \hbar \sqrt{s(s+1) - m(m+1)} \chi_s^{m+1}, \\ S_- \chi_s^m &= \hbar \sqrt{s(s+1) - m(m-1)} \chi_s^{m-1}. \end{aligned} \quad (7.6)$$

Here χ_s^m is a generic S_z eigenspinor, for any value of s . In the special case $s = 1/2$, these equations reduce to

$$S_+ \chi^- = \hbar \chi^+, \quad S_- \chi^+ = \hbar \chi^-, \quad S_+ \chi^+ = S_- \chi^- = 0. \quad (7.7)$$

These four matrix equations, each of which is really two ordinary equations, suffice to determine the eight elements of the raising and lowering matrices:

$$S_+ = \hbar \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad S_- = \hbar \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \quad (7.8)$$

But $S_+ = S_x + iS_y$ and $S_- = S_x - iS_y$, so

$$S_x = \frac{1}{2}(S_+ + S_-) = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad S_y = \frac{1}{2i}(S_+ - S_-) = \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \quad (7.9)$$

Note that both of these matrices are Hermitian, with a symmetric real part and an antisymmetric imaginary part. Also, you can easily check (see Problem 7.1) that, together with the S_z matrix above, they obey the angular momentum commutation relations.

If you were to measure the x or y component of this particle’s spin angular momentum, the possible outcomes would be $+\hbar/2$ and $-\hbar/2$, just as for the z component. And indeed, the S_x and S_y matrices both have eigenvalues $\pm\hbar/2$. The

corresponding eigen*spinors* represent the quantum states in which these spin components have well defined S_x or S_y . I'll label these spinors with an extra superscript x or y , that is,

$$S_x \chi^{x+} = +\frac{\hbar}{2} \chi^{x+}, \quad S_x \chi^{x-} = -\frac{\hbar}{2} \chi^{x-}, \quad (7.10)$$

and similarly for $S_y \chi^{y+}$ and $S_y \chi^{y-}$. (We can then think of the symbols χ^+ and χ^- as shorthand for χ^{z+} and χ^{z-} .) To actually find these eigenspinors, we can just introduce symbols for their two components and write out the eigenvalue equation explicitly. For instance, writing the components of χ^{x+} as c_1 and c_2 , we have

$$\frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \frac{\hbar}{2} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}. \quad (7.11)$$

The top row of this matrix equation says simply $c_2 = c_1$, while the bottom row says redundantly $c_1 = c_2$. Any spinor whose two components equal each other is therefore an eigenspinor of S_x with eigenvalue $+\hbar/2$. And a completely analogous construction shows that any spinor with $c_2 = -c_1$ is an eigenspinor with eigenvalue $-\hbar/2$. By convention we usually work with spinors whose top components are real and positive, so that the normalized S_x eigenspinors are

$$\chi^{x+} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \chi^{x-} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \quad (7.12)$$

You can work out the eigenspinors of S_y in exactly the same way (see Problem 7.2).

Figure 7.2 shows the S_z and S_x eigenspinors plotted as two-dimensional vectors. Notice that, aside from the change in labeling, this diagram is identical to Figure 2.21 for the double-well two-state system. These two different systems are mathematically equivalent, or what mathematicians call *isomorphic*: each has just two basis states, with an infinite variety of complex mixtures of these states also allowed. Any such system is called a *qubit*.

Problem 7.1. Check that the 2×2 matrices for S_x , S_y , and S_z obey the angular momentum commutation relations, equations 7.1.

Problem 7.2. Find the normalized eigenspinors χ^{y+} and χ^{y-} of the S_y matrix. Use the (arbitrary) convention that the first component of each be real and positive. Show that they are orthogonal, being careful to complex-conjugate one of them when you take the inner product (see equation 4.53).

Problem 7.3. Use the methods of this section to work out the matrices and eigenspinors for $s = 1$ (a so-called *spin-1 particle*). In this case there are three possible values of the spin angular momentum along any direction, so the spinors live in a three-dimensional vector space and the matrices measure 3×3 . As usual, work in the basis of the S_z eigenspinors. It is conventional to order them so that

$$\chi^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \chi^0 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \chi^{-1} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad (7.13)$$

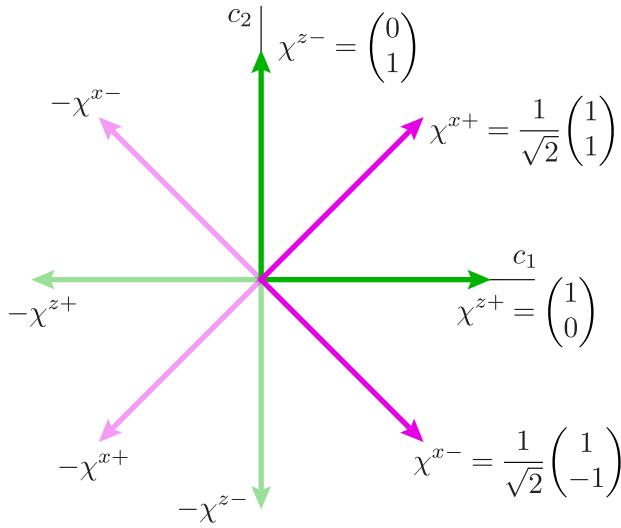


Figure 7.2: The spinors χ^{z+} , χ^{z-} , χ^{x+} , and χ^{x-} , with the component c_1 plotted horizontally and c_2 plotted vertically. An overall factor of -1 has no physical effect on a quantum state, so the negated spinors represent the same four physical states. A normalized spinor has unit length on this diagram, but the diagram cannot show spinors whose components are complex.

where the superscripts indicate the values of m_s . Start by writing down the S_z and S^2 matrices, then work out the matrices for S_+ and S_- , and use these to find S_x and S_y . Check that the commutation relations, equations 7.1, are satisfied. Finally, find the normalized eigenspinors of S_x and S_y , following the convention that the first nonzero component of each spinor is real and positive. (You could, of course, repeat this exercise for $s = 3/2$, $s = 2$, and so on. Mathematicians call this process *working out the finite-dimensional representations of the rotation group*.)

Measurement probabilities

If a spin-1/2 particle is initially in some arbitrary state,

$$\chi = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \quad (7.14)$$

and we then measure¹ its spin along a particular axis, we can calculate the probability of each possible outcome as the square of the component along the corresponding eigenstate, according to the Born rule.

For measurements of S_z this calculation is especially simple: The probabilities

¹Don't worry for now about exactly how you would make such a measurement. I'll describe one way of doing it in Section 7.2.

of getting the results $+\hbar/2$ and $-\hbar/2$, respectively, are

$$\begin{aligned} P(+\hbar/2) &= |\chi^+ \cdot \chi|^2 = \left| (1 \ 0) \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \right|^2 = |c_1|^2, \\ P(-\hbar/2) &= |\chi^- \cdot \chi|^2 = \left| (0 \ 1) \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \right|^2 = |c_2|^2. \end{aligned} \quad (7.15)$$

Here I'm assuming that χ is normalized, that is, $|c_1|^2 + |c_2|^2 = 1$. So, for example, if the system is initially in one of the S_x eigenstates, which have $|c_1|^2 = |c_2|^2 = 1/2$, then we have a 50-50 chance of measuring S_z to be either $+\hbar/2$ or $-\hbar/2$.

As a more interesting example, consider the initial state

$$\chi = \begin{pmatrix} 1/2 \\ \sqrt{3/4} \end{pmatrix} = \begin{pmatrix} \sqrt{1/4} \\ \sqrt{3/4} \end{pmatrix}. \quad (7.16)$$

If we measure this particle's S_z value, the two outcome probabilities are $P(+\hbar/2) = 1/4$ and $P(-\hbar/2) = 3/4$. Figure 7.3 shows this spinor and its projections onto the S_z eigenspinors. If, instead, we measure the x component of the spin for this same initial state, the corresponding probabilities would be

$$\begin{aligned} P(+\hbar/2) &= |\chi^{x+} \cdot \chi|^2 = \left| \frac{1}{\sqrt{2}} (1 \ 1) \begin{pmatrix} 1/2 \\ \sqrt{3/4} \end{pmatrix} \right|^2 = \frac{2 + \sqrt{3}}{4} \approx 0.933, \\ P(-\hbar/2) &= |\chi^{x-} \cdot \chi|^2 = \left| \frac{1}{\sqrt{2}} (1 \ -1) \begin{pmatrix} 1/2 \\ \sqrt{3/4} \end{pmatrix} \right|^2 = \frac{2 - \sqrt{3}}{4} \approx 0.067. \end{aligned} \quad (7.17)$$

Figure 7.3 also shows these projections, and we plainly see that χ has a much larger component in the χ^{x+} direction than in the χ^{x-} direction. The χ^{x-} component

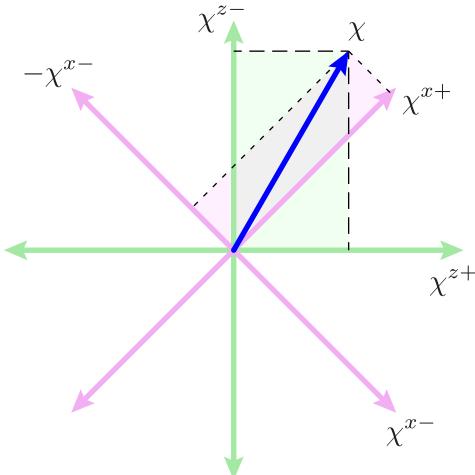


Figure 7.3: Projections of the spinor χ (equation 7.16) onto the eigenspinors of S_z (long-dashed lines) and S_x (short-dashed lines).

happens to be negative in this case, but the minus sign has no physical consequence because we square the vector component to get the measurement probability.

All of the spinor components in the preceding examples have been real, but complex-valued spinor components are perfectly permissible and often unavoidable. Unfortunately, Figures 7.2 and 7.3 suffer from the serious drawback that they cannot show spinors with complex components. Fortunately, the algebraic formula for the Born rule works fine whether or not the spinor components are real; just be sure to complex-conjugate one of the two spinors whenever you take the dot product, as prescribed by equation 4.53.

Problem 7.4. Suppose that a spin-1/2 particle is in the state

$$\chi = \begin{pmatrix} \sqrt{3/5} \\ -\sqrt{2/5} \end{pmatrix}.$$

If you now measure the z component of its spin, what are the possible outcomes and their probabilities? What if instead you measure the x component of its spin? Plot this spinor on a diagram like Figure 7.3, showing its decomposition into components in both the S_z and S_x eigenbases. Does the diagram seem consistent with your calculated answers?

Problem 7.5. In the text above I argued that if a spin-1/2 particle is in either S_x eigenstate and you measure S_z , you have a 50-50 chance of getting $+\hbar/2$ or $-\hbar/2$. Argue that the same is true if the particle is initially in an S_z eigenstate and you measure S_x . Then use the S_y eigenspinors from Problem 7.2 to determine the probabilities of S_x and S_z outcomes when the particle is in an S_y eigenstate, and vice-versa.

Problem 7.6. Suppose that a spin-1/2 particle is in the state

$$\chi = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1+i \end{pmatrix}.$$

Check that this spinor is normalized. Then find the probability of each possible outcome for measurements of the x , y , and z components of this particle's spin.

Problem 7.7. Run the Spins Laboratory web app at <http://physics.weber.edu/schroeder/software/Spins.html>. This app simulates a beam of spin-1/2 particles that enter an “analyzer,” which splits them into two beams based on whether a particular spin component has the value $+\hbar/2$ (labeled \uparrow) or $-\hbar/2$ (labeled \downarrow). Simulated counters then record the particles’ arrival, counting how many are in each of the separated beams. Thus, this experiment effectively measures the spin of each particle along the chosen direction. You can change the direction by clicking/tapping on the large letter X on the analyzer. The particle gun, on the left, can prepare the particles randomly (R) or in one of four specific initial states (labeled 1 through 4). Use the simulation to find the normalized spinors corresponding to each of these four initial states. To do this, you’ll need to make measurements of S_x , S_y , and S_z . To resolve the overall phase ambiguities, follow the convention that the first component of each spinor is real and positive. Be sure to document your work and make your reasoning clear.

The Bloch sphere

Although you can't plot spinors with complex components directly on Figure 7.2 or 7.3, there is an alternate visualization tool that doesn't suffer from this drawback.

The trick is to recognize that any normalized two-component spinor can be written in the form

$$\chi = \begin{pmatrix} \cos \frac{\theta}{2} \\ e^{i\phi} \sin \frac{\theta}{2} \end{pmatrix}, \quad (7.18)$$

for some angles θ and ϕ . To see why this is always possible, recall first that an *overall* phase factor $e^{i\delta}$ has no physical consequence, so if the top component of our spinor is negative or complex, then without loss of generality we can always multiply by some $e^{i\delta}$ (that is, rotate *both* components by some angle δ in the complex plane) to make the top component real and positive. Then, because the top component must be between 0 and 1, we can write it as the cosine of some angle between 0 and $\pi/2$, or equivalently, $\cos(\theta/2)$ where θ is between 0 and π . Once we write the top component in this form, normalization requires that the bottom component have magnitude $\sin(\theta/2)$, and the phase factor $e^{i\phi}$ then allows for the bottom component to have any complex phase, where ϕ can vary from 0 to 2π .

We now interpret the angles θ and ϕ as the usual polar and azimuthal angles in a spherical coordinate system, plotting our spinor as a point on the unit sphere, as shown in Figure 7.4. This sphere is called the *Bloch sphere*, after Swiss-American physicist Felix Bloch, 1905–1983. Every two-component spinor χ corresponds to a unique point on the Bloch sphere, and every point on the Bloch sphere corresponds to a unique physical state (though not to a unique χ , because $e^{i\delta}\chi$, for any δ , would

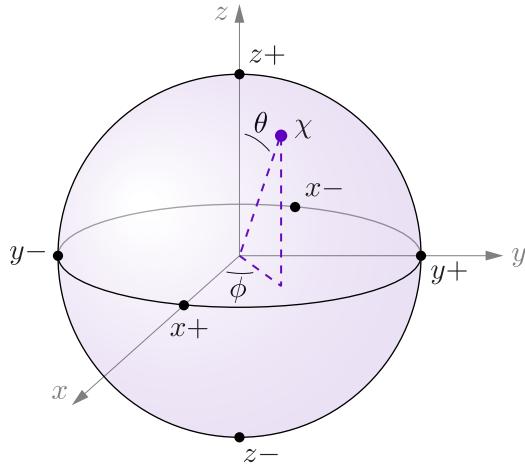


Figure 7.4: Representation of spin-1/2 states on the Bloch sphere. An arbitrary state χ is located on the sphere using the coordinates θ and ϕ defined by equation 7.18. The six eigenstates of S_x , S_y , and S_z are in their natural locations, where the rectangular coordinate axes intersect the sphere.

correspond to the same point on the Bloch sphere).

To build some intuition for the Bloch sphere, let's interpret some particular points on it. At the “north pole” $\theta = 0$, so according to equation 7.18 the corresponding spinor is $\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \chi^{z+}$. At the “south pole” $\theta = \pi$, so the spinor's top component is 0 and its bottom component is $e^{i\phi}$; this is a χ^{z-} eigenstate for any ϕ , but we might as well set $\phi = 0$ and say it's $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. On the “equator,” at $\theta = \pi/2$, the spinor's top component is $1/\sqrt{2}$; its bottom component has this same magnitude but varies in phase with ϕ , so this component is $1/\sqrt{2}$ on the $+x$ axis, $i/\sqrt{2}$ on the $+y$ axis, $-1/\sqrt{2}$ on the $-x$ axis, and $-i/\sqrt{2}$ on the $-y$ axis. By equation 7.12 and Problem 7.2, these states are therefore χ^{x+} , χ^{y+} , χ^{x-} , and χ^{y-} , respectively.

By now you've probably figured out the overall pattern:

- Any point (θ, ϕ) on the Bloch sphere corresponds to the “up” or $+\hbar/2$ eigenstate for the spin component along that particular direction.
- The “down” or $-\hbar/2$ eigenstate for the same spin component is at the point antipodal to the “up” state, precisely on the opposite side of the Bloch sphere.

Problems 7.8 and 7.9 ask you to prove these statements in general.

I hope you agree that the Bloch sphere provides a delightfully intuitive representation of the states of a spin-1/2 particle. In fact it's so intuitive that you may be wondering why we even bother with the representation in terms of two-component spinors, or with pictures like Figures 7.2 and 7.3. The answer is that the Bloch sphere has a couple of serious limitations. First, the Bloch sphere *doubles the angle* between any given pair of states in the actual vector space, so you can't use it directly to construct projections of one state onto another (e.g., for calculating measurement probabilities) as in Figure 7.3. Second, the Bloch sphere does not readily generalize to higher spin ($s = 1$, $s = 3/2$, etc.), so we need spinor technology for those cases.

One other thing the Bloch sphere *can* do, however, is represent other two-state systems (qubits) besides spin-1/2 particles. Once you've expressed the state of any such system as a two-component column vector, you can use equation 7.18 to plot that state on the Bloch sphere. One good example is the double-well two-state system from the end of Chapter 2. Other examples include the polarization states of a photon (see Section 7.3), certain other photon properties, and the states of a superconducting Josephson junction. Despite the striking physical differences between these systems, we can represent them all with exactly the same mathematics (whether complex two-component vectors or points on the Bloch sphere). In the growing discipline of quantum information science, it is now common to ignore the qubits' physical manifestations and focus instead on mathematical algorithms for manipulating collections of qubits of any type—just as traditional computer scientists rarely worry about whether their classical bits are realized as transistors or magnetic domains or fiber-optic light pulses.

Problem 7.8. What if you measure the spin of a particle along an axis that's not x , y , or z ? If we define the axis by an arbitrary unit vector

$$\hat{n} = \hat{i} \sin \theta \cos \phi + \hat{j} \sin \theta \sin \phi + \hat{k} \cos \theta$$

(see equations 6.2), then the corresponding spin operator is

$$S_{\hat{n}} = \vec{S} \cdot \hat{n} = S_x \sin \theta \cos \phi + S_y \sin \theta \sin \phi + S_z \cos \theta. \quad (7.19)$$

We can also call this operator $S_{\theta\phi}$. Write this operator, for a spin-1/2 particle, explicitly as a 2×2 matrix (in terms of θ and ϕ). Then prove that the spinor χ of equation 7.18 is an eigenspinor of this matrix with eigenvalue $+\hbar/2$. (You'll need to use some trig identities. I suggest writing everything in terms of $\sin(\theta/2)$ and $\cos(\theta/2)$.)

Problem 7.9. For a given θ and ϕ on the Bloch sphere, what are the angles that define the antipodal direction? Write the spinor for this antipodal direction in the form of equation 7.18, then prove that it is orthogonal to the spinor for the original direction (θ, ϕ) . Conclude by interpreting this antipodal spinor as an eigenspinor of the $S_{\theta\phi}$ operator of Problem 7.8.

Problem 7.10. For the special case where $\phi = 0$, we can refer to the general spin operator of equation 7.19 simply as S_θ . What are the eigenspinors of this operator, in terms of θ ? Plot these eigenspinors, for a generic choice of θ , on a copy of Figure 7.2. How do the angles on this diagram relate to the physical angle θ ?

Problem 7.11. For the special case where $\phi = 0$, we can refer to the general spin operator of equation 7.19 simply as S_θ . Suppose that a spin-1/2 particle is in the state χ^{z+} , and you measure S_θ for $\theta = 2\pi/3 = 120^\circ$. What are the probabilities of the outcomes $+\hbar/2$ and $-\hbar/2$?

Problem 7.12. Draw a Bloch sphere for the double-well two-state system of Section 2.5, working in the basis where the states ψ_1 and ψ_2 are represented by $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, respectively. Label these states on the sphere, and also label the states ψ_L , ψ_R , ψ_+ , and ψ_- of equations 2.33 and 2.34.

7.2 Spins in magnetic fields

When a particle with spin also has an electric charge (as does an electron or proton or any nucleus), or has electrically charged constituents (as does an atom or a neutron), the spin gives rise to a *magnetic dipole moment*. That is, the particle acts as a tiny bar magnet, feeling a torque when it's placed in an external magnetic field.

The dipole moment is a vector, denoted $\vec{\mu}$, which we define to point in the direction of a bar magnet's north-seeking pole (picture a compass needle). If the external magnetic field is \vec{B} , then the torque causes the dipole's energy to be lowest when it's aligned parallel to \vec{B} , and highest when it's antiparallel. In general, the energy is

$$E_{\text{dipole}} = -\vec{\mu} \cdot \vec{B}. \quad (7.20)$$

So $\vec{\mu}$ has SI units of joules per tesla, and its magnitude is the amount of work required to twist the dipole from pointing parallel to the field to pointing perpendicular, or from perpendicular to antiparallel, per unit field strength.

For any given particle, $\vec{\mu}$ will be directly proportional to the spin vector:

$$\vec{\mu} = \gamma \vec{S}, \quad (7.21)$$

where γ is a constant of proportionality, called the *gyromagnetic ratio*, which depends on the type of particle. In the case of a classical spinning object, it's straightforward to calculate γ if you know how the object's mass and charge are distributed; if the mass m and charge q are distributed identically, then $\gamma = q/(2m)$. Notice that if q is negative, $\vec{\mu}$ and \vec{S} point in opposite directions. For a quantum particle, a first-principles calculation of γ is beyond the scope of this book. But the total charge divided by the total mass is still a good order-of-magnitude guess, at least for charged particles. For an electron, γ is very close to, but not exactly, $-e/m$ (where $e = 1.6 \times 10^{-19}$ C).

Combining equations 7.20 and 7.21, we can write the Hamiltonian for a spinning particle in a uniform magnetic field as

$$H = -\gamma \vec{B} \cdot \vec{S}. \quad (7.22)$$

This Hamiltonian will give different energies to different spin states, and we can now calculate those energies and the splittings between them. This Hamiltonian will also govern the time evolution of an arbitrary initial spin state, analogously to the Hamiltonian for a particle moving through space.

Without loss of generality, let's take \vec{B} to point along the z axis. Then the Hamiltonian is simply a constant times S_z :

$$H = -\gamma B S_z. \quad (7.23)$$

The eigenvalues of H are $-\gamma B m_s \hbar$, and since m_s varies by increments of 1, the splitting between adjacent energy levels is $\gamma \hbar B$. The corresponding eigenstates are, by the usual convention, our basis spinors. In the specific case $s = 1/2$, these eigenstates are

$$\chi^+ = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ with } E = -\frac{\gamma B \hbar}{2}, \quad \text{and} \quad \chi^- = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ with } E = +\frac{\gamma B \hbar}{2}. \quad (7.24)$$

If the particle is initially in one of these eigenstates, then its time dependence is simply an overall wiggle factor, $\exp(\pm i \gamma B t / 2)$, which has no effect on any measurement outcomes.

Suppose, however, that this spin-1/2 particle is initially in some other state—say χ^{x+} (equation 7.12):

$$\chi(t=0) = \chi^{x+} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}} \left[\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right]. \quad (7.25)$$

This state is a mixture of the two S_z eigenstates, each of which, in the presence of the magnetic field pointing in the z direction, evolves with a different wiggle factor:

$$\chi(t) = \frac{1}{\sqrt{2}} \left[\begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{i \gamma B t / 2} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{-i \gamma B t / 2} \right] = \frac{1}{\sqrt{2}} \begin{pmatrix} e^{i \gamma B t / 2} \\ e^{-i \gamma B t / 2} \end{pmatrix}. \quad (7.26)$$

To interpret this time-dependent spinor we can plug in some particular t values and see what we get. For instance, when $\gamma B t = 4\pi$ (or any multiple of 4π), it comes back to simply χ^{x+} . So it's tempting to conclude that the particle's spin state oscillates with a period of $4\pi/(\gamma B)$.

In fact, the spin state does oscillate, but the period isn't really $4\pi/(\gamma B)$. For notice that after just *half* this time, that is, when $\gamma B t = 2\pi$, each of the two wiggle factors equals $e^{i\pi} = -1$, so $\chi = -\chi^{x+}$, which is physically the same state as at $t = 0$. The *physical* oscillation period, therefore, is only $2\pi/(\gamma B)$, corresponding to an angular frequency of

$$\omega_L = \gamma B, \quad (7.27)$$

called the *Larmor frequency* (after Irish physicist Joseph Larmor, 1857–1942).

To see what happens at intermediate times during this oscillation, it's a bit easier if we factor out the unphysical phase factor once and for all. Notice that we can write $\chi(t)$ in the form

$$\chi(t) = e^{i\gamma B t/2} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ e^{-i\gamma B t} \end{pmatrix}, \quad (7.28)$$

so that the part of the spinor in the parentheses matches our usual convention of making the top component real and positive. The overall factor $e^{i\gamma B t/2}$, because it is common to both spinor components, is an unphysical phase that we can ignore. We can now recognize that when $\gamma B t = \pi$ (that is, after half the physical oscillation period), $\chi(t)$ is physically equivalent to χ^{x-} . Also, as you might expect, after a quarter of the physical oscillation period, $\chi(t)$ is physically equivalent to one of the S_y eigenstates (which one depends on the sign of γ).

The oscillation of particle's spin state in an external magnetic field is called *Larmor precession*, and you can easily work out other examples—with other initial states, other field directions, or higher values of s . Although I usually find it easiest to write the time dependence in terms of explicit wiggle factors, we can also express it using the time-dependent Schrödinger equation:

$$i\hbar \frac{\partial \chi}{\partial t} = H\chi. \quad (7.29)$$

Although the H operator is very different here than for the spatial wavefunctions that we worked with in earlier chapters, the TDSE has exactly the same form as long as we write it abstractly in terms of H .

Larmor precession plays a critical role in current research into quantum computing. If we use a spin-1/2 particle to represent each qubit of information, then we can change the value of a qubit by briefly turning on a magnetic field to rotate its state by some desired amount. For example, if we use χ^{z+} to represent the number 0 and χ^{z-} to represent the number 1, we can turn a 0 into a 1, or a 1 into a 0, by turning on a magnetic field in a direction perpendicular to z (for example, the x or y direction) for just enough time to accomplish half of a full oscillation. Similarly, by turning on the field for half as long as that—a quarter of an oscillation

period—we can put the qubit into a superposition state that contains equal parts 0 and 1. Our choice of the direction of \vec{B} will determine exactly which superposition state this is (that is, the relative phase of the 0 and 1 contributions). As long as we have precise control over the direction of \vec{B} and the time it remains turned on, we can put the qubit into any state we like.

Problem 7.13. Consider a spin-1 (that is, $s = 1$) particle in a magnetic field \vec{B} that points in the $+z$ direction.

- (a) What are this particle's energy eigenvalues (in terms of γ and B) and the corresponding eigenspinors?
- (b) If the particle has $\gamma > 0$ and starts out in the state with $S_x = +\hbar$ (that is, in the state χ_1^{x+}), what is the sequence of S_x and S_y eigenstates that it oscillates through, and what are the period and angular frequency of its Larmor precession, in terms of γB ?
- (c) What happens if instead the particle starts in the state with $S_x = 0$?

The Stern-Gerlach experiment

So far in this section I've been assuming that the external magnetic field \vec{B} is uniform (in space). If the field is *nonuniform*, then in addition to exerting a torque on our spinning dipole it also exerts a net force: $\vec{F} = \nabla(\vec{\mu} \cdot \vec{B})$. You've felt this type of force every time you've held two bar magnets near each other.

The net force from a nonuniform \vec{B} field gives us a way to separate particles with different values of, say, μ_z . The typical procedure is to put the particles into a horizontal beam, moving in the y direction, that passes through a nonuniform magnetic field pointing in the z direction (see Figure 7.5). If the z component of the field is also nonuniform, then the particles are deflected in the z direction by varying amounts, depending on their μ_z values. Such experiments were first performed with silver atoms by Stern and Gerlach in the mid-1920s.

A Stern-Gerlach magnet “entangles” a particle’s spin state with its spatial wavefunction, allowing us to determine the spin by measuring the position. Although

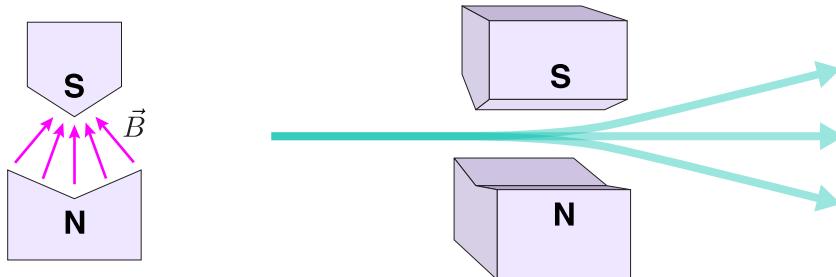


Figure 7.5: Left: End view of a Stern-Gerlach magnet, with poles shaped to produce an upward magnetic field that is nonuniform—in this case stronger as you go up. Right: Particles passing through such a magnet feel a vertical force proportional to their μ_z values, so a beam of particles is split into multiple beams, one with each possible value of μ_z .

we often speak as if the magnet “measures” the particle’s spin, that’s not quite right because we still need a separate detector of some sort to measure the particle’s position. In fact, we could use a succession of similar magnets to recombine the beams, and if no record is kept of which path a particle follows, then no measurement is performed and the particle’s original spin state (before it entered the device) is unchanged. You can use the Spins Laboratory web app, at <http://physics.weber.edu/schroeder/software/Spins.html>, to explore a variety of scenarios involving successive Stern-Gerlach devices with varying orientations.

7.3 Photon polarization

The photon is a “spin-1” ($s = 1$) particle, so we might expect it to have three independent polarization states. In fact there are only two, for reasons related to how photons always travel at the speed of light. We ordinarily measure a photon’s angular momentum about an axis parallel to its direction of motion, and when we do we obtain either $+\hbar$ or $-\hbar$, corresponding to right-handed and left-handed circularly polarized light. The $m_s = 0$ “middle rung” of the ladder of angular momentum states is missing.

Because it has only two independent polarization states, we normally abandon the usual conventions for angular momentum and instead treat a polarized photon as any other two-state (qubit) system, using two-component complex vectors. In this context the vectors are called *Jones vectors* (after American physicist R. Clark Jones, 1916–2004). For our basis states we use not the circularly polarized states but rather the states of horizontal and vertical *linear* polarization, measured relative to some arbitrary reference direction:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \leftrightarrow \text{horizontal}, \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} \leftrightarrow \text{vertical}. \quad (7.30)$$

The real linear combinations of these vectors then represent diagonally polarized states, for instance,

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \leftrightarrow +45^\circ \text{ diagonal}, \quad \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \leftrightarrow -45^\circ \text{ diagonal}. \quad (7.31)$$

The left- and right-handed circularly polarized states are represented by the complex vectors

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \leftrightarrow \text{left-handed}, \quad \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix} \leftrightarrow \text{right-handed}. \quad (7.32)$$

Unfortunately there is some ambiguity in these correspondences, because positive and negative angle definitions depend on whether one is viewing the photon from ahead or behind. If you study optics further you’ll need to get these conventions straight, but for our purposes here we can live with the ambiguity.

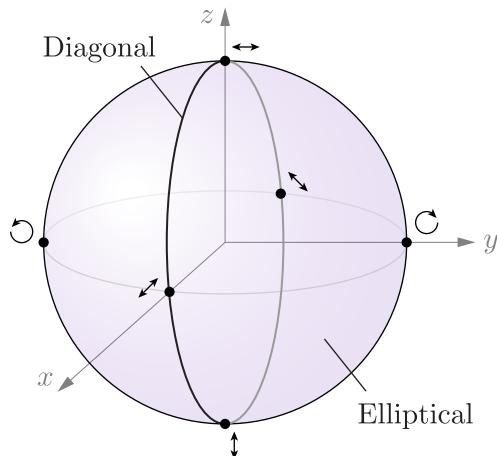


Figure 7.6: Representation of photon polarization states on the Bloch sphere. Horizontal and vertical polarizations are at the north and south poles, respectively, while diagonal polarizations are along the circle in the xz plane. The two circular polarizations are where the $\pm y$ axis intersects the sphere, and everywhere else on the sphere corresponds to a type of elliptical polarization.

Figure 7.6 shows the full set of allowed photon polarization states, represented as points on the surface of the Bloch sphere (also called the Poincaré sphere in this context). As with a spin-1/2 particle, the Bloch sphere has the advantage of being in one-to-one correspondence with the full set of physical states. But unlike spin 1/2, here the correspondence between directions on the Bloch sphere and the associated directions in physical space is rather arbitrary and unnatural. That correspondence is much more direct, at least for the diagonal polarizations, when we simply plot the Jones vector components on a pair of rectangular axes (see Problem 7.14).

Photons are often easier to work with in the laboratory than spin-1/2 particles. Instead of fancy magnets, atomic beams, and vacuum systems, you can do equivalent experiments with inexpensive lasers, polarizing filters, and other optical components. A common calcite crystal will split an unpolarized beam of light into separate vertically and horizontally polarized beams, analogous to the role of a Stern-Gerlach device for beams of spin-1/2 atoms. Half-wave plates and quarter-wave plates, also made of so-called birefringent materials, can change one polarization state into another, in analogy to Larmor precession of a spin-1/2 particle in a magnetic field. Finally, because photons travel at the speed of light, they are handy for sending messages and for performing fascinating experiments in which two entangled particles are separated by a great distance before being measured.

Problem 7.14. Plot the Jones vectors for vertical, horizontal, and $\pm 45^\circ$ diagonal polarizations on a pair of rectangular axes, with the top component on the horizontal axis and the bottom component on the vertical axis, as I did for spin-1/2 spinors in Figure 7.2. Discuss the relative advantages and disadvantages of this visualization, compared to the Bloch sphere in Figure 7.6.

7.4 Addition of angular momenta

Often we work with composite systems—atoms, nuclei, and so on—whose total angular momentum is the combination of multiple contributions from constituent particles. Even a single constituent particle can contribute both orbital and spin angular momentum, and with multiple constituents there can be many contributions to the total. In these cases we need to know how the various eigenstates of the total angular momentum are related to the eigenstates of the constituents' individual angular momenta.

The simplest such case is when there are just two spin-1/2 particles, such as the electron and proton in a hydrogen atom (assuming for simplicity that there is no orbital angular momentum). This system's spin state lives in a four-dimensional vector space, for which the most natural basis consists of the states in which each particle has well-defined S_z . One way to write these states is

$$\chi^+ \chi^+, \quad \chi^+ \chi^-, \quad \chi^- \chi^+, \quad \chi^- \chi^-, \quad (7.33)$$

where my notational convention is that the spinor on the left is for particle 1 (perhaps the electron) and the spinor on the right is for particle 2 (perhaps the proton). These juxtapositions of spinors are actually *tensor product* states, as described in Section 4.4 and in Appendix A.4. In a little while I'll show how to represent them as four-component spinors, but that isn't actually necessary for now.

The system's total angular momentum vector \vec{S} is simply the sum of \vec{S} for each particle, so we can write the operator relation

$$\vec{S} = \vec{S}^{(1)} + \vec{S}^{(2)}, \quad (7.34)$$

where the superscripts in parentheses indicate particles 1 and 2, respectively, while an operator without a superscript is for the system as a whole. The same relation holds for each component of \vec{S} , including S_z , so we can easily show that each of the states in 7.33 is an eigenstate of total S_z :

$$\begin{aligned} S_z \chi^{(1)} \chi^{(2)} &= (S_z^{(1)} + S_z^{(2)}) \chi^{(1)} \chi^{(2)} \\ &= (S_z^{(1)} \chi^{(1)}) \chi^{(2)} + \chi^{(1)} (S_z^{(2)} \chi^{(2)}) \\ &= (m_1 + m_2) \hbar \chi^{(1)} \chi^{(2)}, \end{aligned} \quad (7.35)$$

where m_1 and m_2 are the individual S_z quantum numbers and I've used the fact that $S_z^{(1)}$ acts only on the state of particle 1, while $S_z^{(2)}$ acts only on the state of particle 2. Thus, the total S_z values of the four states listed in 7.33 are, respectively, \hbar , 0, 0, and $-\hbar$.

This list of S_z values is that of a system with $s = 1$, although such a system would have just one $m = 0$ state, and we instead have two. But the states with $m = \pm 1$ are unambiguous:

$$\chi_1^1 = \chi^+ \chi^+, \quad \chi_1^{-1} = \chi^- \chi^-, \quad (7.36)$$

where on the left I've written the spinor for the combined system, using the notation χ_s^m in analogy with Y_l^m . Now, to find the state χ_1^0 , we can just act on one of these with a ladder operator. For instance, the lowering operator applied to the first of these states gives

$$\begin{aligned} S_- \chi_1^1 &= (S_-^{(1)} + S_-^{(2)}) \chi^+ \chi^+ \\ &= (S_-^{(1)} \chi^+) \chi^+ + \chi^+ (S_-^{(2)} \chi^+). \end{aligned} \quad (7.37)$$

Applying equation 7.6 to both sides of this equation, we obtain

$$\sqrt{2} \hbar \chi_1^0 = (\hbar \chi^-) \chi^+ + \chi^+ (\hbar \chi^-), \quad (7.38)$$

or

$$\chi_1^0 = \frac{1}{\sqrt{2}} (\chi^- \chi^+ + \chi^+ \chi^-). \quad (7.39)$$

Here again it's understood, in each term, that the χ on the left is for particle 1 and the χ on the right is for particle 2. If we apply the lowering operator again to equation 7.39 we get χ_1^{-1} as expected. Alternatively, we could start with the state $\chi_1^{-1} = \chi^- \chi^-$ and apply the raising operator once to get equation 7.39, and again to get χ_1^1 .

The χ_1^0 state is therefore a 50-50 *mixture* of the basis states $\chi^+ \chi^-$ and $\chi^- \chi^+$. Meanwhile, there is still an orthogonal mixture in which the + sign in equation 7.39 is replaced with a -. That state also has $m = 0$, but it's not part of the $s = 1$ quantum ladder at all. In fact it constitutes a one-rung quantum ladder of its own, that is, a state with $s = 0$:

$$\chi_0^0 = \frac{1}{\sqrt{2}} (\chi^- \chi^+ - \chi^+ \chi^-). \quad (7.40)$$

You can easily check that if you act on this state with either the raising or the lowering operator, you get zero.

The three states with $s = 1$ (equations 7.36 and 7.39) are collectively called the *triplet* states of the combined system, while the $s = 0$ state (equation 7.40) is called the *singlet*. In summary:

$$\chi_0^0 = \frac{1}{\sqrt{2}} (\chi^- \chi^+ - \chi^+ \chi^-) \quad \text{"singlet"} \quad (7.41)$$

$$\left. \begin{array}{l} \chi_1^1 = \chi^+ \chi^+ \\ \chi_1^0 = \frac{1}{\sqrt{2}} (\chi^- \chi^+ + \chi^+ \chi^-) \\ \chi_1^{-1} = \chi^- \chi^- \end{array} \right\} \text{"triplet"} \quad (7.42)$$

To be fully confident of the s values I've assigned to these states, you can also construct the S^2 operator for the combined system and check that it gives $2\hbar^2$ when acting on any of the triplet states, and 0 when acting on the singlet.

Notice that the three triplet states are all symmetric (unchanged) under the operation of interchanging the two particles, while the singlet is antisymmetric. These properties are critical for understanding the behavior of identical particles. For example, the two electrons in a helium atom are identical fermions, so their overall quantum state must be antisymmetric under interchange. This overall state includes both a spin state and a spatial wavefunction. If the spatial wavefunction is symmetric (the ground state or the other “parahelium” states discussed in Section 6.6), then the spin state must be the antisymmetric singlet. On the other hand, if the spatial wavefunction is antisymmetric (“orthohelium”), then the spin state must be one of the three triplet states or some linear combination thereof.

Also notice that both the singlet χ_0^0 state and the triplet χ_1^0 state are *entangled*. We can't factor either of them into an unambiguous state for one particle and an unambiguous state for the other. Rather, to describe the state of either particle we must describe the state of both, including the implied correlations. For either of these states, for instance, if you were to measure S_z for particle 1 you could obtain either $+\hbar/2$ or $-\hbar/2$, with equal probability, and the same is true if you were to measure S_z for particle 2. But if you measure S_z for *both* particles, you're 100% certain to obtain opposite results. The system's *total* S_z is well defined (even before any measurements), even though S_z for either particle individually is not. For the singlet state (but not for χ_1^0), this is also true for the spin component along any other axis.

The numerical coefficients $1/\sqrt{2}$, $-1/\sqrt{2}$, and so on, which relate the individual-particle basis states to the system states with well-defined total s , are called *Clebsch-Gordan coefficients*. When the individual particles have spins higher than $1/2$, these coefficients can be more intricate—but you can work them out using ladder operators just as in the example above (see Problem 7.19). Alternatively, you can look them up in tables. In general, if the individual particles have quantum numbers s_1 and s_2 , then the combined system can have $s = s_1 + s_2$, or $s = s_1 + s_2 - 1$, and so on down to $s = 0$ or $s = 1/2$, depending on whether $s_1 + s_2$ is an integer or half-integer.

Problem 7.15. Apply the lowering operator again to the state χ_1^0 of equation 7.39, and check that you get the expected result.

Problem 7.16. Apply the raising operator to the state $\chi_1^{-1} = \chi^- \chi^-$, and check that you get the same result as in equation 7.39. Then apply the raising operator again, and check that you get the expected result.

Problem 7.17. Check that applying either the raising or the lowering operator to the singlet state, equation 7.40, gives zero.

Problem 7.18. Rewrite the singlet state, equation 7.40, in terms of the two-component S_x eigenspinors χ^{x+} and χ^{x-} . Do this by brute-force substitution, but comment on why the result was, aside from a possible overall minus sign, inevitable.

Problem 7.19. Consider a system consisting of two spin-1 particles, that is, two particles that each have $s = 1$. (Alternatively, the two particles could each have one unit of orbital angular momentum, in which case we would use the symbol l instead

of s .) In this problem you will “add” these two spins, working out all of the Clebsch-Gordan coefficients for this system.

- (a) List all of the basis states of the combined system for which the z component of each *individual* particle’s spin is well defined. For instance, one of these states is $\chi_1^1 \chi_1^1$, in which each particle’s S_z value is $+\hbar$. As usual, the spinor on the left applies to particle 1, and the spinor on the right applies to particle 2.
- (b) Now list all of the states of the combined system for which the magnitude and z component of the *total* (spin) angular momentum are well defined. Denote each of these states with a single χ symbol, using appropriate subscripts and superscripts. Check that this list has the same number of entries as the list in part (a).
- (c) Two of the states in your first list are in one-to-one correspondence with two of the states in the second list. Which ones? Why?
- (d) Starting with either of the two states that you just identified, apply the (total) raising or lowering operator repeatedly to find expressions for the remaining states that have the same total s value, in terms of the part (a) basis states. (You’ll need to repeatedly use equations 7.6.)
- (e) To work out the expressions for the remaining part (b) states, use the fact that each of these states must be orthogonal to all of the others, along with occasional use of a raising or lowering operator.
- (f) Check your answers using the table of Clebsch-Gordan coefficients at [https://en.wikipedia.org/wiki/Table_of_Clebsch%20Gordan_coefficients](https://en.wikipedia.org/wiki/Table_of_Clebsch-Gordan_coefficients).

Matrix representations

Addition of angular momenta is a pretty abstract idea, so it may be helpful to confirm the above analysis using explicit representations of the states as four-component column spinors, and of the operators as 4×4 matrices. For variety I’ll also show how to avoid using raising and lowering operators, focusing instead on the S^2 matrix.

First let’s (arbitrarily) assign spinors to our four original basis states:

$$\chi^+ \chi^+ = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \chi^+ \chi^- = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \chi^- \chi^+ = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \chi^- \chi^- = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \quad (7.43)$$

Notice that these spinors are tensor products, computed according to the rule

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \otimes \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} c_1 d_1 \\ c_1 d_2 \\ c_2 d_1 \\ c_2 d_2 \end{pmatrix}, \quad \text{e.g.,} \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad (7.44)$$

as in equation 4.65. By convention the *left* component spinor, corresponding to particle 1, gets “expanded” into the high-level “block” structure of the product,

and then the *right* component spinor, corresponding to particle 2, is inserted into each block and multiplied by whichever component of the left spinor is there.

By an analogous construction, we can build the 4×4 versions of operator matrices. The S_z matrix for particle 1 is

$$S_z^{(1)} = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes \mathbb{1} = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad (7.45)$$

where $\mathbb{1}$ is the 2×2 identity matrix. The idea here is that each 2×2 block of the product comes from multiplying a single entry of the first matrix by the entire second matrix. Similarly, the S_z matrix for particle 2 is

$$S_z^{(2)} = \mathbb{1} \otimes \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \quad (7.46)$$

Adding these two matrices then gives the total S_z matrix,

$$S_z = S_z^{(1)} + S_z^{(2)} = \frac{\hbar}{2} \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 \end{pmatrix}. \quad (7.47)$$

You can easily check that this matrix has the correct action on each of the four basis spinors in equation 7.43: They are all eigenspinors, with eigenvalues \hbar , 0, 0, and $-\hbar$, respectively.

Now let's do the S_x matrix. For particle 1 we have

$$S_x^{(1)} = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \mathbb{1} = \frac{\hbar}{2} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad (7.48)$$

while for particle 2 we have

$$S_x^{(2)} = \mathbb{1} \otimes \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (7.49)$$

Adding these matrices gives the total S_x matrix,

$$S_x = S_x^{(1)} + S_x^{(2)} = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}. \quad (7.50)$$

By an entirely analogous construction (see Problem 7.20) you can obtain the S_y matrix,

$$S_y = \frac{\hbar}{2} \begin{pmatrix} 0 & -i & -i & 0 \\ i & 0 & 0 & -i \\ i & 0 & 0 & -i \\ 0 & i & i & 0 \end{pmatrix}. \quad (7.51)$$

Now it's an utterly mindless operation (see Problem 7.20) to square each of the three spin component matrices, then add up the results to obtain S^2 :

$$S^2 = S_x^2 + S_y^2 + S_z^2 = \hbar^2 \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}. \quad (7.52)$$

And this is the matrix we need. Notice that when it acts on the basis spinors $\chi^+\chi^+$ and $\chi^-\chi^-$, it returns these same spinors multiplied by $2\hbar^2$. So these are eigenspinors of S^2 , with eigenvalues $s(s+1)\hbar^2$, for $s=1$. But the other two basis spinors are not eigenspinors of S^2 , due to the 1's in the off-diagonal 2-3 and 3-2 places. Instead, the other two eigenspinors of S^2 are simple linear combinations of $\chi^+\chi^-$ and $\chi^-\chi^+$. One of these linear combinations is the sum:

$$S^2(\chi^+\chi^- + \chi^-\chi^+) = \hbar^2 \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} = 2\hbar^2 \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}. \quad (7.53)$$

This combination is therefore a third eigenspinor of S^2 with eigenvalue $2\hbar^2$, that is, the state with $s=1$ but $m=0$ (no angular momentum along the z axis). (To normalize this spinor we would multiply it by $1/\sqrt{2}$, but this isn't necessary to check that it is an eigenspinor.) On the other hand, the orthogonal linear combination with a relative minus sign gives

$$S^2(\chi^+\chi^- - \chi^-\chi^+) = \hbar^2 \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} = 0 \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \quad (7.54)$$

so this combination has eigenvalue zero, or $s(s+1)\hbar^2$ for $s=0$. This is the singlet state, with no angular momentum at all, along any axis. (Again, to normalize it we would multiply by $1/\sqrt{2}$.)

In summary, the normalized eigenspinors of S_z and S^2 for this system are

$$\chi_0^0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \quad \chi_1^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \chi_1^0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \quad \chi_1^{-1} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \quad (7.55)$$

The first is the singlet, while the other three make up the triplet.

Problem 7.20. Fill in the steps omitted above to obtain the S_y and S^2 matrices, equations 7.51 and 7.52. (You may wish to use Mathematica for this; Appendix B.5 shows how to work with matrices and take tensor products using the `KroneckerProduct` function.)

Problem 7.21. Find the 4×4 matrix representations of the raising and lowering operators for a system of two spin-1/2 particles, by taking tensor products as in the text above. Then check that these operators are related to S_x and S_y in the correct ways, and that they act correctly on each of the four basis states.

Problem 7.22. Take tensor products to find the four-component column spinor representations of the four S_x eigenstates for a system of two spin-1/2 particles: $\chi^{x+}\chi^{x+}$, $\chi^{x+}\chi^{x-}$, and so on. Then check that each of them is an eigenspinor of the two-particle S_x matrix, equation 7.50, with the expected eigenvalue. Which of these states are eigenstates of S^2 ? For those that aren't eigenstates of S^2 , find linear combinations that are.

Appendix A

The mathematics of quantum mechanics

Quantum mechanics describes a physical world that is radically different from the world of classical mechanics. So naturally, in order to express the principles of quantum mechanics, we need some mathematical concepts that go beyond the calculus and differential equations that we use to express the laws of classical physics. These further concepts include probability and statistics, the Dirac delta function, complex numbers, and linear algebra.

All of these mathematical ideas get at least some use in classical physics, and most of them are also taught in standard math courses, so you may already know enough about them to skip this appendix and dive into the main text of this book. On the other hand, if your comfort level with some or all of these mathematical concepts is low, then this appendix is for you. Here I've tried to provide a concise but systematic presentation of each of them, emphasizing the details that matter most in quantum mechanics.¹

A.1 Probability and statistics

Imagine that you have a stack of playing cards (perhaps a bridge hand) consisting of three 9's, two each of 2's, 5's, and 10's, and one each of the numbers 3, 6, 7, and 8. If we denote the card denomination by i and the number of such cards by $n(i)$, then.

$$n(i) = \begin{cases} 3 & \text{for } i = 9, \\ 2 & \text{for } i = 2, 5, 10, \\ 1 & \text{for } i = 3, 6, 7, 8, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.1})$$

¹Much of the inspiration for this appendix comes from Chapter 2 of Daniel T. Gillespie, *A Quantum Mechanics Primer* (International Textbook Co., Scranton, PA, 1970).

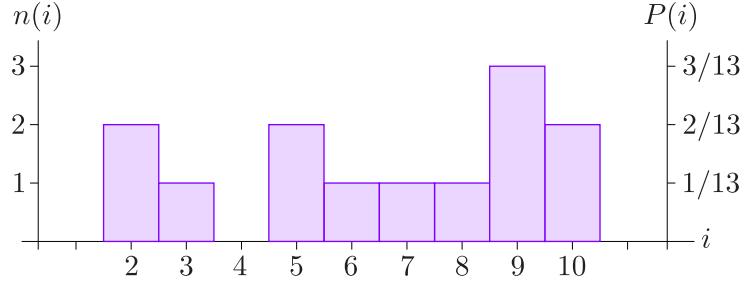


Figure A.1: A histogram of the card counts and probabilities for the example in the text.

Now if you shuffle these cards and draw one at random, the *probability* P of getting a card of denomination i (never mind the suit) is $n(i)$ divided by the total number of cards, $N = 13$:

$$P(i) = \frac{n(i)}{N} = \begin{cases} 3/13 & \text{for } i = 9, \\ 2/13 & \text{for } i = 2, 5, 10, \\ 1/13 & \text{for } i = 3, 6, 7, 8, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.2})$$

Figure A.1 shows a histogram of these card counts and probabilities. Notice that the sum of *all* the probabilities must equal 1:

$$\sum_i P(i) = 1. \quad (\text{A.3})$$

The *average* (or *mean*) of the values of the cards is the sum of all their values, divided by the total number of cards:

$$\begin{aligned} \langle i \rangle &= \frac{2 + 2 + 3 + 5 + 5 + 6 + 7 + 8 + 9 + 9 + 9 + 10 + 10}{13} = \frac{85}{13} \approx 6.54 \\ &= \frac{1}{N} \sum_i i n(i) = \sum_i i P(i). \end{aligned} \quad (\text{A.4})$$

Notice that we can compute the average as the sum over all the i values, *weighted by their probabilities*. Similarly, if we wish, we can compute the average of the *squares* of the i values by summing these squares, weighted by their probabilities:

$$\langle i^2 \rangle = \sum_i i^2 P(i). \quad (\text{A.5})$$

In fact, we can compute the average of *any* function $f(i)$ by adding up all its values, weighted by their probabilities:

$$\langle f(i) \rangle = \sum_i f(i) P(i). \quad (\text{A.6})$$

To estimate the width or “spread” of this distribution of i values, we can subtract each of them from the average, and then average these differences. Well, we *could* do that, but we’d get *zero*, because some of the differences are positive and others are negative, and the average is defined so that these differences average out! So instead, we *square* the differences, average these squares (which are never negative), and then take the square root. The result is the root-mean-square (rms) deviation or *standard deviation*, σ_i :

$$\sigma_i = \sqrt{\frac{\sum_i (i - \langle i \rangle)^2 n(i)}{N}}. \quad (\text{A.7})$$

With a bit of algebra (see Problem A.1) you can show that this definition is equivalent to

$$\sigma_i = \sqrt{\langle i^2 \rangle - \langle i \rangle^2}, \quad (\text{A.8})$$

which is usually the more convenient formula for calculating σ_i .

Problem A.1. Derive the more convenient formula for computing the standard deviation, equation A.8, from the definition A.7.

Problem A.2. For the example described in the text (and depicted in Figure A.1), compute $\langle i^2 \rangle$ and then use equation A.8 to compute the standard deviation. Mark the points on the histogram that correspond to the average and one standard deviation to either side of the average. Are most of the card numbers within one standard deviation of the average? Discuss.

Problem A.3. Using only the numbers 2 through 10, find a collection of numbers for which the standard deviation is as large as possible. Also find a collection for which the standard deviation is as small as possible.

Problem A.4. Find two *different* collections of numbers, each with the same N , the same average, and the same standard deviation. What is the minimum N for which this is possible?

Problem A.5. The *median* of a set of numbers is the number for which half of those in the set lie above it and half lie below it. (a) What is the median of the set of card numbers described in the text and depicted in Figure A.1? (b) Give at least two examples of other sets for which the median is ambiguous or otherwise problematic.

Continuous distributions

Now suppose that instead of a discrete variable i , our cards are labeled with a continuous range of real numbers x . Also our deck contains infinitely many cards, labeled with infinitely many different x values, although certain values (or ranges of values) may still occur more frequently than others. Then n and N are no longer well defined, and the probability $P(x)$ for any *exact* value of x will ordinarily be infinitesimal, but we can still define a probability *density*, $\rho(x)$, so that $\rho(x) dx$ is the probability of obtaining (when you select one card at random) a value between

x and $x + dx$, for infinitesimal dx . More generally, for an arbitrary interval running from $x = a$ to $x = b$, the probability is

$$(\text{Probability of } a < x < b) = \int_a^b \rho(x) dx. \quad (\text{A.9})$$

The probability density is a function *whose purpose in life is to be integrated*. Naturally, $\rho(x)$ must be normalized so that its integral over *all* x equals 1:

$$\int_{-\infty}^{\infty} \rho(x) dx = 1, \quad (\text{A.10})$$

in analogy with equation A.3.

Furthermore, we can calculate the *average* value of x by “adding up” (integrating) all the x values, weighted by their probabilities, in analogy with equation A.4:

$$\langle x \rangle = \int_{-\infty}^{\infty} x \rho(x) dx. \quad (\text{A.11})$$

This is what you would get if you drew many cards from the infinite deck and averaged their numerical values, in the limit of infinitely many trials.²

Similarly, we can calculate the average of the square of x ,

$$\langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 \rho(x) dx, \quad (\text{A.12})$$

and the average of any other function of x ,

$$\langle f(x) \rangle = \int_{-\infty}^{\infty} f(x) \rho(x) dx, \quad (\text{A.13})$$

in analogy with equations A.5 and A.6. And as a measure of the spread of the distribution, we can calculate the standard deviation,

$$\sigma_x = \sqrt{\int (x - \langle x \rangle)^2 \rho(x) dx} = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}, \quad (\text{A.14})$$

in analogy with equations A.7 and A.8.

Problem A.6. Consider a *uniform* probability distribution: $\rho(x) = \rho_0$ (a constant) between $x = -a$ and $x = +a$, and $\rho(x) = 0$ elsewhere. (a) What is ρ_0 in terms of a ? (b) Find the standard deviation of this distribution, in terms of a . (c) If you choose a number according to this probability distribution, what is the chance of it being within one standard deviation of the average? (d) Draw a sketch of this probability distribution, labeling the points one standard deviation to either side of the average.

²We also sometimes call $\langle x \rangle$ the *expectation value* of x , but the term can be misleading because there's no particular reason why you should “expect” to draw a card with this particular value—and for some distributions $\rho(x)$, you might *never* get this particular value.

Problem A.7. Consider a *Gaussian* probability distribution: $\rho(x) = Ae^{-(x/a)^2}$. (a) What is A in terms of a ? (b) Find the standard deviation of this distribution, in terms of a . (c) If you choose a number according to this probability distribution, what is the chance of it being within one standard deviation of the average? (d) Draw a sketch of this probability distribution, labeling the points one standard deviation to either side of the average. (Look up the needed integrals, or use Mathematica. For part (c), do the integral numerically.)

Problem A.8. Sketch a probability distribution $\rho(x)$ for which $\rho = 0$ at $x = \langle x \rangle$.

A.2 The Dirac delta function

When a physical object is localized in a very small space, its density (or probability density or electric charge density) is very large within that space but zero everywhere else. To express this density as a function of position, we *could* use a very narrow, sharply peaked function with some particular shape (perhaps a rectangular function or a Gaussian bell curve, as shown in Figure A.2). But often we don't care about (and perhaps don't know) the actual shape, or even the actual width. Then we can make our lives easier by taking the *limit* in which the density function becomes infinitely narrow but infinitely tall, holding the area under the graph (that is, the total mass, or probability, or electric charge) fixed. This limit of a function is called a *Dirac delta function* (although it's technically not a true function, according to the strict mathematical definition).

In one dimension, we define the Dirac delta function, $\delta(x)$, to put the infinite “spike” at the origin:

$$\delta(x) = \begin{cases} \infty & \text{at } x = 0, \\ 0 & \text{elsewhere.} \end{cases} \quad (\text{A.15})$$

If instead we want the spike at $x = a$, we just shift the argument as we would for any other function of x :

$$\delta(x - a) = \begin{cases} \infty & \text{at } x = a, \\ 0 & \text{elsewhere.} \end{cases} \quad (\text{A.16})$$

Either way, we define the integral of the delta function to be exactly 1:

$$\int_{-\infty}^{\infty} \delta(x) dx = \int_{-\infty}^{\infty} \delta(x - a) dx = 1. \quad (\text{A.17})$$

Here the limits on the integrals don't actually have to be $-\infty$ and ∞ , so long as they include the point where the δ function is infinite.

Manipulating algebraic expressions involving delta functions is usually pretty easy, once you get the hang of it. The most common trick is that wherever you see a delta function multiplying some other function $f(x)$, you can replace this function with its value at whatever point the delta function has its spike:

$$f(x)\delta(x - a) = f(a)\delta(x - a). \quad (\text{A.18})$$

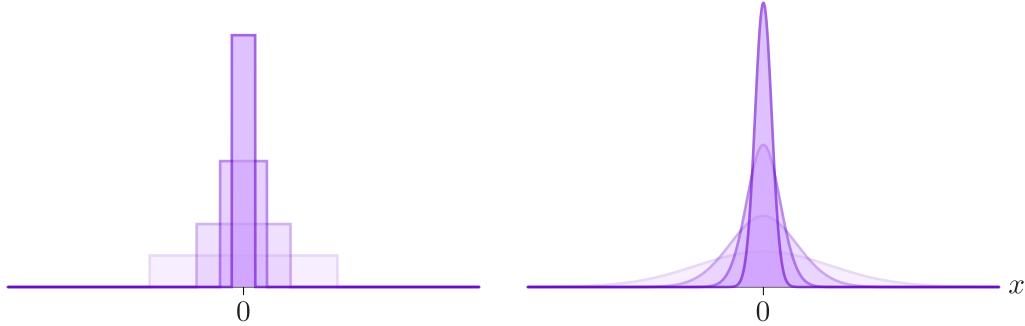


Figure A.2: Sequences of rectangular functions (left) and Gaussian functions (right) that become successively narrower and taller, with the total area under the graph fixed at 1. In the limit of zero width and infinite height, either sequence becomes the Dirac delta function, $\delta(x)$.

This works because the delta function is zero everywhere except $x = a$. Often such expressions occur under integral signs, in which case you can then use equation A.17 to carry out the integral:

$$\int f(x)\delta(x - a) dx = \int f(a)\delta(x - a) dx = f(a), \quad (\text{A.19})$$

provided that the region of integration includes the point a . This result also tells us that if $f(x)$ is itself a delta function with its spike at $x = a$, then the integral is infinite (or as the mathematicians say, “undefined”).

A multidimensional delta function is equivalent to a product of one-dimensional delta functions, one for each dimension. For instance, in three dimensions we can write

$$\delta^{(3)}(\vec{r} - \vec{a}) = \delta(x - a_x)\delta(y - a_y)\delta(z - a_z), \quad (\text{A.20})$$

where the superscript “(3)” simply indicates the dimensionality.

Section 3.1 explains how the delta function comes up in the theory of Fourier transforms, and presents a cool identity (equation 3.6) that relates $\delta(x)$ to the complex function e^{ikx} .

Problem A.9. To better understand the integration formula A.19, let’s approximate the delta function using a rectangular function of the form shown in Figure A.2 (left). Let $\rho_a(x)$ denote such a rectangular function, equal to a constant between $x = -a$ and $x = +a$, with a suitable height ρ_0 . The function $\rho_a(x - b)$ then has the same shape, shifted to the right by b . For definiteness, let’s take the arbitrary function $f(x)$ in equation A.19 to be a power law, x^n . Compute the integral

$$\int_{-\infty}^{\infty} x^n \rho_a(x - b) dx$$

exactly, then take the limit $a \rightarrow 0$ and check that you get the expected result.

Problem A.10. Repeat the previous problem but instead use a Gaussian to approximate the delta function: $\rho_a(x) = Ae^{-(x/a)^2}$ (where A is a normalization constant that you need to determine). Carrying out the integral for an arbitrary power law x^n is awkward, so just do the case $n = 2$. Look up the needed integral(s), or use Mathematica.

A.3 Complex numbers

A *complex number* is an ordered *pair* of real numbers: a sort of two-dimensional number or, if you like, a vector in an abstract two-dimensional space. If c represents a complex number and a and b are its two components, then for many purposes we could write simply

$$c = (a, b), \quad (\text{A.21})$$

listing the components in order as we would do for any other two-dimensional vector. We define the *magnitude* $|c|$ of a complex number (also called its *modulus*, especially by mathematicians) in the same way as for any other vector, using the Pythagorean formula:

$$|c| = \sqrt{a^2 + b^2}, \quad (\text{A.22})$$

where a and b are the two components of c . And as with any other vector, we can add, negate, and subtract complex numbers, component-wise. And we can multiply a complex number by a real number, just as we multiply vectors by real “scalars.”

The difference between complex numbers and ordinary two-dimensional vectors is that for complex numbers we also define a new type of multiplication. For any two complex numbers $c_1 = (a_1, b_1)$ and $c_2 = (a_2, b_2)$ we define the product as

$$c_1 c_2 = (a_1 a_2 - b_1 b_2, a_1 b_2 + b_1 a_2), \quad (\text{A.23})$$

where again I’ve listed the two components in order. I’ll try to motivate this strange formula below. First, however, let me show you how to remember it. Define the symbol i so that $i^2 = -1$. Then write each complex number c in the symbolic format

$$c = a + ib, \quad (\text{A.24})$$

so $c_1 = a_1 + ib_1$ and $c_2 = a_2 + ib_2$. Using this notation, we can work out the product $c_1 c_2$ using the ordinary rules of algebra:

$$\begin{aligned} c_1 c_2 &= (a_1 + ib_1)(a_2 + ib_2) \\ &= a_1 a_2 + ia_1 b_2 + ib_1 a_2 + i^2 b_1 b_2 \\ &= (a_1 a_2 - b_1 b_2) + i(a_1 b_2 + b_1 a_2), \end{aligned} \quad (\text{A.25})$$

where in the last step I substituted -1 for i^2 . Notice that this result is equivalent to what I wrote in equation A.23. Similarly, by writing complex numbers in the form $a + ib$ and then treating i as an ordinary algebraic symbol, we obtain the correct rules for addition, negation, and subtraction of complex numbers. We can

also define the operation of division in the natural way, so that c_1/c_2 is the complex number that, when multiplied by c_2 , gives c_1 .

Because there is no real number whose square is -1 , the symbol i is called the “imaginary” unit. The two components a and b of $c = a + ib$ are called its *real part* and *imaginary part*, respectively. But *please* don’t take these terms too literally! Again: A complex number is simply a two-component vector with a funny rule for multiplication.³ Its so-called “real” and “imaginary” parts are both equally legitimate. If you’re uncomfortable with the idea of a number whose square is -1 (or if you ever need to program a computer to do complex arithmetic using a language that doesn’t already have this feature), just treat a complex number as a pair of real numbers and use equation A.23 for multiplication.

In mathematical expressions, we denote the real and imaginary parts of a complex number c as $\text{Re } c$ and $\text{Im } c$, respectively. We can therefore express any complex number as

$$c = (\text{Re } c) + i(\text{Im } c), \quad (\text{A.26})$$

avoiding the introduction of new letters of the alphabet such as a and b . If $\text{Im } c = 0$ then we say that c is a *pure real number*, while if $\text{Re } c = 0$ then we say that c is a *pure imaginary number*.

We define the *complex conjugate*, denoted c^* , as the complex number whose real part is the same as that of c but whose imaginary part has the opposite sign:

$$c^* = (\text{Re } c) - i(\text{Im } c). \quad (\text{A.27})$$

This trivial-seeming operation is surprisingly useful, largely because multiplying c^* by c gives us the squared magnitude of c :

$$c^*c = [(\text{Re } c) - i(\text{Im } c)][(\text{Re } c) + i(\text{Im } c)] = (\text{Re } c)^2 + (\text{Im } c)^2 = |c|^2. \quad (\text{A.28})$$

When working with complex conjugates, it is useful to know the identities

$$(c_1 + c_2)^* = c_1^* + c_2^*, \quad (\text{A.29})$$

$$(c_1 c_2)^* = c_1^* c_2^*, \quad (\text{A.30})$$

$$(c^*)^* = c. \quad (\text{A.31})$$

More generally, to evaluate the complex conjugate of a complicated expression, you just put a $*$ on every symbol that stands for a complex number, “cancel” any double $*$ ’s, and change i to $-i$ wherever it appears explicitly.

Problem A.11. Prove that multiplication of complex numbers is commutative: $c_1 c_2 = c_2 c_1$.

Problem A.12. Prove equations A.29 through A.31.

³See also <https://www.xkcd.com/2028/>.

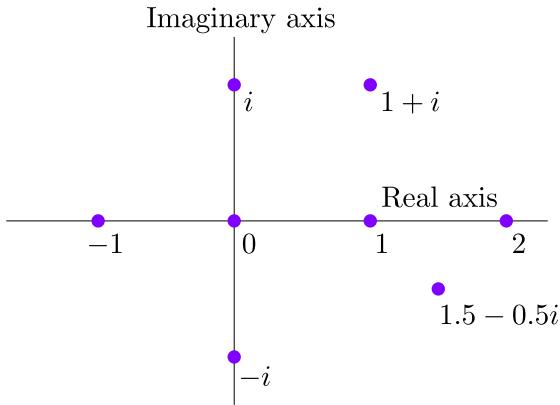


Figure A.3: The complex plane, with some examples of complex numbers plotted as points. (Alternatively, we could plot them as vectors with their tails at the origin.)

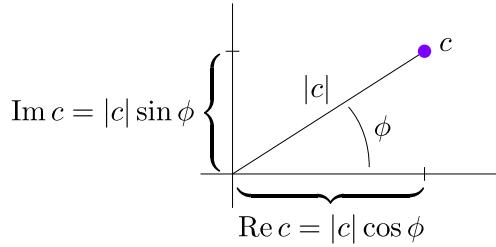


Figure A.4: In polar coordinates, the magnitude (or *modulus*) of a complex number c is denoted $|c|$, and its phase angle (or *argument*) is often denoted ϕ .

The complex plane

To visualize a complex number, we plot its two components in a two-dimensional space, called the *complex plane*. By convention, we plot the real part along the horizontal axis and the imaginary part along the vertical axis. Figure A.3 shows some examples.

Often it is more useful to express the location of a number in the complex plane using polar coordinates, as shown in Figure A.4. The distance of the plotted point from the origin (that is, the length of the two-dimensional “vector”) is the magnitude $|c|$. The angle, measured counter-clockwise from the real axis, is called the *phase* (or among mathematicians, the *argument*) of the number, and often denoted ϕ . We can then write the conversions from polar to rectangular coordinates as

$$\text{Re } c = |c| \cos \phi, \quad \text{Im } c = |c| \sin \phi. \quad (\text{A.32})$$

Using these polar-to-rectangular conversions, we can write any complex number in terms of its polar coordinates:

$$c = |c|(\cos \phi + i \sin \phi). \quad (\text{A.33})$$

This polar form is often more convenient, and more intuitively helpful, than the rectangular form, equation A.26. Moreover, we can write the polar form more compactly using Euler's formula,

$$e^{i\phi} = \cos \phi + i \sin \phi, \quad (\text{A.34})$$

so that equation A.33 becomes simply

$$c = |c|e^{i\phi}. \quad (\text{A.35})$$

If $|c| = 1$, then we sometimes call the number a *pure phase*.

It's best to think of Euler's formula (equation A.34) as the *definition* of e to an imaginary power. But this definition is *useful* because it implies the same algebraic rules that apply to ordinary exponential functions:

$$e^{i(\phi_1+\phi_2)} = e^{i\phi_1}e^{i\phi_2}, \quad (\text{A.36})$$

$$e^{in\phi} = (e^{i\phi})^n. \quad (\text{A.37})$$

Another useful identity is

$$(e^{i\phi})^* = e^{-i\phi}, \quad (\text{A.38})$$

which again says that to take the complex conjugate, you replace i by $-i$ wherever it appears.

Polar coordinates are especially helpful for understanding the strange formula A.23 for multiplication. If we write our two complex numbers as $c_1 = |c_1|e^{i\phi_1}$ and $c_2 = |c_2|e^{i\phi_2}$, then their product is

$$c_1 c_2 = |c_1|e^{i\phi_1}|c_2|e^{i\phi_2} = |c_1||c_2|e^{i(\phi_1+\phi_2)}, \quad (\text{A.39})$$

where I've used equation A.36 in the last step. This result says that in order to multiply two complex numbers, we *multiply* their magnitudes and *add* their phases (see Figure A.5). Geometrically, therefore, multiplication by a complex number c consists of two operations: a *stretch* by a factor of $|c|$, and a *rotation* by the phase angle of c . (If $|c| < 1$, as is often the case in quantum mechanics, then the stretch is actually a shrink.)

Problem A.13. Use the definition A.34 to prove formulas A.36, A.37, and A.38. For the first and third, you'll need to use some trig identities. For the second, assume that n is an integer and use “induction,” first proving that it works for a particular small n and then showing that *if* it works for n then it must also work for $n + 1$.

Problem A.14. Draw pictures to illustrate, in the complex plane, the operations of addition, negation, and subtraction of complex numbers.

Problem A.15. Describe geometrically, in the complex plane, the operation of multiplying an arbitrary complex number by i .

Problem A.16. What is the geometrical interpretation of the operation of *division* of two complex numbers? Illustrate the process with a picture similar to Figure A.5, and describe the process in words.

Problem A.17. Draw a picture in the complex plane that shows how to take the square root of a complex number. What is \sqrt{i} ?

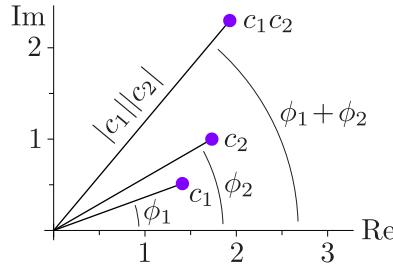


Figure A.5: The geometrical interpretation of the product of two complex numbers. In order to compute the product, we *multiply* the magnitudes of the two numbers, while we *add* their phase angles. In the example shown here, $|c_1| = 1.5$ and $|c_2| = 2.0$.

Complex-valued functions

A complex-valued *function* $\psi(x)$ is just a pair of ordinary functions, one for the real part and one for the imaginary part:

$$\psi(x) = f(x) + ig(x) = \operatorname{Re} \psi(x) + i \operatorname{Im} \psi(x). \quad (\text{A.40})$$

We can combine such functions algebraically just as you would expect, using complex arithmetic. We can also do calculus on complex-valued functions:

$$\frac{d\psi}{dx} = \frac{df}{dx} + i \frac{dg}{dx}, \quad \int \psi(x) dx = \int f(x) dx + i \int g(x) dx. \quad (\text{A.41})$$

All the rules of calculus are what they should be, because derivatives and integrals are ultimately defined in terms of differences and sums, which we've already defined to operate separately on the two components of a complex number.

There are several ways to visualize a complex-valued function. The most straightforward is simply to plot the real and imaginary parts separately. The method I've generally used in this book, however, emphasizes the magnitude and phase of each function value rather than the real and imaginary parts. The trick is to use color hues to represent the phases, as shown in Figure A.6. We associate the three primary colors with three equally spaced directions in the complex plane, conventionally putting red on the positive real axis ($\phi = 0$), green at $\phi = 120^\circ$, and blue at $\phi = 240^\circ$. (This arrangement is also conventional in computer graphics systems.) The secondary colors (yellow, cyan, magenta) are half-way between the primaries, with intermediate blends filling in all the other directions. Cyan then falls on the negative real axis, while the positive imaginary axis corresponds to a yellowish green, and the negative imaginary axis corresponds to purple. For a function $\psi(x)$ of just one variable, we can then plot $|\psi|$ or $|\psi|^2$ on the vertical axis and fill the colors in below the curve to indicate the phases. For a function of two variables we can use the saturation (or alternatively, the brightness) to indicate $|\psi|$ or $|\psi|^2$, so that zero corresponds to white (or black), while the most intense colors correspond to some maximum $|\psi|$ value.

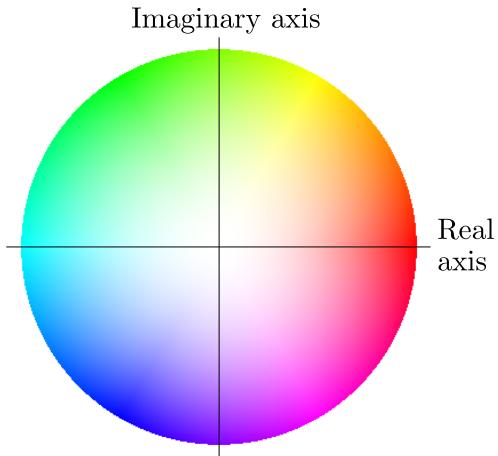


Figure A.6: The conventional assignment of color hues to indicate phase angles in the complex plane. Red is on the positive real axis, yellowish green on the positive imaginary axis, cyan on the negative real axis, and purple on the negative imaginary axis. In this image I have also made the color saturation proportional to the squared magnitude of the complex number, so that zero corresponds to white and full saturation corresponds to some maximum value of the magnitude, at the circle's outer edge.

A.4 Vector spaces

Sometimes abstraction is a good thing, because it lets us learn a set of concepts just once, and then apply them to many specific cases. The abstract framework that applies to every quantum system is that of a *vector space*.

A vector space is a generalization of the system of ordinary three-dimensional vectors. It consists of some collection of objects called *vectors* that we can add, negate, and subtract, together with a collection of numbers called *scalars* that we can multiply the vectors by, in an operation called *scalar multiplication*. All of these operations have to obey the usual algebraic rules (commutative law, distributive law, and so on), and, crucially, a vector space must be *closed* under all of these operations, so that adding any two vectors, or multiplying any vector by any scalar, results in another vector.

The most familiar vectors are simply n -tuples of real numbers, such as (a, b) or (a, b, c) , which we can also represent as arrows with magnitude and direction. If the vectors are two-dimensional we get the simplest nontrivial vector space; three-dimensional vectors live in a different vector space. The generalization to dimensions higher than 3 is straightforward, although the vectors then become hard to draw.

But in quantum mechanics we need to generalize further, to vectors that aren't n -tuples at all. For example, we often use certain types of *functions* as our “vectors,” with the usual rules for adding functions, negating functions, subtracting functions, and scalar-multiplying functions. The functions can be defined on the entire real line, or on just a finite line segment, or on a two-dimensional plane or a higher-

dimensional space. Don't confuse the dimension of the function domain with the dimension of the resulting vector space! Vector spaces of functions (sometimes called *function spaces*) are often *infinite*-dimensional, in a sense that I'll define below.

The other generalization in quantum mechanics is that our scalars will be *complex* numbers rather than reals. This makes even the smallest (two-dimensional) vector space pretty hard to visualize!

In quantum mechanics we also require that our vector space have an *inner product* operation, which combines any two vectors to produce a scalar. If we denote the inner product of vectors α and β as $\langle \alpha | \beta \rangle$, then this operation must have the following properties:

$$\langle \beta | \alpha \rangle = \langle \alpha | \beta \rangle^*, \quad (\text{A.42})$$

$$\langle \alpha | \beta_1 + \beta_2 \rangle = \langle \alpha | \beta_1 \rangle + \langle \alpha | \beta_2 \rangle, \quad (\text{A.43})$$

$$\langle \alpha_1 + \alpha_2 | \beta \rangle = \langle \alpha_1 | \beta \rangle + \langle \alpha_2 | \beta \rangle, \quad (\text{A.44})$$

$$\langle \alpha | c\beta \rangle = c \langle \alpha | \beta \rangle, \quad (\text{A.45})$$

$$\langle c\alpha | \beta \rangle = c^* \langle \alpha | \beta \rangle, \quad (\text{A.46})$$

where in the last two equations c is an arbitrary scalar. Aside from the occasional appearance of a * for complex conjugation, these properties are the same as for the dot product of ordinary three-dimensional vectors. To summarize the last four relations we say that the inner product is *bilinear*.⁴ The * in equation A.42 ensures that the inner product of any vector with itself is a real number; we call the square root of this number the length or *norm* of the vector. The * in equation A.46 is required for consistency with equations A.42 and A.45.

If the norm of a vector is 1, we call it a *unit vector*, or say that it is *normalized*. If the inner product of two vectors is 0, we say they are *orthogonal*. And if we have a collection of vectors $\{\alpha_i\}$ such that any vector in the space can be written as a linear combination of them,

$$\beta = c_1 \alpha_1 + c_2 \alpha_2 + \dots \quad (\text{A.47})$$

(for an arbitrary vector β and some set of scalars $\{c_i\}$), then we say the set $\{\alpha_i\}$ *spans* the vector space, or is *complete*, or forms a *basis*. We will almost always work with basis vectors that are normalized and mutually orthogonal; in that case we call the collection an *orthonormal basis*.

The *dimension* of a vector space is the number of mutually orthogonal vectors needed to form a complete basis. For the vector space of n -tuples of numbers, the dimension is simply n . But for function spaces the dimension is often infinite.

Examples of vector spaces

I'll use the standard symbol \mathbb{C}^n to represent the vector space of n -tuples of complex numbers. So \mathbb{C}^2 is the vector space of all ordered pairs of complex numbers, (c_1, c_2) ;

⁴More precisely, the inner product is linear in its second argument and *antilinear* in its first argument, due to the * in equation A.46.

\mathbb{C}^3 is the space of all ordered triples of complex numbers, (c_1, c_2, c_3) ; and so on.

The inner product of two vectors in \mathbb{C}^n is simply the familiar dot product, except that we complex-conjugate each component of the vector on the *left*. For instance, if α is the vector (c_1, c_2, c_3) and β is the vector (d_1, d_2, d_3) , then

$$\langle \alpha | \beta \rangle = c_1^* d_1 + c_2^* d_2 + c_3^* d_3 = \sum_i c_i^* d_i. \quad (\text{A.48})$$

You can easily check that this definition satisfies all the requirements of equations A.42 through A.46.

The dimension of the space \mathbb{C}^n is simply n , as we can check by constructing the standard orthonormal basis, in which each basis vector has a single component equal to 1 with the rest of the components equal to 0. For instance, the standard basis for \mathbb{C}^3 is

$$(1, 0, 0), \quad (0, 1, 0), \quad (0, 0, 1). \quad (\text{A.49})$$

These vectors are normalized, they are mutually orthogonal, and an arbitrary vector of the form (c_1, c_2, c_3) can be written as a linear combination of them.

You may or may not be surprised to learn that *every* n -dimensional vector space with complex scalars (and with the same finite n) is equivalent to \mathbb{C}^n . By “equivalent,” I mean that we can put all the vectors in one-to-one correspondence, and this correspondence is preserved under addition, scalar multiplication, and so on. (The technical term for this kind of equivalence is *isomorphism*.) So once you understand \mathbb{C}^n , you know everything there is to know about finite-dimensional vector spaces (with complex scalars).

Things get more complicated with vector spaces of infinite dimension.

Our prototype of an infinite-dimensional vector space is the collection of all “square-integrable” complex-valued functions of x , where x is a (real) variable ranging from $-\infty$ to ∞ . The square-integrable condition says that for every function $\alpha(x)$ in our space, the normalization integral

$$\int_{-\infty}^{\infty} |\alpha(x)|^2 dx \quad (\text{A.50})$$

must be finite. (So in quantum mechanics language, this is the space of all *normalizable* wavefunctions for a particle in one dimension, plus the function $\alpha(x) = 0$.) This space is an example of what is called a *Hilbert space* (a particular type of infinite-dimensional vector space), so to denote it I will use the (nonstandard⁵) symbol \mathbb{H}^1 , where the 1 indicates that we’re talking about functions of just one variable (x). The space of square-integrable complex-valued functions of n variables would then be denoted \mathbb{H}^n .

⁵Mathematicians call this space L^2 , where the 2 indicates that we *square* a function to take its norm, and the L indicates that we compute the norm using what’s called Lebesgue integration. But to us physicists that 2 goes without saying, and the difference between a Lebesgue integral and an ordinary (Riemann) integral is of no consequence. What we *do* care about is whether our functions depend on one variable or several. And for whatever reason, the name *Hilbert space* has caught on among physicists.

In analogy to equation A.48, the inner product of two \mathbb{H}^1 functions $\alpha(x)$ and $\beta(x)$ would be

$$\langle \alpha | \beta \rangle = \int_{-\infty}^{\infty} \alpha^*(x) \beta(x) dx. \quad (\text{A.51})$$

The sum becomes an integral, because the discrete index i has become a continuous variable x . Again, it is not hard to show that this definition satisfies all the requirements of equations A.42 through A.46. The restriction to square-integrable functions ensures that all inner products are finite.

One set of basis “functions” for \mathbb{H}^1 would be the Dirac delta functions $\delta(x - a)$, for all possible values of a . These form a complete set, so any \mathbb{H}^1 function can be expressed as a linear combination of them. However, they are not actual functions and are not themselves members of \mathbb{H}^1 . Similarly, the complex exponential functions e^{ikx} , for all real numbers k , form a complete set but are not square-integrable and so are not themselves members of \mathbb{H}^1 . An example of a basis set of actual \mathbb{H}^1 functions would be the harmonic oscillator energy eigenfunctions described in Section 2.4. If we restrict our vector space to functions that are nonzero only within the finite interval $0 < x < a$, then a convenient set of basis functions is the sine waves $\sin(n\pi x/a)$, for positive integers n (see Section 1.4). In any case, we need an infinite number of basis functions to span the vector space, so the dimension of the space is infinite.

Problem A.18. Find three vectors that form an orthonormal basis for \mathbb{C}^3 , without using any of the trivial unit vectors listed in A.49.

Problem A.19. The vector space \mathbb{H}^1 includes all normalizable functions of x . What about the set of all normalized functions of x , that is, all allowed wavefunctions for a quantum particle in one dimension: Is this set a vector space? Why or why not?

Combining vector spaces

Suppose we have *two* vector spaces, not necessarily of the same dimension. We can then combine them to make a larger vector space by building products of the form $\alpha\beta$, where α is a vector from the first space and β is a vector from the second. The new vector space, called a *tensor product space*, consists of all such products *and* all possible linear combinations of these products.

But what kind of “products” am I talking about here? If α and β are functions (of variables with different names, perhaps x and y), then $\alpha\beta$ would just be an ordinary product, which is now a function of all of the variables on which α and β depend. Alternatively, if α is a function while β is an n -tuple of complex numbers (that is, a member of \mathbb{C}^n), or vice-versa, then $\alpha\beta$ would again be a familiar type of product, resulting in an n -tuple of *functions*. But if α is a member of \mathbb{C}^m and β is a member of \mathbb{C}^n , then $\alpha\beta$ is really the *tensor product*, often denoted $\alpha \otimes \beta$, and defined as the list of all possible products of the components of α and β , with one component taken from each. So, for example, if $\alpha = (c_1, c_2)$ and $\beta = (d_1, d_2, d_3)$,

then

$$\alpha \otimes \beta = (c_1, c_2) \otimes (d_1, d_2, d_3) = (c_1 d_1, c_1 d_2, c_1 d_3, c_2 d_1, c_2 d_2, c_2 d_3). \quad (\text{A.52})$$

(The order in which I've listed the components of the tensor product is actually arbitrary, but once you pick an order you need to use it consistently.) Again, the tensor product space consists of *all* such possible products, *and* all possible linear combinations of these products.

As you can see, the dimension of a finite-dimensional tensor product space is equal to the product of the dimensions of the two original spaces. We write

$$\mathbb{C}^m \otimes \mathbb{C}^n = \mathbb{C}^{mn}, \quad (\text{A.53})$$

where now the \otimes symbol has the more abstract meaning of combining the entire spaces (by taking tensor products of their respective vectors).

If either or both of the two original vector spaces is infinite-dimensional, then so is the tensor product space. There's no special symbol for the tensor product of \mathbb{C}^m with \mathbb{H}^n , but a tensor product of two \mathbb{H} spaces gives a larger \mathbb{H} space:

$$\mathbb{H}^m \otimes \mathbb{H}^n = \mathbb{H}^{m+n}. \quad (\text{A.54})$$

So, for example, we can think of \mathbb{H}^2 , the space of all square-integrable functions of two variables, as a tensor product of two \mathbb{H}^1 spaces.

A.5 Matrices

A *matrix* is a rectangular array of numbers, arranged in rows and columns:

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots \\ A_{21} & A_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}. \quad (\text{A.55})$$

The individual numbers A_{ij} are called the *elements* of the matrix. Notice that, by convention, the first subscript indicates the row number while the second subscript indicates the column number. If the number of rows equals the number of columns we say the matrix is *square*. In quantum mechanics all our matrices are either square, or have just a single row or a single column. A matrix with just one row is also called a *row vector*, while a matrix with just one column is also called a *column vector*.

If two matrices have the same dimensions (same numbers of rows and columns), then we can add them, element by element:

$$A + B = \begin{pmatrix} A_{11} + B_{11} & A_{12} + B_{12} & \cdots \\ A_{21} + B_{21} & A_{22} + B_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}. \quad (\text{A.56})$$

Matrix subtraction works in the same way. We can also negate a matrix by negating each of its elements, and multiply a matrix by a scalar (an ordinary number), by multiplying each element by that scalar.

Matrix multiplication

The interesting operation on matrices is *matrix multiplication*. We define the product AB to be the matrix whose ij element is constructed from the i th row of A and the j th column of B , according to the formula

$$\begin{aligned}(AB)_{ij} &= A_{i1}B_{1j} + A_{i2}B_{2j} + \dots \\ &= \sum_k A_{ik}B_{kj},\end{aligned}\tag{A.57}$$

where the sum runs over all the columns of A and all the rows of B . Naturally this works only if the number of columns in A equals the number of rows in B ; if that's not the case then the product is undefined. Here is an example in which I've highlighted the row of A and the column of B that contribute to the element of their product in row 3 and column 2:

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{pmatrix}.\tag{A.58}$$

To carry out this portion of the multiplication, I imagine picking up the second column of B , rotating it 90 degrees counterclockwise, and laying it along the third row of A , in order to place the elements that get multiplied next to each other. Then just insert + signs between the pairs:

$$\begin{array}{ccc} A_{31} & A_{32} & A_{33} \\ B_{12} & B_{22} & B_{32} \end{array} \rightarrow A_{31}B_{12} + A_{32}B_{22} + A_{33}B_{32} = C_{32}.\tag{A.59}$$

Matrix multiplication is associative (that is, $(AB)C = A(BC)$), and it distributes over matrix addition (for example, $A(B + C) = AB + AC$). But matrix multiplication is *not* commutative: $AB \neq BA$, except in some special cases, when we say that A and B *commute*.

The most common type of matrix multiplication in quantum mechanics is when we multiply a square matrix by a column vector, to obtain another column vector. In three dimensions the explicit formula would be

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} A_{11}c_1 + A_{12}c_2 + A_{13}c_3 \\ A_{21}c_1 + A_{22}c_2 + A_{23}c_3 \\ A_{31}c_1 + A_{32}c_2 + A_{33}c_3 \end{pmatrix}.\tag{A.60}$$

This operation is also called a *linear transformation* on the vector (c_1, c_2, c_3) , because each of the resulting vector components is a linear function of the three components c_1 , c_2 , and c_3 . In general the output vector will differ from the input vector in both length and direction, so we can think of the transformation as a combination of a stretch (or shrink) and a rotation.⁶

⁶For a marvelously visual explanation of matrices as linear transformations, watch the video “Linear Transformations and Matrices” by Grant Sanderson, <https://www.youtube.com/watch?v=kYB8IZa5AuE>, which is part of the Essence of Linear Algebra series on the 3Blue1Brown YouTube channel.

Problem A.20. Consider the following matrices:

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

- (a) Compute the matrix products AB , BA , AC , CA , BC , and CB .
- (b) Which of these pairs, if any, commute?
- (c) Check the associative law, $(AB)C = A(BC)$.
- (d) Check the distributive law, $A(B + C) = AB + AC$.

Problem A.21. A square matrix A is *diagonal* if $A_{ij} = 0$ for all $i \neq j$. Prove that any two diagonal matrices (of the same dimension) commute.

Problem A.22. Solve the linear system

$$\begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 6 \\ 8 \end{pmatrix}$$

for x and y .

Problem A.23. Consider the following matrices:

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix}, \quad D = \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix}.$$

Interpret each of these matrices as a linear transformation, by determining how it transforms various vectors. Start with the unit vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, but be sure to check some others as well. Describe or draw the effects of each transformation.

Special types of matrices

The *transpose* of a matrix, denoted by a superscript T , is what you get when you interchange the matrix's rows and columns:

$$(A^T)_{ij} = A_{ji}. \quad (\text{A.61})$$

If a (square) matrix is its own transpose we say it is *symmetric*, while if it is minus its transpose we say it is *antisymmetric*. For matrices with complex elements we are usually more interested in the *adjoint*, denoted by a \dagger ("dagger") and defined as the complex conjugate of the transpose:

$$(A^\dagger)_{ij} = A_{ji}^*. \quad (\text{A.62})$$

If a (square) matrix is its own adjoint we say it is *Hermitian*; in that case its real part is symmetric while its imaginary part is antisymmetric. (Sometimes the adjoint is called the *Hermitian conjugate*.)

Note that transposing a column vector turns it into a row vector (and vice versa). Moreover, the adjoint of a column vector,

$$\begin{pmatrix} c_1 \\ c_2 \\ \vdots \end{pmatrix}^\dagger = (c_1^* \ c_2^* \ \cdots), \quad (\text{A.63})$$

is just what we need to take its inner product with a second column vector via matrix multiplication, according to equation A.48:

$$\begin{pmatrix} c_1 \\ c_2 \\ \vdots \end{pmatrix}^\dagger \begin{pmatrix} d_1 \\ d_2 \\ \vdots \end{pmatrix} = (c_1^* & c_2^* & \cdots) \begin{pmatrix} d_1 \\ d_2 \\ \vdots \end{pmatrix} = \sum_i c_i^* d_i. \quad (\text{A.64})$$

(Note that a $1 \times n$ matrix multiplied by an $n \times 1$ matrix gives a 1×1 matrix, that is, a single number.)

An *identity matrix* is a square matrix with a 1 in each space along the main diagonal, and a 0 everywhere else:

$$\text{Identity matrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (\text{A.65})$$

Multiplying by such a matrix leaves any other matrix unchanged. The *inverse* of a matrix, denoted by a superscript -1 , is the matrix that we can multiply by to obtain the identity matrix:

$$AA^{-1} = A^{-1}A = \mathbb{1}, \quad (\text{A.66})$$

where $\mathbb{1}$ denotes the identity matrix. If the inverse of a matrix is the same as its transpose, then we say the matrix is *orthogonal*. For matrices with complex elements we are usually more interested in the case where the inverse is equal to the adjoint; we then say the matrix is *unitary*.

Here is a summary of the four related types of special matrices that I've just defined:

$$\begin{aligned} \text{Symmetric: } & A^T = A \\ \text{Hermitian: } & A^\dagger = A \\ \text{Orthogonal: } & A^T = A^{-1} \\ \text{Unitary: } & A^\dagger = A^{-1} \end{aligned} \quad (\text{A.67})$$

Hermitian is the natural complex generalization of symmetric, and unitary is the natural complex generalization of orthogonal.

Problem A.24. Compute the matrix products

$$(2 \ 3) \begin{pmatrix} 4 \\ 5 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 \\ 3 \end{pmatrix} (4 \ 5).$$

Problem A.25. Consider the following matrices:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad C = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Which of these matrices is symmetric? Antisymmetric? Hermitian? Orthogonal? Unitary?

Problem A.26. Prove that the adjoint of the product of two matrices is the product of the adjoints, *in reverse order*: $(AB)^\dagger = B^\dagger A^\dagger$.

Eigenvectors and eigenvalues

When a square matrix A acts on a nonzero column vector α via matrix multiplication, the result is another column vector that can differ from α in both magnitude and direction. In some special cases, however, the direction might not change—only the magnitude. We can then write

$$A\alpha = k\alpha \quad (\text{A.68})$$

for some numerical constant k , and we say that α is an *eigenvector* of A with *eigenvalue* k . (The prefix *eigen* comes from German, and could be translated as “characteristic” or “particular.”) If k is a real number greater than 1, then the action of A stretches α by a factor of k ; if k is a real number less than 1, then instead of a stretch it’s a shrink. We also allow k to be negative, in which case A flips the eigenvector’s direction (in addition to a possible stretch or shrink).

In linear algebra courses they teach you how to find the eigenvalues of a given matrix A by solving what’s called its *characteristic equation*. The idea is to slip an identity matrix (“ $\mathbb{1}$ ”) into the right-hand side of equation A.68, move this term to the left-hand side, and factor out the α to obtain

$$(A - k \cdot \mathbb{1})\alpha = 0, \quad (\text{A.69})$$

where 0 is technically the zero *vector*. They also teach you that equations of this form have solutions if and only if the *determinant* of the matrix acting on α is zero:

$$\det(A - k \cdot \mathbb{1}) = 0. \quad (\text{A.70})$$

This is the “characteristic equation” that I promised. When written out for the case where A is an $n \times n$ matrix, it becomes an n th-order polynomial equation, which is usually hard to solve for $n > 2$. In the 2×2 case the explicit characteristic equation is

$$(A_{11} - k)(A_{22} - k) - A_{12}A_{21} = 0, \quad (\text{A.71})$$

which you can solve for k using the quadratic formula.

A quadratic equation generically has two solutions, so a 2×2 matrix generically has two eigenvalues. Often, though, the solutions are complex, while in some special cases the two solutions can be *degenerate*, that is, identical to each other.

For larger matrices, solving the characteristic equation is usually impractical. In quantum mechanics, though, many of the matrices we work with are sufficiently simple that we can find eigenvalues through physical arguments or through a bit of trial-and-error. For more complicated matrices, there are iterative algorithms for finding eigenvalues numerically, and today’s computers can carry out this process for a remarkably large matrix in a mere fraction of a second. Generically, an $n \times n$ matrix has n eigenvalues, although again there are special cases in which two or more eigenvalues can be degenerate.

Regardless of the dimension, if a matrix is Hermitian then it's not hard to prove that its eigenvalues must all be real numbers. In quantum mechanics we are almost never interested in the eigenvalues of non-Hermitian matrices.

Once we know the eigenvalues of a matrix, it's straightforward (though sometimes tedious) to use equation A.68 to find the eigenvectors. (The iterative algorithms mentioned above find the eigenvalues and eigenvectors simultaneously, so in that case no further step is even necessary.) Notice that if α is an eigenvector with eigenvalue k , then so is any constant times α , so every eigenvector really defines what we might call an *eigendirection*. For a Hermitian matrix, these directions are all mutually orthogonal to each other, so we can use the normalized eigenvectors as an orthonormal basis. (When there are m degenerate eigenvalues we get an m -dimensional space of eigenvectors, from which we can always choose basis vectors that are mutually orthogonal.)

Problem A.27. Show that α and β are eigenvectors of A , where

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \alpha = \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \end{pmatrix}, \quad \beta = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

What are the corresponding eigenvalues? Because A is Hermitian, its eigenvectors must be orthogonal (if, as in this case, their eigenvalues are distinct). Check that α is orthogonal to β , then find a vector that is orthogonal to both α and β , check that it too is an eigenvector of A , and find the corresponding eigenvalue.

Problem A.28. Find the eigenvalues of each of the matrices in Problem A.20, by solving the characteristic equation in each case. Comment briefly on the results.

Problem A.29. Find the eigenvectors and eigenvalues of each of the matrices in Problem A.23, by inspection or trial-and-error if possible (without using the characteristic equation unless you have to). Explain how the results relate to your descriptions of the transformations in that problem.

Problem A.30. What can you say in general about the eigenvalues and eigenvectors of a diagonal matrix?

Appendix B

Mathematica tutorial

This appendix uses a tutorial format to introduce the features of Wolfram Mathematica that are used in this book. If you have never used Mathematica before then I suggest working through the entire tutorial as early as possible. If you are already somewhat proficient with Mathematica then you may be able to skip over some sections, or perhaps use this appendix as a reference for looking up features of the language as you need them. Either way, you should read this appendix with your fingers on the keyboard, typing in every displayed instruction as you come to it. Because I haven't reproduced any of Mathematica's output, you'll get very little out of merely reading without typing. If you're using an electronic version of this book, please resist the temptation to electronically copy and paste the code fragments; typing every character with your own fingers will provide the practice you need to make this powerful tool your own.

B.1 Basic plots and basic syntax

There's no better starting point than a basic plot, of a real-valued function of a single variable. Here's an example, for a function that comes up in Section 2.4:

```
Plot[(4x^4 - 12x^2 + 3)Exp[-x^2/2], {x, -5, 5}]
```

Type this line into Mathematica now, then hit shift-enter to execute it. The plot should appear in a moment.

This single line of code illustrates several of Mathematica's most important features:

- Mathematica code is built out of *functions*, such as `Plot` and `Exp`. Function names are case-sensitive, and all of Mathematica's built-in function names begin with capital letters.
- Function arguments are enclosed in square brackets and separated by commas. Here the `Plot` function has two arguments (both compound structures), while the `Exp` function has just one.

- The first argument of `Plot` is a mathematical expression, built out of the `Exp` function and the operators `+`, `-`, `/`, and `^`. Multiplication requires no operator (just as in standard math notation), although you can always use `*` for clarity. (In a product of two variables such as `a` and `b` you need to write `a*b` or `a b`, so Mathematica won't think you're referring to a single variable named `ab`.)
- In compound math expressions, exponentiation takes precedence over multiplication and division, which in turn take precedence over addition and subtraction. So `-x^2/2` means $-(x^2)/2$, not $-x^{2/2}$. To override the default grouping you can use parentheses. The spaces that I've inserted around the `+` and `-` are merely for readability.
- The second argument of `Plot` is a *list*, delimited by curly braces, with the three elements of the list separated by commas. Here the list specifies the independent variable of the function to plot, followed by the starting and ending values of this variable. I'll say more about lists in Section B.2.

Assigning names

For all but the simplest tasks, you'll want to break up your Mathematica code into multiple statements, to be executed in sequence. The following three lines produce the same output as the single line above, but provide more flexibility:

```
psi = (4x^4 - 12x^2 + 3)Exp[-x^2/2];
xMax = 5;
Plot[psi, {x, -xMax, xMax}]
```

Here I've assigned the names `psi` and `xMax` to the function I want to plot and the maximum x value, so I can more easily modify or reuse these quantities. The semicolons serve as statement separators, and also suppress any displayed output of the lines that precede them. Go ahead and test this code now, putting all three lines into a single Mathematica “cell” (as indicated by the bracket at the window's right margin). Again hit shift-enter to execute the code. Then try modifying `xMax`, and `psi` too if you wish, to see how the plot changes.

Because all of Mathematica's built-in names (such as `Plot` and `Exp`) begin with capital letters, it's a good habit to begin all your own names (such as `psi` and `xMax`) with lower-case letters. Then you can easily tell at a glance which names are Mathematica's and which are yours. You'll also avoid inadvertent name conflicts, without having to learn all of Mathematica's built-in names (about 5000 of them, as of this writing).

Plot options

If you set `xMax` too high in the previous example, you'll notice that Mathematica expands the vertical scale of the plot and clips off the peaks of the function's five “bumps.” This is because Mathematica thinks you want to see more detail in the

small function values that occur at larger values of $|x|$. To override this often-unwanted behavior, you can use the `PlotRange` option:

```
xMax = 12;
Plot[psi, {x, -xMax, xMax}, PlotRange->All]
```

Try this modification now, then try providing an explicit range for the plot, using a two-element list:

```
PlotRange->{-5, 5}
```

Here are a couple of other useful options for modifying the appearance of a `Plot`:

```
Plot[psi, {x, -xMax, xMax}, PlotStyle->{Red, Thick},
      AspectRatio->0.4]
```

You can specify options in any order, always separated by commas. For a complete list of options to use with `Plot`, use the menus to find the Mathematica documentation (“Help → Wolfram Documentation” in my version), then use the search feature to find the `Plot` function and, on its documentation page, scroll down to the “Options” section.

Instead of the color `Red`, you can (of course) specify other pre-defined colors, or define your own colors using either `Hue` or `RGBColor`:

```
Hue[0.75, 1, 1]
RGBColor[0.5, 0, 1]
```

Try each of these, varying the numerical values to see how the color changes. All of the arguments of `Hue` and `RGBColor` should be in the range from 0 to 1, although for the first argument of `Hue` only the difference from the next-lowest integer is used, so the hues cycle repeatedly as this argument increases or decreases. The second and third arguments of `Hue` are saturation and brightness, and if you omit them you’ll get the default of 1 for each.

More functions to plot

Besides `Exp`, Mathematica knows every named math function you’re likely to have heard of: `Sin`, `Cos`, `Tan`, `ArcSin`, `ArcCos`, `ArcTan`, `Sinh`, `Cosh`, `Tanh`, `Sqrt`, `Abs`, `Log`, and more. The trig functions take angles in radians, and the logarithm function defaults to base e , though you can specify any base you want as an additional argument (see the documentation for examples). Mathematica also knows the constants `Pi` and `E` (note the capitalization!). And Mathematica knows all sorts of special functions that come up in advanced mathematics, including Bessel functions, the gamma function, and the orthogonal polynomials described in Sections 2.4, 6.3, and 6.5.

But I’m really getting ahead of myself here. There’s no need to learn about all these functions now, as long as you know how to do simple stuff like plotting a basic sine function:

```
Plot[Sin[10x], {x, 0, Pi}]
```

Another need that arises fairly often is to define (and plot) a function using different formulas for different ranges of the independent variable. I usually do this with Mathematica's `If` function:

```
Plot[If[x<0, 0, x^2], {x, -1, 1}]
```

Notice that `If` returns its second argument if the condition in the first argument is true; otherwise it returns its third argument.

If the function you're plotting is discontinuous, you can use the `Exclusions->None` option to connect the segments:

```
Plot[If[x>-1 && x<1, 0, 1], {x, -2, 2}, Exclusions->None]
```

(Try this code both with and without the `Exclusions` option.) This example also shows how to build a compound condition using `&&` for “and”; the symbol for “or” is `|`. In still more complicated situations you may need to embed one `If` function within another, or look up the more versatile `Piecewise` function.

Function definitions and variable replacements

When you assigned a mathematical expression to the name `psi` a little while ago, you were telling Mathematica to replace “`psi`” with that expression wherever this name subsequently appears. For more flexibility, though, you can instead define your own function. Try this example:

```
gaussian[x_] := Exp[-x^2]
```

Here there are two differences from simply saying `gaussian = Exp[-x^2]`. The first is the replacement of `=` with `:=`, which *delays* Mathematica's evaluation of the right-hand side until later, when the `gaussian` function is actually used. The second difference is appending `[x_]` to the name `gaussian`, indicating that you're actually defining a function that takes a single argument, and that on the right-hand side this argument has the formal name `x`. When you later use the function you can insert any expression you like for this argument, and Mathematica will insert your expression in place of `x`. Try each of these uses, one at a time:

```
gaussian[0.5]
```

```
gaussian[y]
```

```
gaussian[(a-b)/c]
```

```
Plot[gaussian[x-5] + gaussian[x+5], {x, -10, 10}]
```

It can sometimes be hard to decide whether to define a function in this way, or to simply assign a name as in the definition of `psi` above. I usually stick with a simple name assignment for an expression that I won't be needing to modify in the ways illustrated here (e.g., replacing `x` with `y` or something more complicated). And if you *do* want to make such a replacement in a previously named expression, you can do it like this:

```
psi /. x -> 3z      (* replace x with 3z wherever it appears *)
```

I usually pronounce the `/.` operator as “at the point.” Here I’ve also shown how to insert brief comments into Mathematica code (but for more lengthy explanatory comments, see below).

Working with notebooks

By now you’ve executed quite a few Mathematica instructions, many of which are minor modifications of earlier instructions. You may have typed some of these modifications in new cells, below your previous work, or you may have simply edited your previous code and re-executed it in the same cell.

You may be interested to know that as far as Mathematica is concerned, it makes *no difference* where you put each new instruction, or whether you delete earlier instructions. You can put new instructions into existing cells, or between them, or below them, or even in a totally separate window. All name assignments will continue to exist, in Mathematica’s internal state, until you either quit Mathematica, or quit the so-called Kernel (which you can do using a menu command), or explicitly erase them using the `Clear` function. By the same token, when you save your work and reopen it during a new Mathematica session, you’ll need to re-execute any cells containing name assignments that you wish to use in your subsequent work.

A Mathematica document, displayed in a single window, is called a *notebook*, and will be saved with the file extension `.nb`. Each notebook’s contents are divided into cells, as indicated by the brackets at the window’s right margin. Code execution takes place one cell at a time, so it’s a good idea to put a group of closely related instructions into a single cell when they should always be executed together.

You can also create non-executable notebook cells that contain explanatory text and section headings. Go to `Format → Style` (in the menus) to denote the current cell as one of these non-executable types. Mathematica will automatically group cells together, within sections and so on, indicating the group structure with multiple levels of right-margin brackets. If you wish to delete an entire cell or group of cells, click on its bracket to select it and then choose `Clear` from the `Edit` menu. You can also “collapse” a cell group, hiding most of its content, by double-clicking on its bracket. Double-click again to un-collapse the group.

Problem B.1. Plot the function $(1 - \frac{2}{3}r + \frac{2}{27}r^2)e^{-r/3}$, taken from Table 6.2. The variable r is always positive, and you should show all values where the function isn’t negligible. Use a non-default color of your choice.

Problem B.2. Plot the function $(5 \cos^2 \theta - 1) \sin \theta$, taken from Table 6.1. The variable θ ranges from 0 to π . There's no θ key on your keyboard, so either use a variable name like `t` or `theta`, or look up how to input Greek letters in the Mathematica documentation. Use a non-default color and non-default aspect ratio of your choice.

Problem B.3. Use Mathematica to plot the uppermost curve in Figure 3.11, showing both the constant portion on the left and the curved portion on the right, given by equation 3.36. Use the variable name `v0` for the constant V_0 , equal to 0.025. Be sure that the two portions of the graph are connected in your plot. For the square root function you can either use `Sqrt` or just raise to the power 0.5.

Problem B.4. Insert a title cell and several section heading cells into the Mathematica notebook that you've created while working through this tutorial. Use appropriate Style settings, and practice collapsing and expanding the sections by double-clicking on the brackets at the right of the window.

B.2 Lists, tables, and animation

To show multiple graphs at the same time, you can plot a list of functions:

```
Plot[{Sin[x], Cos[x]}, {x, 0, 8Pi}]
```

Mathematica colors the multiple curves differently by default. To specify your own colors, provide a list to the `PlotStyle` option:

```
Plot[{Sin[x], Cos[x]}, {x, 0, 8Pi},
      PlotStyle -> {{Red, Thick}, {Blue, Thick}}]
```

Here I've actually used a list of lists, to specify the thicknesses of the curves as well. Try this code as written, then try it again with simply `PlotStyle -> {Red, Thick}`.

Lists have all sorts of uses besides plotting. You can construct a list using curly braces, and its elements can be numbers, names, functions, other lists, or any expressions at all. To extract or modify a particular element of a list, you put that element's number ("index") in double square-brackets:

```
elements = {hydrogen, helium, lithium, berilium, boron, carbon};
elements[[2]]

elements[[4]] = beryllium; elements
```

Note that the index numbering starts with 1, not 0 as in some computer languages. But try typing "`elements[[0]]`" and see what you get!

Tables

Often you'll want to work with a list of expressions that's built from some pattern. Then, instead of explicitly typing each expression inside curly braces, you can build the list with Mathematica's `Table` function:

```
Table[ $\text{Sin}[n \cdot \pi \cdot x]$ , {n, 1, 9}]
```

Note the syntactical similarity to `Plot`, with a list that specifies the variable to vary and its beginning and ending values. With `Table` this variable changes in increments of 1 by default, but you can specify any other step size as a fourth element of the list. Here is an example that plots just five of these sine functions, with vertical displacements:

```
Plot[Table[ $\text{Sin}[n \cdot \pi \cdot x] + 1.5(n-1)$ , {n, 1, 9, 2}], {x, 0, 1},  
AspectRatio->1]
```

Another option for multiple graphics is to use `Table` to produce a list of plots:

```
Table[Plot[x^n Exp[-x], {x, 0, 10}], {n, 0, 3}]
```

(To eliminate the curly braces and commas in the output, you can look up how to use the `GraphicsRow`, `GraphicsColumn`, and `GraphicsGrid` functions.)

Plotting lists of values

To plot individual points rather than curves, use the `ListPlot` function:

```
data = Table[{x, Exp[-x^2]}, {x, -3, 3, 0.1}];  
ListPlot[data]
```

This example is necessarily contrived; ordinarily you would use `ListPlot` only when you don't have a simple formula that you could just as well feed to `Plot`. You can use the `PlotStyle` option to change the color of the plotted dots, and the `PlotMarkers` option to get other shapes besides dots, but taking full control over the marker appearance is surprisingly difficult in Mathematica.

More often in theoretical work we simply want to connect the dots to obtain a continuous curve:

```
ListPlot[data, Joined->True]
```

Alternatively, from a collection of data you can create an interpolating function, which you can then evaluate and plot (within the range spanned by the data) just as any other function:

```
interp = Interpolation[data];  
Plot[interp[x], {x, -3, 3}]
```

Manipulating and animating a plot

Another way to show how a plot (or any other Mathematica output) depends on a parameter is to use the `Manipulate` function:

```
Manipulate[Plot[ $\text{Sin}[n \cdot \pi \cdot x]$ , {x, 0, 1}], {n, 1, 100, 1}]
```

Note again the syntactic similarity to `Plot` and `Table`, with a list that specifies the variable (here `n`), its beginning and ending values, and the step size (which you can omit for a parameter that varies continuously). As you can see, the output of `Manipulate` includes a slider that you can drag to vary the parameter interactively, along with a tiny “+” button that you can click to reveal animation controls and a text box for fine adjustments. Here is an example that’s nice to animate:

```
Manipulate[Plot[Sin[x]Cos[t] + Sin[2x]Cos[3t], {x, 0, Pi},
    PlotRange->{-2, 2}], {t, 0, 2Pi}]
```

Notice the use of a fixed `PlotRange` to prevent Mathematica from rescaling the graph as `t` changes.

Problem B.5. The New Zealand \$100 banknote famously shows a portrait of Ernest Rutherford along with a plot of the superimposed functions e^{-t} and $1-e^{-t}$, representing the relative amounts of a radioactive substance and its decay product over time. Use Mathematica to produce a similar plot of these functions.

Problem B.6. Use `Table` to make a list of the squares of the first 100 positive integers. Then tell Mathematica to show just element 42 from this list, on a line by itself.

Problem B.7. The energy levels of a hydrogen atom are $-(13.6 \text{ eV})/n^2$, where n is a positive integer. Use `Table` to make a list of the first ten energy levels (as numbers in eV), then plot an energy level diagram, showing each level as a horizontal line. The name and range of the dummy variable on the horizontal axis are arbitrary (I suggest simply `{x, 0, 1}`). Use the option `Ticks->{None, Automatic}` to suppress the labeling of the horizontal axis. Adjust the aspect ratio to make the diagram taller than it is wide. By default, Mathematica will put each line in a different color; see if you can figure out how to use `PlotStyle` to make them all the same color.

Problem B.8. Use Mathematica to reproduce each of the plots in Figure A.2, showing a sequence of rectangular functions and a sequence of Gaussian bell curves. (Don’t worry about the shading or transparency, unless you have time to dig into the documentation to figure out how to do those things.) Then use `Manipulate` to create a plot of each shape with a slider to control the height and width.

B.3 Complex numbers, functions, and plots

To create complex-valued numbers and expressions in Mathematica you simply use a capital `I` for the imaginary unit. Arithmetic then works just as it should:

$$c = (1 + I)(1 + 2I)$$

The functions `Re` and `Im` extract the real and imaginary parts of an expression, while `Abs` gives the modulus (magnitude) of a complex number and `Arg` gives its phase angle in radians:

$$\{\text{Abs}[c], \text{Arg}[c]\}$$

To take the complex conjugate you can type `c*` or `Conjugate[c]`.

Moreover, Mathematica knows all about the complex exponential function $e^{i\theta}$. Here, for example, is a Gaussian wavepacket (see Section 3.2):

```
psi1 = Exp[-x^2] Exp[I*5*x];
Plot[{Re[psi1], Im[psi1]}, {x, -3, 3}, PlotRange->All]
```

Plotting phase as color

The preceding example plots the real and imaginary parts of the function separately, in different colors. Usually, though, the real and imaginary parts are of less concern to us than the magnitude and phase. We therefore plot the magnitude (or squared magnitude) of a complex function on the vertical axis, and fill the area beneath with color hues to indicate the phases. Here is the code to do that:

```
Plot[Abs[psi1], {x, -3, 3},
  PlotPoints -> 300,
  Filling -> Axis,
  ColorFunction -> Function[x, Hue[Arg[psi1]/(2Pi)]],
  ColorFunctionScaling -> False]
```

The critical option here is the `ColorFunction`, which uses the `Hue` function to generate a color from the phase angle (`Arg[psi1]`), scaled by 2π to give a value between 0 and 1. This convention for hue assignments is shown in Figure A.6. The `Function` function is one of Mathematica's trickiest features: it creates an “anonymous” function that associates the independent variable `x` (same that I used in defining `psi1`) with the desired hue. The `ColorFunctionScaling` option turns off the automatic remapping of colors to span the function’s full range (see the following section for examples of when this is desirable), while the `Filling` option extends the coloring down to the x axis. The `PlotPoints` option ensures that Mathematica will evaluate the function at enough points to show the full detail of the color sequence.

Problem B.9. Make a phase-as-color plot of the complex function $y = (x + i)^{-2}$, over as wide a range of x values as seems interesting. Briefly explain why both the magnitudes and the colors have the values they do at a few points.

B.4 Plotting functions of two variables

Mathematica provides three different ways to plot a real-valued function of two variables. With a little work we can also plot a complex-valued function of two variables, using colors to show the phases.

Density plots

A density plot shows the two independent variables along the horizontal and vertical axes, while using a color scheme to indicate the function value at each point. The syntax is the same as for `Plot`, but with a second list specifying the second variable and its range:

```
DensityPlot[(x-y)^2 Exp[-(x^2 + y^2)], {x, -2, 2}, {y, -2, 2}]
```

As with `Plot`, you'll often want to use some names for expressions, and add some non-default options:

```
psi2 = (x-y)^2 Exp[-(x^2 + y^2)];
xMax = 3;
DensityPlot[psi2, {x, -xMax, xMax}, {y, -xMax, xMax},
PlotRange->All, PlotPoints->50, PlotLegends->Automatic]
```

Here the `PlotRange` option is needed to prevent Mathematica from excluding the largest function values (showing them as ugly white patches); `PlotPoints` determines the sampling resolution of the plot in each dimension (with a default value of only about 15, which you'll usually want to increase); and `PlotLegends->Automatic` adds a legend showing how the colors correspond to numerical function values.

But the most interesting option for density plots is `ColorFunction`, which determines the color scheme. The easiest way to override the default color scheme is to use one of Mathematica's 50 or so named color schemes, such as:

```
ColorFunction -> "SunsetColors"
```

Look up “color schemes” in the documentation for a table of these named schemes.

Alternatively, you can create arbitrary color gradients using the `Blend` function. The syntax for a basic white-to-black gradient would be:

```
ColorFunction -> Function[f, Blend[{White, Black}, f]]
```

Here `f` is merely a dummy variable, whose name is arbitrary, representing the plotted function's value.¹ The list of colors provided to `Blend` can include any of Mathematica's named colors, or arbitrary colors defined using `RGBColor` or `Hue`:

```
ColorFunction -> Function[f,
Blend[{Hue[.7, 1, .5], Hue[.9, .1, 1]}, f]]
```

Moreover, the list of colors can be of any length:

```
ColorFunction -> Function[f,
Blend[{Black, Blue, Cyan, White}, f]]
```

¹The `ColorFunction` for a `DensityPlot` is given the *value* of the function being plotted, unlike that for an ordinary `Plot`, which is given the independent variable. To see what variables get passed to the `ColorFunction` in different contexts, look it up in the documentation.

The possibilities are effectively endless.

But just because you *can* color your density plots in arbitrary ways doesn't mean you *should*. With the power to create arbitrary color schemes comes the responsibility to use a scheme that will highlight, rather than obscure, what the plot is actually trying to show. For a function whose values are always positive, it's usually best to use a scheme that changes monotonically either from dark colors to light ones, or from light to dark, but not both. By default, the full range of colors will be mapped to the full range of function values that occur within the plot. For a function whose values can be both positive and negative, it's best to map zero to either white or black (or some other reasonably neutral color), blending into different hues for positive and negative values:

```
DensityPlot[Sin[2*Pi*x] Sin[3*Pi*y], {x, 0, 1}, {y, 0, 1},
  PlotPoints -> 50,
  ColorFunction -> Function[f, Blend[{Cyan, Black, Red}, f]]]
```

Note that a three-color blend will map the middle color to zero only when the plotted function's minimum value is minus its maximum value. When the function's minimum and maximum values are not symmetric about zero, you can shift and scale its values in the `Blend` function to map zero to 0.5 and the most extreme value to either zero or 1, then set `ColorFunctionScaling` to `False`, which maps the full color range to the interval from 0 to 1:

```
psi3 = Sin[Pi*x] Sin[Pi*y] + Sin[2*Pi*x] Sin[2*Pi*y];
DensityPlot[psi3, {x, 0, 1}, {y, 0, 1},
  PlotPoints -> 50,
  ColorFunction ->
  Function[f, Blend[{Cyan, Black, Red}, f*0.32 + 0.5]],
  ColorFunctionScaling -> False]
```

Alternatively, in cases like this you might find it easier to use the method described below for complex-valued functions.

Contour plots

In a contour plot, the values of the two independent variables are again plotted along the horizontal and vertical axes, but now the function values are indicated by contour lines, each of which connects points of a particular value:

```
ContourPlot[(x-y)^2 Exp[-(x^2 + y^2)], {x, -2, 2}, {y, -2, 2}]
```

The syntax is the same as with `DensityPlot`, so you can switch between one and the other with just a few keystrokes. Take a moment now to convert some more of the examples in the previous subsection to contour plots.

One advantage of `ContourPlot` is that when you hover over a contour line with the cursor, Mathematica shows the numerical function value to which that line corresponds.

By default, a `ContourPlot` also includes the same coloring as a `DensityPlot`, except that it is discretized to show a uniform color between adjacent contour lines. Naturally there is an option to omit the shading:

```
ContourShading -> False
```

You can also add an option to change the total number of contour lines:

```
Contours -> 20
```

Surface plots

A surface plot of a function of two variables plots the function values on a third axis, to produce a three-dimensional surface. This surface is then drawn in perspective on a two-dimensional screen or printed page. The Mathematica function to produce such an image is `Plot3D`:

```
Plot3D[Sin[2*Pi*x] Sin[3*Pi*y], {x, 0, 1}, {y, 0, 1}]
```

Here are some options to try (preferably one at a time) that modify the plot's appearance:

```
Mesh -> None,
PlotPoints -> 100,
BoxRatios -> {1, 1, 0.6},
Axes -> False,
Boxed -> False,
ViewPoint -> {-0.5, -2, 1}
```

Note that you can also rotate the view using the mouse (or other pointing device).

Of course you can also modify the color of the surface, but this can get complicated due to the simulated lighting and surface shininess. Some colors look ok without any further changes:

```
PlotStyle -> Hue[.33, .75, 1]
```

But most colors look odd, at least to my eyes, under the default colored light sources. The easiest fix is to switch to "neutral" lighting:

```
Lighting -> "Neutral"
```

This setting has the side effect of making the surface look too shiny (in my opinion), so I usually just turn the shininess completely off:

```
PlotStyle -> {Hue[.75, .25, .85], Specularity[0]}
```

At this point you can experiment with a variety of different hue, saturation, and brightness values.

Complex-valued functions

The trick to making a phase-as-color density plot of a complex-valued function is not to use `DensityPlot` at all, but rather to use `RegionPlot`, which can fill any region in the plane using an arbitrary `ColorFunction`. In general you would specify the region to fill using a condition on `x` and `y`, but we can just fill the entire region using the trivial condition `True`. Here, then, is some code to plot a two-dimensional wavepacket:

```
wp2d = Exp[-(x^2+y^2)] Exp[I(8x-5y)];
RegionPlot[True, {x, -3, 3}, {y, -3, 3},
  PlotPoints -> 100,
  BoundaryStyle -> None,
  ColorFunction ->
    Function[{x, y}, Hue[Arg[wp2d]/(2Pi), 1, Abs[wp2d]]],
  ColorFunctionScaling -> False]
```

As in the example at the end of Section B.3, I've set the hue to correspond to the phase of the function according to Figure A.6. Here I've also set the saturation to 1 and the brightness to the function magnitude, so we get black where the function value is zero or negligible. To map zero to white instead, set the brightness to 1 and the saturation to the function magnitude. This particular function has a maximum magnitude of 1, so the brightest parts of the image are as bright as possible. For other functions you may need to apply a scale factor to optimize the brightness.

An alternative is to make a surface plot of the function magnitude and then color the surface according to phase:

```
Plot3D[Abs[wp2d], {x, -3, 3}, {y, -3, 3},
  PlotRange -> All,
  PlotPoints -> 100,
  ColorFunction ->
    Function[{x, y}, Hue[Arg[wp2d]/(2Pi), Abs[wp2d], 0.75]],
  ColorFunctionScaling -> False,
  Lighting -> "Neutral",
  Mesh -> None]
```

In this example I've made the areas where the function is negligible light gray, since neither black nor white looked very attractive to my eyes.

Plotting from lists of values

To plot a function of two variables from a list of sampled values (perhaps numerically generated), Mathematica provides `ListDensityPlot`, `ListContourPlot`, and `ListPlot3D` functions. All of them work as you would probably expect, and you can look them up in the documentation for details.

Alternatively, you can always use `Interpolation` to turn a list of values into an interpolating function, then plot that function as described in the preceding subsections. This is how you would want to handle a discretely sampled complex-valued function of two variables.

Problem B.10. Make a density plot, a contour plot, and a surface plot of the function $((x - y)^3 - 2x + 3y)^2 e^{-(x^2+y^2)}$, over a wide enough range of x and y values to show its full shape. Use nonstandard but appropriate colors for each plot, chosen to suit your taste.

Problem B.11. Make a density plot of the function $(\sin r)/r$, where $r = \sqrt{x^2 + y^2}$, showing r values out to about 20. Use a color function that maps zero to white, with distinct hues for positive and negative values.

Problem B.12. Use `Plot3D` to make a combined surface plot of the real and imaginary parts of the wavepacket defined above as `wp2d`. That is, show the real and imaginary parts as separate, intersecting surfaces on the same plot. Be sure to use a large enough value of `PlotPoints` to show the functions accurately. Start with the default colors, then switch to two other colors of your choice, adjusting the color parameters as appropriate.

Problem B.13. Make a phase-as-color density plot of the function $\sin(\pi x) \sin(2\pi y) + i \sin(2\pi x) \sin(\pi y)$, for x and y values ranging from 0 to 1 (see Figure 4.8). Make several versions of the plot, mapping zero to either black or white and mapping either the magnitude or the squared magnitude to the brightness/saturation. Be sure to scale the function magnitude so its maximum value is mapped to a brightness/saturation of 1.0. Which version of the plot do you like best?

B.5 Calculus and other math

Mathematica is probably best known for its symbolic manipulation features, including the ability to work out otherwise-laborious derivatives and integrals. Even more useful for physicists, however, are its built-in numerical algorithms for integration, root finding, optimization, differential equations, and matrix operations.

Derivatives

The function to take a symbolic derivative is called simply `D`:

```
psi = (4x^4 - 12x^2 + 3)Exp[-x^2/2];
deriv = D[psi, x]
```

To take the second derivative you can either just apply the `D` function twice, or use this syntax:

```
deriv2 = D[psi, {x, 2}]
```

Often the result of such a calculation will be unnecessarily complicated, but Mathematica also provides a handy `Simplify` function:

```
Simplify[deriv2]
```

Integrals

To take antiderivatives you use the `Integrate` function:

```
Integrate[Sin[x]^2, x]
```

You can also do definite integrals, using a syntax similar to that of `Plot` and `Table`:

```
Integrate[Sin[x]^2, {x, 0, Pi}]
```

Notice that the answer is provided exactly in terms of π . If you'd prefer a numerical answer in decimal approximation, you can apply the `N` function:

```
N[Integrate[Sin[x]^2, {x, 0, Pi}]]
```

Alternatively, you can just use the `NIntegrate` function, which does the integral numerically from the start:

```
NIntegrate[Exp[-x^2], {x, 0, 1}]
```

Naturally, the answers provided by `NIntegrate` are only approximate. By default it tries to be accurate to several significant digits, and therefore it runs into trouble and complains if you ask it to evaluate an integral whose answer is exactly zero:

```
NIntegrate[Sin[x], {x, 0, 2Pi}]
```

By the way, both `Integrate` and `NIntegrate` can handle limits of $\pm\infty$:

```
Integrate[Exp[-x^2], {x, -Infinity, Infinity}]
```

Sums

Sums use the same syntax as integrals:

```
Sum[1/2^n, {n, 0, 10}]
```

The default step size is 1, but you can specify any step size:

```
Sum[1/2^n, {n, 0, 10, 2}]
```

And again you can use `Infinity` as the upper limit:

```
Sum[1/2^n, {n, 0, Infinity}]
```

Root finding and optimization

When you can't perform algebraic manipulations to solve an equation for the variable it determines, numerical root finding is called for. The most familiar examples are transcendental equations, for instance, $\sin x = x/10$. This equation has a “trivial” solution at $x = 0$, but it has other solutions as well, and there's no formula for them. However, you can easily plot both sides to see where the solutions are:

```
Plot[{Sin[x], x/10}, {x, 0, 10}]
```

As you can see, there are three nontrivial solutions for positive x (and three symmetrical solutions for negative x). To find them numerically, use the `FindRoot` function:

```
FindRoot[Sin[x] == x/10, {x, 3}]
```

Notice the *double ==* sign, indicating that this is an equation to be solved, not an attempt to assign a name to an expression. The number 3 is the x value where Mathematica will begin looking for a solution. If you give it a different starting point it may find a different solution, or none at all:

```
FindRoot[Sin[x] == x/10, {x, 5}]
```

Try some different starting points until you have found all three of the positive solutions to this equation. The lesson here is that you should always (if at all possible) plot the graph first, rather than blindly looking for solutions with arbitrary starting points.

A related numerical task is to find the maximum or minimum value of a function. An important example is the distribution of photon energies radiated by a thermal (“blackbody”) source:

```
planckSpectrum = x^3/(Exp[x] - 1);
Plot[planckSpectrum, {x, 0, 12}]
```

If you try to locate the peak of this curve by setting the derivative of the function to zero, you get a transcendental equation. Although you could solve that equation using `FindRoot` (see Problem B.14), it's easier to just use `FindMaximum`:

```
FindMaximum[planckSpectrum, {x, 3}]
```

There's also a `FindMinimum` function that works in the same way.

In more complicated examples, the equations and functions that you need to solve or optimize may themselves involve numerical calculations (e.g., `NIntegrate`). In these cases you'll need to provide not one but two suggested starting points: `{x, 2, 4}`. Then Mathematica will have some idea of how big a step to take, to one side or the other, to get closer to the solution. When you provide just one starting point, Mathematica differentiates the expression(s) to guess a good step size—but

it can't take the derivative of an expression that's defined only numerically. For an example in which two starting points are needed, see Section 3.4.

The answers returned by `FindRoot`, `FindMaximum`, and `FindMinimum` are wrapped inside of short lists and replacement rules. To extract the numbers themselves for further use, you can use double brackets to index into these structures:

```
FindRoot[Sin[x] == x/10, {x, 3}][[1, 2]]
```

Here the appended “`[[1, 2]]`” says to extract the second part of the first part of the resulting expression. The expression itself is a single-element list, whose first part is the replacement rule $x \rightarrow 2.85234$. The second part of *this* expression is the number 2.85234 itself. Notice that the double-bracket notation can extract parts from *any* expression, not just from lists. Here are the instructions to extract the maximum value and its location from the Planck spectrum example:

```
FindMaximum[planckSpectrum, {x, 3}][[1]]
```

```
FindMaximum[planckSpectrum, {x, 3}][[2, 1, 2]]
```

(To see the full structure of an expression, for the purpose of extracting the part you want, you can enclose the expression in `FullForm[]`.)

Ordinary differential equations

One of Mathematica's most powerful and useful numerical functions is `NDSolve`, for numerical solution of differential equations. Here is an example of its use on an ordinary differential equation:

```
solution = NDSolve[{f''[x] == -x*f[x], f[0] == 1, f'[0] == 0},
  f, {x, -2, 10}]
```

The syntax here is a little tricky. The first argument of `NDSolve` is a *list* of equations, specifically the differential equation to solve and its boundary conditions. For a second-order differential equation, the boundary conditions should specify the value and the first derivative of the unknown function at the same point (here $x = 0$). Notice that all of these equations are written with double `==` signs, and that the prime symbol `'` indicates a derivative. The second argument of `NDSolve` is the name of the function to solve for, and the third argument is a list of the independent variable and its minimum and maximum values, as in `Plot` or `Integrate`. Notice that the boundary conditions needn't be specified at one of the actual boundary points.

`NDSolve` returns an interpolating function, awkwardly nested inside a replacement rule and two levels of lists. To extract it for plotting or direct evaluation you can index into this structure with `[[1, 1, 2]]`, but it's usually easier to use the following syntax:

```
Plot[f[x] /. solution, {x, -2, 10}]
```

Vectors and matrices

You can use a Mathematica list to represent a vector with any (finite) number of components. Addition, subtraction, and scalar multiplication then work automatically:

```
vecA = {2, 1, -2};
vecB = {-1, 3, -2};
vecA + vecB

2vecA - vecB
```

To take the dot product of two vectors, use the `Dot` function or simply put a period (.) between them:

```
vecA . vecB
```

To represent a matrix you use a list of lists, with each inner list representing one row. The `MatrixForm` function tells Mathematica to display the nested list as a matrix in the traditional way:

```
matrixM = {{1, 0, 2}, {0, -3, 1}, {2, 1, 2}};
MatrixForm[matrixM]
```

As with vectors, the operations of addition, subtraction, and scalar multiplication of matrices just work. To carry out matrix multiplication you again use the dot symbol (or `Dot` function):

```
matrixM . vecA
```

Notice that the vector is treated as a column vector in this example. Alternatively, you can turn it into a true one-column matrix by enclosing it in another set of curly braces (to make a one-row matrix) and then taking the transpose:

```
columnMatrixA = Transpose[{vecA}]

MatrixForm[matrixM . columnMatrixA]
```

To make a diagonal matrix (with zeros in all the off-diagonal slots), there's a nice shortcut:

```
MatrixForm[DiagonalMatrix[{1, 2, 3}]]
```

The most important matrix operation in quantum mechanics is to find a matrix's eigenvalues and eigenvectors—that is, to find a new basis in which it is diagonal. The Mathematica function for this operation is `Eigensystem` (note the lower-case `s` where you might expect a capital):

```
sxMatrix = {{0, 1, 0}, {1, 0, 1}, {0, 1, 0}}/Sqrt[2];
MatrixForm[sxMatrix]

{eValues, eVectors} = Eigensystem[sxMatrix]
```

I've provided separate names for the lists of eigenvalues and -vectors, which `Eigensystem` wraps in a two-element list. It's easy to check the results:

```
sxMatrix . eVectors[[1]]
```

Notice that the eigenvalues and -vectors are sorted in corresponding order, so `eValues[[1]]` goes with `eVectors[[1]]`.

For matrices consisting of exact integers or algebraic expressions, `Eigensystem` tries to find exact results and does not bother to normalize the eigenvectors. But if the matrix contains even one element that has a decimal point (to indicate a number stored in decimal approximation), then `Eigensystem` will use an approximate numerical algorithm and return answers in decimal approximation, with normalized eigenvectors:

```
Eigensystem[sxMatrix*1.0]
```

The numerical algorithm used in such cases is remarkably fast, so `Eigensystem` can handle numerical matrices with hundreds, or even thousands, of rows and columns.

When we work with a quantum system that's built from two smaller systems, we need to take tensor products of matrices, as discussed in Sections 4.4 and 7.4. Mathematica *has* a function called `TensorProduct`, but it doesn't quite do what we want (try it out if you like); instead the needed Mathematica function is called `KroneckerProduct`:

```
vec1 = {{a}, {b}};
vec2 = {{c}, {d}, {e}};
MatrixForm[KroneckerProduct[vec1, vec2]]

MatrixForm[KroneckerProduct[vec2, vec1]]

sxMatrix = (1/2) {{0, 1}, {1, 0}};
MatrixForm[KroneckerProduct[sxMatrix, IdentityMatrix[2]]]

MatrixForm[KroneckerProduct[IdentityMatrix[2], sxMatrix]]
```

Notice that the *first* matrix in the Kronecker product gets "stretched," so that each of its elements can be multiplied by the *second* matrix in the product.

Problem B.14. Consider the function $x^3/(e^x - 1)$, defined above as `planckSpectrum`.
 (a) Use Mathematica to find the derivative of this function. (b) What do you get when you try to use Mathematica to find the antiderivative of this function? (c) Find the exact value of the definite integral of this function from 0 to ∞ . (d) Use `NIntegrate`

to find what fraction of the total area under the graph of this function lies at x values greater than 5.0. Check your plot of the function to be sure the answer is reasonable.

(e) Apply `FindRoot` to the derivative you found in part (a), to verify the location of the function's maximum that you found earlier using `FindMaximum`.

Problem B.15. Use `Manipulate`, `Plot`, and `Sum` to create an interactive plot of the Fourier series

$$\sum_{\text{odd } n} (-1)^{(n-1)/2} \frac{\sin(n\pi x)}{n^2},$$

where x ranges from 0 to 1 and the sum goes up to some maximum n value that you can vary by dragging the slider. What shape does this series converge to? What is the function's maximum value in that limit?

Problem B.16. The motion of a simple pendulum is governed by the differential equation $d^2\theta/dt^2 = -\sin\theta(t)$. Use `NDSolve` to solve this equation and produce a plot of $\theta(t)$ over a few cycles. How does the plot change as the amplitude increases toward π ?

Problem B.17. Enter the matrices

$$S_x = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad S_y = \frac{1}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad S_z = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

into Mathematica, and then compute (and display using `MatrixForm`) the combinations $S_x S_y - S_y S_x$, $S_y S_z - S_z S_y$, and $S_z S_x - S_x S_z$ (called *commutators*; see Section 5.1). In each case you should find that the commutator of two of the S matrices is proportional to the third, as stated (with some extra factors of \hbar) in equation 7.1.

Problem B.18. Use the `Table` function and the `RandomReal` function (look it up) to generate a 10×10 matrix of random real numbers in the range -1 to 1 . Find the eigenvalues of this matrix, and note that some of them are complex. Then add the matrix to its transpose to obtain a matrix that is symmetric (but otherwise still random). Find the eigenvalues of this matrix, and note that they are all real. Also calculate the dot products of several of the eigenvectors with each other, and note that they are orthogonal (to the accuracy one could reasonably expect).

B.6 Loops and simulations

Despite Mathematica's ability to apply functions to entire lists, there are times when we need to use it like a more traditional programming language, carrying out repeated operations sequentially. For this purpose it provides `While` and `For` functions that "loop" through instructions repeatedly, just as the similarly named structures do in a language like C.

While and For loops

Here is an example showing how to use the `While` function:

```
n = 1;
While[n < 1000,
  Print[n];
  n = 2n]
```

I hasten to point out that there are much better ways to use Mathematica to make a list of the first nine powers of 2. But the point here is just to see how the `While` function works. Its first argument is a condition, while its second argument is a sequence of instructions (just two of them in this case), separated by semicolons. These instructions are executed repeatedly, as long as the condition holds. Notice that the symbol `n` has the value 1024 by the time the loop terminates, but this value never gets printed. Oh yes: the `Print` function is a convenient way of displaying results while the loop is executing.

Alternatively, we can accomplish exactly the same task with a `For` loop:

```
For[n = 1, n < 1000, n = 2n,
    Print[n]]
```

This version is more compact but a little harder to get used to. The four arguments of `For` are an initialization of the variable that controls the loop, a condition on that variable (used the same as in `While`), a statement that modifies the value of the variable, and finally the statement(s) to execute each time through the loop. Perhaps a bit confusingly, the variable-modification statement (here `n = 2n`) appears *before* the body of the loop (here `Print[n]`), but is executed *after* each loop iteration (immediately before the condition is again tested).

This use of `For` is perfectly legitimate, but it's somewhat traditional to use `For` only when the variable-modification statement adds a fixed amount to the variable, rather than multiplying it as in this case. The most common modification is simply to add 1. By the way, Mathematica provides the same shortcuts for operating on a variable as in C: you can say `n += 1` or simply `n++` in place of `n = n + 1`.

Building a list

Sometimes the purpose of a loop is to build a list of data for later use (e.g., plotting). To do this you can initialize the list to contain no elements, then use the `AppendTo` function within the body of the loop:

```
n = 1;
nList = {};
While[n < 1000,
    AppendTo[nList, n];
    n = 2n];
nList
```

Again, this is a poor example because you could construct the resulting list more easily using the `Table` function. Section 2.2 contains a much better example of `While` and `AppendTo`.

Ongoing animation

To observe what's happening while a loop executes, you can wrap the `Dynamic` function around any code whose output should be continuously updated. The most

common application is to a plot:

```
t = 0;
Dynamic[Plot[Sin[x - t], {x, 0, 4Pi}]]
```

Now whenever the parameter `t` is changed, the plot will update automatically. If you change `t` repeatedly in a loop, you get an animated graphic:

```
While[t < 50,
  t += 0.1;
  Pause[0.033]]
```

The `Pause` function, whose argument is in seconds, slows down the animation so we can see it. Here you could even change the condition `t < 50` to simply `True`, in which case the animation would continue indefinitely, until you interrupt it with the Abort Evaluation menu command. But once again this example isn't a good one: Because the sine function is periodic, you could produce the same animated graphic more easily using `Manipulate`. A better example of a dynamically updating plot is in Section 3.3.

Problem B.19. Mathematica has a built-in `Fibonacci` function, but pretend it doesn't. Use a `While` or `For` loop to build a list consisting of the first 20 numbers in the Fibonacci sequence, which begins with 1,1 and continues by adding the last two elements to obtain the next element. You can reference the last element of a list as `list[[-1]]` and the second-to-last element as `list[[-2]]`.

Problem B.20. You can draw simple pictures in Mathematica using the `Graphics` function:

```
Graphics[{Yellow, Disk[{0, 0}, 0.1], Blue, Disk[{1, 0}, 0.05]},
  PlotRange -> {{-2, 2}, {-2, 2}}, Frame -> True]
```

This example draws two filled circles, to represent the sun and a planet. Starting from this code, animate the image to put the planet into motion, in a circular orbit around the sun. (If you know how to write a Newtonian orbit simulation, do so and explore some elliptical orbits.)