

*The Physics of  
Quantum Mechanics*



Daniel F. Styer

# *The Physics of Quantum Mechanics*

Daniel F. Styer  
Schiffer Professor of Physics, Oberlin College

This book is in **draft** form — it is not polished or complete. It needs more problems. I appreciate your comments.

copyright © 19 August 2021 Daniel F. Styer

The copyright holder grants the freedom to copy, modify, convey, adapt, and/or redistribute this work under the terms of the Creative Commons Attribution Share Alike 4.0 International License. A copy of that license is available at <http://creativecommons.org/licenses/by-sa/4.0/legalcode>.

You may freely download this book in pdf format from

<http://www.oberlin.edu/physics/dstyer/ThePhysicsOfQM>.

It is formatted to print nicely on either A4 or U.S. Letter paper. The author receives no monetary gain from your download: it is reward enough for him that you want to explore quantum mechanics.

Instructions for living a life:  
Pay attention.  
Be astonished.  
Tell about it.

— Mary Oliver, *Sometimes*



# Contents

<i>Welcome</i>	1
1. What is Quantum Mechanics About?	5
1.1 Quantization . . . . .	5
1.2 Interference . . . . .	22
1.3 Aharonov-Bohm effect . . . . .	31
1.4 Light on the atoms . . . . .	33
1.5 Entanglement . . . . .	35
1.6 Quantum cryptography . . . . .	48
1.7 What is a qubit? . . . . .	52
2. Forging Mathematical Tools	55
2.1 What is a quantal state? . . . . .	55
2.2 Amplitude . . . . .	57
2.3 Reversal-conjugation relation . . . . .	64
2.4 Establishing a phase convention . . . . .	66
2.5 How can I specify a quantal state? . . . . .	68
2.6 States for entangled systems . . . . .	77
2.7 Are states “real”? . . . . .	82
2.8 What is a qubit? . . . . .	82

3.	Refining Mathematical Tools	85
3.1	Extras . . . . .	85
3.2	Outer products, operators, measurement . . . . .	90
3.3	Photon polarization . . . . .	96
3.4	Lightning linear algebra . . . . .	100
3.5	Problems . . . . .	109
4.	Formalism	113
4.1	The role of formalism . . . . .	113
4.2	The density matrix . . . . .	117
5.	Time Evolution	119
5.1	Operator for time evolution . . . . .	119
5.2	Working with the Schrödinger equation . . . . .	123
5.3	Formal properties of time evolution; Conservation laws . . . . .	135
5.4	Magnetic moment in a uniform magnetic field . . . . .	139
5.5	The neutral K meson . . . . .	139
6.	The Quantum Mechanics of Position	143
6.1	Describing states in continuum systems . . . . .	143
6.2	How does position amplitude change with time? . . . . .	150
6.3	What is wavefunction? . . . . .	158
6.4	Operators and their representations; The momentum basis . . . . .	159
6.5	Time evolution of average quantities . . . . .	169
7.	The Free Particle	173
7.1	Problems . . . . .	173

8. Square Wells	177
8.1 What does an electron look like? . . . . .	178
9. The Simple Harmonic Oscillator	181
9.1 Resume of energy eigenproblem . . . . .	181
9.2 Solution of the energy eigenproblem: Differential equation approach . . . . .	181
9.3 Solution of the energy eigenproblem: Operator factorization approach . . . . .	186
9.4 Problems . . . . .	190
10. Qualitative Solution of Energy Eigenproblems	193
11. Perturbation Theory	195
11.1 The $\mathcal{O}$ notation . . . . .	195
11.2 Perturbation theory for cubic equations . . . . .	198
11.3 Derivation of perturbation theory for the energy eigenproblem . . . . .	201
11.4 Perturbation theory for the energy eigenproblem: Summary of results . . . . .	205
11.5 Problems . . . . .	206
12. Quantum Mechanics in Two and Three Dimensions	211
12.1 More degrees of freedom . . . . .	211
12.2 Vector operators . . . . .	214
12.3 Multiple particles . . . . .	215
12.4 The phenomena of quantum mechanics . . . . .	216

13. Angular Momentum	219
13.1 Solution of the angular momentum eigenproblem . . . . .	219
13.2 Summary of the angular momentum eigenproblem . . . . .	223
13.3 Ordinary differential equations for the $d_{m,m'}^{(j)}(\theta)$ . . . . .	223
13.4 Problems . . . . .	224
14. Central Force Motion	227
14.1 Energy eigenproblem in two dimensions . . . . .	227
14.2 Energy eigenproblem in three dimensions . . . . .	234
14.3 Bound state energy eigenproblem for Coulombic potentials	239
14.4 Summary of the bound state energy eigenproblem for a Coulombic potential . . . . .	243
14.5 Problems . . . . .	244
15. Identical Particles	247
15.1 Many-particle systems in quantum mechanics . . . . .	247
15.2 An antisymmetric basis for the helium problem . . . . .	261
16. Breather	271
16.1 Scaled variables . . . . .	272
16.2 Variational method for finding the ground state energy . .	275
16.3 Problems . . . . .	277
17. Hydrogen	281
17.1 The Stark effect . . . . .	281
18. Helium	291
18.1 Ground state energy of helium . . . . .	291



19. Atoms	297
19.1 Addition of angular momenta . . . . .	297
19.2 Hartree-Fock approximation . . . . .	304
19.3 Atomic ground states . . . . .	305
20. Molecules	307
20.1 The hydrogen molecule ion . . . . .	307
20.2 Problems . . . . .	313
20.3 The hydrogen molecule . . . . .	314
20.4 Can we do better? . . . . .	315
21. WKB: The Quasiclassical Approximation	317
21.1 The connection region . . . . .	319
21.2 Why is WKB the “quasiclassical” approximation? . . . . .	320
21.3 The “power law” potential . . . . .	320
22. The Interaction of Matter and Radiation	329
22.1 Perturbation Theory for the Time Development Problem . . . . .	329
22.2 Setup . . . . .	330
22.3 Light absorption . . . . .	334
22.4 Absorbing incoherent light . . . . .	340
22.5 Absorbing and emitting light . . . . .	341
22.6 Problems . . . . .	346
23. The Territory Ahead	349

Appendix A	Tutorial on Matrix Diagonalization	351
A.1	What's in a name? . . . . .	351
A.2	Vectors in two dimensions . . . . .	352
A.3	Tensors in two dimensions . . . . .	355
A.4	Tensors in three dimensions . . . . .	359
A.5	Tensors in $d$ dimensions . . . . .	360
A.6	Linear transformations in two dimensions . . . . .	361
A.7	What does "eigen" mean? . . . . .	363
A.8	How to diagonalize a symmetric matrix . . . . .	364
A.9	A glance at computer algorithms . . . . .	370
A.10	A glance at non-symmetric matrices and the Jordan form	371
Appendix B	The Spherical Harmonics	377
Appendix C	Radial Wavefunctions for the Coulomb Problem	379
Appendix D	Quantum Mechanics Cheat Sheet	381
<i>Index</i>		385

# Welcome

Why would anyone want to study a book titled *The Physics of Quantum Mechanics*?

Starting in the year 1900, physicists exploring the newly discovered atom found that the atomic world of electrons and protons is not just smaller than our familiar world of trees, balls, and automobiles, it is also fundamentally different in character. Objects in the atomic world obey different rules from those obeyed by a tossed ball or an orbiting planet. These atomic rules are so different from the familiar rules of everyday physics, so counterintuitive and unexpected, that it took more than 25 years of intense research to uncover them.

But it is really only since the year 1990 that physicists have come to appreciate that the rules of the atomic world (now called “quantum mechanics”) are not just *different* from the everyday rules (now called “classical mechanics”). The atomic rules are also far *richer*. The atomic rules provide for phenomena like particle interference and entanglement that are simply absent from the everyday world. Every phenomenon of classical mechanics is also present in quantum mechanics, but the quantum world provides for many additional phenomena.

Here’s an analogy: Some films are in black-and-white and some are in color. It does not malign any black-and-white film to say that a color film has more possibilities, more richness. In fact, black-and-white films are simply one category of color films, because black and white are both colors. Anyone moving from the world of only black-and-white to the world of color is opening up the door to a new world — a world ripe with new possibilities and new expression — without closing the door to the old world.

This same flood of richness and freshness comes from entering the quantum world. It is a difficult world to enter, because we humans have no experience, no intuition, no expectations about this world. Even our language, invented by people living in the everyday world, has no words for the new quantal phenomena — just as a language among a race of the color-blind would have no word for “red”.

Reading this book is not easy: it is like a color-blind student learning about color from a color-blind teacher. The book is just one long argument, building up the structure of a world that we can explore not through touch or through sight or through scent, but only through logic. Those willing to follow and to challenge the logic, to open their minds to a new world, will find themselves richly rewarded.

### *The place of quantum mechanics in nature*

Quantum mechanics is the framework for describing and analyzing small things, like atoms and nuclei. Quantum mechanics also applies to big things, like baseballs and galaxies, but when applied to big things, certain approximations become legitimate: taken together, these are called the *classical approximation* to quantum mechanics, and the result is the familiar classical mechanics.

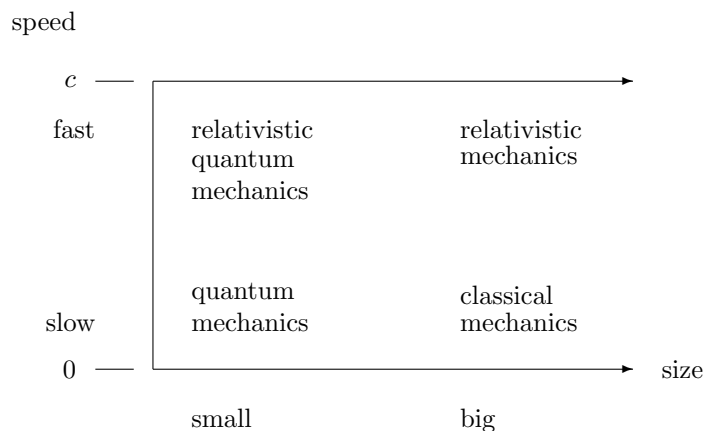
Quantum mechanics is not only less familiar and less intuitive than classical mechanics; it is also harder than classical mechanics. So whenever the classical approximation is sufficiently accurate, we would be foolish not to use it. This leads some to develop the misimpression that quantum mechanics applies to small things, while classical mechanics applies to big things. No. Quantum mechanics applies to all sizes, but classical mechanics is a good approximation to quantum mechanics when it is applied to big things.

For what size is the classical approximation good enough? That depends on the accuracy desired. The higher the accuracy demanded, the more situations will require full quantal treatment rather than approximate classical treatment. But as a rule of thumb, something as big as a DNA strand is almost always treated classically, not quantum mechanically.

This situation is analogous to the relationship between relativistic mechanics and classical mechanics. Relativity applies always, but classical mechanics is a good approximation to relativistic mechanics when applied

to slow things (that is, with speeds much less than light speed  $c$ ). The speed at which the classical approximation becomes legitimate depends upon the accuracy demanded, but as a rule of thumb particles moving less than a quarter of light speed are treated classically.

The difference between the quantal case and the relativistic case is that while relativistic mechanics is less familiar, less comforting, and less expected than classical mechanics, it is no more intricate than classical mechanics. Quantum mechanics, in contrast, is less familiar, less comforting, less expected, *and* more intricate than classical mechanics. This intricacy makes quantum mechanics harder than classical mechanics, yes, but also richer, more textured, more nuanced. Whether to curse or celebrate this intricacy is your choice.



Finally, is there a framework that applies to situations that are both fast and small? There is: it is called “relativistic quantum mechanics” and is closely related to “quantum field theory”. Ordinary non-relativistic quantum mechanics is a good approximation for relativistic quantum mechanics when applied to slow things. Relativistic mechanics is a good approximation for relativistic quantum mechanics when applied to big things. And classical mechanics is a good approximation for relativistic quantum mechanics when applied to big, slow things.

***What you can expect from this book***

This book introduces quantum mechanics at the third- or fourth-year American undergraduate level. It assumes the reader knows about. . . .

This is a book about physics, not about mathematics. The word “physics” derives from the Greek word for “nature”, so the emphasis lies in nature, not in the mathematics we use to describe nature. Thus the book starts with experiments about nature, then builds mathematical machinery to describe nature, then erects a formalism (“postulates”), and then moves on to applications, where the formalism is applied to nature and where the understanding of both nature and formalism is deepened.

The book never abandons its focus on nature. It provides a balanced, interwoven treatment of concepts, formalism, and applications so that each thread reinforces the other. There are many problems at many levels of difficulty, but no problem is there just for “make-work”: each has a “moral to the story”. Some problems are essential to the logical development of the subject: these are labeled (unsurprisingly) “essential”. Other problems promote learning far better than simple reading can: these are labeled “recommended”. Sample problems build both mathematical technique and physical insight.

The book does not merely convey correct ideas, it also refutes misconceptions. Just to get started, I list the most important and most pernicious misconceptions about quantum mechanics: (a) An electron has a position but you don’t know what it is. (b) The only states are energy states. (c) The wavefunction  $\psi(\vec{x}, t)$  is “out there” in space and you could reach out and touch it if only your fingers were sufficiently sensitive.

The object of the biographical footnotes in this book is twofold: First, to present the briefest of outlines of the subject’s historical development, lest anyone get the misimpression that quantum mechanics arose fully formed, like Aphrodite from sea foam. Second, to show that the founders of quantum mechanics were not inaccessible giants, but people with foibles and strengths, with interests both inside and outside of physics, just like you and me.

## Chapter 1

# What is Quantum Mechanics About?

### 1.1 Quantization

We are used to things that vary continuously: An oven can take on any temperature, a recipe might call for any quantity of flour, a child can grow to a range of heights. If I told you that an oven might take on the temperature of 172.1 °C or 181.7 °C, but that a temperature of 173.8 °C was physically impossible, you would laugh in my face.

So you can imagine the surprise of physicists on 14 December 1900, when Max Planck announced that certain features of blackbody radiation (that is, of light in thermal equilibrium) could be explained by assuming that the energy of the light could *not* take on any value, but only certain discrete values. Specifically, Planck found that light of frequency  $\omega$  could take on only the energies of

$$E = \hbar\omega(n + \frac{1}{2}), \quad \text{where } n = 0, 1, 2, 3, \dots, \quad (1.1)$$

and where the constant  $\hbar$  (now called the “reduced Planck constant”) is

$$\hbar = 1.054571817 \times 10^{-34} \text{ J s}. \quad (1.2)$$

(I use modern terminology and the current value for  $\hbar$ , rather than the terminology and value used by Planck in 1900.)

That is, light of frequency  $\omega$  can have an energy of  $3.5\hbar\omega$ , and it can have an energy of  $4.5\hbar\omega$ , but it is physically impossible for this light to have an energy of  $3.8\hbar\omega$ . Any numerical quantity that can take on only discrete values like this is called “quantized”. By contrast, a numerical quantity that can take on any value is called “continuous”.

The photoelectric effect supplies additional evidence that the energy of light comes only in discrete values. And if the energy of light comes in

discrete values, then it's a good guess that the energy of an atom comes in discrete values too. This good guess was confirmed through investigations of atomic spectra (where energy goes into or out of an atom via absorption or emission of light) and through the Franck–Hertz experiment (where energy goes into or out of an atom via collisions).

Furthermore, if the energy of an atom comes in discrete values, then it's a good guess that other properties of an atom — such as its magnetic moment — also take on only discrete values. The theme of this book is that these good guesses have all proved to be correct.

The story of Planck's<sup>1</sup> discovery is a fascinating one, but it's a difficult and elaborate story because it involves not just quantization, but also thermal equilibrium and electromagnetic radiation. The story of the discovery of atomic energy quantization is just as fascinating, but again fraught with intricacies. In an effort to remove the extraneous and dive deep to the heart of the matter, we focus on the magnetic moment of an atom. We will, to the extent possible, do a quantum-mechanical treatment of an atom's magnetic moment while maintaining a classical treatment of all other aspects — such as its energy and momentum and position. (In chapter 6, “The Quantum Mechanics of Position”, we take up a quantum-mechanical treatment of position, momentum, and energy.)

### 1.1.1 *The Stern-Gerlach experiment*

An electric current flowing in a loop produces a magnetic moment, so it makes sense that the electron orbiting (or whatever it does) an atomic nucleus would produce a magnetic moment for that atom. And of course, it also makes sense that physicists would be itching to measure that magnetic moment.

It is not difficult to measure the magnetic moment of, say, a scout compass. Place the magnetic compass needle in a known magnetic field and measure the torque that acts to align the needle with the field. You

---

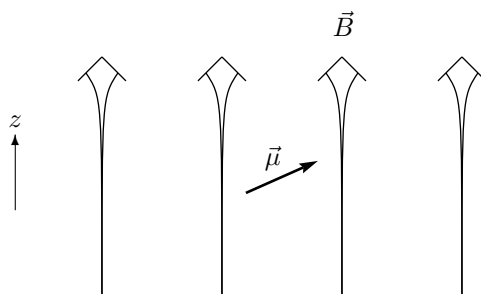
<sup>1</sup>Max Karl Ernst Ludwig Planck (1858–1947) was a German theoretical physicist particularly interested in thermodynamics and radiation. Concerning his greatest discovery, the introduction of quantization into physics, he wrote, “I can characterize the whole procedure as an act of desperation, since, by nature I am peaceable and opposed to doubtful adventures.” [Letter from Planck to R.W. Wood, 7 October 1931, quoted in J. Mehra and H. Reichenberg, *The Historical Development of Quantum Theory* (Springer-Verlag, New York, 1982) volume 1, page 49.]



will need to measure an angle and you might need to look up a formula in your magnetism textbook, but there is no fundamental difficulty.

Measuring the magnetic moment of an atom is a different matter. You can't even see an atom, so you can't watch it twist in a magnetic field like a compass needle. Furthermore, because the atom is very small, you expect the associated magnetic moment to be very small, and hence very hard to measure. The technical difficulties are immense.

These difficulties must have deterred but certainly did not stop Otto Stern and Walter Gerlach.<sup>2</sup> They realized that the twisting of a magnetic moment in a *uniform* magnetic field could not be observed for atomic-sized magnets, and also that the moment would experience zero net force. But they also realized that a magnetic moment in a *non-uniform* magnetic field *would* experience a net force, and that this force could be used to measure the magnetic moment.



*A classical magnetic moment in a non-uniform magnetic field.*

A classical magnetic moment  $\vec{\mu}$ , situated in a magnetic field  $\vec{B}$  that points in the  $z$  direction and increases in magnitude in the  $z$  direction, is subject to a force

$$\mu_z \frac{\partial B}{\partial z}, \quad (1.3)$$

where  $\mu_z$  is the  $z$ -component of the magnetic moment or, in other words, the projection of  $\vec{\mu}$  on the  $z$  axis. (If this is not obvious to you, then work problem 1.1, “Force on a classical magnetic moment”, on page 9.)

<sup>2</sup>Otto Stern (1888–1969) was a Polish-German-Jewish physicist who made contributions to both theory and experiment. He left Germany for the United States in 1933 upon the Nazi ascension to power. Walter Gerlach (1889–1979) was a German experimental physicist. During the Second World War he led the physics section of the Reich Research Council and for a time directed the German effort to build a nuclear bomb.

Stern and Gerlach used this fact to measure the  $z$ -component of the magnetic moment of an atom. First, they heated silver in an electric “oven”. The vaporized silver atoms emerged from a pinhole in one side of the oven, and then passed through a non-uniform magnetic field. At the far side of the field the atoms struck and stuck to a glass plate. The entire apparatus had to be sealed within a good vacuum, so that collisions with nitrogen molecules would not push the silver atoms around. The deflection of an atom away from straight-line motion is proportional to the magnetic force, and hence proportional to the projection  $\mu_z$ . In this ingenious way, Stern and Gerlach could measure the  $z$ -component of the magnetic moment of an atom even though any single atom is invisible.

Before reading on, pause and think about what results you would expect from this experiment.

Here are the results that I expect: I expect that an atom which happens to enter the field with magnetic moment pointing straight up (in the  $z$  direction) will experience a large upward force. Hence it will move upward and stick high up on the glass-plate detector. I expect that an atom which happens to enter with magnetic moment pointing straight down (in the  $-z$  direction) will experience a large downward force, and hence will stick far down on the glass plate. I expect that an atom entering with magnetic moment tilted upward, but not straight upward, will move upward but not as far up as the straight-up atoms, and the mirror image for an atom entering with magnetic moment tilted downward. I expect that an atom entering with horizontal magnetic moment will experience a net force of zero, so it will pass through the non-uniform field undeflected.

Furthermore, I expect that when a silver atom emerges from the oven source, its magnetic moment will be oriented randomly — as likely to point in one direction as in any other. There is only one way to point straight up, so I expect that very few atoms will stick high on the glass plate. There are many ways to point horizontally, so I expect many atoms to pass through undeflected. There is only one way to point straight down, so I expect very few atoms to stick far down on the glass plate.<sup>3</sup>

In summary, I expect that atoms would leave the magnetic field in any of a range of deflections: a very few with large positive deflection, more with a

---

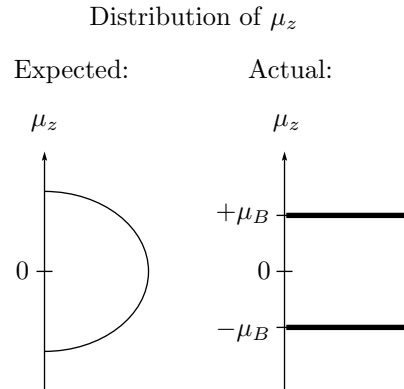
<sup>3</sup>To be specific, this reasoning suggests that the number of atoms with moment tilted at angle  $\theta$  relative to the  $z$  direction is proportional to  $\sin \theta$ , where  $\theta$  ranges from  $0^\circ$  to  $180^\circ$ . You might want to prove this to yourself, but we’ll never use this result so don’t feel compelled.

small positive deflection, a lot with no deflection, some with a small negative deflection, and a very few with large negative deflection. This continuity of deflections reflects a continuity of magnetic moment projections.

In fact, however, this is not what happens at all! The projection  $\mu_z$  does not take on a continuous range of values. Instead, it is quantized and takes on only two values, one positive and one negative. Those two values are called  $\mu_z = \pm\mu_B$  where  $\mu_B$ , the so-called “Bohr magneton”, has the measured value of

$$\mu_B = 9.274\,010\,078 \times 10^{-24} \text{ J/T}, \quad (1.4)$$

with an uncertainty of 3 in the last decimal digit.



The Stern-Gerlach experiment was initially performed with silver atoms but has been repeated with many other types of atoms. When nitrogen is used, the projection  $\mu_z$  takes on one of the four quantized values of  $+3\mu_B$ ,  $+\mu_B$ ,  $-\mu_B$ , or  $-3\mu_B$ . When sulfur is used, it takes on one of the five quantized values of  $+4\mu_B$ ,  $+2\mu_B$ ,  $0$ ,  $-2\mu_B$ , and  $-4\mu_B$ . For no atom do the values of  $\mu_z$  take on the broad continuum of my classical expectation. For all atoms, the projection  $\mu_z$  is quantized.

## Problems

### 1.1 Force on a classical magnetic moment

The force on a classical magnetic moment is most easily calculated using “magnetic charge fiction”: Consider the magnetic moment

to consist of two “magnetic charges” of magnitude  $+m$  and  $-m$ , separated by the position vector  $\vec{d}$  running from  $-m$  to  $+m$ . The magnetic moment is then  $\vec{\mu} = m\vec{d}$ .

- a. Use  $B_+$  for the magnitude of the magnetic field at  $+m$ , and  $B_-$  for the magnitude of the magnetic field at  $-m$ . Show that the net force on the magnetic moment is in the  $z$  direction with magnitude  $mB_+ - mB_-$ .
- b. Use  $d_z$  for the  $z$ -component of the vector  $\vec{d}$ . Show that to high accuracy

$$B_+ = B_- + \frac{\partial B}{\partial z} d_z.$$

Surely, for distances of atomic scale, this accuracy is more than adequate.

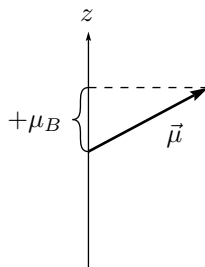
- c. Derive expression (1.3) for the force on a magnetic moment.

### 1.1.2 *The conundrum of projections*

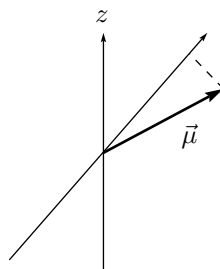
I would expect the projections  $\mu_z$  of a silver atom to take on a continuous range of values. But in fact, these values are quantized: Whenever  $\mu_z$  is measured, it turns out to be either  $+\mu_B$  or  $-\mu_B$ , and never anything else. This is counterintuitive and unexpected, but we can live with the counterintuitive and unexpected — it happens all the time in politics.

However, this fact of quantization appears to result in a logical contradiction, because there are many possible axes upon which the magnetic moment can be projected. The figures on the next page make it clear that it is impossible for *any* vector to have a projection of either  $\pm\mu_B$  on *all* axes!

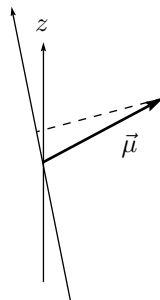
Because if the projection of  $\vec{\mu}$  on the  $z$  axis is  $+\mu_B \dots$



$\dots$  then the projection of  $\vec{\mu}$  on this second axis must be more than  $+\mu_B \dots$



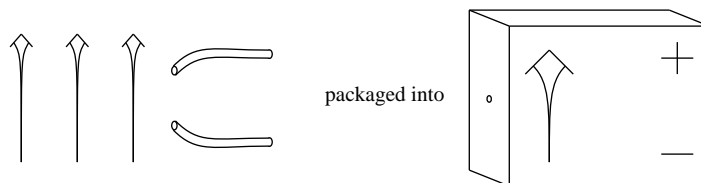
$\dots$  while the projection of  $\vec{\mu}$  on this third axis must be less than  $+\mu_B$ .



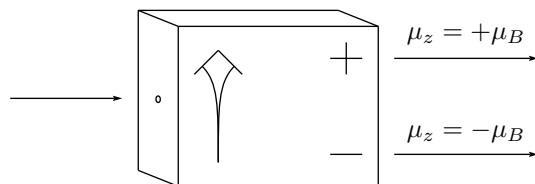
Whenever we measure the magnetic moment, projected onto any axis, the result is either  $+\mu_B$  or  $-\mu_B$ . Yet is it impossible for the projection of any classical arrow on *all* axes to be either  $+\mu_B$  or  $-\mu_B$ ! This seeming

contradiction is called “the conundrum of projections”. We can live with the counterintuitive, the unexpected, the strange, but we cannot live with a logical contradiction. How can we resolve it?

The resolution comes not from meditating on the question, but from experimenting about it. Let us actually measure the projection on one axis, and then on a second. To do this easily, we modify the Stern-Gerlach apparatus and package it into a box called a “Stern-Gerlach analyzer”. This box consists of a Stern-Gerlach apparatus followed by “pipes” that channel the outgoing atoms into horizontal paths.<sup>4</sup> This chapter treats only silver atoms, so we use analyzers with two exit ports.



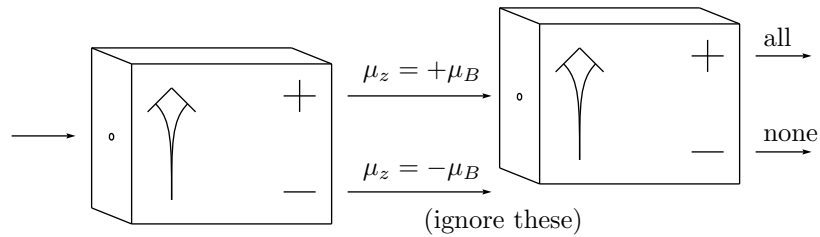
An atom enters a vertical analyzer through the single hole on the left. If it exits through the upper hole on the right (the “+ port”) then the outgoing atom has  $\mu_z = +\mu_B$ . If it exits through the lower hole on the right (the “- port”) then the outgoing atom has  $\mu_z = -\mu_B$ .



<sup>4</sup>In general, the “pipes” will manipulate the atoms through electromagnetic fields, not through touching. One way way to make such “pipes” is to insert a second Stern-Gerlach apparatus, oriented upside-down relative to the first. The atoms with  $\mu_z = +\mu_B$ , which had experienced an upward force in the first half, will experience an equal downward force in the second half, and the net impulse delivered will be zero. But whatever their manner of construction, the pipes must *not* change the magnetic moment of an atom passing through them.

### 1.1.3 Two vertical analyzers

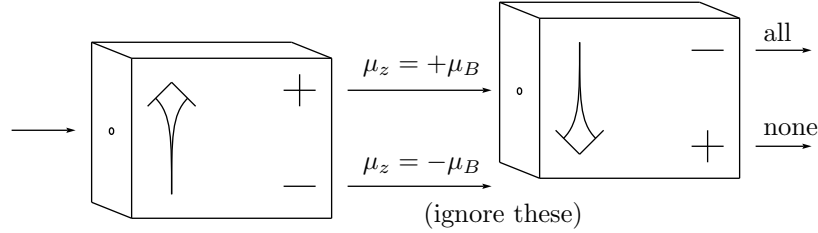
In order to check the operation of our analyzers, we do preliminary experiments. Atoms are fed into a vertical analyzer. Any atom exiting from the + port is then channeled into a second vertical analyzer. That atom exits from the + port of the second analyzer. This makes sense: the atom had  $\mu_z = +\mu_B$  when exiting the first analyzer, and the second analyzer confirms that it has  $\mu_z = +\mu_B$ .



Furthermore, if an atom exiting from the – port of the first analyzer is channeled into a second vertical analyzer, then that atom exits from the – port of the second analyzer.

### 1.1.4 One vertical and one upside-down analyzer

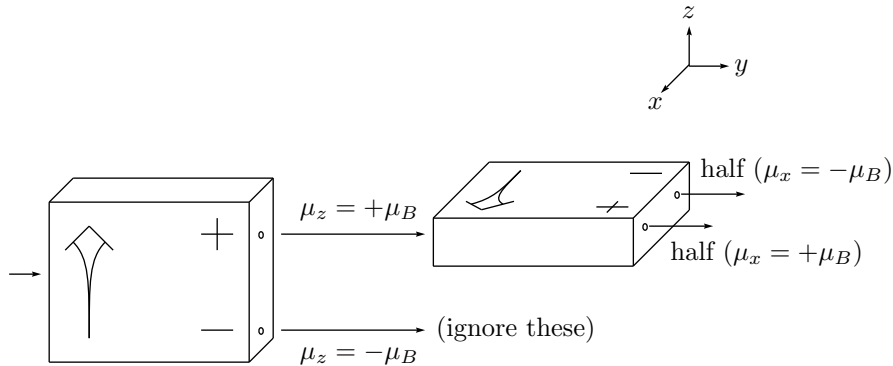
Atoms are fed into a vertical analyzer. Any atom exiting from the + port is then channeled into a second analyzer, but this analyzer is oriented *upside-down*. What happens? If the projection on an upward-pointing axis is  $+\mu_B$  (that is,  $\mu_z = +\mu_B$ ), then the projection on a downward-pointing axis is  $-\mu_B$  (we write this as  $\mu_{(-z)} = -\mu_B$ ). So I expect that these atoms will emerge from the – port of the second analyzer (which happens to be the higher port). And this is exactly what happens.



Similarly, if an atom exiting from the  $-$  port of the first analyzer is channeled into an upside-down analyzer, then that atom emerges from the  $+$  port of the second analyzer.

### 1.1.5 One vertical and one horizontal analyzer

Atoms are fed into a vertical analyzer. Any atom exiting from the  $+$  port is then channeled into a second analyzer, but this analyzer is oriented *horizontally*. The second analyzer doesn't measure the projection  $\mu_z$ , it measures the projection  $\mu_x$ . What happens in this case? Experiment shows that the atoms emerge randomly: half from the  $+$  port, half from the  $-$  port.



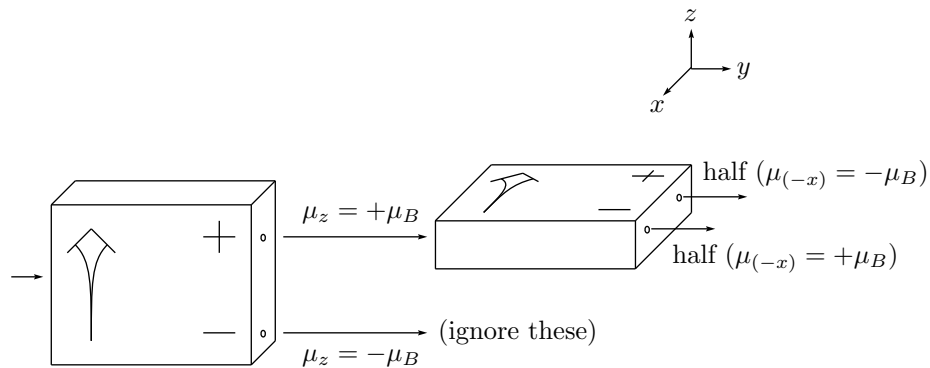
This makes some sort of sense. If a classical magnetic moment were vertically oriented, it would have  $\mu_x = 0$ , and such a classical moment would go straight through a horizontal Stern-Gerlach analyzer. We've seen that atomic magnetic moments *never* go straight through. If you "want" to go straight but are forced to turn either left or right, the best you can do is turn left half the time and right half the time. (Don't take this paragraph



literally... atoms have no personalities and they don't "want" anything. But it is a useful mnemonic.)

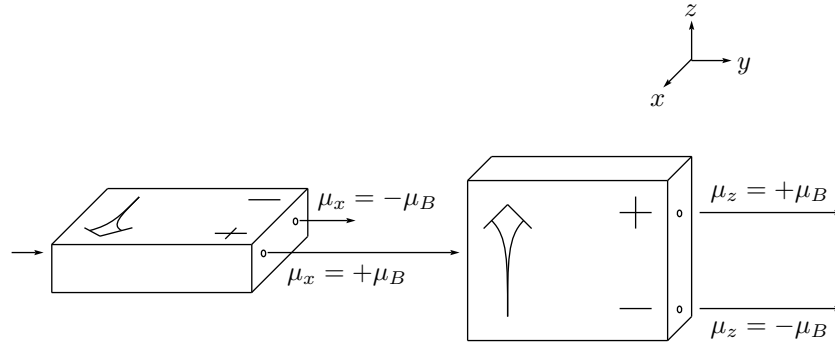
### 1.1.6 *One vertical and one backwards horizontal analyzer*

Perform the same experiment as above (section 1.1.5), except insert the horizontal analyzer in the opposite sense, so that it measures the projection on the negative  $x$  axis rather than the positive  $x$  axis. Again, half the atoms emerge from the  $+$  port, and half emerge from the  $-$  port.



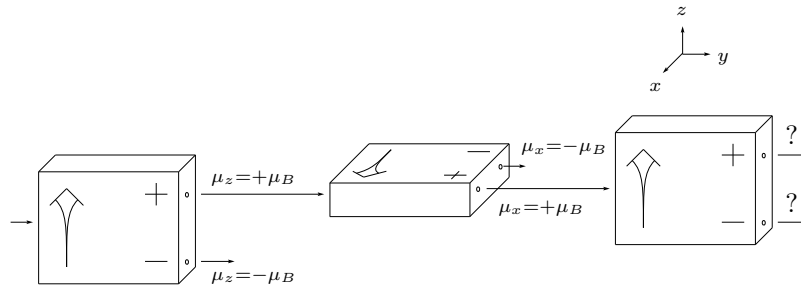
### 1.1.7 *One horizontal and one vertical analyzer*

A  $+x$  analyzer followed by a  $+z$  analyzer is the same apparatus as above (section 1.1.6), except that both analyzers are rotated as a unit by  $90^\circ$  about the  $y$  axis. So of course it has the same result: half the atoms emerge from the  $+$  port, and half emerge from the  $-$  port.



### 1.1.8 Three analyzers

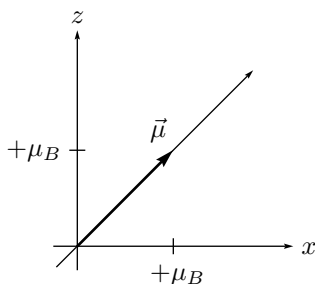
Atoms are fed into a vertical analyzer. Any atom exiting from the + port is then channeled into a horizontal analyzer. Half of these atoms exit from the + port of the horizontal analyzer (see section 1.1.5), and *these* atoms are channeled into a third analyzer, oriented vertically. What happens at the third analyzer?



There are two ways to think of this: (I) When the atom emerged from the + port of the first analyzer, it was determined to have  $\mu_z = +\mu_B$ . When that same atom emerged from the + port of the second analyzer, it was determined to have  $\mu_x = +\mu_B$ . Now we know two projections of the magnetic moment. When it enters the third analyzer, it still has  $\mu_z = +\mu_B$ , so it will emerge from the + port. (II) The last two analyzers in this sequence are a horizontal analyzer followed by a vertical analyzer, and from section 1.1.7 we know what happens in this case: a 50/50 split. That will happen in this case, too.

So, analysis (I) predicts that all the atoms entering the third analyzer will exit through the + port and none through the – port. Analysis (II) predicts that half the atoms will exit through the + port and half through the – port.

Experiment shows that analysis (II) gives the correct result. But what could possibly be wrong with analysis (I)? Let’s go through line by line: “When the atom emerged from the + port of the first analyzer, it was determined to have  $\mu_z = +\mu_B$ .” Nothing wrong here — this is what an analyzer does. “When that same atom emerged from the + port of the second analyzer, it was determined to have  $\mu_x = +\mu_B$ .” Ditto. “Now we know two projections of the magnetic moment.” This has got to be the problem. To underscore that problem, look at the figure below.



If an atom *did* have both  $\mu_z = +\mu_B$  and  $\mu_x = +\mu_B$ , then the projection on an axis rotated  $45^\circ$  from the vertical would be  $\mu_{45^\circ} = +\sqrt{2}\mu_B$ . But the Stern-Gerlach experiment assures us that whenever  $\mu_{45^\circ}$  is measured, the result is either  $+\mu_B$  or  $-\mu_B$ , and never  $+\sqrt{2}\mu_B$ . In summary, *it is not possible for a moment to have a projection on both the z axis and on the x axis*. Passing to the fourth sentence of analysis (I) — “When the atom enters the third analyzer, it still has  $\mu_z = +\mu_B$ , so it will emerge from the + port” — we immediately see the problem. The atom emerging from the + port of the second analyzer does *not* have  $\mu_z = +\mu_B$  — it doesn’t have a projection on the z axis at all.

Because it's easy to fall into misconceptions, let me emphasize what I'm saying and what I'm not saying:

I'm saying that if an atom has a value for  $\mu_x$ , then it doesn't have a value for  $\mu_z$ .

I'm not saying that the atom has a value for  $\mu_z$  but no one knows what it is.

I'm not saying that the atom has a value for  $\mu_z$  but that value is changing rapidly.

I'm not saying that the atom has a value for  $\mu_z$  but that value is changing unpredictably.

I'm not saying that a random half of such atoms have the value  $\mu_z = +\mu_B$  and the other half have the value  $\mu_z = -\mu_B$ .

I'm not saying that the atom has a value for  $\mu_z$  which will be disturbed upon measurement.

*The atom with a value for  $\mu_x$  does not have a value for  $\mu_z$  in the same way that love does not have a color.*

This is a new phenomenon, and it deserves a new name. That name is "indeterminacy". This is perhaps not the best name, because it might suggest, incorrectly, that an atom with a value for  $\mu_x$  has a value for  $\mu_z$  and we merely haven't yet determined what that value is. The English language was invented by people who didn't understand quantum mechanics, so it is not surprising that there are no perfectly appropriate names for quantum mechanical phenomena. This is a defect in our language, not a defect in quantum mechanics or in our understanding of quantum mechanics, and it is certainly not a defect in nature.<sup>5</sup>

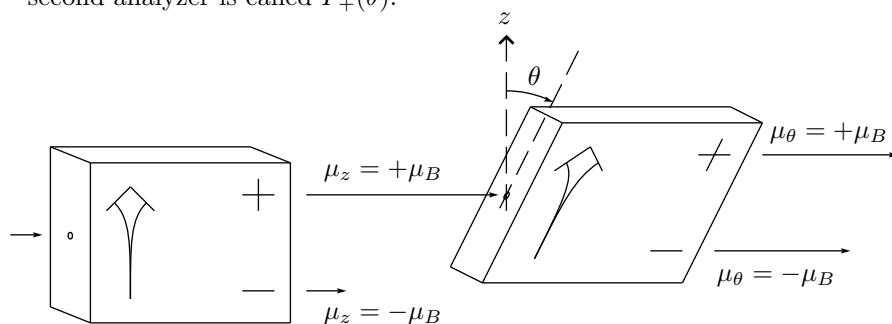
How can a vector have a projection on one axis but not on another? It is the job of the rest of this book to answer that question,<sup>6</sup> but one thing is clear already: The visualization of an atomic magnetic moment as a classical arrow must be wrong.

<sup>5</sup>In exactly the same manner, the name "orange" applies to light within the wavelength range 590–620 nm and the name "red" applies to light within the wavelength range 620–740 nm, but the English language has no word to distinguish the wavelength range 1590–1620 nm from the wavelength range 1620–1740 nm. This is not because optical light is "better" or "more deserving" than infrared light. It is due merely to the accident that our eyes detect optical light but not infrared light.

<sup>6</sup>Preview: In quantum mechanics, the magnetic moment is represented mathematically not by a vector but by a vector operator.

### 1.1.9 The upshot

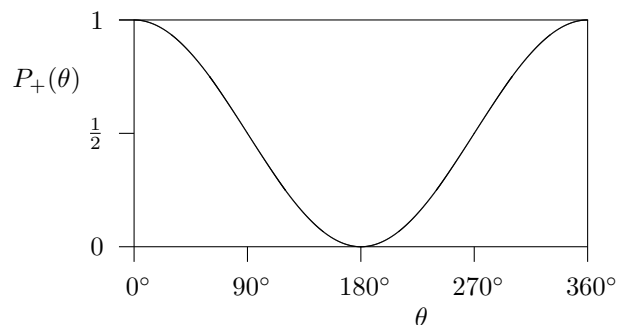
We escape from the conundrum of projections through probability. If an atom has  $\mu_z = +\mu_B$ , and if the projection on some other axis is measured, then the result cannot be predicted with certainty: we instead give probabilities for the various results. If the second analyzer is rotated by angle  $\theta$  relative to the vertical, the probability of emerging from the + port of the second analyzer is called  $P_+(\theta)$ .



We already know some special values: from section 1.1.3,  $P_+(0^\circ) = 1$ ; from section 1.1.5,  $P_+(90^\circ) = \frac{1}{2}$ ; from section 1.1.4,  $P_+(180^\circ) = 0$ ; from section 1.1.6,  $P_+(270^\circ) = \frac{1}{2}$ ; from section 1.1.3,  $P_+(360^\circ) = 1$ . It is not hard to guess the curve that interpolates between these values:

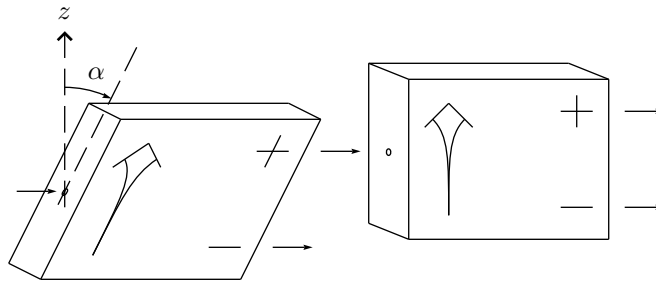
$$P_+(\theta) = \cos^2(\theta/2), \quad (1.5)$$

and experiment confirms this guess.

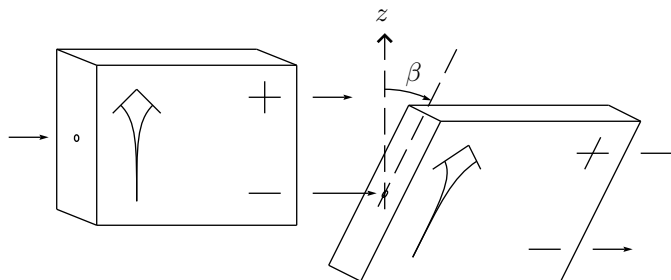


**Problems****1.2 Exit probabilities** (essential problem)

- a. An analyzer is tilted from the vertical by angle  $\alpha$ . An atom leaving its + port is channeled into a vertical analyzer. What is the probability that this atom emerges from the + port? The - port? (*Clue:* Use the “rotate as a unit” concept introduced in section 1.1.7.)

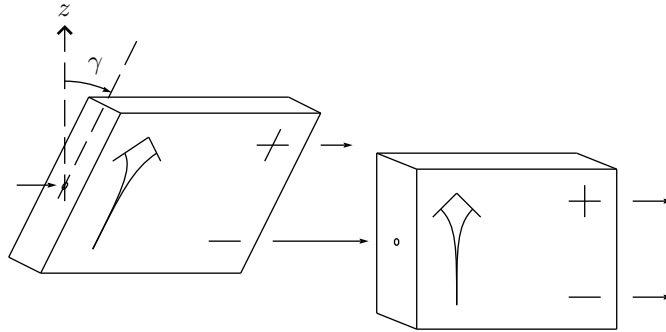


- b. An atom exiting the - port of a vertical analyzer behaves exactly like one exiting the + port of an upside-down analyzer (see section 1.1.4). Such an atom is channeled into an analyzer tilted from the vertical by angle  $\beta$ . What is the probability that this atom emerges from the + port? The - port?



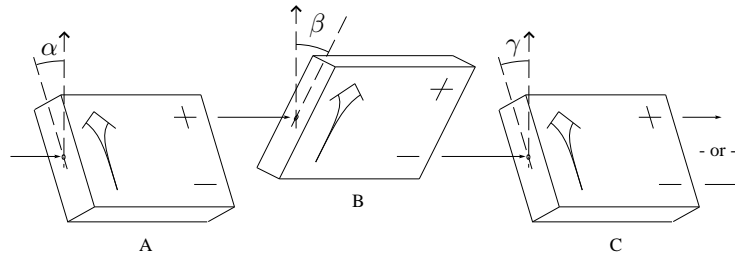
(Problem continues on next page.)

- c. An analyzer is tilted from the vertical by angle  $\gamma$ . An atom leaving its  $-$  port is channeled into a vertical analyzer. What is the probability that this atom emerges from the  $+$  port? The  $-$  port?



### 1.3 Multiple analyzers

An atom with  $\mu_z = +\mu_B$  is channeled through the following line of three Stern-Gerlach analyzers.



Find the probability that it emerges from (a) the  $-$  port of analyzer A; (b) the  $+$  port of analyzer B; (c) the  $+$  port of analyzer C; (d) the  $-$  port of analyzer C.

### 1.4 Properties of the $P_+(\theta)$ function

- An atom exits the  $+$  port of a vertical analyzer; that is, it has  $\mu_z = +\mu_B$ . Argue that the probability of this atom exiting from the  $-$  port of a  $\theta$  analyzer is the same as the probability of it exiting from the  $+$  port of a  $(180^\circ - \theta)$  analyzer.
- Conclude that the  $P_+(\theta)$  function introduced in section 1.1.9

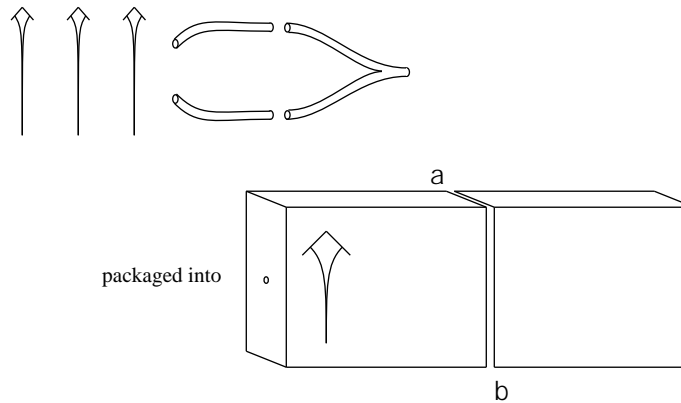
must satisfy

$$P_+(\theta) + P_+(180^\circ - \theta) = 1.$$

- c. Does the experimental result (1.5) satisfy this condition?

## 1.2 Interference

There are more quantum mechanical phenomena to uncover. To support our exploration, we build a new experimental device called the “analyzer loop”.<sup>7</sup> This is nothing but a Stern-Gerlach analyzer followed by “piping” that channels the two exit paths together again.<sup>8</sup>



The device must be constructed to high precision, so that there can be no way to distinguish whether the atom passed through by way of the top

<sup>7</sup>We build it in our minds. The experiments described in this section have never been performed exactly as described here, although researchers are getting close. [See Shimon Machluf, Yonathan Japha, and Ron Folman, “Coherent Stern–Gerlach momentum splitting on an atom chip” *Nature Communications* **4** (9 September 2013) 2424.] We know the results that would come from these experiments because conceptually parallel (but more complex!) experiments *have* been performed on photons, neutrons, atoms, and molecules.

<sup>8</sup>If you followed the footnote on page 12, you will recall that these “pipes” manipulate atoms through electromagnetic fields, not through touching. One way to make them would be to insert two more Stern-Gerlach apparatuses, the first one upside-down and the second one rightside-up relative to the initial apparatus. But whatever the manner of their construction, the pipes must *not* change the magnetic moment of an atom passing through them.

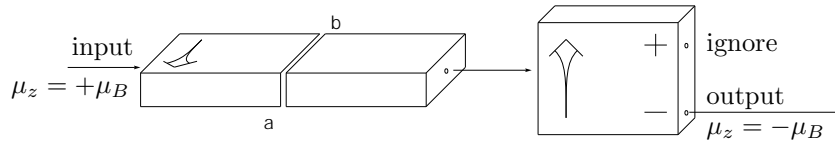


path or the bottom path. For example, the two paths must have the same length: If the top path were longer, then an atom going through via the top path would take more time, and hence there *would* be a way to tell which way the atom passed through the analyzer loop.

In fact, the analyzer loop is constructed so precisely that it doesn't change the character of the atom passing through it. If the atom enters with  $\mu_z = +\mu_B$ , it exits with  $\mu_z = +\mu_B$ . If it enters with  $\mu_x = -\mu_B$ , it exits with  $\mu_x = -\mu_B$ . If it enters with  $\mu_{17^\circ} = -\mu_B$ , it exits with  $\mu_{17^\circ} = -\mu_B$ . It is hard to see why anyone would want to build such a device, because they're expensive (due to the precision demands), and they do absolutely nothing!

Once you made one, however, you could convert it into something useful. For example, you could insert a piece of metal blocking path a. In that case, all the atoms exiting would have taken path b, so (if the analyzer loop were oriented vertically) all would emerge with  $\mu_z = -\mu_B$ .

Using the analyzer loop, we set up the following apparatus: First, channel atoms with  $\mu_z = +\mu_B$  into a horizontal analyzer loop.<sup>9</sup> Then, channel the atoms emerging from that analyzer loop into a vertical analyzer. Ignore atoms emerging from the + port of the vertical analyzer and look for atoms emerging from the - port.



We execute three experiments with this set-up: first we pass atoms through when path a is blocked, then when path b is blocked, finally when neither path is blocked.

### 1.2.1 Path a blocked

- (1) Atoms enter the analyzer loop with  $\mu_z = +\mu_B$ .
- (2) Half of them attempt path a, and end up impaled on the blockage.
- (3) The other half take path b, and emerge from the analyzer loop with  $\mu_x = -\mu_B$ .

---

<sup>9</sup>To make sure that all of these atoms have  $\mu_z = +\mu_B$ , they are harvested from the + port of a vertical analyzer.

- (4) Those atoms then enter the vertical analyzer. Similar to the result of section 1.1.7, half of these atoms emerge from the + port and are ignored. Half of them emerge from the – port and are counted.
- (5) The overall probability of passing through the set-up is  $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ .

If you perform this experiment, you will find that this analysis is correct and that these results are indeed obtained.

### 1.2.2 *Path b blocked*

- (1) Atoms enter the analyzer loop with  $\mu_z = +\mu_B$ .
- (2) Half of them attempt path **b**, and end up impaled on the blockage.
- (3) The other half take path **a**, and emerge from the analyzer loop with  $\mu_x = +\mu_B$ .
- (4) Those atoms then enter the vertical analyzer. Exactly as in section 1.1.7, half of these atoms emerge from the + port and are ignored. Half of them emerge from the – port and are counted.
- (5) The overall probability of passing through the set-up is  $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ .

Once again, experiment confirms these results.

### 1.2.3 *Neither path blocked*

Here, I have not just one, but two ways to analyze the experiment:

*Analysis I:*

- (1) An atom passes through the set-up either via path **b** or via path **a**.
- (2) From section 1.2.1, the probability of passing through via path **b** is  $\frac{1}{4}$ .
- (3) From section 1.2.2, the probability of passing through via path **a** is  $\frac{1}{4}$ .
- (4) Thus the probability of passing through the entire set-up is  $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ .

*Analysis II:*

- (1) Because “the analyzer loop is constructed so precisely that it doesn’t change the character of the atom passing through it”, the atom emerges from the analyzer loop with  $\mu_z = +\mu_B$ .
- (2) When such atoms enter the vertical analyzer, all of them emerge through the + port. (See section 1.1.3.)

(3) Thus the probability of passing through the entire set-up is zero.

These two analyses cannot both be correct. Experiment confirms the result of analysis II, but what could possibly be wrong with analysis I? Item (2) is already confirmed through the experiment of section 1.2.1, item (3) is already confirmed through the experiment of section 1.2.2, and don't tell me that I made a mistake in the arithmetic of item (4). The only thing left is item (1): "An atom passes through the set-up either via path b or via path a." This simple, appealing, common-sense statement must be *wrong!*

Just a moment ago, the analyzer loop seemed like a waste of money and skill. But in fact, a horizontal analyzer loop is an extremely clever way of correlating the path through the analyzer loop with the value of  $\mu_x$ : If the atom has  $\mu_x = +\mu_B$ , then it takes path a. If the atom has  $\mu_x = -\mu_B$ , then it takes path b. If the atom has  $\mu_z = +\mu_B$ , then it *doesn't have a value of  $\mu_x$*  and hence it *doesn't take a path*.

Notice again what I'm saying: I'm not saying the atom takes one path or the other but we don't know which. I'm not saying the atom breaks into two pieces and each half traverses its own path. I'm saying the atom doesn't take a path. The  $\mu_z = +\mu_B$  atoms within the horizontal analyzer loop *do not have a position* in the same sense that *love does not have a color*. If you think of an atom as a smaller, harder version of a classical marble, then you're visualizing the atom incorrectly.

Once again, our experiments have uncovered a phenomenon that doesn't happen in daily life, so there is no word for it in conventional language.<sup>10</sup> Sometimes people say that "the atom takes both paths", but that phrase does not really get to the heart of the new phenomenon. I have asked students to invent a new word to represent this new phenomenon, and my favorite of their many suggestions is "ambivate" — a combination of ambulate and ambivalent — as in "an atom with  $\mu_z = +\mu_B$  ambivates through both paths of a horizontal analyzer loop". While this is a great word, it hasn't caught on. The conventional name for this phenomenon is "quantal interference".

---

<sup>10</sup>In exactly the same way, there was no need for the word "latitude" or the word "longitude" when it was thought that the Earth was flat. The discovery of the near-spherical character of the Earth forced our forebears to invent new words to represent these new concepts. Words do not determine reality; instead reality determines which words are worth inventing.

The name “quantal interference” comes from a (far-fetched) analogy with interference in wave optics. Recall that in the two-slit interference of light, there are some observation points that have a light intensity if light passes through slit a alone, and the same intensity if light passes through slit b alone, but zero intensity if light passes through both slits. This is called “destructive interference”. There are other observation points that have a light intensity if the light passes through slit a alone, and the same intensity if light passes through slit b alone, but four times that intensity if light passes through both slits. This is called “constructive interference”. But in fact the word “interference” is a poor name for this phenomenon as well. It’s adapted from a football term, and football players never (or at least never intentionally) run “constructive interference”.

One last word about language: The device that I’ve called the “analyzer loop” is more conventionally called an “interferometer”. I didn’t use that name at first because that would have given away the ending.

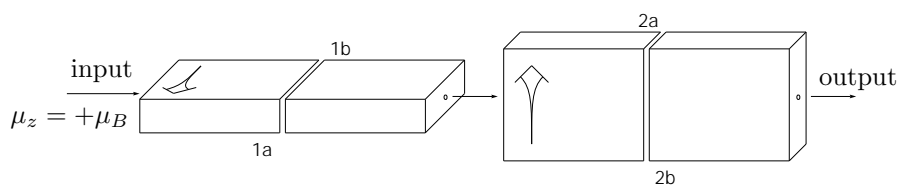
Back on page 6 I said that, to avoid unnecessary distraction, this chapter would “to the extent possible, do a quantum-mechanical treatment of an atom’s magnetic moment while maintaining a classical treatment of all other aspects — such as its energy and momentum and position”. You can see now why I put in that qualifier “to the extent possible”: we have found that within an interferometer, a quantum-mechanical treatment of magnetic moment demands a quantum-mechanical treatment of position as well.

#### 1.2.4 *Sample Problem: Constructive interference*

Consider the same set-up as on page 23, but now ignore atoms leaving the  $-$  port of the vertical analyzer and consider as output atoms leaving the  $+$  port. What is the probability of passing through the set-up when path a is blocked? When path b is blocked? When neither path is blocked?

*Solution:*  $\frac{1}{4}$ ;  $\frac{1}{4}$ ; 1. Because  $\frac{1}{4} + \frac{1}{4} < 1$ , this is an example of constructive interference.

---

**1.2.5 Sample Problem: Two analyzer loops**


Atoms with  $\mu_z = +\mu_B$  are channeled through a horizontal analyzer loop (number 1), then a vertical analyzer loop (number 2). If all paths are open, 100% of the incoming atoms exit from the output. What percentage of the incoming atoms leave from the output if the following paths are blocked?

- |        |               |
|--------|---------------|
| (a) 2a | (d) 1b        |
| (b) 2b | (e) 1b and 2a |
| (c) 1a | (f) 1a and 2b |

---

*Solution:* Only two principles are needed to solve this problem: First, an atom leaving an unblocked analyzer loop leaves in the same condition it had when it entered. Second, an atom leaving an analyzer loop with one path blocked leaves in the condition specified by the path that it took, regardless of the condition it had when it entered. Use of these principles gives the solution in the table on the next page. Notice that in changing from situation (a) to situation (e), you add blockage, yet you increase the output!

paths blocked	input condition	path taken through # 1	intermediate condition	path taken through # 2	output condition	probability of input $\rightarrow$ output
none	$\mu_z = +\mu_B$	"both"	$\mu_z = +\mu_B$	a	$\mu_z = +\mu_B$	100%
2a	$\mu_z = +\mu_B$	"both"	$\mu_z = +\mu_B$	100% blocked at a	none	0%
2b	$\mu_z = +\mu_B$	"both"	$\mu_z = +\mu_B$	a	$\mu_z = +\mu_B$	100%
1a	$\mu_z = +\mu_B$	50% blocked at a 50% pass through b	$\mu_x = -\mu_B$	"both"	$\mu_x = -\mu_B$	50%
1b	$\mu_z = +\mu_B$	50% pass through a 50% blocked at b	$\mu_x = +\mu_B$	"both"	$\mu_x = +\mu_B$	50%
1b and 2a	$\mu_z = +\mu_B$	50% pass through a 50% blocked at b	$\mu_x = +\mu_B$	25% blocked at a 25% pass through b	$\mu_z = -\mu_B$	25%
1a and 2b	$\mu_z = +\mu_B$	50% blocked at a 50% pass through b	$\mu_x = -\mu_B$	25% pass through a 25% blocked at b	$\mu_z = +\mu_B$	25%

### 1.2.6 Sample Problem: Find the flaw

No one would write a computer program and call it finished without testing and debugging their first attempt. Yet some approach physics problem solving in exactly this way: they get to the equation that is “the solution”, stop, and then head off to bed for some well-earned sleep without investigating whether the solution makes sense. This is a loss, because the real fun and interest in a problem comes not from our cleverness in finding “the solution”, but from uncovering what that solution tells us about nature. To give you experience in this reflection step, I’ve designed “find the flaw” problems in which you don’t find the solution, you only test it. Here’s an example.

#### Find the flaw: Tilted analyzer loop

Four students — Aldo, Beth, Celine, and Denzel — work problem 1.5 presented on the next page. All find the same answer for part (a), namely zero, but for parts (b) and (c) they produce four different answers! Their candidate answers are:

	(b)	(c)
Aldo	$\cos^4(\theta/2)$	$\sin^4(\theta/2)$
Beth	$\frac{1}{4} \sin(\theta)$	$\frac{1}{4} \sin(\theta)$
Celine	$\frac{1}{4} \sqrt{2} \sin(\theta/2)$	$\frac{1}{4} \sqrt{2} \sin(\theta/2)$
Denzel	$\frac{1}{2} \sin^2(\theta)$	$\frac{1}{2} \sin^2(\theta)$

Without actually working the problem, provide simple reasons showing that all of these candidates must be wrong.

*Solution:* For the special case  $\theta = 0^\circ$  the correct answers for (b) and (c) are both 0. Aldo’s answer to (b) fails this test.

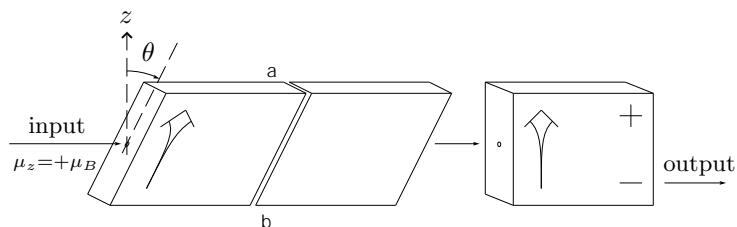
The special case  $\theta = 90^\circ$  was investigated in sections 1.2.1 and 1.2.2: in this case the answers for (b) and (c) are both  $\frac{1}{4}$ . Denzel’s answer fails this test.

Beth’s answer gives negative probabilities when  $180^\circ < \theta < 360^\circ$ . Bad idea!

The answer should not change when  $\theta$  increases by  $360^\circ$ . Celine's answer fails this test. (For example, it gives the answer  $+\frac{1}{4}$  when  $\theta = 90^\circ$  and  $-\frac{1}{4}$  when  $\theta = 450^\circ$ , despite the fact that  $90^\circ$  and  $450^\circ$  are the same angle.)

### Problems

#### 1.5 Tilted analyzer loop (recommended problem)

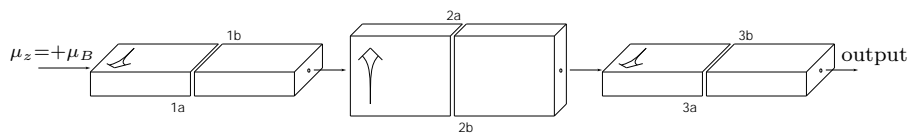


An atom with  $\mu_z = +\mu_B$  enters the analyzer loop (interferometer) shown above, tilted at angle  $\theta$  to the vertical. The outgoing atom enters a  $z$ -analyzer, and whatever comes out the  $-$  port is considered output. What is the probability for passage from input to output when:

- Paths **a** and **b** are both open?
- Path **b** is blocked?
- Path **a** is blocked?

#### 1.6 Three analyzer loops (recommended problem)

Atoms with  $\mu_z = +\mu_B$  are channeled into a horizontal analyzer loop, followed by a vertical analyzer loop, followed by a horizontal analyzer loop.



If all paths are open, 100% of the incoming atoms exit from the output. What percent of the incoming atoms leave from the output if the following paths are blocked?



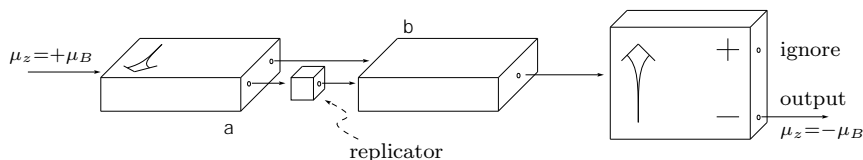
- |        |               |                      |
|--------|---------------|----------------------|
| (a) 3a | (d) 2b        | (g) 1b and 3b        |
| (b) 3b | (e) 1b        | (h) 1b and 3a        |
| (c) 2a | (f) 2a and 3b | (i) 1b and 3a and 2a |

(Note that in going from situation (h) to situation (i) you get *more* output from *increased* blockage.)

### 1.3 Aharonov-Bohm effect

We have seen how to sort atoms using a Stern-Gerlach analyzer, made of a non-uniform magnetic field. But how do atoms behave in a uniform magnetic field? In general, this is an elaborate question, (treated in section 5.4), and the answer will depend on the initial condition of the atom's magnetic moment, on the magnitude of the field, and on the amount of time that the atom spends in the field. But for one special case the answer, determined experimentally, is easy. If an atom is exposed to uniform magnetic field  $\vec{B}$  for exactly the right amount of time (which turns out to be a time of  $\pi\hbar/\mu_B B$ ), then the atom emerges with exactly the same magnetic condition it had initially: If it starts with  $\mu_z = -\mu_B$ , it ends with  $\mu_z = -\mu_B$ . If it starts with  $\mu_x = +\mu_B$ , it ends with  $\mu_x = +\mu_B$ . If it starts with  $\mu_{29^\circ} = +\mu_B$ , it ends with  $\mu_{29^\circ} = +\mu_B$ . Thus for atoms moving at a given speed, we can build a box containing a uniform magnetic field with just the right length so that any atom passing through it will spend just the right amount of time to emerge in the same condition it had when it entered. We call this box a “replicator”.

If you play with one of these boxes you'll find that you can build any elaborate set-up of sources, detectors, blockages, and analyzers, and that inserting a replicator into any path will not affect the outcome of any experiment. But notice that this apparatus list does not include interferometers (our “analyzer loops”)! Build the interference experiment of page 23. Do not block either path. Instead, slip a replicator into one of the two paths a or b — it doesn't matter which.



Without the replicator *no* atom emerges at output. But experiment shows that after inserting the replicator, *all* the atoms emerge at output.

How can this be? Didn't we just say of a replicator that "any atom passing through it will... emerge in the same condition it had when it entered"? Indeed we did, and indeed this is true. But an atom with  $\mu_z = +\mu_B$  doesn't pass through path a or path b — it ambivates through both paths.

If the atom *did* take one path or the other, then the replicator *would* have no effect on the experimental results. The fact that it *does* have an effect is proof that the atom *doesn't* take one path or the other.

The fact<sup>11</sup> that one can perform this remarkable experiment was predicted theoretically (in a different context) by Walter Franz. He announced his result in Danzig (now Gdańsk, Poland) in May 1939, just months before the Nazi invasion of Poland, and his prediction was largely forgotten in the resulting chaos. The effect was rediscovered theoretically by Werner Ehrenberg and Raymond Siday in 1949, but they published their result under the opaque title of "The refractive index in electron optics and the principles of dynamics" and their prediction was also largely forgotten. The effect was rediscovered theoretically a third time by Yakir Aharonov and David Bohm in 1959, and this time it sparked enormous interest, both experimental and theoretical. The phenomenon is called today the "Aharonov-Bohm effect".

### ***Problem***

#### 1.7 **Bomb-testing interferometer**<sup>12</sup> (recommended problem)

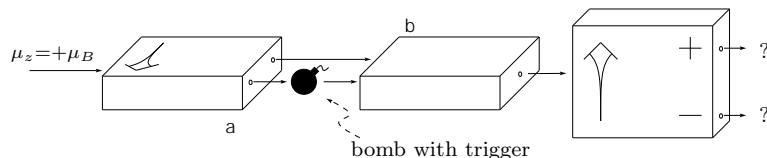
The Acme Bomb Company sells a bomb triggered by the presence of silver, and claims that the trigger is so sensitive that the bomb explodes when its trigger absorbs even a single silver atom. You have heard similar extravagant claims from other manufacturers, so you're suspicious. You purchase a dozen bombs, then shoot individual silver atoms at each in turn. The first bomb tested explodes! The trigger worked as advertised, but now it's useless because it's blasted to pieces. The second bomb tested doesn't explode — the atom slips through a hole in the trigger. This confirms your suspicion that not *all* the triggers are

<sup>11</sup>See B.J. Hiley, "The early history of the Aharonov-Bohm effect" (17 April 2013) <https://arxiv.org/abs/1304.4736>.

<sup>12</sup>Avshalom C. Elitzur and Lev Vaidman, "Quantum mechanical interaction-free measurements" *Foundations of Physics* **23** (July 1993) 987–997.

as sensitive as claimed, so this bomb is useless to you as well. If you continue testing in this fashion, at the end all your good bombs will be blown up and you will be left with a stash of bad bombs.

So instead, you set up the test apparatus sketched here:



An atom with  $\mu_z = +\mu_B$  enters the interferometer. If the bomb trigger has a hole, then the atom ambivates through both paths, arrives at the analyzer with  $\mu_z = +\mu_B$ , and exits the + port of the analyzer. But if the bomb trigger is good, then either (a) the atom takes path a and sets off the bomb, or else (b) the atom takes path b.

- If the bomb trigger is good, what is the probability of option (a)?  
Of option (b)?
- If option (b) happens, what kind of atom arrives at the analyzer?  
What is the probability of that atom exiting through the + port?  
The - port?

Conclusion: If the atom exits through the - port, then the bomb is good. If it exits through the + port then the bomb might be good or bad and further testing is required. But you *can* determine that the bomb trigger is good without blowing it up!

#### 1.4 Light on the atoms

Our conclusion that, under some circumstances, the atom “does not have a position” is so dramatically counterintuitive that you might — no, you *should* — be tempted to test it experimentally. Set up the interference experiment on page 23, but instead of simply allowing atoms to pass through the interferometer, watch to see which path the atom takes through the set-up. To watch them, you need light. So set up the apparatus with lamps trained on the two paths a and b.

Send in one atom. There’s a flash of light at path a.

Another atom. Flash of light at b.

Another atom. Flash at **b** again.

Then **a**, then **a**, then **b**.

You get the drift. Always the light appears at one path or the other. (In fact, the flashes come at random with probability  $\frac{1}{2}$  for a flash at **a** and  $\frac{1}{2}$  for a flash at **b**.) Never is there no flash. Never are there “two half flashes”. The atom always has a position when passing through the interferometer. “So much”, say the skeptics, “for this metaphysical nonsense about ‘the atom takes both paths’.”

But wait. Go back and look at the output of the vertical analyzer. When we ran the experiment with no light, the probability of coming out the  $-$  port was 0. When we turn the lamps on, then the probability of coming out the  $-$  port becomes  $\frac{1}{2}$ .

When the lamps are off, analysis II on page 24 is correct: the atoms ambivate through both paths, and the probability of exiting from the  $-$  port is 0. When the lamps are on and a flash is seen at path **a**, then the atom *does* take path **a**, and now the analysis of section 1.2.2 on page 24 is correct: the probability of exiting from the  $-$  port is  $\frac{1}{2}$ .

The process when the lamps are on is called “observation” or “measurement”, and a lot of nonsense has come from the use of these two words. The important thing is whether the light is present or absent. Whether or not the flashes are “observed” by a person is irrelevant. To prove this to yourself, you may, instead of observing the flashes in person, record the flashes on video. If the lamps are on, the probability of exiting from the  $-$  port is  $\frac{1}{2}$ . If the lamps are off, the probability of exiting from the  $-$  port is 0. Now, after the experiment is performed, you may either destroy the video, or play it back to a human audience, or play it back to a feline audience. Surely, by this point it is too late to change the results at the exit port.

It’s not just light. Any method you can dream up for determining the path taken will show that the atom takes just one path, but that method will also change the output probability from 0 to  $\frac{1}{2}$ . No person needs to actually read the results of this mechanism: as long as the mechanism is at work, as long as it is in principle possible to determine which path is taken, then one path is taken and no interference happens.

What happens if you train a lamp on path **a** but leave path **b** in the dark? In this case a flash means the atom has taken path **a**. No flash means

the atom has taken path **b**. In both cases the probability of passage for the atom is  $\frac{1}{2}$ .

How can the atom taking path **b** “know” that the lamp at path **a** is turned on? The atom initially “sniffs out” *both* paths, like a fog creeping down two passageways. The atom that eventually does take path **b** in the dark started out attempting both paths, and that’s how it “knows” the lamp at path **a** is on. This is called the “Renninger negative-result experiment”.

It is not surprising that the presence or absence of light should affect an atom’s motion: this happens even in classical mechanics. When an object absorbs or reflects light, that object experiences a force, so its motion is altered. For example, a baseball tossed upward in a gymnasium with the overhead lamps off attains a slightly greater height than an identical baseball experiencing an identical toss in the same gymnasium with the overhead lamps on, because the downward-directed light beams push the baseball downward. (This is the same “radiation pressure” that is responsible for the tails of comets. And of course, the effect occurs whenever the lamps are turned on: whether any person actually watches the illuminated baseball is irrelevant.) This effect is negligible for typical human-scale baseballs and tosses and lamps, but atoms are far smaller than baseballs and it is reasonable that the light should alter the motion of an atom more than it alters the motion of a baseball.

One last experiment: Look for the atoms with dim light. In this case, some of the atoms will pass through with a flash. But — because of the dimness — some atoms will pass through without any flash at all. For those atoms passing through with a flash, the probability for exiting the – port is  $\frac{1}{2}$ . For those atoms passing through without a flash, the probability of exiting the – port is 0.

## 1.5 Entanglement

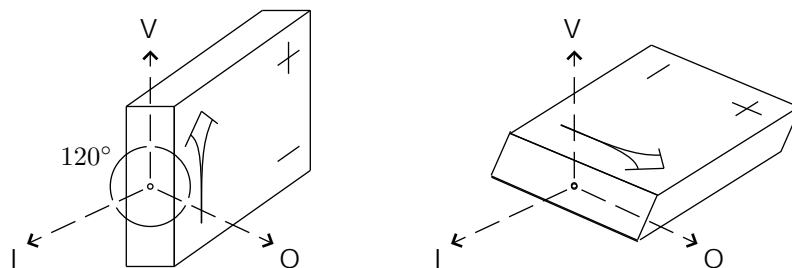
I have claimed that an atom with  $\mu_z = +\mu_B$  *doesn't have* a value of  $\mu_x$ , and that when such an atom passes through a horizontal interferometer, it *doesn't have* a position. You might say to yourself, “These claims are so weird, so far from common sense, that I just can’t accept them. I believe the atom *does* have a value of  $\mu_x$  and *does* have a position, but something else very complicated is going on to make the atom *appear* to lack a  $\mu_x$  and

a position. I don't know what that complicated thing is, but just because I haven't yet thought it up yet doesn't mean that it doesn't exist."

If you think this, you're in good company: Einstein<sup>13</sup> thought it too. This section introduces a new phenomenon of quantum mechanics, and shows that no local deterministic mechanism, no matter how complex or how fantastic, can give rise to all the results of quantum mechanics. Einstein was wrong.

### 1.5.1 *Flipping Stern-Gerlach analyzer*

A new piece of apparatus helps us uncover this new phenomenon of nature. Mount a Stern-Gerlach analyzer on a stand so that it can be oriented either vertically ( $0^\circ$ ), or tilted one-third of a circle clockwise ( $+120^\circ$ ), or tilted one-third of a circle counterclockwise ( $-120^\circ$ ). Call these three orientations V (for vertical), O (for out of the page), or I (for into the page). As an atom approaches the analyzer, select one of these three orientations at random, flip the analyzer to that orientation, and allow the atom to pass through as usual. As a new atom approaches, again select an orientation at random, flip the analyzer, and let the atom pass through. Repeat many times.



*Flipping Stern-Gerlach analyzer. The arrows V, O, and I, oriented  $120^\circ$  apart, all lie within the plane perpendicular to the atom's approach path.*

<sup>13</sup>Although Albert Einstein (1879–1955) is most famous for his work on relativity, he claimed that he had “thought a hundred times as much about the quantum problems as I have about general relativity theory.” (Remark to Otto Stern, reported in Abraham Pais, “Subtle is the Lord...”: *The Science and the Life of Albert Einstein*, [Oxford University Press, Oxford, UK, 1982] page 9.)

What happens if an atom with  $\mu_z = +\mu_B$  enters a flipping analyzer? With probability  $\frac{1}{3}$ , the atom enters a vertical analyzer (orientation V), and in that case it exits the + port with probability 1. With probability  $\frac{1}{3}$ , the atom enters an out-of-the-page analyzer (orientation O), and in that case (see equation 1.5) it exits the + port with probability

$$\cos^2(120^\circ/2) = \frac{1}{4}.$$

With probability  $\frac{1}{3}$ , the atom enter an into-the-page analyzer (orientation I), and in that case it exits the + port with probability  $\frac{1}{4}$ . Thus the overall probability of this atom exiting through the + port is

$$\frac{1}{3} \times 1 + \frac{1}{3} \times \frac{1}{4} + \frac{1}{3} \times \frac{1}{4} = \frac{1}{2}. \quad (1.6)$$

A similar analysis shows that if an atom with  $\mu_z = -\mu_B$  enters the flipping analyzer, it exits the + port with probability  $\frac{1}{2}$ .

You could repeat the analysis for an atom entering with  $\mu_{(+120^\circ)} = +\mu_B$ , but you don't need to. Because the three orientations are exactly one-third of a circle apart, rotational symmetry demands that an atom entering with  $\mu_{(+120^\circ)} = +\mu_B$  behaves exactly as an atom entering with  $\mu_z = +\mu_B$ .

In conclusion, an atom entering in any of the six conditions  $\mu_z = +\mu_B$ ,  $\mu_z = -\mu_B$ ,  $\mu_{(+120^\circ)} = +\mu_B$ ,  $\mu_{(+120^\circ)} = -\mu_B$ ,  $\mu_{(-120^\circ)} = +\mu_B$ , or  $\mu_{(-120^\circ)} = -\mu_B$  will exit through the + port with probability  $\frac{1}{2}$ .

### 1.5.2 EPR source of atom pairs

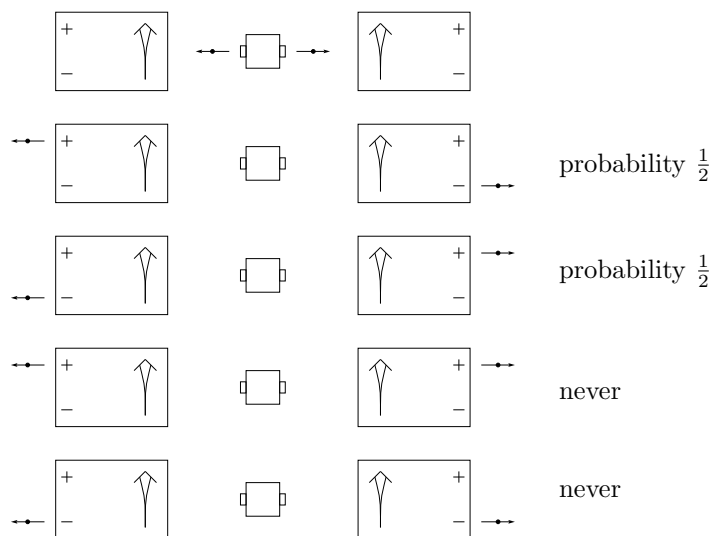
Up to now, our atoms have come from an oven. For the next experiments we need a special source<sup>14</sup> that expels two atoms at once, one moving to the left and the other to the right. For the time being we call this an “EPR” source, which produces an atomic pair in an “EPR” condition. The letters come from the names of those who discovered this condition: Albert Einstein, Boris Podolsky, and Nathan Rosen. After investigating this condition we will develop a more descriptive name.

---

<sup>14</sup>The question of how to build this special source need not concern us at the moment: it is an experimental fact that such sources do exist. One way to make one would start with a diatomic molecule with zero magnetic moment. Cause the molecule to disintegrate and eject the two daughter atoms in opposite directions. Because the initial molecule had zero magnetic moment, the pair of daughter atoms will have the properties of magnetic moment described. In fact, it's easier to build a source, not for a pair of atoms, but for a pair of photons using a process called spontaneous parametric down-conversion.

The following experiments investigate the EPR condition:

(1) Each atom encounters a vertical Stern-Gerlach analyzer. The experimental result: *the two atoms exit through opposite ports*. To be precise: with probability  $\frac{1}{2}$ , the left atom exits + and the right atom exits -, and with probability  $\frac{1}{2}$ , the left atom exits - and the right atom exits +, but it never happens that both atoms exit + or that both atoms exit -.



You might suppose that this is because for half the pairs, the left atom is generated with  $\mu_z = +\mu_B$  while the right atom is generated with  $\mu_z = -\mu_B$ , while for the other half of the pairs, the left atom is generated with  $\mu_z = -\mu_B$  while the right atom is generated with  $\mu_z = +\mu_B$ . This supposition seems suspicious, because it singles out the  $z$  axis as special, but at this stage in our experimentation it's possible.

(2) Repeat the above experiment with horizontal Stern-Gerlach analyzers. The experimental result: Exactly the same as in experiment (1)! The two atoms always exit through opposite ports.

Problem 1.9 on page 47 demonstrates that the results of this experiment rule out the supposition presented at the bottom of experiment (1).

(3) Repeat the above experiment with the two Stern-Gerlach analyzers oriented at  $+120^\circ$ , or with both oriented at  $-120^\circ$ , or with both oriented



at  $57^\circ$ , or for any other angle, as long as both have the same orientation. The experimental result: Exactly the same for any orientation!

(4) In an attempt to trick the atoms, we set the analyzers to vertical, then launch the pair of atoms, then (while the atoms are in flight) switch both analyzers to, say,  $42^\circ$ , and have the atoms encounter these analyzers both with switched orientation. The experimental result: Regardless of what the orientation is, and regardless of when that orientation is set, the two atoms always exit through opposite ports.

Here is one way to picture this situation: The pair of atoms has a total magnetic moment of zero. But whenever the projection of a single atom on any axis is measured, the result must be  $+\mu_B$  or  $-\mu_B$ , never zero. The only way to insure that that total magnetic moment, projected on any axis, sums to zero is the way described above. Do not put too much weight on this picture: like the “wants to go straight” story of section 1.1.5 (page 14), this is a classical story that happens to give the correct result. The definitive answer to any question is always experiment, not any picture or story, however appealing it may be.

These four experiments show that it is impossible to describe the condition of the atoms through anything like “the left atom has  $\mu_z = +\mu_B$ , the right atom has  $\mu_z = -\mu_B$ ”. How *can* we describe the condition of the pair? This will require further experimentation. For now, we say it has an EPR condition.

### 1.5.3 *EPR atom pair encounters flipping Stern-Gerlach analyzers*

A pair of atoms leaves the EPR source, and each atom travels at the same speed to vertical analyzers located 100 meters away. The left atom exits the  $-$  port, the right atom exits the  $+$  port. When the pair is flying from source to analyzer, it’s not correct to describe it as “the left atom has  $\mu_z = -\mu_B$ , the right atom has  $\mu_z = +\mu_B$ ”, but after the atoms leave their analyzers, then this is a correct description.

Now shift the left analyzer one meter closer to the source. The left atom encounters its analyzer before the right atom encounters its. Suppose the left atom exits the  $-$  port, while the right atom is still in flight toward its analyzer. We know that when the right atom eventually does encounter its vertical analyzer, it will exit the  $+$  port. Thus it is correct to describe

the right atom as having “ $\mu_z = +\mu_B$ ”, even though that atom hasn’t yet encountered its analyzer.

Replace the right vertical analyzer with a flipping Stern-Gerlach analyzer. (In the figure below, it is in orientation O, out of the page.) Suppose the left atom encounters its vertical analyzer and exits the  $-$  port. Through the reasoning of the previous paragraph, the right atom now has  $\mu_z = +\mu_B$ . We know that when such an atom encounters a flipping Stern-Gerlach analyzer, it exits the  $+$  port with probability  $\frac{1}{2}$ .



Similarly, if the left atom encounters its vertical analyzer and exits the  $+$  port, the right atom now has  $\mu_z = -\mu_B$ , and once it arrives at its flipping analyzer, it will exit the  $-$  port with probability  $\frac{1}{2}$ . Summarizing these two paragraphs: Regardless of which port the left atom exits, the right atom will exit the opposite port with probability  $\frac{1}{2}$ .

Now suppose that the left analyzer were not vertical, but instead in orientation I, tilted into the page by one-third of a circle. It’s easy to see that, again, regardless of which port the left atom exits, the right atom will exit the opposite port with probability  $\frac{1}{2}$ .

Finally, suppose that the left analyzer is a flipping analyzer. Once again, the two atoms will exit from opposite ports with probability  $\frac{1}{2}$ .

The above analysis supposed that the left analyzer was one meter closer to the source than the right analyzer, but clearly it also works if the right analyzer is one meter closer to the source than the left analyzer. Or one centimeter. One suspects that the same result will hold even if the two analyzers are exactly equidistant from the source, and experiment bears out this suspicion.

In summary: Each atom from this EPR source enters a flipping Stern-Gerlach analyzer.

- (A) The atoms exit from opposite ports with probability  $\frac{1}{2}$ .
- (B) If the two analyzers happen to have the same orientation, the atoms exit from opposite ports.

This is the prediction of quantum mechanics, and experiment confirms this prediction.

#### 1.5.4 *The prediction of local determinism*

Suppose you didn't know anything about quantum mechanics, and you were told the result that "if the two analyzers have the same orientation, the atoms exit from opposite ports." Could you explain it?

I am sure you could. In fact, there are two possible explanations: First, the communication explanation. The left atom enters its vertical analyzer, and notices that it's being pulled toward the + port. It calls up the right atom with its walkie-talkie and says "If your analyzer has orientation I or O then you might go either way, but if your analyzer has orientation V you've got to go to the - port!" This is a possible explanation, but it's not a *local* explanation. The two analyzers might be 200 meters apart, or they might be 200 light-years apart. In either case, the message would have to get from the left analyzer to the right analyzer instantaneously. The walkie-talkies would have to use not radio waves, which propagate at the speed of light, but some sort of not-yet-discovered "insta-rays". Physicists have always been skeptical of non-local explanations, and since the advent of relativity they have grown even more skeptical, so we set this explanation aside. Can you find a local explanation?

Again, I am sure you can. Suppose that when the atoms are launched, they have some sort of characteristic that specifies which exit port they will take when they arrive at their analyzer. This very reasonable supposition, called "determinism", pervades all of classical mechanics. It is similar to saying "If I stand atop a 131 meter cliff and toss a ball horizontally with speed 23.3 m/s, I can predict the angle with which the ball strikes the ground, even though that event will happen far away and long in the future." In the case of the ball, the resulting strike angle is encoded into the initial position and velocity. In the case of the atoms, it's not clear how the

exit port will be encoded: perhaps through the orientation of its magnetic moment, perhaps in some other, more elaborate way. But the method of encoding is irrelevant: if local determinism holds, then something within the atom determines which exit port it will take when it reaches its analyzer.<sup>15</sup> I'll represent this "something" through a code like  $(+ + -)$ . The first symbol means that if the atom encounters an analyzer in orientation  $V$ , it will exit through the  $+$  port. The second means that if it encounters an analyzer in orientation  $O$ , it will exit through the  $+$  port. The third means that if it encounters an analyzer in orientation  $I$ , it will exit through the  $-$  port. The only way to ensure that "if the two analyzers have the same orientation, the atoms exit from opposite ports" is to assume that when the two atoms separate from each other within the source, they have opposite codes. If the left atom has  $(+ - +)$ , the right atom must have  $(- + -)$ . If the left atom has  $(- - -)$ , the right atom must have  $(+ + +)$ . This is the local deterministic scheme for explaining fact (B) that "if the two analyzers have the same orientation, the atoms exit from opposite ports".

But can this scheme explain fact (A)? Let's investigate. Consider first the case mentioned above: the left atom has  $(+ - +)$  and the right atom has  $(- + -)$ . These atoms will encounter analyzers set to any of  $3^2 = 9$  possible pairs of orientations. We list them below, along with with exit ports taken by the atoms. (For example, the third line of the table considers a left analyzer in orientation  $V$  and a right analyzer in orientation  $I$ . The left atom has code  $(+ - +)$ , and the first entry in that code determines that the left atom will exit from the  $V$  analyzer through the  $+$  port. The right atom has code  $(- + -)$ , and the third entry in that code determines that the right atom will exit from the  $I$  analyzer through the  $-$  port.)

---

<sup>15</sup>But remember that in quantum mechanics determinism does not hold. The information *can't* be encoded within the three projections of a classical magnetic moment vector, because at any one instant, the quantum magnetic moment vector has only one projection.

left port	left analyzer	right analyzer	right port	opposite?
+	V	V	-	yes
+	V	O	+	no
+	V	I	-	yes
-	O	V	-	no
-	O	O	+	yes
-	O	I	-	no
+	I	V	-	yes
+	I	O	+	no
+	I	I	-	yes

Each of the nine orientation pairs (VV, OI, etc.) are equally likely, five of the orientation pairs result in atoms exiting from opposite ports, so when atoms of this type emerge from the source, the probability of these atoms exiting from opposite ports is  $\frac{5}{9}$ .

What about a pair of atoms generated with different codes? Suppose the left atom has  $(--+)$  so the right atom must have  $(++-)$ . If you perform the analysis again, you will find that the probability of atoms exiting from opposite ports is once again  $\frac{5}{9}$ .

Suppose the left atom has  $(---)$ , so the right atom must have  $(+++)$ . The probability of the atoms exiting from opposite ports is of course 1.

There are, in fact, just  $2^3 = 8$  possible codes:

code for left atom	probability of exiting opposite
+++	1
--+	$\frac{5}{9}$
+ - +	$\frac{5}{9}$
++-	$\frac{5}{9}$
+ --	$\frac{5}{9}$
- + -	$\frac{5}{9}$
--+	$\frac{5}{9}$
---	1

If the source makes left atoms of only type  $(- - +)$ , then the probability of atoms exiting from opposite ports is  $\frac{5}{9}$ . If the source makes left atoms of only type  $(+ + +)$ , then the probability of atoms exiting from opposite ports is 1. If the source makes left atoms of type  $(- - +)$  half the time, and of type  $(+ + +)$  half the time, then the probability of atoms exiting from opposite ports is halfway between  $\frac{5}{9}$  and 1, namely  $\frac{7}{9}$ . But no matter how the source makes atoms, the probability of atoms exiting from opposite ports must be somewhere between  $\frac{5}{9}$  and 1.

But experiment and quantum mechanics agree: That probability is actually  $\frac{1}{2}$  — and  $\frac{1}{2}$  is *not* between  $\frac{5}{9}$  and 1. No local deterministic scheme — no matter how clever, or how elaborate, or how baroque — can give the result  $\frac{1}{2}$ . There is no “something within the atom that determines which exit port it will take when it reaches its analyzer”. If the magnetic moment has a projection on axis V, then it *doesn't have* a projection on axis O or axis I.

There is a reason that Einstein, despite his many attempts, never produced a scheme that explained quantum mechanics in terms of some more fundamental, local and deterministic mechanism. It is not that Einstein wasn't clever. It is that *no such scheme exists*.

### 1.5.5 *The upshot*

This is a new phenomenon — one totally absent from classical physics — so it deserves a new name, something more descriptive than “EPR”. Einstein called it “spooky action at a distance”.<sup>16</sup> The phenomenon is spooky all right, but this phrase misses the central point that the phenomenon involves “correlations at a distance”, whereas the word “action” suggests “cause-and-effect at a distance”. Schrödinger coined the term “entanglement” for this phenomenon and said it was “not... *one* but rather *the* characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought”.<sup>17</sup> The world has followed Schrödinger and the phenomenon is today called entanglement. We will later investigate entanglement in more detail, but for now we will just call our EPR source a

<sup>16</sup>Letter from Einstein to Max Born, 3 March 1947, *The Born-Einstein Letters* (Macmillan, New York, 1971) translated by Irene Born.

<sup>17</sup>Erwin Schrödinger, “Discussion of probability relations between separated systems” *Mathematical Proceedings of the Cambridge Philosophical Society* **31** (October 1935) 555–563.

“source of entangled atom pairs” and describe the condition of the atom pair as “entangled”.

The failure of local determinism described above is a special case of “Bell’s Theorem”, developed by John Bell<sup>18</sup> in 1964. The theorem has by now been tested experimentally numerous times in numerous contexts (various different angles; various distances between the analyzers; various sources of entangled pairs; various kinds of particles flying apart — gamma rays, or optical photons, or ions). In every test, quantum mechanics has been shown correct and local determinism wrong. What do we gain from these results?

*First*, they show that nature does not obey local determinism. To our minds, local determinism is common sense and any departure from it is weird. Thus whatever theory of quantum mechanics we eventually develop will be, to our eyes, weird. This will be a strength, not a defect, in the theory. The weirdness lies in nature, not in the theory used to describe nature.

Each of us feels a strong psychological tendency to reject the unfamiliar. In 1633, the Holy Office of the Inquisition found Galileo Galilei’s idea that the Earth orbited the Sun so unfamiliar that they rejected it. The inquisitors put Galileo on trial and forced him to abjure his position. From the point of view of nature, the trial was irrelevant, Galileo’s abjuration was irrelevant: the Earth orbits the Sun whether the Holy Office finds that fact comforting or not. It is our job as scientists to change our minds to fit nature; we do not change nature to fit our preconceptions. Don’t make the inquisitors’ mistake.

*Second*, the Bell’s theorem result guides not just our calculations about nature but also our visualizations of nature, and even the very idea of what it means to “understand” nature. Lord Kelvin<sup>19</sup> framed the situation perfectly in his 1884 Baltimore lectures: “I never satisfy myself until I can

<sup>18</sup>John Stewart Bell (1928–1990), a Northern Irish physicist, worked principally in accelerator design, and his investigation of the foundations of quantum mechanics was something of a hobby. Concerning tests of his theorem, he remarked that “The reasonable thing just doesn’t work.” [Jeremy Bernstein, *Quantum Profiles* (Princeton University Press, Princeton, NJ, 1991) page 84.]

<sup>19</sup>William Thomson, the first Baron Kelvin (1824–1907), was an Irish mathematical physicist and engineer who worked in Scotland. He is best known today for establishing the thermodynamic temperature scale that bears his name, but he also made fundamental contributions to electromagnetism. He was knighted for his engineering work on the first transatlantic telegraph cable.

make a mechanical model of a thing. If I can make a mechanical model I can understand it. As long as I cannot make a mechanical model all the way through I cannot understand, and this is why I cannot get the electromagnetic theory.”<sup>20</sup> If we take this as our meaning of “understand”, then the experimental tests of Bell’s theorem assure us that we will never be able to understand quantum mechanics.<sup>21</sup> What is to be done about this? There are only two choices. Either we can give up on understanding, or we can develop a new and more appropriate meaning for “understanding”.

Max Born<sup>22</sup> argued for the first choice: “The ultimate origin of the difficulty lies in the fact (or philosophical principle) that we are compelled to use the words of common language when we wish to describe a phenomenon, not by logical or mathematical analysis, but by a picture appealing to the imagination. Common language has grown by everyday experience and can never surpass these limits.”<sup>23</sup> Born felt that it was impossible to visualize or “understand” quantum mechanics: all you could do was grind through the “mathematical analysis”.

Humans are visual animals, however, and I have found that when we are told not to visualize, we do so anyway. But we do so in an illicit and uncritical way. For example, many people visualize an atom passing through an interferometer as a small, hard, marble, with a definite position, despite the already-discovered fact that this visualization is untenable. Many people visualize a photon as a “ball of light” despite the fact that a photon (as conventionally defined) has a definite energy and hence can never have a position.

It *is* possible to develop a visualization and understanding of quantum mechanics. This *can’t* be done by building a “mechanical model all the way through”. It must be done through *both* analogy and contrast: atoms

<sup>20</sup>William Thomson, “Baltimore lectures on wave theory and molecular dynamics,” in Robert Kargon and Peter Achinstein, editors, *Kelvin’s Baltimore Lectures and Modern Theoretical Physics* (MIT Press, Cambridge, MA, 1987) page 206.

<sup>21</sup>The first time I studied quantum mechanics seriously, I wrote in the margin of my textbook “Good God they do it! But how?” I see now that I was looking for a mechanical mechanism undergirding quantum mechanics. It doesn’t exist, but it’s very natural for anyone to *want* it to exist.

<sup>22</sup>Max Born (1882–1970) was a German-Jewish theoretical physicist with a particular interest in optics. At the University of Göttingen in 1925 he directed Heisenberg’s research which resulted in the first formulation of quantum mechanics. His granddaughter, the British-born Australian actress and singer Olivia Newton-John, is famous for her 1981 hit song “Physical”.

<sup>23</sup>Max Born, *Atomic Physics*, sixth edition (Hafner Press, New York, 1957) page 97.



behave in *some* ways like small hard marbles, in *some* ways like classical waves, and in *some* ways like a cloud or fog of probability. Atoms don't behave *exactly* like any of these things, but if you keep in mind both the analogy and its limitations, then you can develop a pretty good visualization and understanding.

And that brings us back to the name “entanglement”. It's an important name for an important phenomenon, but it suggests that the two distant atoms are connected mechanically, through strings. They aren't. The two atoms are correlated — if the left comes out +, the right comes out –, and vice versa — but they aren't correlated because of some signal sent back and forth through either strings or walkie-talkies. Entanglement involves correlation without causality.

### **Problems**

#### 1.8 **An atom walks into an analyzer**

Execute the “similar analysis” mentioned in the sentence below equation (1.6).

#### 1.9 **A supposition squashed** (essential problem)

If atoms were generated according to the supposition presented below experiment (1) on page 38, then would would happen when they encountered the two horizontal analyzers of experiment (2)?

#### 1.10 **A probability found through local determinism**

Suppose that the codes postulated on page 42 did exist. Suppose also that a given source produces the various possible codes with these probabilities:

code for left atom	probability of making such a pair
+++	1/2
++-	1/4
+--	1/8
--+	1/8

If this given source were used in the experiment of section 1.5.3 with distant flipping Stern-Gerlach analyzers, what would be the probability of the two atoms exiting from opposite ports?

### 1.11 A probability found through quantum mechanics

In the test of Bell's inequality (the experiment of section 1.5.3), what is the probability given by quantum mechanics that, if the orientation settings are different, the two atoms exit from opposite ports?

## 1.6 Quantum cryptography

We've seen a lot of new phenomena, and the rest of this book is devoted to filling out our understanding of these phenomena and applying that understanding to various circumstances. But first, can we use them for anything?

We can. The sending of coded messages used to be the province of armies and spies and giant corporations, but today everyone does it. All transactions through automatic teller machines are coded. All Internet commerce is coded. This section describes a particular, highly reliable encoding scheme and then shows how quantal entanglement may someday be used to implement this scheme. (Quantum cryptography was used to securely transmit voting ballots cast in the Geneva canton of Switzerland during parliamentary elections held 21 October 2007. But it is not today in regular use anywhere.)

In this section I use names conventional in the field of coded messages (called cryptography). Alice and Bob wish to exchange private messages, but they know that Eve is eavesdropping on their communication. How can they encode their messages to maintain their privacy?

### 1.6.1 *The Vernam cipher*

The Vernam cipher or "one-time pad" technique is the only coding scheme proven to be absolutely unbreakable (if used correctly). It does not *rely* on the use of computers — it was invented by Gilbert Vernam in 1919 — but today it is mostly implemented using computers, so I'll describe it in that context.

Data are stored on computer disks through a series of magnetic patches on the disk that are magnetized either "up" or "down". An "up" patch is taken to represent 1, and a "down" patch is taken to represent 0. A string of seven patches is used to represent a character. For example, by a

convention called ASCII, the letter “a” is represented through the sequence 1100001 (or, in terms of magnetizations, up, up, down, down, down, down, up). The letter “W” is represented through the sequence 1010111. Any computer the world around will represent the message “What?” through the sequence

1010111 1101000 1100001 1110100 0111111

This sequence is called the “plaintext”.

But Alice doesn’t want a message recognizable by any computer the world around. She wants to send the message “What?” to Bob in such a way that Eve will *not* be able to read the message, even though Eve has eavesdropped on the message. Here is the scheme invented by Vernam: Before sending her message, Alice generates a string of random 0s and 1s just as long as the message she wants to send — in this case,  $7 \times 5 = 35$  bits. She might do this by flipping 35 coins, or by flipping one coin 35 times. I’ve just done that, producing the random number

0100110 0110011 1010110 1001100 1011100

Then Alice gives Bob a copy of that random number – the “key”.

Instead of sending the plaintext, Alice modifies her plaintext into a coded “ciphertext” using the key. She writes down her plaintext and writes the key below it, then works through column by column. For each position, if the key is 0 the plaintext is left unchanged; but if the key is 1 the plaintext is reversed (from 0 to 1 or vice versa). For the first column, the key is 0, so Alice *doesn’t* change the plaintext: the first character of ciphertext is the same as the first character of plaintext. For the second column, the key is 1, so Alice *does* change the plaintext: the second character of ciphertext is the reverse of the second character of plaintext. Alice goes through all the columns, duplicating the plaintext where the key is 0 and reversing the plaintext where the key is 1.

plaintext:	1010111 1101000 1100001 1110100 0111111
key:	0100110 0110011 1010110 1001100 1011100
ciphertext:	1110001 1011011 0110111 0111000 1100011

Then, Alice sends out her ciphertext over open communication lines.

Now, the ciphertext that Bob (and Eve) receive translates to *some* message through the ASCII convention – in fact, it translates to “q[78c” — but because the key is random, the ciphertext is just as random. Bob decipheres Alice’s message by carrying out the encoding process on the ciphertext, namely, duplicating the ciphertext where the key is 0 and reversing the ciphertext where the key is 1. The result is the plaintext. Eve does not know the key, so she cannot produce the plaintext.

The whole scheme relies on the facts that the key is (1) random and (2) unknown to Eve. The very name “one-time pad” underscores that a key can only be used once and must then be discarded. If a single key is used for two messages, then the second key is not “random” — it is instead perfectly correlated with the first key. There are easy methods to break the code when a key is reused.

Generating random numbers is not easy, and the Vernam cipher demands keys as long as the messages transmitted. As recently as 1992, high-quality computer random-number generators were classified by the U.S. government as munitions, along with tanks and fighter planes, and their export from the country was prohibited.

And of course Eve must not know the key. So there must be some way for Alice to get the key to Bob securely. If they have some secure method for transmitting keys, why don’t they just use that same secure method for sending their messages?

In common parlance, the word “random” can mean “unimportant, not worth considering” (as in “Joe made a random comment”). So it may seem remarkable that a major problem for government, the military, and commerce is the generation and distribution of randomness, but that is indeed the case.

### 1.6.2 *Quantum mechanics to the rescue*

Since quantum mechanics involves randomness, it seems uniquely positioned to solve this problem. Here’s one scheme.

Alice and Bob set up a source of entangled atoms halfway between their two homes. Both of them erect vertical Stern-Gerlach analyzers to detect the atoms. If Alice’s atom comes out +, she will interpret it as a 1, if –, a 0. Bob interprets his atoms in the opposite sense. Since the entangled

atoms always exit from opposite ports, Alice and Bob end up with the same random number, which they use as a key for their Vernam-cipher communications over conventional telephone or computer lines.

This scheme will indeed produce and distribute copious, high-quality random numbers. But Eve can get at those same numbers through the following trick: She cuts open the atom pipe leading from the entangled source to Alice's home, and inserts a vertical interferometer.<sup>24</sup> She watches the atoms pass through her interferometer. If the atom takes path a, Eve knows that when Alice receives that same atom, it will exit from Eve's + port. If the atom takes path b, the opposite holds. Eve gets the key, Eve breaks the code.

It's worth looking at this eavesdropping in just a bit more detail. When the two atoms depart from their source, they are entangled. It is *not* true that, say, Alice's atom has  $\mu_z = +\mu_B$  while Bob's atom has  $\mu_z = -\mu_B$  — the pair of atoms is in the condition we've called "entangled", but the individual atoms themselves are not in *any* condition. However, after Eve sees the atom taking path a of her interferometer, then the two atoms are no longer entangled — now it *is* true that Alice's atom has the condition  $\mu_z = +\mu_B$  while Bob's atom has the condition  $\mu_z = -\mu_B$ . The key received by Alice and Bob will be random whether or not Eve is listening in. To test for eavesdropping, Alice and Bob must examine it in some other way.

Replace Alice and Bob's vertical analyzers with flipping Stern-Gerlach analyzers. After Bob receives his random sequence of pluses and minuses, he sends it to Alice over an open communication line. (Eve will intercept that sequence but it won't do her any good, because Bob sends only the pluses and minuses, not the orientations of his analyzer.) Alice now knows both the results at her analyzer and the results at Bob's analyzer, so she can perform a test of Bell's theorem: If she finds that the probability of atoms coming out opposite is  $\frac{1}{2}$ , then she knows that their atoms have arrived entangled, thus Eve has *not* observed the atoms in transit. If she finds that the probability is between  $\frac{5}{9}$  and 1, then she knows for certain that Eve *is* listening in, and they must not use their compromised key.

Is there some other way for Eve to tap the line? No! If the atom pairs pass the test for entanglement, then no one can know the values of their

<sup>24</sup>Inspired by James Bond, I always picture Eve as exotic beauty in a little black dress slinking to the back of an eastern European café to tap the diplomatic cable which conveniently runs there. But in point of fact Eve would be a computer.

$\mu_z$  projections because those projections don't exist! We have guaranteed that no one has intercepted the key by the interferometer method, or by any other method whatsoever.

Once Alice has tested Bell's theorem, she and Bob still have a lot of work to do. For a key they must use only those random numbers produced when their two analyzers happen to have the same orientations. There are detailed protocols specifying how Alice and Bob must exchange information about their analyzer orientations, in such a way that Eve can't uncover them. I won't describe these protocols because while they tell you how clever people are, they tell you nothing about how nature behaves. But you should take away that entanglement is not merely a phenomenon of nature: it is also a natural resource.

### 1.7 What is a qubit?

We've devoted an entire chapter to the magnetic moment of a silver atom. Perhaps you find this inappropriate: do you really care so much about silver atoms? Yes you do, because the phenomena and principles we've established concerning the magnetic moment of a silver atom apply to a host of other systems: the polarization of a light photon, the hybridization of a benzene molecule, the position of the nitrogen atom within an ammonia molecule, the neutral kaon, and more. Such systems are called "two-state systems" or "spin- $\frac{1}{2}$  systems" or "qubit systems". The ideas we establish concerning the magnetic moment of a silver atom apply equally well to all these systems.

After developing these ideas in the next chapter, we will (in chapter 6, "The Quantum Mechanics of Position") generalize them to continuum systems like the position of an electron.

**Problem**1.12 **Questions** (recommended problem)

Answering questions is an important scientific skill and, like any skill, it is sharpened through practice. This book gives you plenty of opportunities to develop that skill. Asking questions is another important scientific skill.<sup>25</sup> To hone that skill, write down a list of questions you have about quantum mechanics at this point. Be brief and pointed: you will not be graded for number or for verbosity. In future problems, I will ask you to add to your list.

[[For example, one of my questions would be: “Can entanglement be used to send a message from the left analyzer to the right analyzer?”]]

---

<sup>25</sup>“The important thing is not to stop questioning,” said Einstein. “Never lose a holy curiosity.” [Interview by William Miller, “Death of a Genius”, *Life* magazine, volume **38**, number 18 (2 May 1955) pages 61–64 on page 64.]





## Chapter 2

# Forging Mathematical Tools

When you walked into your introductory classical mechanics course, you were already familiar with the phenomena of introductory classical mechanics: flying balls, spinning wheels, colliding billiard balls. Your introductory mechanics textbook didn't need to introduce these things to you, but instead jumped right into describing these phenomena mathematically and explaining them in terms of more general principles.

The first chapter of this textbook made you familiar with the phenomena of quantum mechanics: quantization, interference, and entanglement — at least, insofar as these phenomena are manifest in the behavior of the magnetic moment of a silver atom. You are now, with respect to quantum mechanics, at the same level that you were, with respect to classical mechanics, when you walked into your introductory mechanics course. It is now our job to describe these quantal phenomena mathematically, to explain them in terms of more general principles, and (eventually) to investigate situations more complex than the magnetic moment of one or two silver atoms.

### 2.1 What is a quantal state?

We've been talking about the state of the silver atom's magnetic moment by saying things like “the projection of the magnetic moment on the  $z$  axis is  $\mu_z = -\mu_B$ ” or “ $\mu_x = +\mu_B$ ” or “ $\mu_\theta = -\mu_B$ ”. This notation is clumsy. First of all, it requires you to write down the same old  $\mu$ s time and time again. Second, the most important thing is the axis ( $z$  or  $x$  or  $\theta$ ), and the symbol for the axis is also the smallest and easiest to overlook.

P.A.M. Dirac<sup>1</sup> invented a notation that overcomes these faults. He looked at descriptions like

$$\mu_z = -\mu_B \quad \text{or} \quad \mu_x = +\mu_B \quad \text{or} \quad \mu_\theta = -\mu_B$$

and noted that the only difference from one expression to the other was the axis subscript and the sign in front of  $\mu_B$ . Since the only thing that distinguishes one expression from another is  $(z, -)$ , or  $(x, +)$ , or  $(\theta, -)$ , Dirac thought, these should be the only things we need to write down. He denoted these three states as

$$|z-\rangle \quad \text{or} \quad |x+\rangle \quad \text{or} \quad |\theta-\rangle.$$

The placeholders  $|\ \ \rangle$  are simply ornaments to remind us that we're talking about quantal states, just as the arrow atop  $\vec{r}$  is simply an ornament to remind us that we're talking about a vector. States expressed using this notation are sometimes called “kets”.

Simply establishing a notation doesn't tell us much. Just as in classical mechanics, we say we know a state when we know all the information needed to describe the system now and to predict its future. In our universe the classical time evolution law is

$$\vec{F} = m \frac{d^2 \vec{r}}{dt^2}$$

and so the state is specified by giving both a position  $\vec{r}$  and a velocity  $\vec{v}$ . If nature had instead provided the time evolution law

$$\vec{F} = m \frac{d^3 \vec{r}}{dt^3}$$

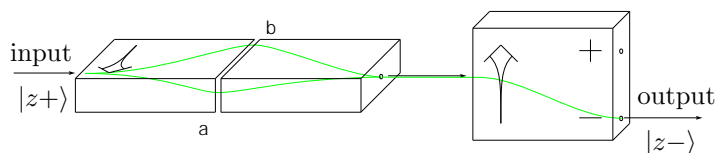
then the state would have been specified by giving a position  $\vec{r}$ , a velocity  $\vec{v}$ , and an acceleration  $\vec{a}$ . The specification of state is dictated by nature, not by humanity, so we can't know how to specify a state until we know the laws of physics governing that state. Since we don't yet know the laws of quantal physics, we can't yet know exactly how to specify a quantal state.

Classical intuition makes us suppose that, to specify the magnetic moment of a silver atom, we need to specify all three components  $\mu_z$ ,  $\mu_x$ , and  $\mu_y$ . We have already seen that nature precludes such a specification: if the magnetic moment has a value for  $\mu_z$ , then it doesn't have a value for  $\mu_x$ ,

<sup>1</sup>The Englishman Paul Adrien Maurice Dirac (1902–1984) in 1928 formulated a relativistically correct quantum mechanical equation that turns out to describe the electron. In connection with this so-called Dirac equation, he predicted the existence of antimatter. Dirac was painfully shy and notoriously cryptic.

and it's absurd to demand a specification for something that doesn't exist. As we learn more and more quantum physics, we will learn better and better how to specify states. There will be surprises. But always keep in mind that (just as in classical mechanics) it is experiment, not philosophy or meditation, and certainly not common sense, that tells us how to specify states.

## 2.2 Amplitude



An atom in state  $|z+\rangle$  ambivates through the apparatus above. We have already seen that, when the atom ambivates in darkness,

$$\begin{aligned} &\text{probability to go from input to output} \neq \\ &\quad \text{probability to go from input to output via path a} \\ &\quad + \text{probability to go from input to output via path b.} \end{aligned} \quad (2.1)$$

On the other hand, it makes sense to associate some sort of “influence to go from input to output via path a” with the path through a and a corresponding “influence to go from input to output via path b” with the path through b. This postulated influence is called “probability amplitude” or just “amplitude”.<sup>2</sup> Whatever amplitude is, its desired property is that

$$\begin{aligned} &\text{amplitude to go from input to output} = \\ &\quad \text{amplitude to go from input to output via path a} \\ &\quad + \text{amplitude to go from input to output via path b.} \end{aligned} \quad (2.2)$$

For the moment, the very existence of amplitude is nothing but a hopeful surmise. Scientists cannot now and indeed never will be able to prove that the concept of amplitude applies to all situations. That's because new situations are being investigated every day, and perhaps tomorrow a new

<sup>2</sup>The name “amplitude” is a poor one, because it is also used for the maximum value of a sinusoidal signal — in the function  $A \sin(\omega t)$ , the symbol  $A$  represents the amplitude — and this sinusoidal signal “amplitude” has nothing to do with the quantal “amplitude”. One of my students correctly suggested that a better name for quantal amplitude would be “proclivity”. But it's too late now to change the word.

situation will be discovered that cannot be described in terms of amplitudes. But as of today, that hasn't happened.

The role of amplitude, whatever it may prove to be, is to calculate probabilities. We establish three desirable rules:

- (1) *From amplitude to probability.* For every possible action there is an associated amplitude, such that

$$\text{probability for the action} = |\text{amplitude for the action}|^2.$$

- (2) *Actions in series.* If an action takes place through several successive stages, the amplitude for that action is the product of the amplitudes for each stage.
- (3) *Actions in parallel.* If an action could take place in several possible ways, the amplitude for that action is the sum of the amplitudes for each possibility.

The first rule is a simple way to make sure that probabilities are always positive. The second rule is a natural generalization of the rule for probabilities in series — that if an action happens through several stages, the probability for the action as a whole is the product of the probabilities for each stage. And the third rule simply restates the “desired property” presented in equation (2.2).

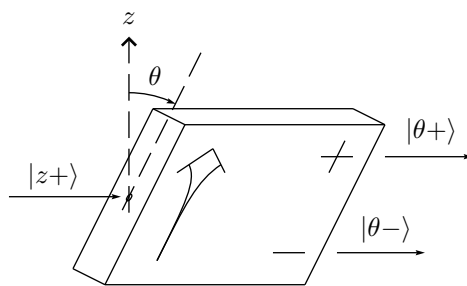
We apply these rules to various situations that we've already encountered, beginning with the interference experiment sketched above. Recall the probabilities already established (first column in table):

	probability	amplitude	amplitude
go from input to output	0	0	0
go from input to output via path a	$\frac{1}{4}$	$\frac{1}{2}$	$+\frac{1}{2}$
go from input to output via path b	$\frac{1}{4}$	$\frac{1}{2}$	$-\frac{1}{2}$

If rule (1) is to hold, then the amplitude to go from input to output must also be 0, while the amplitude to go via a path must have magnitude  $\frac{1}{2}$  (second column in table). According to rule (3), the two amplitudes to go via a and via b must sum to zero, so they cannot both be represented by positive numbers. Whatever mathematical entity is used to represent amplitude, it must enable two such entities, each with non-zero magnitude, to sum to zero. There are many such entities: real numbers, complex

numbers, hypercomplex numbers, and vectors in three dimensions are all possibilities. For this particular interference experiment, it suffices to assign real numbers to amplitudes: the amplitude to go via path a is  $+\frac{1}{2}$ , and the amplitude to go via path b is  $-\frac{1}{2}$ . (Third column in table. The negative sign could have been assigned to path a rather than to path b: this choice is merely conventional.) For other interference experiments complex numbers are required. It turns out that, for all situations yet encountered, one can represent amplitude mathematically as a complex number. Once again, this reflects the results of experiment, not of philosophy or meditation.

The second situation we'll consider is a Stern-Gerlach analyzer.



The amplitude that an atom entering the  $\theta$ -analyzer in state  $|z+\rangle$  exits in state  $|\theta+\rangle$  is called<sup>3</sup>  $\langle\theta+|z+\rangle$ . That phrase is a real mouthful, so the symbol  $\langle\theta+|z+\rangle$  is pronounced “the amplitude that  $|z+\rangle$  is in  $|\theta+\rangle$ ”, even though this briefer pronunciation leaves out the important role of the analyzer.<sup>4</sup> From rule (1), we know that

$$|\langle\theta+|z+\rangle|^2 = \cos^2(\theta/2) \tag{2.3}$$

$$|\langle\theta-|z+\rangle|^2 = \sin^2(\theta/2). \tag{2.4}$$

You can also use rule (1), in connection with the experiments described in

<sup>3</sup>The states appear in the symbol in the opposite sequence from their appearance in the description.

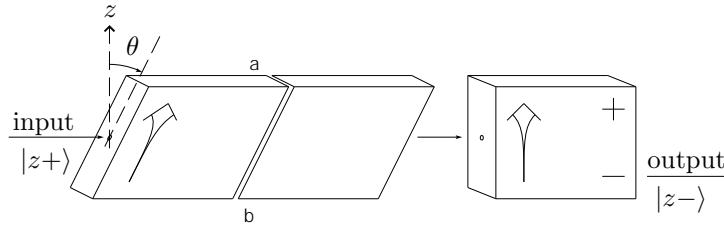
<sup>4</sup>The ultimate source of such problems is that the English language was invented by people who did not understand quantum mechanics, hence they never produced concise, accurate phrases to describe quantal phenomena. In the same way, the ancient phrase “search the four corners of the Earth” is still colorful and practical, and is used today even by those who know that the Earth doesn’t have four corners.

problem 1.2, “Exit probabilities” (on page 20) to determine that

$$\begin{aligned} |\langle z+|\theta+\rangle|^2 &= \cos^2(\theta/2) \\ |\langle z-|\theta+\rangle|^2 &= \sin^2(\theta/2) \\ |\langle \theta+|z-\rangle|^2 &= \sin^2(\theta/2) \\ |\langle \theta-|z-\rangle|^2 &= \cos^2(\theta/2) \\ |\langle z+|\theta-\rangle|^2 &= \sin^2(\theta/2) \\ |\langle z-|\theta-\rangle|^2 &= \cos^2(\theta/2). \end{aligned}$$

Clearly analyzer experiments like these determine the *magnitude* of an amplitude. No analyzer experiment can determine the *phase* of an amplitude. To determine phases, we must perform interference experiments.

So the third situation is an interference experiment.



Rule (2), actions in series, tells us that the amplitude to go from  $|z+\rangle$  to  $|z-\rangle$  via path **a** is the product of the amplitude to go from  $|z+\rangle$  to  $|\theta+\rangle$  times the amplitude to go from  $|\theta+\rangle$  to  $|z-\rangle$ :

$$\text{amplitude to go via path a} = \langle z-|\theta+\rangle \langle \theta+|z+\rangle.$$

Similarly

$$\text{amplitude to go via path b} = \langle z-|\theta-\rangle \langle \theta-|z+\rangle.$$

And then rule (3), actions in parallel, tells us that the amplitude to go from  $|z+\rangle$  to  $|z-\rangle$  is the sum of the amplitude to go via path **a** and the amplitude to go via path **b**. In other words

$$\langle z-|z+\rangle = \langle z-|\theta+\rangle \langle \theta+|z+\rangle + \langle z-|\theta-\rangle \langle \theta-|z+\rangle. \quad (2.5)$$

We know the magnitude of each of these amplitudes from analyzer experiments:

amplitude	magnitude
$\langle z- z+\rangle$	0
$\langle z- \theta+\rangle$	$ \sin(\theta/2) $
$\langle \theta+ z+\rangle$	$ \cos(\theta/2) $
$\langle z- \theta-\rangle$	$ \cos(\theta/2) $
$\langle \theta- z+\rangle$	$ \sin(\theta/2) $

The task now is to assign phases to these magnitudes in such a way that equation (2.5) is satisfied. In doing so we are faced with an embarrassment of riches: there are *many* consistent ways to make this assignment. Here are two commonly used conventions:

amplitude	convention I	convention II
$\langle z- z+\rangle$	0	0
$\langle z- \theta+\rangle$	$\sin(\theta/2)$	$i \sin(\theta/2)$
$\langle \theta+ z+\rangle$	$\cos(\theta/2)$	$\cos(\theta/2)$
$\langle z- \theta-\rangle$	$\cos(\theta/2)$	$\cos(\theta/2)$
$\langle \theta- z+\rangle$	$-\sin(\theta/2)$	$-i \sin(\theta/2)$

There are two things to notice about these amplitude assignments. First, one normally assigns values to physical quantities by experiment, or by calculation, but not “by convention”. Second, both of these conventions show unexpected behaviors: Because the angle  $0^\circ$  is the same as the angle  $360^\circ$ , one would expect that  $\langle 0^\circ+|z+\rangle$  would equal  $\langle 360^\circ+|z+\rangle$ , whereas in fact the first amplitude is  $+1$  and the second is  $-1$ . Because the state  $|180^\circ-\rangle$  (that is,  $|\theta-\rangle$  with  $\theta = 180^\circ$ ) is the same as the state  $|z+\rangle$ , one would expect that  $\langle 180^\circ-|z+\rangle = 1$ , whereas in fact  $\langle 180^\circ-|z+\rangle$  is either  $-1$  or  $-i$ , depending on convention. These two observations underscore the fact that amplitude is a mathematical tool that enables us to calculate physically observable quantities, like probabilities. It is not itself a physical entity. No experiment measures amplitude. Amplitude is not “out there, physically present in space” in the way that, say, a nitrogen molecule is.

A good analogy is that an amplitude convention is like a language. Any language is a human convention: there is no intrinsic connection between a physical horse and the English word “horse”, or the German word “pferd”,

or the Swahili word “farasi”. The fact that language is pure human convention, and that there are multiple conventions for the name of a horse, doesn’t mean that language is unimportant: on the contrary language is an immensely powerful tool. And the fact that language is pure human convention doesn’t mean that you can’t develop intuition about language: on the contrary if you know the meaning of “arachnid” and the meaning of “phobia”, then your intuition for English tells you that “arachnophobia” means fear of spiders. Exactly the same is true for amplitude: it is a powerful tool, and with practice you can develop intuition for it.

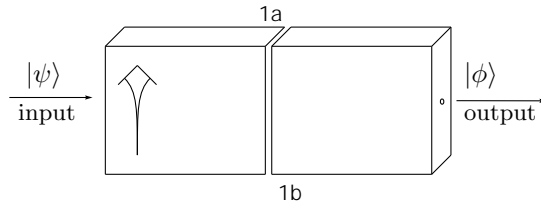
When I introduced the phenomenon of quantal interference on page 25, I said that there was no word or phrase in the English language that accurately represents what’s going on: It’s flat-out wrong to say “the atom takes path a” and it’s flat-out wrong to say “the atom takes path b”. It gives a wrong impression to say “the atom takes no path” or “the atom takes both paths”. I introduced the phrase “the atom ambivates through the two paths of the interferometer”. Now we have a technically correct way of describing the phenomenon: “the atom has an amplitude to take path a and an amplitude to take path b”.

Here’s another warning about language: If an atom in state  $|\psi\rangle$  enters a vertical analyzer, the amplitude for it to exit from the + port is  $\langle z+|\psi\rangle$ . (And of course the amplitude for it exit from the – port is  $\langle z-|\psi\rangle$ .) This is often stated “If the atom is in state  $|\psi\rangle$ , the amplitude of it being in state  $|z+\rangle$  is  $\langle z+|\psi\rangle$ .” This is an acceptable shorthand for the full explanation, which requires thinking about an analyzer experiment, even though the shorthand never mentions the analyzer. But never say “If the atom is in state  $|\psi\rangle$ , the probability of it being in state  $|z+\rangle$  is  $|\langle z+|\psi\rangle|^2$ .” This gives the distinct and incorrect impression that before entering the analyzer, the atom was either in state  $|z+\rangle$  or in state  $|z-\rangle$ , and you just didn’t know which it was. Instead, say “If an atom in state  $|\psi\rangle$  enters a vertical analyzer, the probability of exiting from the + port in state  $|z+\rangle$  is  $|\langle z+|\psi\rangle|^2$ .”

### 2.2.1 *Sample Problem: Two paths*

Find an equation similar to equation (2.5) representing the amplitude to start in state  $|\psi\rangle$  at input, ambivate through a vertical interferometer, and end in state  $|\phi\rangle$  at output.





*Solution:* Because of rule (2), actions in series, the amplitude for the atom to take the top path is the product

$$\langle \phi | z + \rangle \langle z + | \psi \rangle.$$

Similarly the amplitude for it to take the bottom path is

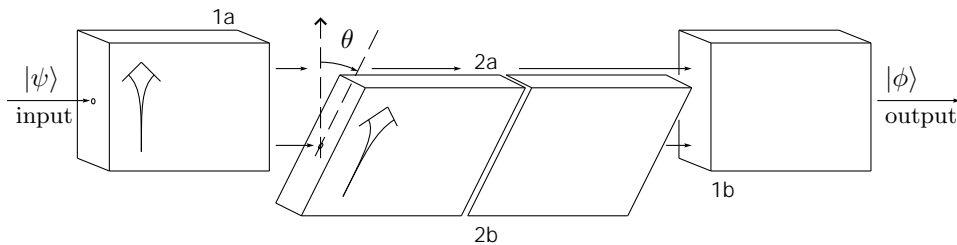
$$\langle \phi | z - \rangle \langle z - | \psi \rangle.$$

Because of rule (3), actions in parallel, the amplitude for it to ambivate through both paths is the sum of these two, and we conclude that

$$\langle \phi | \psi \rangle = \langle \phi | z + \rangle \langle z + | \psi \rangle + \langle \phi | z - \rangle \langle z - | \psi \rangle. \quad (2.6)$$

### 2.2.2 Sample Problem: Three paths

Stretch apart a vertical interferometer, so that the recombining rear end is far from the splitting front end, and insert a  $\theta$  interferometer into the bottom path. Now there are *three* paths from input to output. Find an equation similar to equation (2.5) representing the amplitude to start in state  $|\psi\rangle$  at input and end in state  $|\phi\rangle$  at output.



---

*Solution:*

$$\begin{aligned}
 \langle \phi | \psi \rangle &= \langle \phi | z+ \rangle \langle z+ | \psi \rangle \\
 &+ \langle \phi | z- \rangle \langle z- | \theta+ \rangle \langle \theta+ | z- \rangle \langle z- | \psi \rangle \\
 &+ \langle \phi | z- \rangle \langle z- | \theta- \rangle \langle \theta- | z- \rangle \langle z- | \psi \rangle
 \end{aligned}
 \tag{2.7}$$


---

## Problems

### 2.1 Talking about interference

An atom in state  $|\psi\rangle$  ambivates through a vertical analyzer. We say, appropriately, that “the atom has an amplitude to take the top path and an amplitude to take the bottom path”. Find expressions for those two amplitudes and describe, in ten sentences or fewer, why it is *not* appropriate to say “the atom has probability  $|\langle z+ | \psi \rangle|^2$  to take the top path and probability  $|\langle z- | \psi \rangle|^2$  to take the bottom path”.

### 2.2 Other conventions

Two conventions for assigning amplitudes are given in the table on page 61. Show that if  $\langle z- | \theta+ \rangle$  and  $\langle z- | \theta- \rangle$  are multiplied by phase factor  $e^{i\alpha}$ , and if  $\langle z+ | \theta+ \rangle$  and  $\langle z+ | \theta- \rangle$  are multiplied by phase factor  $e^{i\beta}$  (where  $\alpha$  and  $\beta$  are both real), then the resulting amplitudes are just as good as the original (for either convention I or convention II).

### 2.3 Peculiarities of amplitude

Page 61 pointed out some of the peculiarities of amplitude; this problem points out another. Since the angle  $\theta$  is the same as the angle  $360^\circ + \theta$ , one would expect that  $\langle \theta+ | z+ \rangle$  would equal  $\langle (360^\circ + \theta) | z+ \rangle$ . Show, using either of the conventions given in the table on page 61, that this expectation is false. What is instead correct?

### 2.3 Reversal-conjugation relation

Working with amplitudes is made easier through the theorem that the amplitude to go from state  $|\psi\rangle$  to state  $|\phi\rangle$  and the amplitude to go in the opposite direction are related through complex conjugation:

$$\langle \phi | \psi \rangle = \langle \psi | \phi \rangle^* .
 \tag{2.8}$$

The proof below works for states of the magnetic moment of a silver atom — the kind of states we’ve worked with so far — but in fact the result holds for any quantal system.

The proof relies on three facts: First, the probability for one state to be analyzed into another depends only on the magnitude of the angle between the incoming magnetic moment and the analyzer, and not on the sense of that angle. (An atom in state  $|z+\rangle$  has the same probability of leaving the + port of an analyzer whether it is rotated  $17^\circ$  clockwise or  $17^\circ$  counterclockwise.) Thus

$$|\langle\phi|\psi\rangle|^2 = |\langle\psi|\phi\rangle|^2. \quad (2.9)$$

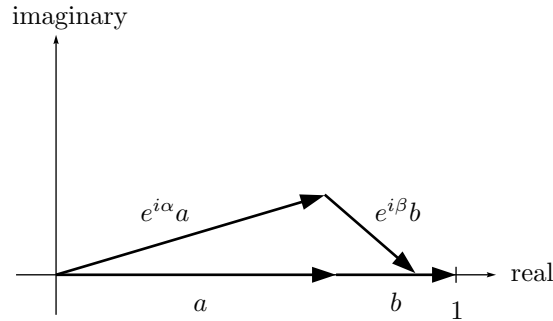
Second, an atom exits an interferometer in the same state in which it entered, so

$$\langle\phi|\psi\rangle = \langle\phi|\theta+\rangle\langle\theta+|\psi\rangle + \langle\phi|\theta-\rangle\langle\theta-|\psi\rangle. \quad (2.10)$$

Third, an atom entering an analyzer comes out somewhere, so

$$1 = |\langle\theta+|\psi\rangle|^2 + |\langle\theta-|\psi\rangle|^2. \quad (2.11)$$

The proof also relies on a mathematical result called “the triangle inequality for complex numbers”: If  $a$  and  $b$  are real numbers with  $a + b = 1$ , and in addition  $e^{i\alpha}a + e^{i\beta}b = 1$ , with  $\alpha$  and  $\beta$  real, then  $\alpha = \beta = 0$ . You can find very general, very abstract, proofs of the triangle inequality, but the complex plane sketch below encapsulates the idea:



From the first fact (2.9), the two complex numbers  $\langle\phi|\psi\rangle$  and  $\langle\psi|\phi\rangle$  have the same magnitude, so they differ only in phase. Write this statement as

$$\langle\phi|\psi\rangle = e^{i\delta}\langle\psi|\phi\rangle^* \quad (2.12)$$

where the phase  $\delta$  is a real number that might depend on the states  $|\phi\rangle$  and  $|\psi\rangle$ . Apply this general result first to the particular state  $|\phi\rangle = |\theta+\rangle$ :

$$\langle\theta+|\psi\rangle = e^{i\delta_+} \langle\psi|\theta+\rangle^*, \quad (2.13)$$

and then to the particular state  $|\phi\rangle = |\theta-\rangle$ :

$$\langle\theta-|\psi\rangle = e^{i\delta_-} \langle\psi|\theta-\rangle^*, \quad (2.14)$$

where the two real numbers  $\delta_+$  and  $\delta_-$  might be different. Our objective is to prove that  $\delta_+ = \delta_- = 0$ .

Apply the second fact (2.10) with  $|\phi\rangle = |\psi\rangle$ , giving

$$\begin{aligned} 1 &= \langle\psi|\theta+\rangle\langle\theta+|\psi\rangle + \langle\psi|\theta-\rangle\langle\theta-|\psi\rangle \\ &= e^{i\delta_+} \langle\psi|\theta+\rangle\langle\psi|\theta+\rangle^* + e^{i\delta_-} \langle\psi|\theta-\rangle\langle\psi|\theta-\rangle^* \\ &= e^{i\delta_+} |\langle\psi|\theta+\rangle|^2 + e^{i\delta_-} |\langle\psi|\theta-\rangle|^2 \\ &= e^{i\delta_+} |\langle\theta+|\psi\rangle|^2 + e^{i\delta_-} |\langle\theta-|\psi\rangle|^2. \end{aligned} \quad (2.15)$$

Compare this result to the third fact (2.11)

$$1 = |\langle\theta+|\psi\rangle|^2 + |\langle\theta-|\psi\rangle|^2 \quad (2.16)$$

and use the triangle inequality with  $a = |\langle\theta+|\psi\rangle|^2$  and  $b = |\langle\theta-|\psi\rangle|^2$ . The two the phases  $\delta_+$  and  $\delta_-$  must vanish, so the “reversal-conjugation relation” is proven.

## 2.4 Establishing a phase convention

Although there are multiple alternative phase conventions for amplitudes (see problem 2.2 on page 64), we will from now on use only phase convention I from page 61:

$$\begin{aligned} \langle z+|\theta+\rangle &= \cos(\theta/2) \\ \langle z-|\theta+\rangle &= \sin(\theta/2) \\ \langle z+|\theta-\rangle &= -\sin(\theta/2) \\ \langle z-|\theta-\rangle &= \cos(\theta/2) \end{aligned} \quad (2.17)$$

In particular, for  $\theta = 90^\circ$  we have

$$\begin{aligned} \langle z+|x+\rangle &= 1/\sqrt{2} \\ \langle z-|x+\rangle &= 1/\sqrt{2} \\ \langle z+|x-\rangle &= -1/\sqrt{2} \\ \langle z-|x-\rangle &= 1/\sqrt{2} \end{aligned} \quad (2.18)$$

This convention has a desirable special case for  $\theta = 0^\circ$ , namely

$$\begin{aligned}\langle z+|\theta+\rangle &= 1 \\ \langle z-|\theta+\rangle &= 0 \\ \langle z+|\theta-\rangle &= 0 \\ \langle z-|\theta-\rangle &= 1\end{aligned}\tag{2.19}$$

but an unexpected special case for  $\theta = 360^\circ$ , namely

$$\begin{aligned}\langle z+|\theta+\rangle &= -1 \\ \langle z-|\theta+\rangle &= 0 \\ \langle z+|\theta-\rangle &= 0 \\ \langle z-|\theta-\rangle &= -1\end{aligned}\tag{2.20}$$

This is perplexing, given that the angle  $\theta = 0^\circ$  is the same as the angle  $\theta = 360^\circ$ ! Any convention will have similar perplexing cases. Such perplexities underscore the fact that amplitudes are important mathematical tools used to calculate probabilities, but are not “physically real”.

Given these amplitudes, we can use the interference result (2.6) to calculate any amplitude of interest:

$$\begin{aligned}\langle \phi|\psi\rangle &= \langle \phi|z+\rangle\langle z+|\psi\rangle + \langle \phi|z-\rangle\langle z-|\psi\rangle \\ &= \langle z+|\phi\rangle^*\langle z+|\psi\rangle + \langle z-|\phi\rangle^*\langle z-|\psi\rangle\end{aligned}\tag{2.21}$$

where in the last line we have used the reversal-conjugation relation (2.8).

### Problems

#### 2.4 Other conventions, other peculiarities

Write what this section would have been had we adopted convention II rather than convention I from page 61. In addition, evaluate the four amplitudes of equation (2.17) for  $\theta = +180^\circ$  and  $\theta = -180^\circ$ .

#### 2.5 Finding amplitudes (recommended problem)

Using the interference idea embodied in equation (2.21), calculate the amplitudes  $\langle \theta+|54^\circ+\rangle$  and  $\langle \theta-|54^\circ+\rangle$  as a function of  $\theta$ . Do these amplitudes have the values you expect for  $\theta = 54^\circ$ ? For  $\theta = 234^\circ$ ? Plot  $\langle \theta+|54^\circ+\rangle$  for  $\theta$  from  $0^\circ$  to  $360^\circ$ . Compare the result for  $\theta = 0^\circ$  and  $\theta = 360^\circ$ .

#### 2.6 Rotations

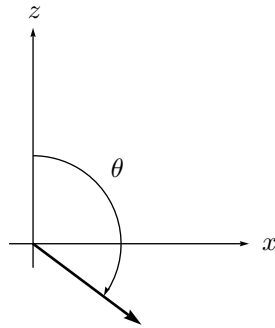
Use the interference idea embodied in equation (2.21) to show that

$$\begin{aligned}\langle x+|\theta+\rangle &= \frac{1}{\sqrt{2}}[\cos(\theta/2) + \sin(\theta/2)] \\ \langle x-|\theta+\rangle &= -\frac{1}{\sqrt{2}}[\cos(\theta/2) - \sin(\theta/2)] \\ \langle x+|\theta-\rangle &= \frac{1}{\sqrt{2}}[\cos(\theta/2) - \sin(\theta/2)] \\ \langle x-|\theta-\rangle &= \frac{1}{\sqrt{2}}[\cos(\theta/2) + \sin(\theta/2)]\end{aligned}\tag{2.22}$$

If and only if you enjoy trigonometric identities, you should then show that these results can be written equivalently as

$$\begin{aligned}
 \langle x+|\theta+\rangle &= \cos((\theta - 90^\circ)/2) \\
 \langle x-|\theta+\rangle &= \sin((\theta - 90^\circ)/2) \\
 \langle x+|\theta-\rangle &= -\sin((\theta - 90^\circ)/2) \\
 \langle x-|\theta-\rangle &= \cos((\theta - 90^\circ)/2)
 \end{aligned}
 \tag{2.23}$$

This makes perfect geometric sense, as the angle relative to the  $x$  axis is  $90^\circ$  less than the angle relative to the  $z$  axis:



## 2.5 How can I specify a quantal state?

We introduced the Dirac notation for quantal states on page 56, but haven't yet fleshed out that notation by specifying a state mathematically. Start with an analogy:

### 2.5.1 How can I specify a position vector?

We are so used to writing down the position vector  $\vec{r}$  that we rarely stop to ask ourselves what it means. But the plain fact is that whenever we measure a length (say, with a meter stick) we find not a vector, but a single number! Experiments measure never the vector  $\vec{r}$  but always a scalar — the dot product between  $\vec{r}$  and some other vector, call it  $\vec{s}$  for “some other”.

If we know the dot product between  $\vec{r}$  and every vector  $\vec{s}$ , then we know everything there is to know about  $\vec{r}$ . Does this mean that to specify  $\vec{r}$ , we

must keep a list of all possible dot products  $\vec{s} \cdot \vec{r}$ ? Of course not... such a list would be infinitely long!

You know that if you write  $\vec{r}$  in terms of an orthonormal basis  $\{\hat{i}, \hat{j}, \hat{k}\}$ , namely

$$\vec{r} = r_x \hat{i} + r_y \hat{j} + r_z \hat{k} \quad (2.24)$$

where  $r_x = \hat{i} \cdot \vec{r}$ ,  $r_y = \hat{j} \cdot \vec{r}$ , and  $r_z = \hat{k} \cdot \vec{r}$ , then you've specified the vector. Why? Because if you know the triplet  $(r_x, r_y, r_z)$  and the triplet  $(s_x, s_y, s_z)$ , then you can easily find the desired dot product

$$\vec{s} \cdot \vec{r} = \begin{pmatrix} s_x & s_y & s_z \end{pmatrix} \begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix} = s_x r_x + s_y r_y + s_z r_z. \quad (2.25)$$

It's a lot more compact to specify the vector through three dot products — namely  $\hat{i} \cdot \vec{r}$ ,  $\hat{j} \cdot \vec{r}$ , and  $\hat{k} \cdot \vec{r}$  — from which you can readily calculate an infinite number of desired dot products, than it is to list all infinity dot products themselves!

### 2.5.2 How can I specify a quantal state?

Like the position vector  $\vec{r}$ , the quantal state  $|\psi\rangle$  cannot by itself be measured. But if we determine (through some combination of analyzer experiments, interference experiments, and convention) the amplitude  $\langle\sigma|\psi\rangle$  for every possible state  $|\sigma\rangle$ , then we know everything there is to know about  $|\psi\rangle$ . Is there some compact way of specifying the state, or do we have to keep an infinitely long list of all these amplitudes?

This nut is cracked through the interference experiment result

$$\langle\sigma|\psi\rangle = \langle\sigma|\theta+\rangle\langle\theta+|\psi\rangle + \langle\sigma|\theta-\rangle\langle\theta-|\psi\rangle, \quad (2.26)$$

which simply says, in symbols, that the atom exits an interferometer in the same state in which it entered (see equation 2.10). It gets hard to keep track of all these symbols, so I'll introduce the names

$$\begin{aligned} \langle\theta+|\psi\rangle &= \psi_+ \\ \langle\theta-|\psi\rangle &= \psi_- \end{aligned}$$

and

$$\begin{aligned} \langle\theta+|\sigma\rangle &= \sigma_+ \\ \langle\theta-|\sigma\rangle &= \sigma_- \end{aligned}$$

From the reversal-conjugation relation, this means

$$\begin{aligned}\langle\sigma|\theta+\rangle &= \sigma_+^* \\ \langle\sigma|\theta-\rangle &= \sigma_-^*.\end{aligned}$$

In terms of these symbols, the interference result (2.26) is

$$\langle\sigma|\psi\rangle = \sigma_+^*\psi_+ + \sigma_-^*\psi_- = (\sigma_+^* \ \sigma_-^*) \begin{pmatrix} \psi_+ \\ \psi_- \end{pmatrix}. \quad (2.27)$$

And this is our shortcut! By keeping track of only two amplitudes,  $\psi_+$  and  $\psi_-$ , for each state, we can readily calculate any amplitude desired. We don't have to keep an infinitely long list of amplitudes.

This dot product result for computing amplitude is so useful and so convenient that sometimes people say the amplitude *is* a dot product. No. The amplitude reflects analyzer experiments, plus interference experiments, plus convention. The dot product is a powerful mathematical tool for computing amplitudes. (A parallel situation: There are many ways to find the latitude and longitude coordinates for a point on the Earth's surface, but the easiest is to use a GPS device. Some people are so enamored of this ease that they call the latitude and longitude the "GPS coordinates". But in fact the coordinates were established long before the Global Positioning System was built.)

### 2.5.3 What is a basis?

For vectors in three-dimensional space, an orthonormal basis such as  $\{\hat{i}, \hat{j}, \hat{k}\}$  is a set of three vectors of unit magnitude perpendicular to each other. As we've seen, the importance of a basis is that every vector  $\vec{r}$  can be represented as a sum over these basis vectors,

$$\vec{r} = r_x\hat{i} + r_y\hat{j} + r_z\hat{k},$$

and hence any vector  $\vec{r}$  can be conveniently represented through the triplet

$$\begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix} = \begin{pmatrix} \hat{i} \cdot \vec{r} \\ \hat{j} \cdot \vec{r} \\ \hat{k} \cdot \vec{r} \end{pmatrix}.$$

For quantal states, we've seen that a set of two states such as  $\{|\theta+\rangle, |\theta-\rangle\}$  plays a similar role, so it too is called a basis. For the magnetic moment of a silver atom, two states  $|a\rangle$  and  $|b\rangle$  constitute a basis whenever  $\langle a|b\rangle = 0$ , and the analyzer experiment of section 1.1.4 shows that



the states  $|\theta+\rangle$  and  $|\theta-\rangle$  certainly satisfy this requirement. In the basis  $\{|a\rangle, |b\rangle\}$  an arbitrary state  $|\psi\rangle$  can be conveniently represented through the pair of amplitudes

$$\begin{pmatrix} \langle a|\psi\rangle \\ \langle b|\psi\rangle \end{pmatrix}.$$

### 2.5.4 Hilbert space

We have learned to express a physical state as a mathematical entity — namely, using the  $\{|a\rangle, |b\rangle\}$  basis, the state  $|\psi\rangle$  is represented as a column matrix of amplitudes

$$\begin{pmatrix} \langle a|\psi\rangle \\ \langle b|\psi\rangle \end{pmatrix}.$$

This mathematical entity is called a “state vector in Hilbert<sup>5</sup> space”.

For example, in the basis  $\{|z+\rangle, |z-\rangle\}$  the state  $|\theta+\rangle$  is represented by

$$\begin{pmatrix} \langle z+|\theta+\rangle \\ \langle z-|\theta+\rangle \end{pmatrix} = \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix}. \quad (2.28)$$

Whereas (in light of equation 2.22) in the basis  $\{|x+\rangle, |x-\rangle\}$  that same state  $|\theta+\rangle$  is represented by the different column matrix

$$\begin{pmatrix} \langle x+|\theta+\rangle \\ \langle x-|\theta+\rangle \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}[\cos(\theta/2) + \sin(\theta/2)] \\ -\frac{1}{\sqrt{2}}[\cos(\theta/2) - \sin(\theta/2)] \end{pmatrix}. \quad (2.29)$$

Write down the interference experiment result twice

$$\begin{aligned} \langle a|\psi\rangle &= \langle a|z+\rangle\langle z+|\psi\rangle + \langle a|z-\rangle\langle z-|\psi\rangle \\ \langle b|\psi\rangle &= \langle b|z+\rangle\langle z+|\psi\rangle + \langle b|z-\rangle\langle z-|\psi\rangle \end{aligned}$$

and then write these two equations as one using column matrix notation

$$\begin{pmatrix} \langle a|\psi\rangle \\ \langle b|\psi\rangle \end{pmatrix} = \begin{pmatrix} \langle a|z+\rangle \\ \langle b|z+\rangle \end{pmatrix} \langle z+|\psi\rangle + \begin{pmatrix} \langle a|z-\rangle \\ \langle b|z-\rangle \end{pmatrix} \langle z-|\psi\rangle.$$

Notice the column matrix representations of states  $|\psi\rangle$ ,  $|z+\rangle$ , and  $|z-\rangle$ , and write this equation as

$$|\psi\rangle = |z+\rangle\langle z+|\psi\rangle + |z-\rangle\langle z-|\psi\rangle. \quad (2.30)$$

---

<sup>5</sup>The German mathematician David Hilbert (1862–1943) made contributions to functional analysis, geometry, mathematical physics, and other areas. He formalized and extended the concept of a vector space. Hilbert and Albert Einstein raced to uncover the field equations of general relativity, but Einstein beat Hilbert by a matter of weeks.

And now we have a new thing under the sun. We never talk about adding together two classical states, nor multiplying them by numbers, but this equation gives us the meaning of such state addition in quantum mechanics. This is a new mathematical tool, it deserves a new name, and that name is “superposition”. Superposition is the mathematical reflection of the physical phenomenon of interference, as in the sentence: “When an atom ambivates through an interferometer, its state is a superposition of the state of an atom taking path **a** and the state of an atom taking path **b**.”

Superposition is not familiar from daily life or from classical mechanics, but there is a story<sup>6</sup> that increases understanding: “A medieval European traveler returns home from a journey to India, and describes a rhinoceros as a sort of cross between a dragon and a unicorn.” In this story the rhinoceros, an animal that is *not* familiar but that *does* exist, is described as intermediate (a “sort of cross”) between two fantasy animals (the dragon and the unicorn) that *are* familiar (to the medieval European) but that *do not* exist.

Similarly, an atom in state  $|z+\rangle$  ambivates through both paths of a horizontal interferometer. This action is *not* familiar but *does* happen, and it is characterized as a superposition (a “sort of cross”) between two actions (“taking path **a**” and “taking path **b**”) that *are* familiar (to all of us steeped in the classical approximation) but that *do not* happen.

In principle, any calculation performed using the Hilbert space representation of states could be performed by considering suitable, cleverly designed analyzer and interference experiments. But it’s a lot easier to use the abstract Hilbert space machinery. (Similarly, any result in electrostatics could be found using Coulomb’s Law, but it’s a lot easier to use the abstract electric field and electric potential. Any calculation involving vectors could be performed graphically, but it’s a lot easier to use abstract components. Any addition or subtraction of whole numbers could be performed by counting out marbles, but it’s a lot easier to use abstract mathematical tools like carrying and borrowing.)

### 2.5.5 Peculiarities of state vectors

Because state vectors are built from amplitudes, and amplitudes have peculiarities (see pages 61 and 67), it is natural that state vectors have sim-

<sup>6</sup>Invented by John D. Roberts, but first published in Robert T. Morrison and Robert N. Boyd, *Organic Chemistry*, second edition (Allyn & Bacon, Boston, 1966) page 318.

ilar peculiarities. For example, since the angle  $\theta$  is the same as the angle  $\theta + 360^\circ$ , I would expect that the state vector  $|\theta+\rangle$  would be the same as the state vector  $|(\theta + 360^\circ)+\rangle$ .

But in fact, in the  $\{|z+\rangle, |z-\rangle\}$  basis, the state  $|\theta+\rangle$  is represented by

$$\begin{pmatrix} \langle z+|\theta+\rangle \\ \langle z-|\theta+\rangle \end{pmatrix} = \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix}, \quad (2.31)$$

so the state  $|(\theta + 360^\circ)+\rangle$  is represented by

$$\begin{aligned} \begin{pmatrix} \langle z+|(\theta + 360^\circ)+\rangle \\ \langle z-|(\theta + 360^\circ)+\rangle \end{pmatrix} &= \begin{pmatrix} \cos((\theta + 360^\circ)/2) \\ \sin((\theta + 360^\circ)/2) \end{pmatrix} \\ &= \begin{pmatrix} \cos(\theta/2 + 180^\circ) \\ \sin(\theta/2 + 180^\circ) \end{pmatrix} = \begin{pmatrix} -\cos(\theta/2) \\ -\sin(\theta/2) \end{pmatrix}. \end{aligned} \quad (2.32)$$

So in fact  $|\theta+\rangle = -|(\theta + 360^\circ)+\rangle$ . Bizarre!

This bizarreness is one facet of a general rule: If you multiply any state vector by a complex number with magnitude unity — a number such as  $-1$ , or  $i$ , or  $\frac{1}{\sqrt{2}}(-1 + i)$ , or  $e^{2.7i}$  — a so-called “complex unit” or “phase factor” — then you get a different state vector that represents the same state. This fact is called “global phase freedom” — you are free to set the overall phase of your state vector for your own convenience. This general rule applies only for multiplying both elements of the state vector by the same complex unit: if you multiply the two elements with *different* complex units, you *will* obtain a vector representing a different state (see problem 2.8 on page 75).

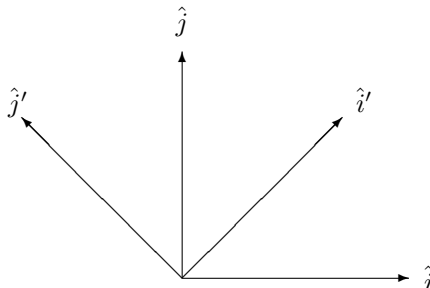
### 2.5.6 Names for position vectors

The vector  $\vec{r}$  is specified in the basis  $\{\hat{i}, \hat{j}, \hat{k}\}$  by the three components

$$\begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix} = \begin{pmatrix} \hat{i} \cdot \vec{r} \\ \hat{j} \cdot \vec{r} \\ \hat{k} \cdot \vec{r} \end{pmatrix}.$$

Because this component specification is so convenient, it is sometimes said that the vector  $\vec{r}$  is not just specified, but is *equal* to this triplet of numbers. That’s false.

Think of the vector  $\vec{r} = 5\hat{i} + 5\hat{j}$ . It is represented in the basis  $\{\hat{i}, \hat{j}, \hat{k}\}$  by the triplet  $(5, 5, 0)$ . But this is not the only basis that exists. In the basis  $\{\hat{i}' = (\hat{i} + \hat{j})/\sqrt{2}, \hat{j}' = (-\hat{i} + \hat{j})/\sqrt{2}, \hat{k}\}$ , that same vector is represented by the triplet  $(5\sqrt{2}, 0, 0)$ . If we had said that  $\vec{r} = (5, 5, 0)$  and that  $\vec{r} = (5\sqrt{2}, 0, 0)$ , then we would be forced to conclude that  $5 = 5\sqrt{2}$  and that  $5 = 0$ !



To specify a position vector  $\vec{r}$ , we use the components of  $\vec{r}$  in a particular basis, usually denoted  $(r_x, r_y, r_z)$ . We often write “ $\vec{r} = (r_x, r_y, r_z)$ ” but in fact that’s not exactly correct. The vector  $\vec{r}$  represents a position — it is independent of basis. The row matrix  $(r_x, r_y, r_z)$  represents the components of that position vector in a particular basis — it is the “name” of the position in a particular basis. Instead of using an equals sign  $=$  we use the symbol  $\doteq$  to mean “represented by in a particular basis”, as in “ $\vec{r} \doteq (5, 5, 0)$ ” meaning “the vector  $\vec{r} = 5\hat{i} + 5\hat{j}$  is represented by the triplet  $(5, 5, 0)$  in the basis  $\{\hat{i}, \hat{j}, \hat{k}\}$ ”.

Vectors are physical things: a caveman throwing a spear at a mammoth was performing addition of position vectors, even though the caveman didn’t understand basis vectors or Cartesian coordinates. The concept of “position” was known to cavemen who did not have any concept of “basis”.

### 2.5.7 Names for quantal states

We’ve been specifying a state like  $|\psi\rangle = |17^\circ+\rangle$  by stating the axis upon which the projection of  $\vec{\mu}$  is definite and equal to  $+\mu_B$  — in this case, the axis tilted  $17^\circ$  from the vertical.

Another way to specify a state  $|\psi\rangle$  would be to give the amplitude that  $|\psi\rangle$  is in any possible state: that is, to list  $\langle\theta+|\psi\rangle$  and  $\langle\theta-|\psi\rangle$  for all values of  $\theta$ :  $0^\circ \leq \theta < 360^\circ$ . One of those amplitudes (in this case  $\langle 17^\circ+|\psi\rangle$ ) will have value 1, and finding this one amplitude would give us back the information in the specification  $|17^\circ+\rangle$ . In some ways this is a more convenient specification because we don’t have to look up amplitudes: they’re right there in the list. On the other hand it is an awful lot of information to have to carry around.

The Hilbert space approach is a third way to specify a state that combines the brevity of the first way with the convenience of the second way. Instead of listing the amplitude  $\langle\sigma|\psi\rangle$  for *every* state  $|\sigma\rangle$  we list only the two amplitudes  $\langle a|\psi\rangle$  and  $\langle b|\psi\rangle$  for the elements  $\{|a\rangle, |b\rangle\}$  of a basis. We've already seen (equation 2.27) how quantal interference then allows us to readily calculate any amplitude.

Just as we said “the position vector  $\vec{r}$  is represented in the basis  $\{\hat{i}, \hat{j}, \hat{k}\}$  as  $(1, 1, 0)$ ” or

$$\vec{r} \doteq (1, 1, 0),$$

so we say “the quantal state  $|\psi\rangle$  is represented in the basis  $\{|z+\rangle, |z-\rangle\}$  as

$$|\psi\rangle \doteq \begin{pmatrix} \langle z+|\psi\rangle \\ \langle z-|\psi\rangle \end{pmatrix}.”$$

### Problems

#### 2.7 Superposition and interference (recommended problem)

On page 72 I wrote that “When an atom ambivates through an interferometer, its state is a superposition of the state of an atom taking path a and the state of an atom taking path b.”

- Write down a superposition equation reflecting this sentence for the interference experiment sketched on page 57.
- Do the same for the interference experiment sketched on page 60.

#### 2.8 Representations (recommended problem)

In the  $\{|z+\rangle, |z-\rangle\}$  basis the state  $|\psi\rangle$  is represented by

$$\begin{pmatrix} \psi_+ \\ \psi_- \end{pmatrix}.$$

(In other words,  $\psi_+ = \langle z+|\psi\rangle$  and  $\psi_- = \langle z-|\psi\rangle$ .)

- If  $\psi_+$  and  $\psi_-$  are both real, show that there is one and only one axis upon which the projection of  $\vec{\mu}$  has a definite, positive value, and find the angle between that axis and the  $z$  axis in terms of  $\psi_+$  and  $\psi_-$ .
- What would change if you multiplied both  $\psi_+$  and  $\psi_-$  by the same phase factor (complex unit)?
- What would change if you multiplied  $\psi_+$  and  $\psi_-$  by different phase factors?

This problem invites the question “What if the ratio of  $\psi_+/\psi_-$  is not pure real?” When you study more quantum mechanics, you will find that in this case the axis upon which the projection of  $\vec{\mu}$  has a definite, positive value is not in the  $x$ - $z$  plane, but instead has a component in the  $y$  direction as well.

### 2.9 Addition of states

Some students in your class wonder “What does it mean to ‘add two quantal states’? You never add two classical states.” For their benefit you decide to write four sentences interpreting the equation

$$|\psi\rangle = a|z+\rangle + b|z-\rangle \quad (2.33)$$

describing why you *can* add quantal states but *can't* add classical states. Your four sentences should include a formula for the amplitude  $a$  in terms of the states  $|\psi\rangle$  and  $|z+\rangle$ .

### 2.10 Names of six states, in two bases

Write down the representations (the “names”) of the states  $|z+\rangle$ ,  $|z-\rangle$ ,  $|x+\rangle$ ,  $|x-\rangle$ ,  $|\theta+\rangle$ , and  $|\theta-\rangle$  in (a) the basis  $\{|z+\rangle, |z-\rangle\}$  and in (b) the basis  $\{|x+\rangle, |x-\rangle\}$ .

### 2.11 More peculiarities of states

Because a vector pointing down at angle  $\theta$  is the same as a vector pointing up at angle  $\theta - 180^\circ$ , I would expect that  $|\theta-\rangle = |(\theta - 180^\circ)+\rangle$ . Show that this expectation is false by uncovering the true relation between these two state vectors.

### 2.12 Translation matrix

(This problem requires background knowledge in the mathematics of matrix multiplication.)

Suppose that the representation of  $|\psi\rangle$  in the basis  $\{|z+\rangle, |z-\rangle\}$  is

$$\begin{pmatrix} \psi_+ \\ \psi_- \end{pmatrix} = \begin{pmatrix} \langle z+|\psi\rangle \\ \langle z-|\psi\rangle \end{pmatrix}.$$

The representation of  $|\psi\rangle$  in the basis  $\{|\theta+\rangle, |\theta-\rangle\}$  is just as good, and we call it

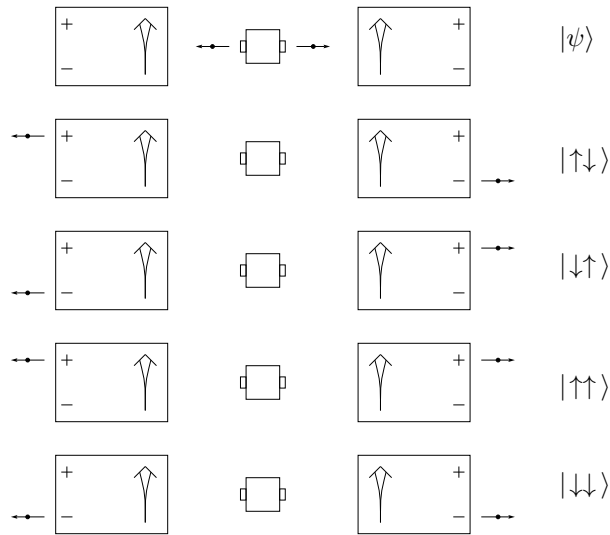
$$\begin{pmatrix} \psi'_+ \\ \psi'_- \end{pmatrix} = \begin{pmatrix} \langle \theta+|\psi\rangle \\ \langle \theta-|\psi\rangle \end{pmatrix}.$$

Show that you can “translate” between these two representations using the matrix multiplication

$$\begin{pmatrix} \psi'_+ \\ \psi'_- \end{pmatrix} = \begin{pmatrix} \cos(\theta/2) & \sin(\theta/2) \\ -\sin(\theta/2) & \cos(\theta/2) \end{pmatrix} \begin{pmatrix} \psi_+ \\ \psi_- \end{pmatrix}.$$

## 2.6 States for entangled systems

In the Einstein-Podolsky-Rosen experiment (1) on page 38, with two vertical analyzers, the initial state is represented by  $|\psi\rangle$ , and various possible final states are represented by  $|\uparrow\downarrow\rangle$  and so forth, as shown below. (In this section all analyzers will be vertical, so we adopt the oft-used convention that writes  $|z+\rangle$  as  $|\uparrow\rangle$  and  $|z-\rangle$  as  $|\downarrow\rangle$ .)



The experimental results tell us that

$$\begin{aligned}
 |\langle\uparrow\downarrow|\psi\rangle|^2 &= \frac{1}{2} \\
 |\langle\downarrow\uparrow|\psi\rangle|^2 &= \frac{1}{2} \\
 |\langle\uparrow\uparrow|\psi\rangle|^2 &= 0 \\
 |\langle\downarrow\downarrow|\psi\rangle|^2 &= 0.
 \end{aligned}
 \tag{2.34}$$

Additional analysis (sketched in problem 15.??, “Normalization of singlet spin state”) is needed to assign phases to these amplitudes. The results are

$$\begin{aligned}
 \langle\uparrow\downarrow|\psi\rangle &= +\frac{1}{\sqrt{2}} \\
 \langle\downarrow\uparrow|\psi\rangle &= -\frac{1}{\sqrt{2}} \\
 \langle\uparrow\uparrow|\psi\rangle &= 0 \\
 \langle\downarrow\downarrow|\psi\rangle &= 0.
 \end{aligned}
 \tag{2.35}$$

Using the generalization of equation (2.30) for a four-state basis, these results tell us that

$$\begin{aligned} |\psi\rangle &= |\uparrow\downarrow\rangle\langle\uparrow\downarrow|\psi\rangle + |\downarrow\uparrow\rangle\langle\downarrow\uparrow|\psi\rangle + |\uparrow\uparrow\rangle\langle\uparrow\uparrow|\psi\rangle + |\downarrow\downarrow\rangle\langle\downarrow\downarrow|\psi\rangle \\ &= \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle). \end{aligned} \quad (2.36)$$

A simple derivation, with profound implications.

### 2.6.1 *State pertains to system, not to atom*

In this entangled situation there is no such thing as an “amplitude for the right atom to exit from the + port,” because the probability for the right atom to exit from the + port depends on whether the left atom exits the + or the – port. The pair of atoms has a state, but the right atom by itself *doesn’t have a state*, in the same way that an atom passing through an interferometer *doesn’t have a position* and that love *doesn’t have a color*.

Leonard Susskind<sup>7</sup> puts it this way: If entangled states existed in auto mechanics as well as quantum mechanics, then an auto mechanic might tell you “I know everything about your car but ... I can’t tell you anything about any of its parts.”

### 2.6.2 *“Collapse of the state vector”*

Set up this EPR experiment with the left analyzer 100 kilometers from the source, and the right analyzer 101 kilometers from the source. As soon as the left atom comes out of its – port, then it is known that the right atom will come out if its + port. The system is no longer in the entangled state  $\frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$ ; instead the left atom is in state  $|\downarrow\rangle$  and the right atom is in state  $|\uparrow\rangle$ . The state of the right atom has changed (some say it has “collapsed”) despite the fact that it is 200 kilometers from the left analyzer that did the state changing!

This fact disturbs those who hold the misconception that states are physical things located out in space like nitrogen molecules, because it seems that information about state has made an instantaneous jump across 200 kilometers. In fact no information has been transferred from left to right: true, Alice at the left interferometer knows that the right atom will

<sup>7</sup>Leonard Susskind and Art Friedman, *Quantum Mechanics: The Theoretical Minimum* (Basic Books, New York, 2014) page xii.



exit the + port 201 kilometers away, but Bob at the right interferometer *doesn't* have this information and won't unless she tells him in some conventional, light-speed-or-slower fashion.<sup>8</sup>

If Alice could in some magical way manipulate her atom to ensure that it would exit the – port, then she *could* send a message instantaneously. But Alice does not possess magic, so she cannot manipulate the left-bound atom in this way. Neither Alice, nor Bob, *nor even the left-bound atom itself* knows from which port it will exit. Neither Alice, nor Bob, *nor even the left-bound atom itself* can influence from which port it will exit.

---

<sup>8</sup>If you are familiar with gauges in electrodynamics, you will find quantal state similar to the Coulomb gauge. In the Coulomb gauge, the electric potential at a point in space changes the instant that any charged particle moves, regardless of how far away that charged particle is. This does not imply that information moves instantly, because electric potential by itself is not measurable. The same applies for quantal state.

### 2.6.3 Measurement and entanglement

Back in section 1.4, “Light on the atoms” (page 33), we discussed the character of “observation” or “measurement” in quantum mechanics. Let’s bring our new machinery concerning quantal states to bear on this situation.

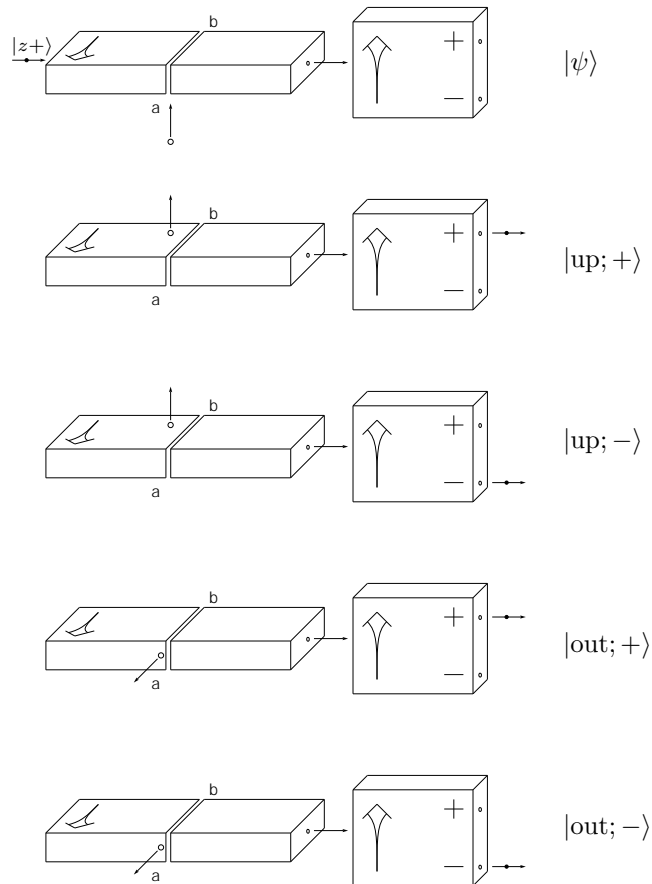
The figure on the next page shows, in the top panel, a potential measurement about to happen. An atom (represented by a black dot) in state  $|z+\rangle$  approaches a horizontal interferometer at the same time that a photon (represented by a white dot) approaches path **a** of that interferometer.

We employ a simplified model in which the photon either misses the atom, in which case it continues undeflected upward, or else the photon interacts with the atom, in which case it is deflected outward from the page. In this model there are four possible outcomes, shown in the bottom four panels of the figure.

After this potential measurement, the system of photon plus atom is in an entangled state: the states shown on the right must list *both* the condition of the photon (“up” or “out”) *and* the condition of the atom (+ or –).

If the photon misses the atom, then the atom must emerge from the + port of the analyzer: there is zero probability that the system has final state  $|\text{up}; -\rangle$ . But if the photon interacts with the atom, then the atom might emerge from either port: there is non-zero probability that the system has final state  $|\text{out}; -\rangle$ . These two states are exactly the same as far as the atom is concerned; they differ only in the position of the photon.

If we focus only on the atom, we would say that something strange has happened (a “measurement” at path **a**) that enabled the atom to emerge from the – port which (in the absence of “measurement”) that atom would never do. But if we focus on the entire system of photon plus atom, then it is an issue of entanglement, not of measurement.



**Problem**

**2.13 Amplitudes for “Measurement and entanglement”**

Suppose that, in the “simplified model” for measurement and entanglement, the probability for photon deflection is  $\frac{1}{5}$ . Find the four probabilities  $|\langle \text{up}; + | \psi \rangle|^2$ ,  $|\langle \text{up}; - | \psi \rangle|^2$ ,  $|\langle \text{out}; + | \psi \rangle|^2$ , and  $|\langle \text{out}; - | \psi \rangle|^2$ .

## 2.7 Are states “real”?

This is a philosophical question for which there is no specific meaning and hence no specific answer. But in my opinion, states are mathematical tools that enable us to efficiently and accurately calculate the probabilities that can be found through repeated analyzer experiments, interference experiments, and indeed all experiments. They are not physically “real”.

Indeed, it is possible to formulate quantum mechanics in such a way that probabilities and amplitudes are found without using the mathematical tool of “state” at all. Richard Feynman and Albert Hibbs do just this in their 1965 book *Quantum Mechanics and Path Integrals*. States do not make an appearance until deep into their book, and even when they do appear they are not essential. The Feynman “sum over histories” formulation described in that book is, for me, the most intuitively appealing approach to quantum mechanics. There is, however, a price to be paid for this appeal: it’s very difficult to work problems in the Feynman formulation.

## 2.8 What is a qubit?

At the end of the last chapter (on page 52) we listed several so-called “two-state systems” or “spin- $\frac{1}{2}$  systems” or “qubit systems”. You might have found these terms strange: There are an infinite number of states for the magnetic moment of a silver atom:  $|z+\rangle$ ,  $|1^\circ+\rangle$ ,  $|2^\circ+\rangle$ , and so forth. Where does the name “two-state system” come from? You now see the answer: it’s short for “two-basis-state system”.

The term “spin” originated in the 1920s when it was thought that an electron was a classical charged rigid sphere that created a magnetic moment through spinning about an axis. A residual of that history is that people still call<sup>9</sup> the state  $|z+\rangle$  by the name “spin up” and by the symbol  $|\uparrow\rangle$ , and the state  $|z-\rangle$  by “spin down” and  $|\downarrow\rangle$ . (Sometimes the association is made in the opposite way.) Meanwhile the state  $|x+\rangle$  is given the name “spin sideways” and the symbol  $|\rightarrow\rangle$ .

<sup>9</sup>The very most precise and pedantic people restrict the term “spin” to elementary particles, such as electrons and neutrinos. For composite systems like the silver atom they speak instead of “the total angular momentum  $\vec{J}$  of the silver atom in its ground state, projected on a given axis, and divided by  $\hbar$ .” For me, the payoff in precision is not worth the penalty in polysyllables.

Today, two-basis-state systems are more often called “qubit” systems from the term used in quantum information processing. In a classical computer, like the ones we use today, a bit of information can be represented physically by a patch of magnetic material on a disk: the patch magnetized “up” is interpreted as a 1, the patch magnetized “down” is interpreted as a 0. Those are the only two possibilities. In a quantum computer, a qubit of information can be represented physically by the magnetic moment of a silver atom: the atom in state  $|z+\rangle$  is interpreted as  $|1\rangle$ , the atom in state  $|z-\rangle$  is interpreted as  $|0\rangle$ . But the atom might be in any (normalized) superposition  $a|1\rangle + b|0\rangle$ , so rather than two possibilities there are an infinite number.

Furthermore, qubits can interfere with and become entangled with other qubits, options that are simply unavailable to classical bits. With more states, and more ways to interact, quantum computers can only be faster than classical computers, and even as I write these possibilities are being explored.

In today’s state of technology, quantum computers are hard to build, and they may never live up to their promise. But maybe they will.

Chapters 1 and 2 have focused on two-basis-state systems, but of course nature provides other systems as well. For example, the magnetic moment of a nitrogen atom (mentioned on page 9) is a “four-basis-state” system, where one basis is

$$|z; +2\rangle, \quad |z; +1\rangle, \quad |z; -1\rangle, \quad |z; -2\rangle. \quad (2.37)$$

And chapter 6 shifts our focus to a system with an infinite number of basis states.

### ***Problem***

#### 2.14 **Questions** (recommended problem)

Update your list of quantum mechanics questions that you started at problem 1.12 on page 53. Write down new questions and, if you have uncovered answers to any of your old questions, write them down briefly.

[[For example, one of my questions would be: “I’d like to see a proof that the global phase freedom mentioned on page 73, which obviously changes the amplitudes computed, does not change any experimentally accessible result.”]]



## Chapter 3

# Refining Mathematical Tools

### 3.1 Extras

#### Summing states

We have said that “if we determine the amplitude  $\langle\sigma|\psi\rangle$  for every possible state  $|\sigma\rangle$ , then we know everything about  $|\psi\rangle$  that there is to know.” So, for example, if two particular states  $|\psi_1\rangle$  and  $|\psi_2\rangle$  have the same amplitudes  $\langle\sigma|\psi_1\rangle = \langle\sigma|\psi_2\rangle$  for every state  $|\sigma\rangle$ , then the two states must be the same:  $|\psi_1\rangle = |\psi_2\rangle$ . In short, we can just erase the leading  $\langle\sigma|$ s from both sides.

Apply this idea to a more elaborate equation like the interference result

$$\langle\sigma|\psi\rangle = \langle\sigma|\theta+\rangle\langle\theta+|\psi\rangle + \langle\sigma|\theta-\rangle\langle\theta-|\psi\rangle. \quad (3.1)$$

Since this holds for every state  $|\sigma\rangle$ , we can erase the leading  $\langle\sigma|$ s to find

$$|\psi\rangle = |\theta+\rangle\langle\theta+|\psi\rangle + |\theta-\rangle\langle\theta-|\psi\rangle. \quad (3.2)$$

What is this supposed to mean? We have never before summed two quantal states! Physically, its meaning is nothing new: it just says if an atom ambivates through the two branches of a  $\theta$ -interferometer, it exits in the same state ( $|\psi\rangle$ ) that it entered. An atom entering in state  $|\theta+\rangle$  goes via branch a, and exits in state  $|\theta+\rangle$ . An atom entering in state  $|\theta-\rangle$  goes via branch b, and exits in state  $|\theta-\rangle$ . An atom entering in state  $|\psi\rangle$  goes via branch a with amplitude  $\langle\theta+|\psi\rangle$  and via branch b with amplitude  $\langle\theta-|\psi\rangle$ , and exits in state  $|\psi\rangle$ .

Mathematically, however, equation (3.2) implies some new kind of addition, an addition of states. It's easiest to flesh out the mathematical meaning by replacing the states  $|\psi\rangle$ ,  $|\theta+\rangle$ , and  $|\theta-\rangle$  with their representations as

column matrices in the  $\{|z+\rangle, |z-\rangle\}$  basis. The interference equation (3.2) above becomes

$$\begin{pmatrix} \langle z+|\psi\rangle \\ \langle z-|\psi\rangle \end{pmatrix} = \begin{pmatrix} \langle z+|\theta+\rangle \\ \langle z-|\theta+\rangle \end{pmatrix} \langle \theta+|\psi\rangle + \begin{pmatrix} \langle z+|\theta-\rangle \\ \langle z-|\theta-\rangle \end{pmatrix} \langle \theta-|\psi\rangle. \quad (3.3)$$

The first line of this equation is just equation (3.1) with  $|\sigma\rangle = |z+\rangle$ , while the second line is just equation (3.1) with  $|\sigma\rangle = |z-\rangle$ .

We have introduced the idea of adding states through interference experiments, and then forged the mathematical tools to describe those experiments, including state addition. It is possible, however, to go the other direction, and just talk about summing two states such as

$$\alpha|a\rangle + \beta|b\rangle. \quad (3.4)$$

*Exercise:* Show that for the above sum to produce a state, the identity

$$|\alpha|^2 + |\beta|^2 + 2\Re\{\alpha^*\beta\langle a|b\rangle\} = 1$$

must apply.

When looked at in this way — mathematics first, experiments afterwards — the concept of state summation and the phenomenon of interference are often called “superposition”.

The whole idea of summing states is new. In Newtonian mechanics, two states are never summed. That is because summing of states represents the phenomenon of interference, and in classical mechanics interference doesn’t happen.

When we learned how to add vectors, we learned to add them both geometrically (by setting them tail to head and drawing a vector from the first tail to the last head) and through components. The same holds for adding quantal states: You can add them physically, through interference experiments, or through components.

The equation

$$\vec{r} = \hat{i}r_x + \hat{j}r_y + \hat{k}r_z = \hat{i}(\hat{i} \cdot \vec{r}) + \hat{j}(\hat{j} \cdot \vec{r}) + \hat{k}(\hat{k} \cdot \vec{r})$$

for geometrical vectors is useful and familiar. The equation

$$|\psi\rangle = |\theta+\rangle\langle\theta+|\psi\rangle + |\theta-\rangle\langle\theta-|\psi\rangle.$$

for state vectors is just as useful and will soon be just as familiar.



*Alternative approach — sum of three or more states:*

Consider the three path experiment at equation (2.7) and think about how you would describe the situation if you walked into the experiment exactly when the atom was ambivating through the middle of the apparatus. If the initial state were  $|\psi\rangle$ , you would say the atom has amplitude  $\langle z+|\psi\rangle$  of ambivating through the top path (1a), amplitude  $\langle\theta+|z-\rangle\langle z-|\psi\rangle$  of ambivating through the middle path (2a), and amplitude  $\langle\theta-|z-\rangle\langle z-|\psi\rangle$  of ambivating through the bottom path (2b). Since an atom in the top path has state  $|z+\rangle$ , an atom in the middle path has state  $|\theta+\rangle$ , and an atom in the bottom path has state  $|\theta-\rangle$ , we say that the atom in state  $|\psi\rangle$  has been dissected into the three states through

$$\begin{aligned} |\psi\rangle &= |z+\rangle\langle z+|\psi\rangle \\ &\quad + |\theta+\rangle\langle\theta+|z-\rangle\langle z-|\psi\rangle \\ &\quad + |\theta-\rangle\langle\theta-|z-\rangle\langle z-|\psi\rangle \end{aligned} \tag{3.5}$$

Now we know what it means to add *three* states — not necessarily basis states. By using combinations of more and more interferometers, we could build up devices that would sum more and more states. The mathematics of state summation is a formal expression of the physics of interference.

**Analysis of states into basis states**

For all states  $|\psi\rangle$ ,  $|\phi\rangle$ , and any basis  $|a\rangle$ ,  $|b\rangle$ :

$$\langle\phi|\psi\rangle = \langle\phi|a\rangle\langle a|\psi\rangle + \langle\phi|b\rangle\langle b|\psi\rangle.$$

This equation is exactly the interference experiment: If the atom goes through both branch  $a$  and through branch  $b$ , then it emerges unchanged.

In particular, we write this equation for  $|\phi\rangle = |z+\rangle$  and then for  $|\phi\rangle = |z-\rangle$ :

$$\begin{pmatrix} \langle z+|\psi\rangle \\ \langle z-|\psi\rangle \end{pmatrix} = \begin{pmatrix} \langle z+|a\rangle \\ \langle z-|a\rangle \end{pmatrix} \langle a|\psi\rangle + \begin{pmatrix} \langle z+|b\rangle \\ \langle z-|b\rangle \end{pmatrix} \langle b|\psi\rangle.$$

This equation is the representation, in the  $\{|z+\rangle, |z-\rangle\}$  basis, of

$$|\psi\rangle = |a\rangle\langle a|\psi\rangle + |b\rangle\langle b|\psi\rangle. \tag{3.6}$$

What is this equation supposed to mean? The quantities  $|\psi\rangle$ ,  $|a\rangle$ , and  $|b\rangle$  represent states, the quantities  $\langle a|\psi\rangle$  and  $\langle b|\psi\rangle$  represent complex numbers. This is the first time we have ever added states. What does it mean? It means nothing more nor less than the interferometer experiment. (Just

as cavemen were able to spear mammoths through vector addition, even though they didn't know about basis states or the coordinate representations of vectors in a particular basis, so atoms are able to interfere through the above equation, even though they don't know about bases or representations. The representations make it easier to work with position vectors or with quantal states, but they aren't required.)

Reinforcing the meaning of this equation by writing it down for the basis  $\{|\theta+\rangle, |\theta-\rangle\}$ : When I write

$$|\psi\rangle = |\theta+\rangle\langle\theta+|\psi\rangle + |\theta-\rangle\langle\theta-|\psi\rangle,$$

I mean that if an atom passes through a  $\theta$ -analyzer, it has amplitude  $\langle\theta+|\psi\rangle$  to behave like an atom in state  $|\theta+\rangle$ , and it has amplitude  $\langle\theta-|\psi\rangle$  to behave like an atom in state  $|\theta-\rangle$ .

There's an easier way to derive the equation (3.6). If two states  $|\alpha\rangle$  and  $|\beta\rangle$  have the same amplitudes

$$\langle\phi|\alpha\rangle = \langle\phi|\beta\rangle$$

for all states  $|\phi\rangle$ , then they are the same state:

$$|\alpha\rangle = |\beta\rangle.$$

(Two atoms that behave identically under *all* circumstances are in the same state.) In other words, because the state  $|\phi\rangle$  is arbitrary, we can erase the symbol  $\langle\phi|$  from both sides of the equation. Applying this principle to

$$\langle\phi|\psi\rangle = \langle\phi|a\rangle\langle a|\psi\rangle + \langle\phi|b\rangle\langle b|\psi\rangle,$$

where the state  $|\phi\rangle$  is arbitrary, results in

$$|\psi\rangle = |a\rangle\langle a|\psi\rangle + |b\rangle\langle b|\psi\rangle.$$

This equation is analogous to the relation for position vectors that

$$\vec{r} = \hat{i}x + \hat{j}y + \hat{k}z.$$

Example: If the basis  $\{|a\rangle, |b\rangle\}$  is  $\{|z+\rangle, |z-\rangle\}$ , then

$$|\psi\rangle = |z+\rangle\langle z+|\psi\rangle + |z-\rangle\langle z-|\psi\rangle.$$

Using the representations for  $|z+\rangle$  and  $|z-\rangle$  in the  $\{|z+\rangle, |z-\rangle\}$  basis, this equation becomes

$$|\psi\rangle \doteq \begin{pmatrix} 1 \\ 0 \end{pmatrix} \langle z+|\psi\rangle + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \langle z-|\psi\rangle = \begin{pmatrix} \langle z+|\psi\rangle \\ \langle z-|\psi\rangle \end{pmatrix}.$$

### Change of basis

Suppose the two amplitudes  $\langle z+|\psi\rangle$  and  $\langle z-|\psi\rangle$  are known. Then we can easily find the amplitudes  $\langle\theta+|\psi\rangle$  and  $\langle\theta-|\psi\rangle$ , for any value of  $\theta$ , through

$$\begin{aligned}\langle\theta+|\psi\rangle &= \langle\theta+|z+\rangle\langle z+|\psi\rangle + \langle\theta+|z-\rangle\langle z-|\psi\rangle \\ \langle\theta-|\psi\rangle &= \langle\theta-|z+\rangle\langle z+|\psi\rangle + \langle\theta-|z-\rangle\langle z-|\psi\rangle\end{aligned}$$

These two equations might seem arcane, but in fact each one just represents the interference experiment performed with a vertical analyzer: The state  $|\psi\rangle$  is unaltered if the atom travels through the two branches of an vertical interferometer, that is via the upper  $z+$  branch *and* the lower  $z-$  branch. And if the state is unaltered then the amplitude to go to state  $|\theta+\rangle$  is of course also unaltered.

The pair of equations is most conveniently written as a matrix equation

$$\begin{pmatrix} \langle\theta+|\psi\rangle \\ \langle\theta-|\psi\rangle \end{pmatrix} = \begin{pmatrix} \langle\theta+|z+\rangle & \langle\theta+|z-\rangle \\ \langle\theta-|z+\rangle & \langle\theta-|z-\rangle \end{pmatrix} \begin{pmatrix} \langle z+|\psi\rangle \\ \langle z-|\psi\rangle \end{pmatrix}.$$

The  $2 \times 1$  column matrix on the right side is called the representation of state  $|\psi\rangle$  in the basis  $\{|z+\rangle, |z-\rangle\}$ . The  $2 \times 1$  column matrix on the left side is called the representation of state  $|\psi\rangle$  in the basis  $\{|\theta+\rangle, |\theta-\rangle\}$ . The square  $2 \times 2$  matrix is independent of the state  $|\psi\rangle$ , and depends only on the geometrical relationship between the initial basis  $\{|z+\rangle, |z-\rangle\}$  and the final basis  $\{|\theta+\rangle, |\theta-\rangle\}$ :

$$\begin{pmatrix} \langle\theta+|z+\rangle & \langle\theta+|z-\rangle \\ \langle\theta-|z+\rangle & \langle\theta-|z-\rangle \end{pmatrix} = \begin{pmatrix} \cos(\theta/2) & \sin(\theta/2) \\ -\sin(\theta/2) & \cos(\theta/2) \end{pmatrix}.$$

### Terms concerning quantum states

For atoms in state  $|z+\rangle$ , the probability of measuring  $\mu_\theta$  and finding  $\mu_\theta = +\mu_B$  is  $\cos^2(\theta/2)$ . We say “The projection probability from  $|z+\rangle$  to  $|\theta+\rangle$  is  $\cos^2(\theta/2)$ .” This situation is frequently, but incorrectly, described as “The probability that an atom in state  $|z+\rangle$  is in state  $|\theta+\rangle$  is  $\cos^2(\theta/2)$ .”

If the projection probability from  $|A\rangle$  to  $|B\rangle$  is zero, and vice versa, the two states are *orthogonal*. (For example,  $|z+\rangle$  and  $|z-\rangle$  are orthogonal, whereas  $|z+\rangle$  and  $|x-\rangle$  are not.)

Given a set of states  $\{|A\rangle, |B\rangle, \dots, |N\rangle\}$ , this set is said to be *complete* if an atom in *any* state is analyzed into one state of this set. In other words, it is complete if

$$\sum_{i=A}^N (\text{projection probability from any given state to } |i\rangle) = 1.$$

(For example, the set  $\{|\theta+\rangle, |\theta-\rangle\}$  is complete.)

### General definition of basis

We say that a set of states  $\{|a\rangle, |b\rangle, \dots, |n\rangle\}$  is a *basis* if both of the following apply:

- An atom in *any* state is analyzed into one element of this set. That is, for any state  $|\psi\rangle$

$$|\langle a|\psi\rangle|^2 + |\langle b|\psi\rangle|^2 + \dots + |\langle n|\psi\rangle|^2 = 1. \quad (3.7)$$

- There is zero amplitude for one element to be another element. That is

$$\begin{aligned} \langle a|b\rangle = 0, \langle a|c\rangle = 0, \dots, \langle a|n\rangle = 0, \\ \langle b|c\rangle = 0, \dots, \langle b|n\rangle = 0, \\ \text{etc.} \end{aligned} \quad (3.8)$$

For example, the set  $\{|\theta+\rangle, |\theta-\rangle\}$  is a basis for any value of  $\theta$ . The set  $\{|z+\rangle, |x-\rangle\}$  is not a basis.

## 3.2 Outer products, operators, measurement

### The outer product

Let's go back to our equation that represents the interference experiment: For any states  $|\phi\rangle$  and  $|\psi\rangle$ , and for any pair of basis states  $|a\rangle$  and  $|b\rangle$ ,

$$\langle \phi|\psi\rangle = \langle \phi|a\rangle\langle a|\psi\rangle + \langle \phi|b\rangle\langle b|\psi\rangle.$$

Now effect the divorce of amplitudes into inner product of states:

$$\langle \phi|\psi\rangle = \langle \phi| \left\{ |a\rangle\langle a| + |b\rangle\langle b| \right\} |\psi\rangle. \quad (3.9)$$

Our question: What's that thing between curly brackets?

In any particular basis,  $|a\rangle$  is represented by a  $2 \times 1$  column matrix, while  $\langle a|$  is represented by a  $1 \times 2$  row matrix. Thus the product  $|a\rangle\langle a|$  is represented by a  $2 \times 2$  square matrix. Similarly for  $|b\rangle\langle b|$ . Thus, in any particular basis, the thing between curly brackets is represented by a  $2 \times 2$  matrix.

If this confuses you, then think of it this way. If

$$|\alpha\rangle \doteq \begin{pmatrix} \alpha_a \\ \alpha_b \end{pmatrix} \quad \text{and} \quad |\beta\rangle \doteq \begin{pmatrix} \beta_a \\ \beta_b \end{pmatrix},$$

then

$$\langle\alpha| \doteq (\alpha_a^* \ \alpha_b^*) \quad \text{and} \quad \langle\beta| \doteq (\beta_a^* \ \beta_b^*).$$

Thus the “inner product” is the  $1 \times 1$  matrix

$$\langle\alpha|\beta\rangle = (\alpha_a^* \ \alpha_b^*) \begin{pmatrix} \beta_a \\ \beta_b \end{pmatrix} = \alpha_a^* \beta_a + \alpha_b^* \beta_b,$$

while the “outer product” is represented by the  $2 \times 2$  matrix

$$|\alpha\rangle\langle\beta| \doteq \begin{pmatrix} \alpha_a \\ \alpha_b \end{pmatrix} (\beta_a^* \ \beta_b^*) = \begin{pmatrix} \alpha_a \beta_a^* & \alpha_a \beta_b^* \\ \alpha_b \beta_a^* & \alpha_b \beta_b^* \end{pmatrix}.$$

A piece of terminology:  $|\alpha\rangle\langle\beta|$  is called an *operator* and the square matrix that represents it in a particular basis is called a *matrix*. The two terms are often used interchangeably, but if you care to make the distinction then this is how to make it. It’s conventional to symbolize operators with hats, like  $\hat{A}$ .

With these ideas in place, we see what’s inside the curly brackets of expression (3.9) — it’s the identity operator

$$\hat{1} = |a\rangle\langle a| + |b\rangle\langle b|,$$

and this holds true for any basis  $\{|a\rangle, |b\rangle\}$ .

We check this out two ways. First, in the basis  $\{|z+\rangle, |z-\rangle\}$ , we find the representation for the operator

$$|z+\rangle\langle z+| + |z-\rangle\langle z-|.$$

Remember that in this basis

$$|z+\rangle \doteq \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{while} \quad |z-\rangle \doteq \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

so

$$|z+\rangle\langle z+| \doteq \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1 \ 0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \quad (3.10)$$

Meanwhile

$$|z-\rangle\langle z-| \doteq \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0 \ 1) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \quad (3.11)$$

Thus

$$|z+\rangle\langle z+| + |z-\rangle\langle z-| \doteq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Yes! As required, this combination is the identity matrix, which is of course the representation of the identity operator.

For our second check, in the basis  $\{|z+\rangle, |z-\rangle\}$  we find the representation for the operator

$$|\theta+\rangle\langle\theta+| + |\theta-\rangle\langle\theta-|.$$

Remember that in this basis

$$|\theta+\rangle \doteq \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} \quad \text{while} \quad |\theta-\rangle \doteq \begin{pmatrix} -\sin(\theta/2) \\ \cos(\theta/2) \end{pmatrix},$$

so

$$\begin{aligned} |\theta+\rangle\langle\theta+| &\doteq \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} \begin{pmatrix} \cos(\theta/2) & \sin(\theta/2) \end{pmatrix} & (3.12) \\ &= \begin{pmatrix} \cos^2(\theta/2) & \cos(\theta/2)\sin(\theta/2) \\ \sin(\theta/2)\cos(\theta/2) & \sin^2(\theta/2) \end{pmatrix}. \end{aligned}$$

Meanwhile

$$\begin{aligned} |\theta-\rangle\langle\theta-| &\doteq \begin{pmatrix} -\sin(\theta/2) \\ \cos(\theta/2) \end{pmatrix} \begin{pmatrix} -\sin(\theta/2) & \cos(\theta/2) \end{pmatrix} & (3.13) \\ &= \begin{pmatrix} \sin^2(\theta/2) & -\sin(\theta/2)\cos(\theta/2) \\ -\cos(\theta/2)\sin(\theta/2) & \cos^2(\theta/2) \end{pmatrix}. \end{aligned}$$

(As a check, notice that when  $\theta = 0$ , equation (3.12) reduces to equation (3.10), and equation (3.13) reduces to equation (3.11).) Thus

$$\begin{aligned} |\theta+\rangle\langle\theta+| + |\theta-\rangle\langle\theta-| &\doteq \begin{pmatrix} \cos^2(\theta/2) & \cos(\theta/2)\sin(\theta/2) \\ \sin(\theta/2)\cos(\theta/2) & \sin^2(\theta/2) \end{pmatrix} \\ &\quad + \begin{pmatrix} \sin^2(\theta/2) & -\sin(\theta/2)\cos(\theta/2) \\ -\cos(\theta/2)\sin(\theta/2) & \cos^2(\theta/2) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

Yes! Once again this combination is the identity matrix.

### Measurement

What happens when an atom in state  $|\psi\rangle$  passes through a  $\theta$ -analyzer? Or, what is the same thing, what happens when an atom in state  $|\psi\rangle$  is measured to find the projection of  $\boldsymbol{\mu}$  on the  $\theta$  axis? (We call the projection of  $\boldsymbol{\mu}$  on the  $\theta$  axis  $\mu_\theta$ .)

The atom enters the analyzer in state  $|\psi\rangle$ . It has two possible fates:

- It emerges from the + port, in which case the atom has been measured to have  $\mu_\theta = +\mu_B$ , and it emerges in state  $|\theta+\rangle$ . This happens with probability  $|\langle\theta+|\psi\rangle|^2$ .
- It emerges from the - port, in which case the atom has been measured to have  $\mu_\theta = -\mu_B$ , and it emerges in state  $|\theta-\rangle$ . This happens with probability  $|\langle\theta-|\psi\rangle|^2$ .

What is the average value of  $\mu_\theta$ ?

$$\begin{aligned}
 \langle\mu_\theta\rangle &= (+\mu_B)|\langle\theta+|\psi\rangle|^2 + (-\mu_B)|\langle\theta-|\psi\rangle|^2 \\
 &= (+\mu_B)\langle\theta+|\psi\rangle^*\langle\theta+|\psi\rangle + (-\mu_B)\langle\theta-|\psi\rangle^*\langle\theta-|\psi\rangle \\
 &= (+\mu_B)\langle\psi|\theta+\rangle\langle\theta+|\psi\rangle + (-\mu_B)\langle\psi|\theta-\rangle\langle\theta-|\psi\rangle \\
 &= \langle\psi|\left\{(+\mu_B)|\theta+\rangle\langle\theta+| + (-\mu_B)|\theta-\rangle\langle\theta-|\right\}|\psi\rangle
 \end{aligned}$$

In the last line we have effected the divorce — writing amplitudes in terms of inner products between states.

Given the last line, it makes sense to define an operator associated with the measurement of  $\mu_\theta$ , namely

$$\hat{\mu}_\theta = (+\mu_B)|\theta+\rangle\langle\theta+| + (-\mu_B)|\theta-\rangle\langle\theta-|,$$

so that

$$\langle\mu_\theta\rangle = \langle\phi|\hat{\mu}_\theta|\phi\rangle.$$

Notice what we've done here: To find the average value of  $\mu_\theta$  for a particular atom, we've split up the problem into an operator  $\hat{\mu}_\theta$  involving only the measuring device and a state  $|\psi\rangle$  involving only the atomic state.

Example: What is the matrix representation of  $\hat{\mu}_\theta$  in the basis  $\{|z+\rangle, |z-\rangle\}$ ? We have already found representations for the outer product  $|\theta+\rangle\langle\theta+|$  in equation (3.12) and for the outer product  $|\theta-\rangle\langle\theta-|$  in equation (3.13). Using these expressions

$$\begin{aligned}
 \hat{\mu}_\theta &= (+\mu_B)|\theta+\rangle\langle\theta+| + (-\mu_B)|\theta-\rangle\langle\theta-| \\
 &\doteq (+\mu_B)\begin{pmatrix} \cos^2(\theta/2) & \cos(\theta/2)\sin(\theta/2) \\ \sin(\theta/2)\cos(\theta/2) & \sin^2(\theta/2) \end{pmatrix} \\
 &\quad + (-\mu_B)\begin{pmatrix} \sin^2(\theta/2) & -\sin(\theta/2)\cos(\theta/2) \\ -\cos(\theta/2)\sin(\theta/2) & \cos^2(\theta/2) \end{pmatrix} \\
 &= \mu_B\begin{pmatrix} \cos^2(\theta/2) - \sin^2(\theta/2) & 2\cos(\theta/2)\sin(\theta/2) \\ 2\cos(\theta/2)\sin(\theta/2) & \sin^2(\theta/2) - \cos^2(\theta/2) \end{pmatrix} \\
 &= \mu_B\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}
 \end{aligned}$$

where in the last line I have used the trigonometric half-angle formulas that everyone learned in high school and then forgot. (I forgot them too, but I know where to look them up.)

In particular, using the values  $\theta = 0$  and  $\theta = 90^\circ$ ,

$$\hat{\mu}_z = \mu_B \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{and} \quad \hat{\mu}_x = \mu_B \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and furthermore

$$\hat{\mu}_\theta = \cos \theta \hat{\mu}_z + \sin \theta \hat{\mu}_x.$$

Which is convenient because the unit vector  $\hat{\mathbf{r}}$  in the direction of  $\theta$  is

$$\hat{\mathbf{r}} = \cos \theta \hat{\mathbf{k}} + \sin \theta \hat{\mathbf{i}}.$$

So, knowing the operator associated with a measurement, we can easily find the resulting average value for any given state when measured. But we often want to know more than the average. We want to know also the standard deviation. Indeed we would like to know everything about the measurement: the possible results, the probability of each result, the state the system will be in after the measurement is performed. Surprisingly, all this information is wrapped up within the measurement operator as well.

We know that there are only two states that have a definite value of  $\mu_\theta$ , namely  $|\theta+\rangle$  and  $|\theta-\rangle$ . How do these states behave when acted upon by the operator  $\hat{\mu}_\theta$ ?

$$\begin{aligned} \hat{\mu}_\theta |\theta+\rangle &= \{(+\mu_B)|\theta+\rangle\langle\theta+| + (-\mu_B)|\theta-\rangle\langle\theta-|\} |\theta+\rangle \\ &= (+\mu_B)|\theta+\rangle\langle\theta+|\theta+\rangle + (-\mu_B)|\theta-\rangle\langle\theta-|\theta+\rangle \\ &= (+\mu_B)|\theta+\rangle(1) + (-\mu_B)|\theta-\rangle(0) \\ &= (+\mu_B)|\theta+\rangle \end{aligned}$$

In other words, when the operator  $\hat{\mu}_\theta$  acts upon the state  $|\theta+\rangle$ , the result is  $(+\mu_B)$  times that same state  $|\theta+\rangle$  — and  $(+\mu_B)$  is exactly the result that we would always obtain if we measured  $\mu_\theta$  for an atom in state  $|\theta+\rangle$ ! A parallel result holds for  $|\theta-\rangle$ .

To convince you of how rare this phenomena is, let me apply the operator  $\hat{\mu}_\theta$  to some other state, say  $|z+\rangle$ . The result is

$$\begin{aligned} \hat{\mu}_\theta |z+\rangle &= \{(+\mu_B)|\theta+\rangle\langle\theta+| + (-\mu_B)|\theta-\rangle\langle\theta-|\} |z+\rangle \\ &= (+\mu_B)|\theta+\rangle\langle\theta+|z+\rangle + (-\mu_B)|\theta-\rangle\langle\theta-|z+\rangle \\ &= (+\mu_B)|\theta+\rangle(\cos(\theta/2)) + (-\mu_B)|\theta-\rangle(-\sin(\theta/2)). \end{aligned}$$



But

$$\begin{aligned} |\theta+\rangle &= |z+\rangle\langle z+|\theta+\rangle + |z-\rangle\langle z-|\theta+\rangle = |z+\rangle(\cos(\theta/2)) + |z-\rangle(\sin(\theta/2)) \\ |\theta-\rangle &= |z+\rangle\langle z+|\theta-\rangle + |z-\rangle\langle z-|\theta-\rangle = |z+\rangle(-\sin(\theta/2)) + |z-\rangle(\cos(\theta/2)), \end{aligned}$$

so

$$\begin{aligned} \hat{\mu}_\theta|z+\rangle &= (+\mu_B)|\theta+\rangle(\cos(\theta/2)) + (-\mu_B)|\theta-\rangle(-\sin(\theta/2)) \\ &= \mu_B[|z+\rangle(\cos^2(\theta/2) - \sin^2(\theta/2)) + |z-\rangle(2\cos(\theta/2)\sin(\theta/2))] \\ &= \mu_B[|z+\rangle\cos\theta + |z-\rangle\sin\theta], \end{aligned}$$

where in the last line I have again used the half-remembered half-angle formulas.

The upshot is that most of the time,  $\hat{\mu}_\theta$  acting upon  $|z+\rangle$  does *not* produce a number times  $|z+\rangle$  — most of the time it produces some combination of  $|z+\rangle$  and  $|z-\rangle$ . In fact the only case in which  $\hat{\mu}_\theta$  acting upon  $|z+\rangle$  produces a number times  $|z+\rangle$  is when  $\sin\theta = 0$ , that is when  $\theta = 0$  or when  $\theta = 180^\circ$ .

The states when  $\hat{\mu}_\theta$  acting upon  $|\psi\rangle$  produces a number times the original state  $|\psi\rangle$  are rare: they are called *eigenstates*. The associated numbers are called *eigenvalues*. We have found the two eigenstates of  $\hat{\mu}_\theta$ : they are  $|\theta+\rangle$  with eigenvalue  $+\mu_B$  and  $|\theta-\rangle$  with eigenvalue  $-\mu_B$ .

$$\begin{aligned} \hat{\mu}_\theta|\theta+\rangle &= (+\mu_B)|\theta+\rangle && \text{eigenstate } |\theta+\rangle \text{ with eigenvalue } +\mu_B \\ \hat{\mu}_\theta|\theta-\rangle &= (-\mu_B)|\theta-\rangle && \text{eigenstate } |\theta-\rangle \text{ with eigenvalue } -\mu_B \end{aligned}$$

The eigenstates are the states with definite values of  $\mu_\theta$ . And the eigenvalues are those values!

### Summary: The quantum theory of measurement

This summarizes the quantum theory of measurement as applied to the measurement of  $\mu$  projected onto the unit vector in the direction of  $\theta$ :

The operator  $\hat{\mu}_\theta$  has two eigenstates which constitute a complete and orthogonal basis:

$$\begin{aligned} &\text{state } |\theta+\rangle \text{ with eigenvalue } +\mu_B \\ &\text{state } |\theta-\rangle \text{ with eigenvalue } -\mu_B \end{aligned}$$

(a) If you measure  $\mu_\theta$  of an atom in an eigenstate of  $\hat{\mu}_\theta$ , then the number you measure will be the corresponding eigenvalue, and the atom will remain in that eigenstate.

(b) If you measure  $\mu_\theta$  of an atom in an arbitrary state  $|\psi\rangle$ , then the number you measure will be one of the two eigenvalues of  $\hat{\mu}_\theta$ : It will be  $+\mu_B$  with probability  $|\langle\theta+|\psi\rangle|^2$ , it will be  $-\mu_B$  with probability  $|\langle\theta-|\psi\rangle|^2$ . If the value measured was  $+\mu_B$ , then the atom will leave in state  $|\theta+\rangle$ , if the value measured was  $-\mu_B$ , then the atom will leave in state  $|\theta-\rangle$ .

*Exercise:* Show that (a) follows from (b).

### Are states and operators real?

This is a philosophical question for which there's no specific meaning and hence no specific answer. But in my opinion, states and operators are mathematical tools that enable us to efficiently and accurately calculate the probabilities that we find through repeated measurement experiments, interference experiments, and indeed all experiments. They are not “real”.

Indeed, it is possible to formulate quantum mechanics in such a way that probabilities and amplitudes are found without using the mathematical tools of “state” and “operator” at all. Richard Feynman and Albert Hibbs do just this in their book *Quantum Mechanics and Path Integrals*. States and operators do not make an appearance until deep into this book, and even when they do appear it is only to make a connection between Feynman's formulation and more traditional formulations of quantum mechanics — states and operators are not essential. In my opinion, this Feynman “sum over histories” formulation is the most intuitively appealing approach to quantum mechanics. There is, however, a price to be paid for this appeal: it's very difficult to work problems in the Feynman formulation.

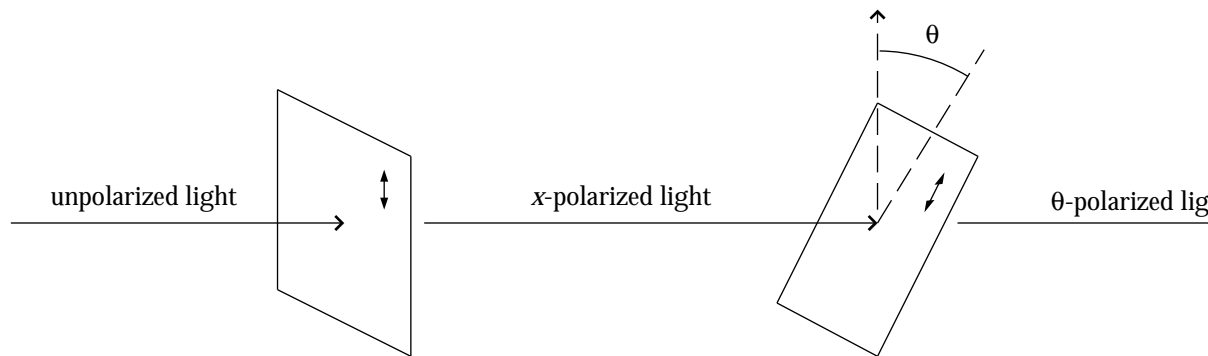
For more extensive discussion, see N. David Mermin, “What's bad about this habit?” *Physics Today*, **62** (5), May 2009, pages 8–9, and discussion about this essay in *Physics Today*, **62** (9), September 2009, pages 10–15.

### 3.3 Photon polarization

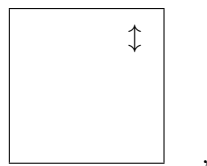
This book has developed the principles of quantum mechanics using a particular system, the magnetic moment of a silver atom, which has two basis states. Another system with two basis states is polarized light. I did not use this system mainly because photons are less familiar than atoms. These problems develop the quantum mechanics of photon polarization much as the text developed the quantum mechanics of magnetic moment.

One cautionary note: There is always a tendency to view the photon as a little bundle of electric and magnetic fields, a “wave packet” made up of these familiar vectors. This view is completely incorrect. In quantum electrodynamics, in fact, the electric field is a classical macroscopic quantity that takes on meaning only when a large number of photons are present.

### 3.15 Classical description of polarized light



When a beam of unpolarized light passes through an ideal polarizing sheet (represented by the symbol

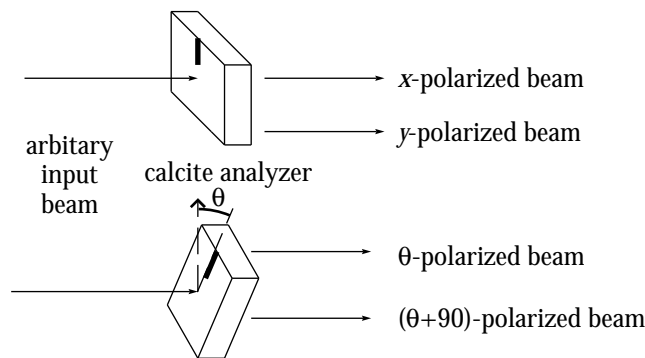


where the arrow shows the polarizing axis), the emerging beam is of lower intensity and it is “polarized”, that is, the electric field vector undulates but points only parallel or antiparallel to the polarizing axis. When a beam of vertically polarized light (an “ $x$ -polarized beam”) is passed through an ideal polarizing sheet with polarizing axis oriented at an angle  $\theta$  to the vertical, the beam is reduced in intensity and emerges with an electric field undulating parallel to the sheet’s polarizing axis (a “ $\theta$ -polarized beam”). The sheet performs these feats by absorbing any component of electric field perpendicular to its polarizing axis. Show

that if the incoming  $x$ -polarized beam has intensity  $I_0$ , then the outgoing  $\theta$ -polarized beam has intensity  $I_0 \cos^2 \theta$ . Show that this expression gives the expected results when  $\theta$  is  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$  or  $270^\circ$ .

### 3.16 Quantal description of polarized light: Analyzers

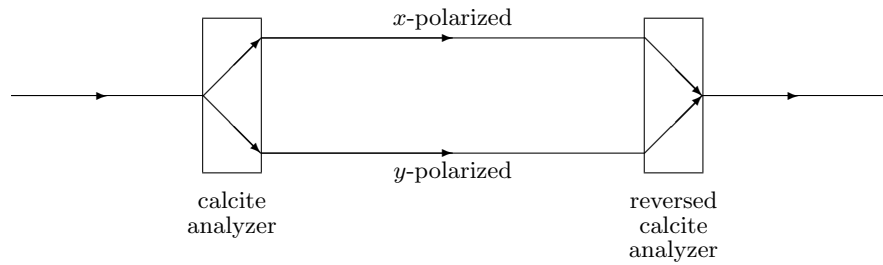
In quantum mechanics, a photon state is described by three quantities: energy, direction of motion, and polarization. We ignore the first two quantities for now. There are an infinite number of possible polarization states: the photons in an  $x$ -polarized beam are all in the  $|x\rangle$  state, the photons in a  $\theta$ -polarized beam ( $0^\circ \leq \theta < 180^\circ$ ) are all in the  $|\theta\rangle$  state, etc. In the quantum description, when an  $|x\rangle$  photon encounters a polarizing sheet oriented at an angle  $\theta$  to the vertical, then either it is absorbed (with probability  $\sin^2 \theta$ ) or else it emerges as a  $|\theta\rangle$  photon (with probability  $\cos^2 \theta$ ). A polarizing sheet is thus not an analyzer: whereas an analyzer would split the incident beam into two (or more) beams, the polarizing sheet absorbs one of the beams that an analyzer would emit. An analyzer can be constructed out of any material that exhibits double refraction. It is conventional to use a simple calcite crystal:



What are the probabilities  $|\langle x|\theta\rangle|^2$ ,  $|\langle x|\theta + 90^\circ\rangle|^2$ ? Design a pair of experiments to show that the states  $\{|\theta\rangle, |\theta + 90^\circ\rangle\}$  constitute a basis.

### 3.17 Interference

As usual, two analyzers (one inserted backwards) make up an analyzer loop.

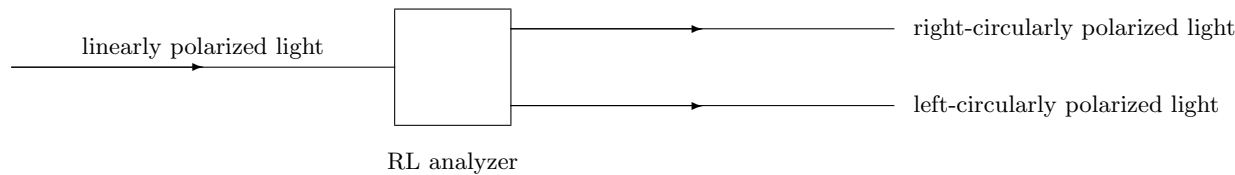


Invent a series of experiments that demonstrates quantum interference. (I used input photons in state  $|x\rangle$ , passed through an analyzer loop rotated at angle  $\theta$  to the vertical, followed by a vertical analyzer. But you might develop some other arrangement.) Show that the results of these experiments, plus the results of problem 2.16, are consistent with the amplitudes

$$\begin{aligned} \langle x|\theta\rangle &= \cos\theta & \langle x|\theta + 90^\circ\rangle &= -\sin\theta \\ \langle y|\theta\rangle &= \sin\theta & \langle y|\theta + 90^\circ\rangle &= \cos\theta. \end{aligned} \quad (3.14)$$

### 3.18 Circular polarization

Just as it is possible to analyze any light beam into  $x$ - and  $y$ -polarized beams, or into  $\theta$ - and  $\theta + 90^\circ$ -polarized beams, so it is possible to analyze any beam into right- and left-circularly polarized beams. You might remember from classical optics that any linearly polarized beam splits half-and-half into right- and left-circularly polarized light when so analyzed.



Quantum mechanics maintains that right- and left-circularly polarized beams are made up of photons in the  $|R\rangle$  and  $|L\rangle$  states, respectively. The projection amplitudes thus have magnitudes

$$\begin{aligned} |\langle R|\ell p\rangle| &= 1/\sqrt{2} \\ |\langle L|\ell p\rangle| &= 1/\sqrt{2} \end{aligned} \quad (3.15)$$

where  $|\ell p\rangle$  is any linearly polarized state. An RL analyzer loop is described through the equation

$$\langle\theta|R\rangle\langle R|x\rangle + \langle\theta|L\rangle\langle L|x\rangle = \langle\theta|x\rangle = \cos\theta. \quad (3.16)$$

Show that no real valued projection amplitudes can satisfy both relations (3.15) and (3.16), but that the complex values

$$\begin{aligned} \langle L|\theta\rangle &= e^{i\theta}/\sqrt{2} & \langle L|x\rangle &= 1/\sqrt{2} \\ \langle R|\theta\rangle &= e^{-i\theta}/\sqrt{2} & \langle R|x\rangle &= 1/\sqrt{2} \end{aligned} \quad (3.17)$$

are satisfactory!

### 3.4 Lightning linear algebra

Linear algebra provides many of the mathematical tools used in quantum mechanics. This section will scan through and summarize linear algebra to drive home the main points...it won't attempt to prove things or to develop the theory in the most elegant form using the smallest number of assumptions.

Scalars: either real numbers ( $x$ ) or complex numbers ( $z$ )

Vectors: notation  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$ , or  $|\psi\rangle$ ,  $|\phi\rangle$ ,  $|\chi\rangle$

In addition, there must be a rule for multiplying a vector by a scalar and a rule for adding vectors, so that  $\mathbf{a} + x\mathbf{b}$  is a vector.

I won't define "vector" any more than I defined "number". But I will give some examples:

arrows in  $N$ -dimensional space

$n$ -tuples, with real entries or with complex entries

polynomials

functions

$n \times m$  matrices

### Inner product

The “inner product” is a function from the ordered pairs of vectors to the scalars,

$$\text{I.P.}(\mathbf{a}, \mathbf{b}) = \text{a real or complex number}, \quad (3.18)$$

that satisfies

$$\text{I.P.}(\mathbf{a}, \mathbf{b} + \mathbf{c}) = \text{I.P.}(\mathbf{a}, \mathbf{b}) + \text{I.P.}(\mathbf{a}, \mathbf{c}) \quad (3.19)$$

$$\text{I.P.}(\mathbf{a}, z\mathbf{b}) = z\text{I.P.}(\mathbf{a}, \mathbf{b}) \quad (3.20)$$

$$\text{I.P.}(\mathbf{a}, \mathbf{b}) = [\text{I.P.}(\mathbf{b}, \mathbf{a})]^* \quad (3.21)$$

$$\text{I.P.}(\mathbf{a}, \mathbf{a}) > 0 \quad \text{unless } \mathbf{a} = \mathbf{0} \quad (3.22)$$

It follows from equation (3.21) that  $\text{I.P.}(\mathbf{a}, \mathbf{a})$  is real. Equation (3.22) demands also that it's positive.

Why is there a complex conjugation in equation (3.21)? Why not just demand that  $\text{I.P.}(\mathbf{a}, \mathbf{b}) = \text{I.P.}(\mathbf{b}, \mathbf{a})$ ? The complex conjugation is needed for consistency with (3.22). If it weren't there, then

$$\text{I.P.}(i\mathbf{a}, i\mathbf{a}) = (i \cdot i)\text{I.P.}(\mathbf{a}, \mathbf{a}) = -\text{I.P.}(\mathbf{a}, \mathbf{a}) < 0.$$

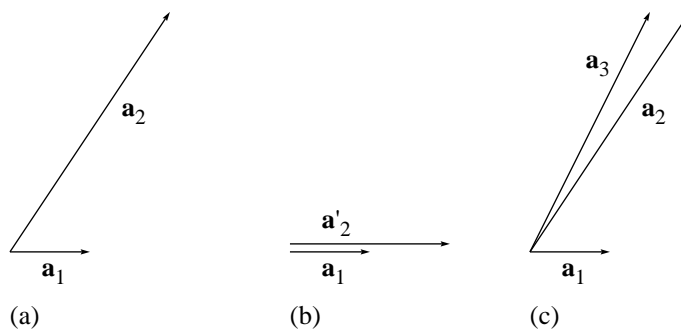
*Notation:*  $\text{I.P.}(\mathbf{a}, \mathbf{b}) = (\mathbf{a}, \mathbf{b}) = \mathbf{a} \cdot \mathbf{b}$ ,  $\text{I.P.}(|\phi\rangle, |\psi\rangle) = \langle \phi | \psi \rangle$ .

*Definition:* The norm of  $|\psi\rangle$  is  $\sqrt{\langle \psi | \psi \rangle}$ .

### Making new vectors from old

Given some vectors, say  $\mathbf{a}_1$  and  $\mathbf{a}_2$ , what vectors can you build from them using scalar multiplication and vector addition?

Example: arrows in the plane.



In (a), any arrow in the plane can be built out of  $\mathbf{a}_1$  and  $\mathbf{a}_2$ . In other words, any arrow in the plane can be written in the form  $\mathbf{r} = r_1\mathbf{a}_1 + r_2\mathbf{a}_2$ . We say that “the set  $\{\mathbf{a}_1, \mathbf{a}_2\}$  spans the plane”.

In (b), we cannot build the whole plane from  $\mathbf{a}_1$  and  $\mathbf{a}'_2$ . These two vectors do not span the plane.

In (c), the set  $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$  spans the plane, but the set is redundant: you don't *need* all three. You can build  $\mathbf{a}_3$  from  $\mathbf{a}_1$  and  $\mathbf{a}_2$ :  $\mathbf{a}_3 = \mathbf{a}_2 - \frac{1}{2}\mathbf{a}_1$ , so anything that can be built from  $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$  can also be built from  $\{\mathbf{a}_1, \mathbf{a}_2\}$ .

The set  $\{\mathbf{a}_1, \mathbf{a}_2\}$  is “linearly independent”, the set  $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$  is not. *Linearly independent*: You can't build any member of the set out of the other members.

So any arrow  $\mathbf{r}$  in the plane has a unique representation in terms of  $\{\mathbf{a}_1, \mathbf{a}_2\}$  but not in terms of  $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ . For example,

$$\begin{aligned}\mathbf{r} &= 2\mathbf{a}_3 = -1\mathbf{a}_1 + 2\mathbf{a}_2 + 0\mathbf{a}_3 \\ &= 0\mathbf{a}_1 + 0\mathbf{a}_2 + 2\mathbf{a}_3\end{aligned}$$

*Basis*: A spanning set of linearly independent vectors. (That is, a minimum set of building blocks from which any vector you want can be constructed. In any given basis, there is a unique representation for an arbitrary vector.)

It's easy to see that all bases have the same number of elements, and this number is called the dimensionality,  $N$ .

The easiest basis to work with is an *orthonormal basis*: A basis  $\{|1\rangle, |2\rangle, \dots, |N\rangle\}$  is orthonormal if

$$\langle n|m\rangle = \delta_{n,m}. \quad (3.23)$$

For any basis an arbitrary vector  $|\psi\rangle$  can be written

$$|\psi\rangle = \psi_1|1\rangle + \psi_2|2\rangle + \dots + \psi_N|N\rangle = \sum_{n=1}^N \psi_n|n\rangle, \quad (3.24)$$

but for many bases it's hard to find the coefficients  $\psi_n$ . For an orthonormal basis, however, it's easy. Take the inner product of basis element  $|m\rangle$  with  $|\psi\rangle$ , giving

$$\langle m|\psi\rangle = \sum_{n=1}^N \psi_n \langle m|n\rangle = \sum_{n=1}^N \psi_n \delta_{m,n} = \psi_m. \quad (3.25)$$



Thus the expansion (3.24) is

$$|\psi\rangle = \sum_{n=1}^N |n\rangle \langle n|\psi\rangle. \quad (3.26)$$

You have seen this formula in the context of arrows. For example, using two-dimensional arrows with the orthonormal basis  $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}\}$ , also called  $\{\mathbf{e}_x, \mathbf{e}_y\}$ , you know that

$$\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y,$$

where

$$x = \mathbf{r} \cdot \mathbf{e}_x \quad \text{and} \quad y = \mathbf{r} \cdot \mathbf{e}_y.$$

Thus

$$\mathbf{r} = \mathbf{e}_x(\mathbf{e}_x \cdot \mathbf{r}) + \mathbf{e}_y(\mathbf{e}_y \cdot \mathbf{r}),$$

which is just an instance of the more general expression (3.26).

### Representations

Any vector  $|\psi\rangle$  is completely specified by the  $N$  numbers  $\psi_1, \psi_2, \dots, \psi_N$  (that is, the  $N$  numbers  $\langle n|\psi\rangle$ ). We say that in the basis  $\{|1\rangle, |2\rangle, \dots, |N\rangle\}$ , the vector  $|\psi\rangle$  is *represented by* the column matrix

$$\begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_N \end{pmatrix} = \begin{pmatrix} \langle 1|\psi\rangle \\ \langle 2|\psi\rangle \\ \vdots \\ \langle N|\psi\rangle \end{pmatrix}. \quad (3.27)$$

It is very easy to manipulate vectors through their representations, so representations are used often. So often, that some people go overboard and say that the vector  $|\psi\rangle$  is *equal* to this column matrix. This is false. The matrix representation is a name for the vector, but is not equal to the vector — much as the word “tree” is a name for a tree, but is not the same as a tree. The symbol for “is represented by” is  $\doteq$ , so we write

$$|\psi\rangle \doteq \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_N \end{pmatrix} = \begin{pmatrix} \langle 1|\psi\rangle \\ \langle 2|\psi\rangle \\ \vdots \\ \langle N|\psi\rangle \end{pmatrix}. \quad (3.28)$$

What can we do with representations? Here's a way to connect an inner product, which is defined solely through the list of properties (3.19)–(3.22), to a formula in terms of representations.

$$\begin{aligned}
 & \langle \phi | \psi \rangle && \llbracket \text{using (3.26) ...} \rrbracket \\
 = & \langle \phi | \left\{ \sum_n |n\rangle \langle n| \psi \right\} \rangle && \llbracket \text{using (3.19) ...} \rrbracket \\
 = & \sum_n \langle \phi | n \rangle \langle n | \psi \rangle && \llbracket \text{using (3.21) ...} \rrbracket \\
 = & \sum_n \phi_n^* \psi_n \\
 = & (\phi_1^* \quad \phi_2^* \quad \cdots \quad \phi_N^*) \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_N \end{pmatrix}
 \end{aligned}$$

We will sometimes say that  $\langle \phi |$  is the “dual vector” to  $|\phi\rangle$  and is represented by the row matrix

$$(\phi_1^* \quad \phi_2^* \quad \cdots \quad \phi_N^*). \quad (3.29)$$

#### *Transformation of representations*

In the orthonormal basis  $\{|1\rangle, |2\rangle, \dots, |N\rangle\}$ , the vector  $|\psi\rangle$  is represented by

$$\begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_N \end{pmatrix}. \quad (3.30)$$

But in the different orthonormal basis  $\{|1'\rangle, |2'\rangle, \dots, |N'\rangle\}$ , the vector  $|\psi\rangle$  is represented by

$$\begin{pmatrix} \psi'_1 \\ \psi'_2 \\ \vdots \\ \psi'_N \end{pmatrix}. \quad (3.31)$$

How are these two representations related?

$$\begin{aligned}\psi'_n &= \langle n'|\psi\rangle \\ &= \langle n'|\left\{\sum_m |m\rangle\langle m|\psi\rangle\right\} \\ &= \sum_m \langle n'|m\rangle\langle m|\psi\rangle\end{aligned}$$

so

$$\begin{pmatrix} \psi'_1 \\ \psi'_2 \\ \vdots \\ \psi'_N \end{pmatrix} = \begin{pmatrix} \langle 1'|1\rangle & \langle 1'|2\rangle & \cdots & \langle 1'|N\rangle \\ \langle 2'|1\rangle & \langle 2'|2\rangle & \cdots & \langle 2'|N\rangle \\ \vdots & & & \vdots \\ \langle N'|1\rangle & \langle N'|2\rangle & \cdots & \langle N'|N\rangle \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_N \end{pmatrix}. \quad (3.32)$$

### Operators

A *linear operator*  $\hat{A}$  is a function from vectors to vectors

$$\hat{A} : |\psi\rangle \mapsto |\phi\rangle \quad \text{or in other words} \quad |\phi\rangle = \hat{A}|\psi\rangle, \quad (3.33)$$

with the property that

$$\hat{A}(z_1|\psi_1\rangle + z_2|\psi_2\rangle) = z_1\hat{A}|\psi_1\rangle + z_2\hat{A}|\psi_2\rangle. \quad (3.34)$$

If you know how  $\hat{A}$  acts upon each member of a basis set  $\{|1\rangle, |2\rangle, \dots, |N\rangle\}$ , then you know everything there is to know about  $\hat{A}$ , because for any vector  $|\psi\rangle$

$$\hat{A}|\psi\rangle = \hat{A}\left\{\sum_n \psi_n |n\rangle\right\} = \sum_n \psi_n \hat{A}|n\rangle, \quad (3.35)$$

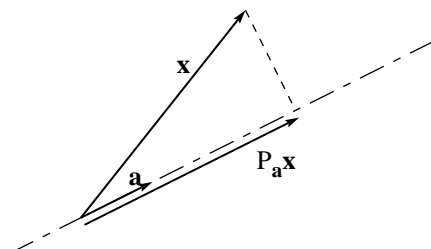
and the vectors  $\hat{A}|n\rangle$  are known.

Examples of linear operators:

- The identity operator:  $\hat{1}|\psi\rangle = |\psi\rangle$ .
- Rotations in the plane. (Linear because the sum of the rotated arrows is the same as the rotation of the summed arrows.)
- The “projection operator”  $\hat{P}_{|a\rangle}$ , defined in terms of some fixed vector  $|a\rangle$  as

$$\hat{P}_{|a\rangle} : |\psi\rangle \mapsto (\langle a|\psi\rangle) |a\rangle \quad (3.36)$$

This is often used for vectors  $|a\rangle$  of norm 1, in which case, for arrows in space, it looks like:



- More generally, for a given  $|a\rangle$  and  $|b\rangle$  the operator

$$\hat{S} : |\psi\rangle \mapsto (\langle b|\psi\rangle) |a\rangle \quad (3.37)$$

is linear.

Operators may not commute. That is, we might well have

$$\hat{A}_1 \hat{A}_2 |\psi\rangle \neq \hat{A}_2 \hat{A}_1 |\psi\rangle. \quad (3.38)$$

Non-linear operators also exist, such as  $\hat{N} : |\psi\rangle \mapsto (\langle \psi|\psi\rangle) |\psi\rangle$ , but are not so important in applications to quantum mechanics. [[This is the source of the name *linear* algebra. For non-linear operators, knowledge of the action of  $\hat{N}$  on the basis vectors is *not* sufficient to define the operator. It is a mystery why all the exact operators in quantum mechanics are linear.] ]

### Outer products

Recall the operator

$$\begin{aligned} \hat{S} : |\psi\rangle &\mapsto (\langle b|\psi\rangle) |a\rangle = |a\rangle \langle b|\psi\rangle \\ \hat{S}|\psi\rangle &= |a\rangle \langle b|\psi\rangle \end{aligned} \quad (3.39)$$

We will write the operator  $\hat{S}$  as  $|a\rangle \langle b|$  and call it “the outer product of  $|a\rangle$  and  $|b\rangle$ ”. This means neither more nor less than the defining equation (3.39).

For any orthonormal basis  $\{|1\rangle, |2\rangle, \dots, |N\rangle\}$ , consider the operator

$$\hat{T} \equiv |1\rangle \langle 1| + |2\rangle \langle 2| + \dots + |N\rangle \langle N|. \quad (3.40)$$

The effect of this operator on an arbitrary vector  $|\psi\rangle$  is given in equation (3.26), which shows that  $\hat{T}|\psi\rangle = |\psi\rangle$  for any  $|\psi\rangle$ . Hence my favorite equation

$$\hat{1} = \sum_n |n\rangle\langle n|. \quad (3.41)$$

This might look like magic, but it means nothing more than equation (3.26): that a vector may be resolved into its components. The operator of equation (3.41) simply represents the act of chopping a vector into its components and reassembling them. It is the mathematical representation of an analyzer loop!

#### Representations of operators

Operators are represented by  $N \times N$  matrices. If

$$|\phi\rangle = \hat{A}|\psi\rangle, \quad (3.42)$$

then

$$\begin{aligned} \langle n|\phi\rangle &= \langle n|\hat{A}|\phi\rangle \\ &= \langle n|\hat{A}\hat{1}|\phi\rangle \\ &= \langle n|\hat{A} \left\{ \sum_m |m\rangle\langle m| \right\} |\phi\rangle \\ &= \sum_m \langle n|\hat{A}|m\rangle\langle m|\psi\rangle, \end{aligned} \quad (3.43)$$

or, in matrix form,

$$\begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{pmatrix} = \begin{pmatrix} \langle 1|\hat{A}|1\rangle & \langle 1|\hat{A}|2\rangle & \cdots & \langle 1|\hat{A}|N\rangle \\ \langle 2|\hat{A}|1\rangle & \langle 2|\hat{A}|2\rangle & \cdots & \langle 2|\hat{A}|N\rangle \\ \vdots & \vdots & & \vdots \\ \langle N|\hat{A}|1\rangle & \langle N|\hat{A}|2\rangle & \cdots & \langle N|\hat{A}|N\rangle \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_N \end{pmatrix}. \quad (3.44)$$

The matrix  $\mathbf{M}$  that represents operator  $\hat{A}$  in this particular basis has elements  $M_{n,m} = \langle n|\hat{A}|m\rangle$ .

In a different basis, the same operator  $\hat{A}$  will be represented by a different matrix. You can figure out for yourself how to transform the matrix representation of an operator in one basis into the matrix representation of that operator in a second basis. But it's not all that important to do so. Usually you work in the abstract operator notation until you've figured out the easiest basis to work with, and then work in only that basis.

*Unitary operators*

If the norm of  $\hat{U}|\psi\rangle$  equals the norm of  $|\psi\rangle$  for all  $|\psi\rangle$ , then  $\hat{U}$  should be called “norm preserving” but in fact is called “unitary”. The rotation operator is unitary.

*Hermitian conjugate*

For every operator  $\hat{A}$  there is a unique operator  $\hat{A}^\dagger$ , the “Hermitian<sup>1</sup> conjugate” (or “Hermitian adjoint”) of  $\hat{A}$  such that

$$\langle\psi|\hat{A}^\dagger|\phi\rangle = \langle\phi|\hat{A}|\psi\rangle^* \quad (3.45)$$

for all vectors  $|\psi\rangle$  and  $|\phi\rangle$ .

If the matrix elements for  $\hat{A}$  are  $M_{n,m}$ , then the matrix elements for  $\hat{A}^\dagger$  are  $K_{n,m} = M_{m,n}^*$ .

*Hermitian operators*

If  $\hat{H}^\dagger = \hat{H}$ , then  $\hat{H}$  is said to be *Hermitian*. Matrix representations of Hermitian operators have  $M_{n,m} = M_{m,n}^*$ . Hermitian operators are important in quantum mechanics because if an operator is to correspond to an observable, then that operator must be Hermitian.

*Theorem:* If  $\hat{H}$  is Hermitian, then: (a) All of its eigenvalues are real. (b) There is an orthonormal basis consisting of eigenvectors of  $\hat{H}$ .

*Corollaries:* If the orthonormal basis mentioned in (b) is  $\{|1\rangle, |2\rangle, \dots, |N\rangle\}$ , and  $\hat{H}|n\rangle = \lambda_n|n\rangle$ , then

$$\hat{H} = \lambda_1|1\rangle\langle 1| + \lambda_2|2\rangle\langle 2| + \dots + \lambda_N|N\rangle\langle N|. \quad (3.46)$$

The matrix representation of  $\hat{H}$  in this basis is diagonal:

$$\hat{H} \doteq \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \lambda_N \end{pmatrix}. \quad (3.47)$$

---

<sup>1</sup>Charles Hermite (1822-1901), French mathematician who contributed to number theory, orthogonal polynomials, elliptic functions, quadratic forms, and linear algebra. Teacher of Hadamard and Poincaré, father-in-law of Picard.

### 3.5 Problems

#### 3.19 Change of basis

The set  $\{|a\rangle, |b\rangle\}$  is an orthonormal basis.

- a. Show that the set  $\{|a'\rangle, |b'\rangle\}$ , where

$$|a'\rangle = +\cos\phi|a\rangle + \sin\phi|b\rangle$$

$$|b'\rangle = -\sin\phi|a\rangle + \cos\phi|b\rangle$$

is also an orthonormal basis. (The angle  $\phi$  is simply a parameter — it has no physical significance.)

- b. Write down the transformation matrix from the  $\{|a\rangle, |b\rangle\}$  basis representation to the  $\{|a'\rangle, |b'\rangle\}$  basis representation.

(If you suspect a change of basis is going to help you, but you're not sure how or why, this change often works, so it's a good one to try first. You can adjust  $\phi$  to any parameter you want, but it's been my experience that it is most often helpful when  $\phi = 45^\circ$ .)

#### 3.20 Change of representation, I

If the set  $\{|a\rangle, |b\rangle\}$  is an orthonormal basis, then the set  $\{|a'\rangle, |b'\rangle\}$ , where  $|a'\rangle = |b\rangle$  and  $|b'\rangle = |a\rangle$  is also an orthonormal basis — it's just a reordering of the original basis states. Find the transformation matrix. If state  $|\psi\rangle$  is represented in the  $\{|a\rangle, |b\rangle\}$  basis as

$$\begin{pmatrix} \psi_a \\ \psi_b \end{pmatrix},$$

then how is this state represented in the  $\{|a'\rangle, |b'\rangle\}$  basis?

#### 3.21 Change of representation, II

Same as the previous problem, but use  $|a'\rangle = i|a\rangle$  and  $|b'\rangle = -i|b\rangle$ .

#### 3.22 Inner product

You know that the inner product between two position unit vectors is the cosine of the angle between them. What is the inner product between the states  $|z+\rangle$  and  $|\theta+\rangle$ ? Does the geometrical interpretation hold?

#### 3.23 Outer product

Using the  $\{|z+\rangle, |z-\rangle\}$  basis representations

$$|\psi\rangle \doteq \begin{pmatrix} \psi_+ \\ \psi_- \end{pmatrix} \quad |\phi\rangle \doteq \begin{pmatrix} \phi_+ \\ \phi_- \end{pmatrix} \quad |\theta+\rangle \doteq \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} \quad |\theta-\rangle \doteq \begin{pmatrix} -\sin(\theta/2) \\ \cos(\theta/2) \end{pmatrix},$$

write representations for  $|\theta+\rangle\langle\theta+|$  and  $|\theta-\rangle\langle\theta-|$ , then for  $\langle\phi|\theta+\rangle\langle\theta+|\psi\rangle$  and  $\langle\phi|\theta-\rangle\langle\theta-|\psi\rangle$ , and finally verify that

$$\langle\phi|\psi\rangle = \langle\phi|\theta+\rangle\langle\theta+|\psi\rangle + \langle\phi|\theta-\rangle\langle\theta-|\psi\rangle.$$

### 3.24 Measurement operator

Write the representation of the  $\hat{\mu}_\theta$  operator

$$\hat{\mu}_\theta = (+\mu_B)|\theta+\rangle\langle\theta+| + (-\mu_B)|\theta-\rangle\langle\theta-|$$

in the  $\{|z+\rangle, |z-\rangle\}$  basis. Using this representation, verify that  $|\theta+\rangle$  and  $|\theta-\rangle$  are eigenvectors.

### 3.25 The trace

For any  $N \times N$  matrix  $A$  (with components  $a_{ij}$ ) the *trace* of  $A$  is defined by

$$\text{tr}(A) = \sum_{i=1}^N a_{ii}$$

Show that  $\text{tr}(AB) = \text{tr}(BA)$ , and hence that  $\text{tr}(ABCD) = \text{tr}(DABC) = \text{tr}(CDAB)$ , etc. (the so-called “cyclic invariance” of the trace). However, show that  $\text{tr}(ABC)$  does not generally equal  $\text{tr}(CBA)$  by constructing a counterexample. (Assume all matrices to be square.)

### 3.26 The outer product

Any two complex  $N$ -tuples can be multiplied to form an  $N \times N$  matrix as follows: (The star represents complex conjugation.)

$$x = (x_1 \ x_2 \ \dots \ x_N)$$

$$y = (y_1 \ y_2 \ \dots \ y_N)$$

$$x \otimes y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} (y_1^* \ y_2^* \ \dots \ y_N^*) = \begin{pmatrix} x_1 y_1^* & x_1 y_2^* & \dots & x_1 y_N^* \\ x_2 y_1^* & x_2 y_2^* & \dots & x_2 y_N^* \\ \vdots & \vdots & \ddots & \vdots \\ x_N y_1^* & x_N y_2^* & \dots & x_N y_N^* \end{pmatrix}.$$

This so-called “outer product” is quite different from the familiar “dot product” or “inner product”

$$x \cdot y = (x_1^* \ x_2^* \ \dots \ x_N^*) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = x_1^* y_1 + x_2^* y_2 + \dots + x_N^* y_N.$$

Write a formula for the  $i, j$  component of  $x \otimes y$  and use it to show that  $\text{tr}(y \otimes x) = x \cdot y$ .



### 3.27 Pauli matrix algebra

Three important matrices are the *Pauli matrices*:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

(Sometimes they are called  $\sigma_1, \sigma_2, \sigma_3$  and other times they are called  $\sigma_x, \sigma_y, \sigma_z$ .)

- a. Show that the four matrices  $\{I, \sigma_1, \sigma_2, \sigma_3\}$ , where

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

constitute a basis for the set of  $2 \times 2$  matrices, by showing that any matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

can be written as

$$A = z_0 I + z_1 \sigma_1 + z_2 \sigma_2 + z_3 \sigma_3.$$

Produce formulas for the  $z_i$  in terms of the  $a_{ij}$ .

- b. Show that

- i.  $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = I^2 = I$
- ii.  $\sigma_i \sigma_j = -\sigma_j \sigma_i$  for  $i \neq j$
- iii.  $\sigma_1 \sigma_2 = i \sigma_3$  (a)
- $\sigma_2 \sigma_3 = i \sigma_1$  (b)
- $\sigma_3 \sigma_1 = i \sigma_2$  (c)

Note: Equations (b) and (c) are called “cyclic permutations” of equation (a), because in each equation, the indices go in the order



and differ only by starting at different points in the “merry-go-round.”

- c. Show that for any complex numbers  $c_1, c_2, c_3$ ,

$$(c_1 \sigma_1 + c_2 \sigma_2 + c_3 \sigma_3)^2 = (c_1^2 + c_2^2 + c_3^2) I.$$

### 3.28 Diagonalizing the Pauli matrices

Find the eigenvalues and corresponding (normalized) eigenvectors for all three Pauli matrices.

### 3.29 Exponentiation of Pauli matrices

Define exponentiation of matrices through

$$e^M = \sum_{n=0}^{\infty} \frac{M^n}{n!}.$$

a. Show that

$$e^{z\sigma_i} = \cosh(z)I + \sinh(z)\sigma_i \quad \text{for } i = 1, 2, 3.$$

(Hint: Look up the series expansions of sinh and cosh.)

b. Show that

$$e^{(\sigma_1 + \sigma_3)} = \cosh(\sqrt{2})I + \frac{\sinh(\sqrt{2})}{\sqrt{2}}(\sigma_1 + \sigma_3).$$

c. Prove that  $e^{\sigma_1}e^{\sigma_3} \neq e^{(\sigma_1 + \sigma_3)}$ .

### 3.30 Hermitian operators

a. Show that if  $\hat{A}$  is a linear operator and  $(a, \hat{A}a)$  is real for all vectors  $a$ , then  $\hat{A}$  is Hermitian. (Hint: Employ the hypothesis with  $a = b + c$  and  $a = b + ic$ .)

b. Show that any operator of the form

$$\hat{A} = c_a|a\rangle\langle a| + c_b|b\rangle\langle b| + \cdots + c_z|z\rangle\langle z|,$$

where the  $c_n$  are real constants, is Hermitian.

c. You know that if an operator is Hermitian then all of its eigenvalues are real. Show that the converse is false by producing a counterexample. (Hint: Try a  $2 \times 2$  upper triangular matrix.)

### 3.31 Unitary operators

Show that all the eigenvalues of a unitary operator have square modulus unity.

### 3.32 Commutator algebra

Prove that

$$[\hat{A}, b\hat{B} + c\hat{C}] = b[\hat{A}, \hat{B}] + c[\hat{A}, \hat{C}]$$

$$[a\hat{A} + b\hat{B}, \hat{C}] = a[\hat{A}, \hat{C}] + b[\hat{B}, \hat{C}]$$

$$[\hat{A}, \hat{B}\hat{C}] = \hat{B}[\hat{A}, \hat{C}] + [\hat{A}, \hat{B}]\hat{C}$$

$$[\hat{A}\hat{B}, \hat{C}] = \hat{A}[\hat{B}, \hat{C}] + [\hat{A}, \hat{C}]\hat{B}$$

$$0 = [\hat{A}, [\hat{B}, \hat{C}]] + [\hat{B}, [\hat{C}, \hat{A}]] + [\hat{C}, [\hat{A}, \hat{B}]] \quad (\text{the "Jacobi identity"}).$$

## Chapter 4

# Formalism

Look at *A Quantum Mechanics Primer* by Daniel T. Gillespie, pages 1 through 70.

### 4.1 The role of formalism

We started off trying to follow the behavior of a silver atom as it passed through various magnetic fields, and we ended up with an elaborate mathematical structure of state vectors, Hilbert space, operators, and eigenstates. This is a good time to step back and focus, not on the formalism, but on what the formalism is good for: what it does, what it doesn't do, and why we should care. We do so by looking at a different mathematical formalism for a more familiar physical problem.

Here's the physical problem: Suppose I count out 178 marbles and put them in an empty bucket. Then I count out 252 more marbles and put them in the same bucket. How many marbles are in the bucket?

There are a number of ways to solve this problem. First, by experiment: One could actually count out and place the marbles, and then count the number of marbles in the bucket at the end of the process. Second, by addition using Arabic numerals, using the rules for addition of three-digit numbers ("carrying") that we all learned in elementary school. Third, by the trick of writing

$$178 + 252 = 180 + 250 = 430$$

which reduces the problem to two-digit addition. Fourth, by converting from Arabic numerals in base 10 (decimal) to Arabic numerals in base 8

(octal) and adding the octal numerals:

$$178_{(\text{dec})} + 252_{(\text{dec})} = 262_{(\text{oct})} + 374_{(\text{oct})} = 656_{(\text{oct})} = 430_{(\text{dec})}.$$

Fifth, by converting to Roman numerals and adding them using the Roman addition rules that are simple and direct, but that you probably didn't learn in elementary school. Sixth, by converting to Mayan numerals and adding them using rules that are, to you, even less familiar. If you think about it, you'll come up with other methods.

The formal processes of Arabic numeral addition, Roman numeral addition, and Mayan numeral addition are interesting only because they give the same result as the experimental method of counting out marbles. That is, these formal, mathematical processes matter only because they reflect something about the physical world. (It's clear that addition using decimal Arabic numerals is considerably easier — and cheaper — than actually doing the experiment. If you were trained in octal or Roman or Mayan numerals, then you'd also find executing those algorithms easier than doing the experiment.)

Does the algorithm of “carrying” tell us anything about addition? For example, does it help us understand what's going on when we count out the total number of marbles in the bucket at the end of the experiment? I would answer “no”. The algorithm of carrying tells us not about addition, but about how we represent numbers using Arabic numerals with decimal positional notation (“place value”). The “carry digits” are a convenient mathematical tool to help calculate the total number of marbles in the bucket. The amount of carrying involved differs depending upon whether the addition is performed in decimal or in octal. It is absurd to think that one could look into the bucket and identify which marbles were involved in the carry and which were not! Nevertheless, you can and should develop an intuition about whether or not a carry will be needed when performing a sum. Indeed, when we wrote  $178 + 252$  as  $180 + 250$ , we did so precisely to avoid a carry.

There are many ways to find the sum of two integers. These different methods differ in ease of use, in familiarity, in concreteness, in ability to generalize to negative, fractional, and imaginary numbers. So you might prefer one method to another. But you can't say that one method is right and another is wrong: the significance of the various methods is, in fact, that they all produce the same answer, and that that answer is the same as the number of marbles in the bucket at the end of the process.

As with marbles in a bucket, so with classical mechanics. You know several formalisms — several algorithms — for solving problems in classical mechanics: the Newtonian formalism, the Lagrangian formalism, the Hamiltonian formalism, Poisson brackets, etc. These formal, mathematical processes are significant only because they reflect something about the physical world.

The mathematical manipulations involved in solving a particular problem using Newton's force-oriented method differ dramatically from the mathematical manipulations involved in solving that same problem using Hamilton's energy-oriented method, but the two answers will always be the same. Just as one can convert integers from a representation as decimal Arabic numerals to a representation as octal Arabic numerals, or as Roman numerals, or as Mayan numerals, so one can add any constant to a Hamiltonian and obtain a different Hamiltonian that is just as good as the original. Poisson brackets don't actually exist out in nature — you can never perform an experiment to measure the numerical value of a Poisson bracket — but they are convenient mathematical tools that help us calculate the values of positions that we *can* measure.

Although Lagrangians, Hamiltonians, and Poisson brackets are features of the algorithm, not features of nature, it is nevertheless possible to develop intuition concerning Lagrangians, Hamiltonians, and Poisson brackets. You might call this “physical intuition” or you might call it “mathematical intuition” or “algorithmic intuition”. Regardless of what you call it, it's a valuable thing to learn.

These different methods for solving classical problems differ in ease of use, in familiarity, in concreteness, in ability to generalize to relativistic and quantal situations. So you might prefer one method to another. But you can't say that one method is right and another is wrong: the significance of the various methods is, in fact, that they all produce the same answer, and that that answer is the same as the classical behavior exhibited by the system in question.

As with marbles in a bucket, and as with classical mechanics, so with quantum mechanics. This chapter has developed an elegant and formidable formal apparatus representing quantal states as vectors in Hilbert space and experiments as operators in Hilbert space. This is not the only way of solving problems in quantum mechanics: One could go back to the fundamental rules for combining amplitudes in series and in parallel (page 58), just as

one could go back to solving arithmetic problems by throwing marbles into a bucket. Or one could develop more elaborate and more formal ways to solve quantum mechanics problems, just as one could use the Lagrangian or Hamiltonian formulations in classical mechanics. This book will not treat these alternative formulations of quantum mechanics: the path integral formulation (Feynman), the phase space formulation (Wigner), the density matrix formulation (for an introduction, see section 4.2), the variational formulation, the pilot wave formulation (de Broglie-Bohm), or any of the others. But be assured that these alternative formulations exist, and their existence proves that kets and operators are features of the algorithmic tools we use to solve quantum mechanical problems, not features of nature.

The mathematical manipulations involved in solving a particular problem using the Hilbert space formalism differ dramatically from the mathematical manipulations involved in solving that same problem using the rules for combining amplitudes in series and in parallel, but the two answers will always be the same. In almost all cases the Hilbert space formalism is far easier to apply, and that's why we use it. We use it so often that we can fall into the trap of thinking that kets and operators are features of nature, not features of an algorithm. But remember that just as one can convert integers from a representation as decimal Arabic numerals to a representation as octal Arabic numerals, or as Roman numerals, or as Mayan numerals, so one can multiply any state vector by a constant of modulus unity to obtain a different state vector that is just as good as the original. State vectors don't actually exist out in nature — you can never perform an experiment to measure the numerical value of a state vector (or even of an amplitude) — but they are convenient mathematical tools that help us calculate the values of probabilities that we *can* measure.

Many students, faced with the formidable mathematical formalism of quantum mechanics, fall into the trap of despair. “How can nature possibly be so sophisticated and formal?” This is the same trap as wondering “How can marbles know the algorithm for carrying in the addition of decimal Arabic numerals?” Nature doesn't know anything about Hilbert space, just as marbles don't know anything about carrying. The fact that the formalism of quantum mechanics is more sophisticated than the formalism of addition, or the formalism of classical mechanics, simply reflects the two facts (noted previously on pages XXX and YYY) that quantum mechanics is far removed from common sense, and that quantum mechanics is stupendously rich.

## 4.2 The density matrix

### 4.1 Definition

Consider a system in quantum state  $|\psi\rangle$ . Define the operator

$$\hat{\rho} = |\psi\rangle\langle\psi|,$$

called the *density matrix*, and show that the expectation value of the observable associated with operator  $\hat{A}$  in  $|\psi\rangle$  is

$$\text{tr}\{\hat{\rho}\hat{A}\}.$$

### 4.2 Statistical mechanics

Frequently physicists don't know exactly which quantum state their system is in. (For example, silver atoms coming out of an oven are in states of definite  $\mu$  projection, but there is no way to know which state any given atom is in.) In this case there are two different sources of measurement uncertainty: first, we don't know what state they system is in (statistical uncertainty, due to our ignorance) and second, even if we did know, we couldn't predict the result of every measurement (quantum uncertainty, due to the way the world works). The density matrix formalism neatly handles both kinds of uncertainty at once.

If the system could be in any of the states  $|a\rangle, |b\rangle, \dots, |i\rangle, \dots$  (not necessarily a basis set), and if it has probability  $p_i$  of being in state  $|i\rangle$ , then the density matrix

$$\hat{\rho} = \sum_i p_i |i\rangle\langle i|$$

is associated with the system. Show that the expectation value of the observable associated with  $\hat{A}$  is still given by

$$\text{tr}\{\hat{\rho}\hat{A}\}.$$

### 4.3 Trace of the density matrix

Show that  $\text{tr}\{\hat{\rho}\} = 1$ . (This can be either a long and tedious proof, or a short and insightful one.)





## Chapter 5

# Time Evolution

### 5.1 Operator for time evolution

You now are at the point in quantum mechanics where you were when you first stepped into the door of your classical mechanics classroom: you know what you're trying to calculate.

But! How to calculate it? If quantum mechanics is to have a classical limit, then quantal states have to change with time. We write this time dependence explicitly as

$$|\psi(t)\rangle. \quad (5.1)$$

We seek the equations that govern this time evolution, the ones parallel to the classical time development equations, be they the Newtonian equations

$$\sum \vec{F} = m\vec{a} \quad (5.2)$$

or the Lagrange equations

$$\frac{\partial L}{\partial q_i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} = 0 \quad (5.3)$$

or the Hamilton equations

$$\frac{\partial H}{\partial q_i} = -\dot{p}_i, \quad \frac{\partial H}{\partial p_i} = \dot{q}_i. \quad (5.4)$$

Assume the existence of some “time development operator”  $\hat{U}(\Delta t)$  such that

$$|\psi(t + \Delta t)\rangle = \hat{U}(\Delta t)|\psi(t)\rangle. \quad (5.5)$$

You might think that this statement is so general that we haven't assumed anything — we've just said that things are going to change with time. In

fact we've made a big assumption: just by our notation we've assumed that the time-development operator  $\hat{U}$  is linear, independent of the state  $|\psi\rangle$  that's evolving. That is, we've assumed that the *same* operator will time-evolve *any* different state. (The operator will, of course, depend on which system is evolving in time: the number of particles involved, their interactions, their masses, the value of the magnetic field in which they move, and so forth.)

By virtue of the meaning of time, we expect the operator  $\hat{U}(\Delta t)$  to have these four properties:

- (1)  $\hat{U}(\Delta t)$  is unitary.
- (2)  $\hat{U}(\Delta t_2)\hat{U}(\Delta t_1) = \hat{U}(\Delta t_2 + \Delta t_1)$ .
- (3)  $\hat{U}(\Delta t)$  is dimensionless.
- (4)  $\hat{U}(0) = \hat{1}$ .

And it's also reasonable to assume that the time-development operator can be expanded in a Taylor series:

$$\hat{U}(\Delta t) = \hat{U}(0) + \hat{A}\Delta t + \hat{B}(\Delta t)^2 + \dots \quad (5.6)$$

We know that  $\hat{U}(0) = \hat{1}$ , and we'll write the quadratic and higher-order terms as  $\hat{B}(\Delta t)^2 + \dots = \mathcal{O}(\Delta t^2) \dots$  which is read “terms of order  $\Delta t^2$  and higher” or just as “terms of order  $\Delta t^2$ ”. Finally, we'll write  $\hat{A}$  in a funny way so that

$$\hat{U}(\Delta t) = \hat{1} - \frac{i}{\hbar}\hat{H}\Delta t + \mathcal{O}(\Delta t^2). \quad (5.7)$$

I could just say, “we define  $\hat{H} = i\hbar\hat{A}$ ” but that just shunts aside the important question — why is this a useful definition? There are two reasons: First, the operator  $\hat{H}$  turns out to be Hermitian. (We will prove this in this section.) Second, because it's Hermitian, it can represent a measured quantity. When we investigate the classical limit, we will see that it corresponds to the classical energy.<sup>1</sup>

[[The energy operator is called “the Hamiltonian” and represented by the letter  $\hat{H}$  in honor of William Rowan Hamilton, who first pointed out the central role that energy can play in time development in the formal theory of classical mechanics. Hamilton (1805–1865) made important contributions to mathematics, optics, classical mechanics, and astronomy. At the age

<sup>1</sup>For now, you can just use dimensional analysis to see that it has the correct *dimensions* for energy.

of 22 years, while still an undergraduate, he was appointed professor of astronomy at his university and the Royal Astronomer of Ireland. He was not related to the American founding father Alexander Hamilton.]]

*Theorem:* The operator  $\hat{H}$  defined above is Hermitian.

*Proof:* The proof uses the fact that the norm of  $|\psi(t + \Delta t)\rangle$  equals the norm of  $|\psi(t)\rangle$ :

$$|\psi(t + \Delta t)\rangle = |\psi(t)\rangle - \frac{i}{\hbar} \Delta t \underbrace{\hat{H}|\psi(t)\rangle}_{\equiv |\psi_H(t)\rangle} + \mathcal{O}(\Delta t^2). \quad (5.8)$$

Thus

$$\begin{aligned} \langle \psi(t + \Delta t) | \psi(t + \Delta t) \rangle &= \left[ \langle \psi(t) | + \frac{i}{\hbar} \Delta t \langle \psi_H(t) | + \mathcal{O}(\Delta t^2) \right] \left[ |\psi(t)\rangle - \frac{i}{\hbar} \Delta t |\psi_H(t)\rangle + \mathcal{O}(\Delta t^2) \right] \\ &= \langle \psi(t) | \psi(t) \rangle + \frac{i}{\hbar} \Delta t \left[ \langle \psi_H(t) | \psi(t) \rangle - \langle \psi(t) | \psi_H(t) \rangle \right] + \mathcal{O}(\Delta t^2) \end{aligned} \quad (5.10)$$

$$1 = 1 + \frac{i}{\hbar} \Delta t \left[ \langle \psi(t) | \psi_H(t) \rangle^* - \langle \psi(t) | \psi_H(t) \rangle \right] + \mathcal{O}(\Delta t^2) \quad (5.11)$$

$$0 = \frac{i}{\hbar} \Delta t \left[ \langle \psi(t) | \hat{H} | \psi(t) \rangle^* - \langle \psi(t) | \hat{H} | \psi(t) \rangle \right] + \mathcal{O}(\Delta t^2). \quad (5.12)$$

This equation has to hold for all values of  $\Delta t$ , so the quantity in square brackets must vanish! That is,

$$\langle \psi(t) | \hat{H} | \psi(t) \rangle^* = \langle \psi(t) | \hat{H} | \psi(t) \rangle \quad (5.13)$$

for all vectors  $|\psi(t)\rangle$ . A simple exercise (problem 2.30, part a) then shows that operator  $\hat{H}$  is Hermitian. ■

We have written the time-development equation as

$$|\psi(t + \Delta t)\rangle = |\psi(t)\rangle - \frac{i}{\hbar} \Delta t \hat{H} |\psi(t)\rangle + \mathcal{O}(\Delta t^2). \quad (5.14)$$

Rearrangement gives

$$\frac{|\psi(t + \Delta t)\rangle - |\psi(t)\rangle}{\Delta t} = -\frac{i}{\hbar} \hat{H} |\psi(t)\rangle + \mathcal{O}(\Delta t). \quad (5.15)$$

In the limit  $\Delta t \rightarrow 0$ , this gives

$$\frac{d|\psi(t)\rangle}{dt} = -\frac{i}{\hbar} \hat{H} |\psi(t)\rangle \quad (5.16)$$

which is an important result known as the Schrödinger<sup>2</sup> equation!

<sup>2</sup>Erwin Schrödinger (1887–1961) was interested in physics, biology, philosophy, and Eastern religion. Born in Vienna, he held physics faculty positions in Germany, Poland, and Switzerland. In 1926 he discovered the time-development equation that now bears his name. This led, in 1927, to a prestigious appointment in Berlin. In 1933, disgusted with the Nazi regime, he left Berlin for Oxford, England. He held several positions in various cities before ending up in Dublin. There, in 1944, he wrote a book titled *What is Life?* which is widely credited for stimulating interest in what had been a backwater of science: biochemistry.

*Time evolution of projection probabilities*

*Theorem:* If  $|\phi\rangle$  is a time-independent state and  $\hat{P}_\phi = |\phi\rangle\langle\phi|$  is its associated projection operator, then

$$\frac{d}{dt}|\langle\phi|\psi(t)\rangle|^2 = -\frac{i}{\hbar}\langle[\hat{P}_\phi, \hat{H}]\rangle. \quad (5.17)$$

*Proof:*

$$\begin{aligned} \frac{d}{dt}|\langle\phi|\psi(t)\rangle|^2 &= \frac{d}{dt}[\langle\phi|\psi(t)\rangle\langle\phi|\psi(t)\rangle^*] \\ &= \left[\langle\phi|\frac{d}{dt}|\psi(t)\rangle\right]\langle\phi|\psi(t)\rangle^* + \langle\phi|\psi(t)\rangle\left[\langle\phi|\frac{d}{dt}|\psi(t)\rangle\right]^* \end{aligned}$$

But  $\langle\phi|\frac{d}{dt}|\psi(t)\rangle = -\frac{i}{\hbar}\langle\phi|\hat{H}|\psi(t)\rangle$ , so

$$\begin{aligned} \frac{d}{dt}|\langle\phi|\psi(t)\rangle|^2 &= -\frac{i}{\hbar}\left[\langle\phi|\hat{H}|\psi(t)\rangle\langle\phi|\psi(t)\rangle^* - \langle\phi|\psi(t)\rangle\langle\phi|\hat{H}|\psi(t)\rangle^*\right] \\ &= -\frac{i}{\hbar}\left[\langle\psi(t)|\phi\rangle\langle\phi|\hat{H}|\psi(t)\rangle - \langle\psi(t)|\hat{H}|\phi\rangle\langle\phi|\psi(t)\rangle\right] \\ &= -\frac{i}{\hbar}\left[\langle\psi(t)|\left\{|\phi\rangle\langle\phi|\hat{H} - \hat{H}|\phi\rangle\langle\phi|\right\}|\psi(t)\rangle\right] \\ &= -\frac{i}{\hbar}\langle\psi(t)|[\hat{P}_\phi, \hat{H}]|\psi(t)\rangle \quad \blacksquare \end{aligned}$$

*Lemma:* Suppose  $\hat{A}$  and  $\hat{B}$  are commuting Hermitian operators. If  $|a\rangle$  is an eigenvector of  $\hat{A}$  and  $\hat{P}_a = |a\rangle\langle a|$ , then  $[\hat{P}_a, \hat{B}] = 0$ .

*Proof:* From the compatibility theorem, there is an eigenbasis  $\{|b_n\rangle\}$  of  $\hat{B}$  with  $|b_1\rangle = |a\rangle$ . Write  $\hat{B}$  in diagonal form as

$$\hat{B} = \sum_n b_n |b_n\rangle\langle b_n|.$$

Then

$$\hat{B}|b_1\rangle\langle b_1| = \sum_n |b_n\rangle\langle b_n|b_1\rangle\langle b_1| = \sum_n |b_n\rangle\delta_{n,1}\langle b_1| = b_1|b_1\rangle\langle b_1|$$

while

$$|b_1\rangle\langle b_1|\hat{B} = \sum_n |b_1\rangle\langle b_1|b_n\rangle\langle b_n| = \sum_n |b_1\rangle\delta_{1,n}\langle b_n| = b_1|b_1\rangle\langle b_1|.$$

## 5.2 Working with the Schrödinger equation

Quantal states evolve according to the Schrödinger time-development equation

$$\frac{d}{dt}|\psi(t)\rangle = -\frac{i}{\hbar}\hat{H}|\psi(t)\rangle. \quad (5.18)$$

We have shown that the linear operator  $\hat{H}$  is Hermitian and has the dimensions of energy. I've stated that we are going to show, when we discuss the classical limit, that the operator  $\hat{H}$  corresponds to energy, and this justifies the name “Hamiltonian operator”. That's still not much knowledge! This is just as it was in classical mechanics: Time development is governed by

$$\sum \mathbf{F} = m\mathbf{a}, \quad (5.19)$$

but this doesn't help you until you know what forces are acting. Similarly, in quantum mechanics the Schrödinger equation is true but doesn't help us until we know how to find the Hamiltonian operator.

We find the Hamiltonian operator in quantum mechanics in the same way that we find the force function in classical mechanics: by appeal to experiment, to special cases, to thinking about the system and putting the pieces together. It's a creative task to stitch together the hints that we know to find a Hamiltonian. Sometimes in this book I'll be able to guide you down this creative path. Sometimes, as in great art, the creative process came through a stroke of genius that can only be admired and not explained.

### Representations of the Schrödinger equation

As usual, we become familiar with states through their components, that is through their representations in a particular basis:

$$|\psi(t)\rangle = \sum_n \psi_n |n\rangle. \quad (5.20)$$

We know that  $|\psi(t)\rangle$  changes with time on the left-hand side, so something has to change with time on the right-hand side. Which is it, the expansion coefficients  $\psi_n$  or the basis states  $|n\rangle$ ? The choice has nothing to do with nature — it is purely formal. All our experimental results will depend on  $|\psi(t)\rangle$ , and whether we ascribe the time development to the expansion coefficients or to the basis states is merely a matter of convenience. There are three common conventions, called “pictures”: In the “Schrödinger picture”, the expansion coefficients change with time while the basis states don't. In the “Heisenberg picture” the reverse is true. In the “interaction picture” both expansion coefficients and basis states change with time.

time constant	time dependent	name
$\{ n\rangle\}$	$\psi_n(t)$	Schrödinger picture
$\psi_n$	$\{ n(t)\rangle\}$	Heisenberg picture
nothing	$\psi_n(t), \{ n(t)\rangle\}$	interaction picture

This book will use the Schrödinger picture, but be aware that this is mere convention.

In the Schrödinger picture, the expansion coefficients  $\langle n|\psi(t)\rangle = \psi_n(t)$  change in time according to

$$\frac{d}{dt}\langle n|\psi(t)\rangle = -\frac{i}{\hbar}\langle n|\hat{H}|\psi(t)\rangle = -\frac{i}{\hbar}\sum_m \langle n|\hat{H}|m\rangle\langle m|\psi(t)\rangle, \quad (5.21)$$

or, in other words, according to

$$\frac{d\psi_n(t)}{dt} = -\frac{i}{\hbar}\sum_m H_{n,m}\psi_m(t) \quad \text{where, recall} \quad H_{n,m} = H_{m,n}^*. \quad (5.22)$$

### A system with one basis state

Consider a system with one basis state — say, a motionless hydrogen atom in its electronic ground state, which we call  $|1\rangle$ . Then

$$|\psi(t)\rangle = \psi_1(t)|1\rangle$$

If the initial state happens to be

$$|\psi(0)\rangle = |1\rangle,$$

then the time development problem is

$$\text{Initial condition: } \psi_1(0) = 1$$

$$\text{Differential equation: } \frac{d\psi_1(t)}{dt} = -\frac{i}{\hbar}E_g\psi_1(t),$$

where  $E_g = \langle 1|\hat{H}|1\rangle$  is the energy of the ground state.

The solution is straightforward:

$$\psi_1(t) = 1e^{-(i/\hbar)E_g t}$$

or, in other words,

$$|\psi(t)\rangle = e^{-(i/\hbar)E_g t}|1\rangle. \quad (5.23)$$

Because two state vectors that differ only in phase represent the same state, the *state* doesn't change even though the *coefficient*  $\psi_1(t)$  does change with time. The system says always in the ground state.

When I was in high school, my chemistry teacher said that “an atom is a pulsating blob of probability”. He was thinking of this equation, with the expansion coefficient  $\psi_1(t)$  changing in time as

$$e^{-(i/\hbar)E_g t} = \cos((E_g/\hbar)t) - i \sin((E_g/\hbar)t). \quad (5.24)$$

On one hand you know that this function “pulsates” — that is, changes in time periodically with period  $2\pi\hbar/E_g$ . On the other hand you know also that this function represents an irrelevant overall phase — for example, it has no effect on any probability at all. My high school chemistry teacher was going overboard in ascribing physical reality to the mathematical tools we use to describe reality.

*Exercise:* Change energy zero. You know the energy zero is purely conventional so changing the energy zero shouldn’t change anything in the physics. And indeed it changes only the phase, which is also purely conventional. In the words of my high school chemistry teacher this changes the “pulsation” rate — but it doesn’t change anything about the behavior of the hydrogen atom.

### A system with two basis states: The silver atom

Consider a system with two basis states — say, a silver atom in a uniform vertical magnetic field. Take the two basis states to be

$$|1\rangle = |z+\rangle \quad \text{and} \quad |2\rangle = |z-\rangle. \quad (5.25)$$

It’s very easy to write down the differential equation

$$\frac{d}{dt} \begin{pmatrix} \psi_1(t) \\ \psi_2(t) \end{pmatrix} = -\frac{i}{\hbar} \begin{pmatrix} H_{1,1} & H_{1,2} \\ H_{2,1} & H_{2,2} \end{pmatrix} \begin{pmatrix} \psi_1(t) \\ \psi_2(t) \end{pmatrix} \quad (5.26)$$

but it’s much harder to see what the elements in the Hamiltonian matrix should be — that is, it’s hard to guess the Hamiltonian operator.

The classical energy for this system is

$$U = -\boldsymbol{\mu} \cdot \mathbf{B} = -\mu_z B. \quad (5.27)$$

Our guess for the quantum Hamiltonian is simply to change quantities into operators

$$\hat{H} = -\hat{\mu}_z B \quad (5.28)$$

where

$$\hat{\mu}_z = (+\mu_B)|z+\rangle\langle z+| + (-\mu_B)|z-\rangle\langle z-| \quad (5.29)$$

is the quantum mechanical operator corresponding to the observable  $\mu_z$ . (See equation 3.2.) In this equation  $B$  is not an operator but simply a number, the magnitude of the classical magnetic field in which the silver atom is immersed. You might think that we should quantize the magnetic field as well as the atomic magnetic moment, and indeed a full quantum-mechanical treatment would have to include the quantum theory of electricity and magnetism. That's a task for later. For now, we'll accept the Hamiltonian (5.28) as a reasonable starting point, and indeed it turns out to describe this system to high accuracy, although not perfectly.<sup>3</sup>

It is an easy exercise to show that in the basis

$$\{|z+\rangle, |z-\rangle\} = \{|1\rangle, |2\rangle\},$$

the Hamiltonian operator (5.28) is represented by the matrix

$$\begin{pmatrix} H_{1,1} & H_{1,2} \\ H_{2,1} & H_{2,2} \end{pmatrix} = \begin{pmatrix} -\mu_B B & 0 \\ 0 & +\mu_B B \end{pmatrix}. \quad (5.30)$$

Thus the differential equations (5.26) become

$$\begin{aligned} \frac{d\psi_1(t)}{dt} &= -\frac{i}{\hbar}(-\mu_B B)\psi_1(t) \\ \frac{d\psi_2(t)}{dt} &= -\frac{i}{\hbar}(+\mu_B B)\psi_2(t). \end{aligned}$$

The solutions are straightforward:

$$\begin{aligned} \psi_1(t) &= \psi_1(0)e^{-(i/\hbar)(-\mu_B B)t} \\ \psi_2(t) &= \psi_2(0)e^{-(i/\hbar)(+\mu_B B)t}. \end{aligned}$$

Stuff about initial state  $|z+\rangle$ .

Suppose the initial state is

$$|x+\rangle = |z+\rangle\langle z+|x+\rangle + |z-\rangle\langle z-|x+\rangle = |z+\rangle\frac{1}{\sqrt{2}} + |z-\rangle\frac{1}{\sqrt{2}}.$$

Then

$$\psi_1(0) = \frac{1}{\sqrt{2}} \quad \psi_2(0) = \frac{1}{\sqrt{2}}$$

so

$$|\psi(t)\rangle = \frac{1}{\sqrt{2}}e^{-(i/\hbar)(-\mu_B B)t}|z+\rangle + \frac{1}{\sqrt{2}}e^{-(i/\hbar)(+\mu_B B)t}|z-\rangle.$$

---

<sup>3</sup>If you want perfection, you'll need to go into some discipline other than science.



So the atom is produced in state  $|x+\rangle$ , then is exposed to a vertical magnetic field for time  $t$ , and ends up in the state mentioned above. If we now measure  $\mu_x$ , what is the probability that it's  $+\mu_B$  again? That probability is the square of the amplitude

$$\begin{aligned} \langle x+ | \psi(t) \rangle &= \frac{1}{\sqrt{2}} e^{-(i/\hbar)(-\mu_B B)t} \langle x+ | z+\rangle + \frac{1}{\sqrt{2}} e^{-(i/\hbar)(+\mu_B B)t} \langle x+ | z-\rangle \\ &= \frac{1}{\sqrt{2}} e^{-(i/\hbar)(-\mu_B B)t} \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} e^{-(i/\hbar)(+\mu_B B)t} \frac{1}{\sqrt{2}} \\ &= \frac{1}{2} \left[ e^{-(i/\hbar)(-\mu_B B)t} + e^{-(i/\hbar)(+\mu_B B)t} \right] \\ &= \frac{1}{2} \left[ 2 \cos((1/\hbar)(\mu_B B)t) \right] \\ &= \cos\left(\frac{\mu_B B}{\hbar} t\right) \end{aligned}$$

The probability is

$$|\langle x+ | \psi(t) \rangle|^2 = \cos^2\left(\frac{\mu_B B}{\hbar} t\right) \quad (5.31)$$

which starts at one when  $t = 0$ , then goes down to zero, then goes back up to one, with an oscillation period of

$$\frac{\pi\hbar}{\mu_B B}.$$

This phenomena is called “Rabi oscillation” — it is responsible for the workings of atomic clocks.<sup>4</sup>

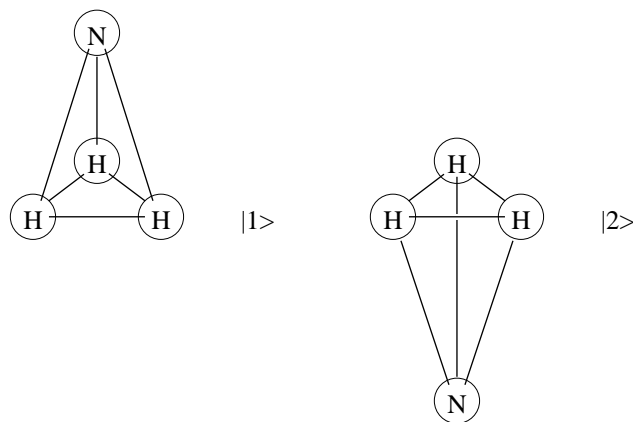
### Another two-state system: The ammonia molecule

Another system with two basis states is the ammonia molecule  $\text{NH}_3$ . If we ignore translation and rotation, and assume that the molecule is rigid,<sup>5</sup> then there are still two possible states for the molecule: state  $|1\rangle$  with the nitrogen atom pointing up, and state  $|2\rangle$  with the nitrogen atom

<sup>4</sup>Isidor Isaac Rabi (1898–1988) won the Nobel Prize for his discovery of nuclear magnetic resonance, but he also contributed to the invention of the laser and the atomic clock. No capsule biography suffices because he did so much. If you read Jeremy Bernstein’s profile of Rabi in the *New Yorker* (13 and 20 October 1975) you will see that he was a *very* clever man.

<sup>5</sup>That is, ignore vibration. These approximations seem, at first glance, to be absurd. They are in fact excellent approximations, because the tunneling happens so fast that the molecule doesn’t have time to translate, rotate, or vibrate to any significant extent during one cycle of tunneling.

pointing down. These are states of definite position for the nitrogen atom, but not states of definite energy (stationary states) because there is some amplitude for the nitrogen atom to tunnel from the “up” position to the “down” position. That is, if you start with the atom in state  $|1\rangle$ , then some time later it might be in state  $|2\rangle$ , because the nitrogen atom tunneled through the plane of hydrogen atoms.



What is the implication of such tunneling for the Hamiltonian matrix? The matrix we dealt with in equation (5.30) was diagonal, and hence the two differential equations split up (“decoupled”) into one involving  $\psi_1(t)$  and another involving  $\psi_2(t)$ . These were independent: If a system started out in the state  $|1\rangle$  (i.e.  $\psi_1(t) = e^{-(i/\hbar)H_{1,1}t}$ ,  $\psi_2(t) = 0$ ), then it stayed there forever. We’ve just said that this is *not* true for the ammonia molecule, so the Hamiltonian matrix must *not* be diagonal.

The Hamiltonian matrix in the  $\{|1\rangle, |2\rangle\}$  basis has the form

$$\begin{pmatrix} H_{1,1} & H_{1,2} \\ H_{2,1} & H_{2,2} \end{pmatrix} = \begin{pmatrix} E & Ae^{i\phi} \\ Ae^{-i\phi} & E \end{pmatrix}. \quad (5.32)$$

The two off-diagonal elements must be complex conjugates of each other because the matrix is Hermitian. It’s reasonable that the two on-diagonal elements are equal because the states  $|1\rangle$  and  $|2\rangle$  are mirror images and hence  $\langle 1|\hat{H}|1\rangle = \langle 2|\hat{H}|2\rangle$ .

For this Hamiltonian, the Schrödinger equation is

$$\frac{d}{dt} \begin{pmatrix} \psi_1(t) \\ \psi_2(t) \end{pmatrix} = -\frac{i}{\hbar} \begin{pmatrix} E & Ae^{i\phi} \\ Ae^{-i\phi} & E \end{pmatrix} \begin{pmatrix} \psi_1(t) \\ \psi_2(t) \end{pmatrix}. \quad (5.33)$$

It's hard to see how to solve this pair of differential equations. The matrix is not diagonal, so the differential equation for  $\psi_1(t)$  involves the unknown function  $\psi_2(t)$ , and the differential equation for  $\psi_2(t)$  involves the unknown function  $\psi_1(t)$ . However, while it's hard to solve in this initial basis, it would be easy to solve in a basis where the matrix is diagonal.

**To diagonalize an  $N \times N$  Hermitian matrix  $M$ :**

(1) In initial basis, the matrix representation of  $\hat{A}$  is  $M$ . The eigenvectors of  $\hat{A}$  satisfy  $\hat{A}|e_n\rangle = \lambda_n|e_n\rangle$ .

(2) Find  $N$  eigenvalues by solving the  $N^{\text{th}}$  order polynomial equation

$$\det |M - \lambda I| = 0.$$

(3) Find the representation  $\mathbf{e}_n$  of the eigenvector  $|e_n\rangle$  by solving  $N$  simultaneous linear equations

$$M\mathbf{e}_n = \lambda_n\mathbf{e}_n.$$

[In the above equation,  $M$  is an  $N \times N$  matrix,  $\mathbf{e}_n$  is an  $N \times 1$  matrix (the  $N$  unknowns), and  $\lambda_n$  is a known number (determined in the previous step).]

(4) In the basis  $\{|e_1\rangle, |e_2\rangle, \dots, |e_N\rangle\}$ , the matrix representation of  $\hat{A}$  is diagonal

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \lambda_N \end{pmatrix}.$$

Let's carry out these steps for the ammonia molecule problem.

1. The Hamiltonian is represented in the initial basis  $\{|1\rangle, |2\rangle\}$  by

$$M = \begin{pmatrix} E & Ae^{i\phi} \\ Ae^{-i\phi} & E \end{pmatrix}$$

2. Find the eigenvalues.

$$\det \begin{vmatrix} E - \lambda & Ae^{i\phi} \\ Ae^{-i\phi} & E - \lambda \end{vmatrix} = 0$$

$$(E - \lambda)^2 - A^2 = 0$$

$$(E - \lambda)^2 = A^2$$

$$E - \lambda = \pm A$$

$$\lambda = E \pm A$$

$$\lambda_1 = E + A \tag{5.34}$$

$$\lambda_2 = E - A \tag{5.35}$$

As required, the eigenvalues are real.

**3.** Find the eigenvectors.

We start with the eigenvector for  $\lambda_1 = E + A$ :

$$\begin{aligned} \mathbf{M}\mathbf{e}_1 &= \lambda_1\mathbf{e}_1 \\ (\mathbf{M} - \lambda_1\mathbf{I})\mathbf{e}_1 &= 0 \\ \begin{pmatrix} E - \lambda_1 & Ae^{i\phi} \\ Ae^{-i\phi} & E - \lambda_1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \begin{pmatrix} -A & Ae^{i\phi} \\ Ae^{-i\phi} & -A \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \begin{pmatrix} -1 & e^{i\phi} \\ e^{-i\phi} & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ -x + e^{i\phi}y &= 0 \\ e^{-i\phi}x - y &= 0 \end{aligned}$$

These two are not independent equations! They cannot be. There are many eigenvectors because if, say

$$\begin{pmatrix} 1 \\ 5 \end{pmatrix}$$

is an eigenvector, then so are

$$\begin{pmatrix} -1 \\ -5 \end{pmatrix}, \quad \begin{pmatrix} 2 \\ 10 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 3i \\ 15i \end{pmatrix},$$

and infinitely more eigenvectors.

The solution is  $y = e^{-i\phi}x$ , so

$$\mathbf{e}_1 = \begin{pmatrix} x \\ e^{-i\phi}x \end{pmatrix}.$$

Although I could choose any value of  $x$  that I wanted, it is most convenient to work with normalized eigenvectors, for which

$$\begin{aligned} |x|^2 + |y|^2 &= 1 \\ |x|^2 + |e^{-i\phi}x|^2 &= 1 \\ 2|x|^2 &= 1 \end{aligned}$$

This equation has many solutions. I could pick

$$x = \frac{1}{\sqrt{2}} \quad \text{or} \quad x = -\frac{1}{\sqrt{2}} \quad \text{or} \quad x = \frac{i}{\sqrt{2}} \quad \text{or} \quad x = \frac{1+i}{2}$$

but there's no advantage to picking a solution with all sorts of unneeded symbols. So I choose the first possibility and write

$$\mathbf{e}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ e^{-i\phi} \end{pmatrix}.$$

This is the representation of  $|e_1\rangle$  in the basis  $\{|1\rangle, |2\rangle\}$ .

*Exercise.* Show that an eigenvector associated with  $\lambda_2 = E - A$  is

$$|e_2\rangle \doteq \mathbf{e}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ e^{-i\phi} \end{pmatrix}.$$

*Exercise.* Verify that  $\langle e_1 | e_2 \rangle = 0$ .

In summary,

$$\begin{aligned} |e_1\rangle &= \frac{1}{\sqrt{2}} [ +|1\rangle + e^{-i\phi}|2\rangle ] \\ |e_2\rangle &= \frac{1}{\sqrt{2}} [ -|1\rangle + e^{-i\phi}|2\rangle ]. \end{aligned} \quad (5.36)$$

*Exercise.* Show that  $\{|e_1\rangle, |e_2\rangle\}$  constitute a spanning set by building  $|1\rangle$  and  $|2\rangle$  out of  $|e_1\rangle$  and  $|e_2\rangle$ . (Answer:  $|1\rangle = \frac{1}{\sqrt{2}}(|e_1\rangle - |e_2\rangle)$ ,  $|2\rangle = \frac{1}{\sqrt{2}}e^{i\phi}(|e_1\rangle + |e_2\rangle)$ .)

What are these states like?

- States  $|1\rangle$  and  $|2\rangle$  have definite positions for the nitrogen atom, namely “up” or “down”. But they don't have definite energies. These states are sketched on page 128.
- States  $|e_1\rangle$  and  $|e_2\rangle$  have definite energies, namely  $E + A$  or  $E - A$ . But they don't have definite positions for the nitrogen atom. They can't be sketched using classical ink. (For a molecule in this state the nitrogen atom is like a silver atom passing through “both branches” of an interferometer — the atom does not have a definite position.)

4. In the basis  $\{|e_1\rangle, |e_2\rangle\}$ , the matrix representation of the Hamiltonian is

$$\begin{pmatrix} E + A & 0 \\ 0 & E - A \end{pmatrix}.$$

It's now straightforward to solve the differential equations. Using the notation

$$|\psi(t)\rangle = \bar{\psi}_1(t)|e_1\rangle + \bar{\psi}_2(t)|e_2\rangle,$$

the time development differential equations are

$$\begin{aligned}\frac{d\bar{\psi}_1(t)}{dt} &= -\frac{i}{\hbar}(E+A)\bar{\psi}_1(t) \\ \frac{d\bar{\psi}_2(t)}{dt} &= -\frac{i}{\hbar}(E-A)\bar{\psi}_2(t)\end{aligned}$$

with the immediate solutions

$$\begin{aligned}\bar{\psi}_1(t) &= \bar{\psi}_1(0)e^{-(i/\hbar)(E+A)t} \\ \bar{\psi}_2(t) &= \bar{\psi}_2(0)e^{-(i/\hbar)(E-A)t}.\end{aligned}$$

Thus

$$|\psi(t)\rangle = e^{-(i/\hbar)Et} \left[ e^{-(i/\hbar)At}\bar{\psi}_1(0)|e_1\rangle + e^{+(i/\hbar)At}\bar{\psi}_2(0)|e_2\rangle \right]. \quad (5.37)$$

(It is surprising that this time evolution result — and indeed the result of any possible experiment — is independent of the phase  $\phi$  of the off-diagonal element of the Hamiltonian. This surprise is explained in problem 5.8.)

Let's try out this general solution for a particular initial condition. Suppose the nitrogen atom starts out “up” — that is,

$$|\psi(0)\rangle = |1\rangle, \quad (5.38)$$

and we ask for the probability of finding it “down” — that is,  $|\langle 2|\psi(t)\rangle|^2$ .

The initial expansion coefficients in the  $\{|e_1\rangle, |e_2\rangle\}$  basis are (see equations (5.36))

$$\begin{aligned}\bar{\psi}_1(0) &= \langle e_1|\psi(0)\rangle = \langle e_1|1\rangle = \frac{1}{\sqrt{2}} \\ \bar{\psi}_2(0) &= \langle e_2|\psi(0)\rangle = \langle e_2|1\rangle = -\frac{1}{\sqrt{2}}\end{aligned}$$

so

$$|\psi(t)\rangle = \frac{1}{\sqrt{2}}e^{-(i/\hbar)Et} \left[ e^{-(i/\hbar)At}|e_1\rangle - e^{+(i/\hbar)At}|e_2\rangle \right].$$

The amplitude to find the nitrogen atom “down” is

$$\begin{aligned}\langle 2|\psi(t)\rangle &= \frac{1}{\sqrt{2}}e^{-(i/\hbar)Et} \left[ e^{-(i/\hbar)At}\langle 2|e_1\rangle - e^{+(i/\hbar)At}\langle 2|e_2\rangle \right] \\ &= \frac{1}{\sqrt{2}}e^{-(i/\hbar)Et} \left[ e^{-(i/\hbar)At} \left( \frac{1}{\sqrt{2}}e^{-i\phi} \right) - e^{+(i/\hbar)At} \left( \frac{1}{\sqrt{2}}e^{-i\phi} \right) \right] \\ &= \frac{1}{2}e^{-i\phi}e^{-(i/\hbar)Et} \left[ e^{-(i/\hbar)At} - e^{+(i/\hbar)At} \right] \\ &= \frac{1}{2}e^{-i\phi}e^{-(i/\hbar)Et} \left[ -2i \sin \left( (1/\hbar)At \right) \right] \\ &= -ie^{-i\phi}e^{-(i/\hbar)Et} \sin \left( \frac{A}{\hbar}t \right)\end{aligned}$$

and thus the probability of finding the nitrogen atom “down” is

$$|\langle 2|\psi(t)\rangle|^2 = \sin^2\left(\frac{A}{\hbar}t\right). \quad (5.39)$$

GRAPH HERE

This oscillation has period

$$\frac{\pi\hbar}{A} = \frac{2\pi\hbar}{\Delta E}$$

where  $\Delta E$  represents the energy splitting between the two energy eigenvalues,  $E + A$  and  $E - A$ .

This oscillation is at the heart of the MASER (Microwave Amplification by Stimulated Emission of Radiation).

### Problems

#### 5.1 Probability of no change

In equation (5.39) we found the probability that the nitrogen atom began in the “up” position (equation 5.38) and finished in the “down” position. Find the amplitude and the probability that the nitrogen atom will finish in the “up” position, and verify that these two probabilities sum to 1.

#### 5.2 Tunneling for small times

Equation (5.37) solves the time evolution problem completely, for all time. But it doesn’t give a lot of insight into what’s “really going on”. This problem provides some of that missing insight.

- a. When the time involved is short, we can approximate time evolution through

$$|\psi(\Delta t)\rangle = \left[ \hat{1} - \frac{i}{\hbar} \hat{H} \Delta t + \dots \right] |\psi(0)\rangle. \quad (5.40)$$

Show that this equation, represented in the  $\{|1\rangle, |2\rangle\}$  basis, is

$$\begin{pmatrix} \psi_1(\Delta t) \\ \psi_2(\Delta t) \end{pmatrix} \approx \begin{pmatrix} 1 - (i/\hbar)E\Delta t & -(i/\hbar)Ae^{i\phi}\Delta t \\ -(i/\hbar)Ae^{-i\phi}\Delta t & 1 - (i/\hbar)E\Delta t \end{pmatrix} \begin{pmatrix} \psi_1(0) \\ \psi_2(0) \end{pmatrix}. \quad (5.41)$$

- b. Express the initial condition  $|\psi(0)\rangle = |1\rangle$ , used above at equation (5.38), in the  $\{|1\rangle, |2\rangle\}$  basis, and show that, for small times,

$$\begin{pmatrix} \psi_1(\Delta t) \\ \psi_2(\Delta t) \end{pmatrix} \approx \begin{pmatrix} 1 - (i/\hbar)E\Delta t \\ -(i/\hbar)Ae^{i\phi}\Delta t \end{pmatrix}. \quad (5.42)$$

- c. This shows that the system starts with amplitude 1 for being in state  $|1\rangle$ , but that amplitude “seeps” (or “diffuses” or “hops”) from  $|1\rangle$  into  $|2\rangle$ . In fact, the amplitude to be found in  $|2\rangle$  after a small time  $\Delta t$  has passed is  $-(i/\hbar)Ae^{i\phi}\Delta t$ . What is the probability of being found in  $|2\rangle$ ? What is the condition for a “small” time?
- d. Show that the same probability results from approximating result (5.39) for small times.

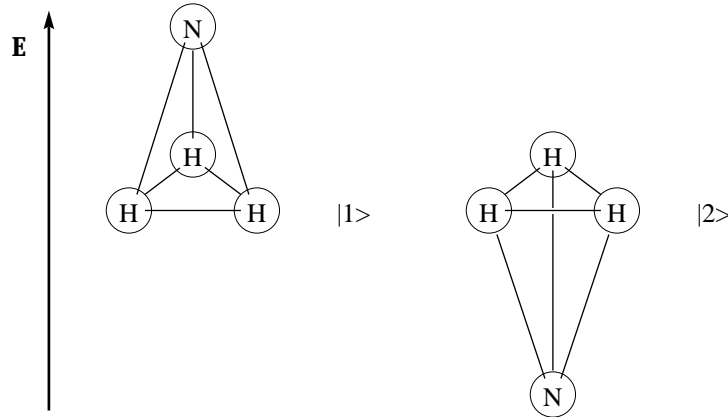
In a normal diffusion process – such as diffusion of blue dye from one water cell into an adjacent water cell – the dye spreads out uniformly and then net diffusion stops. But in this quantal amplitude diffusion, the amplitude is complex-valued. As such, the diffusion of more amplitude into the second cell can result, through destructive interference, in a decreased amplitude in the second cell. This interference gives rise to the oscillatory behavior demonstrated in equation (5.39).

- e. While this approach does indeed provide a lot of insight, it also raises a puzzle. What, according to equation (5.42), is the probability of being found in the initial state  $|1\rangle$  after a short time has passed? Conclude that the total probability is greater than 1! We will resolve this paradox in problem 11.1.



### 5.3 Ammonia molecule in an electric field

Place an ammonia molecule into an external electric field  $\mathcal{E}$  perpendicular to the plane of hydrogen atoms.



Now the states  $|1\rangle$  and  $|2\rangle$  are no longer symmetric, so we can no longer assume that  $\langle 1|\hat{H}|1\rangle = \langle 2|\hat{H}|2\rangle$ . Indeed, the proper matrix representation of  $\hat{H}$  in the  $\{|1\rangle, |2\rangle\}$  basis is

$$\begin{pmatrix} E + p\mathcal{E} & Ae^{i\phi} \\ Ae^{-i\phi} & E - p\mathcal{E} \end{pmatrix},$$

where  $p$  is interpreted as the molecular dipole moment. (Negative charge migrates toward the nitrogen atom.)

- Find the eigenvalues  $e_1$  and  $e_2$  of  $\hat{H}$ . (Check against the results (5.35) that apply when  $\mathcal{E} = 0$ .)
- Find the eigenvectors  $|e_1\rangle$  and  $|e_2\rangle$  in terms of  $|1\rangle$  and  $|2\rangle$ . (Check against the results (5.36).)
- If a molecule is initially in state  $|1\rangle$ , find the probability that it will be found in state  $|2\rangle$  as a function of time.

### 5.3 Formal properties of time evolution; Conservation laws

Quantal states evolve according to the Schrödinger time-development equation

$$\frac{d|\psi(t)\rangle}{dt} = -\frac{i}{\hbar}\hat{H}|\psi(t)\rangle. \quad (5.43)$$

The Hamiltonian operator  $\hat{H}$  is Hermitian, with eigenvectors  $\{|e_n\rangle\}$  and eigenvalues  $e_n$ :

$$\hat{H}|e_n\rangle = e_n|e_n\rangle. \quad (5.44)$$

These are called the “energy eigenstates” or “states of definite energy”.

**Theorem I:** *Energy eigenstates are stationary states.*

If  $|\psi(0)\rangle = (\text{number})|e_n\rangle$ , then  $|\psi(t)\rangle = (\text{number}')|e_n\rangle$ , where both numbers have square modulus unity.

Because of this result, the energy eigenstates are also called “stationary states”: once you’re in one of them, you stay.

*Proof:* A formal proof will be given in the proof of theorem II. This informal proof provides less rigor and more insight.

Start at time  $t = 0$  and step forward a small amount of time  $\Delta t$ :

$$\frac{\Delta|\psi\rangle}{\Delta t} \approx -\frac{i}{\hbar}\hat{H}|\psi(0)\rangle \quad (5.45)$$

$$= -\frac{i}{\hbar}\hat{H}(\text{number})|e_n\rangle \quad (5.46)$$

$$= (\text{stuff})|e_n\rangle. \quad (5.47)$$

$$\Delta|\psi\rangle = (\text{stuff})\Delta t|e_n\rangle. \quad (5.48)$$

That is, the change in the state vector is parallel to the initial state vector, so the new state vector  $|\psi(\Delta t)\rangle = |\psi(0)\rangle + \Delta|\psi\rangle$  is again parallel to the initial state vector, and all three vectors are parallel to  $|e_n\rangle$ . Repeat for as many time steps as desired. ■

The vector  $|\psi(\Delta t)\rangle$  is not only parallel to the vector  $|\psi(0)\rangle$ , but it also has the same norm. (Namely unity.) This can’t happen for regular position vectors multiplied by real numbers. The only way to multiply a vector by a number, and get a different vector with the same norm, is to multiply by a complex number.

**Theorem II:** *Formal solution of the Schrödinger equation.*

$$\text{If } |\psi(0)\rangle = \sum_n \psi_n(0)|e_n\rangle, \text{ then } |\psi(t)\rangle = \sum_n \psi_n(0)e^{-(i/\hbar)e_n t}|e_n\rangle.$$

*Proof:* In component form, the Schrödinger equation is

$$\frac{d\psi_n(t)}{dt} = -\frac{i}{\hbar} \sum_m H_{n,m} \psi_m(t).$$

In the energy eigenbasis,

$$H_{n,m} = \begin{cases} e_n & n = m \\ 0 & n \neq m \end{cases} = e_n \delta_{n,m}.$$

Thus

$$\frac{d\psi_n(t)}{dt} = -\frac{i}{\hbar} \sum_m e_n \delta_{n,m} \psi_m(t) = -\frac{i}{\hbar} e_n \psi_n(t)$$

and

$$\psi_n(t) = \psi_n(0) e^{-(i/\hbar)e_n t}. \quad \blacksquare$$

So, this is how states change with time! What about measurements? We will first find how average values change with time, then look at “the whole shebang” – not just the average, but also the full distribution.

*Definition:* The operator  $\hat{A}\hat{B} - \hat{B}\hat{A}$  is called “the commutator of  $\hat{A}$  and  $\hat{B}$ ” and represented by  $[\hat{A}, \hat{B}]$ .

Theorem III: *Time evolution of averages.*

$$\frac{d\langle \hat{A} \rangle}{dt} = -\frac{i}{\hbar} \langle [\hat{A}, \hat{H}] \rangle.$$

*Proof:* (Using mathematical notation for inner products.)

$$\begin{aligned} \frac{d}{dt} \langle \hat{A} \rangle &= \frac{d}{dt} \left( \psi(t), \hat{A} \psi(t) \right) \\ &= \left( \frac{d\psi(t)}{dt}, \hat{A} \psi(t) \right) + \left( \psi(t), \hat{A} \frac{d\psi(t)}{dt} \right) \\ &= \left( -\frac{i}{\hbar} \hat{H} \psi(t), \hat{A} \psi(t) \right) + \left( \psi(t), \hat{A} \left[ -\frac{i}{\hbar} \hat{H} \psi(t) \right] \right) \quad \llbracket \text{use the fact that } \hat{H} \text{ is Hermitian} \rrbracket \\ &= \frac{i}{\hbar} \left( \psi(t), \hat{H} \hat{A} \psi(t) \right) - \frac{i}{\hbar} \left( \psi(t), \hat{A} \hat{H} \psi(t) \right) \\ &= -\frac{i}{\hbar} \left( \psi(t), [\hat{A} \hat{H} - \hat{H} \hat{A}] \psi(t) \right) \\ &= -\frac{i}{\hbar} \langle [\hat{A}, \hat{H}] \rangle \quad \blacksquare \end{aligned}$$

*Corollary:* If  $\hat{A}$  commutes with  $\hat{H}$ , then  $\langle \hat{A} \rangle$  is constant.

However, just because the average of a measurement doesn't change with time doesn't necessarily mean that *nothing* about the measurement changes with time. To fully specify the results of a measurement, you must also list the possible results, the eigenvalues  $a_n$ , and the probability of getting that result, namely  $|\langle a_n | \psi(t) \rangle|^2$ . The eigenvalues  $a_n$  are time constant, but how do the probabilities change with time?

*Theorem IV: Time evolution of projection probabilities.*

If  $|\phi\rangle$  is a time-independent state and  $\hat{P}_\phi = |\phi\rangle\langle\phi|$  is its associated projection operator, then

$$\frac{d}{dt} |\langle\phi|\psi(t)\rangle|^2 = -\frac{i}{\hbar} \langle[\hat{P}_\phi, \hat{H}]\rangle. \quad (5.49)$$

*Proof:*

$$\begin{aligned} \frac{d}{dt} |\langle\phi|\psi(t)\rangle|^2 &= \frac{d}{dt} [\langle\phi|\psi(t)\rangle\langle\phi|\psi(t)\rangle^*] \\ &= \left[ \langle\phi|\frac{d}{dt}|\psi(t)\rangle \right] \langle\phi|\psi(t)\rangle^* + \langle\phi|\psi(t)\rangle \left[ \langle\phi|\frac{d}{dt}|\psi(t)\rangle \right]^* \end{aligned}$$

But  $\langle\phi|\frac{d}{dt}|\psi(t)\rangle = -\frac{i}{\hbar} \langle\phi|\hat{H}|\psi(t)\rangle$ , so

$$\begin{aligned} \frac{d}{dt} |\langle\phi|\psi(t)\rangle|^2 &= -\frac{i}{\hbar} \left[ \langle\phi|\hat{H}|\psi(t)\rangle\langle\phi|\psi(t)\rangle^* - \langle\phi|\psi(t)\rangle\langle\phi|\hat{H}|\psi(t)\rangle^* \right] \\ &= -\frac{i}{\hbar} \left[ \langle\psi(t)|\phi\rangle\langle\phi|\hat{H}|\psi(t)\rangle - \langle\psi(t)|\hat{H}|\phi\rangle\langle\phi|\psi(t)\rangle \right] \\ &= -\frac{i}{\hbar} \left[ \langle\psi(t)| \left\{ |\phi\rangle\langle\phi|\hat{H} - \hat{H}|\phi\rangle\langle\phi| \right\} |\psi(t)\rangle \right] \\ &= -\frac{i}{\hbar} \langle\psi(t)|[\hat{P}_\phi, \hat{H}]|\psi(t)\rangle \quad \blacksquare \end{aligned}$$

*Lemma:* Suppose  $\hat{A}$  and  $\hat{B}$  are commuting Hermitian operators. If  $|a\rangle$  is an eigenvector of  $\hat{A}$  and  $\hat{P}_a = |a\rangle\langle a|$ , then  $[\hat{P}_a, \hat{B}] = 0$ .

*Proof:* From the compatibility theorem, there is an eigenbasis  $\{|b_n\rangle\}$  of  $\hat{B}$  with  $|b_1\rangle = |a\rangle$ . Write  $\hat{B}$  in diagonal form as

$$\hat{B} = \sum_n b_n |b_n\rangle\langle b_n|.$$

Then

$$\hat{B}|b_1\rangle\langle b_1| = \sum_n |b_n\rangle\langle b_n|b_1\rangle\langle b_1| = \sum_n |b_n\rangle\delta_{n,1}\langle b_1| = b_1|b_1\rangle\langle b_1|$$

while

$$|b_1\rangle\langle b_1|\hat{B} = \sum_n |b_1\rangle\langle b_1|b_n\rangle\langle b_n| = \sum_n |b_1\rangle\delta_{1,n}\langle b_n| = b_1|b_1\rangle\langle b_1|. \quad \blacksquare$$

*Corollary:* If  $\hat{A}$  commutes with  $\hat{H}$ , then nothing about the measurement of  $\hat{A}$  changes with time.

*Definition:* The observable associated with such an operator is said to be *conserved*.

Note: All these results apply to time evolution uninterrupted by measurements.

## 5.4 Magnetic moment in a uniform magnetic field

### 5.5 The neutral K meson

You know that elementary particles are characterized by their mass and charge, but that two particles of identical mass and charge can still behave differently. Physicists have invented characteristics such as “strangeness” and “charm” to label (not explain!) these differences. For example, the difference between the electrically neutral K meson  $K^0$  and its antiparticle the  $\bar{K}^0$  is described by attributing a strangeness of +1 to the  $K^0$  and of -1 to the  $\bar{K}^0$ .

Most elementary particles are completely distinct from their antiparticles: an electron never turns into a positron! Such a change is prohibited by charge conservation. However this prohibition does not extend to the neutral K meson precisely because it is neutral. In fact, there is a time-dependent amplitude for a  $K^0$  to turn into a  $\bar{K}^0$ . We say that the  $K^0$  and the  $\bar{K}^0$  are the two basis states for a two-state system. This two-state system has an observable strangeness, represented by an operator, and we have a  $K^0$  when the system is in an eigenstate of strangeness with eigenvalue +1, and a  $\bar{K}^0$  when the system is in an eigenstate of strangeness with eigenvalue -1. When the system is in other states it does not have a definite value of strangeness, and cannot be said to be “a  $K^0$ ” or “a  $\bar{K}^0$ ”. The two strangeness eigenstates are denoted  $|K^0\rangle$  and  $|\bar{K}^0\rangle$ .

#### 5.4 Strangeness

Write an outer product expression for the strangeness operator  $\hat{S}$ , and find its matrix representation in the  $\{|K^0\rangle, |\bar{K}^0\rangle\}$  basis. Note that this matrix is just the Pauli matrix  $\sigma_3$ .

#### 5.5 Charge Parity

Define an operator  $\widehat{CP}$  that turns one strangeness eigenstate into the other:

$$\widehat{CP}|K^0\rangle = |\bar{K}^0\rangle, \quad \widehat{CP}|\bar{K}^0\rangle = |K^0\rangle.$$

(CP stands for “charge parity”, although that’s not important here.) Write an outer product expression and a matrix representation (in the  $\{|K^0\rangle, |\bar{K}^0\rangle\}$  basis) for the  $\widehat{CP}$  operator. What is the connection between this matrix and the Pauli matrices? Show that the normalized eigenstates of CP are

$$|K_U\rangle = \frac{1}{\sqrt{2}} (|K^0\rangle + |\bar{K}^0\rangle),$$

$$|K_S\rangle = \frac{1}{\sqrt{2}} (|K^0\rangle - |\bar{K}^0\rangle).$$

(The U and S stand for unstable and stable, but that’s again irrelevant because we’ll ignore K meson decay.)

#### 5.6 The Hamiltonian

The time evolution of a neutral K meson is governed by the “weak interaction” Hamiltonian

$$\hat{H} = e\hat{1} + f\widehat{CP}.$$

(There is no way for you to derive this. I’m just telling you.) Show that the numbers  $e$  and  $f$  must be real.

#### 5.7 Time evolution

Neutral K mesons are produced in states of definite strangeness because they are produced by the “strong interaction” Hamiltonian that conserves strangeness. Suppose one is produced at time  $t = 0$  in state  $|K^0\rangle$ . Solve the Schrödinger equation to find its state for all time afterwards. Why is it easier to solve this problem using  $|K_U\rangle, |K_S\rangle$  vectors rather than  $|K^0\rangle, |\bar{K}^0\rangle$  vectors? Calculate and plot the probability of finding the meson in state  $|K^0\rangle$  as a function of time.

[[The neutral K meson system is extraordinarily interesting. I have oversimplified by ignoring decay. More complete treatments can be found in Ashok Das & Adrian Melissinos, *Quantum Mechanics* (Gordon and Breach,

New York, 1986) pages 172–173; R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, volume III (Addison-Wesley, Reading, Massachusetts, 1965) pages 11-12–20; Gordon Baym, *Lectures on Quantum Mechanics* (W.A. Benjamin, Reading, Massachusetts, 1969), pages 38–45; and Harry J. Lipkin, *Quantum Mechanics: New Approaches to Selected Topics* (North-Holland, Amsterdam, 1986) chapter 7.]

### 5.8 The most general two-state Hamiltonian

We've seen a number of two-state systems by now: the spin states of a spin- $\frac{1}{2}$  atom, the polarization states of a photon, the CP states of a neutral K-meson. [For more two-state systems, see R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, volume III (Addison-Wesley, Reading, Massachusetts, 1965) chapters 9, 10, and 11.] This problem investigates the most general possible Hamiltonian for any two-state system.

Because the Hamiltonian must be Hermitian, it must be represented by a matrix of the form

$$\begin{pmatrix} a & c \\ c^* & b \end{pmatrix}$$

where  $a$  and  $b$  are real, but  $c = |c|e^{i\gamma}$  might be complex. Thus the Hamiltonian is specified through four real numbers:  $a$ ,  $b$ , magnitude  $|c|$ , and phase  $\gamma$ . This seems at first glance to be the most general Hamiltonian.

But remember that states can be modified by an arbitrary overall phase. If the initial basis is  $\{|1\rangle, |2\rangle\}$ , show that in the new basis  $\{|1\rangle, |2'\rangle\}$ , where  $|2'\rangle = e^{-i\gamma}|2\rangle$ , the Hamiltonian is represented by the matrix

$$\begin{pmatrix} a & |c| \\ |c| & b \end{pmatrix}$$

which is pure real and which is specified through only three real numbers.





## Chapter 6

# The Quantum Mechanics of Position

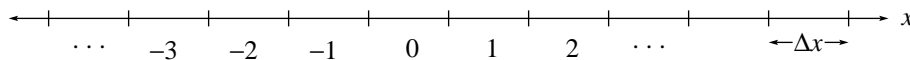
### 6.1 Describing states in continuum systems

At the start of this book we said we'd begin by treating only the magnetic moment of the atom quantum mechanically, and that once we got some grounding on the physical concepts and mathematical tools of quantum mechanics in this situation, we'd move on to the quantal treatment of other properties of the atom — such as its position. This led us to develop quantum mechanics for systems with two basis states. This was a very good thing, and we learned a lot about quantum mechanics, and also about practical applications like atomic clocks and MASERs.

All good things must come to an end, but in this case we're ending one good thing to come onto an even better thing, namely the quantum mechanics of a continuum system. The system we'll pick is a particle moving in one dimension. For the time being we'll ignore the atom's magnetic moment and internal constitution, and focus only on its position. Later in the book we'll treat both position and magnetic moment together.

#### *Course-grained description*

The situation is a point particle moving in one dimension. We start off with a course-grained description of the particle's position: we divide the line into an infinite number of bins, each of width  $\Delta x$ . (We will later take the limit as the bin width vanishes and the number of bins grows to compensate.)



If we ask “In which bin is the particle positioned?” the answer might be “It’s not in any of them. The particle doesn’t have a position.” Not all states have definite positions. On the other hand, there are *some* states that *do* have definite positions. If the particle has a definite position within bin 5 we say that it is in state  $|5\rangle$ .

I maintain that the set of states  $\{|n\rangle\}$  with  $n = 0, \pm 1, \pm 2, \pm 3, \dots$  constitutes a basis, because the set is:

- Orthonormal. If the particle is in one bin, then it’s not in any of the others. The mathematical expression of this property is

$$\langle n|m\rangle = \delta_{n,m}.$$

- Complete. If the particle does have a position, then it has a position within one of the bins. The mathematical expression of this property is

$$\sum_{n=-\infty}^{\infty} |n\rangle\langle n| = \hat{1}.$$

If the particle has no definite position, then it is in a state  $|\psi\rangle$  that is a superposition of basis states (Have I used this word before?)

$$|\psi\rangle = \sum_{n=-\infty}^{\infty} \psi_n |n\rangle \quad (6.1)$$

where

$$\psi_n = \langle n|\psi\rangle \quad \text{so} \quad \sum_{n=-\infty}^{\infty} |\psi_n|^2 = 1. \quad (6.2)$$

The quantity  $|\psi_5|^2$  is the probability that, if the position of the particle is measured (perhaps by shining a light down the one-dimensional axis), the particle will be found within bin 5. We should always say

“ $|\psi_5|^2$  is the probability of finding the particle in bin 5”,

because the word “finding” suggests the whole story: Right now the particle has no position, but after you measure the position then it will have a position, and the probability that this position falls within bin 5 is  $|\psi_5|^2$ . This phrase is totally accurate but it’s a real mouthful. Instead one frequently hears

“ $|\psi_5|^2$  is the probability that the particle is in bin 5”.

This is technically wrong. Before the position measurement, when the particle is in state  $|\psi\rangle$ , the particle *doesn’t have* a position. It has no probability of being in bin 5, or bin 6, or any other bin, just as love doesn’t have probability 0.5 of being red, 0.3 of being green, and 0.2 of being blue. Love doesn’t have a color, and the particle in state  $|\psi\rangle$  doesn’t have a position.

Because the second, inaccurate, phrase is shorter than the first, correct, phrase, it is often used despite its falseness. You may use it too, as long as you don’t believe it.

Similarly, the most accurate statement is

“ $\psi_5$  is the amplitude for finding the particle in bin 5”,

but you will frequently hear the brief and inaccurate

“ $\psi_5$  is the amplitude that the particle is in bin 5”

instead.

### ***Successively finer-grained descriptions***

Suppose we want a more accurate description of the particle’s position properties. We can get it using a smaller value for the bin width  $\Delta x$ . Still more accurate descriptions come from using still smaller values of  $\Delta x$ . Ultimately I can come up with a sequence of ever smaller bins homing in on the position of interest, say  $x_0$ . For all values of  $\Delta x$ , I will call the bin straddling  $x_0$  by the name “bin  $k$ ”. The relevant question seems at first to be: What is the limit

$$\lim_{\Delta x \rightarrow 0} |\psi_k|^2 ?$$

In fact, this is not an interesting question. The answer to that question is “zero”. For example: Suppose you are presented with a narrow strip

of lawn, 1000 meters long, which contains seven four-leaf clovers, scattered over the lawn at random. The probability of finding a four-leaf clover within a 2-meter wide bin is

$$\frac{7}{1000 \text{ m}}(2 \text{ m}) = 0.014.$$

The probability of finding a four-leaf clover within a 1-meter wide bin is

$$\frac{7}{1000 \text{ m}}(1 \text{ m}) = 0.007.$$

The probability of finding a four-leaf clover within a 1-millimeter wide bin is

$$\frac{7}{1000 \text{ m}}(0.001 \text{ m}) = 0.000007.$$

As the bin width goes to zero, the probability goes to zero as well. (Put another way, the probability of finding a four-leaf clover at a point along the strip of lawn is zero, because that probability is

$$\frac{7}{\text{number of points}},$$

and the number of points along the strip is infinite.)

The interesting question concerns not the bin probability, which always goes to zero, but the *probability density*, that is, the probability of finding the particle per length.

*Exercise.* What is the probability density for finding a four-leaf clover in the strip of lawn described above? Be sure to include the dimensions in your answer.

The probability per length of finding the particle at  $x_0$ , called the probability density at  $x_0$ , is the finite quantity

$$\lim_{\Delta x \rightarrow 0} \frac{|\psi_k|^2}{\Delta x}. \quad (6.3)$$

(Remember that the limit goes through a sequence of bins  $k$ , every one of which straddles the target point  $x_0$ .) In this expression both the numerator and denominator go to zero, but they approach zero in such a way that the ratio goes to a finite quantity. In other words, for small values of  $\Delta x$ , we have

$$|\psi_k|^2 \approx (\text{constant})\Delta x, \quad (6.4)$$

where that constant is the probability density for finding the particle at point  $x_0$ .

We know that amplitudes are more general than probabilities, because probabilities give the results for measurement experiments, but amplitudes give the results for both interference and measurement experiments. What does equation (6.4) say about bin amplitudes? It says that for small values of  $\Delta x$

$$\psi_k \approx (\text{constant})' \sqrt{\Delta x} \quad (6.5)$$

whence the limit

$$\lim_{\Delta x \rightarrow 0} \frac{\psi_k}{\sqrt{\Delta x}}$$

exists. This limit defines the quantity, a function of  $x_0$ ,

$$\lim_{\Delta x \rightarrow 0} \frac{\psi_k}{\sqrt{\Delta x}} = \psi(x_0).$$

If I were naming this quantity, I would have named it “amplitude density”. But for historical reasons it has a different name, namely “the wavefunction”.

The wavefunction evaluated at  $x_0$  is often called “the amplitude for the particle to have position  $x_0$ ”, but that’s not exactly correct, because an amplitude squared is a probability whereas a wavefunction squared is a probability density. Instead this phrase is just shorthand for the more accurate phrase “ $\psi(x_0)\sqrt{\Delta x}$  is the amplitude for finding the particle in an interval of short length  $\Delta x$  straddling position  $x_0$ , when the position is measured”.

*Exercise.* Show that the wavefunction for a point particle in one dimension has the dimensions  $1/\sqrt{\text{length}}$ .

### ***Working with wavefunctions***

When we were working with discrete systems, we said that the inner product could be calculated through

$$\langle \phi | \psi \rangle = \sum_n \phi_n^* \psi_n.$$

How does this pull over into continuum systems?

For any particular stage in the sequence of ever-smaller bins, the inner product is calculated through

$$\langle \phi | \psi \rangle = \sum_{i=-\infty}^{\infty} \phi_i^* \psi_i.$$

Prepare to take the  $\Delta x \rightarrow 0$  limit by writing

$$\langle \phi | \psi \rangle = \sum_{i=-\infty}^{\infty} \frac{\phi_i^*}{\sqrt{\Delta x}} \frac{\psi_i}{\sqrt{\Delta x}} \Delta x.$$

Then

$$\langle \phi | \psi \rangle = \lim_{\Delta x \rightarrow 0} \sum_{i=-\infty}^{\infty} \frac{\phi_i^*}{\sqrt{\Delta x}} \frac{\psi_i}{\sqrt{\Delta x}} \Delta x = \int_{-\infty}^{+\infty} \phi^*(x) \psi(x) dx.$$

*Exercise.* What is the normalization condition for a wavefunction?

### Basis states

When we went through the process of looking at finer and finer course-grainings, that is, taking  $\Delta x \rightarrow 0$  and letting the number of bins increase correspondingly, we were not changing the physical state of the particle. Instead, we were just obtaining more and more accurate descriptions of that state. How? By using a larger and larger basis!<sup>1</sup> The sequence of intervals implies a sequence of basis states  $|k\rangle$ . What is the limit of that sequence?

One way to approach this question is to look at the sequence

$$\lim_{\Delta x \rightarrow 0} \psi_k = \lim_{\Delta x \rightarrow 0} \langle k | \psi \rangle = \left[ \lim_{\Delta x \rightarrow 0} \langle k | \right] | \psi \rangle. \quad (6.6)$$

(Where, in the last step, we have acknowledged that in the sequence of finer-grained approximations involves changing the basis states  $|k\rangle$ , not the state of the particle  $|\psi\rangle$ .) This approach is not helpful because the limit always vanishes.

More useful is to look at the sequence

$$\lim_{\Delta x \rightarrow 0} \frac{\psi_k}{\sqrt{\Delta x}} = \lim_{\Delta x \rightarrow 0} \frac{\langle k | \psi \rangle}{\sqrt{\Delta x}} = \left[ \lim_{\Delta x \rightarrow 0} \frac{\langle k |}{\sqrt{\Delta x}} \right] | \psi \rangle = \psi(x_0). \quad (6.7)$$

<sup>1</sup>You might object that the basis was not really getting bigger — it started out with an infinite number of bins and at each stage in the process always has an infinite number of bins. I will reply that in some sense it has a “larger infinity” than it started with. If you want to make this sense rigorous and precise, take a mathematics course that studies transfinite numbers.

This sequence motivates the definition of the “position basis state”

$$|x_0\rangle = \lim_{\Delta x \rightarrow 0} \frac{|k\rangle}{\sqrt{\Delta x}}. \quad (6.8)$$

This new entity  $|x_0\rangle$  is not quite the same thing as the basis states like  $|k\rangle$  that we’ve seen up to now, just as  $\psi(x_0)$  is not quite the same thing as an amplitude. For example,  $|k\rangle$  is dimensionless while  $|x_0\rangle$  has the dimensions of  $1/\sqrt{\text{length}}$ . Mathematicians call the entity  $|x_0\rangle$  not a “basis state” but a “rigged basis state”. The word “rigged” carries the nautical connotation — a rigged ship is one outfitted for sailing and ready to move into action — and not the unsavory connotation — a rigged election is an unfair one. These are again fascinating mathematical questions<sup>2</sup> but this is not a mathematics book, so we won’t make a big fuss over the distinction.

Completeness relation for continuum basis states:

$$\hat{1} = \sum_{i=-\infty}^{\infty} |i\rangle\langle i| = \lim_{\Delta x \rightarrow 0} \sum_{i=-\infty}^{\infty} \frac{|i\rangle}{\sqrt{\Delta x}} \frac{\langle i|}{\sqrt{\Delta x}} \Delta x = \int_{-\infty}^{+\infty} |x\rangle\langle x| dx. \quad (6.9)$$

Orthogonality relation for continuum basis states:

$$\begin{aligned} \langle i|j\rangle &= \delta_{i,j} \\ \langle x|y\rangle &= 0 \quad \text{when } x \neq y \\ \langle x|x\rangle &= \lim_{\Delta x \rightarrow 0} \frac{\langle i|i\rangle}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{1}{\Delta x} = \infty \\ \langle x|y\rangle &= \delta(x-y). \end{aligned}$$

Just as the wavefunction is related to an amplitude but is not a true amplitude, and a rigged basis state  $|x\rangle$  is related to a basis state but is not a true basis state, so the inner product result  $\delta(x-y)$ , the Dirac delta function, is related to a function but is not a true function. Mathematicians call it a “generalized function” or a “Schwartz distribution”.

### *Comparison of discrete and continuous basis states*

---

<sup>2</sup>If you find them interesting, take a course in rigged Hilbert spaces.

**Discrete**basis states  $|n\rangle$ ; dimensionless

$$\psi_n = \langle n|\psi\rangle$$

 $\psi_n$  is dimensionless

$$\sum_n |\psi_n|^2 = 1$$

$$\langle n|m\rangle = \delta_{n,m}$$

$$\langle\phi|\psi\rangle = \sum_n \phi_n^* \psi_n$$

$$\sum_n |n\rangle\langle n| = \hat{1}$$

**Continuous**basis states  $|x\rangle$ ; dimensions  $\frac{1}{\sqrt{\text{length}}}$ 

$$\psi(x) = \langle x|\psi\rangle$$

 $\psi(x)$  has dimensions  $\frac{1}{\sqrt{\text{length}}}$ 

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

$$\langle x|y\rangle = \delta(x-y)$$

$$\langle\phi|\psi\rangle = \int_{-\infty}^{+\infty} \phi^*(x)\psi(x) dx$$

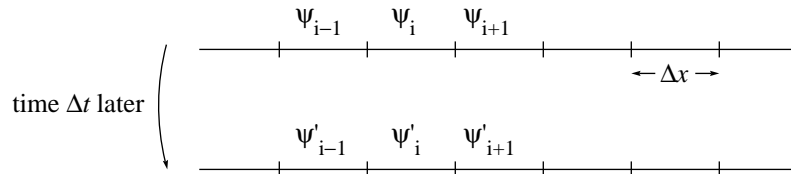
$$\int_{-\infty}^{+\infty} |x\rangle\langle x| dx = \hat{1}$$

*Exercise:* Show that  $\langle\phi|\psi\rangle = \int_{-\infty}^{+\infty} \phi^*(x)\psi(x) dx$ . Hint:  $\langle\phi|\psi\rangle = \langle\phi|\hat{1}|\psi\rangle$ .

**6.2 How does position amplitude change with time?**

In classical mechanics, the equation telling us how position changes with time is  $\vec{F} = m\vec{a}$ . It is not possible to derive  $\vec{F} = m\vec{a}$ , but it *is* possible to motivate it.

The role of this section is to uncover the quantal equivalent of  $\vec{F} = m\vec{a}$ : namely the equation telling us how position amplitude changes with time. As with  $\vec{F} = m\vec{a}$ , it is possible to motivate this equation but not to prove it. As such, the arguments in this section are suggestive, not definitive. Indeed, in some circumstances the arguments are false (e.g. for a single charged particle in a magnetic field, or for a pair of entangled particles).





**Normalization requirement**

The amplitude for the particle to be within bin  $i$  is initially  $\psi_i$ , and after time  $\Delta t$  it changes to  $\psi'_i = \psi_i + \Delta' \psi_i$ . (In this section, change with time is denoted  $\Delta' \psi$ , while change with space is denoted  $\Delta \psi$ .) Because the probability that the particle is in *some* bin is one, the bin amplitudes are normalized to

$$\sum_i |\psi_i|^2 = 1$$

and

$$\sum_i |\psi'_i|^2 = 1.$$

The second equation can be written

$$1 = \sum_i \psi_i'^* \psi_i' = \sum_i (\psi_i^* + \Delta' \psi_i^*) (\psi_i + \Delta' \psi_i) = \sum_i (\psi_i^* \psi_i + \psi_i^* \Delta' \psi_i + \Delta' \psi_i^* \psi_i + \Delta' \psi_i^* \Delta' \psi_i).$$

The first term on the far right sums to exactly 1, due to initial normalization. The next two terms are of the form  $z + z^* = 2\Re\{z\}$ , so

$$0 = \sum_i 2\Re\{\psi_i^* \Delta' \psi_i\} + \Delta' \psi_i^* \Delta' \psi_i.$$

When we go to the limit of very small  $\Delta t$ , then  $\Delta' \psi_i$  will be very small, and  $\Delta' \psi_i^* \Delta' \psi_i$ , as the product of two very small quantities, will be ultra small. Thus we neglect it and conclude that, due to normalization,

$$\Re \left\{ \sum_i \psi_i^* \Delta' \psi_i \right\} = 0. \quad (6.10)$$

We can change this to a relation about wavefunction rather than bin amplitude by remembering that, if  $x_i$  is the point at the center of bin  $i$ , then

$$\psi(x_i) = \lim_{\Delta x \rightarrow 0} \frac{\psi_i}{\sqrt{\Delta x}} \quad (6.11)$$

For very small bins, equation (6.10) becomes

$$\Re \left\{ \sum_i \psi^*(x_i) \sqrt{\Delta x} \Delta' \psi(x_i) \sqrt{\Delta x} \right\} = 0$$

or

$$\Re \left\{ \int_{-\infty}^{+\infty} \psi^*(x) \Delta' \psi(x) dx \right\} = 0. \quad (6.12)$$

### *The flow of amplitude*

Deductions from the preservation of normalization are important but purely formal... they don't tell us anything about the *physics* that's going on as time evolves. We begin with a very reasonable surmise:

$$\psi'_i = A_i\psi_{i-1} + B_i\psi_i + C_i\psi_{i+1}.$$

This says nothing more<sup>3</sup> than that the amplitude to be in bin  $i$  at the end of the time interval is the sum of

the amplitude to be in bin  $i - 1$  initially ( $\psi_{i-1}$ ) times the amplitude to flow right ( $A_i$ )  
 plus  
 the amplitude to be in bin  $i$  initially ( $\psi_i$ ) times the amplitude to stay in that bin ( $B_i$ )  
 plus  
 the amplitude to be in bin  $i + 1$  initially ( $\psi_{i+1}$ ) times the amplitude to flow left ( $C_i$ ).

The only important assumption we've made in writing down this surmise is that only adjacent bins are important: surely a reasonable assumption if the time interval  $\Delta t$  is short. (Some people like to call  $A_i$  and  $C_i$  "hopping amplitudes" rather than "flow amplitudes".)

Note that the change amplitudes  $A_i$ ,  $B_i$ , and  $C_i$  are *independent* of the position bin amplitudes  $\psi_{i-1}$ ,  $\psi_i$ , and  $\psi_{i+1}$ . That is,  $A_i$  represents the amplitude to flow right regardless of what amplitude is originally in bin  $i - 1$ . In other words,  $A_i$ ,  $B_i$ , and  $C_i$  depend on the *situation* (e.g. the mass of the particle, the forces applied to the particle) but not on the *state*.

We surmise further that the flow amplitudes are independent of position and of direction, so all the  $A_i$  and  $C_i$  are independent of  $i$ , and equal to each other. This surmise seems at first to be silly: surely if the particle moves on a line containing a hill and a valley, the flow will be more likely downhill than uphill. However, this observation shows only that  $A_i\psi_{i-1}$  will differ from  $C_i\psi_{i+1}$ , not that  $A_i$  will differ from  $C_i$ . We know that motion can happen even if there are no hills and valleys — that "a particle in motion remains at motion in the absence of an external force" — and the flow amplitudes concern this part of motion, the motion without external force.

<sup>3</sup>Compare the rules for combining amplitude on page 58.

(The surmise that left flow amplitude equals right flow amplitude does, in fact, turn out to be false for a charged particle in a magnetic field.) On the other hand, the hill vs. valley argument means that  $B_i$  will depend on position.

Finally, realize that the amplitudes  $A$  and  $B_i$  will depend on  $\Delta x$  and  $\Delta t$ : we expect that the flow amplitude  $A$  will increase with increasing  $\Delta t$  (more time, more flow), and decrease with increasing  $\Delta x$  (with fat bins the flow at boundaries is less significant).

With these surmises in place, we have

$$\psi'_i = A\psi_{i-1} + B_i\psi_i + A\psi_{i+1}. \quad (6.13)$$

Now, I write  $B_i$  in a funny way as  $B_i = -2A + 1 + D_i$ . I do this so that the equation will turn into

$$\Delta'\psi_i = \psi'_i - \psi_i = A(\psi_{i-1} - \psi_i) + D_i\psi_i + A(\psi_{i+1} - \psi_i), \quad (6.14)$$

which emphasizes amplitude differences rather than amplitude totals. In terms of the differences sketched below

this equation is

$$\Delta'\psi_i = -A\Delta\psi_L + D_i\psi_i + A\Delta\psi_R. \quad (6.15)$$

Writing this way, in terms of differences, sets us up for taking derivatives:

$$\Delta\psi_R - \Delta\psi_L = \Delta x \left( \frac{\Delta\psi_R}{\Delta x} - \frac{\Delta\psi_L}{\Delta x} \right).$$

The ratio  $\Delta\psi_R/\Delta x$  clearly relates to a spatial derivative taken at the right boundary of bin  $i$ . Furthermore

$$\Delta\psi_R - \Delta\psi_L = (\Delta x)^2 \left( \frac{\frac{\Delta\psi_R}{\Delta x} - \frac{\Delta\psi_L}{\Delta x}}{\Delta x} \right)$$

just as clearly relates to a second spatial derivative taken at the center of bin  $i$ .

At some point we need to switch over from talking about bin amplitude to talking about wavefunction, and this is a convenient point. Divide both sides of equation (6.15) by  $\sqrt{\Delta x}$  and use equation (6.11) to write (in an approximation that grows increasingly accurate as  $\Delta x \rightarrow 0$ )

$$\Delta' \psi(x_i) \approx -A(\Delta x)^2 \left( \frac{\partial^2 \psi(x)}{\partial x^2} \right)_{x=x_i} + D_i \psi(x_i)$$

While I have written this equation for the point at the center of bin  $i$ , of course it holds for any point. Defining  $D(x_i) = D_i$  gives

$$\Delta' \psi(x) \approx -A(\Delta x)^2 \frac{\partial^2 \psi}{\partial x^2} + D(x) \psi(x). \quad (6.16)$$

### Using the normalization equation

This is a good time to use result (6.12), the consequence of the normalization requirement. Applying (6.16) in (6.12) shows that

$$\int_{-\infty}^{+\infty} \psi^*(x) \Delta' \psi(x) dx = -A(\Delta x)^2 \int_{-\infty}^{+\infty} \psi^*(x) \frac{\partial^2 \psi}{\partial x^2} dx + \int_{-\infty}^{+\infty} \psi^*(x) D(x) \psi(x) dx \quad (6.17)$$

is pure imaginary. This requirement holds for *all wavefunctions*  $\psi(x)$ , and for *all situations* regardless of  $D(x)$ , so each of the two terms on the right must be pure imaginary. (That is, we cannot count on a real part in first term on the right to cancel a real part in the second term on the right, because if they happened to cancel for one function  $D(x)$ , they *wouldn't* cancel for a different function  $D(x)$ , but the normalization condition has to hold for *all possible* functions  $D(x)$ .)

The first integral on the right-hand side of (6.17) can be performed by parts:

$$\int_{-\infty}^{+\infty} \psi^*(x) \frac{\partial^2 \psi}{\partial x^2} dx = \left[ \psi^*(x) \frac{\partial \psi}{\partial x} \right]_{x=-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \frac{\partial \psi^*}{\partial x} \frac{\partial \psi}{\partial x} dx$$

The part in square brackets vanishes... otherwise  $\psi(x)$  is not normalized. The remaining integral is of the form

$$\int f^*(x) f(x) dx$$

which is pure real. Thus the constant  $A$  must be pure imaginary.

The second integral on the right-hand side of (6.17) is

$$\int_{-\infty}^{+\infty} \psi^*(x) D(x) \psi(x) dx$$

which must be imaginary for all wavefunctions  $\psi(x)$ , even wavefunctions that are pure real. Thus  $D(x)$  must be pure imaginary.

We have found that the amplitudes  $A$  and  $D(x)$  must be pure imaginary, so we define the pure real quantities  $a$  and  $d(x)$  through

$$A = ia \quad \text{and} \quad D(x) = id(x).$$

The discrete-time amplitude equation (6.16) becomes

$$\Delta' \psi(x) \approx i \left[ -a(\Delta x)^2 \frac{\partial^2 \psi}{\partial x^2} + d(x)\psi(x) \right]. \quad (6.18)$$

### Dimensional analysis

Let's find more about the quantity  $a$ , which is dimensionless. It's not plausible for the quantity  $a$  to depend on the phase of the moon, or the national debt. It can only depend on  $\Delta x$ ,  $\Delta t$ , the particle mass  $m$ , and Planck's constant  $\hbar$ . (It makes sense that  $a$  should depend on the inertia of the particle  $m$ , as we've already pointed out that this part of the Hamiltonian is involved with flow.)

quantity	dimensions
$\Delta x$	$[\ell]$
$\Delta t$	$[t]$
$m$	$[m]$
$\hbar$	$[m][\ell]^2/[t]$

The quantity  $a(\Delta x)^2$  must be finite in the limit  $\Delta x \rightarrow 0$ , so  $a$  must depend on  $\Delta x$  through the proportionality

$$a \propto \frac{1}{(\Delta x)^2} \quad \text{dimensions of right-hand side: } \frac{1}{[\ell]^2}.$$

To make  $a$  dimensionless we'll need to cancel the dimensions of length. The only way to do this is through  $\hbar$ :

$$a \propto \frac{\hbar}{(\Delta x)^2} \quad \text{dimensions of right-hand side: } \frac{[m]}{[t]}.$$

Now we need to cancel out the dimensions of mass and time. Again there is only one way to do this:

$$a \propto \frac{\hbar}{(\Delta x)^2} \frac{\Delta t}{m} \quad \text{dimensions of right-hand side: none.}$$

In short

$$a = \frac{\Delta t}{(\Delta x)^2} \frac{\hbar}{m} n_d$$

where  $n_d$  is a dimensionless real number. Note that, as anticipated immediately before equation (6.13), the quantity  $a$  increases with  $\Delta t$  and decreases with  $\Delta x$ .

With our new understanding we write equation (6.18) as

$$\Delta' \psi(x) \approx i \left[ -\frac{\hbar n_d}{m} \Delta t \frac{\partial^2 \psi}{\partial x^2} + d(x) \psi(x) \right]$$

or

$$\frac{\Delta' \psi(x)}{\Delta t} \approx i \left[ -\frac{\hbar n_d}{m} \frac{\partial^2 \psi}{\partial x^2} + \frac{d(x)}{\Delta t} \psi(x) \right]$$

which is conventionally written

$$\frac{\Delta' \psi(x)}{\Delta t} \approx -\frac{i}{\hbar} \left[ \frac{\hbar^2 n_d}{m} \frac{\partial^2 \psi}{\partial x^2} - \frac{\hbar d(x)}{\Delta t} \psi(x) \right].$$

(This form has the advantage that the part in square brackets has the dimensions of energy times the dimensions of  $\psi$ .)

The function  $\hbar d(x)/\Delta t$  has the dimensions of energy, and we call it  $v(x)$ . Now taking the two limits  $\Delta x \rightarrow 0$  and  $\Delta t \rightarrow 0$ , we find

$$\frac{\partial \psi(x, t)}{\partial t} = -\frac{i}{\hbar} \left[ \frac{\hbar^2 n_d}{m} \frac{\partial^2 \psi(x, t)}{\partial x^2} - v(x) \psi(x, t) \right]. \quad (6.19)$$

*Exercise:* Does it make physical sense that the “stay at home bin amplitude”  $D_i$  (see equation 6.14) should increase with increasing  $\Delta t$ ?

### **Classical limit**

To complete the specification of this equation, we must find values for  $n_d$  and  $v(x)$ . This can be done by applying the equation to a massive particle starting with a pretty-well defined position and seeing how that pretty-well defined position changes with time. In this so-called classical limit, the results of quantum mechanics must go over to match the results of classical mechanics. We are not yet equipped to do this, but we will find in section 6.5 that enforcing the classical limit gives the result that  $n_d = -1/2$  and  $v(x)$  is the negative of the classical potential energy function  $V(x)$ .

This latter result astounds me. The classical potential energy function derives from considering a particle with a definite location. Why should it have anything to do with quantum mechanics? I don't know, but it surely does.

We will see that the first part of the Hamiltonian corresponds to kinetic energy, and sure enough we've been relating it to "flow" or "hopping". Again, I am astounded that the quantal expression corresponding to kinetic energy is so different from the classical expression, just as I am astounded that the quantal expression corresponding to potential energy is so similar to the classical expression. Again, it's true whether I find it astounding or not.

### **Conclusion**

The wavefunction  $\psi(x; t)$  evolves in time according to

$$\frac{\partial\psi(x, t)}{\partial t} = -\frac{i}{\hbar} \left[ -\frac{\hbar^2}{2m} \frac{\partial^2\psi(x, t)}{\partial x^2} + V(x)\psi(x, t) \right], \quad (6.20)$$

where  $V(x)$  is the classical potential energy function. This equation was discovered in a completely different way by the 38-year-old Erwin Schrödinger during the Christmas season of 1925, at the alpine resort of Arosa, Switzerland, in the company of "an old girlfriend [from] Vienna", while his wife stayed at home in Zürich. It is called the Schrödinger equation, and it plays the same central role in quantum mechanics that  $\vec{F} = m\vec{a}$  plays in classical mechanics.

Do not think that we have derived the Schrödinger equation... instead we have taken it to pieces to see how it works.

### **Bibliography:**

R.P. Feynman, R.B. Leighton, and M. Sands, *The Feynman Lectures on Physics*, volume 3: Quantum Mechanics (Addison-Wesley, Reading, Massachusetts, 1965) pages 16-1–16-4.

Gordon Baym, *Lectures on Quantum Mechanics* (Benjamin, Reading, Massachusetts, 1969) pages 46–53.

W. Moore, *Schrödinger: Life and Thought* (Cambridge University Press, 1989) page 194.

### 6.3 What is wavefunction?

We have introduced the tool of wavefunction (or “amplitude density”). Wavefunction is sort of like magnetic field in that you can’t touch it or taste it or smell it, but in fact is even more abstract. For one thing wavefunction is complex-valued, not real-valued. For another it is determined, to some extent, by convention. We will see a third soon: The wavefunction for a single particle moving in three-dimensional space is a function of three-dimensional space:  $\psi(\vec{x})$ . But the wavefunction for two particles, A and B, each moving in three-dimensional space, is NOT  $\psi_A(\vec{x})$  plus  $\psi_B(\vec{x})$ , instead it is a function  $\psi(\vec{x}_A, \vec{x}_B)$ . That is, the wavefunction for two particles is a function in a six-dimensional space! (You might recall from a classical mechanics course that this space is called *configuration space*.)

This question has gnawed at people from the very beginnings of quantum mechanics: In the summer of 1926, Erich Hückel<sup>4</sup> composed the ditty, presented here in the free translation by Felix Bloch<sup>5</sup>

Erwin with his  $\psi$  can do  
 Calculations quite a few.  
 But one thing has not been seen:  
 Just what does  $\psi$  really mean?

Rather than worry about what wavefunction is, I recommend that you avoid traps of what wavefunction is not. It can’t be measured. It doesn’t exist in physical space. It is dependent on convention. It is a mathematical tool like the scalar and vector potentials.  $\psi$  is a step in an algorithm: it has no more physical significance than the intermediates of a multiplication.

<sup>4</sup>Erich Hückel (1896–1980) was a German physicist whose work in molecular orbitals resulted in the first successful treatment of the carbon-carbon double bond.

<sup>5</sup>Felix Bloch (1905–1983) was a Jewish-Swiss-American physicist who made contribution to the quantum theory of solids and elsewhere. He won the Nobel Prize for his work in nuclear magnetic resonance. His memory of this poem comes from his “Reminiscences of Heisenberg and the early days of quantum mechanics” [*Physics Today* **29**(12) 23–27 (December 1976)].



## 6.4 Operators and their representations; The momentum basis

### *The position operator and functions of the position operator*

The position operator is called  $\hat{x}$ . If we know the action of  $\hat{x}$  on every member of the  $\{|x\rangle\}$  basis (or any other basis!), then we know everything about the operator. But we do know that!

$$\hat{x}|x'\rangle = x'|x'\rangle.$$

Furthermore, we can find the action of  $\hat{x}^2$  on every member of the  $\{|x\rangle\}$  basis as follows:

$$\hat{x}^2|x'\rangle = \hat{x}(\hat{x}|x'\rangle) = \hat{x}(x'|x'\rangle) = x'(\hat{x}|x'\rangle) = x'(x'|x'\rangle) = (x')^2|x'\rangle.$$

Similarly, for any integer power  $n$ ,

$$\hat{x}^n|x'\rangle = (x')^n|x'\rangle.$$

*Exercise:* Prove this using mathematical induction.

If  $f(x)$  is a scalar function with Taylor series

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n, \quad (6.21)$$

then we define the *operator*  $f(\hat{x})$  through

$$f(\hat{x}) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} \hat{x}^n. \quad (6.22)$$

(This enables us to find operators for quantities like  $e^x$ .) The upshot is that for these operators, the position basis states are eigenstates:

$$f(\hat{x})|x'\rangle = f(x')|x'\rangle.$$

We've been examining the action of operators like  $f(\hat{x})$  on position basis states. What if they act upon some other state? We find out by expanding the general state  $|\psi\rangle$  into position states:

$$\begin{aligned} f(\hat{x})|\psi\rangle &= f(\hat{x})\hat{1}|\psi\rangle \\ &= f(\hat{x}) \int_{-\infty}^{+\infty} |x'\rangle \langle x'| dx' |\psi\rangle \\ &= \int_{-\infty}^{+\infty} f(\hat{x})|x'\rangle \langle x'|\psi\rangle dx' \\ &= \int_{-\infty}^{+\infty} |x'\rangle f(x') \langle x'|\psi\rangle dx'. \end{aligned}$$

To get a feel for this result, we look for the representation of the state  $f(\hat{x})|\psi\rangle$  in the  $\{|x\rangle\}$  basis:

$$\begin{aligned}\langle x|f(\hat{x})|\psi\rangle &= \int_{-\infty}^{+\infty} \langle x|x'\rangle f(x')\langle x'|\psi\rangle dx' \\ &= \int_{-\infty}^{+\infty} \delta(x-x')f(x')\psi(x') dx' \\ &= f(x)\psi(x).\end{aligned}$$

IMPORTANT RESULT: The representation of an operator  $f(\hat{x})$  in the position basis is

$$\langle x|f(\hat{x})|\psi\rangle = f(x)\langle x|\psi\rangle. \quad (6.23)$$

And, as we've seen, if we know  $\langle x|\hat{A}|\psi\rangle$  for general  $|\psi\rangle$  and for general  $x$ , then we know everything there is to know about the operator.

The relation between a function-of-position operator and its position basis representation is simple: erase the hats!

$$|\phi\rangle = f(\hat{x})|\psi\rangle \iff \phi(x) = f(x)\psi(x). \quad (6.24)$$

Another application:

$$\langle \phi|f(\hat{x})|\psi\rangle = \int_{-\infty}^{+\infty} dx \langle \phi|x\rangle x|f(\hat{x})|\psi\rangle = \int_{-\infty}^{+\infty} \phi^*(x)f(x)\psi(x) dx. \quad (6.25)$$

So you might think we're home free. But no, because...

### ***There are other operators***

For example, the Hamiltonian operator, defined in terms of its components in the position basis, is

$$\langle x|\hat{H}|\psi\rangle = \left[ -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \right] \langle x|\psi\rangle. \quad (6.26)$$

The logical definition of the momentum operator is through

$$\hat{H} = \frac{\hat{p}^2}{2m} + V(\hat{x}), \quad (6.27)$$

so

$$\langle x|\hat{p}^2|\psi\rangle = -\hbar^2 \frac{\partial^2}{\partial x^2} \langle x|\psi\rangle. \quad (6.28)$$

IMPORTANT RESULT: Define the momentum operator  $\hat{p}$  in terms of its components / representation in the position basis as

$$\langle x|\hat{p}|\psi\rangle = -i\hbar\frac{\partial}{\partial x}\langle x|\psi\rangle. \quad (6.29)$$

The operator with “+i” rather than “-i” out in front would have the same square, but would not have the correct classical limit. (See problem 6.7 and the second exercise below.)

*Exercise:* Would the phase-shifted convention

$$\langle x|\hat{p}|\psi\rangle = -i\hbar e^{i\delta}\frac{\partial}{\partial x}\langle x|\psi\rangle,$$

where  $\delta$  is pure real, be acceptable?

*Exercise — Sign of the momentum operator:* The function  $\psi_R(x; t) = Ae^{i(+kx-\omega t)}$  represents a wave moving to the right, while  $\psi_L(x; t) = Ae^{i(-kx-\omega t)}$  represents a wave moving to the left. (Take  $k$  to be positive.) Apply each of our two candidate momentum operators

$$\hat{p}_1 \doteq -i\hbar\frac{\partial}{\partial x} \quad \text{and} \quad \hat{p}_2 \doteq +i\hbar\frac{\partial}{\partial x}$$

to both of these functions, and show that the first candidate makes more sense.

Answer:

$$\langle x|\hat{p}_1|\psi_R\rangle = -i\hbar\frac{\partial}{\partial x}Ae^{i(+kx-\omega t)} = -i\hbar(+ik)Ae^{i(+kx-\omega t)} = (+\hbar k)\psi_R(x; t)$$

$$\langle x|\hat{p}_1|\psi_L\rangle = -i\hbar\frac{\partial}{\partial x}Ae^{i(-kx-\omega t)} = -i\hbar(-ik)Ae^{i(-kx-\omega t)} = (-\hbar k)\psi_L(x; t)$$

$$\langle x|\hat{p}_2|\psi_R\rangle = +i\hbar\frac{\partial}{\partial x}Ae^{i(+kx-\omega t)} = +i\hbar(+ik)Ae^{i(+kx-\omega t)} = (-\hbar k)\psi_R(x; t)$$

$$\langle x|\hat{p}_2|\psi_L\rangle = +i\hbar\frac{\partial}{\partial x}Ae^{i(-kx-\omega t)} = +i\hbar(-ik)Ae^{i(-kx-\omega t)} = (+\hbar k)\psi_L(x; t)$$

Thus the eigenvalues for these four situations are:

candidate	wave	eigenvalue
$\hat{p}_1$	rightward moving	$+\hbar k$
$\hat{p}_1$	leftward moving	$-\hbar k$
$\hat{p}_2$	rightward moving	$-\hbar k$
$\hat{p}_2$	leftward moving	$+\hbar k$

Candidate 1 associates the rightward moving wave with a positive momentum eigenvalue and the leftward moving wave with a negative momentum eigenvalue. Candidate 2 does the opposite. Since we intuitively

associate rightward motion with positive momentum, candidate 1 is superior.

Check on  $\hat{p}^2$ :

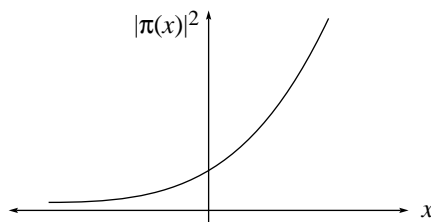
$$\begin{aligned}
 \langle x|\hat{p}^2|\psi\rangle &= \langle x|\hat{p}\hat{p}|\psi\rangle && \text{[[define } |\phi\rangle = \hat{p}|\psi\rangle\text{]]} \\
 &= \langle x|\hat{p}|\phi\rangle \\
 &= -i\hbar \frac{\partial}{\partial x} \langle x|\phi\rangle \\
 &= -i\hbar \frac{\partial}{\partial x} \left( \langle x|\hat{p}|\psi\rangle \right) \\
 &= -i\hbar \frac{\partial}{\partial x} \left( -i\hbar \frac{\partial}{\partial x} \langle x|\psi\rangle \right) \\
 &= -\hbar^2 \frac{\partial^2}{\partial x^2} \langle x|\psi\rangle
 \end{aligned}$$

Now that we have the momentum operator, we will of course want to find its eigenstates  $|p\rangle$ ! (Purists will point out that these are not actually eigenstates, but rigged eigenstates.)

**Find the position representation  $\pi(x) = \langle x|p\rangle$  of the momentum eigenstates**

$$\begin{aligned}
 \hat{p}|p\rangle &= \lambda|p\rangle \\
 \langle x|\hat{p}|p\rangle &= \lambda\langle x|p\rangle \\
 -i\hbar \frac{\partial}{\partial x} \langle x|p\rangle &= \lambda\langle x|p\rangle \\
 -i\hbar \frac{\partial \pi(x)}{\partial x} &= \lambda\pi(x) \\
 \frac{\partial \pi(x)}{\partial x} &= i\frac{\lambda}{\hbar}\pi(x) \\
 \pi(x) &= Ce^{i(\lambda/\hbar)x}
 \end{aligned} \tag{6.30}$$

That's funny. When we solve an eigenproblem, we expect that only a few eigenvalues  $\lambda$  will result. That's what happened with ammonia. But there we had  $2 \times 2$  matrices, and got two eigenvalues, whereas here we have  $\infty \times \infty$  matrices, so we get an infinite number of eigenvalues! The eigenvalue  $\lambda$  can be anything... positive, negative, even complex! A complex-valued  $\lambda$  will result in a probability density looking like this:



with an infinite pile of probability density off to the right or off to the left. This seems unphysical. Furthermore, complex values of  $\lambda$  would result in a non-Hermitian momentum operator, so we reject them. (Remember that in this section we are *not* making rigorous mathematical derivations, instead we are seeking sensible definitions.<sup>6</sup> Complex-valued eigenvalues  $\lambda$  for the momentum operator are not sensible.<sup>7</sup>)

The constant  $C$  is just an overall normalization constant. The best convention is (see problem 6.1)

$$C = \frac{1}{\sqrt{2\pi\hbar}}. \quad (6.31)$$

In summary, the operator  $\hat{p}$  has eigenvectors  $|p\rangle$  (technically, rigged vectors) satisfying

$$\hat{p}|p\rangle = p|p\rangle \quad (6.32)$$

$$\langle x|p\rangle = \frac{1}{\sqrt{2\pi\hbar}} e^{i(p/\hbar)x}. \quad (6.33)$$

<sup>6</sup>“Here and elsewhere in science, as stressed not least by Henri Poincaré, that view is out of date which used to say, ‘Define your terms before you proceed.’ All the laws and theories of physics, including the Lorentz force law [ $\vec{F} = q\vec{E} + q\vec{v} \times \vec{B}$ ], have this deep and subtle character, that they both define the concepts they use (here  $\vec{E}$  and  $\vec{B}$ ) and make statements about these concepts. Contrariwise, the absence of some body of theory, law, and principle deprives one of the means properly to define or even to use concepts. Any forward step in human knowledge is truly creative in this sense: that theory, concept, law, and method of measurement — forever inseparable — are born into the world in union.” C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (W.H. Freeman and Company, San Francisco, 1973) page 71.

<sup>7</sup>In exactly the same way, when you solve this classical trajectory problem — “Carol stands atop a 96 meter cliff and tosses a baseball at speed 45 m/s and angle  $33^\circ$  above the horizontal. When does the baseball hit the ground?” — you find two solutions: 7.6 s and  $-2.6$  s. Even though the negative number is a mathematically correct solution, you reject it on physical grounds.

*Exercise:* Show that  $|p\rangle$  has the dimensions of  $1/\sqrt{\text{momentum}}$ . What are the dimensions of  $\langle x|p\rangle$ ?

Problem 6.1 will show that the momentum states are orthonormal

$$\langle p|p'\rangle = \delta(p - p') \quad (6.34)$$

and complete

$$\hat{1} = \int_{-\infty}^{+\infty} |p\rangle\langle p| dp, \quad (6.35)$$

and hence the set  $\{|p\rangle\}$  constitutes a continuum (“rigged”) basis.

### **Representing states in the momentum basis**

We have been dealing with a state  $|\psi\rangle$  through its representation in the position basis, that is, through its wavefunction (or position representation)

$$\psi(x) = \langle x|\psi\rangle. \quad (6.36)$$

It is equally legitimate to deal with that state through its representation in the momentum basis, that is, through its so-called momentum wavefunction (or momentum representation)

$$\tilde{\psi}(p) = \langle p|\psi\rangle. \quad (6.37)$$

Either representation carries complete information about the state  $|\psi\rangle$ , so you can obtain one from the other

$$\tilde{\psi}(p) = \langle p|\psi\rangle = \int_{-\infty}^{+\infty} \langle p|x\rangle\langle x|\psi\rangle dx = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{+\infty} e^{-i(p/\hbar)x} \psi(x) dx \quad (6.38)$$

$$\psi(x) = \langle x|\psi\rangle = \int_{-\infty}^{+\infty} \langle x|p\rangle\langle p|\psi\rangle dp = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{+\infty} e^{+i(p/\hbar)x} \tilde{\psi}(p) dp \quad (6.39)$$

In short, the position and momentum wavefunctions are related to each other through a Fourier transform!

### **Representing operators in the momentum basis**

It is easy to represent momentum-related operators in the momentum basis. For example, using the fact the  $\hat{p}$  is Hermitian,

$$\langle p|\hat{p}|\psi\rangle = [\langle\psi|\hat{p}|p\rangle]^* = [p\langle\psi|p\rangle]^* = p\langle p|\psi\rangle. \quad (6.40)$$

More generally, for any function of the momentum operator,

$$\langle p|f(\hat{p})|\psi\rangle = f(p)\langle p|\psi\rangle. \quad (6.41)$$

It's a bit more difficult to find the momentum representation of the position operator, that is, to find  $\langle p|\hat{x}|\psi\rangle$ . But we can do it, using a slick trick called "parametric differentiation".

First, I'll introduce parametric differentiation in a purely mathematical context. suppose you need to evaluate the integral

$$\int_0^\infty x e^{-kx} \cos x \, dx$$

but you can only remember that

$$\int_0^\infty e^{-kx} \cos x \, dx = \frac{k}{k^2 + 1}.$$

You can differentiate both sides with respect to the parameter  $k$  finding

$$\begin{aligned} \frac{\partial}{\partial k} \int_0^\infty e^{-kx} \cos x \, dx &= \frac{\partial}{\partial k} \frac{k}{k^2 + 1} \\ \int_0^\infty \frac{\partial e^{-kx}}{\partial k} \cos x \, dx &= \frac{(k^2 + 1) - k(2k)}{(k^2 + 1)^2} \\ \int_0^\infty (-x e^{-kx}) \cos x \, dx &= \frac{-k^2 + 1}{(k^2 + 1)^2} \\ \int_0^\infty x e^{-kx} \cos x \, dx &= \frac{k^2 - 1}{(k^2 + 1)^2} \end{aligned}$$

This is a lot easier than any other method I can think of to evaluate this integral.

Go back to the problem of finding  $\langle p|\hat{x}|\psi\rangle$ :

$$\begin{aligned}
\langle p|\hat{x}|\psi\rangle &= \langle p|\hat{x}\hat{1}|\psi\rangle \\
&= \int_{-\infty}^{+\infty} \langle p|\hat{x}|x\rangle \langle x|\psi\rangle dx \\
&= \int_{-\infty}^{+\infty} \langle p|x\rangle x \langle x|\psi\rangle dx \\
&= \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{+\infty} e^{-i(p/\hbar)x} x \langle x|\psi\rangle dx \quad \text{[[Now use parametric differentiation!]]} \\
&= \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{+\infty} \frac{\hbar}{-i} \frac{\partial}{\partial p} \left[ e^{-i(p/\hbar)x} \right] \langle x|\psi\rangle dx \\
&= +i\hbar \frac{1}{\sqrt{2\pi\hbar}} \frac{\partial}{\partial p} \left[ \int_{-\infty}^{+\infty} e^{-i(p/\hbar)x} \langle x|\psi\rangle dx \right] \\
&= +i\hbar \frac{\partial}{\partial p} \left[ \int_{-\infty}^{+\infty} \langle p|x\rangle \langle x|\psi\rangle dx \right] \\
&= +i\hbar \frac{\partial}{\partial p} \langle p|\psi\rangle \tag{6.42}
\end{aligned}$$

There's a nice symmetry to this result, making it easy to remember: The momentum operator, represented in the position basis, is

$$\langle x|\hat{p}|\psi\rangle = -i\hbar \frac{\partial}{\partial x} \psi(x) \tag{6.43}$$

while the position operator, represented in the momentum basis, is

$$\langle p|\hat{x}|\psi\rangle = +i\hbar \frac{\partial}{\partial p} \tilde{\psi}(p). \tag{6.44}$$

*Exercise:* Show that

$$|\psi\rangle = \int_{-\infty}^{+\infty} \psi(x) |x\rangle dx = \int_{-\infty}^{+\infty} \tilde{\psi}(p) |p\rangle dp. \tag{6.45}$$

Verify that both of these relations have the correct dimensions.

### Problems

#### 6.1 The states $\{|p\rangle\}$ constitute a continuum basis

At equation (6.30) we showed that the inner product  $\langle x|p\rangle$  must have the form

$$\langle x|p\rangle = C e^{i(p/\hbar)x} \tag{6.46}$$

where  $C$  may be chosen for convenience.



a. Show that the operator

$$\hat{A} = \int_{-\infty}^{\infty} |p\rangle\langle p| dp \quad (6.47)$$

is equal to

$$2\pi\hbar|C|^2\hat{1} \quad (6.48)$$

by evaluating

$$\langle\phi|\hat{A}|\psi\rangle = \langle\phi|\hat{1}\hat{A}\hat{1}|\psi\rangle \quad (6.49)$$

for arbitrary states  $|\psi\rangle$  and  $|\phi\rangle$ . Hints: Set the first  $\hat{1}$  equal to  $\int_{-\infty}^{\infty} |x\rangle\langle x| dx$ , the second  $\hat{1}$  equal to  $\int_{-\infty}^{\infty} |x'\rangle\langle x'| dx'$ . The identity

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx} dk \quad (6.50)$$

(see Griffiths equation [2.144] on page 77) for the Dirac delta function is useful here. Indeed, this is one of the most useful equations to be found anywhere!

b. Using the conventional choice  $C = 1/\sqrt{2\pi\hbar}$ , show that

$$\langle p|p'\rangle = \delta(p - p'). \quad (6.51)$$

The expression (6.50) is again helpful.

## 6.2 Peculiarities of continuum basis states

Recall that the elements of a continuum basis set are peculiar in that they possess dimensions. That is not their only peculiarity. For any ordinary state  $|\psi\rangle$ , the wavefunction  $\psi(x) = \langle x|\psi\rangle$  satisfies

$$\int_{-\infty}^{\infty} \psi^*(x)\psi(x) dx = 1. \quad (6.52)$$

Show that the states  $|x'\rangle$  and  $|p\rangle$  cannot obey this normalization.

## 6.3 Hermiticity of the momentum operator

Show that the momentum operator is Hermitian over the space of states  $|\psi\rangle$  that have wavefunction  $\psi(x)$  which vanish at  $x = \pm\infty$ . Hint:

$$\langle\phi|\hat{p}|\psi\rangle = \int_{-\infty}^{\infty} \phi^*(x) \left( -i\hbar \frac{d\psi(x)}{dx} \right) dx. \quad (6.53)$$

Integrate by parts.

## 6.4 Commutator of $\hat{x}$ and $\hat{p}$

Show that  $[\hat{x}, \hat{p}] = i\hbar$  by showing that  $\langle\phi|[\hat{x}, \hat{p}]|\psi\rangle = i\hbar\langle\phi|\psi\rangle$  for arbitrary  $|\phi\rangle$  and  $|\psi\rangle$ . Hints: First evaluate  $\langle x|\hat{p}\hat{x}|\psi\rangle$  and  $\langle x|\hat{x}\hat{p}|\psi\rangle$ . It helps to define  $|\chi\rangle = \hat{x}|\psi\rangle$ .

### 6.5 Momentum representation of the Schrödinger equation

You know that the Schrödinger equation

$$\frac{d|\psi(t)\rangle}{dt} = -\frac{i}{\hbar}\hat{H}|\psi(t)\rangle \quad (6.54)$$

has the position representation

$$\frac{\partial\langle x|\psi(t)\rangle}{\partial t} = -\frac{i}{\hbar}\langle x|\hat{H}|\psi(t)\rangle \quad (6.55)$$

or

$$i\hbar\frac{\partial\psi(x;t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi(x;t)}{\partial x^2} + V(x)\psi(x;t). \quad (6.56)$$

In this problem you will uncover the corresponding equation that governs the time development of

$$\tilde{\psi}(p;t) = \langle p|\psi(t)\rangle. \quad (6.57)$$

The left hand side of equation (6.54) is straightforward because

$$\langle p|\frac{d}{dt}|\psi(t)\rangle = \frac{\partial\tilde{\psi}(p;t)}{\partial t}. \quad (6.58)$$

To investigate the right hand side of equation (6.54) write

$$\hat{H} = \frac{1}{2m}\hat{p}^2 + \hat{V} \quad (6.59)$$

where  $\hat{p}$  is the momentum operator and  $\hat{V}$  the potential energy operator.

- a. Use the Hermiticity of  $\hat{p}$  to show that

$$\langle p|\hat{H}|\psi(t)\rangle = \frac{p^2}{2m}\tilde{\psi}(p;t) + \langle p|\hat{V}|\psi(t)\rangle. \quad (6.60)$$

Now we must investigate  $\langle p|\hat{V}|\psi(t)\rangle$ .

- b. Show that

$$\langle p|\hat{V}|\psi(t)\rangle = \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} e^{-i(p/\hbar)x} V(x)\psi(x;t) dx \quad (6.61)$$

by inserting the proper form of  $\hat{1}$  at the proper location.

- c. Define the (modified) Fourier transform  $\tilde{V}(p)$  of  $V(x)$  through

$$\tilde{V}(p) = \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} e^{-i(p/\hbar)x} V(x) dx \quad (6.62)$$

$$= \int_{-\infty}^{\infty} \langle p|x\rangle V(x) dx. \quad (6.63)$$

Note that  $\tilde{V}(p)$  has funny dimensions. Show that

$$V(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} e^{i(p/\hbar)x} \tilde{V}(p) dp \quad (6.64)$$

$$= \int_{-\infty}^{\infty} \langle x|p\rangle \tilde{V}(p) dp. \quad (6.65)$$

You may use either forms (6.62) and (6.64), in which case the proof employs equation (6.50), or forms (6.63) and (6.65), in which case the proof involves completeness and orthogonality of basis states.

d. Hence show that

$$\langle p|\hat{V}|\psi(t)\rangle = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \tilde{V}(p-p') \tilde{\psi}(p';t) dp'. \quad (6.66)$$

(Caution! Your intermediate expressions will probably involve three distinct variables that you'll want to call "p". Put primes on two of them!)

e. Put everything together to see that  $\tilde{\psi}(p;t)$  obeys the integro-differential equation

$$\frac{\partial \tilde{\psi}(p;t)}{\partial t} = -\frac{i}{\hbar} \left[ \frac{p^2}{2m} \tilde{\psi}(p;t) + \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} dp' \tilde{V}(p-p') \tilde{\psi}(p';t) \right]. \quad (6.67)$$

This form of the Schrödinger equation is particularly useful in the study of superconductivity.

## 6.5 Time evolution of average quantities

In our general treatment of time evolution we found that for any measurable with operation  $\hat{A}$ , the average value  $\langle \hat{A} \rangle_t$  changed with time according to

$$\frac{d\langle \hat{A} \rangle_t}{dt} = -\frac{i}{\hbar} \langle [\hat{A}, \hat{H}] \rangle_t. \quad (6.68)$$

For the systems of this chapter,

$$\hat{H} = \frac{1}{2m} \hat{p}^2 + V(\hat{x}), \quad (6.69)$$

where  $\hat{x}$  and  $\hat{p}$  satisfy the commutation relation

$$[\hat{x}, \hat{p}] = \hat{x}\hat{p} - \hat{p}\hat{x} = i\hbar. \quad (6.70)$$

Knowing this, let's see how the average position  $\langle \hat{x} \rangle_t$  changes with time. We must find

$$[\hat{x}, \hat{H}] = \frac{1}{2m} [\hat{x}, \hat{p}^2] + [\hat{x}, V(\hat{x})].$$

The commutator  $[\hat{x}, V(\hat{x})]$  is easy:

$$[\hat{x}, V(\hat{x})] = \hat{x}V(\hat{x}) - V(\hat{x})\hat{x} = 0.$$

And the commutator  $[\hat{x}, \hat{p}^2]$  is not much harder. We use the known commutator for  $[\hat{x}, \hat{p}]$  to write

$$\hat{x}\hat{p}^2 = (\hat{x}\hat{p})\hat{p} = (\hat{p}\hat{x} + i\hbar)\hat{p} = \hat{p}\hat{x}\hat{p} + i\hbar\hat{p},$$

and then use it again to write

$$\hat{p}\hat{x}\hat{p} = \hat{p}(\hat{x}\hat{p}) = \hat{p}(\hat{p}\hat{x} + i\hbar) = \hat{p}^2\hat{x} + i\hbar\hat{p}.$$

Together we have

$$\hat{x}\hat{p}^2 = \hat{p}^2\hat{x} + 2i\hbar\hat{p}$$

or

$$[\hat{x}, \hat{p}^2] = 2i\hbar\hat{p}.$$

Plugging these commutators into the time-evolution result, we get

$$\frac{d\langle \hat{x} \rangle_t}{dt} = -\frac{i}{\hbar} \frac{1}{2m} 2i\hbar \langle \hat{p} \rangle_t.$$

or

$$\frac{d\langle \hat{x} \rangle_t}{dt} = \frac{\langle \hat{p} \rangle_t}{m}, \quad (6.71)$$

a result that stirs our memories of classical mechanics!

Meanwhile, what happens for average momentum  $\langle \hat{p} \rangle_t$ ?

$$[\hat{p}, \hat{H}] = \frac{1}{2m} [\hat{p}, \hat{p}^2] + [\hat{p}, V(\hat{x})] = [\hat{p}, V(\hat{x})].$$

To evaluate  $[\hat{p}, V(\hat{x})]$  we use the familiar idea that if we know  $\langle x | \hat{A} | \psi \rangle$  for arbitrary  $|x\rangle$  and  $|\psi\rangle$ , then we know everything there is to know about the operator  $\hat{A}$ . In this way, examine

$$\begin{aligned} \langle x | [\hat{p}, V(\hat{x})] | \psi \rangle &= \langle x | \hat{p} V(\hat{x}) | \psi \rangle - \langle x | V(\hat{x}) \hat{p} | \psi \rangle \\ &= -i\hbar \frac{\partial}{\partial x} \langle x | V(\hat{x}) | \psi \rangle - V(x) \langle x | \hat{p} | \psi \rangle \\ &= -i\hbar \frac{\partial}{\partial x} \left[ V(x) \psi(x) \right] - V(x) \left[ -i\hbar \frac{\partial}{\partial x} \psi(x) \right] \\ &= -i\hbar \left[ \frac{\partial V(x)}{\partial x} \psi(x) + V(x) \frac{\partial \psi(x)}{\partial x} - V(x) \frac{\partial \psi(x)}{\partial x} \right] \\ &= -i\hbar \left[ \frac{\partial V(x)}{\partial x} \psi(x) \right]. \end{aligned}$$

Now, the derivative of the classical potential energy function has a name. It's just (the negative of) the classical force function!

$$F(x) = -\frac{\partial V(x)}{\partial x}. \quad (6.72)$$

Continuing the evaluation begun above,

$$\begin{aligned} \langle x | [\hat{p}, V(\hat{x})] | \psi \rangle &= i\hbar [F(x)\psi(x)] \\ &= i\hbar \langle x | F(\hat{x}) | \psi \rangle. \end{aligned}$$

Because this relation holds for any  $|x\rangle$  and for any  $|\psi\rangle$ , we know that the operators are related as

$$[\hat{p}, V(\hat{x})] = i\hbar F(\hat{x}). \quad (6.73)$$

Going back to the time evolution of average momentum,

$$\frac{d\langle \hat{p} \rangle_t}{dt} = -\frac{i}{\hbar} \langle [\hat{p}, \hat{H}] \rangle_t = -\frac{i}{\hbar} i\hbar \langle F(\hat{x}) \rangle_t$$

or

$$\frac{d\langle \hat{p} \rangle_t}{dt} = \langle F(\hat{x}) \rangle_t, \quad (6.74)$$

which is suspiciously close to Newton's second law!

These two results together,

$$\frac{d\langle \hat{x} \rangle_t}{dt} = \frac{\langle \hat{p} \rangle_t}{m} \quad (6.75)$$

$$\frac{d\langle \hat{p} \rangle_t}{dt} = \langle F(\hat{x}) \rangle_t, \quad (6.76)$$

which tug so strongly on our classical heartstrings, are called the Ehrenfest<sup>8</sup> equations. There are two things you should remember about them: First, they are exact (within the assumptions of our derivation: non-relativistic, one-dimensional, no frictional or magnetic forces, etc.). Because they do tug our classical heartstrings, some people get the misimpression that they apply only in the classical limit. That's wrong — if you go back over the derivation you'll see that we never made any such assumption. Second, they are incomplete. This is because (1) knowing  $\langle \hat{x} \rangle_t$  doesn't let you calculate  $\langle F(\hat{x}) \rangle_t$ , because in general  $\langle F(\hat{x}) \rangle_t \neq F(\langle \hat{x} \rangle_t)$ , and because (2) even if you did know both  $\langle \hat{x} \rangle_t$  and  $\langle \hat{p} \rangle_t$ , that would not give you complete knowledge of the state.

<sup>8</sup>Paul Ehrenfest (1880–1933), Austrian and Dutch theoretical physicist. He was known particularly for asking probing questions that clarified the essence and delineated the unsolved problems of any matter under discussion. Particularly in this mode of questioner, he played a central role in the development of relativity, of quantum mechanics, and of statistical mechanics. He died tragically by his own hand.

**Problems****6.6 Alternative derivation.**

Derive result 6.73 by expanding  $V(x)$  in a Taylor series.

**6.7 Choice of sign for momentum operator**

If we had taken the opposite sign choice for the momentum operator at equation (6.29) (call this choice  $\hat{p}_2$ ), then what would have been the commutator  $[\hat{x}\hat{p}_2]$ ? What would have been the result 6.71?

**6.8 Quantities in the Hamiltonian**

When we derived equation (6.19) we were left with an undetermined number  $n_d$  and an undetermined function  $v(x)$ . Repeat the derivation of the Ehrenfest equations with this form of the Schrödinger equation to determine that number and function by demanding the correct classical limit.

## Chapter 7

# The Free Particle

### 7.1 Problems

#### 7.1 Energy eigenstates

In lecture we examined the behavior of a free particle in a state of definite momentum. Such states have a definite energy, but they are not the only possible states of definite energy.

- a. Show that the state

$$|\rho(0)\rangle = A|+p_0\rangle + B|-p_0\rangle, \quad (7.1)$$

where  $|A|^2 + |B|^2 = 1$ , has definite energy  $E(p_0) = p_0^2/2m$ . (That is,  $|\rho(0)\rangle$  is an energy eigenstate with eigenvalue  $p_0^2/2m$ ).

- b. Show that the “wavefunction” corresponding to  $|\rho(t)\rangle$  evolves in time as

$$\rho(x;t) = \frac{1}{\sqrt{2\pi\hbar}} \left[ A e^{i(+p_0x - E(p_0)t)/\hbar} + B e^{i(-p_0x - E(p_0)t)/\hbar} \right]. \quad (7.2)$$

I use the term wavefunction in quotes because  $\rho(x;t)$  is not  $\langle x|\text{normal state}\rangle$  but rather a sum of two terms like  $\langle x|\text{continuum basis state}\rangle$ .

- c. Show that the “probability density”  $|\rho(x;t)|^2$  is independent of time and given by

$$|\rho(x;t)|^2 = \frac{1}{2\pi\hbar} \left[ 1 + 2\Re\{A^*B\} \cos\left(\frac{2p_0x}{\hbar}\right) + 2\Im\{A^*B\} \sin\left(\frac{2p_0x}{\hbar}\right) \right]. \quad (7.3)$$

#### 7.2 A useful integral

Using  $\int_{-\infty}^{\infty} e^{-u^2} du = \sqrt{\pi}$ , show that

$$\int_{-\infty}^{\infty} e^{-u^2\alpha^2/2} e^{iuy} du = \frac{\sqrt{2\pi}}{\alpha} e^{-y^2/2\alpha^2} \quad (7.4)$$

where  $\alpha$  may be complex, but  $\Re\{\alpha^2\} > 0$ . *Hint:* Complete the square by writing

$$-\frac{u^2\alpha^2}{2} + iuy = -\left(\frac{u\alpha}{\sqrt{2}} - i\frac{y}{\sqrt{2}\alpha}\right)^2 - \frac{y^2}{2\alpha^2}.$$

*Note:* If  $c$  is a real number independent of  $x$ , you know that

$$\lim_{x \rightarrow \infty} (x + c) = \infty.$$

You might think that a different limit would result if the additive constant  $c$  were complex, but in fact, that is not the case:

$$\lim_{x \rightarrow \infty} (x + ic) = \infty.$$

It is not unusual for the limit of a sequence of complex numbers to be real.

### 7.3 A somewhat less useful integral

Given  $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$ , show that

$$\int_{-\infty}^{\infty} x^2 e^{-x^2} dx = \frac{\sqrt{\pi}}{2}. \quad (7.5)$$

*Hint:*  $\int_{-\infty}^{\infty} x^2 e^{-x^2} dx = 2 \int_0^{\infty} x^2 e^{-x^2} dx$ , then integrate by parts.

### 7.4 Static properties of a Gaussian wavepacket

Consider the wavefunction

$$\psi(x; 0) = \frac{A}{\sqrt{\sigma}} e^{-x^2/2\sigma^2} e^{i(p_0/\hbar)x}. \quad (7.6)$$

- Show that the wavefunction is properly normalized when  $A = 1/\sqrt[4]{\pi}$ .
- Show that in this state  $\langle \hat{x} \rangle = 0$  (trivial), and  $\Delta x = \sqrt{\langle (\hat{x} - \langle \hat{x} \rangle)^2 \rangle} = \sigma/\sqrt{2}$  (easy).
- Use equation (7.4) to show that

$$\tilde{\psi}(p; 0) = A\sqrt{\frac{\sigma}{\hbar}} e^{-(p-p_0)^2\sigma^2/2\hbar^2}. \quad (7.7)$$

- Hence show that  $\langle \hat{p} \rangle = p_0$  and  $\Delta p = \hbar/(\sqrt{2}\sigma)$ .

### 7.5 Force-free motion of a Gaussian wavepacket

A particle with the initial wavefunction given in the previous problem evolves as

$$\tilde{\psi}(p; t) = e^{-(i/\hbar)E(p)t} \tilde{\psi}(p, 0) \quad (7.8)$$

so that

$$\psi(x; t) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} e^{i(px - E(p)t)/\hbar} \tilde{\psi}(p; 0) dp. \quad (7.9)$$



- a. Plug in  $\tilde{\psi}(p; 0)$  and change the integration variable to  $k$  where  $\hbar k = p - p_0$  in order to show that

$$\psi(x; t) = A \sqrt{\frac{\sigma}{2\pi}} e^{i(p_0 x - E(p_0)t)/\hbar} \int_{-\infty}^{\infty} e^{-k^2(\sigma^2 + i\frac{\hbar t}{m})/2} e^{ik(x - \frac{p_0}{m}t)} dk. \quad (7.10)$$

*Hint:* Change variable first to  $p' = p - p_0$ , then to  $k = p'/\hbar$ .

- b. Define the complex dimensionless quantity

$$\beta = 1 + i \frac{\hbar t}{m\sigma^2} \quad (7.11)$$

and evaluate the integral using equation (7.4), giving

$$\psi(x; t) = A \frac{1}{\sqrt{\sigma\beta}} e^{i(p_0 x - E(p_0)t)/\hbar} e^{-(x - \frac{p_0}{m}t)^2/2\sigma^2\beta}. \quad (7.12)$$

- c. Hence show that

$$|\psi(x; t)|^2 = \frac{1}{\sqrt{\pi}\sigma|\beta|} e^{-(x - \frac{p_0}{m}t)^2/\sigma^2|\beta|^2}. \quad (7.13)$$

By comparing  $|\psi(x; t)|^2$  with  $|\psi(x; 0)|^2$ , read off the results

$$\langle \hat{x} \rangle = \frac{p_0}{m}t, \quad \Delta x = \frac{\sigma|\beta|}{\sqrt{2}} = \frac{\sigma}{\sqrt{2}} \sqrt{1 + \left(\frac{\hbar}{m\sigma^2}t\right)^2}. \quad (7.14)$$

(No computation is required!)



## Chapter 8

# Square Wells

Will include both infinite and finite square wells. See the problem on “Scaling” below.

### 8.1 The infinite square well

You are no doubt familiar with the energy eigenproblem for the infinite square well, but a short review is never-the-less in order. Consider a well of width  $L$ , and place the origin at the left edge of the well. The mathematical problem is to find values  $E_n$  and functions  $\eta_n(x)$  such that a solution to

$$-\frac{\hbar^2}{2m} \frac{d^2\eta_n}{dx^2} = E_n\eta_n(x) \quad (8.1)$$

has

$$\eta_n(0) = \eta_n(L) = 0. \quad (8.2)$$

Show that the eigenvalues are

$$E_n = n^2 \frac{\pi^2 \hbar^2}{2mL^2} \quad n = 1, 2, 3, 4, \dots \quad (8.3)$$

and the eigenfunctions are

$$\eta_n(x) = \sqrt{\frac{2}{L}} \sin\left(n\pi \frac{x}{L}\right). \quad (8.4)$$

Be sure to explain carefully why negative and zero values of  $n$  are not used. Note that  $\eta_n(x)$  is even under reflection about the well center for  $n$  odd, odd for  $n$  even.

### 8.2 Ground state energy for the infinite square well

Make up a problem like problem 9.1.

### 8.3 Characteristics of the ground energy level

The ground state energy for the infinite square well is

$$\frac{\pi^2 \hbar^2}{2mL^2}.$$

Does it make sense that...

- ... this energy vanishes as  $\hbar \rightarrow 0$ ? (*Clue:* Consider the classical limit.)
- ... this energy vanishes as  $L \rightarrow \infty$ ? (*Clue:* Think about the Heisenberg indeterminacy relation. Compare problem 8.2.)
- ... this energy varies as  $1/m$ ?

### 8.4 Scaling

We've seen that a normalizable solution of the energy eigenequation for a square well relies on a tradeoff between the well width and the candidate energy  $E$ . It makes sense that a suitable change in width can be offset by a suitable change in  $E$ . This problem explores that tradeoff.

Suppose  $\eta(x)$  is a solution to

$$-\frac{\hbar^2}{2m} \frac{d^2 \eta}{dx^2} + V(x)\eta(x) = E\eta(x)$$

with a particular energy  $E$ . Now consider a different potential energy function  $U(x) = s^2 V(sx)$ . For example, if  $s = 2$  then  $U(x)$  has the same shape as  $V(x)$ , but has half the width and four times the height. Call  $\sigma(x) = \eta(sx)$ . Show that  $\sigma(x)$  solves the energy eigenproblem for potential energy  $U(x)$ , but with a different energy. Find that energy.

### 8.1 What does an electron look like?

One electron is in an infinite square well, in energy eigenstate  $n = 2$ , with a node right in the center of the well. The electron has probability  $\frac{1}{2}$  of being found in the left half of the well, and probability  $\frac{1}{2}$  of being found in the right half of the well, but no probability density at all of being found at the center. How can the electron move from the left half to the right half without passing through the center?

This question betrays a deep-seated misconception about quantum mechanics. It arises from incorrectly thinking that the electron in energy eigenstate  $n = 2$  has a definite position that we don't know (or that is

changing erratically). That's a wrong picture: the electron doesn't have a position.

The "passing through nodes" question doesn't have an answer because the question assumes an erroneous picture for the character of an electron. It is as silly and as unanswerable as the question "If love is blue and passion is red-hot, how can passionate love exist?"

A similar conundrum is this one: "Suppose I start with an electron in a state so that it has equal probability of being anywhere in a box. If I shine a strong light throughout the entire box I will find the electron at only one point. But what happens if I shine the light on only the left half of the box, and don't find the electron? I now know that the electron is somewhere in the right half. How could the light, shining where the electron isn't, affect the electron?" This conundrum (called the Renninger negative-result experiment) has the same resolution as the "passage through nodes" conundrum: namely, the conundrum arises from an incorrect visualization of the electron as a hard, tiny marble. Before the light shining, the electron didn't have a position. Instead the amplitude to be found in the left half had the same magnitude as the amplitude to be found in the right half.

### ***Problems***

#### **8.5 Paradox?**

- a. The year is 1492, and you are discussing with a friend the radical idea that the earth is round. "This idea can't be correct," objects your friend, "because it contains a paradox. If it were true, then a traveler moving always due east would eventually arrive back at his starting point. Anyone can see that that's not possible!" Convince your friend that this paradox is not an internal inconsistency in the round-earth idea, but an inconsistency between the round-earth idea and the picture of the earth as a plane, a picture which your friend has internalized so thoroughly that he can't recognize it as an approximation rather than the absolute truth.
- b. The year is 2092, and you are discussing with a friend the radical idea of quantal interference. "This idea can't be correct," objects your friend, "because it contains a paradox. If it were true, then an atom passing through branch a would have to know whether branch b were open or blocked. Anyone can see that that's not

possible!” Convince your friend that this paradox is not an internal inconsistency in quantum mechanics, but an inconsistency between quantal ideas and the picture of an atom as a hard little marble that always has a definite position, a picture which your friend has internalized so thoroughly that he can’t recognize it as an approximation rather than the absolute truth.

## Chapter 9

# The Simple Harmonic Oscillator

### 9.1 Resume of energy eigenproblem

The energy eigenproblem for the simple harmonic oscillator is

$$-\frac{\hbar^2}{2m} \frac{d^2 \eta_n(x)}{dx^2} + \frac{m\omega^2}{2} x^2 \eta_n(x) = E_n \eta_n(x). \quad (9.1)$$

This is a second-order linear ordinary differential equation, and the theory of differential equations assures us that for every value of  $E_n$ , there are two linearly independent solutions to this equation.

This does *not*, however, mean that every  $E_n$  is an energy eigenvalue with two energy eigenfunctions. Nearly all of these solutions turn out to be unnormalizable,

$$\int_{-\infty}^{+\infty} \eta^*(x) \eta(x) dx = \infty,$$

so they do not represent physical states. The problem of solving the energy eigenproblem is simply the problem of plowing through the vast haystack of solutions of (9.1) to find those few needles with finite norm.

### 9.2 Solution of the energy eigenproblem: Differential equation approach

*Problem:* Given  $m$  and  $\omega$ , find values  $E_n$  such that the corresponding solutions  $\eta_n(x)$  of

$$-\frac{\hbar^2}{2m} \frac{d^2 \eta_n(x)}{dx^2} + \frac{m\omega^2}{2} x^2 \eta_n(x) = E_n \eta_n(x) \quad (9.2)$$

are normalizable wavefunctions. Such  $E_n$  are the energy eigenvalues, and the corresponding solutions  $\eta_n(x)$  are energy eigenfunctions.

*Strategy:* The following four-part strategy is effective for most differential equation eigenproblems:

- (1) Convert to dimensionless variable.
- (2) Remove asymptotic behavior of solutions.
- (3) Find non-asymptotic behavior using the series method.
- (4) Invoke normalization to terminate the series as a polynomial.

In this treatment, I'll play fast and loose with asymptotic analysis. But everything I'll do is both reasonable and rigorously justifiable. (C.M. Bender and S.A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* McGraw-Hill, New York, 1978.)

1. *Convert to dimensionless variable:* The only combination of  $m$ ,  $\omega$ , and  $\hbar$  with the dimensions of [length] is  $\sqrt{\hbar/m\omega}$ . Hence define the dimensionless variable proportional to length

$$q = \sqrt{\frac{m\omega}{\hbar}} x. \quad (9.3)$$

In terms of this variable, the ordinary differential equation (9.2) is

$$\frac{d^2\eta_n(q)}{dq^2} + \left( \frac{2E_n}{\hbar\omega} - q^2 \right) \eta_n(q) = 0. \quad (9.4)$$

*Exercise:* We're using this equation merely as a stepping-stone to reach the full answer, but in fact it contains a lot of information already. For example, suppose we had two electrons in two far-apart simple harmonic oscillators, the second one with three times the "stiffness" of the first (that is, the spring constants are related through  $k^{(2)} = 3k^{(1)}$ ). We don't yet know the energy of the fourth excited state for either oscillator, yet we can easily find their ratio. What is it?

2. *Remove asymptotic behavior of solutions:* Consider the limit as  $q^2 \rightarrow \infty$ . In this limit, the ODE (9.4) "becomes"

$$\frac{d^2\eta_n(q)}{dq^2} - q^2\eta_n(q) = 0, \quad (9.5)$$

but it is hard to solve even this simplified equation! Fortunately, it's not necessary to find an exact solution, only to find the asymptotic character of the solutions.



Pick the trial solution

$$f_n(q) = e^{-q^2/2}. \quad (9.6)$$

When we test to see whether this is a solution, we find

$$\begin{aligned} & \frac{d^2 f_n(q)}{dq^2} - q^2 f_n(q) \\ &= \left( q^2 e^{-q^2/2} - e^{-q^2/2} \right) - q^2 e^{-q^2/2} = -e^{-q^2/2} \end{aligned}$$

So the function (9.6) does *not* solve the ODE (9.5). On the other hand, the amount by which it “misses” solving (9.5) is small in the sense that

$$\lim_{q^2 \rightarrow \infty} \frac{d^2 f/dq^2 - q^2 f}{q^2 f} = \lim_{q^2 \rightarrow \infty} \frac{-e^{-q^2/2}}{q^2 e^{-q^2/2}} = \lim_{q^2 \rightarrow \infty} \frac{-1}{q^2} = 0.$$

A similar result holds for  $g_n(x) = e^{+q^2/2}$ .

Our conclusion is that, in the limit  $q^2 \rightarrow \infty$ , the solution  $\eta_n(q)$  behaves like

$$\eta_n(q) \approx Ae^{-q^2/2} + Be^{+q^2/2}.$$

If  $B \neq 0$ , then  $\eta_n(q)$  will not be normalizable because the probability density would become infinite as  $q^2 \rightarrow \infty$ . Thus the solutions we want — the normalizable solutions — behave like

$$\eta_n(q) \approx Ae^{-q^2/2}$$

in the limit that  $q^2$  becomes very large.

The three paragraphs above motivate us to define a new function  $v_n(q)$  through

$$\eta_n(q) = e^{-q^2/2} v_n(q). \quad (9.7)$$

(I *could* have just produced this definition by fiat, without motivation. But then you wouldn’t know how to come up with the proper motivation yourself when you’re faced with a new and unfamiliar differential equation.)

In terms of this new function, the exact ODE (9.4) becomes

$$\frac{d^2 v_n(q)}{dq^2} - 2q \frac{dv_n(q)}{dq} + \left( \frac{2E_n}{\hbar\omega} - 1 \right) v_n(q) = 0. \quad (9.8)$$

For brevity we introduce the shorthand notation

$$e_n = \frac{2E_n}{\hbar\omega} - 1. \quad (9.9)$$

3. *Find non-asymptotic behavior using the series method:* Okay, but how are we going to solve equation (9.8) for  $v_n(q)$ ? Through the power series method!

Try a solution of the form

$$\begin{aligned}
 v(q) &= \sum_{k=0}^{\infty} a_k q^k \\
 v'(q) &= \sum_{k=0}^{\infty} k a_k q^{k-1} & qv'(q) &= \sum_{k=0}^{\infty} k a_k q^k \\
 v''(q) &= \sum_{k=0}^{\infty} k(k-1) a_k q^{k-2} & \llbracket \text{note that first two terms vanish } \dots \rrbracket \\
 &= \sum_{k=2}^{\infty} k(k-1) a_k q^{k-2} & \llbracket \text{change summation index to } k' = k - 2 \dots \rrbracket \\
 &= \sum_{k'+2=2}^{\infty} (k'+2)(k'+1) a_{k'+2} q^{k'} & \llbracket \text{rename dummy index } k' \text{ to } k \dots \rrbracket \\
 &= \sum_{k=0}^{\infty} (k+2)(k+1) a_{k+2} q^k
 \end{aligned}$$

Then equation (9.8) becomes

$$\sum_{k=0}^{\infty} [(k+2)(k+1)a_{k+2} - 2ka_k + e_n a_k] q^k = 0. \quad (9.10)$$

All of the terms in square brackets must vanish, whence the recursion relation

$$a_{k+2} = \frac{2k - e_n}{(k+2)(k+1)} a_k \quad k = 0, 1, 2, \dots \quad (9.11)$$

Like any second order ODE, equation (9.8) has two linearly independent solutions:

- An *even solution* of equation (9.8) comes by taking  $a_0 = 1$ ,  $a_1 = 0$ . It is

$$v^{(e)}(q) = 1 - \frac{e_n}{2!} q^2 + \frac{(e_n - 4)e_n}{4!} q^4 - \frac{(e_n - 8)(e_n - 4)e_n}{6!} q^6 + \dots$$

- An *odd solution* of equation (9.8) comes by taking  $a_0 = 0$ ,  $a_1 = 1$ . It is

$$v^{(o)}(q) = q - \frac{e_n - 2}{3!} q^3 + \frac{(e_n - 6)(e_n - 2)}{5!} q^5 - \frac{(e_n - 10)(e_n - 6)(e_n - 2)}{7!} q^7 + \dots$$

What is the asymptotic behavior of such solutions  $v_n(q)$  as  $q^2 \rightarrow \infty$ ? Well, the large  $q$  behavior will be dominated by the high-order terms of the series. Generally, as  $k \rightarrow \infty$ ,

$$\frac{a_{k+2}}{a_k} = \frac{2k - e_n}{(k+2)(k+1)} \rightarrow \frac{2}{k}. \quad (9.12)$$

Compare this behavior to the expansion

$$e^{q^2} = b_0 + b_2q^2 + b_4q^4 + \dots \quad (9.13)$$

which has

$$\frac{b_{k+2}}{b_k} = \frac{1}{(k/2) + 1} \rightarrow \frac{2}{k}. \quad (9.14)$$

So whenever this happens,

$$v_n(q) \approx e^{q^2} \quad \text{and} \quad \eta_n(q) = e^{-q^2/2}v_n(q) \approx e^{q^2/2},$$

Thus giving us the very same unnormalizable behavior we've been trying so hard to avoid!

Is there no way to repair the situation?

4. *Invoke normalization to terminate the series as a polynomial:* The candidate wavefunction  $\eta_n(q)$  is not normalizable when  $a_{k+2}/a_k \rightarrow 2/k$  (see equation (9.12)). There is only one way to avoid this limit: when the series for  $v_n(q)$  terminates as a polynomial. This termination occurs when, for some non-negative integer  $n$ , we have  $2n = e_n$  whence (by recursion relation (9.11)),  $a_k = 0$  for all  $k > n$ . Hence the only physical states correspond to energies with

$$2n = e_n = \frac{2E_n}{\hbar\omega} - 1.$$

Or, rephrasing,

Energy (eigen)states can exist only if they correspond to the energy (eigen)values

$$E_n = \hbar\omega(n + \frac{1}{2}) \quad n = 0, 1, 2, 3, \dots \quad (9.15)$$

What are the wavefunctions of the energy eigenstates?

For  $n$  even,  $v_n^{(e)}(q)$  terminates and  $v_n^{(o)}(q)$  doesn't.  
 For  $n$  odd,  $v_n^{(o)}(q)$  terminates and  $v_n^{(e)}(q)$  doesn't.

By tradition one defines the *Hermite*<sup>1</sup> polynomial of  $n$ th order  $H_n(q)$ :

$$n \text{ even: } H_n(q) = (-1)^{n/2} \frac{n!}{(n/2)!} v_n^{(e)}(q) \quad (9.16)$$

$$n \text{ odd: } H_n(q) = (-1)^{(n-1)/2} \frac{2n!}{((n-1)/2)!} v_n^{(o)}(q) \quad (9.17)$$

so that

$$\eta_n(x) = C_n e^{-q^2/2} H_n(q) \quad q = \sqrt{\frac{m\omega}{\hbar}} x \quad (9.18)$$

where  $C_n$  is a normalization factor.

### 9.3 Solution of the energy eigenproblem: Operator factorization approach

The differential equation approach works. It's hard. It's inefficient in that we find an infinite number of solutions and then throw most of them away. It's dependent on a particular representation. Worst of all, it's hard to use. For example, suppose we wanted to find the expected value of the potential energy in the  $n$ -th energy eigenstate. We would find

$$\langle \hat{U} \rangle_n = \frac{m\omega^2}{2} \langle \eta_n | \hat{x}^2 | \eta_n \rangle = \frac{m\omega^2}{2} \int_{-\infty}^{+\infty} x^2 \eta_n^2(x) dx = \frac{m\omega^2}{2} \frac{\hbar}{m\omega} \frac{\int_{-\infty}^{+\infty} q^2 e^{-q^2} H_n^2(q) dq}{\int_{-\infty}^{+\infty} e^{-q^2} H_n^2(q) dq}.$$

Unless you happen to love integrating the Hermite polynomials, these last two integrals are intimidating.

I'll show you a method, invented by Dirac (or was it Schrödinger?), which avoids all these problems. On the other hand the method is hard to motivate: It clearly springs from the mind of genius.

Start with the Hamiltonian

$$\hat{H} = \frac{1}{2m} \hat{p}^2 + \frac{m\omega^2}{2} \hat{x}^2. \quad (9.19)$$

Since we're in a mathematical mode, it makes sense to define the dimensionless operators

$$\hat{X} = \sqrt{\frac{m\omega}{2\hbar}} \hat{x} \quad \text{and} \quad \hat{P} = \frac{1}{\sqrt{2m\hbar\omega}} \hat{p}, \quad (9.20)$$

<sup>1</sup>Biographical information on Charles Hermite is given on page 108.

which satisfy

$$[\hat{X}, \hat{P}] = \sqrt{\frac{m\omega}{2\hbar}} \frac{1}{\sqrt{2m\hbar\omega}} [\hat{x}, \hat{p}] = \frac{i}{2} \hat{1}, \quad (9.21)$$

and write

$$\hat{H} = \hbar\omega(\hat{X}^2 + \hat{P}^2). \quad (9.22)$$

Now, one of the most fundamental tools of problem solving is to break something complex into its simpler pieces. (“All Gaul is divided into three parts.”) If we had an expression like

$$x^2 - p^2$$

you might well break it into simpler pieces as

$$(x - p)(x + p).$$

Slightly less intuitive would be to express

$$x^2 + p^2$$

as

$$(x - ip)(x + ip).$$

But in our case, we’re factoring an *operator*, and we have to ask concerning the expression

$$(\hat{X} - i\hat{P})(\hat{X} + i\hat{P}) = \hat{X}^2 + i\hat{X}\hat{P} - i\hat{P}\hat{X} + \hat{P}^2 = \hat{X}^2 + i[\hat{X}, \hat{P}] + \hat{P}^2 = \hat{X}^2 + \hat{P}^2 - \frac{1}{2}\hat{1}. \quad (9.23)$$

So we haven’t quite succeeded in factorizing our Hamiltonian — there’s a bit left over due to non-commuting operators — but the result is

$$\hat{H} = \hbar\omega[(\hat{X} - i\hat{P})(\hat{X} + i\hat{P}) + \frac{1}{2}]. \quad (9.24)$$

From here, define

$$\hat{a} = \hat{X} + i\hat{P}. \quad (9.25)$$

The Hermitian adjoint of  $\hat{a}$  is

$$\hat{a}^\dagger = \hat{X} - i\hat{P}. \quad (9.26)$$

Note that the operators  $\hat{a}$  and  $\hat{a}^\dagger$  are not Hermitian. There is no observable corresponding to  $\hat{a}$ . The commutator is

$$[\hat{a}, \hat{a}^\dagger] = \hat{1}. \quad (9.27)$$

*Exercise:* Verify the above commutator.

And in terms of  $\hat{a}$  and  $\hat{a}^\dagger$ , the Hamiltonian is

$$\hat{H} = \hbar\omega(\hat{a}^\dagger\hat{a} + \frac{1}{2}). \quad (9.28)$$

*Our task:* Using only the fact that  $[\hat{a}, \hat{a}^\dagger] = \hat{1}$ , where  $\hat{a}^\dagger$  is the Hermitian adjoint of  $\hat{a}$ , solve the energy eigenproblem for  $\hat{H} = \hbar\omega(\hat{a}^\dagger\hat{a} + \frac{1}{2})$ .

We will do this by solving the eigenproblem for the operator  $\hat{N} = \hat{a}^\dagger\hat{a}$ . Once these are known, we can immediately read off the solution for the eigenproblem for  $\hat{H}$ . So, we look for the eigenvectors  $|n\rangle$  with eigenvalues  $n$  such that

$$\hat{N}|n\rangle = n|n\rangle. \quad (9.29)$$

Because  $\hat{N}$  is Hermitian, its eigenvalues are real. Furthermore, they are positive because (where we define the vector  $|\phi\rangle$  through  $|\phi\rangle = \hat{a}|n\rangle$ )

$$n = \langle n|\hat{N}|n\rangle = \langle n|\hat{a}^\dagger\hat{a}|n\rangle = \langle n|\hat{a}^\dagger|\phi\rangle = \langle\phi|\hat{a}|n\rangle^* = \langle\phi|\phi\rangle^* \geq 0. \quad (9.30)$$

Now I don't know much about energy state  $|n\rangle$ , but I do know that at least one exists. So for this particular one, I can ask "What is  $\hat{a}|n\rangle$ ?" Well,

$$\begin{aligned} \hat{a}|n\rangle &= \hat{1}\hat{a}|n\rangle \\ &= (\hat{a}\hat{a}^\dagger - \hat{a}^\dagger\hat{a})\hat{a}|n\rangle \\ &= \hat{a}\hat{N}|n\rangle - \hat{N}\hat{a}|n\rangle \\ &= n\hat{a}|n\rangle - \hat{N}\hat{a}|n\rangle. \end{aligned}$$

So if I define  $|\phi\rangle = \hat{a}|n\rangle$  (an unnormalized vector), then

$$\begin{aligned} |\phi\rangle &= n|\phi\rangle - \hat{N}|\phi\rangle \\ \hat{N}|\phi\rangle &= n|\phi\rangle - |\phi\rangle = (n-1)|\phi\rangle. \end{aligned}$$

In other words, the vector  $|\phi\rangle$  is an eigenvector of  $\hat{N}$  with eigenvalue  $n-1$ . Wow!

$$|\phi\rangle = C|n-1\rangle.$$

We need to find the normalization constant  $C$ :

$$\begin{aligned} \langle\phi|\phi\rangle &= |C|^2\langle n-1|n-1\rangle = |C|^2 \\ \langle\phi|\phi\rangle &= \langle n|\hat{a}^\dagger\hat{a}|n\rangle = \langle n|\hat{N}|n\rangle = n. \end{aligned}$$

So  $C = \sqrt{n}$  and

$$\hat{a}|n\rangle = \sqrt{n}|n-1\rangle \quad (9.31)$$

The operator  $\hat{a}$  is called a “lowering operator”.

So, we started off with one eigenstate  $|n\rangle$ . We applied  $\hat{a}$  to get another eigenstate — with smaller eigenvalue. We can apply  $\hat{a}$  to this new state to get yet another eigenstate with an even smaller eigenvalue. But this seems to raise a paradox. We saw at equation (9.30) that the eigenvalues were positive or zero. This seems present a mechanism for getting negative eigenvalues — in fact, eigenvalues as small as desired! For example if we started with a state of eigenvalue 2.3, we could lower it to produce a state of eigenvalue 1.3. We could lower this to produce a state of eigenvalue 0.3, and we could lower once more to produce a state of eigenvalue  $-0.7$ . But we know there are no states with negative eigenvalues! Thus there can't be any states of eigenvalue 2.3 to start off with.

However, if we start with a state of eigenvalue 2, we could lower that to get  $|1\rangle$ , lower that to get  $|0\rangle$ , and what happens when we try to lower  $|0\rangle$ ? From equation (9.31), we find

$$\hat{a}|0\rangle = \sqrt{0}|-1\rangle = 0.$$

When we lower the state  $|0\rangle$ , we don't get the state  $|-1\rangle$ . Instead we get nothing!

In conclusion, there are no fractional eigenvalues. The only eigenvalues are the non-negative integers.

We've gotten a lot out of the use of  $\hat{a}$ . What happens when we use  $\hat{a}^\dagger$ ?

$$\begin{aligned}\hat{a}^\dagger|n\rangle &= \hat{a}^\dagger\hat{1}|n\rangle \\ &= \hat{a}^\dagger(\hat{a}\hat{a}^\dagger - \hat{a}^\dagger\hat{a})|n\rangle \\ &= \hat{N}\hat{a}^\dagger|n\rangle - \hat{a}^\dagger\hat{N}|n\rangle \\ &= \hat{N}\hat{a}^\dagger|n\rangle - n\hat{a}^\dagger|n\rangle.\end{aligned}$$

So if I define  $|\chi\rangle = \hat{a}^\dagger|n\rangle$  (an unnormalized vector), then

$$\begin{aligned}|\chi\rangle &= \hat{N}|\chi\rangle - n|\chi\rangle \\ \hat{N}|\chi\rangle &= n|\chi\rangle + |\chi\rangle = (n+1)|\chi\rangle.\end{aligned}$$

In other words, the vector  $|\chi\rangle$  is an eigenvector of  $\hat{N}$  with eigenvalue  $n+1$ :

$$|\phi\rangle = C|n+1\rangle.$$

The operator  $\hat{a}^\dagger$  is a “raising operator”!

*Exercise:* Find the normalization constant  $C$  and conclude that

$$\hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle \quad (9.32)$$

The eigenproblem is solved entirely. Given only  $[\hat{a}, \hat{a}^\dagger] = \hat{1}$ , where  $\hat{a}^\dagger$  is the Hermitian adjoint of  $\hat{a}$ , the operator

$$\hat{H} = \hbar\omega(\hat{a}^\dagger\hat{a} + \frac{1}{2})$$

has eigenstates  $|0\rangle, |1\rangle, |2\rangle, \dots$  with eigenvalues  $\hbar\omega(\frac{1}{2}), \hbar\omega(1 + \frac{1}{2}), \hbar\omega(2 + \frac{1}{2}), \dots$ . These eigenstates are related through

$$\begin{aligned}\hat{a}|n\rangle &= \sqrt{n}|n-1\rangle && \text{“lowering operator”} \\ \hat{a}^\dagger|n\rangle &= \sqrt{n+1}|n+1\rangle && \text{“raising operator”}\end{aligned}$$

The operators  $\hat{a}$  and  $\hat{a}^\dagger$  are collectively called “ladder operators” or “elevator operators”.

## 9.4 Problems

### 9.1 Ground state of the simple harmonic oscillator

You may have been surprised that the lowest possible energy for the simple harmonic oscillator was  $E_0 = \frac{1}{2}\hbar\omega$  rather than  $E_0 = 0$ . This exercise attempts to explain the non-zero ground state energy in seat-of-the-pants, semiclassical terms rather than in rigorous, formal, mathematical terms. It then goes on to use these ideas plus the uncertainty principle to guess at a value for the ground state energy. You may abhor such non-rigorous arguments, but you must be able to do them in order to make informed guesses about the behavior of systems that are too complicated to yield to rigorous mathematical methods.

In classical mechanics the SHO ground state has zero potential energy (the particle is at the origin) and zero kinetic energy (it is motionless). However in quantum mechanics if a particle is localized precisely at the origin, and hence has zero potential energy, then it has a considerable spread of momentum values and hence a non-zero kinetic energy (or, to be precise, a non-zero expectation value for kinetic energy). The kinetic energy can be reduced by decreasing the spread of momentum values, but only by increasing the spread of position values and hence by increasing the (expected value of the) potential energy. The ground state is the state in which this trade off between kinetic and potential energies results in a minimum total energy.

Assume that the spread in position extends over some distance  $d$  about the origin (i.e. the particle will very likely be found between  $x = -d/2$  and  $x = +d/2$ ). This will result in a potential energy somewhat less



than

$$\frac{1}{2}m\omega^2 \left(\frac{d}{2}\right)^2.$$

This argument is not intended to be rigorous, so let's forget the "some-what less" part of the last sentence. Furthermore, a position spread of  $\Delta x = d$  implies through the uncertainty principle a momentum spread of  $\Delta p \geq \hbar/2d$ . (The expected value of the momentum is zero.) Continuing in our non-rigorous vein, let's set  $\Delta p = \hbar/2d$  and kinetic energy equal to

$$\frac{1}{2m} \left(\frac{\Delta p}{2}\right)^2.$$

Sketch potential energy, kinetic energy and total energy as a function of  $d$ . Find the minimum value of  $E(d)$  and compare with the true ground state energy  $E_0 = \frac{1}{2}\hbar\omega$ . (Note that if  $\hbar$  were zero, the energy minimum would fall at  $E(d) = 0$ !)

### 9.2 Expressions for simple harmonic oscillator ladder operators

Show that the lowering operator  $\hat{a}$  has the outer product expression

$$\hat{a} = \sum_{n=0}^{\infty} \sqrt{n} |n-1\rangle\langle n|$$

and the matrix representation (in the energy basis)

$$\begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & 0 \\ 0 & 0 & \sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & \sqrt{3} & 0 & \cdots \\ 0 & 0 & 0 & 0 & \sqrt{4} & \\ 0 & 0 & 0 & 0 & 0 & \\ & & \vdots & & & \ddots \end{pmatrix}.$$

Write down the outer product expression and matrix representation for  $\hat{a}^\dagger$ .

### 9.3 Ladder operators for the simple harmonic oscillator

- Express  $\hat{x}$  and  $\hat{p}$  in terms of  $\hat{a}$  and  $\hat{a}^\dagger$ .
- Calculate the following simple harmonic oscillator matrix elements:

$$\begin{aligned} &\langle m|\hat{a}|n\rangle && \langle m|\hat{p}|n\rangle && \langle m|\hat{x}\hat{p}|n\rangle \\ &\langle m|\hat{a}^\dagger|n\rangle && \langle m|\hat{x}^2|n\rangle && \langle m|\hat{p}\hat{x}|n\rangle \\ &\langle m|\hat{x}|n\rangle && \langle m|\hat{p}^2|n\rangle && \langle m|\hat{H}|n\rangle \end{aligned}$$

- c. Show that the expectation value of the potential energy in a SHO energy eigenstate equals the expectation value of the kinetic energy in that state. (Recall that for a classical simple harmonic oscillator, the time averaged potential energy equals the time averaged kinetic energy.)
- d. Find  $\Delta x$ ,  $\Delta p$ , and  $\Delta x \Delta p$  for the energy eigenstate  $|n\rangle$ .

#### 9.4 Simple harmonic oscillator states

Use scaled variables throughout this problem

- a. Concerning the ground energy state: What is  $\eta_0(x)$  at  $x = 0.5$ ? What is the probability density  $\rho_0(x)$  there?
- b. Concerning the first excited energy state: What is  $\eta_1(x)$  at  $x = 0.5$ ? What is the probability density  $\rho_1(x)$  there?
- c. Concerning the “50–50 combination”  $\psi_A(x) = (\rho_0(x) + \rho_1(x))/\sqrt{2}$ : What is  $\psi_A(x)$  at  $x = 0.5$ ? What is the probability density  $\rho_A(x)$  there?
- d. Concerning another “50–50 combination”  $\psi_B(x) = (\rho_0(x) - \rho_1(x))/\sqrt{2}$ : What is  $\psi_B(x)$  at  $x = 0.5$ ? What is the probability density  $\rho_B(x)$  there?
- e. Veronica argues that “Probability is central to quantum mechanics, so the probability density of any 50–50 combination of  $\eta_0(x)$  and  $\eta_1(x)$  will be half-way between  $\rho_0(x)$  and  $\rho_1(x)$ .” Prove Veronica wrong. What phenomenon of quantum mechanics has she ignored?
- f. (Optional, for the mathematically inclined.) Prove that for any 50–50 combination of  $\eta_0(x)$  and  $\eta_1(x)$ , the probability density at  $x$  will range from  $\rho_A(x)$  to  $\rho_B(x)$ . (*Clue:* Use the triangle inequality.)

Chapter 10

**Qualitative Solution of Energy  
Eigenproblems**



## Chapter 11

# Perturbation Theory

### 11.1 The $\mathcal{O}$ notation

Approximations are an important part of physics, and an important part of approximation is to ensure their reliability and consistency. The  $\mathcal{O}$  notation (pronounced “the big-oh notation”) is a practical tool for making approximations reliable and consistent.

The technique is best illustrated through an example. Suppose you desire an approximation for

$$f(x) = \frac{e^{-x}}{1-x} \quad (11.1)$$

valid for small values of  $x$ , that is, for  $x \ll 1$ . You know that

$$e^{-x} = 1 - x + \frac{1}{2}x^2 - \frac{1}{6}x^3 + \dots \quad (11.2)$$

and that

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots, \quad (11.3)$$

so it seems that reasonable approximations are

$$e^{-x} \approx 1 - x \quad (11.4)$$

and

$$\frac{1}{1-x} \approx 1 + x, \quad (11.5)$$

whence

$$\frac{e^{-x}}{1-x} \approx (1-x)(1+x) = 1 - x^2. \quad (11.6)$$

Let's try out this approximation at  $x_0 = 0.01$ . A calculator shows that

$$\frac{e^{-x_0}}{1-x_0} = 1.0000503\dots \quad (11.7)$$

while the value for the approximation is

$$1 - x_0^2 = 0.9999000. \quad (11.8)$$

This is a very poor approximation indeed... the deviation from  $f(0) = 1$  is even of the wrong sign!

Let's do the problem over again, but this time keeping track of exactly how much we've thrown away while making each approximation. We write

$$e^{-x} = 1 - x + \frac{1}{2}x^2 - \frac{1}{6}x^3 + \dots \quad (11.9)$$

as

$$e^{-x} = 1 - x + \frac{1}{2}x^2 + \mathcal{O}(x^3), \quad (11.10)$$

where the notation  $\mathcal{O}(x^3)$  stands for the small terms that we haven't bothered to write out explicitly. The symbol  $\mathcal{O}(x^3)$  means "terms that are about the magnitude of  $x^3$ , or smaller" and is pronounced "terms of order  $x^3$ ". The  $\mathcal{O}$  notation will allow us to make controlled approximations in which we keep track of exactly how good the approximation is.

Similarly, we write

$$\frac{1}{1-x} = 1 + x + x^2 + \mathcal{O}(x^3), \quad (11.11)$$

and find the product

$$f(x) = [1 - x + \frac{1}{2}x^2 + \mathcal{O}(x^3)] \times [1 + x + x^2 + \mathcal{O}(x^3)] \quad (11.12)$$

$$= [1 - x + \frac{1}{2}x^2 + \mathcal{O}(x^3)] \quad (11.13)$$

$$+ [1 - x + \frac{1}{2}x^2 + \mathcal{O}(x^3)] x \quad (11.14)$$

$$+ [1 - x + \frac{1}{2}x^2 + \mathcal{O}(x^3)] x^2 \quad (11.15)$$

$$+ [1 - x + \frac{1}{2}x^2 + \mathcal{O}(x^3)] \mathcal{O}(x^3). \quad (11.16)$$

Note, however, that  $x \times \frac{1}{2}x^2 = \mathcal{O}(x^3)$ , and that  $x^2 \times \mathcal{O}(x^3) = \mathcal{O}(x^3)$ , and so forth, whence

$$f(x) = [1 - x + \frac{1}{2}x^2 + \mathcal{O}(x^3)] \quad (11.17)$$

$$+ [x - x^2 + \mathcal{O}(x^3)] \quad (11.18)$$

$$+ [x^2 + \mathcal{O}(x^3)] \quad (11.19)$$

$$+ \mathcal{O}(x^3) \quad (11.20)$$

$$= 1 + \frac{1}{2}x^2 + \mathcal{O}(x^3). \quad (11.21)$$

Thus we have the approximation

$$f(x) \approx 1 + \frac{1}{2}x^2. \quad (11.22)$$

Furthermore, we know that this approximation is accurate to terms of order  $\mathcal{O}(x^2)$  (i.e. that the first neglected terms are of order  $\mathcal{O}(x^3)$ ). Evaluating this approximation at  $x_0 = 0.01$  gives

$$1 + \frac{1}{2}x_0^2 = 1.0000500, \quad (11.23)$$

far superior to our old approximation.

What went wrong on our first try? The  $-x^2$  in approximation (11.6) is the same as the  $-x^2$  on line (11.18). However, lines (11.17) and (11.19) demonstrate that there were other terms of about the same size (i.e. other “terms of order  $x^2$ ”) that we neglected in our first attempt.

The  $\mathcal{O}$  notation is superior to the “dot notation” (such as  $\dots$ ) in that dots stand for “a bunch of small terms”, but the dots don’t tell you just how small they are. The symbol  $\mathcal{O}(x^3)$  also stands for “a bunch of small terms”, but in addition it tells you precisely how small those terms are. The  $\mathcal{O}$  notation allows us to approximate in a consistent manner, unlike the uncontrolled approximations where we ignore a “small term” without knowing whether we have already retained terms that are even smaller.

### **Problem**

#### **11.1 Tunneling for small times — $\mathcal{O}$ notation version**

Problem 5.2, part e, raised the paradox that, according to an approximation produced using truncation rather than  $\mathcal{O}$  notation, the total probability was greater than 1. This problem resolves the paradox using  $\mathcal{O}$  notation.

- a. Approximate time evolution through

$$|\psi(\Delta t)\rangle = \left[ \hat{1} - \frac{i}{\hbar} \hat{H} \Delta t - \frac{1}{2\hbar^2} \hat{H}^2 (\Delta t)^2 + \mathcal{O}(\Delta t^3) \right] |\psi(0)\rangle. \quad (11.24)$$

Find the representation of this equation in the  $\{|1\rangle, |2\rangle\}$  basis.

- b. Conclude that for initial condition  $|\psi(0)\rangle = |1\rangle$ ,

$$\begin{pmatrix} \psi_1(\Delta t) \\ \psi_2(\Delta t) \end{pmatrix} = \begin{pmatrix} 1 - (i/\hbar)E\Delta t - (1/2\hbar^2)(E^2 + A^2)(\Delta t)^2 + \mathcal{O}(\Delta t^3) \\ -(i/\hbar)Ae^{-i\phi}\Delta t - (1/\hbar^2)EAe^{-i\phi}(\Delta t)^2 + \mathcal{O}(\Delta t^3) \end{pmatrix}. \quad (11.25)$$

- c. Find the resulting probabilities for the system to be found in  $|1\rangle$  and in  $|2\rangle$ , correct to second order in  $\Delta t$ , and show that these probabilities sum to 1, correct to second order in  $\Delta t$ .

## 11.2 Perturbation theory for cubic equations

Perturbation theory is any technique for approximately solving one problem, when an exact solution for a similar problem is available.

It's a general mathematical technique, applicable to many problems. (It was first developed in the context of classical mechanics: We have an exact solution for the problem two gravitating bodies, such as the ellipse of the Earth orbiting the Sun. But we don't have an exact solution for the problem of three gravitating bodies, such as the Earth plus the Sun plus Jupiter. Perturbation theory was developed to understand how the attraction by Jupiter "perturbed" the motion of the Earth away from the pure elliptical orbit that it would execute if Jupiter didn't exist.) Before we apply perturbation theory to quantum mechanics, we'll apply it in a simpler, and purely mathematical, context.

I wish to solve the cubic equation

$$x^3 - 4.001x + 0.002 = 0. \quad (11.26)$$

There is a formula for finding the three roots of a cubic equation, and we could use it to solve this problem. On the other hand, that formula is very complicated and awkward. And while there's no straightforward exact solution to the problem as stated, that problem is very close to the problem

$$x^3 - 4x = 0, \quad (11.27)$$

which *does* have straightforward exact solutions, namely

$$0, \pm 2. \quad (11.28)$$

Can I use the exact solution of this "nearby" problem to find an approximate solution for the problem of interest?

I'll write the cubic equation as the sum of a part we can solve plus a "small" perturbing part, namely

$$x^3 - 4x + (-0.001x + 0.002) = 0. \quad (11.29)$$

I place the word "small" in quotes because its meaning is not precisely clear. On one hand, for a typical value of  $x$ , say  $x = 1$ , the "big" part is  $-3$  while the small part is only  $0.001$ . On the other hand, for the value  $x = 0$ , the "big" part is zero and the "small" part is  $0.002$ . So for some values of  $x$  the "small" part is bigger than the "big" part. Mathematicians spend a lot of time figuring out a precise meaning of "big" versus "small"



in this context, but we don't need to follow their figurings. It's enough for us that the perturbing part is, in some general way, small compared to the remaining part of the problem, the part that we can solve exactly.

To save space, I'll introduce the constant  $T$  to mean "thousandths", and write our problem as

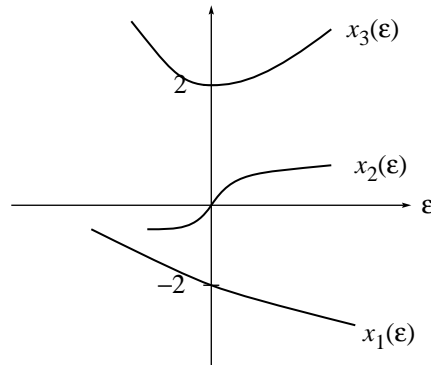
$$x^3 - 4x + T(-x + 2) = 0. \quad (11.30)$$

And now I'll generalize this problem by inserting a variable  $\epsilon$  in front of the "small" part:

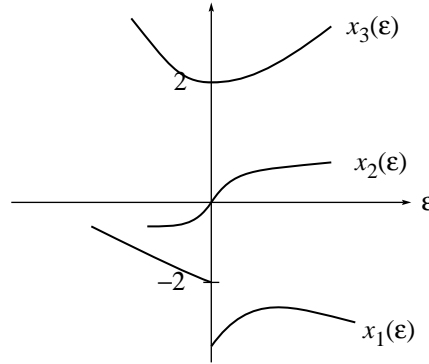
$$x^3 - 4x + \epsilon T(-x + 2) = 0. \quad (11.31)$$

The variable  $\epsilon$  enables us to interpolate smoothly from the problem we're interested in, with  $\epsilon = 1$ , to the problem we know how to solve, with  $\epsilon = 0$ .

Instead of solving one cubic equation, the problem with  $\epsilon = 1$ , we're going to try to solve an infinite number of cubic equations, those with  $0 \leq \epsilon \leq 1$ . For example, I can call the smallest of these solutions  $x_1(\epsilon)$ . I don't know much about  $x_1(\epsilon)$  — I know only that  $x_1(0) = -2$  — but I have an expectation: I expect that  $x_1(\epsilon)$  will behave smoothly as a function of  $\epsilon$ , for example something like this



and I expect that it won't have jumps or kinks like this



Because of this expectation, I expect that I can write  $x_1(\epsilon)$  as a Taylor series:

$$x_1(\epsilon) = \sum_{i=1}^{\infty} a_i \epsilon^i \quad (11.32)$$

$$= -2 + a_1 \epsilon + a_2 \epsilon^2 + \mathcal{O}(\epsilon^3) \quad (11.33)$$

This function  $x_1(\epsilon)$  has to satisfy

$$x_1^3(\epsilon) - (4 + \epsilon T)x_1(\epsilon) + 2\epsilon T = 0. \quad (11.34)$$

I can write the middle term above as an expansion in powers of  $\epsilon$  using equation (11.33):

$$\begin{aligned} -4x_1(\epsilon) &= 8 - \epsilon(4a_1) && - \epsilon^2(4a_2) && + \mathcal{O}(\epsilon^3) \\ -\epsilon T x_1(\epsilon) &= + \epsilon(2T) && - \epsilon^2(Ta_1) && + \mathcal{O}(\epsilon^3) \\ -(4 + \epsilon T)x_1(\epsilon) &= 8 + \epsilon(-4a_1 + 2T) && + \epsilon^2(-4a_2 - Ta_1) && + \mathcal{O}(\epsilon^3) \end{aligned}$$

With just a bit more effort, I can work out the left-most term in equation (11.34) as an expansion:

$$\begin{aligned} x_1^2(\epsilon) &= 4 - \epsilon(4a_1) + \epsilon^2(-4a_2 + a_1^2) + \mathcal{O}(\epsilon^3) \\ x_1^3(\epsilon) &= -8 - \epsilon(-12a_1) + \epsilon^2(12a_2 - 6a_1^2) + \mathcal{O}(\epsilon^3) \end{aligned}$$

So finally, I have worked out the expansion of every term in equation (11.34):

$$\begin{aligned} x_1^3(\epsilon) &= -8 - \epsilon(-12a_1) + \epsilon^2(12a_2 - 6a_1^2) + \mathcal{O}(\epsilon^3) \\ -(4 + \epsilon T)x_1(\epsilon) &= 8 + \epsilon(-4a_1 + 2T) + \epsilon^2(-4a_2 - Ta_1) + \mathcal{O}(\epsilon^3) \\ 2\epsilon T &= + \epsilon(2T) \end{aligned}$$

Summing the three equations above must, according to equation (11.34), produce zero:

$$\begin{aligned} 0 &= (-8 + 8) + \epsilon(12a_1 - 4a_1 + 4T) + \epsilon^2(12a_2 - 6a_1^2 - 4a_2 - Ta_1) + \mathcal{O}(\epsilon^3) \\ 0 &= (-8 + 8) + \epsilon(8a_1 + 4T) + \epsilon^2(8a_2 - 6a_1^2 - Ta_1) + \mathcal{O}(\epsilon^3) \end{aligned}$$

Now, because the expression on the right must vanish for any value of  $\epsilon$ , all the coefficients must vanish. First we must have that  $(-8 + 8) = 0$ , which checks out. Then the term linear in  $\epsilon$  must vanish, so

$$(8a_1 + 4T) = 0 \quad \text{whence} \quad a_1 = -\frac{1}{2}T.$$

And the term quadratic in  $\epsilon$  must vanish, so

$$(8a_2 - 6a_1^2 - Ta_1) = 0 \quad \text{whence} \quad a_2 = \frac{3}{4}a_1^2 + \frac{1}{8}Ta_1 = \frac{1}{8}T^2.$$

The expansion for  $x_1(\epsilon)$  is thus

$$x_1(\epsilon) = -2 - \frac{1}{2}T\epsilon + \frac{1}{8}T^2\epsilon^2 + \mathcal{O}(\epsilon^3)$$

If we set  $\epsilon = 1$  and ignore the terms  $\mathcal{O}(\epsilon^3)$ , we find

$$x_1(1) \approx -2.000399875$$

and comparison to the exact solution of the cubic equation (which is much more difficult to work through) shows that this result is accurate to one part in a billion.

### 11.3 Derivation of perturbation theory for the energy eigenproblem

#### *Approach*

To solve the energy eigenproblem for the Hamiltonian  $\hat{H}^{(0)} + \hat{H}'$ , where the solution

$$\hat{H}^{(0)}|n^{(0)}\rangle = E_n^{(0)}|n^{(0)}\rangle \quad (11.35)$$

is known and where  $\hat{H}'$  is “small” compared with  $\hat{H}^{(0)}$ , we set

$$\hat{H}(\epsilon) = \hat{H}^{(0)} + \epsilon\hat{H}' \quad (11.36)$$

and then find  $|n(\epsilon)\rangle$  and  $E_n(\epsilon)$  such that

$$\hat{H}(\epsilon)|n(\epsilon)\rangle = E_n(\epsilon)|n(\epsilon)\rangle \quad (11.37)$$

and

$$\langle n(\epsilon)|n(\epsilon)\rangle = 1. \quad (11.38)$$

*Intermediate goal*

Find  $|\bar{n}(\epsilon)\rangle$  and  $E_n(\epsilon)$  such that

$$\hat{H}(\epsilon)|\bar{n}(\epsilon)\rangle = E_n(\epsilon)|\bar{n}(\epsilon)\rangle \quad (11.39)$$

and

$$\langle n^{(0)}|\bar{n}(\epsilon)\rangle = 1. \quad (11.40)$$

Then our final goal will be

$$|n(\epsilon)\rangle = \frac{|\bar{n}(\epsilon)\rangle}{\langle \bar{n}(\epsilon)|\bar{n}(\epsilon)\rangle^{1/2}}. \quad (11.41)$$

Remarkably, it often turns out to be good enough to reach our intermediate goal of finding  $|\bar{n}(\epsilon)\rangle$ , and one can then invent tricks for extracting information from these unnormalized eigenstates.

*Initial assumption*

We make the standard perturbation theory guess:

$$|\bar{n}(\epsilon)\rangle = |n^{(0)}\rangle + \epsilon|\bar{n}^{(1)}\rangle + \epsilon^2|\bar{n}^{(2)}\rangle + \mathcal{O}(\epsilon^3) \quad (11.42)$$

$$E_n(\epsilon) = E_n^{(0)} + \epsilon E_n^{(1)} + \epsilon^2 E_n^{(2)} + \mathcal{O}(\epsilon^3) \quad (11.43)$$

[Note that the set  $\{|\bar{n}^{(1)}\rangle\}$  is not complete, or orthonormal, or any other good thing.]

*Consequences of the magnitude choice*

The choice  $\langle n^{(0)}|\bar{n}(\epsilon)\rangle = 1$  gives rise to interesting and useful consequences. First, take the inner product of  $|n^{(0)}\rangle$  with equation (11.42)

$$\begin{aligned} \langle n^{(0)}|\bar{n}(\epsilon)\rangle &= \langle n^{(0)}|n^{(0)}\rangle + \epsilon\langle n^{(0)}|\bar{n}^{(1)}\rangle + \epsilon^2\langle n^{(0)}|\bar{n}^{(2)}\rangle + \mathcal{O}(\epsilon^3) \\ 1 &= 1 + \epsilon\langle n^{(0)}|\bar{n}^{(1)}\rangle + \epsilon^2\langle n^{(0)}|\bar{n}^{(2)}\rangle + \mathcal{O}(\epsilon^3) \end{aligned}$$

Because this relationship holds for *all* values of  $\epsilon$ , the coefficient of each  $\epsilon^m$  must vanish:

$$\langle n^{(0)}|\bar{n}^{(m)}\rangle = 0 \quad m = 1, 2, 3, \dots \quad (11.44)$$

Whence

$$\begin{aligned}
\langle \bar{n}(\epsilon) | \bar{n}(\epsilon) \rangle &= \left[ \langle n^{(0)} | + \epsilon \langle \bar{n}^{(1)} | + \epsilon^2 \langle \bar{n}^{(2)} | + \mathcal{O}(\epsilon^3) \right] \left[ |n^{(0)}\rangle + \epsilon |\bar{n}^{(1)}\rangle + \epsilon^2 |\bar{n}^{(2)}\rangle + \mathcal{O}(\epsilon^3) \right] \\
&= \langle n^{(0)} | n^{(0)} \rangle + \epsilon \left[ \langle \bar{n}^{(1)} | n^{(0)} \rangle + \langle n^{(0)} | \bar{n}^{(1)} \rangle \right] + \epsilon^2 \left[ \langle \bar{n}^{(2)} | n^{(0)} \rangle + \langle \bar{n}^{(1)} | \bar{n}^{(1)} \rangle + \langle n^{(0)} | \bar{n}^{(2)} \rangle \right] + \mathcal{O}(\epsilon^3) \\
&= 1 + \epsilon \left[ 0 + 0 \right] + \epsilon^2 \left[ 0 + \langle \bar{n}^{(1)} | \bar{n}^{(1)} \rangle + 0 \right] + \mathcal{O}(\epsilon^3) \\
&= 1 + \epsilon^2 \langle \bar{n}^{(1)} | \bar{n}^{(1)} \rangle + \mathcal{O}(\epsilon^3).
\end{aligned} \tag{11.45}$$

In other words, while the vector  $|\bar{n}(\epsilon)\rangle$  is not exactly normalized, it is “nearly normalized” — the norm differs from 1 by small, second-order terms.

#### *Developing the perturbation expansion*

What came before was just warming up. We now go and plug our expansion guesses, equations (11.42) and (11.43) into

$$\hat{H}(\epsilon)|n(\epsilon)\rangle = E_n(\epsilon)|n(\epsilon)\rangle \tag{11.46}$$

to find

$$\begin{aligned}
&\left[ \hat{H}^{(0)} + \epsilon \hat{H}' \right] \left[ |n^{(0)}\rangle + \epsilon |\bar{n}^{(1)}\rangle + \epsilon^2 |\bar{n}^{(2)}\rangle + \mathcal{O}(\epsilon^3) \right] \\
&= \left[ E_n^{(0)} + \epsilon E_n^{(1)} + \epsilon^2 E_n^{(2)} + \mathcal{O}(\epsilon^3) \right] \left[ |n^{(0)}\rangle + \epsilon |\bar{n}^{(1)}\rangle + \epsilon^2 |\bar{n}^{(2)}\rangle + \mathcal{O}(\epsilon^3) \right]
\end{aligned} \tag{11.47}$$

Separating out powers of  $\epsilon$  gives

$$\hat{H}^{(0)}|n^{(0)}\rangle = E_n^{(0)}|n^{(0)}\rangle \tag{11.48}$$

$$\hat{H}^{(0)}|\bar{n}^{(1)}\rangle + \hat{H}'|n^{(0)}\rangle = E_n^{(1)}|n^{(0)}\rangle + E_n^{(0)}|\bar{n}^{(1)}\rangle \tag{11.49}$$

$$\hat{H}^{(0)}|\bar{n}^{(2)}\rangle + \hat{H}'|\bar{n}^{(1)}\rangle = E_n^{(2)}|n^{(0)}\rangle + E_n^{(1)}|\bar{n}^{(1)}\rangle + E_n^{(0)}|\bar{n}^{(2)}\rangle \tag{11.50}$$

and so forth.

#### *Finding the first-order energy shifts*

How do we extract useful information from these expansion equations? Let's focus on what we know and what we want to find. We know  $\hat{H}^{(0)}$ ,  $\hat{H}'$ ,  $|n^{(0)}\rangle$ , and  $E_n^{(0)}$ . From equation (11.49) we will find  $E_n^{(1)}$  and  $|\bar{n}^{(1)}\rangle$ .

Knowing these, from equation (11.50) we will find  $E_n^{(2)}$  and  $|\bar{n}^{(2)}\rangle$ . And so forth.

To find the energy shifts  $E_n^{(1)}$ , we multiply equation (11.49) by  $\langle n^{(0)}|$  to find

$$\begin{aligned}\langle n^{(0)}|\hat{H}^{(0)}|\bar{n}^{(1)}\rangle + \langle n^{(0)}|\hat{H}'|n^{(0)}\rangle &= E_n^{(1)}\langle n^{(0)}|n^{(0)}\rangle + E_n^{(0)}\langle n^{(0)}|\bar{n}^{(1)}\rangle \\ E_n^{(0)}\langle n^{(0)}|\bar{n}^{(1)}\rangle + \langle n^{(0)}|\hat{H}'|n^{(0)}\rangle &= E_n^{(1)} + E_n^{(0)}\langle n^{(0)}|\bar{n}^{(1)}\rangle\end{aligned}\quad (11.51)$$

Or,

$$E_n^{(1)} = \langle n^{(0)}|\hat{H}'|n^{(0)}\rangle. \quad (11.52)$$

Often you need only these energies, not the states, and you can stop here. But if you do need the states...

#### *Finding the first-order state shifts*

We will find the state shifts  $|\bar{n}^{(1)}\rangle$  by finding all the components of  $|\bar{n}^{(1)}\rangle$  in the unperturbed basis  $\{|m^{(0)}\rangle\}$ .

Multiply equation (11.49) by  $\langle m^{(0)}|$  ( $m \neq n$ ) to find

$$\begin{aligned}\langle m^{(0)}|\hat{H}^{(0)}|\bar{n}^{(1)}\rangle + \langle m^{(0)}|\hat{H}'|n^{(0)}\rangle &= E_n^{(1)}\langle m^{(0)}|n^{(0)}\rangle + E_n^{(0)}\langle m^{(0)}|\bar{n}^{(1)}\rangle \\ E_m^{(0)}\langle m^{(0)}|\bar{n}^{(1)}\rangle + \langle m^{(0)}|\hat{H}'|n^{(0)}\rangle &= 0 + E_n^{(0)}\langle m^{(0)}|\bar{n}^{(1)}\rangle \\ \langle m^{(0)}|\hat{H}'|n^{(0)}\rangle &= (E_n^{(0)} - E_m^{(0)})\langle m^{(0)}|\bar{n}^{(1)}\rangle\end{aligned}\quad (11.53)$$

Now, if the state  $|n^{(0)}\rangle$  is non-degenerate, then  $E_m^{(0)} \neq E_n^{(0)}$  and we can divide both sides to find

$$\langle m^{(0)}|\bar{n}^{(1)}\rangle = \frac{\langle m^{(0)}|\hat{H}'|n^{(0)}\rangle}{E_n^{(0)} - E_m^{(0)}} \quad (m \neq n) \quad (11.54)$$

But we already know, from equation (11.44), that

$$\langle n^{(0)}|\bar{n}^{(1)}\rangle = 0. \quad (11.55)$$

So now all the projections  $\langle m^{(0)}|\bar{n}^{(1)}\rangle$  are known, and therefore the vector is known:

$$|\bar{n}^{(1)}\rangle = \sum_m |m^{(0)}\rangle \langle m^{(0)}|\bar{n}^{(1)}\rangle \quad (11.56)$$

In conclusion — if  $|n^{(0)}\rangle$  is non-degenerate

$$|\bar{n}^{(1)}\rangle = \sum_{m \neq n} |m^{(0)}\rangle \frac{\langle m^{(0)}|\hat{H}'|n^{(0)}\rangle}{E_n^{(0)} - E_m^{(0)}}. \quad (11.57)$$

### 11.4 Perturbation theory for the energy eigenproblem: Summary of results

Given: Solution for the  $\hat{H}^{(0)}$  eigenproblem:

$$\hat{H}^{(0)}|n^{(0)}\rangle = E_n^{(0)}|n^{(0)}\rangle \quad \langle n^{(0)}|n^{(0)}\rangle = 1. \quad (11.58)$$

Find: Solution for the  $\hat{H}^{(0)} + \epsilon\hat{H}'$  eigenproblem:

$$(\hat{H}^{(0)} + \epsilon\hat{H}')|n(\epsilon)\rangle = E_n(\epsilon)|n(\epsilon)\rangle \quad \langle n(\epsilon)|n(\epsilon)\rangle = 1. \quad (11.59)$$

Define the “matrix elements”

$$\langle n^{(0)}|\hat{H}'|m^{(0)}\rangle = H'_{nm}. \quad (11.60)$$

The solutions are (provided  $|n^{(0)}\rangle$  is not degenerate):

$$E_n(\epsilon) = E_n^{(0)} + \epsilon H'_{nn} + \epsilon^2 \sum_{m \neq n} \frac{H'_{nm}H'_{mn}}{E_n^{(0)} - E_m^{(0)}} + \mathcal{O}(\epsilon^3) \quad (11.61)$$

$$\begin{aligned} |n(\epsilon)\rangle = & |n^{(0)}\rangle \\ & + \epsilon \sum_{m \neq n} |m^{(0)}\rangle \frac{H'_{mn}}{E_n^{(0)} - E_m^{(0)}} \\ & + \epsilon^2 \left[ \sum_{m \neq n} \sum_{\ell \neq n} |m^{(0)}\rangle \frac{H'_{m\ell}H'_{\ell n}}{(E_n^{(0)} - E_m^{(0)})(E_n^{(0)} - E_\ell^{(0)})} \right. \\ & \left. - \sum_{m \neq n} |m^{(0)}\rangle \frac{H'_{nn}H'_{mn}}{(E_n^{(0)} - E_m^{(0)})^2} - |n^{(0)}\rangle \frac{1}{2} \sum_{m \neq n} \frac{H'_{nm}H'_{mn}}{(E_n^{(0)} - E_m^{(0)})^2} \right] \\ & + \mathcal{O}(\epsilon^3) \end{aligned} \quad (11.62)$$

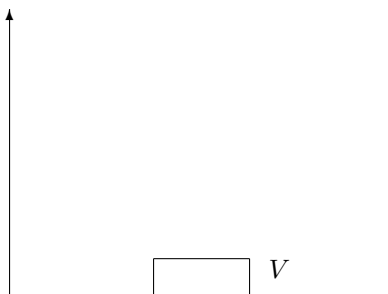
#### *Rules of thumb concerning perturbation theory*

- There is no guarantee that the series is convergent, or even asymptotic.
- But experience says “stop at the first non-vanishing energy correction”.
- The wavefunctions produced are notoriously poor. How can the energies be good when the wavefunctions are poor? See section 16.2.
- The technique is generally useful for many mathematical problems: classical mechanics, fluid mechanics, etc. Even for solving cubic equations!
- Technique is never guaranteed to succeed, but it is likely to fail (and perhaps fail silently!) if there are degenerate energy states. In this case  $E_n^{(0)} = E_m^{(0)}$ , so second-order term perhaps diverges, despite the fact that the first-order term  $\langle n^{(0)}|\hat{H}'|n^{(0)}\rangle$  looks perfectly fine.

## 11.5 Problems

### 11.2 Square well with a bump

An infinite square well of width  $L$  (problem 8.1) is perturbed by putting in a bit of potential of height  $V$  and width  $a$  in the middle of the well. Find the first order energy shifts for all the energy eigenstates, and the first order perturbed wavefunction for the ground state (your result will be an infinite series). (Note: Many of the required matrix elements will vanish! Before you integrate, ask yourself whether the integrand is odd.) When  $a = L$  the perturbed problem can be solved exactly. Compare the perturbed energies with the exact energies and the perturbed ground state wavefunction with the exact ground state wavefunction.



### 11.3 Anharmonic oscillator

- a. Show that for the simple harmonic oscillator,

$$\langle m | \hat{x}^3 | n \rangle = \sqrt{\left(\frac{\hbar}{2m\omega}\right)^3} \left[ \sqrt{n(n-1)(n-2)} \delta_{m,n-3} + 3\sqrt{n^3} \delta_{m,n-1} \right. \\ \left. + 3\sqrt{(n+1)^3} \delta_{m,n+1} + \sqrt{(n+1)(n+2)(n+3)} \delta_{m,n+3} \right].$$

- b. Recall that the simple harmonic oscillator is always an approximation. The real problem always has a potential  $V(x) = \frac{1}{2}kx^2 + bx^3 + cx^4 + \dots$ . The contributions beyond  $\frac{1}{2}kx^2$  are called “anharmonic terms”. Ignore all the anharmonic terms except for  $bx^3$ .



Show that to leading order the  $n^{\text{th}}$  energy eigenvalue changes by

$$-\frac{b^2}{\hbar\omega} \left( \frac{\hbar}{2m\omega} \right)^3 (30n^2 + 30n + 11). \quad (11.63)$$

Note that these shifts are not “small” when  $n$  is large, in which case it is not appropriate to truncate the perturbation series at leading order. Explain physically why you don’t expect the shifts to be small for large  $n$ .

#### 11.4 Slightly relativistic simple harmonic oscillator

You know that the concept of potential energy is not applicable in relativistic situations. One consequence of this is that the only fully relativistic quantum theories possible are quantum field theories. However there do exist situations where a particle’s motion is “slightly relativistic” (say,  $v/c \sim 0.1$ ) and where the force responds quickly enough to the particle’s position that the potential energy concept has approximate validity. For a mass on a spring, this situation hold when the spring’s response time is much less than the period.

- a. Show that a reasonable approximate Hamiltonian for such a “slightly relativistic SHO” is

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{m\omega^2}{2} \hat{x}^2 - \frac{1}{8c^2 m^3} \hat{p}^4. \quad (11.64)$$

- b. Show that

$$\langle m | \hat{p}^4 | 0 \rangle = \left( \frac{m\hbar\omega}{2} \right)^2 (3\delta_{m,0} - 6\sqrt{2}\delta_{m,2} + 2\sqrt{6}\delta_{m,4}). \quad (11.65)$$

- c. Calculate the leading non-vanishing energy shift of the ground state due to this relativistic perturbation.  
 d. Calculate the leading corrections to the ground state eigenvector  $|0\rangle$ .

#### 11.5 Two-state systems

The most general Hamiltonian for a two state system (e.g. spin  $\frac{1}{2}$ , neutral K meson, ammonia molecule) is represented by

$$a_0 I + a_1 \sigma_1 + a_3 \sigma_3 \quad (11.66)$$

where  $a_0$ ,  $a_1$ , and  $a_3$  are real numbers and the  $\sigma$ ’s are Pauli matrices. (See problem 58.)

- a. Assume  $a_3 = 0$ . Solve the energy eigenproblem.  
 b. Now assume  $a_3 \ll a_0 \approx a_1$ . Use perturbation theory to find the leading order shifts in the energy eigenvalues and eigenstates.

- c. Find the energy eigenvalues exactly and show that they agree with the perturbation theory results when  $a_3 \ll a_0 \approx a_1$ .

### 11.6 Degenerate perturbation theory in a two-state system

Consider a two state system with a Hamiltonian represented in some basis by

$$a_0 I + a_1 \sigma_1 + a_3 \sigma_3. \quad (11.67)$$

We shall call the basis for this representation the “initial basis”. This exercise shows how to use perturbation theory to solve (approximately) the energy eigenproblem in the case  $a_0 \gg a_1 \approx a_3$ .

$$\hat{H}^{(0)} = \begin{pmatrix} a_0 & 0 \\ 0 & a_0 \end{pmatrix} \quad \hat{H}' = \begin{pmatrix} a_3 & a_1 \\ a_1 & -a_3 \end{pmatrix} \quad (11.68)$$

In this case the unperturbed Hamiltonian is degenerate. The initial basis

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \quad (11.69)$$

is a perfectly acceptable energy eigenbasis (both states have energy  $a_0$ ), but the basis

$$\left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}, \quad (11.70)$$

for example, is just as good.

- a. Show that if the non-degenerate formula  $E_n^{(1)} = \langle n^{(0)} | \hat{H}' | n^{(0)} \rangle$  were applied (or rather, misapplied) to this problem, then the formula would produce different energy shifts depending upon which basis was used!

Which, if either, are the true energy shifts? The answer comes from equation (11.53), namely

$$(E_n^{(0)} - E_m^{(0)}) \langle m^{(0)} | \hat{n}^{(1)} \rangle = \langle m^{(0)} | \hat{H}' | n^{(0)} \rangle \quad \text{whenever } m \neq n. \quad (11.71)$$

This equation was derived from the fundamental assumption that  $|n(\epsilon)\rangle$  and  $E_n(\epsilon)$  could be expanded in powers of  $\epsilon$ . If the unperturbed states  $|n^{(0)}\rangle$  and  $|m^{(0)}\rangle$  are degenerate, then  $E_n^{(0)} = E_m^{(0)}$  and the above equation demands that

$$\langle m^{(0)} | \hat{H}' | n^{(0)} \rangle = 0 \quad \text{whenever } m \neq n \text{ and } E_n^{(0)} = E_m^{(0)}. \quad (11.72)$$

If this does not apply, then the fundamental assumption must be wrong. And this answers the question of which basis to use! Consistency demands the use of a basis in which the perturbing Hamiltonian is diagonal. (The Hermiticity of  $\hat{H}'$  guarantees that such a basis exists.)

- b. Without finding this diagonalizing basis, find the representation of  $\hat{H}'$  in it.
- c. Find the representation of  $\hat{H}^{(0)}$  in the diagonalizing basis. (Trick question.)
- d. What are the energy eigenvalues of the full Hamiltonian  $\hat{H}^{(0)} + \hat{H}'$ ? (Not “correct to some order in perturbation theory,” but the exact eigenvalues!)
- e. Still without explicitly producing the diagonalizing basis, show that the states in that basis are exact energy eigenstates of the full Hamiltonian.
- f. (Optional) If you’re ambitious, you may now go ahead and show that the (normalized) diagonalizing basis vectors are

$$\frac{1}{\sqrt{2}\sqrt{a_1^2 + a_3^2 - a_3\sqrt{a_1^2 + a_3^2}}} \begin{pmatrix} +a_1 \\ -a_3 + \sqrt{a_1^2 + a_3^2} \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \quad (73)$$

$$\frac{1}{\sqrt{2}\sqrt{a_1^2 + a_3^2 + a_3\sqrt{a_1^2 + a_3^2}}} \begin{pmatrix} -a_1 \\ +a_3 + \sqrt{a_1^2 + a_3^2} \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} \quad (74)$$

where

$$\tan\theta = \frac{a_1}{a_3 + \sqrt{a_1^2 + a_3^2}}. \quad (11.75)$$

*Coda:* Note the reasoning of degenerate perturbation theory: We expand about the basis that diagonalizes  $\hat{H}'$  because expansion about any other basis is immediately self-contradictory, not because this basis is guaranteed to produce a sensible expansion. As usual in perturbation theory, we have no guarantee that this expansion makes sense. We do, however, have a guarantee that any other expansion does not make sense.



## Chapter 12

# Quantum Mechanics in Two and Three Dimensions

### 12.1 More degrees of freedom

Let's think of the process of adding degrees of freedom.

First consider a spinless particle in one dimension:

- (1) The particle's state is described by a vector  $|\psi\rangle$ .
- (2) The vector has dimension  $\infty$ , reflecting the fact that any basis, for example the basis  $\{|x\rangle\}$ , has  $\infty$  elements. (No basis is better than another other basis — for every statement below concerning position there is a parallel statement concerning momentum — but for concreteness we'll discuss only position.)
- (3) These basis elements are orthonormal,

$$\langle x|x'\rangle = \delta(x - x'), \quad (12.1)$$

and complete

$$\hat{1} = \int_{-\infty}^{+\infty} dx |x\rangle\langle x|. \quad (12.2)$$

[[These two equations may seem recondite, formal, and purely mathematical, but in fact they embody the direct, physical results of measurement experiments: Completeness reflects the fact that when the particle's position is measured, it is found to have a position. Orthonormality reflects the fact that when the particle's position is measured, it is found in only one position. *Statement should be refined. Connection between completeness and interference?*]]

- (4) The state  $|\psi\rangle$  is represented (in the position basis) by the numbers  $\langle x|\psi\rangle = \psi(x)$ . In symbols

$$|\psi\rangle \doteq \langle x|\psi\rangle = \psi(x). \quad (12.3)$$

- (5) When the position is measured, the probability of finding the particle at a position within  $dx$  about  $x_0$  is

$$|\psi(x_0)|^2 dx. \quad (12.4)$$

Now consider a spin- $\frac{1}{2}$  particle in one dimension:

- (1) The particle's state is described by a vector  $|\psi\rangle$ .  
 (2) The vector has dimension  $\infty \times 2$ , reflecting the fact that any basis, for example the basis  $\{|x, +\rangle, |x, -\rangle\}$ , has  $\infty \times 2$  elements. (No basis is better than another other basis — for every statement below concerning position plus projection on a vertical axis there is a parallel statement concerning momentum plus projection of a horizontal axis — but for concreteness we'll discuss only position plus projection of a vertical axis.) [For example, the state  $|5, +\rangle$  represents a particle at position 5 with spin  $+$ . The state

$$\frac{1}{\sqrt{2}}[|5, +\rangle - |7, -\rangle]$$

represents a particle with amplitude  $1/\sqrt{2}$  to be at position 5 with spin  $+$  and amplitude  $-1/\sqrt{2}$  to be at position 7 with spin  $-$ , but with no amplitude to be at position 5 with spin  $-$ , and no amplitude to be at position 6 with any spin.]

- (3) These basis elements are orthonormal,

$$\begin{aligned} \langle x, + | x', + \rangle &= \delta(x - x') \\ \langle x, + | x', - \rangle &= 0 \\ \langle x, i | x', j \rangle &= \delta(x - x') \delta_{i,j} \end{aligned} \quad (12.5)$$

and complete

$$\begin{aligned} \hat{1} &= \int_{-\infty}^{+\infty} dx |x, +\rangle \langle x, +| + \int_{-\infty}^{+\infty} dx |x, -\rangle \langle x, -| \\ \hat{1} &= \sum_{i=+,-} \int_{-\infty}^{+\infty} dx |x, i\rangle \langle x, i| \end{aligned} \quad (12.6)$$

- (4) The state  $|\psi\rangle$  is represented (in this basis) by the numbers

$$\begin{pmatrix} \langle x, + | \psi \rangle \\ \langle x, - | \psi \rangle \end{pmatrix} = \begin{pmatrix} \psi_+(x) \\ \psi_-(x) \end{pmatrix}. \quad (12.7)$$

- (5) When both the spin projection and the position are measured, the probability of finding the particle with spin up and at a position within  $dx$  about  $x_0$  is

$$|\psi_+(x_0)|^2 dx. \quad (12.8)$$

The proper way of expressing the representation of the state  $|\psi\rangle$  in the  $\{|x, +\rangle, |x, -\rangle\}$  basis is through the so-called “spinor” above, namely

$$|\psi\rangle \doteq \begin{pmatrix} \psi_+(x) \\ \psi_-(x) \end{pmatrix}.$$

Sometimes you’ll see this written instead as

$$|\psi\rangle \doteq \psi_+(x)|+\rangle + \psi_-(x)|-\rangle.$$

Ugh! This is bad notation, because it confuses the state (something like  $|\psi\rangle$ , a vector) with the representation of a state in a particular basis (something like  $\langle x, i|\psi\rangle$ , a set of amplitudes). Nevertheless, you’ll see it used.

This example represents the way to add degrees of freedom to a description, namely by using a larger basis set. In this case I’ve merely doubled the size of the basis set, by including spin. I could also add a second dimension by adding the possibility of motion in the  $y$  direction, and so forth.

Consider a spinless particle in three dimensions:

- (1) The particle’s state is described by a vector  $|\psi\rangle$ .
- (2) The vector has dimension  $\infty^3$ , reflecting the fact that any basis, for example the basis  $\{|x, y, z\rangle\}$  — which is also written as  $\{|\mathbf{x}\rangle\}$  — has  $\infty^3$  elements. (No basis is better than another other basis — for every statement below concerning position there is a parallel statement concerning momentum — but for concreteness we’ll discuss only position.)
- (3) These basis elements are orthonormal,

$$\langle x, y, z|x', y', z'\rangle = \delta(x - x')\delta(y - y')\delta(z - z'), \quad (12.9)$$

which is also written as

$$\langle \mathbf{x}|\mathbf{x}'\rangle = \delta(\mathbf{x} - \mathbf{x}'). \quad (12.10)$$

In addition, the basis elements are complete

$$\hat{1} = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \int_{-\infty}^{+\infty} dz |x, y, z\rangle \langle x, y, z|, \quad (12.11)$$

which is also written as

$$\hat{1} = \int_{-\infty}^{+\infty} d^3x |\mathbf{x}\rangle \langle \mathbf{x}|. \quad (12.12)$$

- (4) The state  $|\psi\rangle$  is represented (in the position basis) by the numbers  $\langle \mathbf{x}|\psi\rangle = \psi(\mathbf{x})$  (a complex-valued function of a vector argument).
- (5) When the position is measured, the probability of finding the particle at a position within  $d^3x$  about  $\mathbf{x}_0$  is

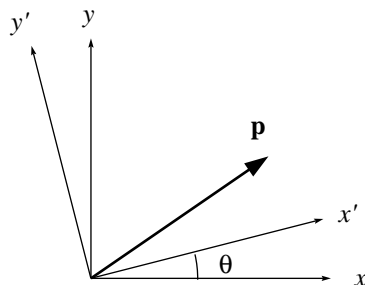
$$|\psi(\mathbf{x}_0)|^2 d^3x. \quad (12.13)$$

## 12.2 Vector operators

So much for states... what about operators?

The general idea of a vector is that it's "something like an arrow". But in what way like an arrow? If you work with the components of a vector, how can the components tell you that they represent something that's "like an arrow"?

Consider the vector momentum  $\mathbf{p}$ . If the coordinate axes are  $x$  and  $y$ , the components of the vector  $\mathbf{p}$  are  $p_x$  and  $p_y$ . But if the coordinate axes are  $x'$  and  $y'$ , then the components of the vector  $\mathbf{p}$  are  $p_{x'}$  and  $p_{y'}$ . It's the same vector, but it has different components using different coordinate axes.



How are these two sets of coordinates related? It's not hard to show that they're related through

$$\begin{aligned} p_{x'} &= p_x \cos \theta + p_y \sin \theta \\ p_{y'} &= -p_x \sin \theta + p_y \cos \theta \end{aligned} \quad (12.14)$$

(There's a similar but more complicated formula for three-dimensional vectors.)

We use this same formula for change of coordinates under rotation whether it's a position vector or a velocity vector or a momentum vector, despite the fact that position, velocity, and momentum are very different in character. It is in this sense that position, velocity, and momentum are all "like an arrow" and it is in this way that the components of a vector show that the entity behaves "like an arrow".



Now, what is a “vector operator”? In two dimensions, it’s a set of two operators that transform under rotation just as the two components of a vector do:

$$\begin{aligned}\hat{p}_{x'} &= \hat{p}_x \cos \theta + \hat{p}_y \sin \theta \\ \hat{p}_{y'} &= -\hat{p}_x \sin \theta + \hat{p}_y \cos \theta\end{aligned}\quad (12.15)$$

(There’s a similar but more complicated formula for three-dimensional vector operators.)

Meanwhile, a “scalar operator” is one that doesn’t change when the coordinate axes are rotated.

For every vector operator there is a scalar operator

$$\hat{p}^2 = \hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2. \quad (12.16)$$

### 12.3 Multiple particles

In section 12.1 we considered adding spin and spatial degrees of freedom for a single particle. But the same scheme works for adding additional particles. (There are peculiarities that apply to the identical particles — see chapter 15 — so in this section we’ll consider non-identical particles.)

Consider a system of two spinless particles (call them red and green) moving in one dimension:

- (1) The system’s state is described by a vector  $|\psi\rangle$ .
- (2) The vector has dimension  $\infty^2$ , reflecting the fact that any basis, for example the basis  $\{|x_R, x_G\rangle\}$  has  $\infty^2$  elements. (No basis is better than another other basis — for every statement below concerning two positions there is a parallel statement concerning two momenta — but for concreteness we’ll discuss only position.)
- (3) These basis elements are orthonormal,

$$\langle x_R, x_G | x'_R, x'_G \rangle = \delta(x_R - x'_R) \delta(x_G - x'_G). \quad (12.17)$$

In addition, the basis elements are complete

$$\hat{1} = \int_{-\infty}^{+\infty} dx_R \int_{-\infty}^{+\infty} dx_G |x_R, x_G\rangle \langle x_R, x_G|. \quad (12.18)$$

- (4) The state  $|\psi\rangle$  is represented (in the position basis) by the numbers  $\langle x_R, x_G | \psi \rangle = \psi(x_R, x_G)$  (a complex-valued function of a two-variable argument).

- (5) When the positions of both particles are measured, the probability of finding the red particle within a window of width  $dx_A$  about  $x_A$  and the green particle within a window of width  $dx_B$  about  $x_B$  is

$$|\psi(x_A, x_B)|^2 dx_A dx_B. \quad (12.19)$$

## 12.4 The phenomena of quantum mechanics

We started (chapter 1) with the phenomena of quantum mechanics: quantization, probability, interference, and entanglement. We used these phenomena to build up the formalism of quantum mechanics: amplitudes, state vectors, operators, etc. (chapter 2).

We've been working at the level of formalism for so long that we're in danger of forgetting the phenomena that underlie the formalism: For example in this chapter we discussed how the formalism of quantum mechanics applies to continuum systems in three dimensions. It's time to return to the level of phenomena and ask how the phenomena of quantum mechanics generalize to continuum systems in three dimensions.

### *Interference*

Interference of a particle — experiments of Tonomura:

A. Tonomura, J. Endo, T. Matsuda, T. Kawasaki, and H. Ezawa, "Demonstration of single-electron buildup of an interference pattern," *American Journal of Physics*, **57** (1989) 117–120.

<http://www.hqrd.hitachi.co.jp/em/doubleslit.cfm>

### *Entanglement*

How does one describe the state of a single classical particle moving in one dimension? It requires two numbers: a position and a momentum (or a position and a velocity). Two particles moving in one dimension require merely that we specify the state of each particle: four numbers. Similarly specifying the state of three particles require six numbers and  $N$  particles require  $2N$  numbers. Exactly the same specification counts hold if the particle moves relativistically.

How, in contrast, does one describe the state of a single quantal particle moving in one dimension? A problem arises at the very start, here, because the specification is given through a complex-valued wavefunction  $\psi(x)$ . Technically the specification requires an infinite number of numbers! Let's approximate the wavefunction through its value on a grid of, say, 100 points. This suggests that a specification requires 200 real numbers, a complex number at each grid point, but one number is taken care of through the overall phase of the wavefunction, and one through normalization. The specification actually requires 198 independent real numbers.

How does one describe the state of two quantal particles moving in one dimension? Now the wavefunction is a function of two variables  $\psi(x_A, x_B)$ . (This wavefunction might factorize into a function of  $x_A$  alone times a function of  $x_B$  alone, but it might not. If it does factorize, the two particles are unentangled, if it does not, the two particles are entangled. In the general quantal case a two-particle state is *not* specified by giving the state of each individual particle, because the individual particles might not have states.) The wavefunction of the system is a function of two-dimensional configuration space, so an approximation of the accuracy established previously requires a  $100 \times 100$  grid of points. Each grid point carries one complex number, and again overall phase and normalization reduce the number of real numbers required by two. For two particles the specification requires  $2 \times (100)^2 - 2 = 19998$  independent real numbers.

Similarly, specifying the state of  $N$  quantal particles moving in one dimension requires a wavefunction in  $N$ -dimensional configuration space which (for a grid of the accuracy we've been using) is specified through  $2 \times (100)^N - 2$  independent real numbers.

The specification of a quantal state not only requires more real numbers than the specification of the corresponding classical state, but that number increases exponentially rather than linearly with the number of particles  $N$ .

The fact that a quantal state holds more information than a classical state is the fundamental reason that a quantal computer is (in principle) faster than a classical computer, and the basis for much of quantum information theory.

Relativity is different from classical physics, but no more complicated. Quantum mechanics, in contrast, is both *different* from and *richer* than classical physics. You may refer to this richness using terms like "splendor",

or “abounding”, or “intricate”, or “ripe with possibilities”. Or you may refer to it using terms like “complicated”, or “messy”, or “full of details likely to trip the innocent”. It’s your choice how to react to this richness, but you can’t deny it.

## Chapter 13

# Angular Momentum

### 13.1 Solution of the angular momentum eigenproblem

We solved the simple harmonic oscillator energy eigenproblem twice: once using a straightforward but laborious differential equation technique, and then again using an operator-factorization technique that was much easier to implement, but which involved unmotivated creative leaps. We'll do the same with the angular momentum eigenproblem, but in the opposite sequence.

Here's the problem:

*Given Hermitian operators  $\hat{J}_x$ ,  $\hat{J}_y$ ,  $\hat{J}_z$  obeying*

$$[\hat{J}_x, \hat{J}_y] = i\hbar\hat{J}_z, \quad \text{and cyclic permutations} \quad (13.1)$$

*find the eigenvalues and eigenvectors for one such operator, say  $\hat{J}_z$ .*

Any other component of angular momentum, say  $\hat{J}_x$  or  $\hat{J}_{42^\circ}$ , will have exactly the same eigenvalues, and eigenvectors with the same structure.

Note that the we are to solve the problem using *only* the commutation relations — we are not to use, say, the expression for the angular momentum operator in the position basis, or the relationship between angular momentum and rotation.

Strangely, our first step is to slightly expand the problem. (I warned you that the solution would not take a straightforward, “follow your nose” path.)

*Define*

$$\hat{J}^2 = \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2 \quad (13.2)$$

and note that

$$[\hat{J}^2, \hat{J}_i] = 0 \quad \text{for } i = x, y, z. \quad (13.3)$$

Because  $\hat{J}^2$  and  $\hat{J}_z$  commute, they have a basis of simultaneous eigenvectors. We expand the problem to find these simultaneous eigenvectors  $|\beta, \mu\rangle$ , which satisfy

$$\hat{J}^2|\beta, \mu\rangle = \hbar^2\beta|\beta, \mu\rangle \quad (13.4)$$

$$\hat{J}_z|\beta, \mu\rangle = \hbar\mu|\beta, \mu\rangle \quad (13.5)$$

We define the values  $\beta$  and  $\mu$  in this way so that  $\beta$  and  $\mu$  will be dimensionless. (If this is not obvious to you, show it now. Also, if it's not obvious that the equations (13.3) follow from the equations (13.1), you should show that, too. What is the commutator  $[\hat{J}^2, \hat{J}_{28^\circ}]$ ?)

Start off by noting that

$$(\hat{J}_x^2 + \hat{J}_y^2)|\beta, \mu\rangle = (\hat{J}^2 - \hat{J}_z^2)|\beta, \mu\rangle = \hbar^2(\beta - \mu^2)|\beta, \mu\rangle. \quad (13.6)$$

Now the first operator  $(\hat{J}_x^2 + \hat{J}_y^2)$  would be  $(\hat{J}_x - i\hat{J}_y)(\hat{J}_x + i\hat{J}_y)$  if  $\hat{J}_x$  and  $\hat{J}_y$  were numbers. The factorization is not in fact quite that clean, because those operators are not in fact numbers. But we use this to inspire the definitions

$$\hat{J}_- = \hat{J}_x - i\hat{J}_y \quad \text{and} \quad \hat{J}_+ = \hat{J}_x + i\hat{J}_y \quad (13.7)$$

so that

$$\hat{J}_-\hat{J}_+ = \hat{J}_x^2 + \hat{J}_y^2 + i(\hat{J}_x\hat{J}_y - \hat{J}_y\hat{J}_x) = \hat{J}_x^2 + \hat{J}_y^2 + i[\hat{J}_x, \hat{J}_y] = \hat{J}_x^2 + \hat{J}_y^2 - \hbar\hat{J}_z. \quad (13.8)$$

This tells us that

$$\hat{J}_-\hat{J}_+|\beta, \mu\rangle = (\hbar^2\beta - \hbar^2\mu^2 - \hbar^2\mu)|\beta, \mu\rangle = \hbar^2(\beta - \mu(\mu + 1))|\beta, \mu\rangle. \quad (13.9)$$

We have immediately that

$$\langle\beta, \mu|\hat{J}_-\hat{J}_+|\beta, \mu\rangle = \hbar^2(\beta - \mu(\mu + 1)). \quad (13.10)$$

But if we define

$$|\phi\rangle = \hat{J}_+|\beta, \mu\rangle \quad \text{then} \quad \langle\phi| = \langle\beta, \mu|\hat{J}_-$$

then equation (13.10) is just the expression for  $\langle\phi|\phi\rangle$ , and we know that for any vector  $\langle\phi|\phi\rangle \geq 0$ . Thus

$$\beta \geq \mu(\mu + 1). \quad (13.11)$$

With these preliminaries out of the way, we investigate the operator  $\hat{J}_+$ . First, its commutation relations:

$$[\hat{J}^2, \hat{J}_+] = 0, \quad (13.12)$$

$$[\hat{J}_z, \hat{J}_+] = [\hat{J}_z, \hat{J}_x] + i[\hat{J}_z, \hat{J}_y] = (i\hbar\hat{J}_y) + i(-i\hbar\hat{J}_x) = \hbar\hat{J}_+. \quad (13.13)$$

Then, use the commutation relations to find the effect of  $\hat{J}_+$  on  $|\beta, \mu\rangle$ . If we again define  $|\phi\rangle = \hat{J}_+|\beta, \mu\rangle$ , then

$$\hat{J}^2|\phi\rangle = \hat{J}^2\hat{J}_+|\beta, \mu\rangle = \hat{J}_+\hat{J}^2|\beta, \mu\rangle = \hbar^2\beta\hat{J}_+|\beta, \mu\rangle = \hbar^2\beta|\phi\rangle, \quad (13.14)$$

$$\hat{J}_z|\phi\rangle = \hat{J}_z\hat{J}_+|\beta, \mu\rangle = (\hat{J}_+\hat{J}_z + \hbar\hat{J}_+)|\beta, \mu\rangle = \hbar\mu\hat{J}_+|\beta, \mu\rangle + \hbar\hat{J}_+|\beta, \mu\rangle = \hbar(\mu + 1)|\phi\rangle. \quad (13.15)$$

That is, the vector  $|\phi\rangle$  is an eigenvector of  $\hat{J}^2$  with eigenvalue  $\beta$  and an eigenvector of  $\hat{J}_z$  with eigenvalue  $\mu + 1$ . In other words,

$$\hat{J}_+|\beta, \mu\rangle = C|\beta, \mu + 1\rangle \quad (13.16)$$

where  $C$  is a normalization factor to be determined.

To find  $C$ , we contrast

$$\langle\phi|\phi\rangle = |C|^2\langle\beta, \mu|\beta, \mu\rangle = |C|^2 \quad (13.17)$$

with the result of equation (13.10), namely

$$\langle\phi|\phi\rangle = \langle\beta, \mu|\hat{J}_-\hat{J}_+|\beta, \mu\rangle = \hbar^2(\beta - \mu(\mu + 1)). \quad (13.18)$$

From this we may select  $C = \hbar\sqrt{\beta - \mu(\mu + 1)}$  so that

$$\hat{J}_+|\beta, \mu\rangle = \hbar\sqrt{\beta - \mu(\mu + 1)}|\beta, \mu + 1\rangle. \quad (13.19)$$

In short, the operator  $\hat{J}_+$  applied to  $|\beta, \mu\rangle$  acts as a *raising operator*: it doesn't change the value of  $\beta$ , but it increases the value of  $\mu$  by 1.

Parallel reasoning applied to  $\hat{J}_-$  shows that

$$\hat{J}_-|\beta, \mu\rangle = \hbar\sqrt{\beta - \mu(\mu - 1)}|\beta, \mu - 1\rangle. \quad (13.20)$$

In short, the operator  $\hat{J}_-$  applied to  $|\beta, \mu\rangle$  acts as a *lowering operator*: it doesn't change the value of  $\beta$ , but it decreases the value of  $\mu$  by 1.

At first it might appear that we could use these raising or lowering operators to ascend to infinitely high heavens or to dive to infinitely low depths, but that appearance is incorrect. Equation (13.11),

$$\beta \geq \mu(\mu + 1), \quad (13.21)$$

will necessarily be violated for sufficiently high or sufficiently low values of  $\mu$ . Instead, there must be some maximum value of  $\mu$  — call it  $\mu_{\max}$  —

such that an attempt to raise  $|\beta, \mu_{\max}\rangle$  results not in a vector proportional to  $|\beta, \mu_{\max} + 1\rangle$ , but results instead in 0. It is clear from equation (13.19) that this value of  $\mu$  satisfies

$$\beta - \mu_{\max}(\mu_{\max} + 1) = 0. \quad (13.22)$$

And it's equally clear from equation (13.20) that there is a minimum value  $\mu_{\min}$  satisfying

$$\beta - \mu_{\min}(\mu_{\min} - 1) = 0. \quad (13.23)$$

Solving these two equations simultaneously, we find that

$$\mu_{\max} = -\mu_{\min} \quad \text{with} \quad \mu_{\max} \geq 0 \quad (13.24)$$

and that

$$\beta = \mu_{\max}(\mu_{\max} + 1). \quad (13.25)$$

But there's more. Because we raise or lower  $\mu$  by 1 with each application of  $\hat{J}_+$  or  $\hat{J}_-$ , the value of  $\mu_{\max}$  must be an integer above  $\mu_{\min}$ :

$$\begin{aligned} \mu_{\max} &= \mu_{\min} + (\text{an integer}) \\ 2\mu_{\max} &= (\text{an integer}) \\ \mu_{\max} &= \frac{\text{an integer}}{2} \geq 0 \end{aligned} \quad (13.26)$$

Common practice is to call the half-integer  $\mu_{\max}$  by the name  $j$ , and the half-integer  $\mu$  by the name  $m$ . And common practice is to label the angular momentum state not as  $|\beta, \mu\rangle$  but as  $|j, m\rangle$ , which contains equivalent information. Using these conventions, the solution to the angular momentum eigenvalue problem is:

The eigenvalues of  $\hat{J}^2$  are

$$\hbar^2 j(j+1) \quad j = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots \quad (13.27)$$

For a given  $j$ , the eigenvalues of  $\hat{J}_z$  are

$$\hbar m \quad m = -j, -j+1, \dots, j-1, j. \quad (13.28)$$

The eigenstates  $|j, m\rangle$  are related through the operators

$$\hat{J}_+ = \hat{J}_x + i\hat{J}_y \quad \hat{J}_- = \hat{J}_x - i\hat{J}_y \quad (13.29)$$

by

$$\hat{J}_+ |j, m\rangle = \hbar \sqrt{j(j+1) - m(m+1)} |j, m+1\rangle \quad (13.30)$$

$$\hat{J}_- |j, m\rangle = \hbar \sqrt{j(j+1) - m(m-1)} |j, m-1\rangle. \quad (13.31)$$

### Problems:

- (1) Write out the "parallel reasoning" that results in equation (13.20).
- (2) The simultaneous solution of equations (13.22) and (13.23) results in two possible solutions, namely (13.24) and  $\mu_{\min} = \mu_{\max} + 1$ . Why do we reject this second solution? Why do we, in equation (13.24), insert the proviso  $\mu_{\max} \geq 0$ ?



### 13.2 Summary of the angular momentum eigenproblem

Given  $[\hat{J}_x, \hat{J}_y] = i\hbar\hat{J}_z$ , and cyclic permutations, the eigenvalues of  $\hat{J}^2$  are

$$\hbar^2 j(j+1) \quad j = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots$$

For a given  $j$ , the eigenvalues of  $\hat{J}_z$  are

$$\hbar m \quad m = -j, -j+1, \dots, j-1, j.$$

The eigenstates  $|j, m\rangle$  are related through the operators

$$\hat{J}_+ = \hat{J}_x + i\hat{J}_y \quad \hat{J}_- = \hat{J}_x - i\hat{J}_y$$

by

$$\begin{aligned} \hat{J}_+ |j, m\rangle &= \hbar\sqrt{j(j+1) - m(m+1)} |j, m+1\rangle \\ \hat{J}_- |j, m\rangle &= \hbar\sqrt{j(j+1) - m(m-1)} |j, m-1\rangle. \end{aligned}$$

### 13.3 Ordinary differential equations for the $d_{m,m'}^{(j)}(\theta)$

$$\begin{aligned} d_{m,m'}^{(j)}(\theta) &= \langle j, m | e^{-i(\hat{J}_y/\hbar)\theta} |j, m'\rangle \\ \frac{d}{d\theta} \left[ d_{m,m'}^{(j)}(\theta) \right] &= \langle j, m | e^{-i(\hat{J}_y/\hbar)\theta} (-i\hat{J}_y/\hbar) |j, m'\rangle \\ &= -\frac{1}{2\hbar} \langle j, m | e^{-i(\hat{J}_y/\hbar)\theta} (\hat{J}_+ - \hat{J}_-) |j, m'\rangle \\ &= -\frac{1}{2} \sqrt{j(j+1) - m'(m'+1)} \langle j, m | e^{-i(\hat{J}_y/\hbar)\theta} |j, m'+1\rangle \\ &\quad + \frac{1}{2} \sqrt{j(j+1) - m'(m'-1)} \langle j, m | e^{-i(\hat{J}_y/\hbar)\theta} |j, m'-1\rangle \end{aligned}$$

$$\frac{d}{d\theta} \left[ d_{m,m'}^{(j)}(\theta) \right] = -\frac{1}{2} \sqrt{j(j+1) - m'(m'+1)} d_{m,m'+1}^{(j)}(\theta) + \frac{1}{2} \sqrt{j(j+1) - m'(m'-1)} d_{m,m'-1}^{(j)}(\theta)$$

For a given  $j$  and  $m$ , these are  $2j+1$  coupled first-order ODEs. In matrix form they are

$$\frac{d}{d\theta} \begin{pmatrix} d_{m,j}^{(j)}(\theta) \\ d_{m,j-1}^{(j)}(\theta) \\ d_{m,j-2}^{(j)}(\theta) \\ d_{m,j-3}^{(j)}(\theta) \\ \vdots \\ d_{m,-j+1}^{(j)}(\theta) \\ d_{m,-j}^{(j)}(\theta) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & \sqrt{j} & 0 & 0 & \dots & 0 & 0 \\ -\sqrt{j} & 0 & \sqrt{2j-1} & 0 & \dots & 0 & 0 \\ 0 & -\sqrt{2j-1} & 0 & \sqrt{3j-3} & \dots & 0 & 0 \\ 0 & 0 & -\sqrt{3j-3} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & \sqrt{j} \\ 0 & 0 & 0 & 0 & \dots & -\sqrt{j} & 0 \end{pmatrix} \begin{pmatrix} d_{m,j}^{(j)}(\theta) \\ d_{m,j-1}^{(j)}(\theta) \\ d_{m,j-2}^{(j)}(\theta) \\ d_{m,j-3}^{(j)}(\theta) \\ \vdots \\ d_{m,-j+1}^{(j)}(\theta) \\ d_{m,-j}^{(j)}(\theta) \end{pmatrix}.$$

### 13.4 Problems

#### 13.1 Trivial pursuit

- Show that if an operator commutes with two components of an angular momentum vector, it commutes with the third as well.
- If  $\hat{J}_x$  and  $\hat{J}_z$  are represented by matrices with pure real entries (as is conventionally the case, see problem 13.2), show that  $\hat{J}_y$  is represented by a matrix with pure imaginary entries.

#### 13.2 Matrix representations for spin- $\frac{1}{2}$

If we are interested only in a particle's angular momentum, and not in its position, momentum, etc., then for a spin- $\frac{1}{2}$  particle the basis  $\{|\frac{1}{2}, \frac{1}{2}\rangle, |\frac{1}{2}, -\frac{1}{2}\rangle\}$  spans the relevant states. These states are usually denoted simply  $\{|\uparrow\rangle, |\downarrow\rangle\}$ . Recall that the matrix representation of operator  $\hat{A}$  in this basis is

$$\begin{pmatrix} \langle\uparrow|\hat{A}|\uparrow\rangle & \langle\uparrow|\hat{A}|\downarrow\rangle \\ \langle\downarrow|\hat{A}|\uparrow\rangle & \langle\downarrow|\hat{A}|\downarrow\rangle \end{pmatrix}, \quad (13.32)$$

and recall also that this isn't always the easiest way to find a matrix representation.

- Find matrix representations in the  $\{|\uparrow\rangle, |\downarrow\rangle\}$  basis of  $\hat{S}_z$ ,  $\hat{S}_+$ ,  $\hat{S}_-$ ,  $\hat{S}_x$ ,  $\hat{S}_y$ , and  $\hat{S}^2$ . Note the reappearance of the Pauli matrices!
- Find normalized column matrix representations for the eigenstates of  $\hat{S}_x$ :

$$\hat{S}_x|\rightarrow\rangle = +\frac{\hbar}{2}|\rightarrow\rangle \quad (13.33)$$

$$\hat{S}_x|\leftarrow\rangle = -\frac{\hbar}{2}|\leftarrow\rangle. \quad (13.34)$$

#### 13.3 Rotations and spin- $\frac{1}{2}$

Verify explicitly that

$$|\rightarrow\rangle = e^{-i(\hat{S}_y/\hbar)(+\pi/2)}|\uparrow\rangle, \quad (13.35)$$

$$|\leftarrow\rangle = e^{-i(\hat{S}_y/\hbar)(-\pi/2)}|\uparrow\rangle. \quad (13.36)$$

(Problems 2.27 through 2.29 are relevant here.)

#### 13.4 Spin-1 projection amplitudes

- (Easy.) Prove that

$$d_{m,m'}^{(j)}(\theta) = [d_{m',m}^{(j)}(-\theta)]^*. \quad (13.37)$$

b. Show that the  $d_{m,m'}^{(j)}(\theta)$  with  $j = 1$  are

$$\begin{aligned}d_{1,1}^{(1)}(\theta) &= +\frac{1}{2}(\cos \theta + 1) & d_{1,0}^{(1)}(\theta) &= -\frac{1}{\sqrt{2}} \sin \theta & d_{1,-1}^{(1)}(\theta) &= -\frac{1}{2}(\cos \theta - 1) \\d_{0,1}^{(1)}(\theta) &= +\frac{1}{\sqrt{2}} \sin \theta & d_{0,0}^{(1)}(\theta) &= \cos \theta & d_{0,-1}^{(1)}(\theta) &= -\frac{1}{\sqrt{2}} \sin \theta \\d_{-1,1}^{(1)}(\theta) &= -\frac{1}{2}(\cos \theta - 1) & d_{-1,0}^{(1)}(\theta) &= +\frac{1}{\sqrt{2}} \sin \theta & d_{-1,-1}^{(1)}(\theta) &= +\frac{1}{2}(\cos \theta + 1)\end{aligned}$$



## Chapter 14

# Central Force Motion

### 14.1 Energy eigenproblem in two dimensions

In one dimension, the energy eigenproblem is

$$-\frac{\hbar^2}{2M} \frac{d^2 \eta_n(x)}{dx^2} + V(x) \eta_n(x) = E_n \eta_n(x). \quad (14.1)$$

The generalization to two dimensions is straightforward:

$$-\frac{\hbar^2}{2M} \left[ \frac{\partial^2 \eta_n(x, y)}{\partial x^2} + \frac{\partial^2 \eta_n(x, y)}{\partial y^2} \right] + V(x, y) \eta_n(x, y) = E_n \eta_n(x, y). \quad (14.2)$$

The part in square brackets is called “the Laplacian of  $\eta_n(x, y)$ ” and represented by the symbol “ $\nabla^2$ ” as follows

$$\left[ \frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2} \right] \equiv \nabla^2 f(x, y). \quad (14.3)$$

Thus the “mathematical form” of the energy eigenproblem is

$$\nabla^2 \eta_n(\vec{r}) = -\frac{2M}{\hbar^2} [E_n - V(\vec{r})] \eta_n(\vec{r}). \quad (14.4)$$

Suppose  $V(x, y)$  is a function of distance from the origin  $r$  only. Then it makes sense to use polar coordinates  $r$  and  $\theta$  rather than Cartesian coordinates  $x$  and  $y$ . What is the expression for the Laplacian in polar coordinates? This can be uncovered through the chain rule, and it’s pretty hard to do. Fortunately, you can look up the answer:

$$\nabla^2 f(\vec{r}) = \left[ \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial f(r, \theta)}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 f(r, \theta)}{\partial \theta^2} \right]. \quad (14.5)$$

Thus, the partial differential equation to be solved is

$$\left[ \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial \eta_n(r, \theta)}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 \eta_n(r, \theta)}{\partial \theta^2} \right] = -\frac{2M}{\hbar^2} [E_n - V(r)] \eta_n(r, \theta) \quad (14.6)$$

or

$$\left\{ \frac{\partial^2}{\partial \theta^2} + r \frac{\partial}{\partial r} \left( r \frac{\partial}{\partial r} \right) + \frac{2M}{\hbar^2} r^2 [E_n - V(r)] \right\} \eta_n(r, \theta) = 0. \quad (14.7)$$

For convenience, we wrap up all the  $r$  dependence into one piece by defining the linear operator

$$\mathcal{L}_n(r) \equiv r \frac{\partial}{\partial r} \left( r \frac{\partial}{\partial r} \right) + \frac{2M}{\hbar^2} r^2 [E_n - V(r)] \quad (14.8)$$

and write the above as

$$\left\{ \frac{\partial^2}{\partial \theta^2} + \mathcal{L}_n(r) \right\} \eta_n(r, \theta) = 0. \quad (14.9)$$

There are at least two ways to approach the above equation: the Fourier series method and the separation of variables method. We'll try each one in turn.

#### Fourier series:

Because increasing the angle  $\theta$  by  $2\pi$  brings you to the same point where you started, the function  $\eta_n(r, \theta)$  is periodic in  $\theta$  with period  $2\pi$ . And the theory of Fourier series teaches that any such function can be written in the form

$$a_0 + \sum_{n=1}^{\infty} (a_n \cos n\theta + b_n \sin n\theta). \quad (14.10)$$

But, because

$$\cos n\theta = \frac{e^{in\theta} + e^{-in\theta}}{2} \quad \text{and} \quad \sin n\theta = \frac{e^{in\theta} - e^{-in\theta}}{2i},$$

this series can also be written in the form

$$\sum_{\ell=-\infty}^{\infty} c_{\ell} e^{i\ell\theta}. \quad (14.11)$$

[When dealing with real functions, the form (14.10) has obvious attractions. But the form (14.11) is always more compact, and when dealing with complex-valued functions it's more natural as well.] The Fourier theorem (14.11) is just another way of saying that the set of functions

$$e^{i\ell\theta} \quad \text{with} \quad \ell = 0, \pm 1, \pm 2, \pm 3, \dots \quad (14.12)$$

constitute a basis for the set of continuous functions with periodicity  $2\pi$ .

Applying this Fourier theorem to the problem at hand, we write the energy eigenfunction as

$$\eta_n(r, \theta) = \sum_{\ell=-\infty}^{\infty} R_{n,\ell}(r) e^{i\ell\theta} \quad (14.13)$$

and then take

$$\begin{aligned} 0 &= \left\{ \frac{\partial^2}{\partial \theta^2} + \mathcal{L}_n(r) \right\} \eta_n(r, \theta) \\ &= \left\{ \frac{\partial^2}{\partial \theta^2} + \mathcal{L}_n(r) \right\} \sum_{\ell=-\infty}^{\infty} R_{n,\ell}(r) e^{i\ell\theta} \\ &= \sum_{\ell=-\infty}^{\infty} \left\{ \frac{\partial^2}{\partial \theta^2} + \mathcal{L}_n(r) \right\} R_{n,\ell}(r) e^{i\ell\theta} \\ &= \sum_{\ell=-\infty}^{\infty} [-\ell^2 R_{n,\ell}(r) + \mathcal{L}_n(r) R_{n,\ell}(r)] e^{i\ell\theta}. \end{aligned} \quad (14.14)$$

Because the basis set  $\{e^{i\ell\theta}\}$  is complete (or, equivalently, because Fourier series are unique) the expression in square brackets above must vanish, for every value of  $\ell$ , whence

$$\mathcal{L}_n(r) R_{n,\ell}(r) - \ell^2 R_{n,\ell}(r) = 0 \quad \text{for } \ell = 0, \pm 1, \pm 2, \pm 3, \dots \quad (14.15)$$

We no longer have a partial differential equation. Instead we have an infinite number of ordinary differential equations

$$\left[ r \frac{d}{dr} \left( r \frac{d}{dr} \right) + \frac{2M}{\hbar^2} r^2 [E_n - V(r)] - \ell^2 \right] R_{n,\ell}(r) = 0. \quad (14.16)$$

### Separation of variables:

Our equation

$$\left\{ \frac{\partial^2}{\partial \theta^2} + \mathcal{L}_n(r) \right\} \eta_n(r, \theta) = 0. \quad (14.17)$$

is a linear partial differential equation, so we cast around for solutions knowing that a linear combination of solutions will also be a solution, and hoping that we will cast our net wide enough to catch all the elements of a basis. We cast around using the technique of “separation of variables”, namely by looking for solutions of the form

$$\eta_n(r, \theta) = R(r)\Theta(\theta). \quad (14.18)$$

Plugging this form into the PDE gives

$$\begin{aligned} R(r)\Theta''(\theta) + \Theta(\theta)\mathcal{L}_n(r)R(r) &= 0 \\ \frac{\Theta''(\theta)}{\Theta(\theta)} + \frac{\mathcal{L}_n(r)R(r)}{R(r)} &= 0 \end{aligned} \quad (14.19)$$

Through the usual separation-of-variables argument, we recognize that if a function of  $r$  alone plus a function of  $\theta$  alone sum to zero, where  $r$  and  $\theta$  are independent variables, then both functions must be equal to a constant:

$$\frac{\mathcal{L}_n(r)R(r)}{R(r)} = -\frac{\Theta''(\theta)}{\Theta(\theta)} = \text{const.} \quad (14.20)$$

First, look at the angular part:

$$\Theta''(\theta) = -\text{const} \Theta(\theta). \quad (14.21)$$

This is the differential equation for a mass on a spring! The two linearly independent solutions are

$$\Theta(\theta) = \sin(\sqrt{\text{const}} \theta) \quad \text{or} \quad \Theta(\theta) = \cos(\sqrt{\text{const}} \theta). \quad (14.22)$$

Now, the boundary condition for this ODE is just that the function must come back to itself if  $\theta$  increases by  $2\pi$ :

$$\Theta(\theta) = \Theta(2\pi + \theta). \quad (14.23)$$

If you think about this for a minute, you'll see that this means  $\sqrt{\text{const}}$  must be an integer. The negative integers don't give us anything new, so we'll take

$$\sqrt{\text{const}} = \ell \quad \text{where} \quad \ell = 0, 1, 2, \dots \quad (14.24)$$

In summary, the solution to the angular problem is

$$\begin{array}{cccccc} \ell = 0 & \ell = 1 & \ell = 2 & \ell = 3 & \dots & \\ \Theta(\theta) & 1 & \sin \theta \text{ or } \cos \theta & \sin 2\theta \text{ or } \cos 2\theta & \sin 3\theta \text{ or } \cos 3\theta & \dots \end{array}$$

Alternatively, we can take linear combinations of the above to produce the set of solutions

$$\Theta(\theta) = e^{i\ell\theta} \quad \text{for} \quad \ell = 0, \pm 1, \pm 2, \pm 3, \dots \quad (14.25)$$

Now examine the radial part of the problem:

$$\frac{\mathcal{L}_n(r)R(r)}{R(r)} = \text{const} = \ell^2. \quad (14.26)$$



Because the radial operator depends on  $n$ , the solutions  $R(r)$  will depend on both  $n$  and  $\ell$ , so we denote them by  $R_{n,\ell}(r)$ . They solve the equation

$$\mathcal{L}_n(r)R_{n,\ell}(r) - \ell^2 R_{n,\ell}(r) = 0 \quad \text{for } \ell = 0, \pm 1, \pm 2, \pm 3, \dots \quad (14.27)$$

We no longer have a partial differential equation. Instead we have an infinite number of ordinary differential equations

$$\left[ r \frac{d}{dr} \left( r \frac{d}{dr} \right) + \frac{2M}{\hbar^2} r^2 [E_n - V(r)] - \ell^2 \right] R_{n,\ell}(r) = 0. \quad (14.28)$$

### Two routes converge:

We are at the end of the bifurcation. Both the “Fourier series” route and the “Separation of variables” route have arrived at the same destination. This is not a coincidence. The mathematical theory of Sturm-Liouville problems<sup>1</sup> assures us that for a wide class of partial differential equations, the result of separation of variables will give rise to a complete set of functions, like the trigonometric functions that form a basis for the continuous functions with period  $2\pi$ .

Write the resulting one-variable ODE as

$$\begin{aligned} & \left\{ \frac{1}{r} \frac{d}{dr} \left( r \frac{d}{dr} \right) + \frac{2M}{\hbar^2} [E_n - V(r)] - \frac{\ell^2}{r^2} \right\} R_{n,\ell}(r) = 0 \\ & \left\{ \frac{1}{r} \frac{d}{dr} \left( r \frac{d}{dr} \right) + \frac{2M}{\hbar^2} \left[ E_n - V(r) - \frac{\hbar^2 \ell^2}{2M r^2} \right] \right\} R_{n,\ell}(r) = 0 \end{aligned} \quad (14.29)$$

I want to compare this differential equation with another one-variable differential equation, namely the one for the energy eigenvalue problem in one dimension:

$$\left\{ \frac{d^2}{dx^2} + \frac{2M}{\hbar^2} [E - V(x)] \right\} \eta(x) = 0. \quad (14.30)$$

The parts to the right are rather similar, but the parts to the left — the derivatives — are rather different.

In addition, the one-dimensional energy eigenfunction satisfies the normalization

$$\int_{-\infty}^{\infty} |\eta(x)|^2 dx = 1, \quad (14.31)$$

---

<sup>1</sup>Charles-François Sturm (1803–1855) was a French mathematician who also helped make the first experimental determination of the speed of sound in water. Joseph Liouville (1809–1882), another French mathematician, made contributions in complex analysis, number theory, differential geometry, and classical mechanics. He was elected to the French Constituent Assembly of 1848 which established the Second Republic.

whereas the two-dimensional energy eigenfunction satisfies the normalization

$$\begin{aligned} \int |\eta(x, y)|^2 dx dy &= 1 \\ \int_0^\infty dr \int_0^{2\pi} r d\theta |R_{n,\ell}(r)e^{i\ell\theta}|^2 &= 1 \\ 2\pi \int_0^\infty dr r |R_{n,\ell}(r)|^2 &= 1. \end{aligned} \quad (14.32)$$

This suggests that the true analog of the one-dimensional  $\eta(x)$  is not  $R_{n,\ell}(r)$ , but rather

$$u_{n,\ell}(r) = \sqrt{r}R_{n,\ell}(r). \quad (14.33)$$

Furthermore,

$$\text{if } u(r) = \sqrt{r}R(r), \quad \text{then } \frac{1}{r} \frac{d}{dr} (rR'(r)) = \frac{1}{\sqrt{r}} \left( u''(r) + \frac{1}{4} \frac{u(r)}{r^2} \right). \quad (14.34)$$

Using this change of function, the radial equation (14.29) becomes

$$\begin{aligned} \left\{ \frac{d^2}{dr^2} + \frac{1}{4} \frac{1}{r^2} + \frac{2M}{\hbar^2} \left[ E - V(r) - \frac{\hbar^2}{2M} \frac{\ell^2}{r^2} \right] \right\} u(r) &= 0 \\ \left\{ \frac{d^2}{dr^2} + \frac{2M}{\hbar^2} \left[ E - V(r) - \frac{\hbar^2}{2M} \left( \ell^2 - \frac{1}{4} \right) \frac{1}{r^2} \right] \right\} u(r) &= 0 \\ \left\{ \frac{d^2}{dr^2} + \frac{2M}{\hbar^2} \left[ E - V(r) - \frac{\hbar^2 (\ell - \frac{1}{2})(\ell + \frac{1}{2})}{2M} \frac{1}{r^2} \right] \right\} u(r) &= 0 \end{aligned} \quad (14.35)$$

In this form, the radial equation is exactly like a one-dimensional energy eigenproblem, except that where the one-dimensional problem has the function  $V(x)$ , the radial problem has the function

$$V(r) + \hbar^2 (\ell - \frac{1}{2})(\ell + \frac{1}{2}) / 2Mr^2.$$

These two functions play parallel mathematical roles in the two problems. To emphasize these similar roles, we define an “effective potential energy” for the radial problem, namely

$$V_{\text{eff}}(r) = V(r) + \frac{\hbar^2 (\ell - \frac{1}{2})(\ell + \frac{1}{2})}{2M} \frac{1}{r^2}. \quad (14.36)$$

Don’t read too much into the term “effective potential energy.” No actual potential energy will depend upon  $\hbar$ , or upon the separation constant  $\ell$ ! I’m not saying that  $V_{\text{eff}}(r)$  is a potential energy function, merely that it plays the mathematical role of one in solving this eigenproblem.

Now that the radial equation (14.35) is in exact correspondence with the one-dimensional equation (14.30), we can solve this eigenproblem using the same “curve toward or away from axis” techniques that we developed for the one-dimensional problem in chapter 10. (Or any other technique which works for the one-dimensional problem.) The resulting eigenfunctions and eigenvalues will, of course, depend upon the value of the separation constant  $\ell$ , because the effective potential depends upon the value of  $\ell$ . And as always, for each  $\ell$  there will be many eigenvalues and eigenfunctions, which we will label by index  $n = 1, 2, 3, \dots$  calling them  $u_{n,\ell}(r)$  with eigenvalue  $E_{n,\ell}$ .

Finally, note that the effective potential energy for  $\ell = +5$  is the same as the effective potential energy for  $\ell = -5$ . Thus the eigenfunctions  $u_{n,+5}(r)$  and eigenvalues  $E_{n,+5}$  will be identical to the eigenfunctions  $u_{n,-5}(r)$  and eigenvalues  $E_{n,-5}$ .

This is a really charming result. We haven't yet specified the potential energy function  $V(r)$ , so we can't yet determine, say,  $E_{7,+5}$  or  $E_{7,-5}$ . Yet we know that these two energy eigenvalues will be equal! Whenever there are two different eigenfunctions, in this case

$$\frac{u_{n,+5}(r)}{\sqrt{r}} e^{+i5\theta} \quad \text{and} \quad \frac{u_{n,+5}(r)}{\sqrt{r}} e^{-i5\theta},$$

attached to the same eigenvalue, the eigenfunctions are said to be *degenerate*. I don't know how such a disparaging term came to be attached to such a charming result, but it has been. [[Consider better placement of this remark.]]

### Summary:

To solve the two-dimensional energy eigenproblem for a radially-symmetric potential energy  $V(r)$ , namely

$$-\frac{\hbar^2}{2M} \nabla^2 \eta(\vec{r}) + V(r) \eta(\vec{r}) = E \eta(\vec{r}), \quad (14.37)$$

first solve the radial energy eigenproblem

$$-\frac{\hbar^2}{2M} \frac{d^2 u(r)}{dr^2} + \left[ V(r) + \frac{\hbar^2 (\ell - \frac{1}{2})(\ell + \frac{1}{2})}{2M r^2} \right] u(r) = E u(r) \quad (14.38)$$

for  $\ell = 0, 1, 2, \dots$ . For a given  $\ell$ , call the resulting energy eigenfunctions and eigenvalues  $u_{n,\ell}(r)$  and  $E_{n,\ell}$  for  $n = 1, 2, 3, \dots$ . Then the two-dimensional solutions are

$$\text{For } \ell = 0: \quad \eta(r, \theta) = \frac{u_{n,0}(r)}{\sqrt{r}} \quad \text{with energy } E_{n,0} \quad (14.39)$$

and

$$\text{For } \ell = 1, 2, 3, \dots: \quad \eta(r, \theta) = \frac{u_{n,\ell}(r)}{\sqrt{r}} e^{+i\ell\theta} \text{ and } \eta(r, \theta) = \frac{u_{n,\ell}(r)}{\sqrt{r}} e^{-i\ell\theta} \quad \text{with energy } E_{n,\ell}. \quad (14.40)$$

Alternatively, for the last equation we can use

$$\text{For } \ell = 1, 2, 3, \dots: \quad \eta(r, \theta) = \frac{u_{n,\ell}(r)}{\sqrt{r}} \sin(\ell\theta) \text{ and } \eta(r, \theta) = \frac{u_{n,\ell}(r)}{\sqrt{r}} \cos(\ell\theta) \quad \text{with energy } E_{n,\ell}. \quad (14.41)$$

**Reflection:**

So we've reduced the two-dimensional problem to a one-dimensional problem. How did this miracle occur? Two things happened:

- The original eigenvalue problem was of the form

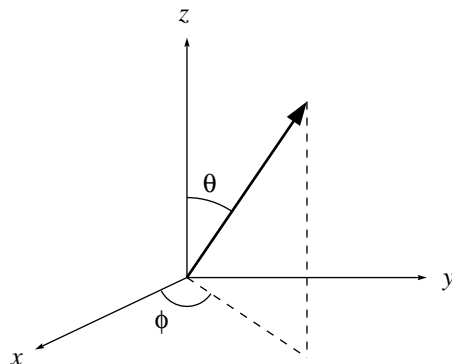
$$\{\text{angular operator} + \text{radial operator}\} \eta_n(r, \theta) = 0. \quad (14.42)$$

- There was an angular operator eigenbasis  $\{\Phi_\ell(\theta)\}$  such that

$$\{\text{angular operator}\} \Phi_\ell(\theta) = \text{number } \Phi_\ell(\theta). \quad (14.43)$$

## 14.2 Energy eigenproblem in three dimensions

Can we get the same miracle to occur in three dimensions?



The energy eigenproblem is

$$-\frac{\hbar^2}{2M}\nabla^2\eta_n(\vec{x}) + V(r)\eta_n(\vec{x}) = E_n\eta_n(\vec{x}), \quad (14.44)$$

and the Laplacian in spherical coordinates is

$$\begin{aligned} \nabla^2 &= \frac{1}{r^2} \left[ \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \left( \sin\theta \frac{\partial}{\partial\theta} \right) + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial\phi^2} \right] \\ &= \frac{1}{r^2} \left[ \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{\partial}{\partial\mu} \left( (1-\mu^2) \frac{\partial}{\partial\mu} \right) + \frac{1}{1-\mu^2} \frac{\partial^2}{\partial\phi^2} \right]. \end{aligned} \quad (14.45)$$

Here  $\phi$  ranges from 0 to  $2\pi$ , and  $\theta$  ranges from 0 to  $\pi$ . It is often convenient to use, in place of  $\theta$ , the variable

$$\mu = \cos\theta \quad \text{where } \mu \text{ ranges from } -1 \text{ to } +1. \quad (14.46)$$

(The situation  $\mu = +1$  corresponds to the “north pole” of the spherical coordinate system – the positive  $z$  axis – while the situation  $\mu = -1$  corresponds to the “south pole” – the negative  $z$  axis.) The energy eigenproblem is then

$$\left\{ \nabla^2 + \frac{2M}{\hbar^2} [E_n - V(r)] \right\} \eta_n(\vec{x}) = 0, \quad (14.47)$$

or

$$\left\{ \frac{\partial}{\partial\mu} \left( (1-\mu^2) \frac{\partial}{\partial\mu} \right) + \frac{1}{1-\mu^2} \frac{\partial^2}{\partial\phi^2} + \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{2M}{\hbar^2} r^2 [E_n - V(r)] \right\} \eta_n(\vec{x}) = 0, \quad (14.48)$$

and this expression is indeed in the desired form (14.42).

Now, is there a basis of angular operator eigenfunctions as required by (14.43)?

We seek a complete set of functions on the unit sphere  $\{y_\lambda(\mu, \phi)\}$  such that

$$\{\text{angular operator}\}y_\lambda(\mu, \phi) = \left\{ \frac{\partial}{\partial\mu} \left( (1-\mu^2) \frac{\partial}{\partial\mu} \right) + \frac{1}{1-\mu^2} \frac{\partial^2}{\partial\phi^2} \right\} y_\lambda(\mu, \phi) = \lambda y_\lambda(\mu, \phi). \quad (14.49)$$

These functions, like any function of angle  $\phi$ , must be periodic in  $\phi$  with period  $2\pi$ , so

$$y_\lambda(\mu, \phi) = \sum_{m=-\infty}^{\infty} p_{\lambda,m}(\mu) e^{im\phi}. \quad (14.50)$$

In these terms, the eigenproblem (14.49) becomes

$$\sum_{m=-\infty}^{\infty} \left\{ \left[ \frac{\partial}{\partial\mu} \left( (1-\mu^2) \frac{\partial}{\partial\mu} \right) - \frac{m^2}{1-\mu^2} - \lambda \right] p_{\lambda,m}(\mu) \right\} e^{im\phi} = 0. \quad (14.51)$$

Because of the uniqueness of Fourier series (compare the reasoning below equation (14.14)) each term in curly brackets must vanish individually. For each  $m$

$$\left[ \frac{d}{d\mu} \left( (1 - \mu^2) \frac{d}{d\mu} \right) - \frac{m^2}{1 - \mu^2} - \lambda \right] p_{\lambda, m}(\mu) = 0 \quad m = 0, \pm 1, \pm 2, \dots \quad (14.52)$$

This differential equation is called the “generalized Legendre<sup>2</sup> equation.” It can be solved using the power-series method — details are left to problem X.Y. There are, of course, two linearly independent solutions for all values of  $m$  and  $\lambda$ . However, most of those solutions diverge at either  $\mu = 1$ , or  $\mu = -1$ , or both. (That is, either at the north or south pole, or both.) Solutions are finite for all  $\mu$  from  $-1$  to  $+1$ , inclusive, if and only if

$$\lambda = -\ell(\ell + 1) \quad \ell = 0, 1, 2, 3, \dots \quad (14.53)$$

and

$$m = -\ell, -\ell + 1, \dots, 0, \dots, \ell - 1, \ell. \quad (14.54)$$

In these cases, the finite solution is called the “associated Legendre function”  $P_\ell^m(\mu)$ .

These functions have a lot of interesting properties, but the main point is that we can use them to define a set of functions on the unit sphere called the *spherical harmonics*

$$Y_\ell^m(\theta, \phi) \quad \text{or} \quad Y_\ell^m(\mu, \phi) = \sqrt{\frac{2\ell + 1}{4\pi} \frac{(\ell - m)!}{(\ell + m)!}} P_\ell^m(\mu) e^{im\phi}. \quad (14.55)$$

(Some authors use different prefactors.)

The spherical harmonics satisfy

$$\{\text{angular operator}\} Y_\ell^m(\mu, \phi) = \left\{ \frac{\partial}{\partial \mu} \left( (1 - \mu^2) \frac{\partial}{\partial \mu} \right) + \frac{1}{1 - \mu^2} \frac{\partial^2}{\partial \phi^2} \right\} Y_\ell^m(\mu, \phi) = -\ell(\ell + 1) Y_\ell^m(\mu, \phi) \quad (14.56)$$

and are complete in the sense that

*Theorem:* If  $f(\mu, \phi)$  is a differentiable function on the unit sphere then

$$f(\mu, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} f_{\ell, m} Y_\ell^m(\mu, \phi) \quad \text{where} \quad f_{\ell, m} = \int_0^{2\pi} d\phi \int_{-1}^1 d\mu (Y_\ell^m(\mu, \phi))^* f(\mu, \phi). \quad (14.57)$$

<sup>2</sup>Adrien-Marie Legendre (1752–1833) made contributions throughout mathematics. He originated the “least squares” method of curve fitting. One notable episode from his life is that the French government denied him the pension he had earned when he refused to endorse a government-supported candidate for an honor.

The above paragraph is precisely analogous to the Fourier series result that the “trigonometric” functions  $e^{i\ell\theta}$  satisfy

$$\left\{ \frac{\partial^2}{\partial\theta^2} \right\} e^{i\ell\theta} = -\ell^2 e^{i\ell\theta} \quad (14.58)$$

and are complete in the sense that

*Theorem:* If  $f(\theta)$  is a differentiable function on the unit circle (i.e. with periodicity  $2\pi$ ) then

$$f(\theta) = \sum_{\ell=-\infty}^{\infty} f_{\ell} e^{i\ell\theta} \quad \text{where} \quad f_{\ell} = \frac{1}{2\pi} \int_0^{2\pi} d\theta (e^{i\ell\theta})^* f(\theta). \quad (14.59)$$

There are a lot of special functions, many of which are used only in very specialized situations. But the spherical harmonics are just as important in three dimensional problems as the trigonometric functions are in two dimensional problems. Spherical harmonics are used in quantum mechanics, in electrostatics, in acoustics, in signal processing, in seismology, and in mapping (to keep track of the deviations of the Earth’s shape from spherical). They are as important as sines and cosines. It’s worth becoming familiar with them.

Now that we have a complete set of eigenfunctions for the angular operator, we can carry out the “reduction to one dimension” process in three dimensions just as we did in two dimensions. The energy eigenvalue problem is

$$\left\{ -\frac{\hbar^2}{2M} \nabla^2 + V(r) - E_n \right\} \eta_n(\vec{r}) = 0. \quad (14.60)$$

Write

$$\eta_n(r, \mu, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} R_{n,\ell,m}(r) Y_{\ell}^m(\mu, \phi) \quad (14.61)$$

to prove that  $R_{n,\ell,m}(r)$  satisfies

$$\left\{ \frac{1}{r^2} \frac{d}{dr} \left( r^2 \frac{d}{dr} \right) + \frac{2M}{\hbar^2} \left[ E_n - V(r) - \frac{\hbar^2}{2M} \frac{\ell(\ell+1)}{r^2} \right] \right\} R_{n,\ell,m}(r) = 0. \quad (14.62)$$

Note that the differential equation is independent of  $m$ , so the solution must also be independent of  $m$ . Thus we drop the subscript  $m$  and write  $R_{n,\ell}(r)$ .

The energy eigenfunction satisfies the normalization

$$\begin{aligned} \int |\eta(x, y, z)|^2 dx dy dz &= 1 \\ \int_0^\infty dr \int_{-1}^1 d\mu \int_0^{2\pi} r^2 d\phi |R_{n,\ell}(r) Y_\ell^m(\mu, \phi)|^2 &= 1 \\ \int_0^\infty dr r^2 |R_{n,\ell}(r)|^2 &= 1. \end{aligned} \quad (14.63)$$

This suggests that the true analog to a one-dimensional wavefunction is  $u_{n,\ell}(r) = rR_{n,\ell}(r)$ , and sure enough  $u_{n,\ell}(r)$  satisfies the equation

$$\left\{ -\frac{\hbar^2}{2M} \frac{d^2}{dr^2} + \left[ V(r) + \frac{\hbar^2 \ell(\ell+1)}{2Mr^2} \right] \right\} u_{n,\ell}(r) = E_{n,\ell} u_{n,\ell}(r). \quad (14.64)$$

### Summary:

To solve the three-dimensional energy eigenproblem for a spherically-symmetric potential energy  $V(r)$ , namely

$$-\frac{\hbar^2}{2M} \nabla^2 \eta(\vec{r}) + V(r) \eta(\vec{r}) = E \eta(\vec{r}), \quad (14.65)$$

first solve the radial energy eigenproblem

$$-\frac{\hbar^2}{2M} \frac{d^2 u(r)}{dr^2} + \left[ V(r) + \frac{\hbar^2 \ell(\ell+1)}{2M} \frac{1}{r^2} \right] u(r) = E u(r) \quad (14.66)$$

for  $\ell = 0, 1, 2, \dots$ . For a given  $\ell$ , call the resulting energy eigenfunctions and eigenvalues  $u_{n,\ell}(r)$  and  $E_{n,\ell}$  for  $n = 1, 2, 3, \dots$ . Then the three-dimensional solutions are

$$\eta_{n,\ell,m}(r, \theta, \phi) = \frac{u_{n,\ell}(r)}{r} Y_\ell^m(\theta, \phi) \quad \text{with energy } E_{n,\ell}, \quad (14.67)$$

where  $m$  takes on the  $2\ell + 1$  values  $-\ell, -\ell + 1, \dots, 0, \dots, \ell - 1, \ell$ . Notice that the  $2\ell + 1$  different solutions for a given  $n$  and  $\ell$ , but with different  $m$ , are degenerate.

*Problem:* Show that the probability density  $|Y_\ell^m(\theta, \phi)|^2$  associated with any spherical harmonic is “axially symmetric,” that is, independent of rotations about the  $z$  axis, that is, independent of  $\phi$ .



### 14.3 Bound state energy eigenproblem for Coulombic potentials

*Problem:* Given a (reduced) mass  $M$  and a Coulombic potential energy  $V(r) = -k/r$ , find the negative values  $E_{n,\ell}$  such that the corresponding solutions  $U_{n,\ell}(r)$  of

$$\left\{ -\frac{\hbar^2}{2M} \frac{d^2}{dr^2} + \left[ -\frac{k}{r} + \frac{\hbar^2 \ell(\ell+1)}{2Mr^2} \right] \right\} U_{n,\ell}(r) = E_{n,\ell} U_{n,\ell}(r) \quad (14.68)$$

are normalizable wavefunctions

$$\int_0^\infty |U_{n,\ell}(r)|^2 dr = 1. \quad (14.69)$$

*Strategy:* Same as for the simple harmonic oscillator eigenproblem:

- (1) Convert to dimensionless variable.
- (2) Remove asymptotic behavior of solutions.
- (3) Find non-asymptotic behavior using the series method.
- (4) Invoke normalization to terminate the series as a polynomial.

1. *Convert to dimensionless variable:* Only one length can be constructed from  $M$ ,  $k$ , and  $\hbar$ . It is

$$a = \frac{\hbar^2}{kM}. \quad (14.70)$$

For the hydrogen problem

$$M = \frac{m_p m_e}{m_p + m_e} \approx m_e \quad \text{and} \quad k = \frac{e^2}{4\pi\epsilon_0},$$

so this length is approximately

$$\frac{\hbar^2}{m_e} \frac{4\pi\epsilon_0}{e^2} \equiv a_0 \equiv \text{“the Bohr radius”} = 0.0529 \text{ nm}. \quad (14.71)$$

Convert to the dimensionless variable

$$\rho = \frac{r}{a} \quad (14.72)$$

and the dimensionless wavefunction

$$u_{n,\ell}(\rho) = \sqrt{a} U_{n,\ell}(a\rho). \quad (14.73)$$

The resulting eigenproblem is

$$\left\{ -\frac{d^2}{d\rho^2} - \frac{2}{\rho} + \frac{\ell(\ell+1)}{\rho^2} \right\} u_{n,\ell}(\rho) = \frac{E_{n,\ell}}{k^2 M / 2\hbar^2} u_{n,\ell}(\rho) \quad (14.74)$$

with

$$\int_0^\infty |u_{n,\ell}(\rho)|^2 d\rho = 1. \quad (14.75)$$

It's clear that the energy

$$\frac{k^2 M}{2\hbar^2} \quad (14.76)$$

is the characteristic energy for this problem. For hydrogen, its value is approximately

$$\left(\frac{e^2}{4\pi\epsilon_0}\right)^2 \frac{m_e}{2\hbar^2} \equiv \text{Ry} \equiv \text{“the Rydberg”} = 13.6 \text{ eV.}$$

Thus it is reasonable, for brevity, to define the dimensionless energy parameter

$$\mathcal{E}_{n,\ell} = \frac{E_{n,\ell}}{k^2 M/2\hbar^2}. \quad (14.77)$$

Furthermore, for the bound state problem  $\mathcal{E}_{n,\ell}$  is negative so we define

$$b_{n,\ell}^2 = -\mathcal{E}_{n,\ell} \quad (14.78)$$

and the eigenproblem becomes

$$\left\{ \frac{d^2}{d\rho^2} + \frac{2}{\rho} - \frac{\ell(\ell+1)}{\rho^2} + b_{n,\ell}^2 \right\} u_{n,\ell}(\rho) = 0 \quad (14.79)$$

with

$$\int_0^\infty |u_{n,\ell}(\rho)|^2 d\rho = 1. \quad (14.80)$$

## 2. Remove asymptotic behavior of solutions:

*Note:* In this section we will show that as  $\rho \rightarrow 0$ ,

$$u_{n,\ell}(\rho) \approx \rho^{\ell+1}, \quad (14.81)$$

and that as  $\rho \rightarrow \infty$ ,

$$u_{n,\ell}(\rho) \approx e^{-b_{n,\ell}\rho}, \quad (14.82)$$

so we will set

$$u_{n,\ell}(\rho) = \rho^{\ell+1} e^{-b_{n,\ell}\rho} v_{n,\ell}(\rho) \quad (14.83)$$

and then solve an ODE for  $v_{n,\ell}(\rho)$ . As far as rigor is concerned we could have just pulled the change-of-function (14.83) out of a hat. Thus this section is motivational and doesn't need to be rigorous.

Because equation (14.79) has problems (or, formally, a “regular singular point”) at  $\rho = 0$ , it pays to find the asymptotic behavior when  $\rho \rightarrow 0$  as well as when  $\rho \rightarrow \infty$ .

2A. *Find asymptotic behavior as  $\rho \rightarrow 0$ :* The ODE is

$$\left\{ \frac{d^2}{d\rho^2} + \left[ \frac{2}{\rho} - \frac{\ell(\ell+1)}{\rho^2} - b_{n,\ell}^2 \right] \right\} u_{n,\ell}(\rho) = 0. \quad (14.84)$$

As  $\rho \rightarrow 0$  the term in square brackets is dominated (unless  $\ell = 0$ ) by  $-\ell(\ell+1)/\rho^2$ . The equation

$$\left\{ \frac{d^2}{d\rho^2} - \frac{\ell(\ell+1)}{\rho^2} \right\} u(\rho) = 0 \quad (14.85)$$

is solved by

$$u(\rho) = A\rho^{\ell+1} + B\rho^{-\ell}. \quad (14.86)$$

However, it's not healthy to keep factors like  $\rho^{-\ell}$  around, because

$$\int_0^{\rho_0} \rho^{-2\ell} d\rho = \frac{1}{-2\ell+1} \left[ \frac{1}{\rho^{2\ell-1}} \right]_0^{\rho_0} = \infty \quad [\text{for } \ell > \frac{1}{2}], \quad (14.87)$$

so wavefunctions with  $\rho^{-\ell}$  prefactors tend to be unnormalizable. (Here  $\rho_0$  is just any positive number.) Thus the wavefunction must behave as

$$u(\rho) \approx A\rho^{\ell+1} \quad (14.88)$$

as  $\rho \rightarrow 0$ .

Our arguments have relied upon  $\ell \neq 0$ , but it turns out that by stupid good luck the result (14.88) applies when  $\ell = 0$  as well. However, it's rather hard to prove this, and since this section is really just motivation anyway, I'll not pursue the matter.

2B. *Find asymptotic behavior as  $\rho \rightarrow \infty$ :* In this case, the square bracket term in equation (14.84) is dominated by  $-b_{n,\ell}^2$ , so the approximate ODE is

$$\left\{ \frac{d^2}{d\rho^2} - b_{n,\ell}^2 \right\} u_{n,\ell}(\rho) = 0 \quad (14.89)$$

with solutions

$$u_{n,\ell}(\rho) = Ae^{-b_{n,\ell}\rho} + be^{+b_{n,\ell}\rho}. \quad (14.90)$$

Clearly, normalization requires that  $B = 0$ , so the wavefunction has the expected exponential cutoff for large  $\rho$ .

In this way, we have justified the definition of  $v_{n,\ell}(\rho)$  in equation (14.83). Plugging (14.83) into ODE (14.79), we find that  $v_{n,\ell}(\rho)$  satisfies the ODE

$$\left\{ \rho \frac{d^2}{d\rho^2} + 2[\ell + 1 - b_{n,\ell}\rho] \frac{d}{d\rho} - 2[b_{n,\ell}\ell + b_{n,\ell} - 1] \right\} v_{n,\ell}(\rho) = 0 \quad (14.91)$$

3. *Find non-asymptotic behavior using the series method:* We try out the solution

$$v_{n,\ell}(\rho) = \sum_{k=0}^{\infty} a_k \rho^k \quad (14.92)$$

and readily find that

$$a_{k+1} = \frac{2b_{n,\ell}(k + \ell + 1) - 2}{(k + 1)(k + 2\ell + 2)} a_k \quad k = 0, 1, 2, \dots \quad (14.93)$$

(Note that because  $k$  and  $\ell$  are both non-negative, the denominator never vanishes.)

4. *Invoke normalization to terminate the series as a polynomial:* If the  $a_k$  coefficient never vanishes, then

$$\frac{a_{k+1}}{a_k} \rightarrow \frac{2b_{n,\ell}}{k} \quad \text{as } k \rightarrow \infty. \quad (14.94)$$

As in the SHO, this leads to  $v(\rho) \approx e^{2b_{n,\ell}\rho}$  as  $\rho \rightarrow \infty$ , which is pure disaster. To avoid catastrophe, we must truncate the series as a  $k$ th order polynomial by demanding

$$b_{n,\ell} = \frac{1}{k + \ell + 1} \quad k = 0, 1, 2, \dots \quad (14.95)$$

Thus  $b_{n,\ell}$  is always the reciprocal of the integer

$$n = k + \ell + 1 \quad (14.96)$$

and

$$\mathcal{E}_{n,\ell} = -b_{n,\ell}^2 = -\frac{1}{n^2} \quad n = 1, 2, 3, \dots \quad (14.97)$$

We have found the permissible bound state energies!

What are the eigenfunctions? The solution  $v_{n,\ell}(\rho)$  that is a polynomial of order  $k = n - \ell - 1$  has a name: it is the Laguerre<sup>3</sup> polynomial

$$L_{n-\ell-1}^{2\ell+1}((2/n)\rho). \quad (14.98)$$

It would be nicer to have a more direct notation like our own  $v_{n,\ell}(\rho)$ , but Laguerre died before quantum mechanics was born, so he could not have known how to make his notation convenient for the quantum mechanical Coulomb problem. The Laguerre polynomials are just one more class of special functions not worth knowing much about.

All together, the energy eigenfunctions are

$$\frac{\eta_{n,\ell,m}(\rho, \theta, \phi) = [\text{constant}] \rho^\ell e^{-\rho/n} L_{n-\ell-1}^{2\ell+1}((2/n)\rho) Y_\ell^m(\theta, \phi). \quad (14.99)$$

<sup>3</sup>Edmond Laguerre (1834–1886), French artillery officer and mathematician, made contributions to analysis and especially geometry.

*Degeneracy*

Recall that each  $v_{n,\ell}(\rho)$  already has an associated  $2\ell + 1$ -fold degeneracy. In addition, each  $\ell$  gives rise to an infinite number of eigenvalues:

$$\mathcal{E}_{n,\ell} = -\frac{1}{(k + \ell + 1)^2} \quad k = 0, 1, 2, \dots \quad (14.100)$$

In tabular form

$$\begin{array}{l} \ell = 0 \text{ gives } n = 1, 2, 3, 4, \dots \\ \ell = 1 \text{ gives } n = 2, 3, 4, \dots \\ \ell = 2 \text{ gives } n = 3, 4, \dots \\ \vdots \end{array}$$

So...

$$\begin{array}{l} \ell = 0 \text{ (degeneracy 1) gives } \mathcal{E}_{n,\ell} = -1, -\frac{1}{2^2}, -\frac{1}{3^2}, -\frac{1}{4^2}, \dots \\ \ell = 1 \text{ (degeneracy 3) gives } \mathcal{E}_{n,\ell} = -\frac{1}{2^2}, -\frac{1}{3^2}, -\frac{1}{4^2}, \dots \\ \ell = 2 \text{ (degeneracy 5) gives } \mathcal{E}_{n,\ell} = -\frac{1}{3^2}, -\frac{1}{4^2}, \dots \\ \vdots \end{array}$$

Eigenenergies of  $-1/n^2$  are associated with  $n$  different values of  $\ell$ , namely  $\ell = 0, 1, \dots, n - 1$ . The total degeneracy is thus

$$\sum_{\ell=0}^{n-1} (2\ell + 1) = n^2. \quad (14.101)$$

**14.4 Summary of the bound state energy eigenproblem for a Coulombic potential**

A complete set of energy eigenfunctions is  $\eta_{n,\ell,m}(r, \theta, \phi)$

$$\begin{array}{l} \text{where} \quad \quad \quad n = 1, 2, 3, \dots \\ \text{and for each } n \quad \quad \ell = 0, 1, 2, \dots, n - 1 \\ \text{and for each } n \text{ and } \ell \quad m = -\ell, -\ell + 1, \dots, \ell - 1, \ell. \end{array}$$

This wavefunction represents a state of energy

$$E_n = -\frac{k^2 M/2\hbar^2}{n^2},$$

independent of  $\ell$  and  $m$ . Thus energy  $E_n$  has an  $n^2$ -fold degeneracy. In particular, for hydrogen this eigenenergy is nearly

$$E_n = -\frac{\text{Ry}}{n^2}, \quad \text{Ry} = 13.6 \text{ eV}.$$

In addition, the wavefunction  $\eta_{n,\ell,m}(r, \theta, \phi)$  represents a state with an angular momentum squared of  $\hbar^2\ell(\ell + 1)$  and an angular momentum  $z$  component of  $\hbar m$ .

[I recommend that you memorize this summary...it's the sort of thing that frequently comes up on GREs and physics oral exams.]

## 14.5 Problems

### 14.1 Positronium

The “atom” positronium is a bound state of an electron and a positron. Find the allowed energies for positronium.

### 14.2 Operator factorization solution of the Coulomb problem

The bound state energy eigenvalues of the hydrogen atom can be found using the operator factorization method. In reduced units, the radial wave equation is

$$\left[ -\frac{d^2}{d\rho^2} + \frac{\ell(\ell + 1)}{\rho^2} - \frac{2}{\rho} \right] u_{n,\ell}(\rho) \equiv h_\ell u_{n,\ell}(\rho) = \mathcal{E}_{n,\ell} u_{n,\ell}(\rho). \quad (14.102)$$

Introduce the operators

$$D_\pm^{(\ell)} \equiv \frac{d}{d\rho} \mp \frac{\ell}{\rho} \pm \frac{1}{\ell} \quad (14.103)$$

and show that

$$D_-^{(\ell+1)} D_+^{(\ell+1)} = -h_\ell - \frac{1}{(\ell + 1)^2}, \quad D_+^{(\ell)} D_-^{(\ell)} = -h_\ell - \frac{1}{\ell^2}. \quad (14.104)$$

From this, conclude that

$$h_{\ell+1} D_+^{(\ell+1)} u_{n,\ell}(\rho) = \mathcal{E}_{n,\ell} D_+^{(\ell+1)} u_{n,\ell}(\rho) \quad (14.105)$$

whence

$$D_+^{(\ell+1)} u_{n,\ell}(\rho) \propto u_{n,\ell+1}(\rho) \quad (14.106)$$

and  $\mathcal{E}_{n,\ell}$  is independent of  $\ell$ .

Argue that for every  $\mathcal{E}_{n,\ell} < 0$  there is a maximum  $\ell$ . (Hint: examine the effective potential for radial motion.) Call this value  $\ell_{\max}$ , and set  $n = \ell_{\max} + 1$  to show that

$$\mathcal{E}_{n,\ell} = -\frac{1}{n^2}, \quad \ell = 0, \dots, n-1. \quad (14.107)$$

### 14.3 A non-Coulombic central force

The central potential

$$V(r) = -\frac{k}{r} + \frac{c}{r^2} \quad (14.108)$$

is a model (albeit a poor one) for the interaction of the two atoms in a diatomic molecule. (Arnold Sommerfeld called this the “rotating oscillator” potential: see his *Atomic Structure and Spectral Lines*, 3rd ed., 1922, appendix 17.) Steven A. Klein (class of 1989) investigated this potential and found that its energy eigenproblem could be solved exactly.

- Sketch the potential, assuming that  $k$  and  $c$  are both positive.
- Following the method of section 14.3, convert the radial equation of the energy eigenproblem into

$$\left[ -\frac{d^2}{d\rho^2} - \frac{2}{\rho} + \frac{\gamma + \ell(\ell+1)}{\rho^2} \right] u_{n,\ell}(\rho) = \mathcal{E}_{n,\ell} u_{n,\ell}(\rho). \quad (14.109)$$

where  $\gamma = 2cM/\hbar^2$  and where  $\rho$ ,  $\mathcal{E}_{n,\ell}$ , and  $u_{n,\ell}(\rho)$  are to be identified.

- Find two values of  $x$  such that  $x(x+1) = \gamma + \ell(\ell+1)$ . Select whichever one will be most convenient for later use.
- Convince yourself that the solution described in section 14.3 does not depend upon  $\ell$  being an integer, and conclude that the energy eigenvalues are

$$\mathcal{E}_{n,\ell} = \frac{-1}{[n - \ell + \frac{1}{2}(-1 + \sqrt{(2\ell+1)^2 + 4\gamma})]^2} \quad (14.110)$$

where  $n = 1, 2, 3, \dots$  and where for each  $n$ ,  $\ell$  can take on values  $\ell = 0, 1, 2, \dots, n-1$ .

- Verify that this energy spectrum reduces to the Coulomb limit when  $c = 0$ .

### 14.4 The quantum mechanical virial theorem

- Argue that, in an energy eigenstate  $|\eta(t)\rangle$ , the expectation value  $\langle \hat{\mathbf{r}} \cdot \hat{\mathbf{p}} \rangle$  does not change with time.

- b. Hence conclude that  $\langle \eta(t) | [\hat{\mathbf{r}} \cdot \hat{\mathbf{p}}, \hat{H}] | \eta(t) \rangle = 0$ .
- c. Show that  $[\hat{\mathbf{r}} \cdot \hat{\mathbf{p}}, \hat{\mathbf{p}}^2] = 2i\hbar \hat{\mathbf{p}}^2$ , while  $[\hat{\mathbf{r}} \cdot \hat{\mathbf{p}}, V(\hat{\mathbf{r}})] = -i\hbar \hat{\mathbf{r}} \cdot \nabla V(\hat{\mathbf{r}})$ , where  $V(\mathbf{r})$  is any scalar function of the vector  $\mathbf{r}$ . (Hint: For the second commutator, use an explicit position basis representation.)
- d. Suppose the Hamiltonian is

$$\hat{H} = \frac{1}{2m} \hat{p}^2 + V(\hat{r}) = \hat{T} + \hat{V}. \quad (14.111)$$

Define the force function  $\mathbf{F}(\mathbf{r}) = -\nabla V(\mathbf{r})$  and the force operator  $\hat{\mathbf{F}} = \mathbf{F}(\hat{\mathbf{r}})$ . Conclude that, for an energy eigenstate,

$$2\langle \hat{T} \rangle = -\langle \hat{\mathbf{r}} \cdot \hat{\mathbf{F}} \rangle. \quad (14.112)$$

This is the “virial theorem.”

- e. If  $V(\mathbf{r}) = C/r^n$ , show that  $2\langle \hat{T} \rangle = -n\langle \hat{V} \rangle$  for any energy eigenstate, and that

$$\langle \hat{T} \rangle = \frac{n}{n-2} E, \quad \langle \hat{V} \rangle = \frac{-2}{n-2} E, \quad (14.113)$$

for the energy eigenstate with energy  $E$ .

#### 14.5 Research project

Discuss the motion of wavepackets in a Coulombic potential. Does the expectation value of  $\hat{\mathbf{r}}$  follow the classical Kepler ellipse? Is it even restricted to a plane? Does the wavepacket spread out in time (as with the force-free particle) or remain compact (as with the simple harmonic oscillator)?



## Chapter 15

# Identical Particles

Note: Heap algorithm? Permutation group things in an appendix?

Identical particles *not* necessarily interacting, so two particles can be at same point.

### 15.1 Many-particle systems in quantum mechanics

[[This section should be in the chapter on “continuum systems” and then referred to from here.]]

**One particle** moves in one dimension. (Ignore spin.) How can we represent this system’s state?

There are several ways: The ordinary wavefunction  $\psi(x)$  represents the state in terms of the position basis. The momentum wavefunction

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

represents the state in terms of the momentum basis. The energy expansion coefficients

$$c_n = \int_{-\infty}^{+\infty} \eta_n^*(x) \psi(x) dx$$

represent the state in terms of the energy basis. [Meaning that

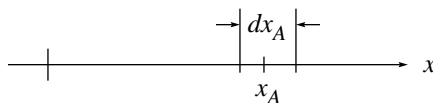
$$\psi(x) = \sum_n c_n \eta_n(x). ]$$

Or we can represent the state in terms of the expectation of position  $\langle x \rangle$ , the expectation of momentum  $\langle p \rangle$ , the indeterminacy in position (which

involves  $\langle x^2 \rangle$ , the indeterminacy in momentum (which involves  $\langle p^2 \rangle$ ), the moments  $\langle x^3 \rangle$  and  $\langle p^3 \rangle$  and so forth, the correlation functions  $\langle xp \rangle$  and  $\langle xp^2x^3p \rangle$  and so forth. You can prove (it's not easy!) that if all these mean values are known then one can reconstruct the wavefunction.

Suppose we know the ordinary position wavefunction. Then if you measure the particle's location, the probability of finding it in a window of width  $dx_A$  about position  $x_A$  is  $|\psi(x_A)|^2 dx_A$ . This is *not* sufficient information to specify the particle's state: it tells you everything there is to know about position, but nothing about momentum or about position-momentum correlation functions.

#### One-particle window



Meanwhile the *amplitude* of finding the particle in this window is  $\psi(x_A) \sqrt{dx_A}$ . If you know the amplitude at every point  $x_A$ , then you *do* have full information about the state.

The normalization is of course

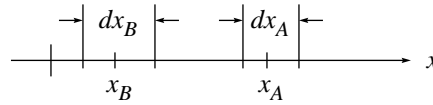
$$\int_{-\infty}^{+\infty} |\psi(x)|^2 dx = 1.$$

*Variations:* If a spin-zero particle moves in three dimensions, the wavefunction is  $\psi(x, y, z)$ . If a spin-half particle ( $s_z = \pm \frac{1}{2}$ ) moves in three dimensions, the wavefunction is  $\psi(x, y, z, s_z)$ , or  $\tilde{\psi}(p_x, p_y, p_z, s_x)$ . In general, when I say things like “the variable  $x$ ”, you will have to generalize in different circumstances to, for example, “the variables  $p_x, p_y, p_z, s_x$ ”.

**Two particles**, say an electron and a neutron, move in one dimension. (Ignore spin.) How can we represent this system's state?

There is now a wavefunction  $\psi(x_A, x_B)$  with the interpretation that if you measure the location of both particles, then the probability of finding the electron in a window of width  $dx_A$  about position  $x_A$ , and the neutron in a window of width  $dx_B$  about position  $x_B$ , is  $|\psi(x_A, x_B)|^2 dx_A dx_B$ .

## Two-particle windows



Note that the letters  $A$  and  $B$  refer to two different positions, not two different particles. The particles are represented by the sequence of arguments: the first argument pertains to the electron, the second argument pertains to the neutron.

I particularly emphasize that the wavefunction applying to the system “electron plus neutron” is *one* function of *two* variables, and is not *two* functions each of *one* variable:

$$\psi_{\text{electron}}(x_A) \quad \text{as well as} \quad \psi_{\text{neutron}}(x_B) \quad \mathbf{NO!}$$

The wavefunction of the system *might* happen to have the factorized form

$$\psi_{\text{electron}}(x_A)\psi_{\text{neutron}}(x_B) \quad \mathbf{PERHAPS}$$

but it does *not necessarily* have this form.

The difference feeds directly into this question: How many (real) numbers does it take to specify a state? In classical mechanics, the answer is straightforward. The state of any single particle is specified through two numbers: the position and momentum of that particle. The state of a collection of several particles is specified through the state of each particle. In summary

particles	real numbers needed to specify classical state
1	2
2	4
3	6
⋮	⋮
$N$	$2N$

In quantum mechanics, the answer is more subtle. To specify the state of even a single particle, one must give the wavefunction  $\psi(x)$ ... an infinite number of complex numbers! For concreteness suppose we approximate this function on a computer, using a grid of 100 points. Then we need 100

complex numbers, that is  $2(100)$  real numbers. But one of these numbers is fixed through the normalization condition, and one is an overall phase that can be set arbitrarily. The end result is that to specify a single-particle wavefunction to this degree of accuracy requires  $2(100) - 2 = 198$  real numbers.

What about two particles? Now we have a wavefunction on a grid of  $100 \times 100$  points, so specifying a two-particle wavefunction to this degree of accuracy requires  $2(100)^2 - 2$  real numbers. This number (19998) is *much larger* than twice 198. To specify the two-particle states, we cannot get away with just specifying two one-particle states. Just as a particle might not have a location, so in a two-particle system an individual particle might not have a state.

In summary

particles	real numbers needed to specify quantal state
1	$2(100) - 2 = 198$
2	$2(100)^2 - 2 = 19998$
3	$2(100)^3 - 2 = 1999998$
$\vdots$	$\vdots$
$N$	$2(100)^N - 2$

Much of the spectacular richness and complexity of the quantum world arises from this rapid increase of information with particle number. (Design of quantum computer.)

**Two identical particles**, say two neutrons, move in one dimension. (Ignore spin.) How can we represent this system's state?

Of course, there is a wavefunction  $\psi(x_A, x_B)$ , but the interpretation is somewhat different. The question is *not* "What is the probability of finding neutron  $\alpha$  within window  $A$  and neutron  $\beta$  within window  $B$ ?" These neutrons are identical, so there is no such thing as "neutron  $\alpha$ " or "neutron  $\beta$ ." The question instead is "What is the probability of finding a neutron within window  $A$  and a neutron within window  $B$ ?" The answer to this question is

$$\begin{cases} 2|\psi(x_A, x_B)|^2 dx_A dx_B & \text{if the windows don't overlap} \\ |\psi(x_A, x_B)|^2 dx_A dx_B & \text{if } x_B = x_A \end{cases}$$

The normalization condition is

$$2 \int_{-\infty}^{+\infty} dx_A \int_{-\infty}^{x_A} dx_B |\psi(x_A, x_B)|^2 = 1 \quad \text{or} \quad \int_{-\infty}^{+\infty} dx_A \int_{-\infty}^{\infty} dx_B |\psi(x_A, x_B)|^2 = 1.$$

If the two particles are identical, then it's certainly true that

$$|\psi(x_A, x_B)|^2 = |\psi(x_B, x_A)|^2.$$

But this condition insures only that the position probabilities are unaffected if you swap the windows. If the two particles are identical, then the same holds for momentum probabilities. In other words, the wavefunctions

$$\psi(x_A, x_B) \text{ and } \psi(x_B, x_A)$$

represent the same state, so

$$\psi(x_A, x_B) = s\psi(x_B, x_A),$$

where  $s$  is a number with modulus unity, not a function of  $x_A$  or  $x_B$ . (The name  $s$  comes from “swap”. We've swapped the subscripts.) Thus, for example,

$$\psi(5, 7) = s\psi(7, 5).$$

But

$$\psi(7, 5) = s\psi(5, 7),$$

so

$$\psi(5, 7) = s^2\psi(5, 7).$$

We conclude that

$$s = \pm 1.$$

In other words, when the wavefunction swaps arguments, it either remains the same or changes sign. In the first case, the wavefunction is called “symmetric under swapping,” (or “under exchange,” or “under interchange”)<sup>1</sup> in the second, “antisymmetric.”

What if there are **three identical particles**? The wavefunction is  $\psi(x_A, x_B, x_C)$  and you can swap either the first and second arguments, or the second and third arguments, or the first and third arguments. The arguments of the next three paragraphs will show that the wavefunction must be either symmetric under each of these three interchanges or else antisymmetric under each of these three interchanges.

<sup>1</sup>I prefer “swap” to emphasize that we're swapping mathematical windows, not exchanging physical particles.

After any swapping, you must produce a wavefunction representing the same state, so any swapping can introduce at most a constant phase factor. Thus

$$\begin{aligned}\psi(x_A, x_B, x_C) &= e^{i\alpha}\psi(x_B, x_A, x_C) \\ &= e^{i\beta}\psi(x_A, x_C, x_B) \\ &= e^{i\gamma}\psi(x_C, x_B, x_A)\end{aligned}$$

The “double swap” argument above shows that  $e^{i\alpha}$  is either  $+1$  or  $-1$ , that  $e^{i\beta}$  is either  $+1$  or  $-1$ , and that  $e^{i\gamma}$  is either  $+1$  or  $-1$ . We can gain more information through repeated swappings that return ultimately to the initial sequence. For example

$$\begin{aligned}\psi(x_A, x_B, x_C) &= e^{i\alpha}\psi(x_B, x_A, x_C) && \text{[[swapping first and second arguments]]} \\ &= e^{i\alpha}e^{i\beta}\psi(x_B, x_C, x_A) && \text{[[swapping second and third arguments]]} \\ &= e^{i\alpha}e^{i\beta}e^{i\gamma}\psi(x_A, x_C, x_B) && \text{[[swapping first and third arguments]]} \\ &= e^{i\alpha}e^{i\beta}e^{i\gamma}e^{i\beta}\psi(x_A, x_B, x_C) && \text{[[swapping second and third arguments]]}\end{aligned}$$

We already know that  $(e^{i\beta})^2 = 1$ , so this argument reveals that  $e^{i\alpha}e^{i\gamma} = 1$ , i.e., these two phase factors are either both  $+1$  or both  $-1$ .

Further arguments of this type will convince you that the three phase factors must either be all  $+1$  or else all  $-1$ . For suppose that

$$\begin{aligned}\psi(x_A, x_B, x_C) &= -\psi(x_B, x_A, x_C) \\ &= +\psi(x_A, x_C, x_B) \\ &= -\psi(x_C, x_B, x_A)\end{aligned}$$

(That is, antisymmetric under swaps of the first and second arguments or the first and third arguments, symmetric under swaps of the second and third arguments.) Then we can go from  $\psi(x_A, x_B, x_C)$  to  $\psi(x_B, x_C, x_A)$  via two different swapping routes:

$$\begin{aligned}\psi(x_A, x_B, x_C) &= (-1)\psi(x_B, x_A, x_C) && \text{[[swapping first and second arguments]]} \\ &= (-1)(+1)\psi(x_B, x_C, x_A) && \text{[[swapping second and third arguments]]}\end{aligned}$$

or

$$\begin{aligned}\psi(x_A, x_B, x_C) &= (-1)\psi(x_C, x_B, x_A) && \text{[[swapping first and third arguments]]} \\ &= (-1)(-1)\psi(x_B, x_C, x_A) && \text{[[swapping first and second arguments]]}\end{aligned}$$

The only function that satisfies both of these conditions is  $\psi(x_A, x_B, x_C) = 0$ .

The other possible “mixed symmetric and antisymmetric” possibility is

$$\begin{aligned}\psi(x_A, x_B, x_C) &= +\psi(x_B, x_A, x_C) \\ &= -\psi(x_A, x_C, x_B) \\ &= +\psi(x_C, x_B, x_A)\end{aligned}$$

but this can be shown impossible by the “two route” argument of the previous paragraph.

- 15.1 **Problem:** Show that the same result applies for functions of four or more arguments by considering first swaps among the first, second, and third arguments; then swaps among the first, second, and fourth arguments; then swaps among the first, second, and fifth arguments; etc.

In conclusion, a wavefunction for any number of identical particles must be either “completely symmetric” (every swap introduces a phase factor of +1) or else “completely antisymmetric” (every swap introduces a phase factor of -1). This is called the “exchange symmetry” of the wavefunction.

- 15.2 **Problem:** If there are two particles, there is one possible swap. If there are three particles, there are three possible swaps. Show that for four particles there are six possible swaps and that for  $N$  particles there are  $N(N - 1)/2$  possible swaps.
- 15.3 **Problem:** Show that the momentum wavefunction has the same interchange symmetry as the position wavefunction (i.e., symmetric or antisymmetric). How about the energy coefficients? (Exactly what does that last question mean?)
- 15.4 **Problem:** Show that exchange symmetry is conserved: If the system starts out in a symmetric state it will remain symmetric at all times in the future, and similarly for antisymmetric.

Given what we’ve said so far, I would guess that a collection of neutrons could start out in a symmetric state (in which case they would be in a symmetric state for all time) or else they could start out in an antisymmetric state (in which case they would be in an antisymmetric state for all time). In fact, however, this is not the case. For suppose you had a

collection of five neutrons in a symmetric state and a different collection of two neutrons in an antisymmetric state. Just by changing which collection is under consideration, you could consider this as one collection of seven neutrons. That collection of seven neutrons would have to be either completely symmetric or completely antisymmetric, and it wouldn't be if the five were in a symmetric state and the two in an antisymmetric state.

So the exchange symmetry has nothing to do with history or with what you consider to be the extent of the collection, but instead depends only on the type of particle. Neutrons, protons, electrons, carbon-13 nuclei, and sigma baryons are always antisymmetric under swapping — they are called *fermions*.<sup>2</sup> Photons, alpha particles, carbon-12 nuclei, and pi mesons are always symmetric under swapping — they are called *bosons*.<sup>3</sup>

Furthermore, all bosons have integral spin and all fermions have half-integral spin. There is a mathematical result in relativistic quantum field theory called “the spin-statistics theorem” that sheds some light on this astounding fact. (See *Pauli and the Spin-Statistics Theorem* by Ian Duck and E.C.G. Sudarshan, and the review of this book by A.S. Wightman in *Am. J. Phys.* **67** (August 1999) 742–746.)

Given their obvious importance, it makes sense to spend some time on the **mathematics of completely symmetric and completely antisymmetric functions**. Given a garden-variety two-variable “seed” function  $f(x_A, x_B)$ , we can build a symmetric function

$$f_S(x_A, x_B) = f(x_A, x_B) + f(x_B, x_A),$$

and an antisymmetric function

$$f_A(x_A, x_B) = f(x_A, x_B) - f(x_B, x_A).$$

Note that these built functions are not necessarily normalized.

<sup>2</sup>Enrico Fermi (1901–1954) of Italy excelled in both experimental and theoretical physics. He directed the building of the first nuclear reactor and produced the first theory of the weak interaction. The Fermi surface in the physics of metals was named in his honor. He elucidated the statistics of what are now called fermions in 1926. He produced so many thoughtful conceptual and estimation problems that such problems are today called “Fermi problems”. I never met him (he died before I was born) but I have met several of his students, and all of them speak of him in that rare tone reserved for someone who is not just a great scientist and a great teacher and a great leader, but also a great human being.

<sup>3</sup>Satyendra Bose (1894–1974) of India made contributions in fields ranging from chemistry to school administration, but his signal contribution was elucidating the statistics of photons. Remarkably, he made this discovery in 1922, three years before Schrödinger developed the concept of wavefunction.



A generalized process works for three-variable functions: The built functions are sums over all  $3!$  permutations of arguments. The function

$$f_S(x_A, x_B, x_C) = f(x_A, x_B, x_C) + f(x_A, x_C, x_B) + f(x_C, x_A, x_B) + f(x_C, x_B, x_A) + f(x_B, x_C, x_A) + f(x_B, x_A, x_C)$$

is completely symmetric while the function

$$f_A(x_A, x_B, x_C) = f(x_A, x_B, x_C) - f(x_A, x_C, x_B) + f(x_C, x_A, x_B) - f(x_C, x_B, x_A) + f(x_B, x_C, x_A) - f(x_B, x_A, x_C)$$

is completely antisymmetric. (These  $3!$  permutations are listed in the sequence called “plain changes” or “the Johnson-Trotter sequence”. This sequence has the benefit that each permutation differs from its predecessor by a single swap of adjacent letters.) This process of building an symmetric function  $f_S$  from arbitrary seed function  $f$  is called “symmetrization”. Similarly for “antisymmetrization”. If the function is a wavefunction, the (anti)symmetrization process is usually understood to include also normalizing the resulting wavefunction.

- 15.5 **Problem:** If the seed  $f(x_A, x_B, x_C)$  happens to be symmetric to begin with, what are the symmetrized and antisymmetrized functions? What if the seed happens to be antisymmetric to begin with?
- 15.6 **Problem:** Show that any two-variable function can be represented as a sum of a symmetric and an antisymmetric function. Can any three-variable function be represented as a sum of a completely symmetric and a completely antisymmetric function?

### *Exchange symmetry and position correlations*

Symmetric implies “huddled together”, antisymmetric implies “spread apart”. (“cluster” / “avoid”)

This is not a result of repulsion. Two electrons, of course, repel each other electrically. This electrical repulsion is reflected through a term in the Hamiltonian of the pair. But the exchange symmetry effect holds even when there is no interaction term in the Hamiltonian. If the two particles are “independent” in that there is no interaction term in the Hamiltonian, in that they don’t interact through a repulsive or attractive force, they still have a tendency to “huddle together” or “spread apart” through the exchange symmetry requirement. (A pair of particles in this situation is said to have “no interaction” in the physics sense of the word “interaction,” even though they do affect each other in the everyday sense of the word “interaction”.)

**Question:** I can see how two electrons, repelling each other through an electric field, can affect each other. But you've just said that two identical particles which don't exert a force on each other nevertheless affect each other. The two particles are not in contact and don't exert a force. What is the mechanism through which one affects the other?

**Answer:** The two particles affect each other through "smelling out" the various positions available to each. Remember that these particles *don't have* positions.

Although this section is titled "Exchange symmetry and position correlations" remember that the symmetry requirement holds also for the momentum wavefunction and for the energy coefficients. Antisymmetric combinations are spread apart in momentum as well as position.

See problem 15.10.

### *Symmetric and antisymmetric bases*

If we want to study identical particles, we'll need to build a basis of symmetric states and a basis of antisymmetric states. Here's how.

Start with a single particle subject to a potential, and solve the energy eigenproblem. Suppose the results are an energy basis of

$$\begin{array}{ll} \eta_1(x) & \epsilon_1 \\ \eta_2(x) & \epsilon_2 \\ \vdots & \vdots \\ \eta_M(x) & \epsilon_M \end{array}$$

In most cases the number of one-particle energy eigenstates  $M$  is infinite, but it's useful to keep that number as a variable anyway. There might or might not be some degeneracies in the system... it doesn't matter.

**Three non-identical particles.** Now suppose there is not one, but three particles subject to this potential, and that they're not identical. We can build a basis of product wavefunctions.

$\eta_1(x_A)\eta_1(x_B)\eta_1(x_C)$	$ 1, 1, 1\rangle$	$\epsilon_1 + \epsilon_1 + \epsilon_1$
$\eta_1(x_A)\eta_1(x_B)\eta_2(x_C)$	$ 1, 1, 2\rangle$	$\epsilon_1 + \epsilon_1 + \epsilon_2$
$\eta_1(x_A)\eta_2(x_B)\eta_1(x_C)$	$ 1, 2, 1\rangle$	$\epsilon_1 + \epsilon_2 + \epsilon_1$
$\eta_2(x_A)\eta_1(x_B)\eta_1(x_C)$	$ 2, 1, 1\rangle$	$\epsilon_2 + \epsilon_1 + \epsilon_1$
$\vdots$	$\vdots$	$\vdots$
$\eta_1(x_A)\eta_7(x_B)\eta_3(x_C)$	$ 1, 7, 3\rangle$	$\epsilon_1 + \epsilon_7 + \epsilon_3$
$\eta_7(x_A)\eta_3(x_B)\eta_1(x_C)$	$ 7, 3, 1\rangle$	$\epsilon_7 + \epsilon_3 + \epsilon_1$
$\vdots$	$\vdots$	$\vdots$
$\eta_M(x_A)\eta_M(x_B)\eta_M(x_C)$	$ M, M, M\rangle$	$\epsilon_M + \epsilon_M + \epsilon_M$

A few remarks: (1) There are  $M^3$  elements in the basis. (2) We have a basis of product wavefunctions, but that doesn't mean that every state is a product state, because an arbitrary state is a sum of basis elements. (3) It's tiring to write always the form in the left column so we abbreviate it through the form in the center column. (4) Sequence matters: the state  $|451\rangle$  is different from the state  $|145\rangle$ . (5) If the three particles don't interact, then this is an energy basis with the eigenvalues shown. But even if they do interact, it's a basis. (6) If the particles don't interact, then there is necessarily degeneracy in this basis. (7) To keep in mind the distinction between this basis for the three-particle system and the basis for the one-particle system from which it is built, we often call the three-particle basis elements "states" and the one-particle basis elements "levels". The levels are the building blocks out of which states are constructed.

**Building a symmetric basis.** Any wavefunction can be expressed as a sum over the above basis,

$$\psi(x_A, x_B, x_C) = \sum_{r=1}^M \sum_{s=1}^M \sum_{t=1}^M c_{r,s,t} \eta_r(x_A) \eta_s(x_B) \eta_t(x_C) = \sum_{r,s,t} c_{r,s,t} |r, s, t\rangle,$$

but if we have three identical bosons, we're not interested in any wavefunction, we're interested in symmetric wavefunctions. To build a symmetric wavefunction, we execute the symmetrization process on  $\psi(x_A, x_B, x_C)$ . Doing so, we conclude that this symmetric wavefunction can be expressed as a sum over the symmetrization of each basis element.

Let's think a bit about the symmetrization of

$$\eta_r(x_A) \eta_s(x_B) \eta_t(x_C) \quad \text{also known as} \quad |r, s, t\rangle.$$

When we introduced the symmetrization process, we permuted the variables (representing position windows)  $x_A$ ,  $x_B$ , and  $x_C$ . But if the seed function

happens to be a product like this, it's obviously the same thing to permute the level indices  $r$ ,  $s$ , and  $t$ . We represent the symmetrization of  $|r, s, t\rangle$  as

$$\hat{\mathcal{S}}|r, s, t\rangle = \text{const} (|r, s, t\rangle + |r, t, s\rangle + |t, r, s\rangle + |t, s, r\rangle + |s, t, r\rangle + |s, r, t\rangle)$$

where “const” is a normalization constant.

If we go through and symmetrize each element of the basis for three non-identical particles, we will find a basis for symmetric states. Let's start with  $|1, 1, 1\rangle$ . This symmetrizes to itself:

$$\hat{\mathcal{S}}|1, 1, 1\rangle = |1, 1, 1\rangle.$$

Next comes  $|1, 1, 2\rangle$ :

$$\hat{\mathcal{S}}|1, 1, 2\rangle = \text{const} (|1, 1, 2\rangle + |1, 2, 1\rangle + |2, 1, 1\rangle + |2, 1, 1\rangle + |1, 2, 1\rangle + |1, 1, 2\rangle) = 2 \text{const} (|1, 1, 2\rangle + |1, 2, 1\rangle + |2, 1, 1\rangle)$$

It's clear, then, that

$$\hat{\mathcal{S}}|1, 1, 2\rangle = \hat{\mathcal{S}}|1, 2, 1\rangle = \hat{\mathcal{S}}|2, 1, 1\rangle,$$

so we must discard two of these three states from our basis. It's equally clear that all states built through symmetrizing any three given levels are the same state. For example

$$\hat{\mathcal{S}}|3, 9, 2\rangle = \hat{\mathcal{S}}|3, 2, 9\rangle = \hat{\mathcal{S}}|2, 3, 9\rangle = \hat{\mathcal{S}}|2, 9, 3\rangle = \hat{\mathcal{S}}|9, 2, 3\rangle = \hat{\mathcal{S}}|9, 3, 2\rangle,$$

and we must discard five of these six states from our basis.

We are left with a basis of

$$\frac{M(M+1)(M+2)}{3!}$$

symmetric elements. One of the neat things about these elements is that they're long ... for example one of them is

$$\frac{1}{\sqrt{3!}} [\eta_1(x_A)\eta_7(x_B)\eta_3(x_C) + \eta_1(x_A)\eta_3(x_B)\eta_7(x_C) + \eta_3(x_A)\eta_1(x_B)\eta_7(x_C) + \eta_3(x_A)\eta_7(x_B)\eta_1(x_C) + \eta_7(x_A)\eta_3(x_B)\eta_1(x_C) + \eta_7(x_A)\eta_1(x_B)\eta_3(x_C)]$$

but to specify them we need only state the three levels that go into building it (the three “building blocks” that go into making it). [This was *not* the case for three non-identical particles.] Consequently one often speaks of this state as “a particle in level 1, a particle in level 7, and a particle in level 3”. This phrase is *not* correct: If a particle *were* in level 7, then it could be distinguished as “the particle in level 7” and hence would not be identical to the other two particles. The correct statement is that the *system* is in the antisymmetric state given above, and that the *individual particles* do not have states. On the other hand, the correct statement is a

mouthful and you may use the “balls in buckets” picture as shorthand — as long as you say it but don’t think it.

**Building an antisymmetric basis.** We can build a basis of states, each of which is antisymmetric, in a parallel manner by antisymmetrizing each element of the basis for non-identical particles and discarding duplicates.

The antisymmetrization of

$$\eta_r(x_A)\eta_s(x_B)\eta_t(x_C) \quad \text{also known as} \quad |r, s, t\rangle,$$

results in

$$\hat{\mathcal{A}}|r, s, t\rangle = \text{const} (|r, s, t\rangle - |r, t, s\rangle + |t, r, s\rangle - |t, s, r\rangle + |s, t, r\rangle - |s, r, t\rangle)$$

where “const” is again a normalization constant.

Let’s start with  $|1, 1, 1\rangle$ . This antisymmetrizes to zero. Same with  $|1, 1, 2\rangle$ :

$$\hat{\mathcal{A}}|1, 1, 2\rangle = \text{const} (|1, 1, 2\rangle - |1, 2, 1\rangle + |2, 1, 1\rangle - |2, 1, 1\rangle + |1, 2, 1\rangle - |1, 1, 2\rangle) = 0.$$

It’s clear, in fact, that *any* basis element with two indices the same will antisymmetrize to zero. The only way to avoid antisymmetrization to zero is for all of the level indices to differ. Furthermore

$$\hat{\mathcal{A}}|r, s, t\rangle = -\hat{\mathcal{A}}|r, t, s\rangle = \hat{\mathcal{A}}|t, r, s\rangle = -\hat{\mathcal{A}}|t, s, r\rangle = \hat{\mathcal{A}}|s, t, r\rangle = -\hat{\mathcal{A}}|s, r, t\rangle$$

so the six distinct basis elements  $|1, 7, 3\rangle$ ,  $|7, 3, 1\rangle$ ,  $|3, 7, 1\rangle$ , etc. all antisymmetrize to the same thing.

Discarding the duplicates results in a basis of

$$\frac{M(M-1)(M-2)}{3!}$$

symmetric elements.

Once again these states have long expressions like

$$\frac{1}{\sqrt{3!}} [\eta_1(x_A)\eta_7(x_B)\eta_3(x_C) - \eta_1(x_A)\eta_3(x_B)\eta_7(x_C) + \eta_3(x_A)\eta_1(x_B)\eta_7(x_C) - \eta_3(x_A)\eta_7(x_B)\eta_1(x_C) + \eta_7(x_A)\eta_3(x_B)\eta_1(x_C) - \eta_7(x_A)\eta_1(x_B)\eta_3(x_C)]$$

but to specify the three-particle state we need only list the one-particle building blocks (“levels”) used in its construction. This results in almost the same “balls in buckets” picture that we drew for symmetric wavefunctions, with the additional restriction that any bucket can contain only one or zero balls. Once again you may use the “balls in buckets” picture as a shorthand,

as long as you keep in mind that it conceals a considerably more intricate process of building and antisymmetrizing.

**Generalizations.** It is easy to generalize this procedure for building antisymmetric and symmetric many-particle basis states out of one-particle levels for any number of particles. The only special case is for two particles, where the symmetric basis has  $M(M+1)/2$  elements and the antisymmetric basis has  $M(M-1)/2$  elements. Putting these two bases together results in a full basis of  $M^2$  elements. This reflects the fact that any function of two variables can be written as the sum of an antisymmetric and a symmetric function. The same is *not* true for systems of three or more particles.

- 15.7 **Problem:** Count the elements in the antisymmetric and symmetric bases for  $N$  particles rather than three. (Continue to use  $M$  levels.) Does your expression have the proper limits when  $N = 1$  and when  $N = M$ ?
- 15.8 **Problem:** Find the normalization constant for  $\hat{S}|7, 3, 7\rangle$ .
- 15.9 **Problem:** Any two-variable function may be written as a sum of a symmetric and an antisymmetric function. Consequently the union of the symmetric basis and the antisymmetric basis is a basis for the set of all two-variable functions. Show that neither of these statements is true for functions of three variables.
- 15.10 **Mean separation**  
 (Be sure to read Griffiths section 5.1.2, “Exchange forces,” before attempting this problem.) Two noninteracting particles, each of mass  $m$ , are in an infinite square well of width  $L$ . The associated one-body energy eigenstates are  $\eta_n(x)$  and  $\eta_m(x)$ , where

$$\eta_n(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi}{L}x\right).$$

Calculate the root-mean-square separation

$$\sqrt{\langle (x_A - x_B)^2 \rangle}$$

if these are

- two non-identical particles, one in state  $\eta_n(x_A)$  and the other in state  $\eta_m(x_B)$
- two identical bosons, in state

$$\frac{1}{\sqrt{2}} [\eta_n(x_A)\eta_m(x_B) + \eta_m(x_A)\eta_n(x_B)]$$

- c. two identical fermions, in state

$$\frac{1}{\sqrt{2}} [\eta_n(x_A)\eta_m(x_B) - \eta_m(x_A)\eta_n(x_B)]$$

Do your results always adhere to our general rule of “symmetric means huddled together; antisymmetric means spread apart”?

- d. *Bonus, for the mathematically inclined.* Solve all three parts above in one step by computing the root-mean-square separation for the two-particle state

$$\cos \theta \eta_n(x_A)\eta_m(x_B) + \sin \theta \eta_m(x_A)\eta_n(x_B)$$

where  $\theta$  is a parameter of no physical significance.

### 15.11 Building basis states

Suppose you had three particles and three “building block” levels (say the orthonormal levels  $\eta_1(x)$ ,  $\eta_3(x)$ , and  $\eta_7(x)$ ). Construct normalized three-particle basis states for the case of

- three non-identical particles
- three identical bosons
- three identical fermions

How many states are there in each basis? Repeat for three particles with four one-particle levels, but in this case simply count and don’t write down all the three-particle states.

## 15.2 An antisymmetric basis for the helium problem

Helium: two electrons and one nucleus. The three-body problem! But wait, the three-body problem hasn’t been solved exactly even in classical mechanics, there’s no hope for an exact solution in quantum mechanics. Does this mean we give up? No. If you give up on a problem you can’t solve exactly, you give up on life.<sup>4</sup> Instead, we look for approximate solutions.

If we take account of the Coulomb forces, but ignore things like the finite size of the nucleus, nuclear motion, relativistic motion of the electron,

<sup>4</sup>Can’t find the exact perfect apartment to rent? Can’t find the exact perfect candidate to vote for? Can’t find the exact perfect friend? Of course you can’t find any of these things. But we get on with our lives accepting imperfections because we realize that the alternatives (homelessness, political corruption, friendlessness) are worse.

spin-orbit effects, and so forth, the Hamiltonian for two electrons and one nucleus is

$$\begin{aligned}\hat{H} &\doteq \left[ -\frac{\hbar^2}{2m_e} \nabla_A^2 - \frac{e^2}{4\pi\epsilon_0} \frac{1}{r_A} \right] + \left[ -\frac{\hbar^2}{2m_e} \nabla_B^2 - \frac{e^2}{4\pi\epsilon_0} \frac{1}{r_B} \right] + \left[ \frac{e^2}{4\pi\epsilon_0} \frac{1}{|\vec{r}_A - \vec{r}_B|} \right] \\ &= \underbrace{\widehat{\text{KE}}_A + \hat{U}_{nA}}_{\equiv \hat{H}_A} + \underbrace{\widehat{\text{KE}}_B + \hat{U}_{nB}}_{\equiv \hat{H}_B} + \underbrace{\hat{U}_{AB}}_{\equiv \hat{H}'}\end{aligned}$$

Recall that in using the subscripts “A” and “B” we are not labeling the *electrons* as “electron A” and “electron B”: the electrons are identical and can’t be labeled. Instead we are labeling the *points in space* where an electron might exist as “point A” and “point B”.

We will look for eigenstates of the partial Hamiltonian  $\hat{H}_A + \hat{H}_B$ . These are not eigenstates of the full Hamiltonian, but they *are* a basis, and they can be used as a place to begin perturbation theory.

### One-particle levels

We begin by finding the one-particle levels (or “orbitals”) for just the Hamiltonian  $\hat{H}_A$ . We combine these with levels for  $\hat{H}_B$  and antisymmetrize the result.

The problem  $\hat{H}_A$  is just the Hydrogen atom Coulomb problem with two changes:

nuclear mass is  $4m_p \implies$  very small effect (“ignore nuclear motion”)

nuclear charge is  $2e \implies$  the Rydberg is  $\text{Ry} = \frac{m_e}{2\hbar^2} \left( \frac{e^2}{4\pi\epsilon_0} \right)^2$ , so  $E_{nA}^{(A)} = -\frac{4\text{Ry}}{n_A^2}$

Similarly, the energy eigenstates for  $\hat{H}_A$  are represented by familiar functions like

$$\eta_{nlm}(\mathbf{r})|\uparrow\rangle \quad \text{or} \quad \eta_{nlm}(\mathbf{r})\chi_+$$

Soon we will need to keep track of  $\hat{H}_A$  versus  $\hat{H}_B$ . A notation like  $\eta_{nlm}(\mathbf{r}_A)|\uparrow\rangle$  is fine for the space part of the eigenstate, but leaves the spin part ambiguous. We will hence use notation like

$$\eta_{nlm}(A)\chi_+(A)$$

to mean the same thing.



[[Notice that the eigenstates don't *have* to take on the factorized form of "space part"  $\times$  "spin part" — for example

$$\frac{1}{\sqrt{2}}[\eta_{200}(\mathbf{r})\chi_+ + \eta_{210}(\mathbf{r})\chi_-]$$

is a perfectly good eigenstate — but that the factorized form is particularly convenient for working with. (If we were to consider spin-orbit coupling, then the eigenstates could *not* take the factorized form.)]]

### ***Antisymmetrization***

Recall how we build an antisymmetrized wavefunction from a product of two one-particle levels,  $\eta_n(A)$  and  $\eta_m(B)$ :

$$\hat{A}\eta_n(A)\eta_m(B) = \frac{1}{\sqrt{2}}[\eta_n(A)\eta_m(B) - \eta_m(A)\eta_n(B)]$$

(The normalization factor  $1/\sqrt{2}$  holds when  $\eta_n(A)$  and  $\eta_m(A)$  are orthogonal.)

Two theorems:

- If you antisymmetrize a product of the same two levels, you end up with zero:

$$\hat{A}\eta_n(A)\eta_n(B) = 0.$$

- If you antisymmetrize two levels in the opposite sequence, you end up with the same state:

$$\hat{A}\eta_n(A)\eta_m(B) = -\hat{A}\eta_m(A)\eta_n(B).$$

### ***The ground state***

The ground levels of  $\hat{H}_A$  and of  $\hat{H}_B$  are both doubly degenerate due to spin. So if you had distinguishable particles, the ground state of  $\hat{H}_A + \hat{H}_B$  would be four-fold degenerate:

distinguishable

$$\begin{aligned} &\eta_{100}(A)\chi_+(A)\eta_{100}(B)\chi_+(B) \\ &\eta_{100}(A)\chi_+(A)\eta_{100}(B)\chi_-(B) \\ &\eta_{100}(A)\chi_-(A)\eta_{100}(B)\chi_+(B) \\ &\eta_{100}(A)\chi_-(A)\eta_{100}(B)\chi_-(B) \end{aligned}$$

But if you have identical fermions, two of these basis states antisymmetrize to zero, and the other two antisymmetrize to the same state:

distinguishable	antisymmetrized
$\eta_{100}(A)\chi_+(A)\eta_{100}(B)\chi_+(B)$	0
$\eta_{100}(A)\chi_+(A)\eta_{100}(B)\chi_-(B)$	$\frac{1}{\sqrt{2}}[\eta_{100}(A)\chi_+(A)\eta_{100}(B)\chi_-(B) - \eta_{100}(A)\chi_-(A)\eta_{100}(B)\chi_+(B)]$
$\eta_{100}(A)\chi_-(A)\eta_{100}(B)\chi_+(B)$	$-\frac{1}{\sqrt{2}}$ [above]
$\eta_{100}(A)\chi_-(A)\eta_{100}(B)\chi_-(B)$	0

Hence the Hamiltonian  $\hat{H}_A + \hat{H}_B$  has a non-degenerate ground state, namely

$$\eta_{100}(A)\eta_{100}(B)\frac{1}{\sqrt{2}}[\chi_+(A)\chi_-(B) - \chi_-(A)\chi_+(B)].$$

It's common to hear things like "In the ground state of Helium, one electron is in one-body level  $|100\rangle$  with spin up and the other is in one-body level  $|100\rangle$  with spin down." This claim is false. The equation makes it clear that "In the ground state of Helium, one electron is in one-body level  $|100\rangle$ , the other is in one-body level  $|100\rangle$ , and the spins are entangled." If the first phrase *were* correct, then you *would* be able to distinguish the two electrons, and they *would not* be identical. But it's not correct.

### ***States built from one ground level***

Now build a state by combining the ground level of one Hamiltonian with  $|nlm\rangle$  from the other. If you had distinguishable particles, this "combination" means a simple multiplication, and there would be eight states (all with the same energy):

distinguishable
$\eta_{100}(A)\chi_+(A)\eta_{nlm}(B)\chi_+(B)$
$\eta_{100}(A)\chi_+(A)\eta_{nlm}(B)\chi_-(B)$
$\eta_{100}(A)\chi_-(A)\eta_{nlm}(B)\chi_+(B)$
$\eta_{100}(A)\chi_-(A)\eta_{nlm}(B)\chi_-(B)$
$\eta_{nlm}(A)\chi_+(A)\eta_{100}(B)\chi_+(B)$
$\eta_{nlm}(A)\chi_+(A)\eta_{100}(B)\chi_-(B)$
$\eta_{nlm}(A)\chi_-(A)\eta_{100}(B)\chi_+(B)$
$\eta_{nlm}(A)\chi_-(A)\eta_{100}(B)\chi_-(B)$

But if you have identical fermions, the “combination” means a multiplication followed by an antisymmetrization. Because of the second theorem concerning antisymmetrization, each of the last four products above antisymmetrize to the same state as one of the first four products. The first four products result in antisymmetrized states as follows:

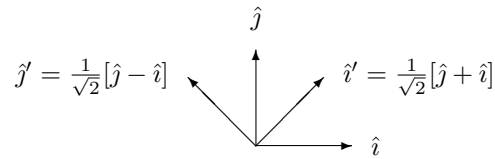
	distinguishable	antisymmetrized
(a)	$\eta_{100}(A)\chi_+(A)\eta_{n\ell m}(B)\chi_+(B)$	$\frac{1}{\sqrt{2}}[\eta_{100}(A)\chi_+(A)\eta_{n\ell m}(B)\chi_+(B) - \eta_{n\ell m}(A)\chi_+(A)\eta_{100}(B)\chi_+(B)]$
(b)	$\eta_{100}(A)\chi_+(A)\eta_{n\ell m}(B)\chi_-(B)$	$\frac{1}{\sqrt{2}}[\eta_{100}(A)\chi_+(A)\eta_{n\ell m}(B)\chi_-(B) - \eta_{n\ell m}(A)\chi_-(A)\eta_{100}(B)\chi_+(B)]$
(c)	$\eta_{100}(A)\chi_-(A)\eta_{n\ell m}(B)\chi_+(B)$	$\frac{1}{\sqrt{2}}[\eta_{100}(A)\chi_-(A)\eta_{n\ell m}(B)\chi_+(B) - \eta_{n\ell m}(A)\chi_+(A)\eta_{100}(B)\chi_-(B)]$
(d)	$\eta_{100}(A)\chi_-(A)\eta_{n\ell m}(B)\chi_-(B)$	$\frac{1}{\sqrt{2}}[\eta_{100}(A)\chi_-(A)\eta_{n\ell m}(B)\chi_-(B) - \eta_{n\ell m}(A)\chi_-(A)\eta_{100}(B)\chi_-(B)]$

Antisymmetrized expressions (a) and (d) readily factor into a space part times a spin part:

$$(a) \implies \frac{1}{\sqrt{2}}[\eta_{100}(A)\eta_{n\ell m}(B) - \eta_{n\ell m}(A)\eta_{100}(B)]\chi_+(A)\chi_+(B)$$

$$(d) \implies \frac{1}{\sqrt{2}}[\eta_{100}(A)\eta_{n\ell m}(B) - \eta_{n\ell m}(A)\eta_{100}(B)]\chi_-(A)\chi_-(B)$$

But expressions (b) and (c) do *not* factor. One thing to do about this is nothing — after all, there’s no *requirement* that the wavefunctions factorize. But another approach is to look for a simple change of basis (remember, these four states all have the same energy of  $\hat{H}_A + \hat{H}_B$ ). Someone (I don’t know who) thought about the favorite change of basis in planar geometry — a rotation of the axes by  $45^\circ$ :



Applying this transformation to the basis elements (“unit vectors”) given through (b) and (c) results in the new basis elements

$$\begin{aligned} \frac{1}{\sqrt{2}}[(b) + (c)] &= \frac{1}{2} \{ [\eta_{100}(A)\eta_{n\ell m}(B) - \eta_{n\ell m}(A)\eta_{100}(B)]\chi_+(A)\chi_-(B) \\ &\quad + [\eta_{100}(A)\eta_{n\ell m}(B) - \eta_{n\ell m}(A)\eta_{100}(B)]\chi_-(A)\chi_+(B) \} \\ &= \frac{1}{\sqrt{2}}[\eta_{100}(A)\eta_{n\ell m}(B) - \eta_{n\ell m}(A)\eta_{100}(B)] \frac{1}{\sqrt{2}}[\chi_+(A)\chi_-(B) + \chi_-(A)\chi_+(B)] \end{aligned}$$

and

$$\begin{aligned} \frac{1}{\sqrt{2}}[(b) - (c)] &= \frac{1}{2} \{ [\eta_{100}(A)\eta_{n\ell m}(B) + \eta_{n\ell m}(A)\eta_{100}(B)]\chi_+(A)\chi_-(B) \\ &\quad - [\eta_{100}(A)\eta_{n\ell m}(B) + \eta_{n\ell m}(A)\eta_{100}(B)]\chi_-(A)\chi_+(B) \} \\ &= \frac{1}{\sqrt{2}} [\eta_{100}(A)\eta_{n\ell m}(B) + \eta_{n\ell m}(A)\eta_{100}(B)] \frac{1}{\sqrt{2}} [\chi_+(A)\chi_-(B) - \chi_-(A)\chi_+(B)]. \end{aligned}$$

This process results in an antisymmetric basis of

$$\begin{aligned} &\frac{1}{\sqrt{2}} [\eta_{100}(A)\eta_{n\ell m}(B) - \eta_{n\ell m}(A)\eta_{100}(B)]\chi_+(A)\chi_+(B) \\ &\frac{1}{\sqrt{2}} [\eta_{100}(A)\eta_{n\ell m}(B) - \eta_{n\ell m}(A)\eta_{100}(B)] \frac{1}{\sqrt{2}} [\chi_+(A)\chi_-(B) + \chi_-(A)\chi_+(B)] \\ &\frac{1}{\sqrt{2}} [\eta_{100}(A)\eta_{n\ell m}(B) - \eta_{n\ell m}(A)\eta_{100}(B)]\chi_-(A)\chi_-(B) \\ &\frac{1}{\sqrt{2}} [\eta_{100}(A)\eta_{n\ell m}(B) + \eta_{n\ell m}(A)\eta_{100}(B)] \frac{1}{\sqrt{2}} [\chi_+(A)\chi_-(B) - \chi_-(A)\chi_+(B)]. \end{aligned}$$

The first three elements are called a “triplet” (with “space antisymmetric, spin symmetric”). The last element is called a “singlet” (with “space symmetric, spin antisymmetric”). This particular basis has three nice properties: (1) Every basis element factorizes into a spatial part times a spin part. (2) Every basis element factorizes into a symmetric part times an antisymmetric part. (3) All three elements in the triplet have identical spatial parts.

The third point means that when we take account of electron-electron repulsion through perturbation theory, we will necessarily find that all three elements of any triplet remain degenerate even when the effects of the sub-Hamiltonian  $\hat{U}_{AB}$  are considered.

[[This process works for combining an arbitrary level  $|n\ell m\rangle$  with a ground state level  $|100\rangle$ . Since this should work for any  $|n\ell m\rangle$ , what happens if we take  $|n\ell m\rangle = |100\rangle$ , the situation we first considered? In particular, what’s up with the normalization?]]

### ***States built from two excited levels***

What happens if we carry out the above process but combining an excited level of one sub-Hamiltonian (say  $\eta_{200}(A)$ ) with an arbitrary level of the other sub-Hamiltonian (say  $\eta_{n\ell m}(B)$ )?

The process goes on in a straightforward way, but it turns out that the resulting eigenenergies are always so high that the atom is unstable:

it decays rapidly to a positive helium atom plus an ejected electron. Such electrons are called “Auger electrons” (pronounced “oh-jey” because Pierre Victor Auger was French) and Auger electron spectroscopy is an important analytical technique in surface and materials science.

### 15.12 Electron-electron repulsion

In class we wrote the (approximate) Hamiltonian for helium as  $\hat{H}_A + \hat{H}_B + \hat{U}_{AB}$  and found antisymmetric energy eigenstates for  $\hat{H}_A + \hat{H}_B$  that we called  $1^1S$ ,  $2^3S$ ,  $2^1P$ , and so forth. Then we qualitatively discussed how the energy associated with such a state would change, under perturbation theory, through the electron-electron repulsion term  $\hat{U}_{AB}$ . Write down expressions for the first-order energy shifts due to  $\hat{U}_{AB}$  for  $2^1S$ ,  $2^3S$ , and  $2^1P$ . (That is, set up the integrals in terms of the one-particle eigenstates  $\eta_{nlm}(\mathbf{r})$ . Do *not* evaluate the integrals.) *Bonus:* Argue that the energy shift for  $2^1P$  is greater than the shift for  $2^1S$ .

### 15.13 Two-electron ions

Apply the techniques of Griffiths, section 7.2, “Ground State of Helium,” to the  $H^-$  and  $Li^+$  ions. Each of these ions has two electrons, like helium, but nuclear charges  $Z = 1$  and  $Z = 3$ , respectively. For each ion find the effective (partially shielded) nuclear charge and determine the best upper bound on the ground state energy.

### 15.14 The meaning of two-particle wavefunctions (Old)

- a. The wavefunction  $\psi(x_A, x_B)$  describes two non-identical particles in one dimension. Does

$$\int_{-\infty}^{\infty} dx_A \int_{-\infty}^{\infty} dx_B |\psi(x_A, x_B)|^2 \quad (15.1)$$

equal one (the usual normalization) or two (the number of particles)? Write integral expressions for:

- i. The probability of finding particle A between  $x_1$  and  $x_2$  and particle B between  $x_3$  and  $x_4$ .
  - ii. The probability of finding particle A between  $x_1$  and  $x_2$ , regardless of where particle B is.
- b. The wavefunction  $\psi(x_A, x_B)$  describes two identical particles in one dimension. Does

$$\int_{-\infty}^{\infty} dx_A \int_{-\infty}^{\infty} dx_B |\psi(x_A, x_B)|^2 \quad (15.2)$$

equal one or two? Assuming that  $x_1 < x_2 < x_3 < x_4$ , write integral expressions for:

- i. The probability of finding one particle between  $x_1$  and  $x_2$  and the other between  $x_3$  and  $x_4$ .
- ii. The probability of finding a particle between  $x_1$  and  $x_2$ .
- c. Look up the definition of “configuration space” in a classical mechanics book. Does the wavefunction inhabit configuration space or conventional three-dimensional position space? For discussion: Does your answer have any bearing upon the question of whether the wavefunction is “physically real” or a “mathematical convenience”? Does it affect your thoughts concerning measurement and the “collapse of the wavepacket”?

**15.15 Symmetric and close together, antisymmetric and far apart (Old)**

In lecture I argued that symmetric wavefunctions describe particles that huddle together while antisymmetric wavefunctions describe particles that avoid one another.

- a. Illustrate this principle as follows: Construct symmetric and antisymmetric two-particle wavefunctions out of the single-particle wavefunctions

$$\eta_1(x) = \sqrt{\frac{2}{L}} \sin\left(\pi \frac{x}{L}\right) \quad \text{and} \quad \eta_2(x) = \sqrt{\frac{2}{L}} \sin\left(2\pi \frac{x}{L}\right), \quad 0 \leq x \leq L, \quad (15.3)$$

which are the first and second energy eigenfunctions for the infinite square well of width  $L$ . For each (anti)symmetrized function make a plot of  $x_A$  and  $x_B$  and shade in regions of high probability density.

- b. Prove that if the two wavefunctions  $\psi(\mathbf{x})$  and  $\phi(\mathbf{x})$  are orthogonal, then the expectation value of  $(\mathbf{x}_A - \mathbf{x}_B)^2$  for the antisymmetric combination of the two wavefunctions is greater than or equal to that for the symmetric combination.

**15.16 Symmetrization and antisymmetrization (mathematical) (Old)**

- a. Show that any two-variable function can be written as the sum of a symmetric function and an antisymmetric function.
- b. Show that this is *not* true for functions of three variables. [Hint: Try the counterexample  $f(x, y, z) = g(x)$ .]
- c. There is a function of three variables that is:
  - i. Antisymmetric under interchange of the first and second variables:  $f(x, y, z) = -f(y, x, z)$ .

- ii. Symmetric under interchange of the second and third variables:  $f(x, y, z) = f(x, z, y)$ .
- iii. Symmetric under interchange of the first and third variables:  $f(x, y, z) = f(z, y, x)$ .

Find this function and show that it is unique.





## Chapter 16

# Breather

Why do we need a breather at this point?

There are no new principles, but lots of applications. The applications will shed light on the principles and the principles will shed light on the applications. I will not attempt to fool you: the applications will be hard. For example, the three-body problem has not been solved in classical mechanics. In the richer, more intricate, world of quantum mechanics, we will not solve it either.

You know from solving problems in classical mechanics that you should think first, before plunging into a hard problem. You know, for example, that if you use the appropriate variables, select the most appropriate coordinate system, or use a symmetry – that you can save untold amounts of labor. (See, for example, George Pólya, *How to Solve it* (Doubleday, Garden City, NY, 1957) Sanjoy Mahajan, *Street-Fighting Mathematics* (MIT Press, Cambridge, MA, 2010).) This rule holds even more so in the more complex world of quantum mechanics.

And that's the role of this chapter. We'll take a breather, pull back from the details, and organize ourselves for facing the difficult problems that lie before us.

Henry David Thoreau, *Walden* (1854): “I went to the woods because I wished to live deliberately, to front only the essential facts of life, and see if I could not learn what it had to teach, and not, when I came to die, discover that I had not lived.”

### 16.1 Scaled variables

Here's the energy eigenproblem for the hydrogen atom (at the level of approximation ignoring collisions, radiation, nuclear mass, nuclear size, spin, magnetic effects, relativity, and the quantum character of the electromagnetic field):

$$\left[ -\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) - \frac{e^2}{4\pi\epsilon_0} \frac{1}{r} \right] \eta(\vec{r}) = E\eta(\vec{r}). \quad (16.1)$$

This section uses dimensional analysis to find the characteristic length and characteristic energy for this problem, then uses scaled variables to express this equation in a more natural and more easily-worked-with form.

Whatever result comes out of this energy eigenequation, whether the result be a length, or an energy, or anything else, the result can only depend on three parameters:  $\hbar$ ,  $m$ , and  $e^2/4\pi\epsilon_0$ . These parameters have the following dimensions:

	dimensions	base dimensions (mass, length, time)
$\hbar$	[Energy×T]	[ML <sup>2</sup> /T]
$m$	[M]	[M]
$e^2/4\pi\epsilon_0$	[Energy×L]	[ML <sup>3</sup> /T <sup>2</sup> ]

How can we build a quantity with the dimensions of length from these three parameters? Well, the quantity will have to involve  $\hbar$  and  $e^2/4\pi\epsilon_0$ , because these are the only parameters that include the dimensions of length, but we'll have to get rid those dimensions of time. We can do that by squaring the first and dividing by the third:

$$\frac{\text{quantity}}{\hbar^2} \Big| \frac{\text{dimensions}}{e^2/4\pi\epsilon_0} \Big| [\text{ML}]$$

And now there's only one way to get rid of the dimension of mass (without reintroducing a dimension of time), namely dividing this quantity by  $m$ :

$$\frac{\text{quantity}}{\hbar^2} \Big| \frac{\text{dimensions}}{m e^2/4\pi\epsilon_0} \Big| [\text{L}]$$

We have uncovered the one and only way to combine these three parameters to produce a quantity with the dimensions of length. We define the *Bohr radius*

$$a_0 \equiv \frac{\hbar^2}{m e^2 / 4\pi\epsilon_0} \approx 0.05 \text{ nm.} \quad (16.2)$$

This quantity sets the typical scale for any length in a hydrogen atom. For example, if I ask for the mean distance from the nucleus to an electron in energy eigenstate  $\eta_{5,4,-3}(\vec{r})$  the answer will be some pure (dimensionless) number times  $a_0$ . If I ask for the uncertainty in  $\hat{x}$  of an electron in state  $\eta_{2,1,0}(\vec{r})$  the answer will be some pure number times  $a_0$ .

Is there a characteristic energy? Yes, it is given through  $e^2/4\pi\epsilon_0$  divided by  $a_0$ . The characteristic energy is

$$\mathcal{E}_0 \equiv \frac{m (e^2/4\pi\epsilon_0)^2}{\hbar^2} = 2\text{Ry.} \quad (16.3)$$

This characteristic energy doesn't have its own name, because we just call it twice the Rydberg energy (the minimum energy required to ionize a hydrogen atom). It plays the same role for energies that  $a_0$  plays for lengths: Any energy value concerning hydrogen will be a pure number times  $\mathcal{E}_0$ .

Now is the time to introduce scaled variables. Whenever I specify a length, I specify that length in terms of some other length. For example, when I say the Eiffel tower is 324 meters tall, I mean that the ratio of the height of the Eiffel tower to the length of the prototype meter bar — that bar stored in a vault in Sèvres, France — is 324.

Now, what is the relevance of the prototype meter bar to atomic phenomena? None! Instead of measuring atomic lengths relative to the prototype meter, it makes more sense to measure them relative to something atomic, namely to the Bohr radius. I define the dimensionless “scaled length”  $\tilde{x}$  as

$$\tilde{x} \equiv \frac{x}{a_0}, \quad (16.4)$$

and it's my preference to measure atomic lengths using this standard, rather than using the prototype meter bar as a standard.

So, what is the energy eigenproblem (16.1) written in terms of scaled lengths? For any function  $f(x)$ , the chain rule of calculus tells us that

$$\frac{\partial f(x)}{\partial x} = \frac{\partial f(\tilde{x})}{\partial \tilde{x}} \frac{\partial \tilde{x}}{\partial x} = \frac{\partial f(\tilde{x})}{\partial \tilde{x}} \frac{1}{a_0}$$

and consequently that

$$\frac{\partial^2 f(x)}{\partial x^2} = \frac{\partial^2 f(\tilde{x})}{\partial \tilde{x}^2} \frac{1}{a_0^2}.$$

Consequently the energy eigenproblem (16.1) is

$$\left[ -\frac{\hbar^2}{2m} \frac{1}{a_0^2} \left( \frac{\partial^2}{\partial \tilde{x}^2} + \frac{\partial^2}{\partial \tilde{y}^2} + \frac{\partial^2}{\partial \tilde{z}^2} \right) - \frac{e^2}{4\pi\epsilon_0} \frac{1}{a_0 \tilde{r}} \right] \eta(\vec{\tilde{r}}) = E \eta(\vec{\tilde{r}}), \quad (16.5)$$

which seems like a nightmare, until you realize that

$$\frac{\hbar^2}{m} \frac{1}{a_0^2} = \frac{e^2}{4\pi\epsilon_0} \frac{1}{a_0} = \mathcal{E}_0.$$

The eigenproblem (16.5) is thus

$$\left[ -\frac{1}{2} \left( \frac{\partial^2}{\partial \tilde{x}^2} + \frac{\partial^2}{\partial \tilde{y}^2} + \frac{\partial^2}{\partial \tilde{z}^2} \right) - \frac{1}{\tilde{r}} \right] \eta(\vec{\tilde{r}}) = \frac{E}{\mathcal{E}_0} \eta(\vec{\tilde{r}}). \quad (16.6)$$

Defining the dimensionless “scaled energy”

$$\tilde{E} \equiv \frac{E}{\mathcal{E}_0}, \quad (16.7)$$

we see immediately that the energy eigenproblem, expressed in scaled variables, is

$$\left[ -\frac{1}{2} \left( \frac{\partial^2}{\partial \tilde{x}^2} + \frac{\partial^2}{\partial \tilde{y}^2} + \frac{\partial^2}{\partial \tilde{z}^2} \right) - \frac{1}{\tilde{r}} \right] \eta(\vec{\tilde{r}}) = \tilde{E} \eta(\vec{\tilde{r}}) \quad (16.8)$$

or

$$\left[ -\frac{1}{2} \tilde{\nabla}^2 - \frac{1}{\tilde{r}} \right] \eta(\vec{\tilde{r}}) = \tilde{E} \eta(\vec{\tilde{r}}). \quad (16.9)$$

Whoa! It’s considerably easier to work with the energy eigenproblem written in this form than it is to work with form (16.1) — there are no  $\hbar$ s and  $e^2/4\pi\epsilon_0$ s to keep track of (and to lose through algebra errors).

The only problem is that there are so many tildes to write down. People get tired of writing tildes, so they just omit them, with the understanding that they are now working with scaled variables rather than traditional variables, and the energy eigenproblem becomes

$$\left[ -\frac{1}{2} \nabla^2 - \frac{1}{r} \right] \eta(\vec{r}) = E \eta(\vec{r}). \quad (16.10)$$

I like to call this process “using scaled variables”. Others call it “measuring length and energy in atomic units”. Still others say that we get equation (16.10) from (16.1) by

$$\text{“setting } \hbar = m = \frac{e^2}{4\pi\epsilon_0} = 1\text{”}.$$

This last phrase is particularly opaque, because taken literally it's absurd. So you must not take it literally: it's a code phrase for the more interesting process of converting to scaled variables and then dropping the tildes.

One last point. Some people call this system not “atomic units” but “natural units”. While these units are indeed the natural system for solving problems in atomic physics, they are not the natural units for solving problems in nuclear physics, or in stellar physics, or in cosmology. And they are particularly unnatural and inappropriate for measuring the heights of towers.

## 16.2 Variational method for finding the ground state energy

Imagine a gymnasium full of fruits

$$\text{smallest fruit} \leq \text{smallest cantaloupe.}$$

Similarly

$$\text{ground state energy} \leq \langle \psi | \hat{H} | \psi \rangle \text{ for any } |\psi\rangle.$$

So try out a bunch of states, turn the crank, find the smallest. Very mechanical.

For example, to estimate the ground state energy of a quartic oscillator  $V(x) = \alpha x^4$ , you could use as trial wavefunctions the Gaussians

$$\psi(x) = \frac{1}{\sqrt[4]{\pi}\sqrt{\sigma}} e^{-x^2/2\sigma^2}.$$

Turn the crank to find  $\langle \psi | \hat{H} | \psi \rangle$ , then minimize to find which value of  $\sigma$  minimizes that expectation value.

Two things to remember: First, it's a mathematical technique useful in many fields, not just in quantum mechanics. Second, it seems merely mechanical, but in fact it relies on picking good trial wavefunctions: you have to gain an intuitive understanding of how the real wavefunction is going to behave, then pick trial wavefunctions which can mimic that behavior. In the words of Forman S. Acton (*Numerical Methods that Work*, 1970, page 252) “In the hands of a Feynman the [variational] technique works like a Latin charm; with ordinary mortals the result is a mixed bag.”

**Sample problem: Variational estimate for the ground state energy of a quartic oscillator**

The trial wavefunction

$$\psi(x) = \frac{1}{\sqrt[4]{\pi}\sqrt{\sigma}} e^{-x^2/2\sigma^2}$$

is normalized. (If you don't know this, you should verify it.) We look for

$$\begin{aligned}
 \langle \psi | \hat{H} | \psi \rangle &= \frac{1}{\sqrt{\pi}\sigma} \int_{-\infty}^{+\infty} e^{-x^2/2\sigma^2} \left[ -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + \alpha x^4 \right] e^{-x^2/2\sigma^2} dx \\
 &= -\frac{\hbar^2}{2m} \frac{1}{\sqrt{\pi}\sigma} \int_{-\infty}^{+\infty} e^{-x^2/2\sigma^2} \left( -\frac{1}{\sigma^2} \right) \left( 1 - \frac{x^2}{\sigma^2} \right) e^{-x^2/2\sigma^2} dx \\
 &\quad + \alpha \frac{1}{\sqrt{\pi}\sigma} \int_{-\infty}^{+\infty} e^{-x^2/2\sigma^2} x^4 e^{-x^2/2\sigma^2} dx \\
 &= \frac{\hbar^2}{2m} \frac{1}{\sqrt{\pi}\sigma^3} \int_{-\infty}^{+\infty} \left( 1 - \frac{x^2}{\sigma^2} \right) e^{-x^2/\sigma^2} dx \\
 &\quad + \alpha \frac{1}{\sqrt{\pi}\sigma} \int_{-\infty}^{+\infty} x^4 e^{-x^2/\sigma^2} dx \\
 &= \frac{\hbar^2}{2m} \frac{1}{\sqrt{\pi}\sigma^2} \int_{-\infty}^{+\infty} (1 - \tilde{x}^2) e^{-\tilde{x}^2} d\tilde{x} + \alpha \frac{\sigma^4}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \tilde{x}^4 e^{-\tilde{x}^2} d\tilde{x}.
 \end{aligned}$$

Already, even before evaluating the integrals, we can see that both integrals are numbers independent of the trial wavefunction width  $\sigma$ . Thus the expected kinetic energy, on the left, decreases with  $\sigma$  while the expected potential energy, on the right, increases with  $\sigma$ . Does this make sense to you?

When you work out (or look up) the integrals, you find

$$\langle \psi | \hat{H} | \psi \rangle = \frac{\hbar^2}{2m} \frac{1}{\sqrt{\pi}\sigma^2} (\sqrt{\pi} - \frac{1}{2}\sqrt{\pi}) + \alpha \frac{\sigma^4}{\sqrt{\pi}} \left( \frac{3}{4}\sqrt{\pi} \right) = \frac{\hbar^2}{2m} \frac{1}{2\sigma^2} + \alpha \frac{3\sigma^4}{4}.$$

If you minimize this energy with respect to  $\sigma$ , you will find that the minimum value (which is, hence, the best upper bound for the ground state energy) is

$$\left( 9 \frac{\hbar^2}{2m} \right)^{2/3} \frac{\alpha^{1/3}}{4}.$$

Problem: Show that the width of the minimum-energy wavefunction is

$$\sigma = \left( \frac{\hbar^2}{2m} \frac{1}{3\alpha} \right)^{1/6}.$$

Added: You can use the variational technique for other states as well: for example, in one dimensional systems, the first excited state is less than or equal to  $\langle \psi | \hat{H} | \psi \rangle$  for all states  $|\psi\rangle$  with a single node.

### 16.3 Problems

#### 16.1 Quantal recurrence in the infinite square well

- Find the period as a function of energy for a classical particle of mass  $m$  in an infinite square well of width  $L$ .
- Show that any wavefunction, regardless of energy, in the same infinite square well is periodic in time with a period

$$\frac{4mL^2}{\hbar\pi}.$$

(This part can be solved knowing only the energy eigenvalues.)

- What happens after one-half of this time has passed? (This part requires some knowledge of the energy eigenfunctions.)

[*Note:* This problem raises deep questions about the character of quantum mechanics and of its classical limit. See D.F. Styer, “Quantum revivals versus classical periodicity in the infinite square well,” *American Journal of Physics* **69** (January 2001) 56–62.]

#### 16.2 Quantal recurrence in the Coulomb problem

Show that in the Coulomb problem, any quantal state consisting of a superposition of two or more bound energy eigenstates with principal quantal numbers  $n_1, n_2, \dots, n_r$  evolves in time with a period of

$$\frac{h}{\text{Ry}} N^2,$$

where Ry is the Rydberg and the integer  $N$  is the least common multiple of  $n_1, n_2, \dots, n_r$ .

#### 16.3 Atomic units

The Schrödinger equation for the Coulomb problem is

$$i\hbar \frac{\partial \Psi(x, y, z, t)}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \Psi(x, y, z, t) - \frac{e^2}{4\pi\epsilon_0 r} \Psi(x, y, z, t).$$

It is clear that the answer to any physical problem can depend only on the three parameters  $\hbar$ ,  $m$ , and  $e^2/4\pi\epsilon_0$ . In section 16.1, we used these ideas to show that any problem that asked for a length had to have an answer which was a dimensionless number times the characteristic length, the so-called Bohr radius

$$a_0 = \frac{4\pi\epsilon_0 \hbar^2}{e^2 m}.$$

- Show that there is only one characteristic energy, i.e. only one way to combine the three parameters to produce a quantity with

the dimensions of energy. (Section 16.1 found *one* way to perform this combination, but I want you to prove that this is the *only* way. *Hint*: Instead of the conventional base dimensions of length, mass, and time, use the unconventional base dimensions of length, mass, and energy.)

- b. Find the characteristic time  $\tau_0$ . What is its numerical value in terms of femtoseconds?
- c. *Bonus*: Show that, in the Bohr model, the period of the innermost orbit is  $2\pi\tau_0$ . What is the period of the  $n$ th orbit?
- d. Estimate the number of heartbeats made in a lifetime by a typical person. If each Bohr model orbit corresponds to a heartbeat, how many “lifetimes of hydrogen” pass in a second?
- e. Write the time-dependent Schrödinger equation in terms of the scaled variables

$$\tilde{r} = \frac{r}{a_0} \quad \text{“lengths measured in atomic units”}$$

and

$$\tilde{t} = \frac{t}{\tau_0} \quad \text{“time measured in atomic units”}.$$

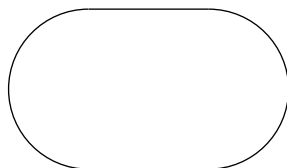
Be sure to use the dimensionless wavefunction

$$\tilde{\Psi}(\tilde{x}, \tilde{y}, \tilde{z}, \tilde{t}) = (a_0)^{3/2} \Psi(x, y, z, t).$$



#### 16.4 Scaling in the stadium problem

The “stadium” problem is often used as a model chaotic system, in both classical and quantum mechanics. [See E.J. Heller, “Bound-State Eigenfunctions of Classically Chaotic Hamiltonian Systems: Scars of Periodic Orbits” *Phys. Rev. Lett.*, **53**, 1515–1518 (1984); S. Tomsovic and E.J. Heller, “Long-Time Semiclassical Dynamics of Chaos: The Stadium Billiard” *Phys. Rev. E*, **47**, 282–299 (1993); E.J. Heller and S. Tomsovic, “Postmodern Quantum Mechanics” *Physics Today*, **46** (7), 38–46 (July 1993).] This is a two-dimensional infinite well shaped as a rectangle with semi-circular caps on opposite ends. Suppose one stadium has the same shape but is exactly three times as large as another. Show that in the larger stadium, wavepackets move just as they do in the smaller stadium, but nine times more slowly. (The initial wavepacket is of course also enlarged three times.) And show that the energy eigenvalues of the larger stadium are one-ninth the energy eigenvalues of the smaller stadium.



#### 16.5 Variational principle for the harmonic oscillator

Find the best bound on the ground state energy of the one-dimensional harmonic oscillator using a trial wavefunction of form

$$\psi(x) = \frac{A}{x^2 + b^2},$$

where  $A$  is determined through normalization and  $b$  is an adjustable parameter. *Hint:* Put the integrals within  $\langle H \rangle$  into dimensionless form so that they are independent of  $A$  and  $b$ , and are “just numbers”: call them  $C_K$  and  $C_P$ . Solve the problem in terms of these numbers, then evaluate the integrals only at the end.

#### 16.6 Solving the Coulomb problem through operator factorization

Griffiths (section 4.2) finds the bound state energy eigenvalues for the Coulomb problem using power series solutions of the Schrödinger equation. Here is another way, based on operator factorization (ladder

operators). In atomic units, the radial wave equation is

$$\left[ -\frac{1}{2} \frac{d^2}{dr^2} + \frac{\ell(\ell+1)}{2r^2} - \frac{1}{r} \right] u_{n,\ell}(r) \equiv h_\ell u_{n,\ell}(r) = \epsilon_{n,\ell} u_{n,\ell}(r)$$

where  $u_{n,\ell}(r)$  is  $r$  times the radial wavefunction. Introduce the operators

$$D_\pm^{(\ell)} \equiv \frac{d}{dr} \mp \frac{\ell}{r} \pm \frac{1}{\ell}.$$

a. Show that

$$D_+^{(\ell)} D_-^{(\ell)} = -2h_\ell - \frac{1}{\ell^2}.$$

and that

$$D_-^{(\ell+1)} D_+^{(\ell+1)} = -2h_\ell - \frac{1}{(\ell+1)^2}$$

b. Conclude that

$$h_{\ell+1} D_+^{(\ell+1)} = D_+^{(\ell+1)} h_\ell,$$

and apply this operator equation to  $u_{n,\ell}(r)$  to show that

$$D_+^{(\ell+1)} u_{n,\ell}(r) \propto u_{n,\ell+1}(r)$$

and that  $\epsilon_{n,\ell}$  is independent of  $\ell$ .

c. Argue that for every  $\epsilon_{n,\ell} < 0$  there is a maximum  $\ell$ . (Hint: Examine the effective potential for radial motion.) Call this  $\ell$  value  $\ell_n$ .

d. Define  $n = \ell_n + 1$  and show that

$$\epsilon_{n,\ell} = -\frac{1}{2n^2} \quad \text{where} \quad \ell = 0, \dots, n-1.$$

(One can also continue this game to find the energy eigenfunctions.)

## Chapter 17

# Hydrogen

Recall the structure of states summarized in section 14.4.

### 17.1 The Stark effect

The unperturbed Hamiltonian, as represented in the position basis, is

$$\hat{H}^{(0)} \doteq -\frac{\hbar^2}{2m}\nabla^2 - \frac{e^2}{4\pi\epsilon_0}\frac{1}{r}. \quad (17.1)$$

An electric field of magnitude  $E$  is applied, and we name the direction of the electric field the  $z$  direction. The perturbing Hamiltonian, again represented in the position basis, is

$$\hat{H}' \doteq eEz = eEr \cos\theta. \quad (17.2)$$

Perturbation theory for the energy eigenvalues tells us that, provided the unperturbed energy state  $|n^{(0)}\rangle$  is non-degenerate,

$$E_n = E_n^{(0)} + \langle n^{(0)}|\hat{H}'|n^{(0)}\rangle + \sum_{m \neq n} \frac{|\langle m^{(0)}|\hat{H}'|n^{(0)}\rangle|^2}{E_n^{(0)} - E_m^{(0)}} + \dots \quad (17.3)$$

Let us apply perturbation theory to the ground state  $|n, \ell, m\rangle = |1, 0, 0\rangle$ . This state is non-degenerate, so equation (17.3) applies without question. A moment's thought will convince you that  $\langle 1, 0, 0|\hat{H}'|1, 0, 0\rangle =$

$eE\langle 1, 0, 0 | \hat{z} | 1, 0, 0 \rangle = 0$ , so the result is

$$\begin{aligned} E_1 &= E_1^{(0)} + \sum_{n=2}^{\infty} \sum_{\ell=0}^{n-1} \sum_{m=-\ell}^{+\ell} \frac{|\langle n, \ell, m | \hat{H}' | 1, 0, 0 \rangle|^2}{E_1^{(0)} - E_n^{(0)}} + \dots \\ &= -Ry + \sum_{n=2}^{\infty} \sum_{\ell=0}^{n-1} \sum_{m=-\ell}^{+\ell} \frac{|eE\langle n, \ell, m | \hat{z} | 1, 0, 0 \rangle|^2}{-Ry + Ry/n^2} + \dots \\ &= -Ry - \frac{e^2 E^2}{Ry} \sum_{n=2}^{\infty} \sum_{\ell=0}^{n-1} \sum_{m=-\ell}^{+\ell} \frac{|\langle n, \ell, m | \hat{z} | 1, 0, 0 \rangle|^2}{1 - 1/n^2} + \dots \quad (17.4) \end{aligned}$$

It would take a lot of work to evaluate the sum here, but one thing is clear: that sum is just some quantity with the dimensions [length<sup>2</sup>], and independent of the field strength  $E$ . So when the electric field is turned on, the ground state energy decreases from the zero-field energy of  $-Ry$ , quadratically with  $E$ . Without even evaluating the sum, we get a lot of important information.

Well, that went well. What if we apply perturbation theory to the first excited state  $|2, 0, 0\rangle$ ? My first thought is that, once again  $\langle 2, 0, 0 | \hat{H}' | 2, 0, 0 \rangle = eE\langle 2, 0, 0 | \hat{z} | 2, 0, 0 \rangle = 0$ , so we'll need to go on to second-order perturbation theory, and hence we'll again find a quadratic Stark effect. The same argument holds for the excited state  $|2, 1, +1\rangle$ , the state  $|7, 5, -3\rangle$  and indeed for *any* energy state.

But that quick and easy argument is wrong. In making it we've forgotten that the equation 17.3 applies only to non-degenerate energy states.<sup>1</sup> The first excited state is four-fold degenerate: the states  $|2, 0, 0\rangle$ ,  $|2, 1, +1\rangle$ ,  $|2, 1, 0\rangle$ , and  $|2, 1, -1\rangle$  all have the same energy, namely  $-Ry/2^2$ . If we were to try to evaluate the sum, we'd have to look at terms like

$$\frac{|\langle 2, 1, 0 | \hat{H}' | 2, 0, 0 \rangle|^2}{E_{2,0,0} - E_{2,1,0}} = \frac{|\langle 2, 1, 0 | \hat{H}' | 2, 0, 0 \rangle|^2}{0},$$

which equals infinity! In our attempt to “get a lot of important information without actually evaluating the sum” we have missed the fact that the sum diverges.

There's only one escape from this trap. We can avoid infinities by making sure that, whenever we have a zero in the denominator, we also have a zero in the numerator. (Author's note to self: Change chapter 11

<sup>1</sup>This is a favorite trick question in physics oral exams.

to show this more rigorously.) That is, we can't perform the perturbation theory expansion using the basis

$$\{|2, 0, 0\rangle, |2, 1, +1\rangle, |2, 1, 0\rangle, |2, 1, -1\rangle\}$$

but we *can* perform it using some new basis, a linear combination of these states, such that in this new basis the matrix elements of  $\hat{H}'$  vanish *except on the diagonal*. In other words, we must diagonalize the  $4 \times 4$  matrix of  $\hat{H}'$ , and perform the perturbation expansion using that new basis rather than the initial basis.

The process, in other words, requires three stages: First find the matrix of  $\hat{H}'$ , then diagonalize it, and finally perform the expansion.

Start by finding the  $4 \times 4$  matrix in the initial basis. Each matrix element will have the form

$$\begin{aligned} \langle a | \hat{H}' | b \rangle &= eE \langle a | \hat{z} | b \rangle \\ &= eE \int_0^{2\pi} d\phi \int_0^\pi \sin \theta d\theta \int_0^\infty r^2 dr \eta_a^*(r, \theta, \phi) r \cos \theta \eta_b(r, \theta, \phi) \end{aligned} \quad (17.5)$$

and they will be arrayed in a matrix like this:

$$\begin{array}{cccc} \langle 200 | & \langle 211 | & \langle 210 | & \langle 21\bar{1} | \\ \left[ \begin{array}{cccc} & & & \\ & & & \\ & & & \\ & & & \end{array} \right] & \begin{array}{l} |200\rangle \\ |211\rangle \\ |210\rangle \\ |21\bar{1}\rangle \end{array} \end{array}$$

(Here the  $m$  value of  $-1$  is shown as  $\bar{1}$  because otherwise it messes up the spacing.)

You might think that there are 16 matrix elements to calculate, that each one is a triple integral, and that the best way to start off is by going to a bar and getting drunk. Courage! The operator is Hermitian, so the subdiagonal elements are the complex conjugates of the corresponding superdiagonal elements — there are only 10 matrix elements to calculate.

The diagonal elements are all proportional to the expectation values of  $\hat{z}$ , and these expectation values vanish for *any* of the traditional Coulomb problem eigenstates  $|n, \ell, m\rangle$ .

$$\begin{array}{cccc} \langle 200 | & \langle 211 | & \langle 210 | & \langle 21\bar{1} | \\ \left[ \begin{array}{cccc} 0 & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{array} \right] & \begin{array}{l} |200\rangle \\ |211\rangle \\ |210\rangle \\ |21\bar{1}\rangle \end{array} \end{array}$$

Remember what the wavefunctions look like:

$$\begin{aligned} |2, 0, 0\rangle &\doteq R_{2,0}(r)Y_0^0(\theta, \phi) \sim 1 \\ |2, 1, +1\rangle &\doteq R_{2,1}(r)Y_1^{+1}(\theta, \phi) \sim \sin\theta e^{+i\phi} \\ |2, 1, 0\rangle &\doteq R_{2,1}(r)Y_1^0(\theta, \phi) \sim \cos\theta \\ |2, 1, -1\rangle &\doteq R_{2,1}(r)Y_1^{-1}(\theta, \phi) \sim \sin\theta e^{-i\phi} \end{aligned}$$

where  $\sim$  means that I've written down the angular dependence but not the radial dependence.

The leftmost matrix element on the top row is

$$\begin{aligned} &\langle 2, 1, +1 | \hat{H}' | 2, 0, 0 \rangle \\ &= eE \int_0^{2\pi} d\phi \int_0^\pi \sin\theta d\theta \int_0^\infty r^2 dr R_{2,1}(r)Y_1^{+1*}(\theta, \phi) r \cos\theta R_{2,0}(r)Y_0^0(\theta, \phi). \end{aligned}$$

There are three integrals here:  $r$ ,  $\theta$ , and  $\phi$ . To do the  $r$  integral I would have to look up the expressions for  $R_{2,1}(r)$  and  $R_{2,0}(r)$ , and then do a gnarly integral. To do the  $\theta$  integral I would have to look up the spherical harmonics and then do an integral not quite so gnarly as the  $r$  integral. But to do the  $\phi$  integral is straightforward: The function  $Y_1^{+1*}(\theta, \phi)$  contributes an  $e^{-i\phi}$  and that's it. The  $\phi$  integral is

$$\int_0^{2\pi} d\phi e^{-i\phi}$$

and this integral is easy to do... it's zero.

$$\begin{array}{cccc|c} \langle 200 | & \langle 211 | & \langle 210 | & \langle 21\bar{1} | & \\ \left[ \begin{array}{cccc} 0 & 0 & & \\ 0 & 0 & & \\ & & 0 & \\ & & & 0 \end{array} \right] & \begin{array}{l} |200\rangle \\ |211\rangle \\ |210\rangle \\ |21\bar{1}\rangle \end{array} \end{array}$$

It's a good thing we put off doing the difficult  $r$  and  $\theta$  integrals, because if we had sweated away working them out, and then found that all we did with those hard-won results was to multiply them by zero, then we'd really need to visit that bar. When I was a child, my Protestant-work-ethic parents told me that when faced with two tasks, I should always "be a man" and do the difficult one first. I'm telling you to do the opposite, because doing the easy task might make you realize that you don't have to do the difficult one.

If you look at the two other matrix elements on the superdiagonal,

$$\langle 2, 1, 0 | \hat{H}' | 2, 1, +1 \rangle \quad \text{and} \quad \langle 2, 1, -1 | \hat{H}' | 2, 1, 0 \rangle,$$

you'll recognize instantly that for each of these two the  $\phi$  integral is

$$\int_0^{2\pi} d\phi e^{+i\phi} = 0.$$

The same holds for  $\langle 2, 1, -1 | \hat{H}' | 2, 0, 0 \rangle$ , so the matrix is shaping up as

$$\begin{array}{cccc|c} \langle 200 | & \langle 211 | & \langle 210 | & \langle 21\bar{1} | & \\ \left[ \begin{array}{cccc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 \\ 0 & & 0 & 0 \end{array} \right] & \begin{array}{l} |200\rangle \\ |211\rangle \\ |210\rangle \\ |21\bar{1}\rangle \end{array} \end{array}$$

and we have just two more elements to calculate.

The matrix element

$$\langle 2, 1, -1 | \hat{H}' | 2, 1, 1 \rangle \sim \int_0^{2\pi} d\phi e^{+2i\phi} = 0,$$

so the only hard integral we have to do is

$$\langle 2, 1, 0 | \hat{H}' | 2, 0, 0 \rangle = eE \langle 2, 1, 0 | \hat{z} | 2, 0, 0 \rangle.$$

The matrix element  $\langle 2, 1, 0 | \hat{z} | 2, 0, 0 \rangle$  is a length, and any length for the Coulomb problem must turn out to be a dimensionless number times the Bohr radius

$$\langle 2, 1, 0 | \hat{H}' | 2, 0, 0 \rangle = eE \langle 2, 1, 0 | \hat{z} | 2, 0, 0 \rangle = eE(\text{number})a_0. \quad (17.6)$$

The only thing that remains to do is to find that dimensionless number. I ask you to do this yourself in problem 17.1 (part a). The answer is  $-3$ . Thus the matrix is

$$-eEa_0 \begin{array}{cccc|c} \langle 200 | & \langle 211 | & \langle 210 | & \langle 21\bar{1} | & \\ \left[ \begin{array}{cccc} 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right] & \begin{array}{l} |200\rangle \\ |211\rangle \\ |210\rangle \\ |21\bar{1}\rangle \end{array} \end{array}$$

and we are done with the first stage of our three-stage problem.

You will be tempted to rush immediately into the problem of diagonalizing this matrix, but “fools rush in where angels fear to tread” (Alexander Pope). If you think about it for an instant, you'll realize that it will be a

lot easier to do the problem if we rearrange the sequence of basis vectors so that the matrix reads

$$-eEa_0 \begin{array}{cccc|c} \langle 200| & \langle 210| & \langle 211| & \langle 21\bar{1}| & \\ \hline 0 & 3 & 0 & 0 & |200\rangle \\ 3 & 0 & 0 & 0 & |210\rangle \\ 0 & 0 & 0 & 0 & |211\rangle \\ 0 & 0 & 0 & 0 & |21\bar{1}\rangle \end{array}$$

Now we start the second stage, diagonalizing the matrix. First, find the eigenvalues:

$$\begin{aligned} 0 &= \det |\mathbf{M} - \lambda\mathbf{I}| \\ &= \det \begin{vmatrix} -\lambda & 3 & 0 & 0 \\ 3 & -\lambda & 0 & 0 \\ 0 & 0 & -\lambda & 0 \\ 0 & 0 & 0 & -\lambda \end{vmatrix} \\ &= -\lambda \det \begin{vmatrix} -\lambda & 0 & 0 \\ 0 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{vmatrix} - 3 \det \begin{vmatrix} 3 & 0 & 0 \\ 0 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{vmatrix} \\ &= \lambda^4 - 3^2 \lambda^2 \\ &= \lambda^2(\lambda^2 - 3^2) \end{aligned}$$

Normally, it's hard to solve a quartic equation, but in this case we can just read off the four solutions:

$$\lambda = +3, -3, 0, 0.$$

The eigenvectors associated with  $\lambda = 0$  and  $\lambda = 0$  are clearly

$$|2, 1, +1\rangle \quad \text{and} \quad |2, 1, -1\rangle.$$

The eigenvector associated with  $\lambda = 3$  will be a linear combination

$$x|2, 0, 0\rangle + y|2, 1, 0\rangle$$

where

$$\begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 3 \begin{pmatrix} x \\ y \end{pmatrix}.$$

Any  $x = y$  is a solution, but I choose the normalized solution so that the eigenvector with eigenvalue 3 is

$$\frac{1}{\sqrt{2}} (|2, 0, 0\rangle + |2, 1, 0\rangle).$$



The parallel process for  $\lambda = -3$  reveals the eigenvector

$$\frac{1}{\sqrt{2}} (-|2, 0, 0\rangle + |2, 1, 0\rangle).$$

[[Why, you will ask, do I use this eigenvector rather than

$$\frac{1}{\sqrt{2}} (|2, 0, 0\rangle - |2, 1, 0\rangle),$$

which is also an eigenvector but which I can write down with fewer pen strokes? The answer is simple personal preference. The version I use is the same one used for geometrical vectors in a plane, and where the change of basis is a  $45^\circ$  rotation. This helps me remember that, even in this recondite and abstruse situation, the process of matrix diagonalization does not change the physical situation, it merely changes the basis vectors we select to help us describe the physical situation.]]

To summarize, in the basis

$$\left\{ \frac{1}{\sqrt{2}} (|2, 0, 0\rangle + |2, 1, 0\rangle), \frac{1}{\sqrt{2}} (-|2, 0, 0\rangle + |2, 1, 0\rangle), |2, 1, +1\rangle, |2, 1, -1\rangle \right\}$$

the matrix representation of the operator  $\hat{H}'$  is

$$-eEa_0 \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

And now, for the final stage, executing perturbation theory starting from this new basis, which I'll call  $\{|a\rangle, |b\rangle, |c\rangle, |d\rangle\}$ . The energy value associated with  $|a\rangle$  is

$$E_2 = E_2^{(0)} + \langle a|\hat{H}'|a\rangle + \sum_m \frac{|\langle a|\hat{H}'|m\rangle|^2}{E_a^{(0)} - E_m^{(0)}} + \dots$$

The first correction we already know: it is  $\langle a|\hat{H}'|a\rangle = -3eEa_0$ . The second correction — the sum — contains terms like

$$\frac{|\langle a|\hat{H}'|b\rangle|^2}{E_a^{(0)} - E_b^{(0)}} = \frac{0}{0}$$

and

$$\frac{|\langle a|\hat{H}'|c\rangle|^2}{E_a^{(0)} - E_c^{(0)}} = \frac{0}{0}$$

and

$$\frac{|\langle a|\hat{H}'|1, 0, 0\rangle|^2}{E_a^{(0)} - E_{1,0,0}^{(0)}} = \frac{\text{something}}{-\frac{3}{4}\text{Ry}}$$

but it contains *no* terms where a number is divided by zero. I will follow the usual rule-of-thumb for perturbation theory, which is to stop at the first non-zero correction and ignore the sum altogether.

Similarly, the leading energy correction associated with  $|b\rangle$  is  $\langle b|\hat{H}'|b\rangle = 3eEa_0$ .

The first-order corrections for  $|c\rangle$  and  $|d\rangle$  vanish, so these states will be subject to a quadratic Stark effect, just like the ground state. I could work them out if I really needed to, but instead I will quote and follow the age-old dictum (modified from “The Lay of the Last Minstrel” by Walter Scott):

Breathes there the man, with soul so dead,  
Who never to himself hath said  
“To hell with it, I’m going to bed.”

### 17.1 The Stark effect

- Find the numerical factor in equation (17.6).
- The “good” energy eigenstates for the  $n = 2$  Stark effect — the states that one should use as unperturbed states in perturbation theory — are

$$\begin{aligned} &|2, 1, +1\rangle \\ &|2, 1, -1\rangle \\ &\frac{1}{\sqrt{2}}(|2, 0, 0\rangle + |2, 1, 0\rangle) \\ &\frac{1}{\sqrt{2}}(-|2, 0, 0\rangle + |2, 1, 0\rangle) \end{aligned}$$

Find the expectation value for position  $\langle \vec{r} \rangle$  in each of these states.

- The expectation value of position is zero in state  $|2, 0, 0\rangle$  and zero in state  $|2, 1, 0\rangle$ , yet it is non-zero in state  $(|2, 0, 0\rangle + |2, 1, 0\rangle)/\sqrt{2}$ . This might seem like a contradiction: After all, if the average position vanishes for two probability densities, then it vanishes for the sum of the two. What great principle of quantum mechanics allows this fact to escape the curse of contradiction? (Answer in one sentence.)
- (**Bonus.**) Describe these four states qualitatively and explain why they are the “good” states for use in the Stark effect.

- e. Consider the Stark effect for the  $n = 3$  states of hydrogen. There are initially nine degenerate states. Construct a  $9 \times 9$  matrix representing the perturbing Hamiltonian. (*Hint:* Before actually working any integrals, use a selection rule to determine the sequence of basis elements that will produce a block diagonal matrix.)
- f. Find the eigenvalues and degeneracies.

### 17.2 Bonus

In the previous problem, on the Stark effect, we had to calculate a lot of matrix elements of the form

$$\int_0^\infty r^2 R_{n,\ell}(r) r R_{n',\ell'}(r) dr.$$

This was possible but (to put it mildly) tedious. Can you think of some easy way to do integrals of this form? Could the operator factorization technique (problem 16.6) give us any assistance? Can you derive any inspiration from our proof of Kramers' relation (problem below)?

### 17.3 Kramers' relation

Kramers' relation states that for any energy eigenstate  $\eta_{n\ell m}(\vec{r})$  of the Coulomb problem, the expected values of  $r^s$ ,  $r^{s-1}$ , and  $r^{s-2}$  are related through

$$\frac{s+1}{n^2} \langle r^s \rangle - (2s+1) a_0 \langle r^{s-1} \rangle + \frac{s}{4} [(2\ell+1)^2 - s^2] a_0^2 \langle r^{s-2} \rangle = 0.$$

- a. Prove Kramers' relation. *Hints:* Use atomic units. Start with the radial equation in form

$$u''(r) = \left[ \frac{\ell(\ell+1)}{r^2} - \frac{2}{r} + \frac{1}{n^2} \right] u(r),$$

and use it to express

$$\int_0^\infty u(r) r^s u''(r) dr$$

in terms of  $\langle r^s \rangle$ ,  $\langle r^{s-1} \rangle$ , and  $\langle r^{s-2} \rangle$ . Then perform that integral by parts to find an integral involving  $u'(r)$  as the highest derivative. Show that

$$\int_0^\infty u(r) r^s u'(r) dr = -\frac{s}{2} \langle r^{s-1} \rangle$$

and that

$$\int_0^\infty u'(r) r^s u'(r) dr = -\frac{2}{s+1} \int_0^\infty u''(r) r^{s+1} u'(r) dr.$$

- b. Use Kramers' relation with  $s = 0$ ,  $s = 1$ ,  $s = 2$ , and  $s = 3$  to find formulas for  $\langle r^{-1} \rangle$ ,  $\langle r \rangle$ ,  $\langle r^2 \rangle$ , and  $\langle r^3 \rangle$ . Note that you could continue indefinitely to find  $\langle r^s \rangle$  for any positive power.
- c. However, you can't use this chain to work downward. Try it for  $s = -1$ , and show that you get a relation between  $\langle r^{-2} \rangle$  and  $\langle r^{-3} \rangle$ , but not either quantity by itself.

## Chapter 18

# Helium

The helium problem is a “three-body problem”. This problem has never been solved exactly even in classical mechanics, and it is hopeless to expect an exact solution in the richer and more intricate regime of quantum mechanics. Does this mean we should give up? Of course not. Most physics problems cannot be solved exactly, but some can be solved approximately well enough to compare theory to experiment, which is itself imperfect. (In the same way, most problems you have with your parents, or with your boy/girlfriend, cannot be solved perfectly. But they can often be solved well enough to continue your relationship.)

### 18.1 Ground state energy of helium

#### *The role of theory*

Jacov Ilich Frenkel (also Yakov Ilich Frenkel or Iakov Ilich Frenkel; 1894–1952) was a prolific physicist. Among other things he coined the term “phonon”. In a review article on the theory of metals (quoted by M.E. Fisher in “The Nature of Critical Points”, Boulder lectures, 1965) he said:

The more complicated the system considered, the more simplified must its theoretical description be. One cannot demand that a theoretical description of a complicated atom, and all the more of a molecule or a crystal, have the same degree of accuracy as of the theory of the simplest hydrogen atom. Incidentally, such a requirement is not only impossible to fulfill but also essentially useless. . . . An exact calculation of the constants characterizing the simplest physical system has essential significance as a test on the correctness of the basic principles of the theory.

However, once it passes this test brilliantly there is no sense in subjecting it to further tests as applied to more complicated systems. The most ideal theory cannot pass such tests, owing to the practically unsurmountable mathematical difficulties unavoidably encountered in applications to complicated systems. In this case all that is demanded of the theory is a correct interpretation of the general character of the quantities and laws pertaining to such a system. The theoretical physicist is in this respect like a cartoonist, who must depict the original, not in all details like a photographic camera, but simplify and schematize it in a way as to disclose and emphasize the most characteristic features. Photographic accuracy can and should be required only of the description of the simplest system. A good theory of complicated systems should represent only a good “caricature” of these systems, exaggerating the properties that are most difficult, and purposely ignoring all the remaining inessential properties.

Which case is the ground state of He?

- 1) Fundamental test of symmetrization postulate.
- 2) Test to see whether QM breaks down for complex systems (Anthony J. Leggett).
- 3) Refinements can involve new physical ideas.
- 4) Physical effects other than ground state energy.

### ***Experiment***

$$E_g = -78.975 \text{ eV.}$$

### ***Theory***

(Summarizing Griffiths 5.2.1 and 7.2.) If we take account of the Coulomb forces, but ignore things like the finite size of the nucleus, nuclear motion, relativistic motion of the electron, spin-orbit effects, and so forth, the Hamiltonian for two electrons and one nucleus is

$$\hat{H} = \hat{H}_A + \hat{H}_B + \hat{U}_{AB} \quad (18.1)$$

where

$$\hat{U}_{AB} = \frac{e^2}{4\pi\epsilon_0} \frac{1}{|\mathbf{r}_A - \mathbf{r}_B|}. \quad (18.2)$$

The ground state wavefunction for H is

$$\eta_{100}(\mathbf{r}) = \frac{1}{\sqrt{\pi}a_0^{3/2}}e^{-r/a_0}. \quad (18.3)$$

But if the nucleus had charge  $+Ze$ , this would be

$$\eta_{100}(\mathbf{r}) = \frac{Z^{3/2}}{\sqrt{\pi}a_0^{3/2}}e^{-Zr/a_0}. \quad (18.4)$$

So the  $\hat{U}_{AB} = 0$  ground state is

$$\eta_{100}(\mathbf{r}_A)\eta_{100}(\mathbf{r}_B) = \frac{Z^3}{\pi a_0^3}e^{-Z(r_A+r_B)/a_0} \quad \text{with} \quad Z = 2. \quad (18.5)$$

This state gives a ground state energy of  $E_g = -8(\text{Ry}) = -109 \text{ eV}$ .

Turning on the electron-electron repulsion, perturbation theory finds  $\langle \hat{U}_{AB} \rangle$  and jacks up  $E_g$  to  $-75 \text{ eV}$ .

The variational method uses the same wavefunction as above, but considers  $Z$  not as 2 but as an adjustable parameter. Interpretation: “shielding” — expect  $1 < Z_{\min} < 2$ . And in fact minimizing  $\langle H \rangle$  with over this class of trial wavefunctions gives  $Z_{\min} = 1.69$  and  $E_g = -77.5 \text{ eV}$ . (Sure enough, an overestimate.) Griffiths stops here and suggests that the rest of the work is humdrum.

### **Further theory**

Review: A. Hibbert, *Rept. Prog. Phys.* **38** (1975) 1222–1225.

**Hylleraas (1929):** Trial wavefunction of form (atomic units)

$$\psi(\mathbf{r}_A, \mathbf{r}_B) = e^{-Z(r_A+r_B)} \sum c_{nlm} (Z(r_A+r_B))^n (Z(r_A-r_B))^{2l} (Z|\mathbf{r}_A-\mathbf{r}_B|)^m.$$

[I won’t go into all the reasons why he picked this trial wavefunction, but...ask why only even powers  $2l$ .] Using  $Z$  and six terms in sum as variational parameters, he got an energy good to 2 parts in 10,000.

This is a good energy. Is there any point in doing better? Yes. Although it gives you a good energy, it gives you a poor wavefunction: Think of a  $d = 2$  landscape with a hidden valley — e.g. a crater, an absolute minimum. The  $d = 2$  landscape represents two variational parameters — by coincidence, the exact wavefunction has the form that you guessed. If you tried just one variational parameter, you’d be walking a line in this landscape. The line could be quite far from the valley bottom while giving

very good elevation estimates for the valley bottom, because the valley is flat at the bottom. [Sketch.]

In fact, you can show that *no* wavefunction of this form, no matter how many terms you pick, can satisfy the Schrödinger Equation — even if you picked an infinite number of terms, you’d never hit the wavefunction right on!

Is there any reason to get the wavefunction right? Yes! For example if you wanted to calculate Stark or Zeeman effect, or spin-orbit, or whatever, you’d need those wavefunctions for doing perturbation theory!

**Kinoshita (1959):** One of the “great fiddlers of physics”. Trial wavefunction of form (atomic units)

$$\psi(\mathbf{r}_A, \mathbf{r}_B) = e^{-Z(r_A+r_B)} \sum c_{nlm} (Z(r_A+r_B))^n \left( Z \frac{r_A - r_B}{|\mathbf{r}_A - \mathbf{r}_B|} \right)^{2l} \left( \frac{|\mathbf{r}_A - \mathbf{r}_B|}{r_A + r_B} \right)^m.$$

He showed that this could satisfy the Schrödinger Equation exactly if sum were infinite. Used 80 terms for accuracy 1 part in 100,000.

**Pekeris (1962):** A different trial wavefunction guaranteed to get the correct form when both electrons are far from nucleus. Used 1078 terms, added fine structure and hyperfine structure, got accuracy 1 part in  $10^9$ .

**Schwartz (1962):** Added terms like  $[Z(r_A + r_B)]^{n/2}$  ... not smooth. Got better energies with 189 terms!

**Frankowski and Pekeris (1966):** Introduced terms like  $\ln^k(Z(r_A + r_B))$  ... not smooth. 246 terms, accuracy 1 part in  $10^{12}$ .

**Kato:** (See Drake, page 155.) Looked at condition for two electrons close, both far from nucleus. In this case it’s like H atom, wavefunction must have cusp. Allow electrons to show this cusp.

State of art: Gordon W.F. Drake, ed. *Atomic, Molecular, and Optical Physics Handbook* page 163. [Reference QC173.A827 1996]

New frontiers: experiment. S.D. Bergeson, *et al.*, “Measurement of the He ground state Lamb shift”, *Phys. Rev. Lett.* **80** (1998) 3475–3478.

New frontiers: theory. S.P. Goldman, “Uncoupling correlated calculations in atomic physics: Very high accuracy and ease,” *Phys. Rev. A* **57** (1998) 677–680. 8066 terms, 1 part in  $10^{18}$ .

New frontiers: Lithium, metallic Hydrogen.



Sometimes people get the impression that variational calculations are dry and mechanical: simply add more parameters to your trial wavefunction, and your results will improve (or at least, they can't get worse). The history of the Helium ground state calculation shows how wrong this impression is. Progress is made by deep thinking about the character of the true wavefunction (What is the character when both electrons are far from the nucleus and far from each other? What is the character when both electrons are far from the nucleus and close to each other?) and then choosing trial wavefunctions that can display (or at least mimic) those characteristics of the true wavefunction.



## Chapter 19

# Atoms

### 19.1 Addition of angular momenta

We often have occasion to add angular momenta. For example, an electron might have orbital angular momentum with respect to the nucleus, but also spin angular momentum. What is the total angular momentum?

Or again, there might be two electrons in an atom, each with orbital angular momentum. What is the total orbital angular momentum of the two electrons?

Or again, there might be an electron with orbital angular momentum relative to the nucleus, but the nucleus moves relative to some origin. What is the total angular momentum of the electron relative to the origin?

This section demonstrates how to perform such additions through a specific example, namely adding angular momentum A with  $\ell_A = 1$  to angular momentum B with  $\ell_B = 2$ . (For the moment, assume that these angular momenta belong to non-identical particles. If the two particles are identical — as in the second example above — then there is an additional requirement that the sum wavefunction be symmetric or antisymmetric under swapping/interchange/exchange.)

First, recall the states for a single angular momentum: There are no states with values of  $\hat{L}_x$ ,  $\hat{L}_y$ ,  $\hat{L}_z$ , and  $\hat{L}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2$  simultaneously, reflecting such facts as that  $\hat{L}_x$  and  $\hat{L}_z$  do not commute. However, because  $\hat{L}^2$  and  $\hat{L}_z$  *do* commute, there *are* states (in fact, a basis of states) that have values of  $\hat{L}^2$  and  $\hat{L}_z$  simultaneously.

For angular momentum A, with  $\ell_A = 1$ , these basis states are

$$\begin{aligned} &|1, +1\rangle \\ &|1, 0\rangle \\ &|1, -1\rangle \end{aligned}$$

where

$$\hat{L}_A^2|\ell_A, m_A\rangle = \hbar^2\ell_A(\ell_A + 1)|\ell_A, m_A\rangle = \hbar^2(1)(2)|\ell_A, m_A\rangle$$

and

$$\hat{L}_{A,z}|\ell_A, m_A\rangle = \hbar m_A|\ell_A, m_A\rangle.$$

These states are called the “ $\ell_A = 1$  triplet”.

For angular momentum B, with  $\ell_B = 2$ , these basis states are

$$\begin{aligned} &|2, +2\rangle \\ &|2, +1\rangle \\ &|2, 0\rangle \\ &|2, -1\rangle \\ &|2, -2\rangle \end{aligned}$$

where

$$\hat{L}_B^2|\ell_B, m_B\rangle = \hbar^2\ell_B(\ell_B + 1)|\ell_B, m_B\rangle = \hbar^2(2)(3)|\ell_B, m_B\rangle$$

and

$$\hat{L}_{B,z}|\ell_B, m_B\rangle = \hbar m_B|\ell_B, m_B\rangle.$$

These states are called the “ $\ell_B = 2$  quintet”.

Now, what sort of states can we have for the sum of these two angular momenta? The relevant total angular momentum operator is

$$\hat{J} = \hat{L}_A + \hat{L}_B$$

so

$$\hat{J}_z = \hat{L}_{A,z} + \hat{L}_{B,z}$$

but

$$\hat{J}^2 \neq \hat{L}_A^2 + \hat{L}_B^2.$$

We can ask for states with values of  $\hat{J}^2$  and  $\hat{J}_z$  simultaneously, but such states will *not* necessarily have values of  $\hat{L}_{A,z}$  and  $\hat{L}_{B,z}$ , because  $\hat{J}^2$  and  $\hat{L}_{A,z}$  do not commute (see problem XXX). For the same reason, we can ask

for states with values of  $\hat{L}_{A,z}$  and  $\hat{L}_{B,z}$  simultaneously, but such states will *not* necessarily have values of  $\hat{J}^2$ .

For most problems, there are two bases that are natural and useful. The first consists of states like  $|\ell_A, m_A\rangle|\ell_B, m_B\rangle$  — simple product states of the bases we discussed above. The second basis consists of states like  $|j, m_J\rangle$ . To find how these are connected, we list states in the first basis according to their associated<sup>1</sup> value of  $m_J$ :

$ \ell_A, m_A\rangle \ell_B, m_B\rangle$	$m_J$
$ 1, +1\rangle 2, +2\rangle$	+3
$ 1, +1\rangle 2, +1\rangle$ $ 1, 0\rangle 2, +2\rangle$	+2   +2
$ 1, +1\rangle 2, 0\rangle$ $ 1, 0\rangle 2, +1\rangle$ $ 1, -1\rangle 2, +2\rangle$	+1   +1   +1
$ 1, +1\rangle 2, -1\rangle$ $ 1, 0\rangle 2, 0\rangle$ $ 1, -1\rangle 2, +1\rangle$	0   0   0
$ 1, +1\rangle 2, -2\rangle$ $ 1, 0\rangle 2, -1\rangle$ $ 1, -1\rangle 2, 0\rangle$	-1   -1   -1
$ 1, 0\rangle 2, -2\rangle$ $ 1, -1\rangle 2, -1\rangle$	-2   -2
$ 1, -1\rangle 2, -2\rangle$	-3

These values of  $m_J$  fall into a natural structure:

There is a heptet of seven states with  $m_J = +3, +2, +1, 0, -1, -2, -3$ . This heptet must be associated with  $j = 3$ .

There is a quintet of five states with  $m_J = +2, +1, 0, -1, -2$ . This quintet must be associated with  $j = 2$ .

There is a triplet of three states with  $m_J = +1, 0, -1$ . This triplet must be associated with  $j = 1$ .

So now we know what the values of  $j$  are! If you think about this problem for general values of  $\ell_A$  and  $\ell_B$ , you will see immediately that the values of  $j$  run from  $\ell_A + \ell_B$  to  $|\ell_A - \ell_B|$ . Often, this is all that's needed.<sup>2</sup> But sometimes you need more. Sometimes you need to express total-angular-momentum states like  $|j, m_J\rangle$  in terms of individual-angular-momentum states like  $|\ell_A, m_A\rangle|\ell_B, m_B\rangle$ .

The basic set-up of our problem comes through the table below:

<sup>1</sup>While the state  $|\ell_A, m_A\rangle|\ell_B, m_B\rangle$  doesn't have a value of  $j$ , it does have a value of  $m_J$ , namely  $m_J = m_A + m_B$ .

<sup>2</sup>In particular, many GRE questions that appear on their face to be deep and difficult only go this far.

$ \ell_A, m_A\rangle \ell_B, m_B\rangle$	$ j, m_J\rangle$
$ 1, +1\rangle_A 2, +2\rangle_B$	$ 3, +3\rangle_J$
$ 1, +1\rangle_A 2, +1\rangle_B$ $ 1, 0\rangle_A 2, +2\rangle_B$	$ 3, +2\rangle_J$ $ 2, +2\rangle_J$
$ 1, +1\rangle_A 2, 0\rangle_B$ $ 1, 0\rangle_A 2, +1\rangle_B$ $ 1, -1\rangle_A 2, +2\rangle_B$	$ 3, +1\rangle_J$ $ 2, +1\rangle_J$ $ 1, +1\rangle_J$
$ 1, +1\rangle_A 2, -1\rangle_B$ $ 1, 0\rangle_A 2, 0\rangle_B$ $ 1, -1\rangle_A 2, +1\rangle_B$	$ 3, 0\rangle_J$ $ 2, 0\rangle_J$ $ 1, 0\rangle_J$
$ 1, +1\rangle_A 2, -2\rangle_B$ $ 1, 0\rangle_A 2, -1\rangle_B$ $ 1, -1\rangle_A 2, 0\rangle_B$	$ 3, -1\rangle_J$ $ 2, -1\rangle_J$ $ 1, -1\rangle_J$
	$ 3, -2\rangle_J$ $ 2, -2\rangle_J$
	$ 3, -3\rangle_J$

Note that we have labeled states like

$$|\ell_A, m_A\rangle|\ell_B, m_B\rangle \text{ as } |\ell_A, m_A\rangle_A|\ell_B, m_B\rangle_B$$

and states like

$$|j, m_J\rangle \text{ as } |j, m_J\rangle_J.$$

Otherwise we might confuse the state  $|2, +1\rangle_B$  on the left side of the second row with the completely different state  $|2, +1\rangle_J$  on the right side of the of the third row. (Some authors solve this notation vexation by writing the states of total angular momentum as  $|j, m_J, \ell_A, \ell_B\rangle$ , taking advantage of the fact that  $\ell_A$  and  $\ell_B$  are the same for all states on the right — and for all states on the left, for that matter. This means every state on the right would be written as  $|j, m_J, 1, 2\rangle$ . For me, it rapidly grows frustrating to tack a “1,2” on to the end of every such state.)

The second line of this table means that the state  $|3, +2\rangle_J$  is some linear combination of the states  $|1, +1\rangle_A|2, +1\rangle_B$  and  $|1, 0\rangle_A|2, +2\rangle_B$ . Similarly for the state  $|2, +2\rangle_J$ . [This is the *meaning* of the assertion made earlier that in the state  $|3, +2\rangle_J$  there is no value for  $m_A$ : The state  $|3, +2\rangle_J$  is a superposition of a state with  $m_A = +1$  and a state with  $m_A = 0$ , but the state  $|3, +2\rangle_J$  itself has no value for  $m_A$ .] Similarly, the state  $|1, +1\rangle_A|2, +1\rangle_B$  is a linear combination of states  $|3, +2\rangle_J$  and  $|2, +2\rangle_J$ .

But what linear combination? We start with the first line of the table. Because there’s only one state on each side, we write

$$|3, +3\rangle_J = |1, +1\rangle_A|2, +2\rangle_B. \quad (19.1)$$

(We could have inserted an overall phase factor of modulus one, such as  $|3, +3\rangle_J = -|1, +1\rangle_A|2, +2\rangle_B$  or  $|3, +3\rangle_J = i|1, +1\rangle_A|2, +2\rangle_B$  or even  $|3, +3\rangle_J = -\sqrt{i}|1, +1\rangle_A|2, +2\rangle_B$ . But this insertion would have only made our lives difficult for no reason.)

Now, to find an expression for  $|3, +2\rangle$ , apply the lowering operator

$$\hat{J}_- = \hat{L}_{A,-} + \hat{L}_{B,-}$$

to both sides of equation (19.1). Remembering that

$$\hat{J}_- |j, m\rangle = \hbar\sqrt{j(j+1) - m(m-1)} |j, m-1\rangle,$$

this lowering gives

$$\begin{aligned} \hat{J}_- |3, +3\rangle_J &= \left[ \hat{L}_{A,-} |1, +1\rangle_A \right] |2, +2\rangle_B \\ &\quad + |1, +1\rangle_A \left[ \hat{L}_{B,-} |2, +2\rangle_B \right] \\ \hbar\sqrt{3(4) - 3(2)} |3, +2\rangle_J &= \left[ \hbar\sqrt{1(2) - 1(0)} |1, 0\rangle_A \right] |2, +2\rangle_B \\ &\quad + |1, +1\rangle_A \left[ \hbar\sqrt{2(3) - 2(1)} |2, +1\rangle_B \right] \\ \sqrt{6} |3, +2\rangle_J &= \left[ \sqrt{2} |1, 0\rangle_A \right] |2, +2\rangle_B + |1, +1\rangle_A \left[ \sqrt{4} |2, +1\rangle_B \right] \\ |3, +2\rangle_J &= \sqrt{\frac{1}{3}} |1, 0\rangle_A |2, +2\rangle_B + \sqrt{\frac{2}{3}} |1, +1\rangle_A |2, +1\rangle_B. \end{aligned} \tag{19.2}$$

Before, we knew only that if the system were in state  $|3, +2\rangle_J$  and we measured  $m_A$ , the result might be 0 or it might be +1. Now we know that the probability of obtaining the result 0 is  $\frac{1}{3}$ , while the probability of obtaining the result +1 is  $\frac{2}{3}$ .

You can continue this process: lower  $|3, +2\rangle_J$  to find an expression for  $|3, +1\rangle_J$ , lower  $|3, +1\rangle_J$  to find an expression for  $|3, 0\rangle_J$ , and so forth. When you get to  $|3, -2\rangle_J$ , you should lower it to find

$$|3, -3\rangle_J = |1, -1\rangle_A |2, -2\rangle_B,$$

and if that's not the result you get, then you made an error somewhere in this long chain.

Now we know how to find expressions for the entire heptet  $|3, m\rangle_J$ , with  $m$  ranging from +3 to -3. But what about the quintet  $|2, m\rangle_J$ , with  $m$  ranging from +2 to -2? If we knew the top member  $|2, +2\rangle_J$ , we could lower away to find the rest of the quintet. But how do we find this starting point?

The trick to use here is orthogonality. We know that

$$|2, +2\rangle_J = \alpha |1, 0\rangle_A |2, +2\rangle_B + \beta |1, +1\rangle_A |2, +1\rangle_B,$$

where  $\alpha$  and  $\beta$  are to be determined, and that

$$|3, +2\rangle_J = \sqrt{\frac{1}{3}} |1, 0\rangle_A |2, +2\rangle_B + \sqrt{\frac{2}{3}} |1, +1\rangle_A |2, +1\rangle_B$$

and that

$$\langle 3, +2 | 2, +2 \rangle_J = 0.$$

We use the orthogonality to find the expansion coefficients  $\alpha$  and  $\beta$ :

$$\begin{aligned} 0 &= \langle 3, +2 | 2, +2 \rangle_J \\ &= \left[ \sqrt{\frac{1}{3}} \langle 1, 0 |_B \langle 2, +2 | + \sqrt{\frac{2}{3}} \langle 1, +1 |_B \langle 2, +1 | \right] \left[ \alpha |1, 0\rangle_A |2, +2\rangle_B + \beta |1, +1\rangle_A |2, +1\rangle_B \right] \\ &= \sqrt{\frac{1}{3}} \alpha \langle 1, 0 |_A \langle 2, +2 | 2, +2 \rangle_B + \sqrt{\frac{1}{3}} \beta \langle 1, 0 |_A \langle 2, +2 | 2, +1 \rangle_B \\ &\quad + \sqrt{\frac{2}{3}} \alpha \langle 1, +1 |_A \langle 2, +1 | 2, +2 \rangle_B + \sqrt{\frac{2}{3}} \beta \langle 1, +1 |_A \langle 2, +1 | 2, +1 \rangle_B \\ &= \sqrt{\frac{1}{3}} \alpha (1) + \sqrt{\frac{1}{3}} \beta (0) + \sqrt{\frac{2}{3}} \alpha (0) + \sqrt{\frac{2}{3}} \beta (1) \\ &= \alpha \sqrt{\frac{1}{3}} + \beta \sqrt{\frac{2}{3}}. \end{aligned}$$

There are, of course, many solutions to this equation, but you can read off a normalized solution, namely

$$\alpha = \sqrt{\frac{2}{3}}, \quad \beta = -\sqrt{\frac{1}{3}}$$

so that

$$|2, +2\rangle_J = \sqrt{\frac{2}{3}} |1, 0\rangle_A |2, +2\rangle_B - \sqrt{\frac{1}{3}} |1, +1\rangle_A |2, +1\rangle_B. \quad (19.3)$$

You could have taken  $|2, +2\rangle_J$  to be the negative of the expression above (or  $i$  times the expression above, or  $\sqrt{i}$  times the expression above, etc.) I recommend against this: life is hard enough on its own, don't go out of your way to deliberately make difficulties for yourself.

Once the expression for  $|2, +2\rangle_J$  is known, we can lower  $m_J$  from  $+2$  all the way to  $-2$  to find expressions for the entire  $j = 2$  quintet.

And then one can find the expression for  $|1, +1\rangle_J$  by demanding that it be orthogonal to  $|3, +1\rangle_J$  and  $|2, +1\rangle_J$ . And once that's found we can lower to find expressions for the entire  $j = 1$  triplet.



In summary, the states of these two angular momenta,  $\ell_A = 1$  and  $\ell_B = 2$ , fall in a Hilbert space with a fifteen-element basis. While there are, of course, an infinite number of bases, the most natural and most useful bases are (1) the states of definite individual angular momenta (the 15 states like  $|\ell_A, m_A\rangle|\ell_B, m_B\rangle$ ) or (2) the states of definite total angular momentum (the 15 states like  $|j, m_j\rangle_j$ ). We now know (in principle) how to express states of the second basis in terms of states in the first basis. The coefficients, like  $\sqrt{1/3}$  and  $\sqrt{2/3}$  in equation (19.2), or  $\sqrt{2/3}$  and  $-\sqrt{1/3}$  in equation (19.3) that implement this change of basis are called Clebsch-Gordan coefficients.<sup>3</sup>

As you can see, it takes a lot of work to compute Clebsch-Gordan coefficients, but fortunately you don't have to do it. There are published tables of Clebsch-Gordan coefficients. Griffiths explains how to use them.

Problem XXX: Commutators. Show that

$$[\hat{J}^2, \hat{L}_{A,z}] = 2i\hbar(\hat{L}_{A,x}\hat{L}_{B,y} - \hat{L}_{A,y}\hat{L}_{B,x}).$$

Without performing any new calculation, find  $[\hat{J}^2, \hat{L}_{B,z}]$ .

---

<sup>3</sup>Alfred Clebsch (1833–1872) and Paul Gordan (1837–1912) were German mathematicians who recognized the importance of these coefficients in the purely mathematical context of invariant theory in about 1868, years before quantum mechanics was discovered. Gordan went on to serve as thesis advisor for Emmy Noether.

## 19.2 Hartree-Fock approximation

For atom with atomic number  $Z$ .

(1) Guess some spherically-symmetric potential energy function that interpolates between

$$\text{for small } r, \quad V(r) \approx -\frac{1}{4\pi\epsilon_0} \frac{Ze^2}{r} \quad (19.4)$$

and

$$\text{for large } r, \quad V(r) \approx -\frac{1}{4\pi\epsilon_0} \frac{e^2}{r}. \quad (19.5)$$

(2) Using all the tricks we've learned about spherically-symmetric potential energy functions, solve (numerically) the energy eigenproblem for the lowest  $Z/2$  one-body energy levels. (If  $Z$  is odd, round up.)

(3) Use the antisymmetrization machinery to combine those levels into the  $Z$ -body ground state.

(4) From the quantal probability density for electrons in configuration space, deduce an electrostatic charge density in position space.

(5) Average that charge density over angle to make it spherically symmetric.

(6) From this spherically-symmetric charge density, use the shell theorem of electrostatics to deduce a spherically-symmetric potential energy function.

(7) Go to step (2)

You'll notice that this process never ends. In practice, you repeat until either you've earned a Ph.D. or you can't stand it any longer.

This is a "mean-field approximation". An electron is assumed to interact with the average (mean) of all the other electrons. Even if you go through this process an infinite number of times, you will never get the fine points of two electrons interacting far from the nucleus and from the other electrons.

Nevertheless, even two or three cycles through this algorithm can produce results in close accord with experiment. This has always surprised me and I think if I understood it I'd discover something valuable about quantum mechanics.

### 19.3 Atomic ground states

In addition to the process described above, you have to worry about spin, and about orbital angular momentum and (when you go on to Hamiltonians more accurate than the above) their interaction.

Friedrich Hund<sup>4</sup> did many such perturbation calculations and noticed regularities that he codified into “Hund’s rules”. Griffith talks about them.

The some aspects of an electronic state are described using a particular notation — called a “term symbol” — which you should know about. A state will have a particular orbital angular momentum  $L$ , spin angular momentum  $S$ , and total angular momentum  $J$ . (It will also have values of  $L_z$ ,  $S_z$ , and  $J_z$ , but they are not recorded in this notation.) You would think these three numbers would be presented as three numbers, but no: they are conventionally presented as the term symbol

$${}^{2S+1}L_J. \quad (19.6)$$

By further convention,  $S$  and  $J$  are given as numbers, while  $L$  is presented as a letter using the S, P, D, F encoding. (In this notation, capital not lower case letters are used. Please don’t ask why.) The ground state of carbon, for example, happens to have  $S = 1$ ,  $L = 1$ , and  $J = 0$ ; it is described as a  ${}^3P_0$  state. One last convention: the spin number is written as a number, but pronounced as a degeneracy. The ground state of carbon is pronounced “triplet pee zero”. The ground state of sodium,  ${}^2S_{1/2}$ , is pronounced “doublet ess one-half”.

The people who write the physics GRE have fallen into the misconception that this term symbol notation tells us something important about nature, rather than about human convention. I recommend that you review the above paragraph the night before you take the GRE.

---

<sup>4</sup>German physicist (1896–1997) who applied quantum mechanics to atoms and molecules, and who discovered quantum tunneling.

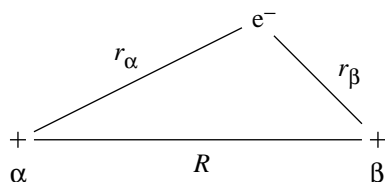


## Chapter 20

# Molecules

### 20.1 The hydrogen molecule ion

The hydrogen molecule ion<sup>1</sup> is two protons and a single electron...  $\text{H}_2^+$ . If we had managed to successfully solve the helium atom problem we would also have solved this one, because it's just three particles interacting through  $1/r^2$  forces. However, you know that this problem has not been exactly solved even in the classical limit. Thus we don't even look for an exact solution: we look for the approximation most applicable to the case of two particles much more massive than the third.



If we take account of the Coulomb forces, but ignore things like the finite size of the nucleus, relativistic motion of the electron, spin-orbit effects, and so forth, the Hamiltonian for one electron and two protons ( $\alpha$  and  $\beta$ ) is

$$\hat{H} = \widehat{\text{KE}}_\alpha + \widehat{\text{KE}}_\beta + \widehat{\text{KE}}_e + \hat{U}_{\alpha\beta} + \hat{U}_{\alpha e} + \hat{U}_{\beta e} \quad (20.1)$$

This is, of course, also the Hamiltonian for the helium atom, or for any three-body problem with pair interactions. Now comes the approximation suitable for the hydrogen molecule ion (but *not* appropriate for the helium

<sup>1</sup>Technically the hydrogen molecule cation.

atom): Assume that the two protons are so massive that they are fixed, and the interaction between them is treated classically. In equations, this approximation demands

$$\widehat{\text{KE}}_\alpha = 0; \quad \widehat{\text{KE}}_\beta = 0; \quad \hat{U}_{\alpha\beta} = U_{\alpha\beta} = \frac{e^2}{4\pi\epsilon_0} \frac{1}{R}. \quad (20.2)$$

The remaining, quantum mechanical, piece of the full Hamiltonian is the electronic Hamiltonian

$$\hat{H}_e = -\frac{\hbar^2}{2m} \nabla^2 - \frac{e^2}{4\pi\epsilon_0} \left( \frac{1}{r_\alpha} + \frac{1}{r_\beta} \right). \quad (20.3)$$

This approximation is called the “Born-Oppenheimer” approximation.

What shall we do with the electronic Hamiltonian? It would be nice to have an analytic solution of the energy eigenproblem. Then we could do precise comparisons between these results and the experimental spectrum of the hydrogen molecule ion, and build on them to study the hydrogen molecule, in exactly the same way that we built on our exact solution for  $\text{He}^+$  to get an approximate solution for He. This goal is hopelessly beyond our reach. [Check out Gordon W.F. Drake, editor, *Atomic, Molecular, and Optical Physics Handbook* (AIP Press, Woodbury, NY, 1996) Reference QC173.A827 1996. There’s a chapter on high-precision calculations for helium, but no chapter on high-precision calculations for the hydrogen molecule ion.] Instead of giving up, we might instead look for an exact solution to the ground state problem. This goal is also beyond our reach. Instead of giving up, we use the variational method to look for an approximate ground state.

Before doing so, however, we notice one exact symmetry of the electronic Hamiltonian that will guide us in our search for approximate solutions. The Hamiltonian is symmetric under the interchange of symbols  $\alpha$  and  $\beta$  or, what is the same thing, symmetric under inversion about the point midway between the two nuclei. Any discussion of parity (see, for example, Gordon Baym *Lectures on Quantum Mechanics* pages 99–101) shows that this means the energy eigenfunctions can always be chosen either odd or even under the interchange of  $\alpha$  and  $\beta$ .

Where will we find a variational trial wavefunction? If nucleus  $\beta$  did not exist, the ground state wavefunction would be the hydrogen ground state wavefunction centered on nucleus  $\alpha$ :

$$\eta_\alpha(\vec{r}) = \frac{1}{\sqrt{\pi a_0^3}} e^{-r_\alpha/a_0} \equiv |\alpha\rangle. \quad (20.4)$$

Similarly if nucleus  $\alpha$  did not exist, the ground state wavefunction would be

$$\eta_\beta(\vec{r}) = \frac{1}{\sqrt{\pi a_0^3}} e^{-r_\beta/a_0} \equiv |\beta\rangle. \quad (20.5)$$

We take as our trial wavefunction a linear combination of these two wavefunctions. This trial wavefunction is called a “linear combination of atomic orbitals” or “LCAO”. So the trial wavefunction is

$$\psi(\vec{r}) = A\eta_\alpha(\vec{r}) + B\eta_\beta(\vec{r}). \quad (20.6)$$

At first glance, it seems that the variational parameters are the complex numbers  $A$  and  $B$ , for a total of four real parameters. However, one parameter is taken up through normalization, and one through overall phase. Furthermore, because of parity the swapping of  $\alpha$  and  $\beta$  can result in at most a change in sign, whence  $B = \pm A$ . Thus our trial wavefunction is

$$\psi(\vec{r}) = A_\pm [\eta_\alpha(\vec{r}) \pm \eta_\beta(\vec{r})], \quad (20.7)$$

where  $A_\pm$  is the normalization constant, selected to be real and positive. (The notation  $A_\pm$  reflects the fact that depending on whether we take the + sign or the – sign, we will get a different normalization constant.)

This might seem like a letdown. We have discussed exquisitely precise variational wavefunction involving hundreds or even thousands of real parameters. Here the only variational parameter is the binary choice: + sign or – sign! Compute  $\langle \hat{H}_e \rangle$  both ways and see which is lower! You don’t even have to take a derivative at the end! Clearly this is a first attempt and more accurate calculations are possible. Rather than give in to despair, however, let’s recognize the limitations and forge on to see what we can discover. At the very least what we learn here will guide us in selecting better trial wavefunctions for our next attempt.

There are only two steps: normalize the wavefunction and evaluate  $\langle \hat{H}_e \rangle$ . However, these steps can be done through a frontal assault (which is likely to get hopelessly bogged down in algebraic details) or through a more subtle approach recognizing that we already know quite a lot about the functions  $\eta_\alpha(\vec{r})$  and  $\eta_\beta(\vec{r})$ , and using this knowledge to our advantage. Let’s use the second approach.

Normalization demands that

$$\begin{aligned} 1 &= |A_\pm|^2 (\langle \alpha | \pm \langle \beta |) (|\alpha \rangle \pm |\beta \rangle) \\ &= |A_\pm|^2 (\langle \alpha | \alpha \rangle \pm \langle \alpha | \beta \rangle \pm \langle \beta | \alpha \rangle + \langle \beta | \beta \rangle) \\ &= 2|A_\pm|^2 (1 \pm \langle \alpha | \beta \rangle) \end{aligned}$$

where in the last step we have used the normalization of  $|\alpha\rangle$  and  $|\beta\rangle$ . The integral  $\langle\alpha|\beta\rangle$  is not easy to calculate, so we set it aside for later by naming it the *overlap integral*

$$I(R) \equiv \langle\alpha|\beta\rangle = \int \eta_\alpha(\vec{r})\eta_\beta(\vec{r}) d^3r. \quad (20.8)$$

In terms of this integral, we can select the normalization to be

$$A_\pm = \frac{1}{\sqrt{2(1 \pm I(R))}}. \quad (20.9)$$

Evaluating the electronic Hamiltonian in the trial wavefunction gives

$$\begin{aligned} \langle\hat{H}_e\rangle &= \frac{(\langle\alpha|\pm\langle\beta|\hat{H}_e(|\alpha\rangle\pm|\beta\rangle))}{2(1\pm I(R))} \\ &= \frac{\langle\alpha|\hat{H}_e|\alpha\rangle\pm\langle\alpha|\hat{H}_e|\beta\rangle\pm\langle\beta|\hat{H}_e|\alpha\rangle+\langle\beta|\hat{H}_e|\beta\rangle}{2(1\pm I(R))} \\ &= \frac{\langle\alpha|\hat{H}_e|\alpha\rangle\pm\langle\beta|\hat{H}_e|\alpha\rangle}{1\pm I(R)} \end{aligned} \quad (20.10)$$

But we have already done large parts of these two integrals:

$$\begin{aligned} \hat{H}_e|\alpha\rangle &= \left[ \widehat{\text{KE}} - \frac{e^2}{4\pi\epsilon_0} \frac{1}{r_\alpha} - \frac{e^2}{4\pi\epsilon_0} \frac{1}{r_\beta} \right] |\alpha\rangle \\ &= \left[ \widehat{\text{KE}} - \frac{e^2}{4\pi\epsilon_0} \frac{1}{r_\alpha} \right] |\alpha\rangle - \frac{e^2}{4\pi\epsilon_0} \frac{1}{r_\beta} |\alpha\rangle \\ &= -\text{Ry} |\alpha\rangle - 2\text{Ry} a_0 \frac{1}{r_\beta} |\alpha\rangle \\ &= -\text{Ry} \left[ |\alpha\rangle + 2\frac{a_0}{r_\beta} |\alpha\rangle \right] \end{aligned} \quad (20.11)$$

whence

$$\langle\alpha|\hat{H}_e|\alpha\rangle = -\text{Ry} \left[ 1 + 2 \left\langle \alpha \left| \frac{a_0}{r_\beta} \right| \alpha \right\rangle \right] \quad (20.12)$$

$$\langle\beta|\hat{H}_e|\alpha\rangle = -\text{Ry} \left[ \langle\beta|\alpha\rangle + 2 \left\langle \beta \left| \frac{a_0}{r_\beta} \right| \alpha \right\rangle \right]. \quad (20.13)$$

On the right-hand side we recognize the overlap integral,  $I(R) = \langle\beta|\alpha\rangle$ , and two new (dimensionless) integrals, which are called the *direct integral*

$$D(R) \equiv \left\langle \alpha \left| \frac{a_0}{r_\beta} \right| \alpha \right\rangle \quad (20.14)$$



and the *exchange integral*

$$X(R) \equiv \left\langle \beta \left| \frac{a_0}{r_\beta} \right| \alpha \right\rangle. \quad (20.15)$$

These two integrals are not easy to work out (I will assign them as homework) but once we do them (plus the overlap integral) we can find the expectation value of the electronic Hamiltonian in the trial wavefunction. It is

$$\begin{aligned} \langle \hat{H}_e \rangle &= -\text{Ry} \frac{1 + 2D(R) \pm I(R) \pm 2X(R)}{1 \pm I(R)} \\ &= -\text{Ry} \left[ 1 + 2 \frac{D(R) \pm X(R)}{1 \pm I(R)} \right]. \end{aligned} \quad (20.16)$$

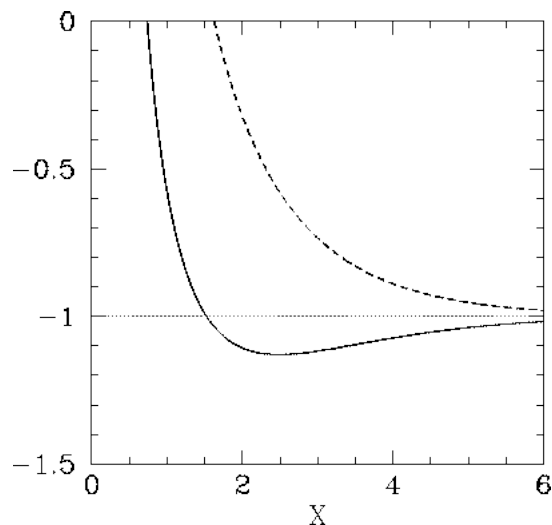
This, remember, is only the electronic part of the Hamiltonian. In the Born-Oppenheimer approximation the nuclear part has no kinetic energy and Coulombic potential energy

$$\frac{e^2}{4\pi\epsilon_0 R} = 2\text{Ry} \frac{a_0}{R}, \quad (20.17)$$

so the upper bound on the total ground state energy is

$$\text{Ry} \left[ 2 \frac{a_0}{R} - 1 - 2 \frac{D(R) \pm X(R)}{1 \pm I(R)} \right]. \quad (20.18)$$

What are the results?



Here the dashed line represents  $-$ , the solid line represents  $+$ .  $X$  means  $R/a_0$ , and the vertical axis is energy in Ry. [When  $R \rightarrow \infty$ , the system is a hydrogen atom (ground state energy  $-Ry$ ) and a clamped proton far away (ground state energy 0).]

### 20.1.1 *Why is + lower energy than -?*

### 20.1.2 *Understanding the integrals*

How can we understand these integrals? This section uses scaled units.

First, all three integrals are always positive.

The overlap integral:  $I(R) = \langle \beta | \alpha \rangle$ .

When  $R \rightarrow \infty$ ,  $I(R)$  approaches zero, exponentially quickly.

When  $R = 0$ ,  $I(R) = 1$ .

The direct integral:  $D(R) = \langle \alpha | 1/r_\beta | \alpha \rangle$ .

When  $R \rightarrow \infty$ ,  $D(R) \rightarrow 1/R$ .

When  $R = 0$ ,  $D(R) = \langle 1/r \rangle = 1$ .

The exchange integral:  $X(R) = \langle \beta | 1/r_\beta | \alpha \rangle$ .

When  $R \rightarrow \infty$ ,  $X(R)$  approaches zero even faster than  $I(R)$  does.

When  $R = 0$ ,  $X(R) = \langle 1/r \rangle = 1$ .

Do the analytic expressions bear these limits out?

$$I(R) = e^{-R} \left( 1 + R + \frac{1}{3} R^2 \right) \quad (20.19)$$

$$D(R) = \frac{1}{R} - \left( 1 + \frac{1}{R} \right) e^{-2R} \quad (20.20)$$

$$X(R) = e^{-R} (1 + R) \quad (20.21)$$

First conclusion: For  $R$  positive,  $I(R) > X(R)$ . Check.

For  $R \rightarrow \infty$ ,  $I(R)$  and  $X(R)$  go to zero exponentially, while  $D(R) \rightarrow 1/R$ . Check.

For  $R \rightarrow 0$ ,

$$I(R) \rightarrow 1 - \frac{1}{6} R^2 + \mathcal{O}(R^3) \quad (20.22)$$

$$X(R) \rightarrow 1 - \frac{1}{2} R^2 + \mathcal{O}(R^3) \quad (20.23)$$

Check, check. But what of  $D(R)$ ? As  $R \rightarrow 0$ , you might say  $D(R) \rightarrow \infty - (1 + \infty)1$ , and the infinities cancel, so you're left with  $D(R) \rightarrow -1$ , but of course that's silly... we've already said that  $D(R)$  is positive. We need to do the limit with some care.

$$\begin{aligned}
 D(R) &= \frac{1}{R} - \left(1 + \frac{1}{R}\right) e^{-2R} \\
 &= \frac{1}{R} - \left(1 + \frac{1}{R}\right) \left(1 + (-2R) + \frac{1}{2}(-2R)^2 + \frac{1}{6}(-2R)^3 + \mathcal{O}((-2R)^4)\right) \\
 &= \frac{1}{R} - \left(1 + \frac{1}{R}\right) \left(1 - 2R + 2R^2 - \frac{4}{3}R^3 + \mathcal{O}(R^4)\right) \\
 &= \frac{1}{R} - \left(1 - 2R + 2R^2 - \frac{4}{3}R^3 + \mathcal{O}(R^4)\right) \\
 &\quad - \frac{1}{R} \left(1 - 2R + 2R^2 - \frac{4}{3}R^3 + \mathcal{O}(R^4)\right) \\
 &= \frac{1}{R} - \left(1 - 2R + 2R^2 - \frac{4}{3}R^3 + \mathcal{O}(R^4)\right) \\
 &\quad - \left(\frac{1}{R} - 2 + 2R - \frac{4}{3}R^2 + \mathcal{O}(R^3)\right) \\
 &= -\left(-1 + \frac{2}{3}R^2 + \mathcal{O}(R^3)\right) \\
 &= 1 - \frac{2}{3}R^2 + \mathcal{O}(R^3). \tag{20.24}
 \end{aligned}$$

All three integrals start at 1 when  $R = 0$ . As  $R$  increases they all take off with zero slope, but drop quadratically:  $I(R)$  is highest, then  $X(R)$ , and  $D(R)$  lowest. But at some point  $D(R)$  crosses the other two. While all three approach zero as  $R \rightarrow \infty$ ,  $D(R)$  does so much more slowly than the other two.

### 20.1.3 Why is $H_2^+$ hard?

Obviously not Pauli exclusion! But if you plot the various contributions, you see that it's classical nuclear repulsion, not "Heisenberg hardness".

## 20.2 Problems

### 20.1 The hydrogen molecule ion: Evaluation of integrals

Evaluate the direct and exchange integrals  $D(R)$  and  $X(R)$ . (*Hint:* Remember that  $\sqrt{x^2} = |x|$ .) Plot as a function of  $R$  the overlap integral,  $I(R)$ , as well as  $D(R)$  and  $X(R)$ .

### 20.2 The hydrogen molecule ion: Thinking about integrals

For the hydrogen molecule ion, find and plot the expectation values of nuclear potential energy, total electronic energy, kinetic electronic energy, and potential electronic energy for the state  $\psi_+(\vec{r})$ , as functions of  $R$ . Do these plots shed any light on our initial question of “Why is stuff hard?” (We gave possible answers of “repulsion hardness,” “Heisenberg hardness,” and “Pauli hardness.”) *Bonus*: The hydrogen molecule ion cannot display Pauli hardness, because it has only one quantal particle. Can you generalize this discussion to the neutral hydrogen molecule?

### 20.3 Improved variational wavefunction

Everett Schlawin (‘09) suggested using “shielded” subwavefunctions like equation (18.4) in place of the subwavefunctions (20.4) and (20.5) that go into making trial wavefunction (20.7). Then there would be a variational parameter  $Z$  in addition to the binary choice of  $+$  or  $-$ . I haven’t tried this, but through the usual variational argument, it can’t be worse than what we’ve tried so far! (That is, the *results* can’t be worse. The *amount of labor* involved can be far, far worse.) Execute this suggestion. Show that this trial wavefunction results in the exact helium ion ground state energy in the case  $R = 0$ .

## 20.3 The hydrogen molecule

When we discussed the helium atom, we had available an exact solution (that is, exact ignoring fine and hyperfine structure) of the helium ion problem. We used the one-body levels of the helium ion problem as building blocks for the two-body helium atom problem. Then we added electron-electron repulsion. You will recall, for example, that the helium atom ground state had the form (where “level” refers to a solution of the one-body helium ion problem)

$$(\text{two electrons in ground level}) \times (\text{spin singlet}) \quad (20.25)$$

while the helium atom first excited state had the form

$$(\text{one electron in ground level, one in first excited level}) \times (\text{spin triplet}). \quad (20.26)$$

We will attempt the same strategy for the hydrogen molecule, but we face a roadblock at the very first step — we lack an exact solution to the hydrogen molecule ion problem! Using LCAO, we have a candidate for a ground state, namely

$$\psi_+(\vec{r}) = A_+[\eta_\alpha(\vec{r}) + \eta_\beta(\vec{r})]. \quad (20.27)$$

## 20.4 Can we do better?

Try out our LCAO upper bound for the electronic ground state energy (20.16) at  $R = 0$ : The result is  $-3$  Ry. But for  $R = 0$  this is just the Helium ion, for which the exact ground state energy is  $-4$  Ry. Sure enough, the variational method produces an upper bound, but it's a poor one.

We've seen before that the trick to getting good variational bounds is to figure out the qualitative character of the true wavefunction and select a trial wavefunction that mimics that character. Friedrich Hund, Robert Mulliken, John C. Slater, and John Lennard-Jones started out by dreaming up a trial wavefunction that could mimic the character of the true wavefunction at  $R = 0$ . Their techniques evolved into what is today called the "molecular orbital method". This is only one of several choices of trial wavefunction. Others are called "valance bond theory" or "the Hückel method" or "the extended Hückel method".

Story about Roald Hoffmann.

All these are primitive, but in synthetic chemistry, you don't need the spectrum, you don't need the ground state energy, all you need to know is which structure has lower energy, and that's the one you'll synthesize.

Today, chemists are much more likely to use a completely different approach, called "density-functional theory". This was developed by the physicist Walter Kohn and made readily accessible through the computer program GAUSSIAN written by the mathematician John Pople. When Kohn and Pople won the Nobel Prize in Chemistry in 1998, I heard some grumbling among chemists that Chemistry Nobel laureates should have taken at least one chemistry course.



## Chapter 21

# WKB: The Quasiclassical Approximation

When I started learning quantum mechanics, I worked a lot of integrals and diagonalized a lot of matrices. But I also vaguely wondered “Why is quantum mechanics true?”. For example, why are there limits on our ability to observe position and momentum? But eventually (while teaching *Applied Quantum Mechanics*) I realized that I had the question backwards. The real question is “We know that interference and entanglement exist. Why don’t we notice them in daily life?” For example, Heisenberg indeterminacy principle answers the question “When is the classical approximation adequate?” That is, the real question concerns the classical limit of quantum mechanics. Ehrenfest’s Theorem shows that classical mechanics *can* be the limit of quantum mechanics, but not that it *has to* be. Research on this topic continues under the name “decoherence”. We approach this subject through the quasiclassical approximation.

The WKB technique finds approximate solutions to the energy eigenproblem in one dimension. It is named for three physicists who independently discovered it: the German Gregor Wentzel, the Dutchman Hendrik Kramers, and the Frenchman Léon Brillouin. In the Netherlands it is known as the KWB approximation, in France as BWK, and in Britain as JWKB (adding a tribute to the English mathematician Sir Harold Jeffreys, who in fact discovered the approximation three years before Wentzel, Kramers, and Brillouin did). In Russia it is known as the quasiclassical approximation, the name that I prefer.

The fact that this approximation was discovered independently four times suggests, correctly, that the idea is pretty straightforward.<sup>1</sup> Focus on

---

<sup>1</sup>The same basic idea can be used in many similar situations: to light moving in a medium where the index of refraction varies slowly, for example.

a region where the potential energy function  $V(x)$  is *constant*. Within that region the eigenfunction of energy  $E$  is, when  $E > V$ , given by

$$\eta(x) = Ae^{\pm ikx} \quad \text{where} \quad \hbar k = \sqrt{2m(E - V)}. \quad (21.1)$$

The plus sign indicates positive momentum, the minus sign negative momentum, and the general solution is of course a linear combination of the two. The wavefunction is sinusoidal oscillatory, with constant wavelength

$$\lambda = \frac{2\pi\hbar}{\sqrt{2m(E - V)}} \quad (21.2)$$

and constant amplitude. Now suppose that  $V(x)$  is *not* constant, but that it varies slowly over the length  $\lambda$ . Then my guess would be that  $\eta(x)$  is almost sinusoidal, but the wavelength and amplitude vary slowly with  $x$ . That is, I would seek oscillatory solutions like

$$\eta(x) = A(x)e^{\pm ik(x)} \quad \text{where} \quad \hbar k(x) = \sqrt{2m(E - V(x))}. \quad (21.3)$$

On the other hand, if the potential energy function  $V(x)$  is constant but  $E < V$ , then the energy eigenfunction is

$$\eta(x) = Ae^{\pm x/d} \quad \text{where} \quad d = \frac{\hbar}{\sqrt{2m(V - E)}}. \quad (21.4)$$

Here  $d$  is the characteristic exponential decay length: If one walks in the direction of decreasing function, then the function diminishes by a factor of  $1/e$  (about  $1/3$ ) every time one steps a distance  $d$ . However when  $V(x)$  is *not* constant, but varies slowly over the length of  $d$ , then  $\eta(x)$  is almost exponential, but the decay length and amplitude vary slowly with  $x$ . That is, I should seek solutions like

$$\eta(x) = A(x)e^{\pm x/d(x)} \quad \text{where} \quad d(x) = \frac{\hbar}{\sqrt{2m(V(x) - E)}}. \quad (21.5)$$

What we have said so far reinforces the qualitative expectations for energy eigenfunction sketching established in chapter XX.

There is one place where this entire scheme is guaranteed to fail. If  $E = V(x)$  then  $\lambda(x) = d(x) = \infty$ , and no potential energy function varies “slowly on the scale of infinity”. The proper handling of these so-called “classical turning points” is the most difficult facet of deriving the quasiclassical approximation. However we will find that once the derivation is done the final result is easy to state and to use.

If you apply these ideas to two- or three-dimensional problems, you find that the classical turning points are now lines (in two dimensions) or



surfaces (in three dimensions). The matching program at turning points becomes a matching program over lines or surfaces (called in this context “caustics”) and the results are neither easy to state nor simple to use. They are connected with classical chaos, and, remarkably, with the theory of the rainbow. Such are the nimble abstractions of mathematics. We will not pursue these avenues in this book.

Define

$$p_c(x) = \sqrt{2m(E - V(x))} \quad (21.6)$$

This function is the “classical momentum”, that is the momentum that a classical particle of energy  $E$  would have if it were located at  $x$ .

### 21.1 The connection region

We have one formula accurate within the classically allowed region, and another accurate within the classically prohibited region. But we lack a formula accurate within the connection region near the classical turning point, where the quasiclassical approximation fails. The job of this section is to find a formula accurate in this region.

GRAPH with blue horizontal line marked  $E$ .  
 olive line slanted from SW to NE  $V(x) = V_x + s(x - x_x)$ .  
 Vertical dashed line  $x_x$   
 arrow to right of dashed line  $x' = x - x_x$

Graph with qualitative  $\eta(x)$  sketch?

cross point  $x_x \dots x$  for “crossing”. So  $V_x = E$ .

$$-\frac{\hbar^2}{2m} \frac{d^2\eta}{dx^2} + V(x)\eta(x) = E\eta(x) \quad (21.7)$$

$$-\frac{\hbar^2}{2m} \frac{d^2\eta}{dx^2} + [E - s(x - x_x)]\eta(x) = E\eta(x) \quad (21.8)$$

In terms of the new variable  $x'$

$$-\frac{\hbar^2}{2m} \frac{d^2\eta}{dx'^2} + sx'\eta(x') = 0 \quad (21.9)$$

There are only two parameters:  $\hbar^2/2m$  and  $s$ . What is the characteristic length for this problem?

quantity	dimensions
$\hbar^2/2m$	[mass][length] <sup>4</sup> /[time] <sup>2</sup>
$s$	[mass][length]/[time] <sup>2</sup>

Clearly the characteristic length is

$$x_0 = \left( \frac{\hbar^2/2m}{s} \right)^{1/3} \quad (21.10)$$

Defining the scaled variable

$$\tilde{x} = x'/x_0 \quad (21.11)$$

we have

$$-\frac{d^2\eta}{d\tilde{x}^2} + \tilde{x}\eta(\tilde{x}) = 0. \quad (21.12)$$

## 21.2 Why is WKB the “quasiclassical” approximation?

The approximation works when the de Broglie wavelength  $h/p$  is much less than the characteristic length  $L_c$  of variations in the potential energy function:

$$\begin{aligned} h/p &\ll L_c \\ p &\gg h/L_c. \end{aligned} \quad (21.13)$$

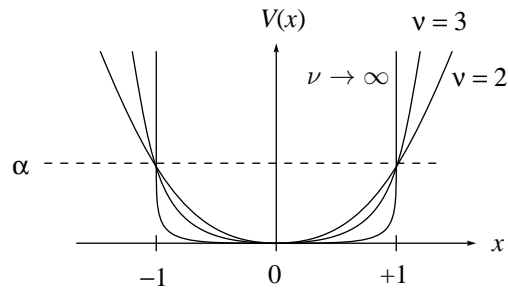
That is, it works for large — classical — values of momentum. Remember that when I say large I don’t mean large on a human scale (say by comparing the momentum of a gnat to the momentum of a semi-truck), I mean large on the scale of  $h/L_c$ . So the momentum could be very small on a human scale yet the WKB approximation would still work very well.

## 21.3 The “power law” potential

While the quasiclassical approximation is difficult to derive, it is straightforward to apply. This section applies the approximation to the so-called “power law” potential energy function,

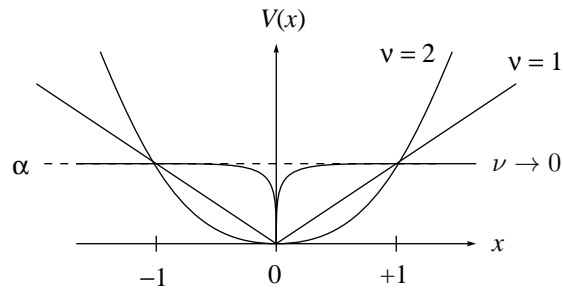
$$V(x) = \alpha|x|^\nu. \quad (21.14)$$

When  $\nu = 2$  this is just the simple harmonic oscillator, which we have studied extensively. When  $\nu > 2$  this potential traces out successively steeper potential wells as  $\nu$  increases:



In the limit  $\nu \rightarrow \infty$ , the power law potential approaches an infinite square well.

Meanwhile, when  $\nu < 2$  this potential traces out successively flatter potential wells as  $\nu$  decreases:

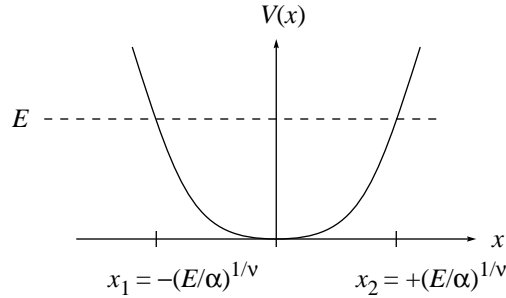


In the limit  $\nu \rightarrow 0$ , the power law potential approaches the flat potential  $V(x) = \alpha$ .

I don't know of any physical system that obeys the power law potential (except for the special cases  $\nu = 0$ ,  $\nu = 2$ , and  $\nu \rightarrow \infty$ ), but it's a good idea to understand quantum mechanics even in cases where it doesn't reflect any physical system.

To apply the quasiclassical approximation, locate the classical turning points at

$$x_1 = -(E/\alpha)^{1/\nu} \quad \text{and} \quad x_2 = +(E/\alpha)^{1/\nu}, \quad (21.15)$$



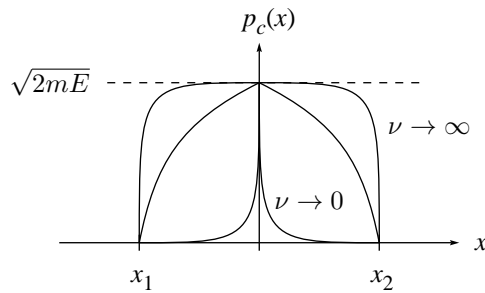
and then perform the integration

$$\int_{x_1}^{x_2} p_c(x) dx = (n - \frac{1}{2})\pi\hbar \quad (21.16)$$

where

$$p_c(x) = \sqrt{2m(E - V(x))} = \sqrt{2m(E - \alpha|x|^{1/\nu})}. \quad (21.17)$$

It's always a good idea to sketch the integrand before executing the integral, and that's what I do here:



So

$$\begin{aligned} \int_{x_1}^{x_2} p_c(x) dx &= \int_{x_1}^{x_2} \sqrt{2m(E - V(x))} dx \\ &= \sqrt{2m} \int_{-(E/\alpha)^{1/\nu}}^{+(E/\alpha)^{1/\nu}} \sqrt{E - \alpha|x|^\nu} dx \\ &= 2\sqrt{2m} \int_0^{+(E/\alpha)^{1/\nu}} \sqrt{E - \alpha x^\nu} dx. \end{aligned}$$

How should one execute this integral? I prefer to integrate over dimensionless variables, so as to separate the physical operation of setting up an integral from the mathematical operation of executing that integral. For that reason I define the dimensionless variable  $u$  through

$$\begin{aligned} \alpha x^\nu &= Eu^\nu, \\ x &= \left(\frac{E}{\alpha}\right)^{1/\nu} u, \\ u &= \left(\frac{\alpha}{E}\right)^{1/\nu} x. \end{aligned}$$

Changing the integral to this variable

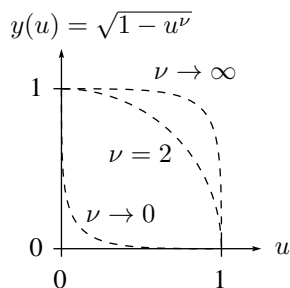
$$\begin{aligned} \int_{x_1}^{x_2} p_c(x) dx &= 2\sqrt{2m} \int_0^1 \sqrt{E - Eu^\nu} \left(\frac{E}{\alpha}\right)^{1/\nu} dx \\ &= 2\sqrt{2mE} \left(\frac{E}{\alpha}\right)^{1/\nu} \int_0^1 \sqrt{1 - u^\nu} dx \\ &= \frac{(8m)^{1/2}}{\alpha^{1/\nu}} E^{(2+\nu)/2\nu} \int_0^1 \sqrt{1 - u^\nu} dx \end{aligned}$$

where the integral here is a numerical function of  $\nu$  independent of  $m$  or  $E$  or  $\alpha$ . Let’s call it

$$I(\nu) = \int_0^1 \sqrt{1 - u^\nu} dx. \quad (21.18)$$

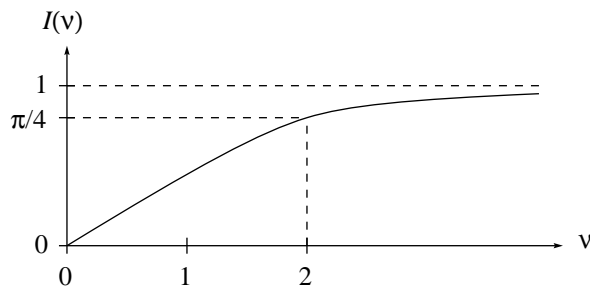
If you try to evaluate this integral in terms of polynomials or trig functions or anything familiar, you will fail. This is a function of  $\nu$  all right, but we’re going to have to uncover its properties on our own without recourse to familiar functions.

Let's start by graphing the integrand.



$I(\nu)$  is the area under the curve. You could produce a table of values through numerical integration, but let's uncover its properties first. It's clear from the graph that  $I(0) = 0$ , that as  $\nu \rightarrow \infty$ ,  $I(\nu) \rightarrow 1$ , and that  $I(\nu)$  increases monotonically.

When  $\nu = 2$ , the integrand  $y$  is  $y = \sqrt{1 - u^2}$  so  $u^2 + y^2 = 1$ ...the integrand traces out a quarter circle of radius 1. The area under this curve is of course  $\pi/4$ . So my first thought is that the function  $I(\nu)$  looks like this:



But I want to investigate one detail further: What is the behavior of  $I(\nu)$  for small values of  $\nu$ ? To find this, I need to understand the behavior

of  $u^\nu$  for small values of  $\nu$ .

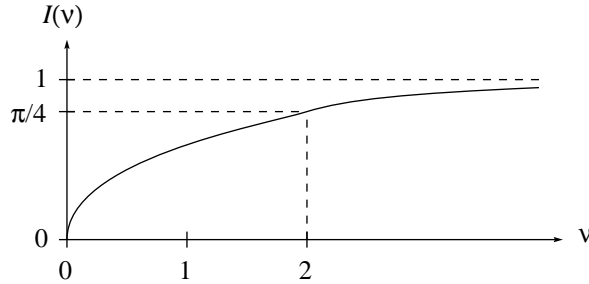
$$\begin{aligned}
 e^x &= 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \dots \\
 u^\nu &= e^{\nu \ln u} = 1 + \nu \ln u + \frac{1}{2}\nu^2 \ln^2 u + \frac{1}{6}\nu^3 \ln^3 u + \dots \\
 1 - u^\nu &= -\nu \ln u - \frac{1}{2}\nu^2 \ln^2 u - \frac{1}{6}\nu^3 \ln^3 u + \dots \\
 \sqrt{1 - u^\nu} &\approx \sqrt{\nu} \sqrt{-\ln u}
 \end{aligned}$$

At first glance it looks very bad to see that negative sign under the square root radical, but then you remember that when  $0 < u < 1$ ,  $\ln u$  is negative, so it’s a good thing that the negative sign is there!

For small values of  $\nu$ ,

$$I(\nu) \approx \sqrt{\nu} \int_0^1 \sqrt{-\ln u} \, du = \sqrt{\nu} (\text{some positive number}). \quad (21.19)$$

Even without knowing the value of that positive number, you know that  $I(\nu)$  takes off from  $\nu = 0$  with infinite slope, like this:



[[You don’t really need the value of “some positive number”, but if you’re insatiably curious, use the substitution  $v = -\ln u$  to find

$$\int_0^1 \sqrt{-\ln u} \, du = \int_\infty^0 \sqrt{v} (-e^{-v}) \, dv = \int_0^\infty v^{1/2} e^{-v} \, dv = \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2},$$

so for small values of  $\nu$ ,

$$I(\nu) \approx \frac{\sqrt{\pi}}{2} \sqrt{\nu}. \quad ]]$$

A formal analysis shows that our integral  $I(\nu)$  can be expressed in terms of gamma functions as

$$I(\nu) = \frac{\sqrt{\pi}}{2 + \nu} \frac{\Gamma\left(\frac{1}{\nu}\right)}{\Gamma\left(\frac{1}{\nu} + \frac{1}{2}\right)},$$

but the graph actually tells you more than this formal expression does. When I was an undergraduate only a very few special functions (for example the  $\Gamma$  function) had been laboriously worked out numerically and tabulated, so it was important to express your integral of interest in terms of one of those few that had been worked out. Now numerical integration is a breeze (your phone is more powerful than the single computer we had on campus when I was an undergraduate), so it's more important to be able to tease information out of the function as we've done here.

In summary, the energy eigenvalues obtained through the quasiclassical approximation

$$(n - \frac{1}{2})\pi\hbar = \frac{(8m)^{1/2}}{\alpha^{1/\nu}} E^{(2+\nu)/2\nu} I(\nu)$$

are

$$E_n = \left[ \frac{\alpha^{1/\nu}}{(8m)^{1/2} I(\nu)} (n - \frac{1}{2})\pi\hbar \right]^{2\nu/(2+\nu)} \quad n = 1, 2, 3, \dots \quad (21.20)$$

You could spend a lot of time probing this equation to find out what it tells us about quantum mechanics. (You could also spend a lot of time looking at the quasiclassical wavefunctions.) I'll content myself with examining the energy eigenvalues for the three special cases  $\nu = 2$ ,  $\nu \rightarrow \infty$ , and  $\nu \rightarrow 0$ .

When  $\nu = 2$  the power-law potential  $V(x) = \alpha x^2$  becomes the simple harmonic oscillator  $V(x) = \frac{1}{2}m\omega^2 x^2$ . Equation (21.20) becomes

$$\begin{aligned} E_n &= \frac{\alpha^{1/2}}{(8m)^{1/2} I(2)} (n - \frac{1}{2})\pi\hbar \\ &= \frac{(\frac{1}{2}m\omega^2)^{1/2}}{(8m)^{1/2} \pi/4} (n - \frac{1}{2})\pi\hbar \\ &= (n - \frac{1}{2})\hbar\omega \quad n = 1, 2, 3, \dots \end{aligned} \quad (21.21)$$

The exact eigenvalues are of course

$$E_n = (n + \frac{1}{2})\hbar\omega \quad n = 0, 1, 2, 3, \dots$$

For the simple harmonic oscillator, the quasiclassical energy eigenvalues are exactly correct. [[The energy eigenfunctions are not.]]



When  $\nu \rightarrow \infty$  the power-law potential becomes an infinite square well of width  $L = 2$ . Equation (21.20) becomes

$$\begin{aligned} E_n &= \left[ \frac{\alpha^{1/\infty}}{(8m)^{1/2} I(\infty)} (n - \frac{1}{2}) \pi \hbar \right]^2 \\ &= \left[ \frac{1}{(8m)^{1/2}} (n - \frac{1}{2}) \pi \hbar \right]^2 \\ &= \frac{\pi^2 \hbar^2}{8m} (n - \frac{1}{2})^2. \end{aligned} \quad (21.22)$$

The exact eigenvalues are (when  $L = 2$ )

$$E_n = \frac{\pi^2 \hbar^2}{2mL^2} n^2 = \frac{\pi^2 \hbar^2}{8m} n^2.$$

Not bad for an approximation.

When  $\nu \rightarrow 0$  the power-law potential becomes the flat, constant potential  $V(x) = \alpha$ . This “free particle” potential admits no bound states. How will the quasiclassical approximation deal with this?

$$\begin{aligned} E_n &= \left[ \frac{\alpha^{1/\nu}}{(8m)^{1/2} I(\nu)} (n - \frac{1}{2}) \pi \hbar \right]^{2\nu/(2+\nu)} \\ &= \frac{\alpha^{2/(2+\nu)}}{(8m)^{\nu/(2+\nu)} I(\nu)^{2\nu/(2+\nu)}} [(n - \frac{1}{2}) \pi \hbar]^{2\nu/(2+\nu)} \\ &\rightarrow \frac{\alpha}{(8m)^0 I(\nu)^\nu} [(n - \frac{1}{2}) \pi \hbar]^0 \\ &\rightarrow \frac{\alpha}{I(\nu)^\nu}. \end{aligned}$$

But what is  $I(\nu)^\nu$  for small  $\nu$ ? We’ve already seen at equation (21.19) that it is  $\sqrt{\nu}^\nu$  (some positive number) $^\nu$ . The right part goes to 1, but

$$\sqrt{\nu}^\nu = \nu^{\nu/2} = e^{\nu \ln \nu/2} \rightarrow e^0 = 1.$$

Thus as  $\nu \rightarrow 0$ ,

$$E_n \rightarrow \alpha \quad \text{for all values of } n. \quad (21.23)$$



## Chapter 22

# The Interaction of Matter and Radiation

Two questions:

- (1) Our theorem says atoms stay in excited energy state forever!
- (2) Absorb light of only one frequency ... what, will absorb light of wavelength 471.3428 nm but not 471.3427 nm?

Strangely, we start our quest to solve these problems by figuring out how to solve differential equations.

### 22.1 Perturbation Theory for the Time Development Problem

By now, you have realized that quantum mechanics is an art of approximations. I make no apologies for this: After all, physics is an art of approximations. (The classical “three-body problem” has never been solved exactly, and never will be.) Indeed, life is an art of approximations. (If you’re waiting for the perfect boyfriend or girlfriend before making a commitment, you’ll be waiting for a long time — and for some, that long wait is a poor solution to the problem of life.)

Furthermore, much of the fun and creativity of theoretical physics comes from finding applicable approximations. If theoretical physics were nothing but turning a mathematical crank to mechanically grind out solutions, it would not be exciting. I do not apologize for the fact that, to do theoretical physics, you have to think!

## 22.2 Setup

Here's our problem:

Solve the initial value problem for the Hamiltonian

$$\hat{H}(t) = \hat{H}^{(0)} + \hat{H}'(t) \quad (22.1)$$

given the solution  $\{|n\rangle\}$  of the unperturbed energy eigenproblem

$$\hat{H}^{(0)}|n\rangle = E_n|n\rangle. \quad (22.2)$$

Here we're thinking of  $\hat{H}'(t)$  as being in some sense "small" compared to the unperturbed Hamiltonian  $\hat{H}^{(0)}$ . One common example is a burst of light shining on an atom. Note also that it doesn't make sense to solve the energy eigenproblem for  $\hat{H}(t)$ , because this Hamiltonian depends upon time, so it doesn't have stationary state solutions!

We solve this problem by expanding the solution  $|\psi(t)\rangle$  in the basis  $\{|n\rangle\}$ :

$$|\psi(t)\rangle = \sum_n C_n(t)|n\rangle \quad \text{where} \quad C_n(t) = \langle n|\psi(t)\rangle. \quad (22.3)$$

Once we know the  $C_n(t)$ , we'll know the solution  $|\psi(t)\rangle$ . Now, the state vector evolves according to

$$\frac{d}{dt}|\psi(t)\rangle = -\frac{i}{\hbar}\hat{H}|\psi(t)\rangle \quad (22.4)$$

so the expansion coefficients evolve according to

$$\begin{aligned} \frac{dC_n(t)}{dt} &= -\frac{i}{\hbar}\langle n|\hat{H}|\psi(t)\rangle \\ &= -\frac{i}{\hbar}\sum_m \langle n|\hat{H}|m\rangle C_m(t) \\ &= -\frac{i}{\hbar}\sum_m \left[ \langle n|\hat{H}^{(0)}|m\rangle + \langle n|\hat{H}'|m\rangle \right] C_m(t) \\ &= -\frac{i}{\hbar}\sum_m \left[ E_m\delta_{m,n} + H'_{n,m} \right] C_m(t) \\ &= -\frac{i}{\hbar} \left[ E_n C_n(t) + \sum_m H'_{n,m} C_m(t) \right] \end{aligned} \quad (22.5)$$

This result is exact: we have yet to make any approximation.

Now, if  $\hat{H}'(t)$  vanished, the solutions would be

$$C_n(t) = C_n(0)e^{-(i/\hbar)E_n t}, \quad (22.6)$$

which motivates us to define new variables  $c_n(t)$  through

$$C_n(t) = c_n(t)e^{-(i/\hbar)E_n t}. \quad (22.7)$$

Because the “bulk of the time development” comes through the  $e^{-(i/\hbar)E_n t}$  term, the  $c_n(t)$  presumably have “less time dependence” than the  $C_n(t)$ . In other words, we expect the  $c_n(t)$  to vary slowly with time.

Plugging this definition into the time development equation (22.5) gives

$$\begin{aligned} & \frac{dc_n(t)}{dt} e^{-(i/\hbar)E_n t} + c_n(t) (-i/\hbar)E_n e^{-(i/\hbar)E_n t} \\ &= -\frac{i}{\hbar} \left[ E_n c_n(t) e^{-(i/\hbar)E_n t} + \sum_m H'_{n,m} c_m(t) e^{-(i/\hbar)E_m t} \right] \end{aligned} \quad (22.8)$$

or

$$\frac{dc_n(t)}{dt} = -\frac{i}{\hbar} \sum_m H'_{n,m} c_m(t) e^{+(i/\hbar)(E_n - E_m)t}. \quad (22.9)$$

Once again, this equation is exact. Its formal solution, given the initial values  $c_n(0)$ , is

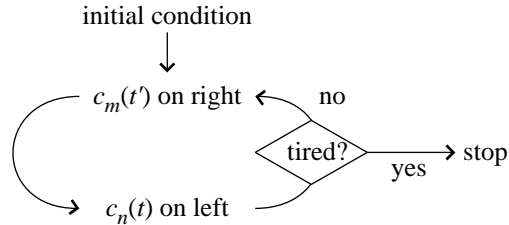
$$c_n(t) = c_n(0) - \frac{i}{\hbar} \sum_m \int_0^t H'_{n,m}(t') c_m(t') e^{+(i/\hbar)(E_n - E_m)t'} dt'. \quad (22.10)$$

This set of equations (one for each basis element) is exact, but at first glance seems useless. The unknown quantities  $c_n(t)$  are present on the left, but also the right-hand sides.

We make progress using our idea that the coefficients  $c_n(t)$  are changing *slowly*. In a very crude approximation, we can think that they're not changing at all. So on the right-hand side of equation (22.10) we plug in not functions, but the constants  $c_m(t') = c_m(0)$ , namely the given initial conditions.

Having made that approximation, we can now perform the integrations and produce, on the left-hand side of equation (22.10), functions of time  $c_n(t)$ . These coefficients aren't exact, because they were based on the crude approximation that the coefficients were constant in time, but they're likely to be better approximations than we started off with.

Now, armed with these more accurate coefficients, we can plug these into the right-hand side of equation (22.10), perform the integration, and produce yet more accurate coefficients on the left-hand side. This process can be repeated over and over, for as long as our stamina lasts.



There is actually a theorem assuring us that this process will converge!

*Theorem (Picard<sup>1</sup>)* If the matrix elements  $H'_{n,m}(t)$  are continuous in time and bounded, and if the basis is finite, then this method converges to the correct solution.

The theorem does not tell us how many iterations will be needed to reach a desired accuracy. In practice, one usually stops upon reaching the first non-zero correction.

In particular, if the initial state is some eigenstate  $|a\rangle$  of the unperturbed Hamiltonian  $\hat{H}^{(0)}$ , then to first order

$$c_n(t) = -\frac{i}{\hbar} \int_0^t H'_{n,a}(t') e^{+(i/\hbar)(E_n - E_a)t'} dt' \quad \text{for } n \neq a \quad (22.11)$$

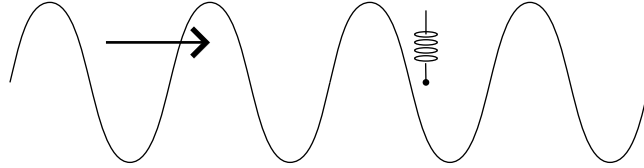
$$c_a(t) = 1 - \frac{i}{\hbar} \int_0^t H'_{a,a}(t') dt'$$

If the system is in energy state  $|a\rangle$  at time zero, then the probability of finding it in energy state  $|b\rangle$  at time  $t$ , through the influence of perturbation  $\hat{H}'(t)$ , is called the *transition probability*

$$P_{a \rightarrow b}(t) = |C_b(t)|^2 = |c_b(t)|^2. \quad (22.12)$$

*Example:* An electron bound to an atom is approximated by a one-dimensional simple harmonic oscillator of natural frequency  $\omega_0$ . The oscillator is in its ground state  $|0\rangle$  and then exposed to light of electric field amplitude  $E_0$  and frequency  $\omega$  for time  $t$ . (The light is polarized in the direction of the oscillations.) What is probability (in first-order perturbation theory) of ending up in state  $|b\rangle$ ?

<sup>1</sup>Émile Picard (1856–1941) made immense contributions to complex analysis and to the theory of differential equations. He wrote one of the first textbooks concerning the theory of relativity, and married the daughter of Charles Hermite.



*Solution part A — What is the Hamiltonian?* If it were a classical particle of charge  $-e$  exposed to electric field  $E_0 \sin \omega t$ , it would experience a force  $-eE_0 \sin \omega t$  and hence have a potential energy of  $eE_0 x \sin \omega t$ . (We can ignore the spatial variation of electric field because the electron is constrained to move only up and down — that’s our “one dimensional” assumption. We can ignore magnetic field for the same reason.)

The quantal Hamiltonian is then

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{m\omega_0^2}{2}\hat{x}^2 + eE_0\hat{x}\sin\omega t. \quad (22.13)$$

We identify the first two terms as the time-independent Hamiltonian  $\hat{H}^{(0)}$  and the last term as the perturbation  $\hat{H}'(t)$ .

*Solution part B — Apply perturbation theory.* The matrix element is

$$H_{n,0}(t) = \langle n|\hat{H}'(t)|0\rangle = eE_0 \sin \omega t \langle n|\hat{x}|0\rangle = eE_0 \sin \omega t \sqrt{\frac{\hbar}{2m\omega_0}} \delta_{n,1}. \quad (22.14)$$

(Remember your raising and lowering operators! See equation (D.31).) Invoking equations (22.11), we obtain

$$c_n(t) = 0 \quad \text{for } n \neq 0, 1 \quad (22.15)$$

$$c_1(t) = -\frac{i}{\hbar} eE_0 \sqrt{\frac{\hbar}{2m\omega_0}} \int_0^t \sin \omega t' e^{i\omega_0 t'} dt' \quad (22.16)$$

$$c_0(t) = 1 \quad (22.17)$$

We will eventually need to perform the time integral in equation (22.16), but even before doing so the main qualitative features are clear: First, probability is *not* conserved within first order perturbation theory. The probability of remaining in the ground state is 1, but the probability of transition to the first excited state is finite! Second, to first order transitions go only to the first excited state. This is an example of a *selection rule*.

The time integral in equation (22.16) will be evaluated at equation (22.30). For now, let's just call it  $I(t)$ . In terms of this integral, the transition probabilities are

$$P_{0 \rightarrow b}(t) = 0 \quad \text{for } b \neq 0, 1 \quad (22.18)$$

$$P_{0 \rightarrow 1}(t) = \frac{e^2 E_0^2}{2m\hbar\omega_0} I(t) I^*(t) \quad (22.19)$$

$$P_{0 \rightarrow 0}(t) = 1 \quad (22.20)$$

### 22.3 Light absorption

How do atoms absorb light?

More specifically, if an electron in atomic energy eigenstate  $|a\rangle$  (usually but not always the ground state) is exposed to a beam of monochromatic, polarized light for time  $t$ , what is the probability of it ending up in atomic energy eigenstate  $|b\rangle$ ? We answer this question to first order in time-dependent perturbation theory.

First, we need to find the effect of light on the electron. We'll treat the light classically — that is, we'll ignore the quantization of the electromagnetic field (quantum electrodynamics) that gives rise to the concept of photons. Consider the light wave (polarized in the  $\hat{k}$  direction, with frequency  $\omega$ ) as an electric field

$$\vec{E}(\vec{r}, t) = E_0 \hat{k} \sin(\vec{k} \cdot \vec{r} - \omega t). \quad (22.21)$$

Presumably, the absorption of light by the atom will result in some sort of diminution of the light beam's electric field, but we'll ignore that. (A powerful beam from a laser will be somewhat diminished when some of the light is absorbed by a single atom, but not a great deal.) The light beam has a magnetic field as well as an electric field, but the magnetic field amplitude is  $B_0 = E_0/c$ , so the electric force is on the order of  $eE_0$  while the magnetic force is on the order of  $evB_0 = e(v/c)E_0$ . Since the electron moves at non-relativistic speeds,  $v/c \ll 1$  and we can ignore the magnetic effect. Finally, the electric field at one side of the atom differs from the electric field at the other side of the atom, but the atom is so small compared to the wavelength of light (atom: about 0.1 nm; wavelength of violet light: about 400 nm) that we can safely ignore this also.

Using these approximations, the force experienced by an electron due to the light beam is

$$\vec{F}(t) = -eE_0 \hat{k} \sin(\omega t), \quad (22.22)$$



so the associated potential energy is

$$U(t) = eE_0 z \sin(\omega t). \quad (22.23)$$

Turning this classical potential energy into a quantal operator gives

$$\hat{H}'(t) = eE_0 \hat{z} \sin(\omega t). \quad (22.24)$$

(Note that the hat  $\hat{k}$  in equation (22.22) signifies unit vector, whereas the hat  $\hat{z}$  in equation (22.24) signifies quantal operator. I'm sorry for any confusion. . . there just aren't enough symbols in the world to represent everything unambiguously!)

Now that we have the quantal operator for the perturbation, we can turn to the time-dependent perturbation theory result (22.11). (Is it legitimate to use perturbation theory in this case? See the problem.)

For all of the atomic energy states  $|a\rangle$  we've considered in this book,

$$H'_{a,a}(t) = \langle a | H'(t) | a \rangle = eE_0 \langle a | \hat{z} | a \rangle \sin(\omega t) = 0, \quad (22.25)$$

whence  $c_a(t) = 1$  and  $P_{a \rightarrow a} = 1$ . Most of the atoms *don't* make transitions.

But what about those that do? For these we need to find the matrix elements

$$H'_{b,a}(t) = \langle b | H'(t) | a \rangle = eE_0 \langle b | \hat{z} | a \rangle \sin(\omega t). \quad (22.26)$$

These are just the  $z_{b,a}$  matrix elements that we calculated for the Stark effect. (And after all, what we're considering here is just the Stark effect with an oscillating electric field.) The transition amplitudes are

$$c_b(t) = -\frac{i}{\hbar} eE_0 \langle b | \hat{z} | a \rangle \int_0^t \sin(\omega t') e^{+(i/\hbar)(E_b - E_a)t'} dt'. \quad (22.27)$$

It is convenient (and conventional!) to follow the lead of Einstein's  $\Delta E = \hbar\omega$  and define

$$E_b - E_a = \hbar\omega_0. \quad (22.28)$$

The time integral is then

$$\begin{aligned} & \int_0^t \sin(\omega t') e^{i\omega_0 t'} dt' \\ &= \int_0^t \frac{e^{+i\omega t'} - e^{-i\omega t'}}{2i} e^{i\omega_0 t'} dt' \\ &= \frac{1}{2i} \left[ \int_0^t e^{i(\omega_0 + \omega)t'} dt' - \int_0^t e^{i(\omega_0 - \omega)t'} dt' \right] \\ &= \frac{1}{2i} \left[ \frac{e^{i(\omega_0 + \omega)t'}}{i(\omega_0 + \omega)} - \frac{e^{i(\omega_0 - \omega)t'}}{i(\omega_0 - \omega)} \right]_0^t \\ &= -\frac{1}{2} \left[ \frac{e^{i(\omega_0 + \omega)t} - 1}{\omega_0 + \omega} - \frac{e^{i(\omega_0 - \omega)t} - 1}{\omega_0 - \omega} \right] \end{aligned} \quad (22.29)$$

Enrico Fermi thought about this expression and realized that in most cases it would not be substantial (as reflected in the fact that  $P_{a \rightarrow a} = 1$ ). The numerators are complex numbers in magnitude between 0 and 2. For light, we're thinking of frequencies  $\omega$  near ZZZ. The only case when this expression is big, is when  $\omega \approx \omega_0$ , and when that's true only the right-hand part is big. So it's legitimate to ignore the left-hand part and write

$$\begin{aligned}
 & \int_0^t \sin(\omega t') e^{i\omega_0 t'} dt' \\
 & \approx -\frac{1}{2} \left[ -\frac{e^{i(\omega_0 - \omega)t} - 1}{\omega_0 - \omega} \right] \\
 & = \frac{1}{2} \left[ \frac{e^{i(\omega_0 - \omega)t/2} e^{i(\omega_0 - \omega)t/2} - e^{-i(\omega_0 - \omega)t/2}}{\omega_0 - \omega} \right] \\
 & = \frac{1}{2} \left[ \frac{e^{i(\omega_0 - \omega)t/2} 2i \sin((\omega_0 - \omega)t/2)}{\omega_0 - \omega} \right] \\
 & = i e^{i(\omega_0 - \omega)t/2} \frac{\sin((\omega_0 - \omega)t/2)}{\omega_0 - \omega} \\
 & = i e^{-i(\omega - \omega_0)t/2} \frac{\sin((\omega - \omega_0)t/2)}{\omega - \omega_0}. \tag{22.30}
 \end{aligned}$$

Plugging this approximation for the integral into equation (22.27) produces

$$c_b(t) = \frac{eE_0 \langle b | \hat{z} | a \rangle}{\hbar} e^{-i(\omega - \omega_0)t/2} \frac{\sin((\omega - \omega_0)t/2)}{\omega - \omega_0}. \tag{22.31}$$

The transition probability is then

$$P_{a \rightarrow b} = \frac{e^2 E_0^2 |\langle b | \hat{z} | a \rangle|^2 \sin^2((\omega - \omega_0)t/2)}{\hbar^2 (\omega - \omega_0)^2}. \tag{22.32}$$

This rule, like all rules,<sup>2</sup> has limits on its applicability: we've already mentioned that it applies when the wavelength of light is much larger than an atom, when the light can be treated classically, when  $\omega \approx \omega_0$ , etc. Most importantly, it applies only when the transition probability is small, because when that probability is large the whole basis of perturbation theory breaks down. You might think that with all these restrictions, it's not a very important result. You'd be wrong. In fact Fermi used it so often that he called it "the golden rule."

<sup>2</sup>A father needs to leave his child at home for a short time. Concerned for his child's safety, he issues the sensible rule "Don't leave home while I'm away." While the father is away, the home catches fire. Should the child violate the rule?

### *Physical implications of Fermi's golden rule*

We have derived Fermi's golden rule, but that's only the start and not the end of our quest to answer the question of "How do atoms absorb light?". What does Fermi's golden rule say about nature? First, we'll think of the formula as a function of frequency  $\omega$  for fixed time  $t$ , then we'll think of the formula as a function of time  $t$  at fixed frequency  $\omega$ .

Write the transition probability as

$$P_{a \rightarrow b} = A \frac{\sin^2((\omega - \omega_0)t/2)}{(\omega - \omega_0)^2} \quad (22.33)$$

where the value of  $A$  is independent of both frequency and time. Clearly, this expression is always positive or zero (good thing!) and is symmetric about the natural transition frequency  $\omega_0$ . The expression is always less than the time-independent "envelope function"  $A/(\omega - \omega_0)^2$ . The transition probability vanishes when

$$\omega - \omega_0 = N\pi/t, \quad N = \pm 2, \pm 4, \pm 6, \dots$$

while it touches the envelope when

$$\omega - \omega_0 = N\pi/t, \quad N = \pm 1, \pm 3, \pm 5, \dots$$

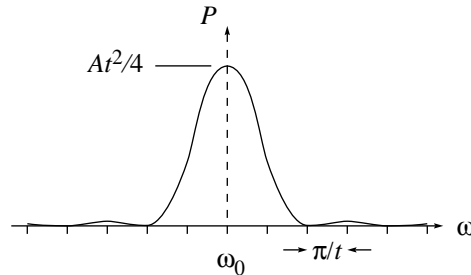
What about when  $\omega = \omega_0$ ? Here you may use l'Hôpital's rule, or the approximation

$$\sin \theta \approx \theta \quad \text{for } \theta \ll 1,$$

but either way you'll find that

$$\text{when } \omega = \omega_0, \quad P_{a \rightarrow b} = At^2/4. \quad (22.34)$$

In short, the transition probability as a function of  $\omega$  looks like this graph:



**Problem:** Show that if the central maximum has value  $P_{\max}$ , then the first touching of the envelope (at  $\omega - \omega_0 = \pi/t$ ) has value  $(4/\pi^2)P_{\max} = 0.405 P_{\max}$ , the second touching (at  $\omega - \omega_0 = 3\pi/t$ ) has value  $(4/9\pi^2)P_{\max} = 0.045 P_{\max}$ , and the third (at  $\omega - \omega_0 = 5\pi/t$ ) has value  $(4/25\pi^2)P_{\max} = 0.016 P_{\max}$ . Notice that these ratios are independent of time.

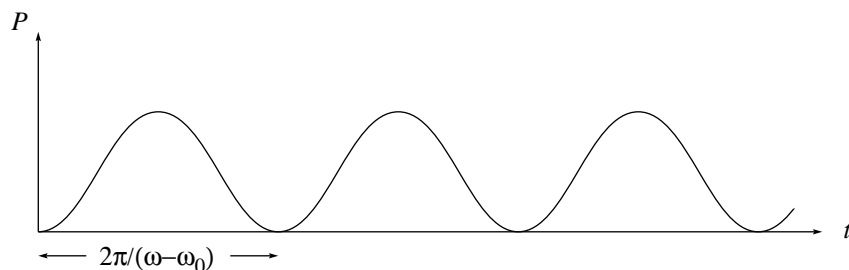
There are several unphysical elements of this graph it gives a result even at  $\omega = 0 \dots$  indeed, even when  $\omega$  is negative! But the formula was derived assuming  $\omega \approx \omega_0$ , so we don't expect it to give physically reasonable results in this regime. In time, the maximum transition probability  $At^2/4$  will grow to be very large, in fact even larger than one! But the formula was derived assuming a small transition probability, and becomes invalid long before such an absurdity happens.

This result may help you with a conundrum. You have perhaps been told something like: "To excite hydrogen from the ground state to the first excited state, a transition with  $\Delta E = \frac{1}{4}\text{Ry}$ , you must supply a photon with energy exactly equal to  $\frac{1}{4}\text{Ry}$ , what is with frequency  $\omega_0 = \frac{1}{4}\text{Ry}/\hbar$ , or in other words with wavelength 364.506 820 nm." You know that no laser produces light with the *exact* wavelength of 364.506 820 nm. If the photon had to have *exactly* that wavelength, there would almost never be a transition. But the laser doesn't need to have *exactly* that wavelength: as you can see, there's some probability of absorbing light that differs a bit from the natural frequency  $\omega_0$ .

**Problem:** Show that the width of the central peak, from zero to zero, is  $4\pi/t$ .

One aspect of the transition probability expression is quite natural: The light most effective at promoting a transition is light with frequency  $\omega$  equal to the transition's natural frequency  $\omega_0$ . Also natural is that the effectiveness decreases as  $\omega$  moves away from  $\omega_0$ , until the transition probability vanishes entirely at  $\omega = \omega_0 \pm 2\pi/t$ . But then a puzzling phenomenon sets in: as  $\omega$  moves still further away from  $\omega_0$ , the transition probability *increases*. This increase is admittedly slight, but nonetheless it exists, and I know of no way to explain it in physical terms. I do point out, however, that this puzzling phenomenon does *not* exist for light pulses of Gaussian form: see problem 22.5, "Gaussian light pulse".

Now, investigate the formula (22.33) as a function of time  $t$  at fixed light frequency  $\omega$ . This seems at first to be a much simpler task, because the graph is trivial:



But now reflect upon the graph. We have a laser set to make transitions from  $|a\rangle$  to  $|b\rangle$ . We turn on the laser, and the probability of that transition increases. So far, so good. Now we keep the laser on, but the probability decreases! And if we keep it on for exactly the right amount of time, there is zero probability for a transition. It's as if we were driving a nail into a board with a hammer. The first few strikes push the nail into the board, but with continued strikes the nail backs out of the board, and it eventually pops out altogether!

How can this be? Certainly, no nail that I've hammered has ever behaved this way! The point is that there are two routes to get from  $|a\rangle$  to  $|a\rangle$ : You can go from  $|a\rangle$  to  $|b\rangle$  and then back to  $|a\rangle$ , or you can stay always in  $|a\rangle$ , that is go from  $|a\rangle$  to  $|a\rangle$  to  $|a\rangle$ . There is an amplitude associated with each route. If these two amplitudes interfere constructively, there is a high probability of remaining in  $|a\rangle$  (a low probability of transitioning to  $|b\rangle$ ). If these two amplitudes interfere destructively, there is a low probability of remaining in  $|a\rangle$  (a high probability of transitioning to  $|b\rangle$ ). This wavy graph is a result of interference of two routes that are, not paths in position space, but routes through energy eigenstates.<sup>3</sup>

This phenomenon is called “Rabi oscillation”, and it's the pulse at the heart of an atomic clock.

<sup>3</sup>This point of view is developed extensively in R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals* (D.F. Styer, emending editor, Dover Publications, Mineola, New York, 2010) pages 116–117, 144–147.

## 22.4 Absorbing incoherent light

For coherent,  $z$ -polarized,  $x$ -directed, long-wavelength, non-magnetic, classical, non-diminishing light, in the approximation of first-order time-dependent perturbation theory, and with  $\omega \approx \omega_0$ , the transition probability is

$$P_{a \rightarrow b} = \frac{e^2 E_0^2}{\hbar^2} |\langle b | \hat{z} | a \rangle|^2 \frac{\sin^2((\omega - \omega_0)t/2)}{(\omega - \omega_0)^2}. \quad (22.35)$$

The classical energy density (average energy per volume) of an electromagnetic wave is  $u = \epsilon_0 E_0^2/2$ , where  $\epsilon_0$  is the famous vacuum permittivity that appears as  $1/(4\pi\epsilon_0)$  in Coulomb's law, so this result is often written

$$P_{a \rightarrow b} = \frac{2e^2 u}{\epsilon_0 \hbar^2} |\langle b | \hat{z} | a \rangle|^2 \frac{\sin^2((\omega - \omega_0)t/2)}{(\omega - \omega_0)^2}. \quad (22.36)$$

What if the light is polarized but not coherent? In this case light comes at varying frequencies. Writing the energy density per frequency as  $\rho(\omega)$ , the transition probability due to light of frequency  $\omega$  to  $\omega + d\omega$  is

$$P'_{a \rightarrow b} = \frac{2e^2 \rho(\omega) d\omega}{\epsilon_0 \hbar^2} |\langle b | \hat{z} | a \rangle|^2 \frac{\sin^2((\omega - \omega_0)t/2)}{(\omega - \omega_0)^2}, \quad (22.37)$$

whence the total transition probability is

$$P_{a \rightarrow b} = \frac{2e^2}{\epsilon_0 \hbar^2} |\langle b | \hat{z} | a \rangle|^2 \int_0^\infty \frac{\sin^2((\omega - \omega_0)t/2)}{(\omega - \omega_0)^2} \rho(\omega) d\omega. \quad (22.38)$$

[[We have assumed that the light components at various frequencies is independent, so that the total transition probability is the sum of the individual transition probabilities. If instead the light components were completely correlated, then the total transition *amplitude* would be the sum of the individual transition *amplitudes*. This is the case in problem 22.5, "Gaussian light pulse". If the light components were incompletely correlated but not completely independent, then a hybrid approach would be needed.]] If  $\rho(\omega)$  is slowly varying relative to the absorption profile (22.33) — which it almost always is — then it is accurate to approximate

$$P_{a \rightarrow b} = \frac{2e^2}{\epsilon_0 \hbar^2} |\langle b | \hat{z} | a \rangle|^2 \rho(\omega_0) \int_{-\infty}^{+\infty} \frac{\sin^2((\omega - \omega_0)t/2)}{(\omega - \omega_0)^2} d\omega, \quad (22.39)$$

where I have changed the lower integration limit from 0 to  $-\infty$ , with negligible change in  $P_{a \rightarrow b}$ , because the integrand nearly vanishes whenever  $\omega < 0$ . Finally, the definite integral

$$\int_{-\infty}^{+\infty} \frac{\sin^2 x}{x^2} dx = \pi$$

gives, for polarized incoherent light,

$$P_{a \rightarrow b} = \frac{\pi e^2}{\epsilon_0 \hbar^2} |\langle b | \hat{z} | a \rangle|^2 \rho(\omega_0) t. \quad (22.40)$$

The primary thing to note about this formula is the absence of Rabi oscillations: it gives a far more familiar *rate* of transition. The second thing is that the rate from  $|b\rangle$  to  $|a\rangle$  is equal to the rate from  $|a\rangle$  to  $|b\rangle$ , which is somewhat unusual: you might think that the rate to lose energy ( $|b\rangle$  to  $|a\rangle$ ) should be greater than the rate to gain energy ( $|a\rangle$  to  $|b\rangle$ ). [Just as it's easier to walk down a staircase than up the same staircase.]

Finally, what if the light is not coherent, not polarized, and not directed? (Such as the light in a room, that comes from all directions.) In this case

$$P_{a \rightarrow b} = \frac{\pi e^2}{3\epsilon_0 \hbar^2} [|\langle b | \hat{x} | a \rangle|^2 + |\langle b | \hat{y} | a \rangle|^2 + |\langle b | \hat{z} | a \rangle|^2] \rho(\omega_0) t. \quad (22.41)$$

## 22.5 Absorbing and emitting light

### *Qualitative quantum electrodynamics*

Of course we want to do better than the treatment above: Instead of treating a quantum mechanical atom immersed in a classical electromagnetic field, we want a full quantum-mechanical treatment of the atom *and* the light. Such a theory — quantum electrodynamics — has been developed and it is a beautiful thing. Because light must travel at speed  $c$  this theory is intrinsically relativistic and, while beautiful, also a very difficult thing. We will not give it a rigorous treatment in this book. But this section motivates the theory and discusses its qualitative character.

Most of this book discusses the quantum mechanics of atoms: The Hamiltonian operator  $\hat{H}_{\text{atom}}$  has energy eigenstates like the ground state  $|a\rangle$  and the excited state  $|b\rangle$ . The system can exist in any linear combination of these states, such as  $(|a\rangle - |b\rangle)/\sqrt{2}$ . If the system starts off in one of the energy states, including the excited state  $|b\rangle$ , it stays there forever.

You can also write down a Hamiltonian operator  $\hat{H}_{\text{EM}}$  for the electromagnetic field. This operator has energy eigenstates. By convention, the ground state is called  $|\text{vacuum}\rangle$ , one excited state is called  $|1 \text{ photon}\rangle$ , an even more excited state is called  $|2 \text{ photons}\rangle$ . The field can also exist in

linear combinations such as  $(|\text{vacuum}\rangle - |2 \text{ photons}\rangle)/\sqrt{2}$ , but this state is *not* a stationary state, and it does *not* have an energy.

You can do the classic things with field energy states: There's an operator for energy and an operator for photon position, but they don't commute. So in the state  $|1 \text{ photon}\rangle$  the photon has an energy but no position. There's a linear combinations of energy states in which the photon *does* have a position, but in these position states the electromagnetic field has no energy.

But there's even more: There is an operator for electric field at a given location. And this operator doesn't commute with either the Hamiltonian or with the photon position operator.<sup>4</sup> So in a state of electric field at some given point, the photon does not have a position, and does not have an energy. Anyone thinking of the photon as a "ball of light" — a wavepacket of electric and magnetic fields — is thinking of a misconception. A photon might have a "pretty well defined" position and a "pretty well defined" energy and a "pretty well defined" field, but it can't have an exact position and an exact energy and an exact field at the same time.

If the entire Hamiltonian were  $\hat{H}_{\text{atom}} + \hat{H}_{\text{EM}}$ , then energy eigenstates of the atom plus field would have the character of  $|a\rangle|2 \text{ photons}\rangle$ , or  $|b\rangle|\text{vacuum}\rangle$  and if you started off in such a state you would stay in it forever. Note particularly the second example: if the atom started in an excited state, it would never decay to the ground state, emitting light.

But since that process (called "spontaneous emission") *does* happen, the Hamiltonian  $\hat{H}_{\text{atom}} + \hat{H}_{\text{EM}}$  *must not* be the whole story. There must be some additional term in the Hamiltonian that involves both the atom and the field: This term is called the "interaction Hamiltonian"  $\hat{H}_{\text{int}}$ . (Sometimes called the "coupling Hamiltonian", because it couples — connects — the atom and the field.) The full Hamiltonian is  $\hat{H}_{\text{atom}} + \hat{H}_{\text{EM}} + \hat{H}_{\text{int}}$ . The state  $|b\rangle|\text{vacuum}\rangle$  is *not* an eigenstate of this full Hamiltonian: If you start off in  $|b\rangle|\text{vacuum}\rangle$ , then at a later time there will be some amplitude to remain in  $|b\rangle|\text{vacuum}\rangle$ , but also some amplitude to be in  $|a\rangle|1 \text{ photon}\rangle$ .

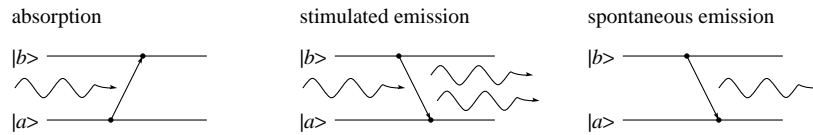
---

<sup>4</sup>It's clear, even without writing down the "EM field Hamiltonian" and the "electric field at a given point" operators, that they do not commute: any operator that commutes with the Hamiltonian is conserved, so if these two operators commuted then the electric field at a given point would never change with time!



**Einstein A and B argument**

Back in 1916, Einstein wanted to know about both absorption *and* emission of light by atoms, and — impatient as always — he didn't want to wait until a full theory of quantum electrodynamics was developed. So he came up with the following argument — one of the cleverest in all of physics.



Einstein said that there were three processes going on, represented schematically in the figure above. In *absorption* of radiation the atom starts in its ground state  $|a\rangle$  and ends in excited state  $|b\rangle$ , while the light intensity at frequency  $\omega_0$  is reduced. Although the reasoning leading to equation (22.41) hadn't yet been performed in 1916, Einstein thought it reasonable that the probability of absorption would be given by some rate coefficient  $B_{ab}$ , times the energy density of radiation with the proper frequency for exciting the atom, times the time:

$$P_{a \rightarrow b} = B_{ab} \rho(\omega_0) t. \quad (22.42)$$

In *stimulated emission* the atom starts in excited state  $|b\rangle$  and, under the influence of light, ends in ground state  $|a\rangle$ . After this happens the light intensity at frequency  $\omega_0$  increases due to the emitted light. In this process the incoming light of frequency  $\omega_0$  “shakes” the atom out of its excited state. Einstein thought the probability for this process would be

$$P_{b \rightarrow a} = B_{ba} \rho(\omega_0) t. \quad (22.43)$$

We know, from equation (22.41), that in fact  $B_{ba} = B_{ab}$ , but Einstein didn't know this so his argument doesn't use this fact.

Finally, in *spontaneous emission* the atom starts in excited state  $|b\rangle$  and ends in ground state  $|a\rangle$ , but it does so without any incoming light to “shake” it. After spontaneous emission the light intensity at frequency  $\omega_0$  increases due to the emitted light. Because this process doesn't rely on incoming light, the probability of it happening doesn't depend on  $\rho(\omega_0)$ . Instead, Einstein thought, the probability would be simply

$$P'_{b \rightarrow a} = At. \quad (22.44)$$

Einstein knew that this process had to happen, because excited atoms in the dark can give off light and go to their ground state, but he didn't have a theory of quantum electrodynamics that would enable him to calculate the rate coefficient  $A$ .

The coefficients  $B_{ab}$ ,  $B_{ba}$ , and  $A$  are independent of the properties of the light, the number of atoms in state  $|a\rangle$ , the number of atoms in state  $|b\rangle$ , etc. — they depend only upon the characteristics of the atom.

Now if you have a bunch of atoms, with  $N_a$  of them in the ground state and  $N_b$  in the excited state, the rate of change of  $N_a$  through these three processes is

$$\frac{dN_a}{dt} = -B_{ab} \rho(\omega_0) N_a + B_{ba} \rho(\omega_0) N_b + AN_b. \quad (22.45)$$

In equilibrium, by definition,

$$\frac{dN_a}{dt} = 0. \quad (22.46)$$

In addition, in thermal equilibrium at temperature  $T$ , the following two facts are true: The first is called “Boltzmann distribution”

$$\frac{N_b}{N_a} = e^{-(E_b - E_a)/k_B T} = e^{-\hbar\omega_0/k_B T}, \quad (22.47)$$

where  $k_B$  is the so-called “Boltzmann constant” that arises frequently in thermal physics. The second is called “energy density for light in thermal equilibrium (backbody radiation)”

$$\rho(\omega) = \frac{\hbar}{\pi^2 c^3} \frac{\omega^3}{e^{\hbar\omega/k_B T} - 1}, \quad (22.48)$$

where  $c$  is the speed of light. [If you have taken a course in statistical mechanics, you have certainly seen the first result. You might think you haven't seen the second result, but in fact it is a property of the ideal Bose gas when the chemical potential  $\mu$  vanishes.]

You might not yet know these two facts, but Einstein did. He combined equation (22.46) and equation (22.45) finding

$$\rho(\omega_0) = \frac{AN_b}{B_{ab}N_a - B_{ba}N_b}.$$

Then he used the Boltzmann distribution (22.47) to produce

$$\rho(\omega_0) = \frac{A}{B_{ab}e^{\hbar\omega_0/k_B T} - B_{ba}} \quad (22.49)$$

and compared that to the blackbody result (22.48) producing

$$\frac{A}{B_{ab}e^{\hbar\omega_0/k_B T} - B_{ba}} = \frac{\hbar}{\pi^2 c^3} \frac{\omega_0^3}{e^{\hbar\omega_0/k_B T} - 1}.$$

This result must hold for all temperatures  $T$ , and the coefficients  $B_{ab}$ ,  $B_{ba}$ , and  $A$  are independent of  $T$ . Thus, Einstein reasoned, we must have

$$B_{ab} = B_{ba} \equiv B \quad (22.50)$$

(which we already knew, but which was a discovery to Einstein) and hence

$$\frac{A}{B(e^{\hbar\omega_0/k_B T} - 1)} = \frac{\hbar}{\pi^2 c^3} \frac{\omega_0^3}{e^{\hbar\omega_0/k_B T} - 1}$$

or, with temperature-dependent parts canceling on both sides,

$$\frac{A}{B} = \frac{\hbar\omega_0^3}{\pi^2 c^3}. \quad (22.51)$$

The result is, of necessity, independent of temperature  $T$ . Einstein's argument uses thermal equilibrium not to discover the macroscopic properties of matter, but as a vehicle to uncover microscopic details about the relation between matter and radiation. We have no way to find  $A$  from first principles, but from the fact that thermal equilibrium exists we can find  $A$  through

$$A = \frac{\hbar\omega_0^3}{\pi^2 c^3} B = \frac{4h}{\lambda_0^3} B. \quad (22.52)$$

I hope you find this argument as astounding, and as beautiful, as I do. It has the character of Einstein: First, it is not technically difficult, but it combines the various elements in a way that I never would have thought of, to produce a result that I thought would require working out full theory of quantum electrodynamics. Second, it turns the problem on its head: The fundamental question is "Will microscopic actions always result in macroscopic thermal equilibrium? If so, how fast will that equilibrium be approached?" Einstein skips over the fundamental question and asks "We know from observation that macroscopic thermal equilibrium does in fact exist. How can we exploit this fact to find out about microscopic actions?"

**Numerical example:** I would expect the stimulated decay rate  $B\rho(\omega_0)$  to exceed the spontaneous emission rate  $A$  (just as a jar on a shelf is more likely to fall off when shaken than when left alone). On the other hand I've found my expectations violated by quantum mechanics so frequently that I can't be sure. What is the ratio of  $A$  to  $B\rho(\omega_0)$  at room temperature for

the transition associated with the red light of a Helium-Neon laser ( $\lambda_0 = 633 \text{ nm}$ )?

Use equation (22.49) to write

$$\frac{B\rho(\omega_0)}{A} = \frac{1}{e^{\hbar\omega_0/k_B T} - 1}. \quad (22.53)$$

Now at room temperature,  $k_B T = \frac{1}{40} \text{ eV}$ , so

$$\frac{\hbar\omega_0}{k_B T} = \frac{hc}{\lambda_0 k_B T} = \frac{1240 \text{ eV}\cdot\text{nm}}{(633 \text{ nm})(\frac{1}{40} \text{ eV})} = 78$$

resulting in

$$\frac{B\rho(\omega_0)}{A} = \frac{1}{e^{78} - 1} = e^{-78} = 10^{-34}.$$

My intuition about shaking has been vindicated! At what temperature will the stimulated and spontaneous rates be equal?

## 22.6 Problems

### 22.1 On being kicked upstairs

A particle in the ground state of an infinite square well is perturbed by a transient effect described by the Hamiltonian (in coordinate representation)

$$H'(x; t) = A_0 \sin\left(\frac{2\pi x}{L}\right) \delta(t), \quad (22.54)$$

where  $A_0$  is a constant with the dimensions of action. What is the probability that after this jolt an energy measurement will find the system in the first excited state?

### 22.2 Second-order time-dependent perturbation theory

At equation (22.16) we treated, to first order in perturbation theory, the problem of a simple harmonic oscillator in its ground state exposed to a sinusoidal external force (with frequency  $\omega$  and amplitude  $eE_0$ ). We concluded that the only non-vanishing first-order transition amplitudes were  $c_0^{(1)}(t) = 1$  and  $c_1^{(1)}(t)$ . (Here the superscript (1) denotes “first-order”.) Show that to second order the non-vanishing transition amplitudes are:

$$c_0^{(2)}(t) = 1 - \frac{i}{\hbar} \int_0^t H'_{01}(t') e^{-i\omega_0 t'} c_1^{(1)}(t') dt', \quad (22.55)$$

$$c_1^{(2)}(t) = -\frac{i}{\hbar} \int_0^t H'_{10}(t') e^{+i\omega_0 t'} c_0^{(1)}(t') dt', \quad (22.56)$$

$$c_2^{(2)}(t) = -\frac{i}{\hbar} \int_0^t H'_{21}(t') e^{+i\omega_0 t'} c_1^{(1)}(t') dt', \quad (22.57)$$

where

$$H'_{01}(t) = H'_{10}(t) = eE_0 \sqrt{\frac{\hbar}{2m\omega_0}} \sin(\omega t), \quad (22.58)$$

and

$$H'_{21}(t) = eE_0 \sqrt{\frac{2\hbar}{2m\omega_0}} \sin(\omega t). \quad (22.59)$$

The integrals for  $c_0^{(2)}(t)$  and  $c_2^{(2)}(t)$  are not worth working out, but it is worth noticing that  $c_2^{(2)}(t)$  involves a factor of  $(eE_0)^2$  (where  $eE_0$  is in some sense “small”), and that  $c_1^{(2)}(t) = c_1^{(1)}(t)$ .

### 22.3 Is light a perturbation?

Is it legitimate to use perturbation theory in the case of light absorbed by an atom? After all, we’re used to thinking of the light from a powerful laser as a big effect, not a tiny perturbation. However, whether an effect is big or small depends on context. Estimate the maximum electric field due to a laser of XX watts, and the electric field at an electron due to its nearby nucleus. Conclude that while the laser is very powerful on a human scale (and you should *not* stick your eye into a laser beam), it is nevertheless very weak on an atomic scale.

### 22.4 Magnitude of transitions

At equation (22.33) we defined

$$A \equiv \frac{e^2 E_0^2 |\langle b | \hat{z} | a \rangle|^2}{\hbar^2}$$

and then noted that it was independent of  $\omega$  and  $t$ , but otherwise ignored it. (Although we used it when we said that the maximum transition probability was  $At^2/4$ .) This problem investigates the character of  $A$ .

The maximum classical force on the electron due to light is  $eE_0$ . A typical force is less, so define the characteristic force due to light as

$$F_{c,L} \equiv \frac{1}{2} eE_0.$$

A typical classical force on the electron due to the nucleus is

$$F_{c,N} \equiv \left( \frac{e^2}{4\pi\epsilon_0} \right) \frac{1}{a_0^2}.$$

Using these two definitions, and taking a typical matrix element  $|\langle b | \hat{z} | a \rangle|$  to be  $a_0$ , show that a typical value of  $A$  is

$$4 \left( \frac{F_{c,L}}{F_{c,N}} \right)^2 \frac{1}{\tau_0^2}.$$

If this excites you, you may also show that the exact value is

$$A = 4 \left( \frac{F_{c,L}}{F_{c,N}} \right)^2 \frac{1}{\tau_0^2} \left| \frac{\langle b | \hat{z} | a \rangle}{a_0} \right|^2.$$

### 22.5 Gaussian light pulse

An atom is exposed to a Gaussian packet of light

$$E(t) = E_0 e^{-t^2/\tau^2} \sin(\omega t). \quad (22.60)$$

At time  $t = -\infty$ , the atom was in state  $|a\rangle$ . Find the amplitude, to first order in perturbation theory, that at time  $t = \infty$  the atom is in state  $|b\rangle$ . *Clue:* Use

$$\int_{-\infty}^{+\infty} e^{ax^2+bx} dx = \sqrt{\frac{\pi}{-a}} e^{-b^2/4a} \quad \text{for } \Re\{a\} \leq 0 \text{ but } a \neq 0.$$

*Answer:*

$$c_b = \frac{eE_0 \langle b|\hat{z}|a\rangle}{\hbar} \left[ -\frac{\sqrt{\pi}}{2\tau} \right] \left[ e^{-\tau^2(\omega+\omega_0)/4} + e^{-\tau^2(\omega-\omega_0)/4} \right].$$

## Chapter 23

# The Territory Ahead

I reckon I got to light out for the territory ahead...

— Mark Twain (last sentence of *Huckleberry Finn*)

This is the last chapter of the book, but not the last chapter of quantum mechanics. There are many fascinating topics that this book hasn't even touched on. Quantum mechanics will — if you allow it — surprise and delight (and mystify) you for the rest of your life.

How to extend what's in this book:

- Relativistic quantum mechanics. (Don't make  $t$  an operator, instead turn  $x$  back to a variable and introduce creation and annihilation operators.)
- Quantum field theory.
- Quantal chaos and the classical limit of quantum mechanics.
- Friction and decay to ground state.
- Atomic, molecular, and solid state physics.

All of these fall solidly within the amplitude framework!





## Appendix A

# Tutorial on Matrix Diagonalization

You know from as far back as your introductory mechanics course that some problems are difficult given one choice of coordinate axes and easy or even trivial given another. (For example, the famous “monkey and hunter” problem is difficult using a horizontal axis, but easy using an axis stretching from the hunter to the monkey.) The mathematical field of linear algebra is devoted, in large part, to systematic techniques for finding coordinate systems that make problems easy. This tutorial introduces the most valuable of these techniques. It assumes that you are familiar with matrix multiplication and with the ideas of the inverse, the transpose, and the determinant of a square matrix. It is also useful to have a nodding acquaintance with the inertia tensor.

This presentation is intentionally non-rigorous. A rigorous, formal treatment of matrix diagonalization can be found in any linear algebra textbook,<sup>1</sup> and there is no need to duplicate that function here. What is provided here instead is a heuristic picture of what’s going on in matrix diagonalization, how it works, and why anyone would want to do such a thing anyway. Thus this presentation complements, rather than replaces, the logically impeccable (“bulletproof”) arguments of the mathematics texts.

Essential problems in this tutorial are marked by asterisks (\*).

### A.1 What’s in a name?

There is a difference between an entity and its name. For example, a tree is made of wood, whereas its name “tree” made of ink. One way to see this is to note that in German, the name for a tree is “Baum”, so the name

changes upon translation, but the tree itself does not change. (Throughout this tutorial, the term “translate” is used as in “translate from one language to another” rather than as in “translate by moving in a straight line”.)

The same holds for mathematical entities. Suppose a length is represented by the number “2” because it is two feet long. Then the same length is represented by the number “24” because it is twenty-four inches long. The same length is represented by two different numbers, just as the same tree has two different names. The representation of a length as a number depends not only upon the length, but also upon the coordinate system used to measure the length.

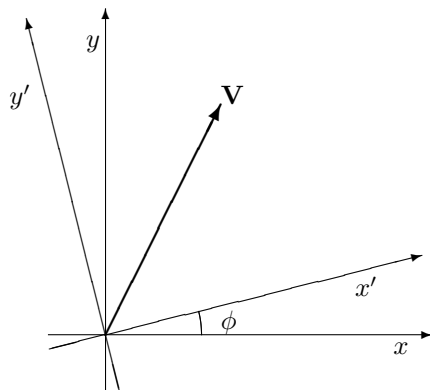
## A.2 Vectors in two dimensions

One way of describing a two-dimensional vector  $\mathbf{V}$  is by giving its  $x$  and  $y$  components in the form of a  $2 \times 1$  column matrix

$$\begin{pmatrix} V_x \\ V_y \end{pmatrix}. \quad (\text{A.1})$$

Indeed it is sometimes said that the vector  $\mathbf{V}$  is *equal* to the column matrix (A.1). This is not precisely correct—it is better to say that the vector is *described by* the column matrix or *represented by* the column matrix or that its *name* is the column matrix. This is because if you describe the vector using a different set of coordinate axes you will come up with a different column matrix to describe the same vector. For example, in the situation shown below the descriptions in terms of the two different coordinate systems are related through the matrix equation

$$\begin{pmatrix} V_{x'} \\ V_{y'} \end{pmatrix} = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} V_x \\ V_y \end{pmatrix}. \quad (\text{A.2})$$



The  $2 \times 2$  matrix above is called the “rotation matrix” and is usually denoted by  $\mathbf{R}(\phi)$ :

$$\mathbf{R}(\phi) \equiv \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}. \quad (\text{A.3})$$

One interesting property of the rotation matrix is that it is always invertible, and that its inverse is equal to its transpose. Such matrices are called *orthogonal*.<sup>1</sup> You could prove this by working a matrix multiplication, but it is easier to simply realize that the inverse of a rotation by  $\phi$  is simply a rotation by  $-\phi$ , and noting that

$$\mathbf{R}^{-1}(\phi) = \mathbf{R}(-\phi) = \mathbf{R}^\dagger(\phi). \quad (\text{A.4})$$

(The dagger represents matrix transposition.)

There are, of course, an infinite number of column matrix representations for any vector, corresponding to the infinite number of coordinate axis rotations with  $\phi$  from 0 to  $2\pi$ . But one of these representations is special: It is the one in which the  $x'$ -axis lines up with the vector, so the column matrix representation is just

$$\begin{pmatrix} V \\ 0 \end{pmatrix}, \quad (\text{A.5})$$

<sup>1</sup>Although all rotation matrices are orthogonal, there are orthogonal matrices that are not rotation matrices: see problem A.4.

where  $V = |\mathbf{V}| = \sqrt{V_x^2 + V_y^2}$  is the magnitude of the vector. This set of coordinates is the preferred (or “canonical”) set for dealing with this vector: one of the two components is zero, the easiest number to deal with, and the other component is a physically important number. You might wonder how I can claim that this representation has full information about the vector: The initial representation (A.1) contains two independent numbers, whereas the preferred representation (A.5) contains only one. The answer is that the preferred representation contains one number (the magnitude of the vector) explicitly while another number (the polar angle of the vector relative to the initial  $x$ -axis) is contained implicitly in the rotation needed to produce the preferred coordinate system.

**A.1 Problem: Right angle rotations**

Verify equation (A.2) in the special cases  $\phi = 90^\circ$ ,  $\phi = 180^\circ$ ,  $\phi = 270^\circ$ , and  $\phi = 360^\circ$ .

**A.2 Problem: The rotation matrix**

- Derive equation (A.2) through purely geometrical arguments.
- Express  $\hat{\mathbf{i}}'$  and  $\hat{\mathbf{j}}'$ , the unit vectors of the  $(x', y')$  coordinate system, as linear combinations of  $\hat{\mathbf{i}}$  and  $\hat{\mathbf{j}}$ . Then use

$$V_{x'} = \mathbf{V} \cdot \hat{\mathbf{i}}' \quad \text{and} \quad V_{y'} = \mathbf{V} \cdot \hat{\mathbf{j}}' \quad (\text{A.6})$$

to derive equation (A.2).

- Which derivation do you find easier?

**A.3 Problem: Rotation to the preferred coordinate system\***

In the preferred coordinate system,  $V_{y'} = 0$ . Use this requirement to show that the preferred system is rotated from the initial system by an angle  $\phi$  with

$$\tan \phi = \frac{V_y}{V_x}. \quad (\text{A.7})$$

For any value of  $V_y/V_x$ , there are *two* angles that satisfy this equation. What is the representation of  $\mathbf{V}$  in each of these two coordinate systems?

**A.4 Problem: A non-rotation orthogonal transformation**

In one coordinate system the  $y$ -axis is vertical and the  $x$ -axis points to the right. In another the  $y'$ -axis is vertical and the  $x'$ -axis points to the left. Find the matrix that translates vector coordinates from one system to the other. Show that this matrix is orthogonal but not a rotation matrix.

**A.5 Problem: Other changes of coordinate\***

Suppose vertical distances (distances in the  $y$  direction) are measured in feet while horizontal distances (distances in the  $x$  direction) are measured in miles. (This system is not perverse. It is used in nearly all American road maps.) Find the matrix that changes the representation of a vector in this coordinate system to the representation of a vector in a system where all distances are measured in feet. Find the matrix that translates back. Are these matrices orthogonal?

**A.6 Problem: Other special representations**

At equation (A.5) we mentioned one “special” (or “canonical”) representation of a vector. There are three others, namely

$$\begin{pmatrix} 0 \\ -V \end{pmatrix}, \quad \begin{pmatrix} -V \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ V \end{pmatrix}. \quad (\text{A.8})$$

If coordinate-system rotation angle  $\phi$  brings the vector representation into the form (A.5), then what rotation angle will result in these three representations?

**A.3 Tensors in two dimensions**

A tensor, like a vector, is a geometrical entity that may be described (“named”) through components, but a  $d$ -dimensional tensor requires  $d^2$  rather than  $d$  components. Tensors are less familiar and more difficult to visualize than vectors, but they are neither less important nor “less physical”. We will introduce tensors through the concrete example of the inertia tensor of classical mechanics (see, for example, reference [2]), but the results we present will be perfectly general.

Just as the two components of a two-dimensional vector are most easily kept track of through a  $2 \times 1$  matrix, so the four components of two-dimensional tensor are most conveniently written in the form of a  $2 \times 2$  matrix. For example, the inertia tensor  $\mathbf{T}$  of a point particle with mass  $m$  located<sup>2</sup> at  $(x, y)$  has components

$$\mathbf{T} = \begin{pmatrix} my^2 & -mxy \\ -mxy & mx^2 \end{pmatrix}. \quad (\text{A.9})$$

(Note the distinction between the tensor  $\mathbf{T}$  and its matrix of components, its “name”,  $T$ .) As with vector components, the tensor components are

<sup>2</sup>Or, to be absolutely precise, the particle located at the point represented by the vector with components  $(x, y)$ .

different in different coordinate systems, although the tensor itself does not change. For example, in the primed coordinate system of the figure on page 353, the tensor components are of course

$$\mathbf{T}' = \begin{pmatrix} my'^2 & -mx'y' \\ -mx'y' & mx'^2 \end{pmatrix}. \quad (\text{A.10})$$

A little calculation shows that the components of the inertia tensor in two different coordinate systems are related through

$$\mathbf{T}' = \mathbf{R}(\phi)\mathbf{T}\mathbf{R}^{-1}(\phi). \quad (\text{A.11})$$

This relation holds for *any* tensor, not just the inertia tensor. (In fact, one way to define “tensor” is as an entity with four components that satisfy the above relation under rotation.) If the matrix representing a tensor is symmetric (i.e. the matrix is equal to its transpose) in one coordinate system, then it is symmetric in all coordinate systems (see problem A.7). Therefore the symmetry is a property of the tensor, not of its matrix representation, and we may speak of “a symmetric tensor” rather than just “a tensor represented by a symmetric matrix”.

As with vectors, one of the many matrix representations of a given tensor is considered special (or “canonical”): It is the one in which the lower left component is zero. Furthermore if the tensor is symmetric (as the inertia tensor is) then in this preferred coordinate system the upper right component will be zero also, so the matrix will be all zeros except for the diagonal elements. Such a matrix is called a “diagonal matrix” and the process of finding the rotation that renders the matrix representation of a symmetric tensor diagonal is called “diagonalization”.<sup>3</sup> We may do an “accounting of information” for this preferred coordinate system just as we did with vectors. In the initial coordinate system, the symmetric tensor had three independent components. In the preferred system, it has two independent components manifestly visible in the diagonal matrix representation, and one number hidden through the specification of the rotation.

#### A.7 Problem: Representations of symmetric tensors\*

Show that if the matrix  $\mathbf{S}$  representing a tensor is symmetric, and if  $\mathbf{B}$  is any orthogonal matrix, then all of the representations

$$\mathbf{B}\mathbf{S}\mathbf{B}^\dagger \quad (\text{A.12})$$

---

<sup>3</sup>An efficient algorithm for diagonalization is discussed in section A.8. For the moment, we are more interested in knowing that a diagonal matrix representation must exist than in knowing how to most easily find that preferred coordinate system.

are symmetric. (Clue: If you try to solve this problem for rotations in two dimensions using the explicit rotation matrix (A.3), you will find it solvable but messy. The clue is that this problem asks you to prove the result in any number of dimensions, and for any orthogonal matrix  $\mathbf{B}$ , not just rotation matrices. This more general problem is considerably easier to solve.)

**A.8 Problem: Diagonal inertia tensor**

The matrix (A.9) represents the inertia tensor of a point particle with mass  $m$  located a distance  $r$  from the origin. Show that the matrix is diagonal in four different coordinate systems: one in which the  $x'$ -axis points directly toward the particle, one in which the  $y'$ -axis points directly away from the particle, one in which the  $x'$ -axis points directly away from the particle, and one in which the  $y'$ -axis points directly toward the particle. Find the matrix representation in each of these four coordinate systems.

**A.9 Problem: Representations of a certain tensor**

Show that a tensor represented in one coordinate system by a diagonal matrix with equal elements, namely

$$\begin{pmatrix} d_0 & 0 \\ 0 & d_0 \end{pmatrix}, \quad (\text{A.13})$$

has the same representation in *all* orthogonal coordinate systems.

**A.10 Problem: Rotation to the preferred coordinate system\***

A tensor is represented in the initial coordinate system by

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}. \quad (\text{A.14})$$

Show that the tensor is diagonal in a preferred coordinate system which is rotated from the initial system by an angle  $\phi$  with

$$\tan(2\phi) = \frac{2b}{a-c}. \quad (\text{A.15})$$

This equation has four solutions. Find the rotation matrix for  $\phi = 90^\circ$ , then show how the four different diagonal representations are related. You do not need to find any of the diagonal representations in terms of  $a$ ,  $b$  and  $c$ . . . just show what the other three are given that one of them is

$$\begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}. \quad (\text{A.16})$$

### A.11 Problem: Inertia tensor in outer product notation

The discussion in this section has emphasized the tensor's matrix representation ("name")  $\mathbb{T}$  rather than the tensor  $\mathbf{T}$  itself.

- a. Define the "identity tensor"  $\mathbf{1}$  as the tensor represented in some coordinate system by

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (\text{A.17})$$

Show that this tensor has the same representation in any coordinate system.

- b. Show that the inner product between two vectors results in a scalar: Namely

if vector  $bfa$  is represented by  $\begin{pmatrix} a_x \\ a_y \end{pmatrix}$  and vector  $bf b$  is represented by  $\begin{pmatrix} b_x \\ b_y \end{pmatrix}$

then the inner product  $\mathbf{a} \cdot \mathbf{b}$  is given through

$$\begin{pmatrix} a_x & a_y \end{pmatrix} \begin{pmatrix} b_x \\ b_y \end{pmatrix} = a_x b_x + a_y b_y,$$

and this inner product is a scalar. (A  $1 \times 2$  matrix times a  $2 \times 1$  matrix is a  $1 \times 1$  matrix.) That is, the vector  $\mathbf{a}$  is represented by different coordinates in different coordinate systems, and the vector  $\mathbf{b}$  is represented by different coordinates in different coordinate systems, but the inner product  $\mathbf{a} \cdot \mathbf{b}$  is the same in all coordinate systems.

- c. In contrast, show that the outer product of two vectors is a tensor: Namely

$$\mathbf{ab} \doteq \begin{pmatrix} a_x \\ a_y \end{pmatrix} \begin{pmatrix} b_x & b_y \end{pmatrix} = \begin{pmatrix} a_x b_x & a_x b_y \\ a_y b_x & a_y b_y \end{pmatrix}.$$

(A  $2 \times 1$  matrix times a  $1 \times 2$  matrix is a  $2 \times 2$  matrix.) That is, show that the representation of  $\mathbf{ab}$  transforms from one coordinate system to another as specified through (A.11).

- d. Show that the inertia tensor for a single particle of mass  $m$  located at position  $\mathbf{r}$  can be written in coordinate-independent fashion as

$$\mathbf{T} = m\mathbf{1}r^2 - m\mathbf{r}\mathbf{r}. \quad (\text{A.18})$$



## A.4 Tensors in three dimensions

A three-dimensional tensor is represented in component form by a  $3 \times 3$  matrix with nine entries. If the tensor is symmetric, there are six independent elements. . . three on the diagonal and three off-diagonal. The components of a tensor in three dimensions change with coordinate system according to

$$\mathbf{T}' = \mathbf{R}\mathbf{T}\mathbf{R}^\dagger, \quad (\text{A.19})$$

where  $\mathbf{R}$  is the  $3 \times 3$  rotation matrix.

A rotation in two dimension is described completely by giving a single angle. In three dimensions more information is required. Specifically, we need not only the amount of the rotation, but we must also know the plane in which the rotation takes place. We can specify the plane by giving the unit vector perpendicular to that plane. Specifying an arbitrary vector in three dimensions requires three numbers, but specifying a unit vector in three dimensions requires only two numbers because the magnitude is already fixed at unity. Thus three numbers are required to specify a rotation in three dimensions: two to specify the rotation's plane, one to specify the rotation's size. (One particularly convenient way to specify a three-dimensional rotation is through the three *Euler angles*. Reference [3] defines these angles and shows how to write the  $3 \times 3$  rotation matrix in terms of these variables. For the purposes of this tutorial, however, we will not need an explicit rotation matrix. . . all we need is to know is the number of angles required to specify a rotation.)

In two dimensions, any symmetric tensor (which has three independent elements), could be represented by a diagonal tensor (with two independent elements) plus a rotation (one angle). We were able to back up this claim with an explicit expression for the angle.

In three dimensions it seems reasonable that any symmetric tensor (six independent elements) can be represented by a diagonal tensor (three independent elements) plus a rotation (three angles). The three angles just have to be selected carefully enough to make sure that they cause the off-diagonal elements to vanish. This supposition is indeed correct, although we will not pause for long enough to prove it by producing explicit formulas for the three angles.

### A.5 Tensors in $d$ dimensions

A  $d$ -dimensional tensor is represented by a  $d \times d$  matrix with  $d^2$  entries. If the tensor is symmetric, there are  $d$  independent on-diagonal elements and  $d(d-1)/2$  independent off-diagonal elements. The tensor components will change with coordinate system in the now-familiar form

$$\mathbf{T}' = \mathbf{R}\mathbf{T}\mathbf{R}^\dagger, \quad (\text{A.20})$$

where  $\mathbf{R}$  is the  $d \times d$  rotation matrix.

How many angles does it take to specify a rotation in  $d$  dimensions? Remember how we went from two dimensions to three: The three dimensional rotation took place “in a plane”, i.e. in a two-dimensional subspace. It required two (i.e.  $d-1$ ) angles to specify the orientation of the plane plus one to specify the rotation within the plane... a total of three angles.

A rotation in four dimensions takes place within a three-dimensional subspace. It requires  $3 = d-1$  angles to specify the orientation of the three-dimensional subspace, plus, as we found above, three angles to specify the rotation within the three-dimensional subspace... a total of six angles.

A rotation in five dimensions requires  $4 = d-1$  angles to specify the four-dimensional subspace in which the rotation occurs, plus the six angles that we have just found specify a rotation within that subspace... a total of ten angles.

In general, the number of angles needed to specify a rotation in  $d$  dimensions is

$$A_d = d-1 + A_{d-1} = d(d-1)/2. \quad (\text{A.21})$$

This is exactly the number of independent off-diagonal elements in a symmetric tensor. It seems reasonable that we can choose the angles to ensure that, in the resulting coordinate system, all the off-diagonal elements vanish. The proof of this result is difficult and proceeds in a very different manner from the plausibility argument sketched here. (The proof involves concepts like eigenvectors and eigenvalues, and it gives an explicit recipe for constructing the rotation matrix. It has the advantage of rigor and the disadvantage of being so technical that it's easy to lose track of the fact that that all you're doing is choosing a coordinate system.)

#### A.12 Problem: Non-symmetric tensors\*

Argue that a non-symmetric tensor can be brought into a “triangular”

representation in which all the elements below the diagonal are equal to zero and all the elements on and above the diagonal are independent. (This is indeed the case, although in general some of the non-zero elements remaining will be complex-valued, and some of the angles will involve rotations into complex-valued vectors.)

## A.6 Linear transformations in two dimensions

Section A.3 considered  $2 \times 2$  matrices as representations of tensors. This section gains additional insight by considering  $2 \times 2$  matrices as representations of linear transformations. It demonstrates how diagonalization can be useful and gives a clue to an efficient algorithm for diagonalization.

A linear transformation is a function from vectors to vectors that can be represented in any given coordinate system as

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \quad (\text{A.22})$$

If the equation above represents (“names”) the transformation in one coordinate system, what is its representation in some other coordinate system? We assume that the two coordinate systems are related through an orthogonal matrix  $\mathbf{B}$  such that

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \mathbf{B} \begin{pmatrix} u \\ v \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x' \\ y' \end{pmatrix} = \mathbf{B} \begin{pmatrix} x \\ y \end{pmatrix}. \quad (\text{A.23})$$

(For example, if the new coordinate system is the primed coordinate system of the figure on page 353, then the matrix  $\mathbf{B}$  that translates from the original to the new coordinates is the rotation matrix  $\mathbf{R}(\phi)$ .) Given this “translation dictionary”, we have

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \mathbf{B} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \quad (\text{A.24})$$

But  $\mathbf{B}$  is invertible, so

$$\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{B}^{-1} \begin{pmatrix} x' \\ y' \end{pmatrix} \quad (\text{A.25})$$

whence

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \mathbf{B} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \mathbf{B}^{-1} \begin{pmatrix} x' \\ y' \end{pmatrix}. \quad (\text{A.26})$$

Thus the representation of the transformation in the primed coordinate system is

$$\mathbf{B} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \mathbf{B}^{-1} \quad (\text{A.27})$$

(compare equation A.11). This equation has a very direct physical meaning. Remember that the matrix  $\mathbf{B}$  translates from the old  $(x, y)$  coordinates to the new  $(x', y')$  coordinates, while the matrix  $\mathbf{B}^{-1}$  translates in the opposite direction. Thus the equation above says that the representation of a transformation in the new coordinates is given by translating from new to old coordinates (through the matrix  $\mathbf{B}^{-1}$ ), then applying the old representation (the “ $a$  matrix”) to those old coordinates, and finally translating back from old to new coordinates (through the matrix  $\mathbf{B}$ ).

The rest of this section considers only transformations represented by symmetric matrices, which we will denote by

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \quad (\text{A.28})$$

Let’s try to understand this transformation as something more than a jumble of symbols awaiting a plunge into the calculator. First of all, suppose the vector  $\mathbf{V}$  maps to the vector  $\mathbf{W}$ . Then the vector  $5\mathbf{V}$  will be mapped to vector  $5\mathbf{W}$ . In short, if we know how the transformation acts on vectors with magnitude unity, we will be able to see immediately how it acts on vectors with other magnitudes. Thus we focus our attention on vectors on the unit circle:

$$x^2 + y^2 = 1. \quad (\text{A.29})$$

A brief calculation shows that the length of the output vector is then

$$\sqrt{u^2 + v^2} = \sqrt{a^2x^2 + b^2 + c^2y^2 + 2b(a+c)xy}, \quad (\text{A.30})$$

which isn’t very helpful. Another brief calculation shows that if the input vector has polar angle  $\theta$ , then the output vector has polar angle  $\varphi$  with

$$\tan \varphi = \frac{b + c \tan \theta}{a + b \tan \theta}, \quad (\text{A.31})$$

which is similarly opaque and messy.

Instead of trying to understand the transformation in its initial coordinate system, let’s instead convert (rotate) to the special coordinate system

in which the transformation is represented by a diagonal matrix. In this system,

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} d_1 x' \\ d_2 y' \end{pmatrix}. \quad (\text{A.32})$$

The unit circle is still

$$x'^2 + y'^2 = 1, \quad (\text{A.33})$$

so the image of the unit circle is

$$\left(\frac{u'}{d_1}\right)^2 + \left(\frac{v'}{d_2}\right)^2 = 1, \quad (\text{A.34})$$

namely an ellipse! This result is transparent in the special coordinate system, but almost impossible to see in the original one.

Note particularly what happens to a vector pointing along the  $x'$  coordinate axis. For example, the unit vector in this direction transforms to

$$\begin{pmatrix} d_1 \\ 0 \end{pmatrix} = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad (\text{A.35})$$

In other words, when the vector is transformed it changes in magnitude, but *not* in direction. Vectors with this property are called *eigenvectors*. It is easy to see that any vector on either the  $x'$  or  $y'$  coordinate axes are eigenvectors.

### A.7 What does “eigen” mean?

If a vector  $\mathbf{x}$  is acted upon by a linear transformation  $\mathbf{B}$ , then the output vector

$$\mathbf{x}' = \mathbf{B}\mathbf{x} \quad (\text{A.36})$$

will usually be skew to the original vector  $\mathbf{x}$ . However, for some very special vectors it might just happen that  $\mathbf{x}'$  is parallel to  $\mathbf{x}$ . Such vectors are called “eigenvectors”. (This is a terrible name because (1) it gives no idea of what eigenvectors are or why they’re so important and (2) it sounds gross. However, that’s what they’re called.) We have already seen, in the previous section, that eigenvectors are related to coordinate systems in which the transformation is particularly easy to understand.

If  $\mathbf{x}$  is an eigenvector, then

$$\mathbf{B}\mathbf{x} = \lambda\mathbf{x}, \quad (\text{A.37})$$

where  $\lambda$  is a scalar called “the eigenvalue associated with eigenvector  $\mathbf{x}$ ”. If  $\mathbf{x}$  is an eigenvector, then any vector parallel to  $\mathbf{x}$  is also an eigenvector with the same eigenvalue. (That is, any vector of the form  $c\mathbf{x}$ , where  $c$  is any scalar, is also an eigenvector with the same eigenvalue.) Sometimes we speak of a “line of eigenvectors”.

The vector  $\mathbf{x} = \mathbf{0}$  is never considered an eigenvector, because

$$\mathbf{B}\mathbf{0} = \lambda\mathbf{0}, \quad (\text{A.38})$$

for any value of  $\lambda$  for any linear transformation. On the other hand, if

$$\mathbf{B}\mathbf{x} = 0\mathbf{x} = \mathbf{0} \quad (\text{A.39})$$

for some non-zero vector  $\mathbf{x}$ , then  $\mathbf{x}$  is an eigenvector with eigenvalue  $\lambda = 0$ .

#### A.13 Problem: Plane of eigenvectors

Suppose  $\mathbf{x}$  and  $\mathbf{y}$  are two non-parallel vectors with the same eigenvalue. (In this case the eigenvalue is said to be “degenerate”, which sounds like an aspersion cast upon the morals of the eigenvalue but which is really just poor choice of terminology again.) Show that any vector of the form  $c_1\mathbf{x} + c_2\mathbf{y}$  is an eigenvector with the same eigenvalue.

### A.8 How to diagonalize a symmetric matrix

We saw in section A.3 that for any  $2 \times 2$  symmetric matrix, represented in its initial basis by, say,

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}, \quad (\text{A.40})$$

a simple rotation of axes would produce a new coordinate system in which the matrix representation is diagonal:

$$\begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}. \quad (\text{A.41})$$

These two matrices are related through

$$\begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} = \mathbf{R}(\phi) \begin{pmatrix} a & b \\ b & c \end{pmatrix} \mathbf{R}^{-1}(\phi), \quad (\text{A.42})$$

where  $R(\phi)$  is the rotation matrix (A.3). Problem A.10 gave a direct way to find the desired rotation. However this direct technique is cumbersome and doesn't generalize readily to higher dimensions. This section presents a different technique, which relies on eigenvalues and eigenvectors, that is more efficient and that generalizes readily to complex-valued matrices and to matrices in any dimension, but that is somewhat sneaky and conceptually roundabout.

We begin by noting that any vector lying along the  $x'$ -axis (of the preferred coordinate system) is an eigenvector. For example, the vector  $5\hat{\mathbf{i}}'$  is represented (in the preferred coordinate system) by

$$\begin{pmatrix} 5 \\ 0 \end{pmatrix}. \quad (\text{A.43})$$

Multiplying this vector by the matrix in question gives

$$\begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} \begin{pmatrix} 5 \\ 0 \end{pmatrix} = d_1 \begin{pmatrix} 5 \\ 0 \end{pmatrix}, \quad (\text{A.44})$$

so  $5\hat{\mathbf{i}}'$  is an eigenvector with eigenvalue  $d_1$ . The same holds for any scalar multiple of  $\hat{\mathbf{i}}'$ , whether positive or negative. Similarly, any scalar multiple of  $\hat{\mathbf{j}}'$  is an eigenvector with eigenvalue  $d_2$ . In short, the two elements on the diagonal in the preferred (diagonal) representation are the two eigenvalues, and the two unit vectors  $\hat{\mathbf{i}}'$  and  $\hat{\mathbf{j}}'$  of the preferred coordinate system are two of the eigenvectors.

Thus finding the eigenvectors and eigenvalues of a matrix gives you the information needed to diagonalize that matrix. The unit vectors  $\hat{\mathbf{i}}'$  and  $\hat{\mathbf{j}}'$  constitute an "orthonormal basis of eigenvectors". The eigenvectors even give the rotation matrix directly, as described in the next paragraph.

Let's call the rotation matrix

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}, \quad (\text{A.45})$$

so that the inverse (transpose) matrix is

$$\mathbf{B}^{-1} = \mathbf{B}^\dagger = \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix}. \quad (\text{A.46})$$

The representation of  $\hat{\mathbf{i}}'$  in the preferred basis is

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad (\text{A.47})$$

so its representation in the initial basis is (see equation A.2)

$$\mathbf{B}^\dagger \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} b_{11} \\ b_{12} \end{pmatrix}. \quad (\text{A.48})$$

Similarly, the representation of  $\hat{\mathbf{j}}'$  in the initial basis is

$$\mathbf{B}^\dagger \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} b_{21} \\ b_{22} \end{pmatrix}. \quad (\text{A.49})$$

Thus the rotation matrix is

$$\mathbf{B} = \begin{pmatrix} \text{initial rep. of } \hat{\mathbf{i}}', \text{ on its side} \\ \text{initial rep. of } \hat{\mathbf{j}}', \text{ on its side} \end{pmatrix}. \quad (\text{A.50})$$

### Example

Suppose we need to find a diagonal representation for the matrix

$$\mathbf{T} = \begin{pmatrix} 7 & 3 \\ 3 & 7 \end{pmatrix}. \quad (\text{A.51})$$

First we search for the special vectors—the eigenvectors—such that

$$\begin{pmatrix} 7 & 3 \\ 3 & 7 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}. \quad (\text{A.52})$$

At the moment, we don't know either the eigenvalue  $\lambda$  or the associated eigenvector  $(x, y)$ . Thus it seems that (bad news) we are trying to solve two equations for three unknowns:

$$\begin{aligned} 7x + 3y &= \lambda x \\ 3x + 7y &= \lambda y \end{aligned} \quad (\text{A.53})$$

Remember, however, that there is not one *single* eigenvector: any multiple of an eigenvector is also an eigenvector. (Alternatively, any vector on the line that extends the eigenvector is another eigenvector.) We only need one of these eigenvectors, so let's take the one that has  $x = 1$  (i.e. the vector on the extension line where it intersects the vertical line  $x = 1$ ). (This technique will fail if we have the bad luck that our actual eigenvector is vertical and hence never passes through the line  $x = 1$ .) So we really have two equations in two unknowns:

$$\begin{aligned} 7 + 3y &= \lambda \\ 3 + 7y &= \lambda y \end{aligned}$$

but note that they are not *linear* equations... the damnable product  $\lambda y$  in the lower right corner means that all our techniques for solving linear equations go right out the window. We *can* solve these two equations for  $\lambda$  and  $y$ , but there's an easier, if somewhat roundabout, approach.



*Finding eigenvalues*

Let's go back to equation (A.52) and write it as

$$\begin{pmatrix} 7 & 3 \\ 3 & 7 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \lambda \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (\text{A.54})$$

Then

$$\begin{pmatrix} 7 & 3 \\ 3 & 7 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (\text{A.55})$$

or

$$\begin{pmatrix} 7 - \lambda & 3 \\ 3 & 7 - \lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (\text{A.56})$$

Let's think about this. It says that for some matrix  $\mathbf{M} = \mathbf{T} - \lambda\mathbf{1}$ , we have

$$\mathbf{M} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (\text{A.57})$$

You know right away one vector  $(x, y)$  that satisfies this equation, namely  $(x, y) = (0, 0)$ . And most of the time, this is the *only* vector that satisfies the equation, because

$$\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{M}^{-1} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (\text{A.58})$$

We appear to have reached a dead end. The solution is  $(x, y) = (0, 0)$ , but the zero vector is not, by definition, considered an eigenvector of any transformation. (Because it always gives eigenvalue zero for any transformation.)

However, if the matrix  $\mathbf{M}$  is *not invertible*, then there will be other solutions to

$$\mathbf{M} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (\text{A.59})$$

in addition to the trivial solution  $(x, y) = (0, 0)$ . Thus we must look for those special values of  $\lambda$  such that the so-called *characteristic matrix*  $\mathbf{M}$  is not invertible. These values come if and only if the determinant of  $\mathbf{M}$  vanishes. For this example, we have to find values of  $\lambda$  such that

$$\det \begin{pmatrix} 7 - \lambda & 3 \\ 3 & 7 - \lambda \end{pmatrix} = 0. \quad (\text{A.60})$$

This is a quadratic equation in  $\lambda$

$$(7 - \lambda)^2 - 3^2 = 0 \quad (\text{A.61})$$

called the *characteristic equation*. Its two solutions are

$$7 - \lambda = \pm 3 \quad (\text{A.62})$$

or

$$\lambda = 7 \pm 3 = 10 \text{ or } 4. \quad (\text{A.63})$$

We have found the two eigenvalues of our matrix!

*Finding eigenvectors*

Let's look now for the eigenvector associated with  $\lambda = 4$ . Equation (A.53)

$$\begin{aligned}7x + 3y &= \lambda x \\3x + 7y &= \lambda y\end{aligned}$$

still holds, but no longer does it look like two equations in three unknowns, because we are now interested in the case  $\lambda = 4$ :

$$\begin{aligned}7x + 3y &= 4x \\3x + 7y &= 4y\end{aligned}$$

Following our nose gives

$$\begin{aligned}3x + 3y &= 0 \\3x + 3y &= 0\end{aligned}$$

and when we see this our heart skips a beat or two. . . a degenerate system of equations! Relax and rest your heart. This system has an infinite number of solutions and it's supposed to have an infinite number of solutions, because any multiple of an eigenvector is also an eigenvector. The eigenvectors associated with  $\lambda = 4$  are any multiple of

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix}. \quad (\text{A.64})$$

An entirely analogous search for the eigenvectors associated with  $\lambda = 10$  finds any multiple of

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (\text{A.65})$$

*Tidying up*

We have the two sets of eigenvectors, but which shall we call  $\hat{\mathbf{i}}'$  and which  $\hat{\mathbf{j}}'$ ? This is a matter of individual choice, but my choice is usually to make the transformation be a rotation (without reflection) through a small positive angle. Our new, preferred coordinate system is related to the original coordinates by a simple rotation of  $45^\circ$  if we choose

$$\hat{\mathbf{i}}' = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \hat{\mathbf{j}}' = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}. \quad (\text{A.66})$$

(Note that we have also “normalized the basis”, i.e. selected the basis vectors to have magnitude unity.) Given this choice, the orthogonal rotation matrix that changes coordinates from the original to the preferred system is (see equation A.50)

$$\mathbf{B} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad (\text{A.67})$$

and the diagonalized matrix (or, more properly, the representation of the matrix in the preferred coordinate system) is

$$\begin{pmatrix} 10 & 0 \\ 0 & 4 \end{pmatrix}. \quad (\text{A.68})$$

You don't believe me? Then multiply out

$$\mathbf{B} \begin{pmatrix} 7 & 3 \\ 3 & 7 \end{pmatrix} \mathbf{B}^\dagger \quad (\text{A.69})$$

and see for yourself.

### Problems

#### A.14 **Problem: Diagonalize a $2 \times 2$ matrix\***

Diagonalize the matrix

$$\begin{pmatrix} 26 & 12 \\ 12 & 19 \end{pmatrix}. \quad (\text{A.70})$$

- Find its eigenvalues.
- Find its eigenvectors, and verify that they are orthogonal.
- Sketch the eigenvectors, and determine the signs and sequence most convenient for assigning axes. (That is, should the first eigenvector you found be called  $\hat{\mathbf{i}}'$ ,  $-\hat{\mathbf{i}}'$ , or  $\hat{\mathbf{j}}'$ ?)
- Find the matrix that translates from the initial basis to the basis of eigenvectors produced in part (c.).
- Verify that the matrix produced in part (d.) is orthogonal.
- Verify that the representation of the matrix above in the basis of eigenvectors is diagonal.
- (Optional.) What is the rotation angle?

#### A.15 **Problem: Eigenvalues of a $2 \times 2$ matrix**

Show that the eigenvalues of

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix} \quad (\text{A.71})$$

are

$$\lambda = \frac{1}{2} \left[ (a + c) \pm \sqrt{(a - c)^2 + 4b^2} \right]. \quad (\text{A.72})$$

Under what circumstances is an eigenvalue complex valued? Under what circumstances are the two eigenvalues the same?

**A.16 Problem: Diagonalize a  $3 \times 3$  matrix**

Diagonalize the matrix

$$\frac{1}{625} \begin{pmatrix} 1182 & -924 & 540 \\ -924 & 643 & 720 \\ 540 & 720 & -575 \end{pmatrix}. \quad (\text{A.73})$$

a. Find its eigenvalues by showing that the characteristic equation is

$$\lambda^3 - 2\lambda^2 - 5\lambda + 6 = (\lambda - 3)(\lambda + 2)(\lambda - 1) = 0. \quad (\text{A.74})$$

b. Find its eigenvectors, and verify that they are orthogonal.

c. Show that the translation matrix can be chosen to be

$$\mathbf{B} = \frac{1}{25} \begin{pmatrix} 20 & -15 & 0 \\ 9 & 12 & -20 \\ 12 & 16 & 15 \end{pmatrix}. \quad (\text{A.75})$$

Why did I use the phrase “the translation matrix can be chosen to be” rather than “the translation matrix is”?

**A.17 Problem: A  $3 \times 3$  matrix eigenproblem**

Find the eigenvalues and associated eigenvectors for the matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{pmatrix}. \quad (\text{A.76})$$

**A.9 A glance at computer algorithms**

Anyone who has worked even one of the problems in section A.8 knows that diagonalizing a matrix is no picnic: there’s a lot of mundane arithmetic involved and it’s very easy to make mistakes. This is a problem ripe for computer solution. One’s first thought is to program a computer to solve the problem using the same technique that we used to solve it on paper: first find the eigenvalues through the characteristic equation, then find the eigenvectors through a degenerate set of linear equations.

This turns out to be a very poor algorithm for automatic computation. The effective algorithm is to choose a matrix  $\mathbf{B}$  such that the off-diagonal elements of

$$\mathbf{B}\mathbf{A}\mathbf{B}^{-1} \quad (\text{A.77})$$

are smaller than the off-diagonal elements of  $\mathbf{A}$ . Then choose another, and another. Go through this process again and again until the off-diagonal elements have been ground down to machine zero. There are many strategies for choosing the series of  $\mathbf{B}$  matrices. These are well-described in any edition of *Numerical Recipes*.<sup>4</sup>

When you need to diagonalize matrices numerically, I urge you to look at *Numerical Recipes* to see what's going on, but I urge you *not* to code these algorithms yourself. These algorithms rely in an essential way on the fact that computer arithmetic is approximate rather than exact, and hence they are quite tricky to implement. Instead of coding the algorithms yourself, I recommend that you use the implementations in either LAPACK<sup>5</sup> (the Linear Algebra PACKage) or EISPACK.<sup>6</sup> These packages are probably the finest computer software ever written, and they are free. They can be obtained through the "Guide to Available Mathematical Software" (GAMS) at <http://gams.nist.gov>.

### A.10 A glance at non-symmetric matrices and the Jordan form

Many of the matrices that arise in applications are symmetric and hence the results of the previous sections are the only ones needed. But every once in a while you do encounter a non-symmetric matrix and this section gives you a guide to treating them. It is just an introduction and treats only  $2 \times 2$  matrices.

Given a non-symmetric matrix, the first thing to do is rotate the axes to make the matrix representation triangular, as discussed in problem A.12:

$$\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}. \quad (\text{A.78})$$

Note that  $b \neq 0$  because otherwise the matrix would be symmetric and we would already be done. In this case vectors on the  $x$ -axis are eigenvectors because

$$\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad (\text{A.79})$$

Are there any other eigenvectors? The equation

$$\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix} \quad (\text{A.80})$$

tells us that

$$\begin{aligned} ax + by &= \lambda x \\ cy &= \lambda y \end{aligned}$$

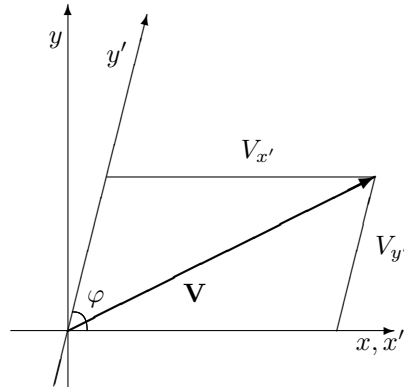
whence  $\lambda = c$  and the eigenvector has polar angle  $\theta$  where

$$\tan \theta = \frac{c - a}{b}. \quad (\text{A.81})$$

Note that if  $c = a$  (the “degenerate” case: both eigenvalues are the same) then  $\theta = 0$  or  $\theta = \pi$ . In this case *all* of the eigenvectors are on the  $x$ -axis.

### *Diagonal form*

We already know that that a rotation of orthogonal (Cartesian) coordinates will not diagonalize this matrix. We must instead transform to a skew coordinate system in which the axes are not perpendicular.



Note that in with oblique axes, the coordinates are given by

$$\mathbf{V} = V_{x'} \hat{\mathbf{i}}' + V_{y'} \hat{\mathbf{j}}' \quad (\text{A.82})$$

but, because  $\hat{\mathbf{i}}'$  and  $\hat{\mathbf{j}}'$  are not perpendicular, it is *not true* that

$$V_{x'} = \mathbf{V} \cdot \hat{\mathbf{i}}'. \quad \text{NO!} \quad (\text{A.83})$$

A little bit of geometry will convince you that the name of the vector  $\mathbf{V}$  changes according to

$$\begin{pmatrix} V_{x'} \\ V_{y'} \end{pmatrix} = \mathbf{B} \begin{pmatrix} V_x \\ V_y \end{pmatrix}, \quad (\text{A.84})$$

where

$$\mathbf{B} = \frac{1}{\sin \varphi} \begin{pmatrix} \sin \varphi & -\cos \varphi \\ 0 & 1 \end{pmatrix}. \quad (\text{A.85})$$

This matrix is *not* orthogonal. In fact its inverse is

$$\mathbf{B}^{-1} = \begin{pmatrix} 1 & \cos \varphi \\ 0 & \sin \varphi \end{pmatrix}. \quad (\text{A.86})$$

Finally, note that we cannot have  $\varphi = 0$  or  $\varphi = \pi$ , because then both  $V_{x'}$  and  $V_{y'}$  would give information about the horizontal component of the vector, and there would be no information about the vertical component of the vector.

What does this say about the representations of tensors (or, equivalently, of linear transformations)? The “name translation” argument of equation (A.27) still applies, so

$$\mathbf{T}' = \mathbf{B}\mathbf{T}\mathbf{B}^{-1}. \quad (\text{A.87})$$

Using the explicit matrices already given, this says

$$\mathbf{T}' = \frac{1}{\sin \varphi} \begin{pmatrix} \sin \varphi & -\cos \varphi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \begin{pmatrix} 1 & \cos \varphi \\ 0 & \sin \varphi \end{pmatrix} = \begin{pmatrix} a & (a-c)\cos \varphi + b\sin \varphi \\ 0 & c \end{pmatrix}. \quad (\text{A.88})$$

To make this diagonal, we need only choose a skew coordinate system where the angle  $\varphi$  gives

$$(a-c)\cos \varphi + b\sin \varphi = 0, \quad (\text{A.89})$$

that is, one with

$$\tan \varphi = \frac{c-a}{b}. \quad (\text{A.90})$$

Comparison with equation (A.81) shows that this simply means that the skew coordinate system should have its axes pointing along two eigenvectors. We have once again found an intimate connection between diagonal

representations and eigenvectors, a connection which is exploited fully in abstract mathematical treatments of matrix diagonalization.

Once again we can do an accounting of information. In the initial coordinate system, the four elements of the matrix contain four independent pieces of information. In the diagonalizing coordinate system, two of those pieces are explicit in the matrix, and two are implicit in the two axis rotation angles needed to implement the diagonalization.

This procedure works almost all the time. But, if  $a = c$ , then it would involve  $\varphi = 0$  or  $\varphi = \pi$ , and we have already seen that this is not an acceptable change of coordinates.

### *Degenerate case*

Suppose our matrix has equal eigenvalues,  $a = c$ , so that it reads

$$\begin{pmatrix} a & b \\ 0 & a \end{pmatrix}. \quad (\text{A.91})$$

If  $b = 0$ , then the matrix is already diagonal. (Indeed, in this case *all* vectors are eigenvectors with eigenvalue  $a$ , and the linear transformation is simply multiplication of each vector by  $a$ ).

But if  $b \neq 0$ , then, as we have seen, the only eigenvectors are on the  $x$ -axis, and it is impossible to make a basis of eigenvectors. Only one thing can be done to make the matrix representation simpler than it stands in equation (A.91), and that is a shift in the scale used to measure the  $y$ -axis.

For example, suppose that in the  $(x, y)$  coordinate system, the  $y$ -axis is calibrated in inches. We wish to switch to the  $(x', y')$  system in which the  $y'$ -axis is calibrated in feet. There is no change in axis orientation or in the  $x$ -axis. It is easy to see that the two sets of coordinates are related through

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1/12 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 12 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} \quad (\text{A.92})$$

This process is sometimes called a “stretching” or a “scaling” of the  $y$ -axis.

The transformation represented by matrix (A.91) in the initial coordinate system is represented in the new coordinate system by

$$\begin{pmatrix} 1 & 0 \\ 0 & 1/12 \end{pmatrix} \begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 12 \end{pmatrix} = \begin{pmatrix} a & 12b \\ 0 & a \end{pmatrix}. \quad (\text{A.93})$$



The choice of what to do now is clear. Instead of scaling the  $y$ -axis by a factor of 12, we can scale it by a factor of  $1/b$ , and produce a new matrix representation of the form

$$\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}. \quad (\text{A.94})$$

Where is the information in this case? In the initial coordinate system, the four elements of the matrix contain four independent pieces of information. In the new coordinate system, two of those pieces are explicit in the matrix, one is implicit in the rotation angle needed to implement the initial triangularization, and one is implicit in the  $y$ -axis scale transformation.

### *The Jordan form*

Remarkably, the situation discussed above for  $2 \times 2$  matrices covers all the possible cases for  $n \times n$  matrices. That is, in  $n$  dimensional space, the proper combination of rotations, skews, and stretches of coordinate axes will bring the matrix representation (the “name”) of any tensor or linear transformation into a form where every element is zero except on the diagonal and on the superdiagonal. The elements on the diagonal are eigenvalues, and each element on the superdiagonal is either zero or one: zero if the two adjacent eigenvalues differ, either zero or one if they are the same. The warning of problem A.12 applies here as well: The eigenvalues on the diagonal may well be complex valued, and the same applies for the elements of the new basis vectors.

### References

<sup>1</sup>For example, Kenneth Hoffman and Ray Kunze, *Linear Algebra*, second edition (Prentice-Hall, Englewood Cliffs, New Jersey, 1971).

<sup>2</sup>For example, Jerry Marion and Stephen Thorton, *Classical Dynamics of Particles and Systems*, fourth edition (Saunders College Publishing, Fort Worth, Texas, 1995) section 11.2.

<sup>3</sup>For example, Jerry Marion and Stephen Thorton, *Classical Dynamics of Particles and Systems*, fourth edition (Saunders College Publishing, Fort Worth, Texas, 1995) section 11.7.

<sup>4</sup>W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes* (Cambridge University Press, Cambridge, U.K., 1992).

<sup>5</sup>E. Anderson, *et al.*, *LAPACK Users' Guide* (SIAM, Philadelphia, 1992).

<sup>6</sup>B.T. Smith, *et al.*, *Matrix Eigensystem Routines—EISPACK Guide* (Springer-Verlag, Berlin, 1976).

## Appendix B

# The Spherical Harmonics

A “function on the unit sphere” is a function  $f(\theta, \phi)$ . Another convenient variable is  $\mu = \cos \theta$ . “Integration over the unit sphere” means

$$\int d\Omega f(\theta, \phi) = \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi f(\theta, \phi) = \int_{-1}^{+1} d\mu \int_0^{2\pi} d\phi f(\theta, \phi).$$

$$\nabla^2 Y_\ell^m(\theta, \phi) = -\frac{1}{r^2} \ell(\ell + 1) Y_\ell^m(\theta, \phi) \quad (\text{B.1})$$

$$\int Y_{\ell'}^{m'*}(\theta, \phi) Y_\ell^m(\theta, \phi) d\Omega = \delta_{\ell', \ell} \delta_{m', m} \quad (\text{B.2})$$

$$f(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} f_{\ell, m} Y_\ell^m(\theta, \phi) \quad \text{where} \quad (\text{B.3})$$

$$f_{\ell, m} = \int Y_\ell^{m*}(\theta, \phi) f(\theta, \phi) d\Omega \quad (\text{B.4})$$

In the table, square roots are always taken to be positive.

$$Y_0^0(\mu, \phi) = \left( \frac{1}{2^2\pi} \right)^{1/2}$$

$$Y_1^0(\mu, \phi) = \left( \frac{3}{2^2\pi} \right)^{1/2} \mu$$

$$Y_1^{\pm 1}(\mu, \phi) = \mp \left( \frac{3}{2^3\pi} \right)^{1/2} \sqrt{1 - \mu^2} e^{\pm i\phi}$$

$$= \left( \frac{3}{2^2\pi} \right)^{1/2} \frac{z}{r}$$

$$= \mp \left( \frac{3}{2^3\pi} \right)^{1/2} \frac{1}{r} (x \pm iy)$$

$$Y_2^0(\mu, \phi) = \left( \frac{5}{2^4\pi} \right)^{1/2} (3\mu^2 - 1)$$

$$Y_2^{\pm 1}(\mu, \phi) = \mp \left( \frac{3 \cdot 5}{2^3\pi} \right)^{1/2} \mu \sqrt{1 - \mu^2} e^{\pm i\phi}$$

$$Y_2^{\pm 2}(\mu, \phi) = \left( \frac{3 \cdot 5}{2^5\pi} \right)^{1/2} (1 - \mu^2) e^{\pm 2i\phi}$$

$$= \left( \frac{5}{2^4\pi} \right)^{1/2} \left( 3 \frac{z^2}{r^2} - 1 \right)$$

$$= \mp \left( \frac{3 \cdot 5}{2^3\pi} \right)^{1/2} \frac{z}{r^2} (x \pm iy)$$

$$= \left( \frac{3 \cdot 5}{2^5\pi} \right)^{1/2} \frac{1}{r^2} (x \pm iy)^2$$

$$Y_3^0(\mu, \phi) = \left( \frac{7}{2^4\pi} \right)^{1/2} (5\mu^3 - 3\mu)$$

$$Y_3^{\pm 1}(\mu, \phi) = \mp \left( \frac{3 \cdot 7}{2^6\pi} \right)^{1/2} (5\mu^2 - 1) \sqrt{1 - \mu^2} e^{\pm i\phi}$$

$$Y_3^{\pm 2}(\mu, \phi) = \left( \frac{3 \cdot 5 \cdot 7}{2^5\pi} \right)^{1/2} \mu (1 - \mu^2) e^{\pm 2i\phi}$$

$$Y_3^{\pm 3}(\mu, \phi) = \mp \left( \frac{5 \cdot 7}{2^6\pi} \right)^{1/2} (1 - \mu^2) \sqrt{1 - \mu^2} e^{\pm 3i\phi}$$

$$= \left( \frac{7}{2^4\pi} \right)^{1/2} \left( 5 \frac{z^3}{r^3} - 3 \frac{z}{r} \right)$$

$$= \mp \left( \frac{3 \cdot 7}{2^6\pi} \right)^{1/2} \left( 5 \frac{z^2}{r^2} - 1 \right) \frac{1}{r} (x \pm iy)$$

$$= \left( \frac{3 \cdot 5 \cdot 7}{2^5\pi} \right)^{1/2} \frac{z}{r^3} (x \pm iy)^2$$

$$= \mp \left( \frac{5 \cdot 7}{2^6\pi} \right)^{1/2} \frac{1}{r^3} (x \pm iy)^3$$

## Appendix C

# Radial Wavefunctions for the Coulomb Problem

Based on Griffiths, page 154, but with scaled variables and with integers factorized.

$$R_{10}(r) = 2e^{-r}$$

$$R_{20}(r) = \frac{1}{\sqrt{2}} \left(1 - \frac{1}{2}r\right) e^{-r/2}$$

$$R_{21}(r) = \frac{1}{\sqrt{2^3 \cdot 3}} r e^{-r/2}$$

$$R_{30}(r) = \frac{2}{\sqrt{3^3}} \left(1 - \frac{2}{3}r + \frac{2}{3^3}r^2\right) e^{-r/3}$$

$$R_{31}(r) = \frac{2^3}{3^3 \sqrt{2 \cdot 3}} \left(1 - \frac{1}{2 \cdot 3}r\right) r e^{-r/3}$$

$$R_{32}(r) = \frac{2^2}{3^4 \sqrt{2 \cdot 3 \cdot 5}} r^2 e^{-r/3}$$

$$R_{40}(r) = \frac{1}{2^2} \left(1 - \frac{3}{2^2}r + \frac{1}{2^3}r^2 - \frac{1}{2^6 \cdot 3}r^3\right) e^{-r/4}$$

$$R_{41}(r) = \frac{\sqrt{5}}{2^4 \sqrt{3}} \left(1 - \frac{1}{2^2}r + \frac{1}{2^4 \cdot 5}r^2\right) r e^{-r/4}$$

$$R_{42}(r) = \frac{1}{2^6 \sqrt{5}} \left(1 - \frac{1}{2^2 \cdot 3}r\right) r^2 e^{-r/4}$$

$$R_{43}(r) = \frac{1}{2^8 \cdot 3 \sqrt{5 \cdot 7}} r^3 e^{-r/4}$$



## Appendix D

# Quantum Mechanics Cheat Sheet

**Delta functions:**

$$\int_{-\infty}^{+\infty} e^{ikx} dk = 2\pi\delta(x) \quad (\text{D.1})$$

$$\int_{-\infty}^{+\infty} e^{i(p/\hbar)x} dp = 2\pi\hbar\delta(x) \quad (\text{D.2})$$

$$\int_{-\infty}^{+\infty} e^{i\omega t} d\omega = 2\pi\delta(t) \quad (\text{D.3})$$

**Fourier transforms:**

$$\tilde{\psi}(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{+\infty} \psi(x) e^{-i(p/\hbar)x} dx \quad (\text{D.4})$$

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{+\infty} \tilde{\psi}(p) e^{+i(p/\hbar)x} dp \quad (\text{D.5})$$

$$\tilde{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \quad (\text{D.6})$$

$$f(t) = \int_{-\infty}^{+\infty} \tilde{f}(\omega) e^{+i\omega t} \frac{d\omega}{2\pi} \quad (\text{D.7})$$

**Gaussian integrals:**

$$\int_{-\infty}^{+\infty} e^{ax^2+bx} dx = \sqrt{\frac{\pi}{-a}} e^{-b^2/4a} \quad \Re\{a\} \leq 0 \quad (\text{D.8})$$

$$\frac{\int_{-\infty}^{+\infty} x^2 e^{-x^2/2\sigma^2} dx}{\int_{-\infty}^{+\infty} e^{-x^2/2\sigma^2} dx} = \sigma^2 \quad (\text{D.9})$$

**Time development:**

$$\frac{d|\psi(t)\rangle}{dt} = -\frac{i}{\hbar}\hat{H}|\psi(t)\rangle \quad (\text{D.10})$$

$$\frac{\partial\psi(\mathbf{x}, t)}{\partial t} = -\frac{i}{\hbar}\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{x})\right]\psi(\mathbf{x}, t) \quad (\text{D.11})$$

$$|\psi(t)\rangle = \sum_n e^{-(i/\hbar)E_n t} c_n |\eta_n\rangle \quad (\text{D.12})$$

$$\frac{d\langle\hat{A}\rangle}{dt} = -\frac{i}{\hbar}\langle[\hat{A}, \hat{H}]\rangle \quad (\text{D.13})$$

**Momentum:**

$$\hat{p} \iff -i\hbar\frac{\partial}{\partial x} \quad (\text{D.14})$$

$$[\hat{x}, \hat{p}] = i\hbar \quad (\text{D.15})$$

$$\langle x|p\rangle = \frac{1}{\sqrt{2\pi\hbar}}e^{i(p/\hbar)x} \quad (\text{D.16})$$

**Dimensions:**

$$\psi(x) \text{ has dimensions } [\text{length}]^{-1/2} \quad (\text{D.17})$$

$$\psi(\mathbf{x}_1, \mathbf{x}_2) \text{ has dimensions } [\text{length}]^{-6/2} \quad (\text{D.18})$$

$$\tilde{\psi}(p) \text{ has dimensions } [\text{momentum}]^{-1/2} \quad (\text{D.19})$$

$$\hbar \text{ has dimensions } [\text{length} \times \text{momentum}] \text{ or } [\text{energy} \times \text{time}] \quad (\text{D.20})$$

**Energy eigenfunction sketching:** (one dimension)

$$n\text{th excited state has } n \text{ nodes} \quad (\text{D.21})$$

$$\text{if classically allowed: regions of high } V(x) \text{ have large amplitude and long wavelength} \quad (\text{D.22})$$

$$\text{if classically forbidden: regions of high } V(x) \text{ have faster cutoff} \quad (\text{D.23})$$

**Infinite square well:** (width  $L$ )

$$\eta_n(x) = \sqrt{2/L} \sin kx \quad k = n\pi/L \quad n = 1, 2, 3, \dots \quad (\text{D.24})$$

$$E_n = \frac{\hbar^2 k^2}{2m} = n^2 \frac{\pi^2 \hbar^2}{2mL^2} \quad (\text{D.25})$$



**Simple harmonic oscillator:** ( $V(x) = \frac{1}{2}Kx^2$ ,  $\omega = \sqrt{K/m}$ )

$$E_n = (n + \frac{1}{2})\hbar\omega \quad n = 0, 1, 2, \dots \quad (\text{D.26})$$

$$[\hat{a}, \hat{a}^\dagger] = \hat{1} \quad (\text{D.27})$$

$$\hat{H} = \hbar\omega(\hat{a}^\dagger\hat{a} + \frac{1}{2}) \quad (\text{D.28})$$

$$\hat{a}|n\rangle = \sqrt{n}|n-1\rangle \quad (\text{D.29})$$

$$\hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle \quad (\text{D.30})$$

$$\hat{x} = \sqrt{\hbar/2m\omega}(\hat{a} + \hat{a}^\dagger) \quad (\text{D.31})$$

$$\hat{p} = -i\sqrt{m\hbar\omega/2}(\hat{a} - \hat{a}^\dagger) \quad (\text{D.32})$$

**Coulomb problem:**

$$E_n = -\frac{\text{Ry}}{n^2} \quad \text{Ry} = \frac{m_e(e^2/4\pi\epsilon_0)^2}{2\hbar^2} = 13.6 \text{ eV} \quad (\text{D.33})$$

$$a_0 = \frac{(e^2/4\pi\epsilon_0)}{2\text{Ry}} = 0.0529 \text{ nm} \quad (\text{Bohr radius}) \quad (\text{D.34})$$

$$\tau_0 = \frac{\hbar}{2\text{Ry}} = 0.0242 \text{ fsec} \quad (\text{characteristic time}) \quad (\text{D.35})$$

**Angular momentum:**

$$[\hat{J}_x, \hat{J}_y] = i\hbar\hat{J}_z, \text{ and cyclic permutations} \quad (\text{D.36})$$

The eigenvalues of  $\hat{J}^2$  are

$$\hbar^2 j(j+1) \quad j = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots \quad (\text{D.37})$$

For a given  $j$ , the eigenvalues of  $\hat{J}_z$  are

$$\hbar m \quad m = -j, -j+1, \dots, j-1, j. \quad (\text{D.38})$$

The eigenstates  $|j, m\rangle$  are related through the operators

$$\hat{J}_+ = \hat{J}_x + i\hat{J}_y \quad \hat{J}_- = \hat{J}_x - i\hat{J}_y \quad (\text{D.39})$$

by

$$\hat{J}_+|j, m\rangle = \hbar\sqrt{j(j+1) - m(m+1)}|j, m+1\rangle \quad (\text{D.40})$$

$$\hat{J}_-|j, m\rangle = \hbar\sqrt{j(j+1) - m(m-1)}|j, m-1\rangle. \quad (\text{D.41})$$

**Spherical harmonics:**

A “function on the unit sphere” is a function  $f(\theta, \phi)$ . Another convenient variable is  $\mu = \cos \theta$ . “Integration over the unit sphere” means

$$\int d\Omega f(\theta, \phi) = \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi f(\theta, \phi) = \int_{-1}^{+1} d\mu \int_0^{2\pi} d\phi f(\theta, \phi).$$

$$\nabla^2 Y_\ell^m(\theta, \phi) = -\frac{1}{r^2} \ell(\ell + 1) Y_\ell^m(\theta, \phi) \quad (\text{D.42})$$

$$\int Y_{\ell'}^{m'*}(\theta, \phi) Y_\ell^m(\theta, \phi) d\Omega = \delta_{\ell', \ell} \delta_{m', m} \quad (\text{D.43})$$

$$f(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} f_{\ell, m} Y_\ell^m(\theta, \phi) \quad \text{where (D.44)}$$

$$f_{\ell, m} = \int Y_\ell^{m*}(\theta, \phi) f(\theta, \phi) d\Omega \quad (\text{D.45})$$

# Index

- ↑↓ symbols, 77, 82
  
- absorption of radiation, 343
- action at a distance, spooky, 44
- Aharonov, Yakir, 32
- Aharonov-Bohm effect, 31–32
- ambivate, 25
- amplitude, 57
  - peculiarities of, 61, 64, 67, 72
  - pronunciation of, 59
  - symbol for, 59
- analogy, 1, 72
- analyzer loop, 22
- analyzer, Stern-Gerlach, 12
- approximation
  - controlled, 197
  - uncontrolled, 197
- atomic units, 272–275
  
- baum, 351
- Bell's Theorem, 45
- Bell, John, 45
- blackbody radiation, 5
- Bloch, Felix, 158
- Bohm, David, 32
- Bohr magneton, 9
- Bohr radius, 273
- Born, Max, 46
- Bose, Satyendra, 254
- boson, 254
  
- characteristic equation, 367
- characteristic matrix, 367
- classical limit of quantum mechanics,
  - 2
- Clebsch, Alfred, 303
- collapse of the state vector, 78
- complex unit, 73
- configuration space, 158
- continuous, 5
- controlled approximation, 197
- conundrum of projections, 10–19
- Coulomb gauge, analogous to quantal state, 79
  
- definite value
  - use of, 180
- degenerate, 233
- delight, 3
- despair, 3, 6
- dimensional analysis, 272–275
- Dirac notation, 56
- Dirac, P.A.M., 56
- disparaging term attached to
  - charming result, 233
  
- effective potential energy function,
  - 232
- Ehrenberg, Werner, 32
- Ehrenfest, Paul, 171
- eigenvalues, 363, 365
- eigenvectors, 363, 365

- Einstein A and B argument, 343–345
- Einstein, Albert, 36, 37, 44, 53, 71, 343–345
- entanglement, 44–47, 50–52, 77–81
  - the word, 44, 47
- EPR, 37, 44
- exchange, 251
  
- Fermi, Enrico, 254, 336
- fermion, 254
- find-the-flaw problems, 29–30
- football, 26
- Frank–Hertz experiment, 6
- Franz, Walter, 32
  
- Galileo Galilei, 45
- Gerlach, Walter, 7
- global phase freedom, 73, 75
- Gordan, Paul, 303
  
- Hückel, Erich, 158
- Hamilton, Alexander, 121
- Hamilton, William Rowan, 120
- Hermite, Charles, 108
- Hilbert space, 71
- Hilbert, David, 71
- history of quantum mechanics, 32, 82
- Holy Office of the Inquisition, 45
- Hund’s rules, 305
- Hund, Friedrich, 305
  
- indeterminacy, 18, 25
- Inquisition, Holy Office of the, 45
- interchange, 251
- interference, 25, 339
  - constructive, 26
  - the physical manifestation of
    - superposition, 72
- intuition regarding quantum mechanics, 45–47, 82
  
- Johnson-Trotter sequence of permutations, 255
  
- Kelvin, 45
- ket, 56
  
- Laguerre, Edmond, 242
- language, 2, 18, 25–26, 34, 44, 47, 57, 62, 64
  - for amplitude, 59
  - swap v. exchange v. interchange, 251
- Legendre, Adrien-Marie, 236
- Liouville, Joseph, 231
- love, 179
  - color of, 18, 25, 78
  
- marble, 46, 179, 180
- matrix mathematics, 351–376
- measurement, 34
- measurement disturbs the system”, 18
- metaphor, 18, 35, 46
- misconceptions, 178
  - a “wheels and gears” mechanism
    - undergirds quantum mechanics, 46
  - all states are energy states, 4
  - amplitude is physically “real”, 4, 61–62, 67, 78–79
  - atom can absorb light only if
    - $\hbar\omega = \Delta E$ , 338
  - collapse of the quantal state”
    - involves (or permits) instantaneous communication, 78–79
  - electron is a small, hard marble, 25
  - indeterminate quantity exists but changes rapidly, 18
  - indeterminate quantity exists but changes unpredictably, 18
  - indeterminate quantity exists but is disturbed upon measurement, 18
  - indeterminate quantity exists but knowledge is lacking, 4, 18, 44, 62
  - indeterminate quantity exists in random shares, 18
  - magnetic moment behaves like a classical arrow, 18
  - photon as ball of light, 46, 342
  - photon is a small, hard marble, 46

- quantum mechanics applies only to small things, 2
  - state of system given through states of each constituent, 78
- misconceptions, catalog of, 4
- natural units, 275
- Noether, Emmy, 303
- observation, 34
- orthogonal matrix, 353
- outer product, 358
- overall phase factor, 73, 75
- perfection, 261
- permutations, 255
- phase factor, 73
- phase freedom, global, 73, 75
- phase, overall, 73, 75
- photon, 341–342
- Picard, Émile, 332
- plain changes sequence of
  - permutations, 255
- Planck constant, 5
- Planck, Max, 5, 6
- Podolsky, Boris, 37
- politics, 10
- Pope, Alexander, 285
- probability, 19
- probability amplitude, 57
- PS
  - test and reflect on your solution, 29–30
- quantization, 5, 9, 10
- quantum computer, 83
- qubit systems, 52, 82
- Rabi, I.I., 127
- Renninger negative-result
  - experiment, 35, 179
- replicator, 31
- reversal conjugation theorem, 64
- richness, 1, 3
- Rosen, Nathan, 37
- rotation matrix, 353
- scaled variables, 272–275
- Schrödinger, Erwin, 44, 121
- shorthand, dangerous: say it but don't think it, 259
- Siday, Raymond, 32
- solutions, exact vs. approximate, 261
- spin, 82
- spin- $\frac{1}{2}$  systems, 52, 82
- spontaneous emission of radiation, 342, 343
- spooky action at a distance, 44
- state
  - definition, 56
  - of entangled system, 78
  - peculiarities of, 72, 76
- state vector, 71
- Stern, Otto, 7, 36
- Stern-Gerlach analyzer, 12
- stimulated emission of radiation, 343
- Sturm, Charles-François, 231
- Sturm-Liouville problems, 231
- superposition, 72, 86
  - the mathematical reflection of interference, 72
- Susskind, Leonard, 78
- swap, 251
- tentative character of science, 57–58
- Thomson, William, 45
- tree, 351
- triangle inequality, 65
- two-state systems, 52, 82
- uncontrolled approximation, 197
- vacuum state, 341
- visualization, 18, 25, 46, 180