# Table of Contents

I. The Schrödinger Equation ............................................................................................... 1  
   A. The Weirdness of Quantum Mechanics ................................................................. 1  
   B. The Schrödinger Equation .................................................................................... 6  
   C. The Meaning of the Wave Function ..................................................................... 9  
   D. The Fourier Transform of the Wave Function .................................................. 11  
   E. Expectation Values and Uncertainties ................................................................. 13  
Problems for Chapter 1 ................................................................................................. 15

II. Solving Schrödinger’s equation ................................................................................... 16  
   A. Solving Schrödinger’s Equation in Free Space .................................................... 16  
   B. The Time-independent Schrödinger equation ..................................................... 18  
   C. Probability Current ............................................................................................... 20  
   D. Reflection from a Step Boundary ....................................................................... 22  
   E. Quantum Tunneling .............................................................................................. 25  
   F. The Infinite Square Well ...................................................................................... 26  
   G. Bound States from a Double Delta potential .................................................... 28  
Problems for Chapter 2 ................................................................................................. 32

III. Hilbert Space and Vector Spaces ................................................................................ 33  
   A. Hilbert Space ......................................................................................................... 33  
   B. Dirac Notation and Covectors .......................................................................... 35  
   C. Coordinate Bases .................................................................................................. 38  
   D. Operators ............................................................................................................. 40  
   E. Commutators ....................................................................................................... 42  
   F. Adjoint Operators ................................................................................................. 43  
   G. Matrix Notation .................................................................................................... 46  
   H. Eigenvalues and Eigenvectors .......................................................................... 48  
   I. Finding Eigenvectors and Eigenvalues .............................................................. 50  
Problems for Chapter 3 ................................................................................................. 53

IV. The Postulates of Quantum Mechanics ...................................................................... 55  
   A. Revisiting Representations ............................................................................... 55  
   B. The Postulates of Quantum Mechanics ............................................................ 56  
   C. Consistency of the Postulates ............................................................................ 59  
   D. Measurements and Reduction of the Wave Function ....................................... 61  
   E. Expectation Values and Uncertainties ............................................................... 64  
   F. Evolution of Expectation Values ........................................................................ 66  
   G. The Time-Independent Schrödinger Equation .................................................. 67  
Problems for Chapter 4 ................................................................................................. 70

V. The Harmonic Oscillator .............................................................................................. 72  
   A. The 1D Harmonic Oscillator ............................................................................. 72  
   B. Working with the Harmonic Oscillator and Coherent States ......................... 75  
   C. Multiple Particles and Harmonic Oscillators .................................................... 78  
   D. The Complex Harmonic Oscillator ................................................................... 81  
Problems for Chapter 5 ................................................................................................. 83
VI. Symmetry .................................................................................................................... 84
   A. The Translation Operator ......................................................................................... 84
   B. Rotation Operators ................................................................................................... 86
   C. Parity and the Finite Square Well ............................................................................ 90
   D. Discrete Rotations .................................................................................................... 93
   E. Discrete Translations and the Kronig-Penney Model .............................................. 95
   F. Continuous Symmetries and Generators .................................................................. 98
   G. Solving Problems with Continuous Symmetries ................................................... 100
Problems for Chapter 6 ............................................................................................... 104

VII. Angular Momentum ................................................................................................ 106
   A. Angular Momentum Commutation Relations ........................................................ 106
   B. Generalized Angular Momentum ........................................................................... 108
   C. Spherically Symmetric Problems ........................................................................... 110
   D. Spherical Harmonics .............................................................................................. 114
   E. The Hydrogen Atom ............................................................................................... 116
   F. Hydrogen-like atoms .............................................................................................. 122
   G. Tables of Spherical Harmonics and Radial Wave Functions ................................. 124
Problems for Chapter 7 ............................................................................................... 125

VIII. Spin and Adding Angular Momentum ................................................................. 127
   A. Rotations revisited .................................................................................................. 127
   B. Total Angular Momentum and Addition of Angular Momentum .......................... 130
   C. Clebsch-Gordan Coefficients ................................................................................. 134
   D. Scalars, Vectors, Tensors ....................................................................................... 136
   E. The Wigner-Eckart Theorem .................................................................................. 140
   F. Products of Spherical Harmonics ........................................................................... 142
Problems for Chapter 8 ............................................................................................... 144

IX. Electromagnetic Forces ............................................................................................ 147
   A. Electric Forces ........................................................................................................ 147
   B. Electromagnetic Fields and Forces ........................................................................ 149
   C. Working with Kinematic Momentum .................................................................... 152
   D. Magnetic Dipole Moments ..................................................................................... 153
   E. Simple Problems with Magnetic Fields .................................................................. 155
   F. The Aharanov-Bohm Effect and Magnetic Monopoles .......................................... 158
Problems for Chapter 9 ............................................................................................... 161

X. Multiple Particles ....................................................................................................... 163
   A. Tensor Products of Vector Spaces ......................................................................... 163
   B. Multiple Particles ................................................................................................... 164
   C. Symmetry and Interchangeable Particles ............................................................... 165
   D. Identical Particles ................................................................................................... 167
   E. Non-Interacting Particles and Fermi Degeneracy Pressure .................................... 170
   F. Atoms ...................................................................................................................... 174
Problems for Chapter 10 ............................................................................................. 178
I. The Schrödinger Equation

A. The Weirdness of Quantum Mechanics

The world is a strange place, but it wasn’t until 1900, when Max Planck was studying the thermal spectrum of light (black body radiation) that we began to understand just how strange it was. Although this represented the formal start of modern quantum mechanics, it is easier to start in 1877, with Hertz’s discovery of the photoelectric effect, which was explained in 1905 by Einstein. This effect demonstrated, in effect, that when you shine a light source of frequency $f$ onto a piece of metal, the light acted not as if it were made of waves, but rather as if it were composed of tiny packets of energy $hf$, where $h$ was a new constant called Planck’s constant, given by

$$h = 6.626 \times 10^{-34} \text{ J} \cdot \text{s} = 4.136 \times 10^{-15} \text{ eV} \cdot \text{s}.$$  

This is the basis of the modern photomultiplier, a device that is capable of detecting individual particles of light, also known as photons. We aren’t really interested in the inner workings of a realistic photomultiplier, but instead will simply treat it as an idealized detector that counts photons.

Although the formula $E = hf$ was the formula originally formulated by Planck, we will find it more useful to rewrite this by writing the frequency $f$ in terms of the angular frequency $\omega = 2\pi f$, and then exchange Planck’s constant for the so-called reduced Planck constant,

$$\hbar = \frac{h}{2\pi} = 1.0546 \times 10^{-34} \text{ J} \cdot \text{s} = 6.582 \times 10^{-16} \text{ eV} \cdot \text{s}.$$  

In terms of which the relationship between energy and angular frequency is given by

$$E = \hbar \omega. \quad (1.1)$$

The basic problem with this discovery is that there was already ample evidence that light was electromagnetic waves (indeed, if they aren’t waves, then what frequency are we discussing?). For example, the electric field of an electromagnetic wave traveling in the $+x$ direction would take the form

$$\mathbf{E}(r,t) = E_0 \cos(kx - \omega t) \quad \text{or} \quad \mathbf{E}(r,t) = E_0 \sin(kx - \omega t). \quad (1.2)$$

For technical mathematical reasons, which represent a convenience when studying electricity and magnetism but will prove crucial here, it is common to combine these two real functions by multiplying the latter by $i$ and then adding them to get the complex field

$$\mathbf{E}(r,t) = E_0 \exp[i(kx - \omega t)]. \quad (1.3)$$

This form of the electric field (together with a corresponding magnetic field) can be shown to satisfy Maxwell’s equations provided $\omega = ck$. With a little help from relativity, another important relationship can be worked out. For particles moving at the speed of light (such as photons), the energy and momentum are related by $E = cp$, where $c$ is the
speed of light and \( p \) is the momentum. Combining these two relations with (1.1), it is a moment’s work to demonstrate

\[
p = \hbar k. \tag{1.4}
\]

Waves have properties very different from particles; for example, particles have a very definite position, whereas waves tend to be spread out, or in essence, have some uncertainty in their position. The waves given by (1.2) or (1.3) have a very definite value of the wave number \( k \), but they are infinite in extent, so that the position \( x \) has no definite value; indeed, for (1.3) the waves are distributed uniformly throughout all of space. Without going into too much detail here, it can be shown that you can combine waves with different wave numbers \( k \) to make a wave packet that is less spread out in space but has an uncertainty in wave number \( k \). Indeed, it is a classical result that for any wave the uncertainties in the position \( x \) and wave number \( k \) must satisfy the inequality

\[
(\Delta x)(\Delta k) \geq \frac{\hbar}{2},
\]

where \( \Delta x \) and \( \Delta k \) represent the uncertainty in the position and wave number respectively. These uncertainties will be defined much more precisely later in the course. If we combine this with (1.4), we obtain a corresponding relationship between the momentum and position of a photon, namely

\[
(\Delta x)(\Delta p) \geq \frac{\hbar}{2}. \tag{1.5}
\]

In contrast, particles are normally described by giving their position and momentum, and hence it is implicitly assumed in classical mechanics that the uncertainty in each of these quantities is zero.

To help us understand some of the difficulties encountered as we move to a quantum picture of the world, a number of gedanken1 experiments may help to clarify the situation. For example, consider a half-silvered mirror; that is, a mirror that lets some light through, but not all. Now, suppose we place a dim source in front of this half-silvered mirror and shine a light on it, detecting the resulting waves using a pair of photomultipliers, as illustrated in Fig. 1-1. What happens in this experiment? Perhaps not surprisingly, the rate of photons appearing in each of the two detectors is exactly half the rate at which they strike the half-silvered mirror. If you inject the photons from the source very slowly, say one at a time, you find that the photons appear in either detector A or detector B – not in both, but always in one of them2. It is as if the half-

---

1 The German word “gedanken” means thought. A gedanken experiment is a thought experiment – one that is not necessarily going to be performed, but instead is used to illustrate a particular point.

2 Actually, realistic detectors never have 100% efficiency, so in real experiments, some photons will not appear in either detector. Though this gedanken experiment cannot be perfectly performed, it is believed that nothing in principle makes such an experiment impossible.
silvered mirror sorts the photons one way or the other, even though there is apparently no way to predict which way any single photon will go.

It is perhaps worth pointing out a couple of reasons for our confusion. If you work out the effects on the wave functions (1.2) or (1.3) of the half silvered mirror, you would find that the wave gets divided equally into two waves, one of which goes into each detector. But experimentally, each photon goes one way or the other. If you changed the half-silvered mirror for one that reflects some different fraction, you will find that the photons again each flow either one way or the other, though the probabilities will shift. Indeed, the probability of the photons going to one detector or the other is proportional to the intensity, which in the complex notation of (1.3) yields

\[ P \propto |E(r, t)|^2 = E_0 \cdot E_0^* \quad (1.6) \]

Indeed, since the power of a wave is proportional to this same quantity, and each of the photons has exactly the same energy (for a plane wave), this result is hardly surprising.

Another experiment will clarify things even more. It is possible, with the addition of a couple of conventional mirrors and an additional half-silvered mirror, to create a configuration along the lines of Fig. 1-2. The light comes from the source, is divided in two by the half-silvered mirror. The light then continues to the two conventional mirrors, which then send it to a half-silvered mirror, which recombines the two waves. It is possible, if the distances are carefully controlled, to arrange things so that according to classical wave theory the waves will combine perfectly at the second half-silvered mirror such that all of the light goes into detector A and none of it goes into detector B. It is important to note that this happens, according to the wave theory, only if the light wave follows both paths; it is the interference of the two halves of the waves at the recombining half-silvered mirror that allows this effect: the two waves conspire to cancel out as they head towards detector B, and they conspire to add together (interfere constructively) as they head towards detector A.

If we do this experiment with light that is low intensity, it is difficult to reconcile the wave and particle pictures. In this picture, half the photons will be deflected by the first half-silvered mirror, and half will pass through. Thus each photon will follow either the upper path or the lower path, not both, and therefore will reach the second half-silvered mirror from the upper path or the lower path, but not both. Since we already know that half-silvered mirrors reflect half the photons and pass the other half, we would expect the photons to again be divided in half, so in the end there are four equally likely paths from the source to the detector, two of them ending up in detector A and two in detector B. But this is not what we find experimentally. We find, instead, that the
photons all end up in detector A, and none of them in detector B. Recall, in the wave theory, this can only happen if the wave passes along both paths, but how can we explain this if we have only one particle at a time? It is as if the particle behaves like a wave between the first mirror and the second, and then behaves like a particle after passing through the second.

To clarify the situation, imagine replacing one of the two conventional mirrors in Fig. 1-2 by something that absorbs photons, say, another photomultiplier, as shown in Fig. 1-3. In this configuration, the system acts exactly as predicted if photons are particles: half of them are deflected by the first half-silvered mirror into detector C, and the other half are subdivided by the second mirror into half, so that ultimately one-fourth of them end up in detectors A and B. But when the second mirror is replaced, the photons disappear again. In this case, adding a path by which the particles could reach detector B causes a decrease in the photons arriving there.

It is important to remember that the interference described in Fig. 1-2 can only occur if the two path lengths are carefully controlled. In particular, if one of the mirrors is repositioned by, say, a modest fraction of a wavelength of light, the interference will be imperfect or even completely destroyed, and many photons will make it into detector B. Fortunately, modern optical equipment can easily be adjusted much more accurately than a fraction of a wavelength, so this experiment (or comparable equivalent experiments) can easily be performed.

What we would like to do is find an experiment where we can actually catch quantum mechanics in the act, so to speak, where we can both measure the path that the light took (upper or lower?) and also we can create interference effects, such as are shown in Fig. 1-2. This proves to be surprisingly difficult. For example, consider the following minor modification of Fig. 1-2: replace one (or both) of the conventional mirrors with a very lightweight mirror; indeed, let’s make it so light that when a single photon bounces off of it, it will cause the mirror to recoil with some measurable velocity.\footnote{Not surprisingly, this experiment is only a gedanken experiment; such a small recoil from a macroscopic mirror cannot actually be measured. What is surprising is that this experiment will prove to be theoretically impossible as well. Fundamentally, any such experiment is doomed to failure.}

Now we inject a single photon into the apparatus, let it pass through and check whether it registered in detector A or B. We also carefully measure the momentum of one of the mirrors after the experiment is done, to see whether the photon rebounded from it, transferring momentum to it. If we set up the distances accurately, then the photon should always appear in detector A, never in B, and if necessary, we can repeat the experiment several times to check that it actually never does appear in B. In the mean
time, for each repetition of the experiment, we can check which path the photon took. Since the interference requires that the photon take both paths, and measuring the recoil of the mirror requires that the photon take one path or the other, it seems we should be able to catch nature in a bind. What will happen?

The answer turns out to be surprisingly subtle. Recall, as I have said before, that the mirror must be placed very carefully to make sure the interference works; indeed, it must be positioned much better than a wavelength. In other words, if there is some uncertainty in the mirror’s position, because we were sloppy in setting up the experiment, we had to be very careful that this uncertainty satisfies

\[ k \Delta x_m \ll \lambda, \]

or using the relationship

\[ 2 \pi \Delta x_m \ll \lambda, \]

we must have

\[ k \langle \Delta x \rangle_m \ll 2 \pi, \]

where \( k \) is the wave number of the light, and \( \Delta x \) is the uncertainty in the position of the mirror.

Of course, we are also trying to measure an incredibly tiny momentum change in the mirror. When it reflects a single photon the photon transfers a momentum of order \( \hbar k \) to the mirror. We must be certain there are no other forces on the mirror that are larger; for example, errant breezes. This can generally be controlled by placing the experiment in vacuum, etc., but we must in addition be very careful that the tiny mirror is not moving with any significant momentum initially, otherwise we might mistake this initial motion for the effects of the recoil of the photon. In particular, we had better make sure that the momentum is known with accuracy smaller than the momentum of the photon, so we need

\[ \langle \Delta p \rangle_m \ll \hbar. \]

We see, therefore, that the initial momentum and position of the mirror must both be known to high accuracy, or the experiment is doomed to failure. Now, if we combine equations (1.7) and (1.8) by multiplying them, we find that the experiment’s success will require

\[ \langle \Delta x \rangle_m \langle \Delta p \rangle_m \ll 2 \pi \hbar. \]

The exact meaning of the “much less than” symbol in (1.9) is ambiguous. Roughly it came about because we insisted that we not have any photons passing into detector B. If, for example, we let the combination \( k \langle \Delta x \rangle_m \) in equation (1.7) be equal to \( \frac{\sqrt{2}}{4} \pi = 1.1 \), we can show that the interference effects are completely destroyed, and half the photons will go into each detector, rather than all into A. So we want inequality (1.9) to be quite a bit smaller, perhaps as small as 0.5 or even less.

Now, if we could do the experiment described, we really would be able to detect which way the photon went and prove that it went both ways. Perhaps not surprisingly, the experiment fails, and the reason is that Eq. (1.9) simply cannot be satisfied. It is impossible to specify the position and momentum of a mirror, even theoretically, with arbitrary precision. The reason is that quantum mechanics applies not only to light, but also to mirrors as well. Mirrors must satisfy the uncertainty relationship (1.5), as well as photons. Indeed, as far as we know, (1.5) applies to electrons, protons, neutrons, atoms,
molecules, mirrors, and even planets. Quantum mechanics is *universal*. All of physics must be rewritten to incorporate quantum mechanics. Everything, and not just photons, acts like waves (at least sometimes), and our goal will be to come up with a wave theory that applies not just to light, but to everything.

Although it is possible (at least in principle) to satisfy either of the two relationships (1.7) or (1.8), it is impossible to satisfy both. If we satisfy (1.7), then interference phenomena will occur, and all (or the vast majority) of the photons will fall into detector A, but we will be incapable of measuring which of the two paths the photon took. If we satisfy (1.8), then we will successfully measure which way the photon went, but the uncertainty in the position of the mirror will destroy the interference. In essence, the process of measuring where the photon went affects the photon. Sometimes it is stated that quantum mechanics simply states that measuring systems disturbs them, changes them, but quantum mechanics is far more profound than this simple statement.

That other particles besides light have wave-like properties was first proposed by deBroglie in 1924, and confirmed in the case of electrons by Davisson and Germer in 1928. In 1930, it was demonstrated for atoms and molecules, and since then for protons and neutrons. It is now believed that everything has both particle and wave properties, though in some cases (gravity, for example), this has not been experimentally demonstrated.

The deBroglie hypothesis was that Eq. (1.4) applies to electrons, and shortly thereafter it was similarly speculated that Eq. (1.1) applies as well. We will simply assume that these two relations apply to *all* particles.

B. The Schrödinger Equation

In late 1925, Erwin Schrödinger was giving a talk about the deBroglie hypothesis to a group of colleagues, when Debye suggested that if you are going to use wave relations to describe electrons, you probably should develop a wave equation, something akin to Maxwell’s equations, to describe that electron. This goal was achieved by Schrödinger, and we attempt here to ‘derive’ Schrödinger’s equation. To obtain this equation, we will have to make a number of assumptions and guesses. In the end, what assumptions we make are not really important; what is important is the result, and we can only ask ourselves whether the result is consistent with what is observed in the world around us, or if it is contradicted by it. Hence we won’t worry too much about the fine mathematical points of our derivation.

We start by assuming that an electron (even a single electron) in one dimension is described by a wave function $\Psi(x,t)$. Furthermore, we will assume, as Schrödinger did, that it is a *complex* wave function. In electricity and magnetism, the introduction of complex electric and magnetic fields is a mathematical convenience, but in quantum mechanics, it will turn out to be a necessity. We will assume that $\Psi(x,t)$ has a single component, unlike the electric or magnetic field, which as vector fields have three components, and also that in free space (no forces or interactions) waves similar to (1.3) will be solutions of our equations. In other words, we are looking for equations which are satisfied by waves of the form
\[ \Psi(x,t) = N \exp\left[ i (kx - \omega t) \right]. \] \hspace{1cm} (1.10)

There is no particular reason to think this will work for arbitrary \( k \) or \( \omega \). In fact, we know that for non-relativistic particles of mass \( m \), we expect there to be a relationship between the momentum \( p \) and the energy \( E \) given by

\[ E = \frac{p^2}{2m}. \] \hspace{1cm} (1.11)

Now, we would like to somehow relate equations (1.10) and (1.11) with the quantum mechanical relations (1.1) and (1.4). We want these in the form of a wave relationship; that is, we want expressions involving things like derivatives of the wave (1.10). Noting that when you take the derivative of (1.10) with respect to \( x \), you get a factor of \( ik \), it is easy to see that

\[ p\Psi(x,t) = \hbar k \Psi(x,t) = -i\hbar \frac{\partial}{\partial x} \Psi(x,t). \] \hspace{1cm} (1.12)

It is tempting to naively “cancel” the wave function from both sides of this equation, and write something like

\[ p \rightarrow -i\hbar \frac{\partial}{\partial x}. \] \hspace{1cm} (1.13)

I have written this with an arrow, rather than an equal sign, because I want to make clear that this should not be viewed as an equality, but rather as a transformation. When we move from the classical to the quantum mechanical, we replace the classical momentum with the momentum operator, the expression on the right. If we perform this replacement twice, it is easy to see that

\[ p^2\Psi(x,t) = (\hbar k)^2 \Psi(x,t) = \left( -i\hbar \frac{\partial}{\partial x} \right)^2 \Psi(x,t) = -\hbar^2 \frac{\partial^2}{\partial x^2} \Psi(x,t). \] \hspace{1cm} (1.14)

In a similar manner, we can convert energy into a time derivative by using (1.1). We find

\[ E\Psi(x,t) = \hbar \omega \Psi(x,t) = i\hbar \frac{\partial}{\partial t} \Psi(x,t), \] \hspace{1cm} (1.15)

which suggests that when we “quantize” a theory, we make the similar substitution

\[ E \rightarrow i\hbar \frac{\partial}{\partial t}. \] \hspace{1cm} (1.16)

We now use (1.11), a classical equation, to relate (1.14) and (1.16) and produce a full wave equation.

\[ E\Psi(x,t) = \frac{p^2}{2m} \Psi(x,t) \rightarrow i\hbar \frac{\partial}{\partial t} \Psi(x,t) = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \Psi(x,t). \]

This, then, is Schrödinger’s one-dimensional equation for a free non-relativistic particle. To clarify the steps we went through in deriving it, it is helpful to put them in the proper order, rather than the hodgepodge derivation we just completed. These steps are:
(1) Start with an equation for the energy in terms of the momentum and position,¹
(2) Multiply it on the right side by the wave function $\Psi(x,t)$,
(3) Replace the momentum by the differential operator (1.13), and
(4) Replace the energy by the differential operator (1.16).

For example, let us apply these steps for a non-free particle, one that is acted on by a force. We need to start with an expression for the energy. Most forces can be written as the derivative of some potential function; that is, in one dimension,

$$F(x,t) = -\frac{\partial}{\partial x} V(x,t).$$

Then the potential just contributes to the energy, and the total energy is

$$E = \frac{p^2}{2m} + V(x,t).$$

Following our prescription, we multiply this on the right by the wave function,

$$E \Psi(x,t) = \frac{p^2}{2m} \Psi(x,t) + V(x,t) \Psi(x,t).$$

We then replace $E$ and $p$ by differential operators (1.13) and (1.15) to give us

$$i\hbar \frac{\partial}{\partial t} \Psi(x,t) = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \Psi(x,t) + V(x,t) \Psi(x,t), \quad (1.17)$$

and we have Schrödinger’s equation, now with a potential.

What do we do if we are in more than one dimension? Well, we simply start with the three-dimensional formula for energy

$$E = \frac{p_x^2}{2m} + V(r,t) = \frac{p_x^2 + p_y^2 + p_z^2}{2m} + V(r,t).$$

We now multiply by the wave function on the right, which, not surprisingly, now will be a function of all three space variables. What we do with the three momentum expressions is not specified by (1.12), but it isn’t hard to guess that the correct procedure is to replace $p_x \rightarrow -i\hbar (\partial/\partial x)$, $p_y \rightarrow -i\hbar (\partial/\partial y)$, and $p_z \rightarrow -i\hbar (\partial/\partial z)$, or more succinctly,

$$p \rightarrow -i\hbar \nabla. \quad (1.18)$$

It is then only a minute’s work to find that Schrödinger’s equation for a particle in 3D takes the form

$$i\hbar \frac{\partial}{\partial t} \Psi(r,t) = -\frac{\hbar^2}{2m} \nabla^2 \Psi(r,t) + V(r,t) \Psi(r,t). \quad (1.19)$$

¹ In the example done already, only the momentum appeared, but in more general cases the position will also appear. The important thing to note is that time derivatives do not appear; the velocity must be first rewritten in terms of the momentum. As we will see later, time may also occasionally appear.
Other possibilities come to mind; for example, what if we have more than one particle? Can we deal with the possibility of having an unknown number of particles? What if the particles are relativistic? What if there are non-conservative forces, like magnetism? Is it possible that the wave function has multiple components, much as the electric and magnetic fields do? All of these are issues that we will deal with in time, but for now we will treat (1.17) and (1.19) as Schrödinger’s equation in 1D and 3D. They will provide ample examples of interesting and sometimes difficult problems to solve, but for now, let us set these aside and ask a bit about the interpretation of the wave function \( \Psi \).

Before we move on, one minor comment is in order. Equations (1.17) and (1.19) are complex equations, because they explicitly contain the imaginary number \( i \). Even if we made the wave function \( \Psi \) real at the initial time \( t = 0 \), it is easy to show that Schrödinger’s equation demands that it will acquire an imaginary part at other times. Hence complex numbers are a necessity, not merely a convenience.

C. The Meaning of the Wave Function

We have derived a wave equation (actually, two different ones) for the wave function \( \Psi (x,t) \), but we never said what \( \Psi (x,t) \) is. In contrast, for example, in electromagnetism, we can define the electric field \( E(r,t) \) as the force on a small test charge divided by that charge. Is there a similar interpretation of \( \Psi (x,t) \)?

The answer is no. There is no known method to actually measure the wave function. Or, at least, there is no known way to measure the wave function without changing it in the very act of measuring it. It can be shown that if we could somehow actually measure the wave function directly, we would be able to produce all kinds of miracles, such as communicating faster than light. But as far as we can tell, quantum mechanics does not allow one to measure the wave function itself.

What you can measure is the location of a particle. However, as I attempted to illustrate with all my gedanken experiments in section A, the wave function must often be more than one place at a time, even though the position of the particle must only be one place. This leads to confusion about what the wave function represents. If you look at equation (1.6) it seems reasonable to associate the square of the amplitude of the wave function with the probability of finding the particle there. For example, in one dimension, we assume the probability density to take the form

\[
\rho(x,t) = |\Psi(x,t)|^2 = \Psi^*(x,t)\Psi(x,t) .
\]

This has units of inverse length, since probability is dimensionless and we are in one dimension. In particular, if we have a region \( [a,b] \) in which we are going to look for the particle, the probability of finding the particle’s position \( x \) in this region will be given by

\[
P(a < x < b) = \int_a^b |\Psi(x,t)|^2 \, dx = \int_a^b \Psi^*(x,t)\Psi(x,t) \, dx .
\]
In particular, if we have one particle, and we look everywhere for it, the particle must be somewhere, so the probability of finding it somewhere must be 1, so we have
\[ 1 = \int_{-\infty}^{\infty} |\Psi(x,t)|^2 \, dx = \int_{-\infty}^{\infty} \Psi^*(x,t) \Psi(x,t) \, dx. \] (1.21)

Because the probability density is found by squaring the amplitude of the wave function, the wave function is sometimes called the probability amplitude.

In three dimensions, similar statements apply. The probability density (now with units of inverse length cubed) is found by squaring the amplitude.
\[ \rho(r,t) = |\Psi(r,t)|^2 = \Psi^*(r,t) \Psi(r,t). \] (1.22)

The probability of finding a particle in a region \( V \) of space is given by
\[ P(r \in V) = \iiint_V |\Psi(r,t)|^2 \, d^3r = \iiint_V \Psi^*(r,t) \Psi(r,t) \, d^3r. \]

If we integrate over all of space, the probability is one
\[ 1 = \iiint_V |\Psi(r,t)|^2 \, d^3r = \iiint_V \Psi^*(r,t) \Psi(r,t) \, d^3r. \] (1.23)

The interpretation of the wave function as a probability density is so distasteful to many physicists that many eschew this formulation of quantum mechanics entirely. For example, in Bohmian quantum mechanics, the wave function is called a pilot wave, a non-detectable function which nonetheless guides the motions of actual particles. In the many worlds interpretation, no interpretation at all is given to the wave function (it is what it is), and probabilities do not appear in the theory at all, or at best are derived as a limiting case. According to Ballentine, quantum mechanics should not be written in terms of the wave function itself, but rather in terms of the density matrix. However, throughout most of these notes I will treat the wave function as if it is real and use the traditional Copenhagen interpretation of quantum mechanics.

A few comments are in order. In classical mechanics, a particle is described at any time completely by giving its position \( x \) and its velocity \( v \), the time derivative of the position. It is not sufficient to give only the position, since its subsequent motion will depend on the initial velocity. In quantum mechanics, the particle is completely described ONLY by the initial wave function \( \Psi(x, t = 0) \). This is because Schrödinger’s equation, (1.17) or (1.19), are first order in time. Given the wave function at \( t = 0 \), we can determine from these equations the first time derivative of the wave function and therefore can determine the wave function a moment later. Repeating the process, we could actually numerically integrate Schrödinger’s equation and determine the wave function at arbitrary time (in some cases, this can actually be practical). Hence the wave function of the particle is completely described at any time by the wave function alone; we don’t need its time derivative.

Another minor concern has to do with the normalization condition (1.21) or (1.23). If we start with an arbitrary wave function at \( t = 0 \) and let it evolve using Schrödinger’s equation, how do we make sure that these equations are satisfied at subsequent times? The answer is that both versions of Schrödinger’s equation
automatically assure conservation of probability, and therefore no additional work is required to maintain the normalization conditions.

A more difficult problem, and one we will not entirely resolve, has to do with measurement. If a particle is described by a spread-out wave function, and we measure its position, what happens? If we measure the same particle repeatedly, will the particle jump from place to place erratically between measurements? The answer to the latter turns out to be no. If you measure a particle, no matter how spread out, and discover that it is at a point \( x \), and then immediately measure the position again, it will still be at \( x \). Of course, the particle may move between measurements, given enough time, so you may have to repeat the experiment pretty quickly (photons, for example, move at the speed of light). This implies that the process of measuring the wave function results in a change of the wave function. This process is described in chapter four.

One last comment has to do with the phase of the wave function. Suppose we take an arbitrary complex wave function \( \Psi \) and multiply it, for example, by a factor of two. The wave function will still satisfy Schrödinger’s equation, but will no longer satisfy the normalization conditions, so it cannot describe anything realistic. In contrast, suppose instead that we multiply the wave function by an arbitrary phase; that is, by a complex number of magnitude one, so we have

\[
\Psi(t) \rightarrow \Psi'(t) = e^{i\theta} \Psi(t),
\]

where \( \theta \) is an arbitrary real number. Then the probability density will remain unchanged

\[
\rho'(t) = |\Psi'(t)|^2 = |e^{i\theta} \Psi(t)|^2 = e^{-i\theta} \Psi^*(t) e^{i\theta} \Psi(t) = \Psi^*(t) \Psi(t) = \rho(t).
\]

In other words, if we detect the position of the particle, it will have the exact same probability distribution. It is also easy to see that if \( \Psi(t) \) satisfies Schrödinger’s equation (1.19), so also will \( \Psi'(t) \). Hence the two wave functions \( \Psi(t) \) and \( \Psi'(t) \) are physically indistinguishable, and we can treat them as identical. In some formulations of quantum mechanics, they are identical; we will treat them as if they are different in principle but identical experimentally.

D. The Fourier Transform of the Wave Function

Suppose at a given time the wave function in one dimension takes the form \( \Psi(x,t_0) = \psi(x) \). Consider now the Fourier transform, as given by equations (A.43):

\[
\tilde{\psi}(k) = \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}} \psi(x) e^{-ikx},
\]

\[
\psi(x) = \int_{-\infty}^{\infty} \frac{dk}{\sqrt{2\pi}} \tilde{\psi}(k) e^{ikx}.
\]

For example, suppose that our wave function takes the form \( \psi(x) \sim e^{ik_0 x} \). Then the Fourier transform will look like \( \tilde{\psi}(k) \sim \delta(k-k_0) \), so the Fourier transform will be
concentrated at $k = k_0$. Furthermore, the momentum, according to eq. (1.4), will be $\hbar k_0$.

Put simply, when the Fourier transform of the wave function is concentrated at a particular value of $k$, the momentum is simply $\hbar$ times that value of $k$.

Now, when the particle is spread out in $x$, we interpreted the wave function as a probability distribution for the position $x$. How do we interpret the Fourier transform wave function $\tilde{\psi}(k)$, when it is spread out in $k$? Probably not surprisingly, the answer is that it represents a probability distribution for the momentum $p$. From (1.24b), we see that if $\tilde{\psi}(k)$ is spread out in $k$, then the wave function $\psi(x)$ is simply a linear combination of many different waves of the form $e^{i k x}$. The square of the amplitude of the Fourier transform tells you the probability density that the momentum is around that value. If you integrate over different $k$ values, you can find out the probability that the momentum lies in a given range. One way to write this is

$$P(\hbar k_A < p < \hbar k_B) = \int_{k_A}^{k_B} dk |\tilde{\psi}(k)|^2.$$ 

Before accepting this equation as valid, the first thing we should check is that the interpretation of this as a probability makes sense. We should demand that the probability of the particle having some momentum is equal to 1, so we find

$$P(-\infty < p < \infty) = \int_{-\infty}^{\infty} dk |\tilde{\psi}(k)|^2 = \int_{-\infty}^{\infty} dx |\psi(x)|^2 = 1,$$

where we have used (A.44) together with the normalization constraint (1.21) to complete the proof.

Indeed, as we look at the wave function and its Fourier transform, there is a kind of complementary relationship between the two. Given either, we can find the other using (1.24). One describes the particle in terms of position, the other in terms of momentum. Both (in general) will represent only a probability distribution for the position or momentum. Either is a complete description of the particle. There is no particular reason to think of $\psi(x)$ as the actual wave function, and $\tilde{\psi}(k)$ as simply something mathematically related to it. Indeed, in the next chapter we will modify our notation so that we can refer to the wave function without specifying which of these two functions we are referring to.
E. Expectation Values and Uncertainties

As we have already said, when you measure the position or momentum of a particle you cannot generally predict exactly what position or momentum it will have. Instead, we get a distribution, a sort of list of probabilities. Although it is most descriptive to give a complete list of the possible probabilities, which for position means knowing $|\psi(x)|^2$, sometimes it is sufficient to ask what the average outcome will be if we perform a measurement.

The expectation value $\bar{a}$ of some quantity $a$ is the average of what you would get if you performed many measurements. If there are only discrete values of $a$ that can occur, the expectation value would take the form

$$\bar{a} = \sum_a a P(a),$$

where $P(a)$ is the probability of some outcome $a$ occurring, and the sum is taken over all values of $a$ that might occur.\(^1\) This sum is simply converted into an integral (and the probability $P$ replaced by the probability density $\rho$) when the variable $a$ is continuous, so we have

$$\bar{a} = \int a \rho(a) \, da.$$  

Let us apply this to the position and the momentum of a single particle in one dimension. The probability density for the particle to have position $x$ is $|\psi(x)|^2$, while for momentum the probability density for the particle to have momentum $\hbar k$ is $|\tilde{\psi}(k)|^2$, so we have

$$\bar{x} = \int x |\psi(x)|^2 \, dx,$$  

$$\bar{p} = \int \hbar k |\tilde{\psi}(k)|^2 \, dk.$$  

Of course, even though this is the average value, the value you get on any given measurement will tend to be different. It is possible to define uncertainties $\Delta x$ and $\Delta p$ as the root-mean-square difference between the expectation value and the value you get on particular measurements. Mathematically, these are given by

$$\Delta x = \left(\bar{x} - x\right)^2 \, dx,$$  

$$\Delta p = \left(\bar{p} - p\right)^2 \, dk.$$  

It is these two quantities that satisfy the famous Heisenberg uncertainty principle, Eq. (1.5), $\Delta x \Delta p \geq \frac{\hbar}{2}$, which we will prove in a later chapter.

The uncertainty principle goes a long way towards understanding how quantum mechanics can solve classical conundrums. Consider, for example, the hydrogen atom,

\(^1\) Note that in this sum, there is no harm in including values which actually cannot occur, since they will have probability zero.
which consists of a heavy proton (which we treat as stationary) and a light electron in orbit around it. Classically, the energy of the hydrogen atom is given by\(^1\)

\[
E = \frac{p^2}{2m} - \frac{k_e e^2}{r},
\]

(1.27)

where \(r\) is the separation of the electron from the proton, \(p\) is the momentum of the electron, and \(m\) is its mass. This is a problem, because the electron can be as close to the atom as it wants to, while keeping its momentum small, and hence the energy can be infinitely negative. Before the advent of quantum mechanics, it was not understood why the electron would not simply spiral into the proton, emitting electromagnetic energy. Quantum mechanically, however, it is impossible to both specify the position of the particle precisely and simultaneously limit the momentum. Suppose, for example, we want the electron to be “near” the proton. We want \(r = 0\), but we cannot demand this precisely, because the electron is a wave function. Instead we ask the electron to be within about a distance \(a\) of the proton. We might therefore approximate

\[
rx a \approx \Delta a \approx (1.28)
\]

We’d also like to get the momentum as low as possible, but because of (1.5), we can’t actually force it to be quite zero. Instead, we approximate

\[
p \approx \Delta p \approx \frac{\hbar}{2\Delta x} = \frac{\hbar}{2a}.
\]

(1.29)

Substituting (1.29) and (1.28) into (1.27), we find that the energy is given by

\[
E \approx \frac{\hbar^2}{8ma^2} - \frac{k_e e^2}{a}.
\]

(1.30)

We now see how quantum mechanics solves the problem. If we make the electron get too close to the proton, the kinetic energy term (the first term) will become very large, and we will end up with a positive energy. We can find the best possible size for the separation by finding where the derivative of (1.30) vanishes, which turns out to be

\[
a \approx \frac{\hbar^2}{4mk_e e^2}.
\]

Substituting this back into (1.30) yields an approximate energy

\[
E \approx -\frac{2mk_e^2 e^4}{\hbar^2}.
\]

This answer turns out to be too large by precisely a factor of four. Part of this can be attributed to the fact that we used an uncertainty relation in one dimension, when we are really working in three; the remaining (smaller) error is simply due to the fact that we worked with an approximation, so we shouldn’t expect to get the answer exactly right.

---

\(^1\) The exact form of the Coulomb interaction, the last term in (1.27), will depend on our choice of units. In this class we will always use SI units, so \(e\) is the charge on the proton and \(k_e\) is Coulomb’s constant.
Problems for Chapter 1

1. A particle of mass $m$ lies in one-dimension in a potential of the form $V(x) = Fx$, where $F$ is constant. The wave function at time $t$ is given by

$$\Psi(x,t) = N(t) \exp \left[ -\frac{1}{2} A(t) x^2 + B(t) x \right],$$

where $N$, $A$, and $B$ are all complex functions of time. Use Schrödinger’s equation to derive equations for the time derivative of the three functions $A$, $B$, and $N$. You do not need to solve these equations.

2. For each of the wave functions in one dimension given below, $N$ and $a$ are positive real numbers. Determine the normalization constant $N$ in terms of $a$, and determine the probability that a measurement of the position of the particle will yield $x > a$.
   (a) $\psi(x) = N/(x + ia)$.
   (b) $\psi(x) = N \exp(-|x|/a)$.
   (c) $\psi(x) = \left\{ \begin{array}{ll} N x^2 (x - 2a) & \text{for } 0 < x < 2a, \\ 0 & \text{otherwise.} \end{array} \right.$

3. An electron in the ground state of hydrogen has, in spherical coordinates, the wave function $\psi(r, \theta, \phi) = N e^{-r/a}$, where $N$ and $a$ are positive constants. Determine the normalization constant $N$ and the probability that a measurement of the position will yield $r > a$. Don’t forget you are working in three dimensions!

4. For each of the normalized wave functions given below, find the Fourier transform $\tilde{\psi}(k)$, and check that it satisfies the normalization condition $\int_{-\infty}^{\infty} |\tilde{\psi}(k)|^2 dk = 1$.
   (a) $\psi(x) = (A/\pi)^{1/4} \exp(iKx - \frac{1}{2} Ax^2)$.
   (b) $\psi(x) = \sqrt{\alpha} \exp(-\alpha |x|)$.

5. For each of the wave functions in question 4, find $\bar{x}$, $\Delta x$, $\bar{p}$, $\Delta p$, and check that the uncertainty relationship $(\Delta x)(\Delta p) \geq \frac{\hbar}{2}$ is satisfied.

6. A particle of mass $m$ lies in the harmonic oscillator potential, given by $V(x) = \frac{1}{2} m \omega^2 x^2$. Later we will solve this problem exactly, but for now, we only want an approximate solution.
   (a) Let the uncertainty in the position be $\Delta x = a$. What is the corresponding minimum uncertainty in the momentum $\Delta p$? Write an expression for the total energy (kinetic plus potential) as a function of $a$.
   (b) Find the minimum of the energy function you found in (a), and thereby estimate the minimum energy (called zero point energy) for a particle in a harmonic oscillator. Your answer should be very simple.
II. Solving Schrödinger’s equation

We now attempt to actually start solving Schrödinger’s equation in one or three dimensions. We start by writing it again (in one or three dimensions):

\[
\frac{i\hbar}{\partial t} \Psi(x,t) = \left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x,t)\right] \Psi(x,t), \\
\frac{i\hbar}{\partial t} \Psi(r,t) = \left[-\frac{\hbar^2}{2m} \nabla^2 + V(r,t)\right] \Psi(r,t).
\]  

(2.1a)

(2.1b)

The first thing I want to note about these equations is that they are linear; that is, the wave function always appears to the first power. Suppose we have managed, somehow, to find two solutions, \( \Psi_1(r,t) \) and \( \Psi_2(r,t) \), to equation (2.1b), then it is easy to show that any linear combination of these functions is also a solution, i.e., so also is the combination

\[
\Psi(r,t) = c_1 \Psi_1(r,t) + c_2 \Psi_2(r,t),
\]

where \( c_1 \) and \( c_2 \) are arbitrary complex constants. Obviously, this can be generalized to any number of functions.

A. Solving Schrödinger’s Equation in Free Space

We now attempt to actually start solving Schrödinger’s equation in three dimensions for the case where there is no potential, so

\[
\frac{i\hbar}{\partial t} \Psi(r,t) = -\frac{\hbar^2}{2m} \nabla^2 \Psi(r,t).
\]

(2.2)

We already know some solutions of this equation; specifically, \( \Psi(r,t) = e^{i\mathbf{k} \cdot \mathbf{r} - i\omega t} \) is known to satisfy Schrödinger’s equation provided \( \hbar \omega = \hbar^2 \mathbf{k}^2 / 2m \). We can add up an arbitrary number of such solutions with different values of \( \mathbf{k} \). Indeed, the most general solution would be to add up all values of \( \mathbf{k} \), in which case the sum becomes an integral, and we therefore consider the solution

\[
\Psi(r,t) = \iiint \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} c(\mathbf{k}) \exp \left( i\mathbf{k} \cdot \mathbf{r} - \frac{i\hbar \mathbf{k}^2}{2m} t \right)
\]

(2.3)

Where \( c(\mathbf{k}) \) is an arbitrary complex function of \( \mathbf{k} \). This is an enormous number of solutions; indeed, as we will demonstrate in a moment, this actually represents all

---

1 The factor of \( (2\pi)^{3/2} \) in the denominator is a convenience, which makes our work a tiny bit simpler without changing the outcome. In one dimension it would be \( \sqrt{2\pi} \).
solutions. To check that it really is a solution, merely substitute (2.3) back into (2.2) and check that the result is true. We find

\[ -\frac{\hbar^2}{2m} \nabla^2 \int f(k) \exp \left( i k \cdot r - \frac{i \hbar k^2 t}{2m} \right) = \int f(k) \exp \left( i k \cdot r - \frac{i \hbar k^2 t}{2m} \right). \]

The most general problem we can be given for a free particle is to specify the initial wave function \( \Psi(r, t=0) = \psi(r) \) and then be asked to find the wave function at subsequent times. Setting \( t = 0 \) in (2.3), we see that

\[ \psi(r) = \int f(k) e^{ik \cdot r}. \]

Comparison with the Fourier transform (A.45c) shows us that

\[ c(k) = \int \psi(r) e^{-ik \cdot r} \]

(2.4)

We can perform the Fourier transform (2.4) for arbitrary wave function \( \psi(r) \), and then substitute the result back into (2.3). Of course, actually performing the integration analytically may be difficult, but often numerical methods will work even when it is difficult or impossible to do these integrals in closed form. Equations (2.3) and (2.4) can be easily generalized to an arbitrary number of dimensions.

As an example, suppose a particle of mass \( m \) in free one-dimensional space has initial wave function

\( \psi(x) = N x \exp \left( -\frac{1}{2} A x^2 \right) \).

where \( A \) and \( N \) are positive real constants. To find the solution of Schrödinger’s equation (2.1a), we first find the constants \( c(k) \) using a formula analogous to (2.4):

\[ c(k) = \int \frac{dx}{\sqrt{2\pi}} e^{-ikx} \psi(x) = N \int \frac{dx}{\sqrt{2\pi}} x \exp \left( -\frac{1}{2} A x^2 - i k x \right). \]

(2.6)

This integral can be performed with the help of eq. (A.29):

\[ \int_{-\infty}^{\infty} x e^{-Ax^2 + Bx} dx = \sqrt{\frac{\pi}{A}} \frac{\partial}{\partial B} \left[ \exp \left( \frac{B^2}{4A} \right) \right] = \frac{B \sqrt{\pi}}{2 A^{3/2}} \exp \left( \frac{B^2}{4A} \right). \]

(2.7)

Making the substitutions \( A \to A/2 \) and \( B \to -ik \), you can easily complete the integral (2.6):

\[ c(k) = -\frac{Nik}{A^{3/2}} \exp \left( -\frac{k^2}{2A} \right). \]
Substituting this into the one-dimensional version of (2.3), we then have

\[ \Psi(x,t) = \int \frac{dk}{\sqrt{2\pi}} c(k) \exp\left(ikx - \frac{ihk^2}{2m} t\right) = -\frac{Ni}{\sqrt{2\pi}A^3} \int k \exp\left(-\frac{k^2}{2A}\right) \exp\left(ikx - \frac{ihk^2}{2m} t\right) dk \]

This integral is again of the type appearing in (2.6), except that the roles of \( x \) and \( k \) have been reversed, and this time we make the substitutions \( A \rightarrow 1/2A + iht/2m \) and \( B \rightarrow ix \), so we have

\[ \Psi(x,t) = -\frac{Ni}{A^{\frac{3}{2}} \sqrt{2\pi}} \frac{ix\sqrt{\pi}}{2(1/2A + iht/2m)^{\frac{3}{2}}} \exp\left(\frac{(ix)^2}{4(1/2A + iht/2m)}\right) \]

It is obvious that this satisfies the boundary condition (2.5); it is less obvious that it satisfies the Schrödinger equation (2.1a).

B. The Time-independent Schrödinger equation

Superposition is a very useful method for solving a wide variety of problems. It is particularly useful in cases when the Schrödinger equation (2.1a) or (2.1b) has no explicit time dependence; that is, when \( V \) is independent of time. Suppose we are asked to find solutions of the equation

\[ i\hbar \frac{\partial}{\partial t} \Psi(r,t) = \left[-\frac{\hbar^2}{2m} \nabla^2 + V(r)\right] \Psi(r,t). \tag{2.8} \]

Let us seek solutions of the form

\[ \Psi(r,t) = \psi(r) \phi(t). \tag{2.9} \]

Substitute this into (2.8), and then divide by both \( \psi(r) \) and \( \phi(t) \).

\[ i\hbar \psi(r) \frac{\partial}{\partial t} \phi(t) = \phi(t) \left[-\frac{\hbar^2}{2m} \nabla^2 + V(r)\right] \psi(r), \]

\[ \frac{i\hbar}{\psi(r)} \frac{\partial}{\partial t} \phi(t) = \frac{1}{\phi(t)} \left[-\frac{\hbar^2}{2m} \nabla^2 + V(r)\right] \psi(r). \tag{2.10} \]

Now, in (2.10), note that left side of the equation is independent of \( r \) and the right side is independent of \( t \). Since they are equal, they must both be independent of both \( r \) and \( t \). We give this constant a name \( E \), and write
\[
\frac{i\hbar}{\phi(t)} \frac{d}{dt} \phi(t) = E = \frac{1}{\psi(r)} \left[ -\frac{\hbar^2}{2m} \nabla^2 + V(r) \right] \psi(r).
\] (2.11)

The left equality is relatively easy to solve. If you multiply both sides by \(dt\) and integrate, you will find\(^1\)

\[
\phi(t) = e^{-iEt/\hbar}
\] (2.12)

This is an equation for a wave with angular frequency \(\omega = E/\hbar\), which allows us to identify \(E\) as the energy of the particle.

The right equality in equation (2.11) can be rewritten in what is known as the \textit{time-independent Schrödinger equation} (in three dimensions) which is

\[
\left[ -\frac{\hbar^2}{2m} \nabla^2 + V(r) \right] \psi(r) = E \psi(r).
\] (2.13)

Of course, we still have to solve (2.13), which will depend on the particular form of \(V(r)\). In general, there will be a list, generally an infinite list (sometimes discrete and sometimes continuous) of such solutions, which we denote with the index \(n\), each with its own energy \(E_n\), so that (2.13) should be written as

\[
\left[ -\frac{\hbar^2}{2m} \nabla^2 + V(r) \right] \psi_n(r) = E_n \psi_n(r).
\] (2.14)

In general, we start with the (hard) problem of solving (2.14), then we substitute the resulting wave functions and (2.12) into (2.9), and finally take an arbitrary linear combination of the resulting solutions. So the most general solution of (2.8) will be

\[
\Psi(r, t) = \sum_n c_n \psi_n(r) e^{-iE_n t/\hbar}.
\] (2.15)

We will leave the problem of finding the constants \(c_n\) for a general initial wave function \(\Psi(r, t = 0)\) to a later chapter. In general, however, we will treat Schrödinger’s equation as solved as soon as we find the wave functions \(\psi_n(r)\) and corresponding energies \(E_n\).

What is the advantage of solving the time-independent Schrödinger equation (2.13) over simply solving the original Schrödinger equation? The primary advantage is that there is one less variable than in the original equation. The reduction of the number of variables is a vast advantage, and can often turn a seemingly intractable problem into a relatively simple one. For example, suppose we are working in one dimension, in which the time-independent Schrödinger equation takes the form

\[
\left[ -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] \psi(x) = E \psi(x).
\] (2.16)

\(^1\) Technically, there is a constant of integration we have ignored here. This constant would become a multiplicative constant in equation (2.12). Ultimately, this multiplicative constant will be included when we use superposition to create the most general solution to (2.12), so there is no harm in momentarily leaving it out.
This is an ordinary differential equation, with no partial derivatives. Numerical methods for solving such ordinary differential equations are fast and highly accurate, so even if we don’t manage to solve (2.16) analytically, we can effectively solve it numerically. As an additional advantage, (2.13) and (2.16) are both real equations, so when solving them we can set aside complex numbers and work with exclusively real functions.

C. Probability Current

In chapter I, section C, we claimed without proof that the normalization conditions (1.21) or (1.23) are preserved by the Schrödinger equation, (2.1a) or (2.1b). We will now demonstrate this and discover an important new concept, the probability current. We start by taking the Schrödinger equation (2.1b) and multiplying by $\psi^*(\mathbf{r},t)$ on the left. We then take the complex conjugate of this equation:

$$\left(\frac{\partial}{\partial t} - \frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r},t)\right)\psi(\mathbf{r},t) = 0,$$

$$\left(\frac{\partial}{\partial t} - \frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r},t)\right)\psi^*(\mathbf{r},t) = 0.$$

We now subtract these two equations. Note that there will be a cancellation of the potential term; this cancellation occurs because $V(\mathbf{r},t)$ is real.

$$i\hbar \left[\psi^*(\mathbf{r},t)\frac{\partial}{\partial t}\psi(\mathbf{r},t) + \psi(\mathbf{r},t)\frac{\partial}{\partial t}\psi^*(\mathbf{r},t)\right] = \frac{\hbar^2}{2m} \left[\psi^*(\mathbf{r},t)\nabla^2\psi^*(\mathbf{r},t) - \psi^*(\mathbf{r},t)\nabla^2\psi(\mathbf{r},t)\right].$$

(2.17)

We note that the left side of this equation is simply a time derivative of the probability density $\rho(\mathbf{r},t) = \psi^*(\mathbf{r},t)\psi(\mathbf{r},t)$. Also, the right side can be written in terms of a divergence, by noting that

$$\nabla \cdot \left[\psi^*(\mathbf{r},t)\nabla\psi(\mathbf{r},t)\right] = \nabla\psi(\mathbf{r},t) \cdot \nabla\psi^*(\mathbf{r},t) + \psi(\mathbf{r},t)\nabla^2\psi^*(\mathbf{r},t),$$

$$\nabla \cdot \left[\psi(\mathbf{r},t)\nabla\psi^*(\mathbf{r},t)\right] = \nabla\psi^*(\mathbf{r},t) \cdot \nabla\psi(\mathbf{r},t) + \psi^*(\mathbf{r},t)\nabla^2\psi(\mathbf{r},t),$$

$$\nabla \cdot \left[\psi^*(\mathbf{r},t)\nabla\psi(\mathbf{r},t) - \psi(\mathbf{r},t)\nabla\psi^*(\mathbf{r},t)\right] = \left[\psi^*(\mathbf{r},t)\nabla^2\psi^*(\mathbf{r},t) - \psi^*(\mathbf{r},t)\nabla^2\psi(\mathbf{r},t)\right].$$

(2.18)

We now define the probability current $\mathbf{j}(\mathbf{r},t)$ by the equation

$$\mathbf{j}(\mathbf{r},t) = \frac{\hbar}{2mi} \left[\psi^*(\mathbf{r},t)\nabla\psi(\mathbf{r},t) - \psi(\mathbf{r},t)\nabla\psi^*(\mathbf{r},t)\right].$$

or the more convenient form

$$\mathbf{j}(\mathbf{r},t) = \frac{\hbar}{m} \text{Im} \left[\psi^*(\mathbf{r},t)\nabla\psi(\mathbf{r},t)\right].$$

(2.19)

Substituting the definition into (2.17) and using (2.18), we see that
Equation (2.20) is a local version of conservation of probability. There is an identical equation in electromagnetism, where $\rho(r,t)$ is the charge density, and $j(r,t)$ is electric current. To see more clearly that (2.20) is conservation of probability, integrate it over a volume $V$, which might represent all of space.

\[
\frac{d}{dt} \iiint_V d^3r \rho(r,t) = -\iiint_V d^3r \nabla \cdot j(r,t).
\]

On the left side, the integral is the probability of a particle lying in the region $V$. On the right side, use Gauss's Law (A.10) to rewrite it as a surface integral.

\[
\frac{d}{dt} P(r \in V) = -\oiint_S \hat{n} \cdot j(r,t) dS.
\]

What (2.21) says is that the total probability of the particle lying in a given volume changes only due to the flow of the probability into or out of the surface of the region. In particular, if the volume $V$ is all of space, and if the wave function is small at infinity, then the right hand side of (2.21) will vanish, and the total probability will remain a constant (equal to 1). So if the wave function is normalized so that (1.23) is satisfied at one time, it will always be satisfied. Note that (2.19) and (2.20) can be trivially generalized to an arbitrary number of dimensions.

To get a better handle on probability current, consider for example the plane wave $e^{ikx-\omega t}$. The probability density at all places and times can be calculated trivially to be $\rho(r,t) = |\Psi|^2$, which might lead one to mistakenly believe that there is no current flow at all. But the current density is not zero; indeed its value is

\[
j(r,t) = \frac{\hbar}{m} |\Psi|^2 \text{Im}[e^{-ikr+i\omega t} \nabla e^{ikr-i\omega t}] = \frac{\hbar}{m} |\Psi|^2 \text{Im}[i ke^{-ikr+i\omega t} e^{ikr-i\omega t}] = \frac{\hbar k}{m} \rho(r,t).
\]

Since $\hbar k$ is the momentum of such a particle, this is just $p/m$, the classical velocity, which means only that the resulting wave has its probability flowing at a steady rate, as we might expect.
D. Reflection from a Step Boundary

Consider a particle impacting from the left with energy \( E \) on a step boundary, as illustrated in Fig. 2-1, given by

\[
V(x) = \begin{cases} 
0 & \text{if } x < 0, \\
V_0 & \text{if } x > 0.
\end{cases}
\]

Classically the particle should continue onwards past the step if \( E > V_0 \), and it should be reflected back to the left if \( E < V_0 \). What happens quantum mechanically?

Since the potential is time independent, we will focus on the time-independent Schrödinger’s equation (2.16). In the region \( x < 0 \) this is just

\[
-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi_I(x) = E \psi_I(x),
\]

where the Roman numeral I just denotes that we are in the region \( x < 0 \). This is just a free particle, and we already know solutions look something like \( e^{ikx} \). If we substitute this into (2.22), we find

\[
\frac{\hbar^2 k^2}{2m} = E.
\]

This equation has two solutions, which I will call \( \pm k \), where

\[
k = \frac{\sqrt{2mE}}{\hbar}. \tag{2.23}
\]

The most general solution for this region will be linear combinations of the two solutions, which will look like

\[
\psi_I(x) = Ae^{ikx} + Be^{-ikx}. \tag{2.24}
\]

Let’s now solve the same equation in the region \( x > 0 \). Schrödinger’s equation (2.16) in this region is

\[
-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi_{II}(x) = (E - V_0) \psi_{II}(x) \tag{2.25}
\]

Let us assume for the moment that \( E > V_0 \). If we define

\[
k' = \sqrt{\frac{2m(E - V_0)}{\hbar}}, \tag{2.26}
\]

then there are two solutions of the form \( e^{\pm ik'x} \), so the general solution takes the form
\[ \psi_{II}(x) = Ce^{ikx} + De^{-ikx}. \] (2.27)

We have solved Schrödinger’s equation in regions I and II. What about the boundary? The potential is everywhere finite, and therefore every term in Schrödinger’s equation will be finite at \( x = 0 \). But Schrödinger’s equation has a second derivative with respect to \( x \). The second derivative can be finite only if the function and its first derivative are continuous, since derivatives are only defined for continuous functions. In other words, we must satisfy the boundary conditions

\[ \psi_I(0) = \psi_{II}(0) \quad \text{and} \quad \psi_I'(0) = \psi_{II}'(0). \] (2.28)

Substituting in our expressions (2.24) and (2.27), this implies

\[ A + B = C + D \quad \text{and} \quad k(A-B) = k'(C-D). \] (2.29)

It is worth stopping a moment to think about the significance of the four terms we have come up with, equations (2.24) and (2.27). To understand their meaning, look at the probability current for each of the four waves:

\[ j_A = |A|^2 \frac{\hbar k}{m}, \quad j_B = -|B|^2 \frac{\hbar k}{m}, \quad j_C = |C|^2 \frac{\hbar k'}{m}, \quad \text{and} \quad j_D = |D|^2 \frac{\hbar k'}{m}. \]

Wave \( A \) has a positive current, which means it is moving to the right, and therefore it represents an incoming wave on the left. Wave \( B \) is a reflected wave, moving away to the left. Wave \( C \) represents the transmitted wave, moving off to the right. And what about \( D \)? It would represent a wave coming in from the right. Such a wave might exist, but we were asking specifically about a wave coming in from the left, so it is irrelevant to this problem, and we assume \( D = 0 \).

We are now prepared to solve our simultaneous equations (2.29). If we solve them for \( B \) and \( C \) in terms of \( A \), we find

\[ C = \frac{2k}{k+k'} A \quad \text{and} \quad B = \frac{k-k'}{k+k'} A. \]

These two equations tell us the magnitude of the two waves, from which we can easily derive the relative size of the probability density. However, we are trying to figure out if the wave is reflected or transmitted, which is asking something about the flow of probability. In short, we want to know what fraction of the probability flowing in is reflected back and what fraction is transmitted forward. These two quantities are denoted \( R \) (for reflected) and \( T \) (for transmitted), and are given by

\[ R \equiv \frac{|j_b|}{j_A} = \frac{|B|^2}{|A|^2} = \left( \frac{k-k'}{k+k'} \right)^2, \quad T \equiv \frac{j_C}{j_A} = \frac{|C|^2 k'}{|A|^2 k} = \frac{4kk'}{(k+k')^2}. \] (2.30)

It is easy to demonstrate that \( R + T = 1 \), which simply means that the wave is certainly either reflected or transmitted, as it must be. These equations can be rewritten in terms of \( E \) and \( V_0 \) with the help of (2.23) and (2.26) if desired.

One concern is that the wave functions we have found are not normalizable. The wave functions (2.24) and (2.27) do not diminish as they go to infinity. Of course, the same could be said of our plane wave solutions we found in section A. However, in

© 2011, Eric D. Carlson 23
section A, we were able to combine plane waves with different energies to form wave packets which we could normalize. The same is true here. Basically, this is possible because the wave solutions we found, (2.24) and (2.27), are not blowing up at infinity, and this made it relatively easy to combine them and make wave packets which do not grow without bounds at infinity.

We have solved the wave function only under the assumption that in (2.25), we have an energy $E$ greater than the potential $V_0$. We need to reconsider the case when $E < V_0$, when classically the wave has insufficient energy to penetrate the barrier. To do so, we reexamine (2.25), which I trivially rewrite in the form

$$ \frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi''_II(x) = (V_0 - E) \psi''_II(x). \quad (2.31) $$

We now define

$$ \alpha = \sqrt{\frac{2m(V_0 - E)}{\hbar}}. $$

Then the solutions of (2.31) will be exponentials of the form

$$ \psi''_II(x) = Ce^{-\alpha x} + De^{\alpha x}. \quad (2.32) $$

It is time to consider the two solutions in more detail before moving onwards. The term $C$ represents a wave that dies out in the region $x > 0$. In contrast, the term $D$ results in a wave function that grows quickly; faster than any polynomial. As a consequence, it will be difficult to build wave packets out of the $D$ term, because the wave grows so fast. Our conclusion is that the $D$ term is non-physical, and we once again assume $D = 0$.

We now substitute (2.24) and (2.32) (with $D = 0$) into the boundary conditions (2.28), which yields the simultaneous equations

$$ A + B = C \quad \text{and} \quad ik(A - B) = -\alpha C. $$

These equations can be solved for $B$ and $C$ in terms of the incoming wave $A$, which yields

$$ B = \frac{k - i\alpha}{k + i\alpha} A \quad \text{and} \quad C = \frac{2k}{k + i\alpha} A. \quad (2.33) $$

Now it’s time to figure out the probability of reflection and transmission. If you substitute a wave function of the form $\psi(x) = Ce^{-\alpha x}$ into the formula for probability current, it is easy to see you just get $j_c = 0$, which makes the transmission probability zero. Hence we quickly find the reflection and transmission coefficients

$$ R = \frac{|j_R|}{j_A} = \left| \frac{B}{A} \right|^2 = \frac{k^2 + \alpha^2}{k^2 + \alpha^2} = 1, \quad T = \frac{j_c}{j_A} = 0. \quad (2.34) $$

Thus when the energy is insufficient to penetrate the barrier classically, it ends up being totally reflected, even though the wave function penetrates a short distance into the forbidden region. When $E$ is greater than $V_0$, when classical theory predicts penetration of the barrier, transmission is possible, even quantum mechanically, but there is also a
possibility of reflection, as given in (2.30). There is even reflection when \( V_0 \) is negative; that is, when the potential barrier is negative. It would be as if you tried to run off a cliff, but instead reflected back! Fig. 2-2 represents a sketch of the reflection probability as a function of \( E/V_0 \).

Even though the particle cannot be transmitted in the case \( E < V_0 \), we do note that there is some penetration of the wave function into the forbidden region, though the wave function rapidly diminishes to zero. This does bring up the possibility that if we had a barrier of finite thickness, would it be possible for it to penetrate the barrier, and escape the other side? This phenomenon, called quantum tunneling, will be explored in the next section.

**E. Quantum Tunneling**

Consider now a barrier of height \( V_0 \) and thickness \( d \), ranging from \( x = -\frac{1}{2}d \) to \( x = \frac{1}{2}d \), as illustrated in Fig. 2-3, with potential

\[
V(x) = \begin{cases} 
V_0 & \text{if } -\frac{1}{2}d < x < \frac{1}{2}d, \\
0 & \text{otherwise.}
\end{cases}
\]

We wish to consider the case of a particle of mass \( m \) and energy \( E < V_0 \) impacting from the left. In a manner similar to before, we define \( k = \sqrt{2mE}/\hbar \) and \( \alpha = \sqrt{2m/(V_0 - E)}/\hbar \). We expect oscillating wave functions \( e^{\pm ikx} \) in the allowed regions I and III, and exponential functions \( e^{\pm \alpha x} \) in the classically forbidden region II. We do not expect there to be an incoming wave on the right, so in region III we exclude the \( e^{-ikx} \) solution from this region. Hence the most general solution will take the form

\[
\psi_I(x) = Ae^{ikx} + Be^{-ikx}, \quad \psi_{II}(x) = Ce^{\alpha x} + De^{-\alpha x}, \quad \psi_{III}(x) = Fe^{ikx}.
\]

As before, we match the functions and their first derivatives at each of the boundaries \( x = \pm \frac{1}{2}d \), leading to four boundary conditions:

\[
Ae^{-ikd/2} + Be^{ikd/2} = Ce^{-\alpha d/2} + De^{\alpha d/2}, \quad ik\left(Ae^{-ikd/2} - Be^{ikd/2}\right) = \alpha\left(Ce^{-\alpha d/2} - De^{\alpha d/2}\right), \quad Ce^{\alpha d/2} + De^{-\alpha d/2} = Fe^{ikd/2}, \quad \alpha\left(Ce^{\alpha d/2} - De^{-\alpha d/2}\right) = ikFe^{ikd/2}.
\]

Figure 2-2: The probability of transmission (blue curve) and reflection (red curve) as a function of the size of the barrier. Note that there is always some reflection except for the trivial case \( V_0 = 0 \). However, if the potential is larger than the energy the wave is completely reflected.

Figure 2-3: The finite barrier of section E.
These are four equations in five unknowns, and we can solve for any of the variables in terms of any of the others. We will be particularly interested in the probability of transmission, which requires that we find $F$. With considerable work one can show that

$$F = \frac{2k\alpha e^{-ikd} A}{2k\alpha \cosh(\alpha d) + i(k^2 - \alpha^2) \sinh(\alpha d)}.$$  

The transmission probability will then be

$$T = \frac{j_F}{j_j} = \left| \frac{F}{A} \right|^2 = \frac{4k^2\alpha^2}{4k^2\alpha^2 + (k^2 + \alpha^2)^2 \sinh^2(\alpha d)} = \frac{4E(V_0 - E)}{4E(V_0 - E) + V_0^2 \sinh^2(\alpha d)}, \quad (2.38)$$

where $\alpha = \sqrt{2m(V_0 - E)}/\hbar$. For a thick barrier, $\alpha d \gg 1$, the hyperbolic sign will exponentially suppress this rate, so the process will be rare, but it never vanishes.

Note that in each of the cases we have studied so far, a free particle (section A), a particle approaching a step barrier (section D), and a finite thickness barrier, we found solutions for all possible energies. The reason this occurred is because in all these cases the energy of the particle was sufficient to get the particle out to infinity. In section F we will explore an alternative case, when the energy is assumed to be less than the energy at infinity. In such cases (called bound states), the wave function will vanish away at infinity, and the energy will turn out to only come in discrete possible values.

F. The Infinite Square Well

Let’s turn our attention now to a different kind of problem, one where the particle’s energy is less than the energy at infinity. Consider a particle in one dimension trapped in an infinite square well, with potential given by

$$V(x) = \begin{cases} 0 & \text{if } 0 < x < a, \\ \infty & \text{otherwise}. \end{cases}$$

We start by trying to solve (2.16). The infinite potential has the effect that the energy will be infinite for any wave function lying outside the allowed region, so certainly we want a wave function that vanishes outside this region. Furthermore, the derivative terms will cause problems if the wave function is not continuous. We therefore demand that the solutions of (2.16) vanish at the two boundaries, $\psi(0) = \psi(a) = 0$.

It remains to solve (2.16) in the allowed region, where the wave function will satisfy

$$E\psi(x) = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x). \quad (2.36)$$
Because (2.36) is a second-order linear differential equation, there will be two linearly
independent solutions. The functions whose second derivative are proportional to their
negatives are $\sin(kx)$ and $\cos(kx)$, so we write\footnote{There is an implicit assumption here that \( E > 0 \). If \( E < 0 \), then the solutions to (2.36) will take the form $e^{\alpha x}$ or $e^{-\alpha x}$, however, no linear combination of these two functions can satisfy both boundary conditions.}

$$\psi(x) = A\sin(kx) + B\cos(kx).$$  (2.37)

Substituting (2.37) into (2.36), it is easy to see that $E = \hbar^2k^2/2m$.

The boundary condition $\psi_n(0) = 0$ will be satisfied only if we pick $B = 0$. Then the boundary condition $\psi_n(a) = 0$ will demand that $\sin(ak) = 0$, which implies that $ak$ will be an integer multiple of $\pi$, so $k = \pi n/a$. Hence we can label our solutions by a single positive integer $n$, in terms of which $\psi_n(x) = A\sin(\pi n x / a)$. We now demand that the wave function be normalized; that is

$$1 = \int_{-a}^{a} |\psi(x)|^2 \, dx = |A|^2 \int_{0}^{a} \sin^2\left(\frac{\pi n x}{a}\right) \, dx = \frac{1}{2}a |A|^2.$$

Solving for $A$, and choosing it to be real and positive, we find $A = \sqrt{2/a}$. Putting everything together, we now have formulas for the eigenstates and energies.

$$\psi_n(x) = \begin{cases} \sqrt{2/a} \sin(\pi n x / a) & \text{for } 0 < x < a, \\ 0 & \text{otherwise,} \end{cases}$$  (2.38a)

$$E_n = \frac{\pi^2 \hbar^2 n^2}{2ma^2}.$$  (2.38b)

The first few wave functions are plotted in Fig. 2-4.

We can generalize this solution to any number of dimensions. For example, suppose a particle of mass $m$ is contained in a three-dimensional potential given by

$$V(x, y, z) = \begin{cases} 0 & \text{if } 0 < x < a, \; 0 < y < b \; \text{and} \; 0 < z < c, \\ \infty & \text{otherwise.} \end{cases}$$

Our wave functions must vanish at the six boundaries. Inside these boundaries, it is easy to see that Schrödinger’s equation is simply

$$-\frac{\hbar^2}{2m} \nabla^2 \psi = E\psi.$$
It isn’t hard to guess the solutions to this equation. Products of sine functions in each of the three dimensions will work nicely. We will have three indices, \( n_x, n_y, n_z \) labeling our states, and the energies will be functions of all of these. The wave function (in the allowed region) and the energies will be given by

\[
\psi_{n_x, n_y, n_z}(x, y, z) = \sqrt{\frac{8}{abc}} \sin\left(\frac{\pi x}{a}\right) \sin\left(\frac{\pi y}{b}\right) \sin\left(\frac{\pi z}{c}\right),
\]

\[
E_{n_x, n_y, n_z} = \frac{\pi^2 \hbar^2}{2m} \left(\frac{n_x^2}{a^2} + \frac{n_y^2}{b^2} + \frac{n_z^2}{c^2}\right).
\]

G. Bound States from a Double Delta potential

Consider a particle of mass \( m \) in one dimension with potential

\[
V(x) = -\lambda \left[\delta(x - \frac{1}{2}a) + \delta(x + \frac{1}{2}a)\right]
\]

so that Schrödinger’s equation takes the form

\[
E\psi(x) = \left[ -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} - \lambda \delta(x - \frac{1}{2}a) - \lambda \delta(x + \frac{1}{2}a) \right] \psi(x) \quad (2.39)
\]

where \( a \) and \( \lambda \) are positive constants. In other words, the particle is free except at two points where the potential is negatively infinite, as sketched in Fig. 2-5. We will be seeking bound states with energy \( E < 0 \). According to classical physics, the particle will be stuck in one of the two infinitely narrow wells. What does quantum physics predict?

Away from the two delta functions, the particle is free, satisfying

\[
E\psi(x) = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x). \quad (2.40)
\]

Keeping in mind that \( E \) is negative, we define

\[
\beta = \sqrt{-2mE} \frac{\hbar}{\hbar}.
\]

We then find that the solutions to (2.40) take the form \( e^{\pm \beta x} \). However, there is no reason to believe that the form of the equation will \textit{match} in the three regions marked I, II, and III in Fig. 2-5, so the coefficients of the three waves will tend to be different. Furthermore, we don’t want solutions which diverge at infinity, so in region III we reject waves that look like \( e^{\beta x} \), and in region I we reject \( e^{-\beta x} \). We therefore guess the solutions will take the form
\[ \psi_1(x) = Ae^{\beta x}, \quad \psi_{II}(x) = Be^{-\beta x} + Ce^{\beta x}, \quad \psi_{III}(x) = De^{-\beta x}. \quad (2.41) \]

What about boundary conditions? It is tempting to assume that the function and its derivative will be continuous, but this is not correct. To figure out what to do at the boundary, integrate (2.39) across one of the boundaries, say from \( \frac{1}{2}a - \epsilon \) to \( \frac{1}{2}a + \epsilon \). We find

\[
E \int_{\frac{1}{2}a - \epsilon}^{\frac{1}{2}a + \epsilon} \psi(x) \, dx = -\frac{\hbar^2}{2m} \int_{\frac{1}{2}a - \epsilon}^{\frac{1}{2}a + \epsilon} \frac{d^2 \psi(x)}{dx^2} \, dx - \lambda \int_{\frac{1}{2}a - \epsilon}^{\frac{1}{2}a + \epsilon} \psi(x) \left[ \delta(x - \frac{1}{2}a) + \delta(x + \frac{1}{2}a) \right] \, dx.
\]

The first term on the right side is easy to evaluate, because the integral of the second derivative of a function is just the first derivative. The integral of a delta function is similarly easy to evaluate; it simply yields \( \psi(x) \) at the boundary point \( x = \frac{1}{2}a \). The other boundary point \( x = -\frac{1}{2}a \) does not contribute because it is not in the range of integration. We have

\[
E \int_{\frac{1}{2}a - \epsilon}^{\frac{1}{2}a + \epsilon} \psi(x) \, dx = -\frac{\hbar^2}{2m} \left[ \psi''_{III} \left( \frac{1}{2}a + \epsilon \right) - \psi''_{II} \left( \frac{1}{2}a - \epsilon \right) \right] - \lambda \psi \left( \frac{1}{2}a \right).
\]

Note that the derivative terms are evaluated using the appropriate functions in each range.

Now, consider the limit as \( \epsilon \to 0 \). The integrand on the left is finite, and the integral will become very small. So this simplifies to

\[
\psi''_{III} \left( \frac{1}{2}a + \epsilon \right) - \psi''_{II} \left( \frac{1}{2}a - \epsilon \right) = -\frac{2m\lambda}{\hbar^2} \psi \left( \frac{1}{2}a \right).
\]

Spend a moment analyzing this equation. It says specifically that the derivative is not continuous at the boundary, it has a finite discontinuity there. However, a finite discontinuity implies that the function must at least be continuous, so we have

\[
\psi'' \left( \frac{1}{2}a \right) = \psi_{III} \left( \frac{1}{2}a \right). \quad (2.42)
\]

This is fortunate, since otherwise in equation (2.42), we wouldn’t know which function to use on the right side of the equation, but we see from (2.43) that it doesn’t matter.

We can similarly integrate (2.39) across the boundary at \( x = -\frac{1}{2}a \). This leads to two more boundary conditions.

\[
\psi''_I \left( -\frac{1}{2}a \right) - \psi''_{II} \left( -\frac{1}{2}a \right) = -\frac{2m\lambda}{\hbar^2} \psi \left( -\frac{1}{2}a \right),
\]

\[
\psi_I \left( -\frac{1}{2}a \right) = \psi_{II} \left( -\frac{1}{2}a \right). \quad (2.44)
\]

Let’s write out equations (2.42), (2.43) and (2.44) in terms of our explicit wave functions (2.41). We find

\[
Ae^{-\beta a/2} = Be^{\beta a/2} + Ce^{-\beta a/2}, \quad (2.45a)
\]

\[
De^{-\beta a/2} = Be^{-\beta a/2} + Ce^{\beta a/2}, \quad (2.45b)
\]
We have here four equations in four unknowns, which one would think would have a unique solution. Indeed, for a general value of the energy \( E \) (and corresponding parameter \( \beta \)), the unique solution to \((2.45)\) is \( A = B = C = D = 0 \). This however is an unacceptable solution. We want non-trivial solutions. To find them, we substitute \((2.45a)\) and \((2.45b)\) into \((2.45c)\) and \((2.45d)\) to yield two, which yields

\[
\frac{m\lambda}{h^2} \left[ Be^{\beta a/2} + Ce^{-\beta a/2} \right] = \beta Be^{\beta a/2} \quad \text{and} \quad \frac{m\lambda}{h^2} \left[ Be^{-\beta a/2} + Ce^{\beta a/2} \right] = \beta Ce^{\beta a/2}.
\]

Note that these equations are identical save for the interchange of the role of \( B \) and \( C \). If either \( B \) or \( C \) vanish, then the other will as well, and the resulting solution is the trivial one. If they don’t vanish, we can divide by either of them, and we can rewrite the two equations as

\[
\frac{B}{C} = \frac{C}{B} = e^{\beta a} \left[ \frac{h^2 \beta}{m\lambda} - 1 \right].
\]

Now, the only numbers that are their own reciprocal are \( \pm 1 \), so substituting this in and rearranging a bit, we must have

\[
\frac{h^2 \beta}{m\lambda} = 1 \pm e^{-\beta a}, \quad (2.46)
\]

The left side of this equation is a simple line. The right side is two functions, one of which starts at 0, the other at 2, and then both asymptotically approach 1. It is easy to see, as illustrated in Fig. 2-6, that the linear function must cross the upper curve eventually, so there is always one solution \( \beta_+ \) to \((2.46)\). The other curve crosses the line at \( \beta = 0 \), but this turns out to lead to only a trivial solution to our equations. However, if \( \lambda \) is not too small, it is easy to show that there will be an additional crossing, which we denote \( \beta_- \). In other words, we have one or two solutions to \((2.46)\), which satisfy the two equations.

Figure 2-6: Graphical solution of eq. \((2.46)\). We are trying to find where one of the straight lines crosses the red (for \( \beta_+ \)) or green (for \( \beta_- \)) curves. The straight lines correspond to \( h^2/ma\lambda = 0.25, 0.50, 0.75, 1.00, 1.25, \) and 1.50 (bottom to top). The solutions for the \( h^2/ma\lambda = 0.50 \) (navy line) are \( a\beta_+ = 2.2177 \) and \( a\beta_- = 1.5936 \) (dotted red and green lines respectively). Note that \( \beta_- \) will exist only if \( h^2/ma\lambda < 1 \).
\[ \frac{\hbar^2 \beta}{m \lambda} = 1 + e^{-\beta, a}, \quad (2.47a) \]
\[ \frac{\hbar^2 \beta}{m \lambda} = 1 - e^{-\beta, a}. \quad (2.47b) \]

Eq. (2.47b) has solutions only if \( \lambda ma > \hbar^2 \). The resulting energies are

\[ E_x = -\frac{\hbar^2 \beta^2}{2m}. \]

These two energies are plotted in Fig. 2-7. Once we have the relevant value of \( \beta \), we know \( B = \pm C \), and we can get the other coefficients from the first two equations of (2.45). If we wish, we can then normalize the overall wave function. The final solution for \( ma\lambda/\hbar^2 = 2 \) are illustrated in Fig. 2-8. Note that one of the solutions is symmetric, and the other anti-symmetric across the origin. As we will later demonstrate, this is a general feature of problems where the potential is symmetric, \( V(x) = V(-x) \).

In summary, in this case we have found that for the double delta function potential, there are bound states only for specific energies. The exact number of solutions depends on the depth of the attractive delta-function potentials; not surprisingly, the number increases as we increase \( \lambda \). This is the general result for an attractive potential: the solutions will exist for only discrete energies. Often there will only be a finite number of such solutions, though we will find when the range of the attraction is long enough we will sometimes find an infinite number of energies.
Problems for Chapter 2

1. A free particle of mass \( m \) in one dimension at \( t = 0 \) has wave function

\[
\Psi(x, t = 0) = \psi(x) = \left(\frac{A}{\pi}\right)^{1/4} \exp\left(iKx - \frac{1}{2} Ax^2\right).
\]

This is identical with chapter 1 problem 4. Find the wave at all subsequent times.

2. One solution of the 2D Harmonic oscillator Schrödinger equation takes the form

\[
\Psi(x, y, t) = (x + iy) e^{-\frac{a(x^2 + y^2)}{2}} e^{-i\omega t}.
\]

(a) Find the probability density \( \rho(x, y, t) \) at all times.
(b) Find the probability current \( j(x, y, t) \) at all times.
(c) Check the local version of conservation of probability, i.e., show that your solution satisfies \( \partial \rho / \partial t + \nabla \cdot j = 0 \).

3. A particle of mass \( m \) lies in the one-dimensional infinite square well, which has potential with allowed region \( 0 < x < a \). At \( t = 0 \), the wave function takes the form

\[
\left(\frac{4}{\sqrt{5a}}\right) \sin^3(\pi x / a).
\]

Rewrite this in the form \( \Psi(x, t = 0) = \sum_i c_i \psi_i(x) \). Find the wave function \( \Psi(x, t) \) at all later times. The identity \( \sin(3\theta) = 3\sin\theta - 4\sin^3\theta \) may be helpful.

4. A particle of mass \( m \) and energy \( E \) scatters from the negative-\( x \) direction off of a delta-function potential: \( V(x) = \lambda \delta(x) \).

(a) For the regions \( x < 0 \) and \( x > 0 \), find general equations for the wave, eliminating any terms that are physically inappropriate.
(b) Integrate Schrödinger’s time-independent equation across the boundary to obtain an equation relating the derivative of the wave function on either side of the boundary. Will the wave function itself be continuous?
(c) Solve the equations and deduce the transmission and reflection coefficients \( T \) and \( R \). Check that \( T + R = 1 \).
(d) A delta function potential \( V(x) = \lambda \delta(x) \) can be thought of as a very narrow potential of height \( V_0 \) and width \( d \), with \( V_0 d = \lambda \). Show that you can get the same result by using (2.38), where you can approximate \( \sinh(\alpha d) \approx \alpha d \) since \( d \) is small.

5. A particle of mass \( m \) is trapped in a potential \( V(x) = -\lambda \delta(x) \). Show that there is only one bound state and find its energy.
III. Hilbert Space and Vector Spaces

Before proceeding further, it is advisable to develop a bit more mathematical structure. Such structure is provided in the context of vector spaces, and more specifically Hilbert space. You are already familiar with ordinary real finite-dimensional vector spaces, like ordinary Euclidean space, in which elements are denoted by triplets of real numbers \( r = (x, y, z) \). Hilbert space, in contrast, is an infinite dimensional complex vector space.

A. Hilbert Space

Consider the set of all complex wave functions \( \psi(r) \) in three-dimensional space, for the moment with no restrictions on it. For any pair of such wave functions, it is useful to define the inner product \( (\phi, \psi) \) by the equation

\[
(\phi, \psi) = \iint \phi^*(r) \psi(r) \, d^3r.
\] (3.1)

The inner product has several special properties which we will take advantage of. The first is that it is linear in its second argument and anti-linear in its first. In other words, if \( c_1 \) and \( c_2 \) are arbitrary complex numbers, it is easy to see from (3.1) that

\[
(\phi, c_1 \psi_1 + c_2 \psi_2) = c_1 (\phi, \psi_1) + c_2 (\phi, \psi_2),
\]

\[
(c_1 \phi_1 + c_2 \phi_2, \psi) = c_1^* (\phi_1, \psi) + c_2^* (\phi_2, \psi).
\] (3.2)

The other property we need is that if we take the inner product of a wave function with itself, it is almost always positive. Specifically,

\[
\|\psi\|^2 \equiv (\psi, \psi) \geq 0, \quad \text{with equality only if} \ \psi = 0.
\] (3.3)

From properties (3.2) and (3.3) it is possible to prove a variety of others. The first tells us how the inner product of any two wave functions changes if you swap them; namely,

\[
(\phi, \psi) = (\psi, \phi)^*.
\]

This identity is trivial to prove using (3.1), but in fact, it can be proven using only (3.2) and (3.3). The other relation is the Schwartz inequality, which takes the form

\[
| (\phi, \psi) |^2 \leq (\phi, \phi) (\psi, \psi).
\] (3.4)

Again, this inequality can be proven using only (3.2) and (3.3); the proof is left as a homework problem.

Now, when considering wave functions for quantum mechanics, we know that we will need to pay special attention to those that are properly normalized, which in our
current notation can be written very simply as \((\psi, \psi) = 1\). However, this will prove too restrictive, so instead we demand only that the wave function be square integrable; that is,

\[
(\psi, \psi) < \infty.
\]

The set of complex wave function satisfying (3.5) will be called *Hilbert space* and denoted \(\mathcal{H}\).

Let \(\psi_1\) and \(\psi_2\) be any elements of Hilbert space. Now, take an arbitrary linear combination \(c_1\psi_1 + c_2\psi_2\), where \(c_1\) and \(c_2\) are complex numbers. A little work will convince you that

\[
(c_1\psi_1 + c_2\psi_2, c_1\psi_1 + c_2\psi_2) = 2|c_1|^2 (\psi_1, \psi_1) + 2|c_2|^2 (\psi_2, \psi_2) - (c_1\psi_1 - c_2\psi_2, c_1\psi_1 - c_2\psi_2).
\]

The first two terms on the right side of (3.6) are finite from (3.5), and the last term is negative, so this cannot be positive infinity (and it is positive, by (3.3)). It follows that \(c_1\psi_1 + c_2\psi_2\) is an element of Hilbert space as well; that is,

\[
\psi_1, \psi_2 \in \mathcal{H} \quad \& \quad c_1, c_2 \in C \implies c_1\psi_1 + c_2\psi_2 \in \mathcal{H},
\]

where \(C\) denotes the set of complex numbers. Any set that satisfies (3.7) is called a complex vector space, so Hilbert space is a complex vector space.\(^1\)

You are probably more familiar with ordinary three-dimensional vectors spaces, which is a real vector space. In ordinary three-dimensional space, if you have two vectors \(\mathbf{v}_1\) and \(\mathbf{v}_2\), we can multiply them by arbitrary real numbers \(r_1\) and \(r_2\) to make a new vector \(r_1\mathbf{v}_1 + r_2\mathbf{v}_2\). The inner product, defined above, plays a role similar to the dot product in ordinary three dimensional space. Ordinary space is said to be three dimensional, because to specify a vector \(\mathbf{v}\), you have to specify three numbers, \(\mathbf{v} = (x, y, z)\). Hilbert space is infinite dimensional, because to specify a wave function \(\psi(\mathbf{r})\) you need to specify the wave function at an infinite number of points.

Although Hilbert space is defined in terms of the relationship (3.5), we will sometimes want to consider functions that lack this restriction, such as plane waves. This can be done. In such cases we are working in a larger vector space, or *superspace* of Hilbert space. At other times, we may wish to add additional restrictions; for example, we might demand that the wave function be finite everywhere, or fall off quickly at large \(\mathbf{r}\), or have one or several derivatives that exist. In such cases we are working in a smaller vector space, or *subspace* of Hilbert space. Most of the time we won’t pay attention to such fine distinctions, and will simply pretend we are always working in Hilbert space.

\(^1\) It is a complex vector space because we combine elements by multiplying them by complex numbers.
B. Dirac Notation and Covectors

Up until now, we have been talking about our wave functions in a specific way, as functions of the position, $\psi (r)$. As we discussed in chapter 2, we can equally well describe a wave function in terms of its Fourier transform $\tilde{\psi} (k)$; indeed, there is nothing superior about the former description over the latter. We would like to regard these as different ways to describe the same wave, without referring explicitly to a coordinate or Fourier representation (just as for ordinary vectors, we refer to them as $r$ without thinking of them in a specific coordinate system). For this we will use Dirac’s bra-ket notation, in which he designated a vector like this: $|\psi\rangle$, which he called a ket. These kets can be multiplied by complex numbers and added together, so we write expressions such as $c_1 |\psi_1\rangle + c_2 |\psi_2\rangle$.

We now need to define the covector space of Hilbert space $\mathcal{H}^*$, defined as the set of all linear mappings from Hilbert space to the complex numbers. In other words, if $f \in \mathcal{H}^*$, then $f (|\psi\rangle)$ is a complex number, and when we say it is linear, we mean

$$f (c_1 |\psi_1\rangle + c_2 |\psi_2\rangle) = c_1 f (|\psi_1\rangle) + c_2 f (|\psi_2\rangle).$$

Now, Dirac denoted elements of the covector space like this: $\langle f |$, which he called a bra. Furthermore, when you put them together, he used a shorthand notation

$$\langle f | |\psi\rangle = \langle f | \psi \rangle,$$

which we now call a bra-ket (think “bracket”). Covectors are defined by how they act on an arbitrary wave function, so if we define $\langle f | \psi \rangle$ for all kets $|\psi\rangle$ we have defined $\langle f |$.

For any two covectors, we can define new covectors by taking an arbitrary linear combination of them; for example,

$$\left( c_1 \langle f_1 | + c_2 \langle f_2 | \right) |\psi\rangle = c_1 \langle f_1 | \psi \rangle + c_2 \langle f_2 | \psi \rangle.$$

This means that covectors also form a vector space, so $\mathcal{H}^*$ is a complex vector space too.

A few examples might help us understand these covectors better:

**Example 1:** For any position $r$ in space, define $\langle r |$ by

$$\langle r | \psi \rangle = \psi (r).$$

That this is a covector can be verified easily:

$$\langle r | (c_1 |\psi_1\rangle + c_2 |\psi_2\rangle) = c_1 \psi_1 (r) + c_2 \psi_2 (r) = c_1 \langle r | \psi_1 \rangle + c_2 \langle r | \psi_2 \rangle.$$

**Example 2:** For any vector $k$, define $\langle k |$ by

$$\langle k | \psi \rangle = \tilde{\psi} (k) = \int \frac{d^3 r}{(2\pi)^{3/2}} e^{-i k \cdot r} \psi (r).$$
Example 3: For any $|\phi\rangle \in \mathcal{H}$, define $\langle \phi |$ by

$$
\langle \phi | \psi \rangle \equiv (\phi, \psi) = \iiint \phi^*(r) \psi(r) d^3r.
$$

(3.13)

Now, there is a technical detail we should probably deal with, namely: that some of these may not converge (and since infinity is not really a complex number, that means they aren’t members of $\mathcal{H}^*$). For example, a function may be infinite at a particular point, but still be square integrable. Hence (3.11) may not be finite for at least some points $r$.

Similarly, it is possible for a wave function to be square integrable, and yet the integral (3.12) will not converge, at least for some values of $k$. Thanks to the Schwartz inequality (3.4), (3.13) always does converge when $|\phi\rangle, |\psi\rangle \in \mathcal{H}$. Indeed, it can be shown that the only bras that yield finite numbers are of the form (3.13). Basically, this demonstrates that there is a one-to-one correspondence between bras and kets within the Hilbert space $\mathcal{H}$, so we can write $\mathcal{H} = \mathcal{H}^*$.

We can perhaps see the problem most clearly if we compare equations (3.12) and (3.13). These two equations look very similar; indeed, if we define

$$
\phi_k(r) = \frac{1}{(2\pi)^{3/2}} e^{i kr},
$$

(3.14)

then we see that

$$
\langle \phi_k | \psi \rangle = \langle k | \psi \rangle,
$$

(3.15)

and hence we could argue that $\langle \phi_k | = \langle k |$. The problem is that (3.14) is not square integrable, and hence isn’t in $\mathcal{H}$. Similarly, we could define

$$
\phi_r(r') = \delta^3(r-r'),
$$

(3.16)

and then we can see that

$$
\langle \phi_r | \psi \rangle = \iiint \delta^3(r-r') \psi(r') d^3r' = \psi(r),
$$

and therefore $\langle \phi_r | = \langle r |$. But once again, (3.16) is not square integrable.

Even though $\langle r |$ and $\langle k |$ are not well-defined on every element of Hilbert space, it is possible to restrict Hilbert space in various ways such that they are well-defined. For example, we could restrict Hilbert space by demanding that the wave functions be finite, or fall off quickly at infinity. If we make such restrictions, we will end up with a smaller vector space $\Omega$ which is a subset of Hilbert space. Such a vector space is called a subspace of Hilbert space, and we write $\Omega \subset \mathcal{H}$. The covector space $\Omega^*$ will then have extra covectors, such as $\langle r |$ and $\langle k |$, so it will be a bigger space, or superspace of $\mathcal{H}^*$.

The relations between these various vector spaces then takes the form

$$
\Omega \subset \mathcal{H} = \mathcal{H}^* \subset \Omega^*.
$$

\footnote{Note that in equation (3.13), $r$ is a dummy variable, which can be renamed at will. In equation (3.16), we renamed it to $r'$ to avoid confusion with $r$.}

© 2011, Eric D. Carlson
For a longer discussion of the relationship between various subspaces of $\mathcal{H}$ and their covector spaces, consult Balantine, but these are mostly issues for mathematicians, and we are interested in physics, so I will generally ignore such details and simply pretend we are working in Hilbert space, even when we use wave functions like (3.14) that are not.

The process of turning a ket into a bra as shown in example 3 is our first example of Hermitian conjugation, which we denote by putting a dagger on the ket, so we have

$$|\phi\rangle^\dagger \equiv \langle \phi |.$$

(3.17)

Since $\mathcal{H} = \mathcal{H}^*$ there is a ket for every bra, and we use the same symbol to work in reverse

$$\langle \phi |^\dagger \equiv |\phi\rangle.$$

(3.18)

Using (3.17) and (3.18), it is trivial to prove that $\left(|\phi\rangle^\dagger\right)^\dagger = |\phi\rangle$ and $\left(\langle \phi |^\dagger\right)^\dagger = \langle \phi |$. Even if we are working with a bra or ket that is not in $\mathcal{H}$, we can use this property to define a corresponding bra or ket, so for example, since $\langle \phi_k | = \langle k |$, it makes sense to define $|k\rangle = |\phi_k\rangle$, so that $\langle r | k \rangle = \langle r | \phi_k \rangle = \phi_k (r)$, so we have

$$\langle r | k \rangle = \frac{e^{ik \cdot r}}{(2\pi)^{3/2}}.$$

(3.19)

Similarly,

$$\langle r' | r \rangle = \delta^3 (r' - r).$$

(3.20)

It is trivial to take the complex conjugate of (3.19) to find $\langle k | r \rangle$, and we can show that

$$\langle k' | k \rangle = \delta^3 (k' - k).$$

(3.21)

Hermitian conjugation acting on a linear combination of kets is a bit non-trivial. Suppose we want the Hermitian conjugate of $c_1 |\phi_1\rangle + c_2 |\phi_2\rangle$. This is a bra, so to understand what it means, we need to let it act on an arbitrary ket. We have

$$ (c_1 |\phi_1\rangle + c_2 |\phi_2\rangle)^\dagger |\psi\rangle = (c_1 \phi_1 + c_2 \phi_2, \psi) = c_1^* \langle \phi_1 | + c_2^* \langle \phi_2 | |\psi\rangle,$$

and since this is true for any wave function $|\psi\rangle$, we have

$$ (c_1 |\phi_1\rangle + c_2 |\phi_2\rangle)^\dagger = c_1^* |\phi_1\rangle + c_2^* |\phi_2\rangle.$$

Hence we see that Hermitian conjugation leaves sums as sums, but takes the complex conjugate of any complex coefficients. We can similarly show that

$$ (c_1 \langle \phi_1 | + c_2 \langle \phi_2 |)^\dagger = c_1^* |\phi_1\rangle + c_2^* |\phi_2\rangle.$$
C. Coordinate Bases

In three dimensions, we know it is often useful to break a vector \( \mathbf{r} \) into three coordinates \((x, y, z)\), where we write \( \mathbf{r} = x\mathbf{\hat{x}} + y\mathbf{\hat{y}} + z\mathbf{\hat{z}} \). In order to do this, we normally define three orthonormal vectors \( \{\mathbf{\hat{x}}, \mathbf{\hat{y}}, \mathbf{\hat{z}}\} \) in terms of which we can write an arbitrary vector. Can we do something similar in Hilbert space?

Any element of a vector space can be written in terms of linear combinations of a subset of that space, called a basis \( \{|\phi_1\>, |\phi_2\>, |\phi_3\>, \ldots\} \). We write, in general,

\[
|\psi\> = c_1 |\phi_1\> + c_2 |\phi_2\> + c_3 |\phi_3\> + \cdots,
\]

where the \( c_i \)'s are complex numbers. This basis is said to be independent if there is no non-trivial combination that adds to make zero. Indeed, if there is a linear combination that adds to zero, then we must have \( c_1 |\phi_1\> + c_2 |\phi_2\> + \cdots + c_n |\phi_n\> = 0 \). Assuming the last term is non-zero, we then see that

\[
|\phi_n\> = -\frac{c_1}{c_n} |\phi_1\> - \frac{c_2}{c_n} |\phi_2\> - \cdots - \frac{c_{n-1}}{c_n} |\phi_{n-1}\>,
\]

and we can rewrite any term in (3.23) without using \( |\phi_n\> \), so we might as well throw it out. If we start from any basis, we can eventually throw out these redundant vectors and get an independent basis.

A basis is called complete if any vector in it can be written in terms of this basis; that is, if any \( |\psi\> \) can be written in the form (3.23). A basis is called orthogonal if the inner product between any two distinct basis vectors is zero, so \( \langle \phi_i | \phi_j \rangle = 0 \) if \( i \neq j \). It is called orthonormal if in addition you get 1 when you take the inner product of a basis vector with itself:

\[
\langle \phi_i | \phi_i \rangle = \delta_{ii}.
\]

It is ideal, when possible, to work with orthonormal bases. If you are given a basis \( \{|\phi_1\>, |\phi_2\>, |\phi_3\>, \ldots\} \) that is not orthonormal, there is a straightforward procedure for converting it to an orthonormal basis \( \{|\phi'_1\>, \ldots\} \). As a first step, we successively define a new orthogonal basis \( \{|\phi'_1\>, \ldots\} \) by the relations

\[
|\phi'_1\> = |\phi_1\>, \quad |\phi'_2\> = |\phi_2\> - \frac{\langle \phi'_1 | \phi_2 \rangle}{\langle \phi'_1 | \phi'_1 \rangle} |\phi'_1\>, \quad \ldots, \quad |\phi'_n\> = |\phi_n\> - \sum_{i=1}^{n-1} \frac{\langle \phi'_i | \phi_n \rangle}{\langle \phi'_i | \phi'_i \rangle} |\phi'_i\>.
\]

We now show by induction that (3.25) leads to an orthogonal basis. Suppose that we already have made sure that \( \{|\phi'_1\>, \ldots, |\phi'_n\>\} \) are orthogonal, and we want to see if \( |\phi'_{n+1}\> \) is orthogonal to all the previous ones. We see that

1 The Kronecker delta function, \( \delta_{ij} \), is defined to equal 1 when \( i = j \) and 0 otherwise.
\[ \langle \phi' | \phi_{n+1}' \rangle = \langle \phi' | \phi_{n+1} \rangle - \sum_{i=1}^{n} \frac{\langle \phi' | \phi_{n+1} \rangle}{\langle \phi'_i | \phi'_i \rangle} \langle \phi'_i | \phi'_i \rangle = \langle \phi' | \phi_{n+1} \rangle - \frac{\langle \phi' | \phi_{n+1} \rangle}{\langle \phi'_i | \phi'_i \rangle} \langle \phi'_i | \phi'_i \rangle = 0. \]

Furthermore, \( |\phi'_{n+1} \rangle \) is not zero, since the \( |\phi'_{i} \rangle \)'s are independent. So we have succeeded in finding an orthogonal basis. To make it orthonormal, simply define

\[ |\phi''_{i} \rangle = \frac{1}{\sqrt{\langle \phi'_i | \phi'_i \rangle}} |\phi'_{i} \rangle. \quad (3.26) \]

It is then easy to show that \( \{|\phi''_{i} \rangle \} \) forms an orthonormal basis.

Given a complete, orthonormal basis \( \{|\phi_{i} \rangle \} \), actually writing an arbitrary wave function \( |\psi \rangle \) in terms of that basis turns out to be surprisingly easy. First write (3.23) in the form

\[ |\psi \rangle = \sum_{i} c_i |\phi_{i} \rangle. \quad (3.27) \]

Now, let the bra \( \langle \phi_{n} | \) act on both sides of the equation. We find

\[ \langle \phi_{n} | \psi \rangle = \langle \phi_{n} | \sum_{i} c_i |\phi_{i} \rangle = \sum_{i} c_i \langle \phi_{n} | \phi_{i} \rangle = \sum_{i} c_i \delta_{i,n} = c_n. \quad (3.28) \]

Hence we have a relatively simple expression for the coefficients \( c_n \), though in fact it may require rather complicated integrals to actually compute (3.28).

It is sometimes the case that orthonormal bases are not labeled by a discrete set of labels, like \( \{|\phi_{i} \rangle \} \), but instead by continuous variables \( \{|\phi_{w} \rangle \} \). In this case, it is a good idea to modify the form of (3.27) and the orthonormal condition (3.24) to

\[ |\psi \rangle = \int c(w) |\phi_{w} \rangle dw \quad (3.29) \]

and

\[ \langle \phi_{v} | \phi_{w} \rangle = \delta (v - w). \quad (3.30) \]

Then the continuum equivalent of (3.28) will be the relationship

\[ c(w) = \langle \phi_{w} | \psi \rangle. \quad (3.31) \]

We have actually already encountered two examples of continuum bases, the position basis \( |r \rangle \) and the Fourier basis \( |k \rangle \). For example, (3.30) appears as equations (3.20) and (3.21). The functions \( c(w) \) take the form \( \psi (r) \) for the basis \( |r \rangle \) and \( \tilde{\psi} (k) \) for the basis \( |k \rangle \).

Given a complete, orthonormal basis, we can then describe any wave function completely by giving its components \( c_i \) or \( c(w) \). Our most common description of a wave function will be to give \( \psi (r) = \langle r | \psi \rangle \), but it will not be uncommon to give \( \tilde{\psi} (k) = \langle k | \psi \rangle \) or even use other bases as well.
D. Operators

An operator is a mapping from kets to kets. For example, if $A$ is an operator, then for any ket $|\psi\rangle$, $A(|\psi\rangle)$ will be another ket. An operator is linear if it has the property

$$A(c_1|\psi_1\rangle + c_2|\psi_2\rangle) = c_1A(|\psi_1\rangle) + c_2A(|\psi_2\rangle)$$

for $c_1$ and $c_2$ arbitrary complex numbers. Though operators do not need to be linear, we virtually always use linear operators in quantum mechanics, and therefore we will generally use the term “operator” to mean “linear operator.” To save time, the combination $A(|\psi\rangle)$ is often abbreviated $A\psi$. Unlike bras and kets, there is no generally accepted notation to distinguish operators from other objects. I will usually use capital letters for operators, but I do not promise to be consistent; indeed, in some cases, I will definitely not use capital letters.

It is not hard to show that operators also form a vector space. If we define the combination $c_1A_1 + c_2A_2$ by its action on an arbitrary ket by

$$(c_1A_1 + c_2A_2)|\psi\rangle = c_1(A_1|\psi\rangle) + c_2(A_2|\psi\rangle)$$

then we have defined this linear combination, and it is in fact another (linear) operator.

We can define the product of two operators $A$ and $B$ by

$$(AB)|\psi\rangle = A(B|\psi\rangle);$$

in other words, let first $B$ act on the ket and then $A$. By this definition, it is easy to prove various properties, such as the associativity of operators. Specifically, we see that

$$((AB)C)|\psi\rangle = (AB)(C|\psi\rangle) = A(B(C|\psi\rangle)) = A((BC)|\psi\rangle) = (A(BC))|\psi\rangle, \quad (3.33)$$

and since we have demonstrated this for an arbitrary ket,

$$(AB)C = A(BC). \quad (3.34)$$

Indeed, with the help of (3.32) and (3.34), it is easy to see that when multiplying operators, there is no point in writing the parentheses in, so normally any of the expressions in (3.33) would be written as simply $ABC|\psi\rangle$. Similarly, it can be shown that the distributive law works the same way, so we can generally use our intuition from algebra and keep things straight. However, one thing that is not generally true is that two operators will commute; in general, $AB \neq BA$. This can lead to difficulties; for example, it is not hard to show that $(A + B)^2 = A^2 + AB + BA + B^2$, but we must be careful not to combine the two terms in the middle into $2AB$, because this will generally give the wrong answer.

It is high time I gave several examples of operators to illustrate what I am talking about.
**Example 1:** Let $|\phi_1\rangle$ and $\langle \phi_2 |$ be any ket and bra, then define the operator $|\phi_1\rangle\langle \phi_2 |$ by

$$|\phi_1\rangle\langle \phi_2 |\langle \psi | \equiv |\phi_1\rangle\langle \phi_2 |\psi\rangle. \quad (3.35)$$

The expression $\langle \phi_2 |\psi\rangle$ is a complex number, and therefore the right side of (3.35) is a ket, and hence this is an operator. It is easy to show it is a linear operator. This operator is sometimes called the *outer product* of $|\phi_1\rangle$ and $\langle \phi_2 |$.

**Example 2:** The three operators $\mathbf{R} = (R_x, R_y, R_z) = (X, Y, Z)$, defined as follows: If $|\psi\rangle$ has wave function $\psi (\mathbf{r})$, we define $X|\psi\rangle$ as the ket with wave function $x\psi (\mathbf{r})$, etc., so

$$\mathbf{R}\psi (\mathbf{r}) \equiv \mathbf{r}\psi (\mathbf{r}). \quad (3.36)$$

**Example 3:** The three operators $\mathbf{P} = (P_x, P_y, P_z)$, which take the derivatives of the wave functions, or more precisely

$$\mathbf{P}\psi (\mathbf{r}) \equiv -i\hbar \nabla \psi (\mathbf{r}). \quad (3.37)$$

**Example 4:** The Hamiltonian operator, defined by

$$H = \frac{\mathbf{P}^2}{2m} + V (\mathbf{R}). \quad (3.38)$$

where $m$ is the mass of the particle, $V$ is the potential, and $\mathbf{P}$ and $\mathbf{R}$ were defined previously. Lest this chapter be completely devoid of physics, I will point out that Schrödinger’s equation in this notation is simply

$$H \Psi (t) = i\hbar \frac{\partial}{\partial t} |\Psi (t)\rangle. \quad (3.39)$$

**Example 5:** The identity operator $\mathbf{1}$, defined as $\mathbf{1}|\psi\rangle \equiv |\psi\rangle$. Indeed, I can multiply this identity operator by any complex number and produce more operators. Even though I used the dramatic notation $\mathbf{1}$ to emphasize this is an operator, this is rarely done, and I’ve never seen, for example, $2$ to represent the operator that is twice as big; it is simply written as $2$. Indeed, I will recklessly write $\mathbf{1}$ as $1$ as soon as equation (3.41).

**Example 6:** The parity operator $\Pi$, which reflects a wave function through the origin, so

$$\Pi\psi (\mathbf{r}) \equiv \psi (-\mathbf{r}). \quad (3.40)$$

Now that we have lots of examples of operators, it is worth pausing a moment to derive the completeness relation, a better way of working with expressions like (3.27) and (3.28). Suppose we have a complete orthonormal basis $\{|\phi\rangle\}$ and an arbitrary wave function $|\psi\rangle$. Then from (3.27) and (3.28) we have
\[
|\psi\rangle = \sum_i c_i |\phi_i\rangle = \sum_i |\phi_i\rangle \langle \psi \rangle = \left( \sum_i |\phi_i\rangle \langle \phi_i | \right) |\psi\rangle.
\]

Since this is true for any ket, it follows that
\[
\sum_i |\phi_i\rangle \langle \phi_i | = 1,
\]
where the 1 on the right represents the identity operator. This is called the completeness relation. We will use it constantly. If we are working in a continuum basis, it will take the form
\[
\int dv |\phi_i\rangle \langle \phi_i | = 1.
\]

For example, for the bases \(|r\rangle\) and \(|k\rangle\) we have
\[
\int d^3r |r\rangle \langle r | = \int d^3k |k\rangle \langle k |.
\]

Though operators are defined by how they turn kets into kets, we can also use them to map bras to bras. We will write such action as \(\langle \phi | A \), which is slightly backwards, or abbreviate it \(\langle \phi | A \), and define it in terms of its action on an arbitrary ket \(|\psi\rangle\) by
\[
\left( \langle \phi | A \right) |\psi\rangle \equiv \langle \phi | (A |\psi\rangle).
\]

We now see why we write this action as \(\langle \phi | A \), since that way (3.43) makes sense. Indeed, we normally do not even write in the parentheses, and instead merely write either side of (3.43) as \(\langle \phi | A |\psi\rangle\). In fact, it is sometimes just as convenient to define an operator by how it affects all bras on the left as it is in terms of how it affects all kets on the right. This will become especially helpful when we define the adjoint operator \(A^\dagger\) in section F.

E. Commutators

As has already been noted, it is common to encounter operators that do not commute. If \(A\) and \(B\) are two operators, we define the commutator of \(A\) and \(B\) as
\[
[A, B] \equiv AB - BA.
\]

For example, let’s find the commutator of \(X\) and \(P_x\). To do so, we let them act on an arbitrary wave function \(|\psi\rangle\). The commutator, therefore, is
\[
[X, P_x] |\psi\rangle = X P_x |\psi\rangle - P_x X |\psi\rangle = x \left( -i\hbar \frac{\partial}{\partial x} \right) |\psi\rangle - \left( -i\hbar \frac{\partial}{\partial x} \right) \left[ x |\psi\rangle \right]
\]
\[
= -i\hbar x \frac{\partial}{\partial x} |\psi\rangle + i\hbar \left[ |\psi\rangle + x \frac{\partial}{\partial x} |\psi\rangle \right] = i\hbar |\psi\rangle.
\]
Since this must be true for any wave function, it follows that \([X, P_x] = i\hbar\). This can be easily generalized to

\[
[R_i, P_j] = i\hbar\delta_{ij}.
\]

(3.45)

We will also commonly encounter the commutator of these operators with themselves, which can easily be seen to be

\[
[R_i, R_j] = [P_i, P_j] = 0.
\]

(3.46)

In general, when we encounter more complicated commutators, it is generally easier to use simple ones like (3.45) and (3.46), together with a few simple rules. The following identities are easily proven:

\[
[B, A] = -[A, B],
\]

(3.47a)

\[
[A + B, C] = [A, C] + [B, C] \quad \text{and} \quad [A, B + C] = [A, B] + [A, C],
\]

(3.47b)

\[
\]

(3.47c)

For example, to calculate the commutator \([P^2, X]\), we would simply find

\[
[P^2, X] = -\sum_i \left[ X, P_i^2 \right] = -\sum_i \left[ P_i X + X P_i \right] P_i = -\sum_i 2i\hbar P_i \delta_{ix} = -2i\hbar P_x.
\]

This generalizes to

\[
[P^2, \mathbf{R}] = -2i\hbar \mathbf{P}.
\]

(3.48)

F. Adjoint Operators

We have defined in section B the Hermitian conjugate of a ket or bra as the corresponding bra or ket, as indicated by (3.17) and (3.18). We also found the Hermitian conjugate of a scalar quantity is the same thing as complex conjugation. We’d now like to define Hermitian conjugation for an arbitrary operator. Let \(A\) be any operator, then define the adjoint operator \(A^\dagger\) by its action on an arbitrary bra:

\[
\langle \psi | A^\dagger = \left( A | \psi \right)^\dagger,
\]

where \( |\psi\rangle = \langle \psi |^\dagger\) is the ket corresponding to the bra \(\langle \psi |\) as defined in (3.13). The combination \(A |\psi\rangle\) is a ket, and hence \((A |\psi\rangle)^\dagger\) is a bra, so \(A^\dagger\) maps bras to bras, as an operator should. It is a linear operator, because
\[(c_1 \langle \psi_1 | + c_2 \langle \psi_2 |) A^\dagger = A \left( c_1 \langle \psi_1 | + c_2 \langle \psi_2 | \right)^\dagger = \left( A \left( c_1 \langle \psi_1 | + c_2 \langle \psi_2 | \right) \right)^\dagger = \left[ c_1^* A | \psi_1 \rangle + c_2^* A | \psi_2 \rangle \right]^\dagger = c_1 \left( A | \psi_1 \rangle \right)^\dagger + c_2 \left( A | \psi_2 \rangle \right)^\dagger = c_1 \langle \psi_1 | A^\dagger + c_2 \langle \psi_2 | A^\dagger. \]

It is easy to show that for any ket and bra \(|\psi\rangle\) and \(\langle \phi|\), we have
\[
\langle \phi| A^\dagger |\psi\rangle = \langle A | \phi \rangle^\dagger |\psi\rangle = \langle A \phi, \psi \rangle = \langle \psi, A \phi \rangle^\ast = \langle \psi | A | \phi \rangle^\ast. \tag{3.49}
\]

Indeed, it is not hard to show that this equation is sufficient to define the adjoint operator. It is pretty easy to show, for example, that the adjoint of the sum of operators is the sum of the adjoints, but the product is a bit trickier to work out. We see that if \(A\) and \(B\) are operators, then,
\[
\langle \psi | (AB)^\dagger \rangle = \left[ \langle AB | \psi \rangle \right]^\dagger = \left[ A \left( B | \psi \rangle \right) \right]^\dagger = \left( \langle B | \psi \rangle \right)^\dagger A^\dagger = \langle \psi | B^\dagger A^\dagger \rangle,
\]
\[
(AB)^\dagger = B^\dagger A^\dagger.
\]

So when we take the adjoint of a product of operators, we get them in the opposite order. In conclusion, we can take the Hermitian conjugate of any expression involving complex numbers, kets, and bras, by the following prescription:

1. If terms are added, you can treat each term separately
2. If factors are multiplied, you must reverse the order of those terms
3. Complex numbers must be complex conjugated
4. Operators must be replaced by their adjoints.

For example, suppose we came across an expression like
\[
|\Psi(t)\rangle = \sum_j e^{-i\omega_j t} |\phi_j\rangle \langle \phi_j | U |\psi\rangle,
\]
then we could take the Hermitian conjugate of this expression to give
\[
\langle \Psi(t) | = \sum_j \langle \psi | U^\dagger |\phi_j\rangle \langle \phi_j | e^{i\omega_j t}.
\]

There might be some confusion here, since the rules seem to be ambiguous; do we treat \(\langle \phi_j | U |\psi\rangle\) as a complex number (which it is) or as a combination of a bra, an operator, and a ket (which it also is)? The answer, thanks to (3.49), is that it doesn’t matter.

There are two categories of operators that deserve special treatment. Some operators are self-adjoint; that is, \(A^\dagger = A\). It follows from (3.49) that such an operator will satisfy
\[
\langle \phi | A |\psi\rangle = \langle \psi | A | \phi \rangle^\ast. \tag{3.50}
\]

Indeed, (3.50) is usually the easiest way to prove that something is self-adjoint. For example, to prove it for the operator \(X\), we note that
\[
\langle \psi | X | \phi \rangle^* = \left[ \iiint \psi^*(\mathbf{r}) x \phi(\mathbf{r}) d^3 \mathbf{r} \right]^* = \iiint \phi^*(\mathbf{r}) x \psi(\mathbf{r}) d^3 \mathbf{r} = \langle \phi | X | \psi \rangle,
\]
so it is self-adjoint. Harder to prove is the momentum operators. We note that
\[
\langle \psi | P_x | \phi \rangle^* - \langle \phi | P_x | \psi \rangle = \left[ \iiint \psi^*(\mathbf{r}) \frac{\partial}{\partial x} \phi(\mathbf{r}) d^3 \mathbf{r} \right]^* - \iiint \phi^*(\mathbf{r}) \frac{\partial}{\partial x} \psi(\mathbf{r}) d^3 \mathbf{r} = -i \hbar \iiint \left[ \psi(\mathbf{r}) \frac{\partial}{\partial x} \phi^*(\mathbf{r}) + \phi^*(\mathbf{r}) \frac{\partial}{\partial x} \psi(\mathbf{r}) \right] d^3 \mathbf{r} = i \hbar \iiint \frac{\partial}{\partial x} \left[ \psi(\mathbf{r}) \phi^*(\mathbf{r}) \right] d^3 \mathbf{r}.
\]
The right side can be integrated over \( x \), leaving us with the expression
\[
\langle \psi | P_x | \phi \rangle^* - \langle \phi | P_x | \psi \rangle = \iiint \left[ \psi(\mathbf{r}) \phi^*(\mathbf{r}) \right]_x=-\infty^\infty dy dz.
\]
If the wave functions lie in Hilbert space, then they must fall off at infinity, and hence this vanishes, demonstrating that \( P_x \) is also self-adjoint. In summary, we have
\[
\mathbf{R}^* = \mathbf{R} \quad \text{and} \quad \mathbf{P}^* = \mathbf{P}.
\]
(3.51)

For reasons that will become apparent later, operators that are self-adjoint are also called Hermitian operators.\(^1\) The other special category of operators are unitary operators, defined by
\[
U^* U = 1
\]
It is easy to show that for any ket \( | \psi \rangle \), the norm of \( | \psi \rangle \) and \( U | \psi \rangle \), defined by (3.3), will be identical. For now, I will state without proof that \( \mathbf{P} \) and \( \mathbf{R} \) are not unitary, whereas multiplication by a complex number that is pure phase, \( e^{i\theta} \), is unitary but not Hermitian. The parity operator, defined by (3.40), is both unitary and Hermitian. We will encounter many other examples of both unitary and Hermitian operators. It is worth noting that the product of two unitary operators is always unitary. In contrast, the product of two Hermitian operators is Hermitian if and only if the two operators commute. In particular, the Hamiltonian, defined by (3.39), is Hermitian.

\(^1\) Ballentine claims that there is a distinction between the two, but thus far, I’ve never encountered a situation where the distinction needs to be made.
G. Matrix Notation

Much of our discussion of bras, kets, and operators has been done in the abstract. It is very helpful to develop a matrix interpretation where these abstract concepts are replaced by relatively concrete matrices of complex numbers.

Suppose we have a complete orthonormal basis \( \{ | \phi_i \rangle \} \), then using the completeness relation (3.41) any ket may be written as

\[
| v \rangle = \sum_i | \phi_i \rangle \langle \phi_i | v \rangle = \sum_i | \phi_i \rangle v_i,
\]

where we have defined \( v_i = \langle \phi_i | v \rangle \), a set of complex numbers. This is simply the components of \( | v \rangle \) in a particular basis. Indeed, it is appropriate to regard the components \( v_i \) as identical to \( | v \rangle \), just as in three dimensions we would write \( \mathbf{r} = (x, y, z) \). We place all the components \( v_i \) together into a “column matrix” and write

\[
| v \rangle = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \end{pmatrix}.
\]

Of course, Hilbert space is infinite dimensional, and hence there will be an infinite number of components in this expression. Nonetheless, we will sometimes find this notation useful, both because it will help us think about what is going using the relatively familiar mathematics of matrices rather than the abstract notation of vector spaces, but also because we will sometimes be able to restrict ourselves and consider only a small subspace of the full Hilbert space that is finite dimensional.

In a similar manner, we can rewrite any bra \( \langle f | \) using (3.41) again, so we have

\[
\langle f | = \sum_i \langle f | \phi_i \rangle \langle \phi_i | = \sum_i f_i \langle \phi_i |,
\]

where \( f_i = \langle f | \phi_i \rangle \). We’ll place these components into a “row matrix” and write

\[
\langle f | = (f_1 \quad f_2 \quad \cdots).
\]

For an operator, we write

\[
A = \sum_i \sum_j | \phi_i \rangle \langle \phi_i | A | \phi_j \rangle \langle \phi_j | = \sum_i \sum_j | \phi_i \rangle A_{ij} \langle \phi_j |,
\]

(3.52)

where \( A_{ij} = \langle \phi_i | A | \phi_j \rangle \). Such components can be placed into a square matrix\(^1\)

\[
A = \begin{pmatrix} A_{11} & A_{12} & \cdots \\ A_{21} & A_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.
\]

\(^1\) Note the convention that the first index \( i \) of \( A_{ij} \) represents the row, and the second \( j \) the column.
Now, consider any combination of bras, kets, and operators that we have been asked to calculate; for example,

$$\langle f \mid A \mid \psi \rangle = \sum_i \sum_j \langle f \mid \phi_i \rangle \langle \phi_i \mid A \rangle \langle A \mid \phi_j \rangle = \sum_j \sum_i f, A_i \psi_j \cdot$$

The expression on the right is simply matrix multiplication, because you are multiplying rows by columns. Hence to combine any combination of bras, kets, and operators, you simply multiply the corresponding matrices.

What does Hermitian conjugation look like in matrix notation? For complex numbers, of course, it just means take their complex conjugate. For a ket, it would become a bra, but since \( \langle \psi \mid \phi_i \rangle = \langle \phi_i \mid \psi \rangle^* \), it also means taking the complex conjugate of the components. Similarly, for a bra, it would become a ket and you would take the complex conjugate of the components. Finally, for an operator \( A \), we see from (3.50) that

$$A_{ij}^* = \langle \phi_i \mid A^\dagger \mid \phi_j \rangle = \langle \phi_i \mid A \rangle^* = A_{ji}^* ,$$

or, using matrix notation, we can write this as

$$A^\dagger = A^* ,$$

where the transpose operator \( T \) denotes swapping rows with columns. In fact, this rule also works for bras, kets, and even for complex numbers (thought of as a \( 1 \times 1 \) matrix), so it is probably best to think of this as the matrix version of all Hermitian conjugation. A Hermitian operator, of course, will be one where \( A_{ij}^* = A_{ji} \), which means that its diagonal components will be real and its off-diagonal ones will come in complex conjugate pairs.

The components of any operator, bra, or ket will depend on the choice of coordinate basis. It is occasionally necessary to change bases. Suppose you have two orthonormal bases, \( \{ | \phi_i \rangle \} \) and \( \{ | \phi_i' \rangle \} \). First define the square matrix

$$V_{ij} = \langle \phi_i \mid \phi_j \rangle . \quad (3.53)$$

It is easy to show that \( V \) is a unitary matrix by considering the components of \( V^\dagger V \):

$$V^\dagger V_{ij} = \sum_k V^\dagger_{ik} V_{kj} = \sum_k V_{ki}^* V_{kj} = \sum_k \langle \phi_k \mid \phi_i \rangle^* \langle \phi_k \mid \phi_j \rangle = \sum_k \langle \phi_k \mid \phi_i \rangle \langle \phi_k \mid \phi_j \rangle = \langle \phi_i \mid \phi_j \rangle = \delta_{ij} ,$$

so \( V^\dagger V = 1 \). Now, to convert, say, the components of an operator \( A \) from the original to the new basis, we have

$$A'_{ij} = \langle \phi_i \mid A \mid \phi_j \rangle = \sum_k \sum_l \langle \phi_i \mid \phi_k \rangle \langle \phi_k \mid A \rangle \langle A \mid \phi_l \rangle \langle \phi_l \mid \phi_j \rangle = \sum_k \sum_l V_{ki}^* A_{kl} V_{lj} = \sum_k \sum_i V_{ki}^* A_{ki} V_{ij} ,$$

or, in matrix notation, \( A' = VAV^\dagger \). We can similarly convert kets and bras, with result

$$A' = V^\dagger AV , \quad \langle w' \rangle = V^\dagger \langle w \rangle , \quad \langle f \rangle^* = \langle f \mid V \rangle . \quad (3.54)$$

It should be emphasized that equation (3.54) relates operators, kets, and bras that are identical, but simply written in different bases. \( A \) and \( A' \) are the same operator, expressed in different bases.
H. Eigenvectors and Eigenvalues

Two concepts that will prove necessary to understanding quantum mechanics are eigenvalues and eigenvectors. An eigenvector of an operator $A$ is a non-zero ket $|v\rangle$ that satisfies the equation

$$A|v\rangle = a|v\rangle,$$

where $a$ is a complex number called the eigenvalue. If $|v\rangle$ is an eigenvector, then so is any multiple of $|v\rangle$, and hence we can multiply $|v\rangle$ by any constant. When possible, we try to normalize our eigenvectors so that $\langle v|v \rangle = 1$.

Eigenvectors are tricky for general operators, but for Hermitian operators they are a bit simpler. Suppose $A$ is a Hermitian operator with eigenvector $|v\rangle$ and eigenvalue $a$. Taking the Hermitian conjugate of equation (3.93), we have, for a Hermitian operator,

$$\langle v|A = \langle v|a^*.$$

It follows that

$$a \langle v|v \rangle = \langle v|a|v\rangle = \langle v|A|v\rangle = \langle v|a^*|v\rangle = a^* \langle v|v \rangle.$$

Since $\langle v|v \rangle \neq 0$, it follows that $a = a^*$, in other words, the eigenvalues of a Hermitian operator are real.

Now, suppose $a_1$ and $a_2$ are two distinct eigenvalues of a Hermitian operator $A$, and $|v_1\rangle$ and $|v_2\rangle$ are the corresponding eigenvectors. Then we can see that

$$a_1 \langle v_2|v_1 \rangle = \langle v_2|a_1|v_1 \rangle = \langle v_2|A|v_1 \rangle = \langle v_2|a_2|v_1 \rangle = a_2 \langle v_2|v_1 \rangle.$$

This implies $(a_1 - a_2) \langle v_2|v_1 \rangle = 0$, and since $a_1$ and $a_2$ were assumed to be distinct, it follows that $\langle v_2|v_1 \rangle = 0$, so eigenvectors of a Hermitian operator with different eigenvalues are orthogonal.

In a finite dimensional space, it can be shown that the eigenvectors of a Hermitian operator form a complete set; that is, every vector can be written as a sum of eigenvectors of the operator, but this is not always true in infinite dimensional spaces. We now define an observable as any Hermitian operator whose eigenvectors form a complete set. Let $A$ be any operator. Make a list of eigenvectors sufficiently large that they are complete. We then use the procedure laid out in section C to produce a complete, independent, orthonormal basis. Because eigenvectors that have different eigenvalues are orthogonal, the procedure laid out in equation (3.25) will only mix kets which have the same eigenvalue. Hence the complete orthonormal basis that results will all still be eigenvectors of $A$. Hence we can choose our basis states to be eigenvectors of $A$ as well. In such a case, it is common to label our states (partly) by their eigenvalues under the operator $A$, so such states might be labeled $|a,n\rangle$, with $a$ telling the eigenvalue under the operator $A$, so $A|a,n\rangle = a|a,n\rangle$ and $n$ distinguishing (when necessary) between different
basis vectors with the same eigenvalue. For example, in a six-dimensional space, $A$ might look like this:

$$A = \begin{pmatrix}
  a_1 & a_1 & 0 \\
  a_1 & a_2 & 0 \\
  0 & a_3 & a_3
\end{pmatrix},$$  \hspace{1cm} (3.55)

where the 0’s indicate that all the non-diagonal elements vanish. In this case, the set of vectors with eigenvalue $a_1$ would then form a three-dimensional subspace, consisting of all vectors that are non-vanishing in the first three components.

We can do better than that. Suppose we have two operators $A$ and $B$ that commute, so $AB = BA$. Then suppose we already have a basis $\{a, n\}$ which are eigenstates of $A$. It is obvious that

$$A(B|a, n\rangle) = BA|a, n\rangle = Ba|a, n\rangle = a(B|a, n\rangle)$$

As a consequence, $B|a, n\rangle$ has eigenvalue $a$ under $A$, and it is an element of the subspace which has this eigenvalue. Hence $B$ keeps vectors within this subspace still within this subspace. For example, if $A$ takes the form given in (3.55), then $B$ would take the form

$$B = \begin{pmatrix}
  b_{11} & b_{12} & b_{13} & 0 \\
  b_{21} & b_{22} & b_{23} & 0 \\
  b_{31} & b_{32} & b_{33} & 0 \\
  0 & b_{44} & b_{55} & b_{66}
\end{pmatrix}.$$  

Such a form is called block diagonal because the only non-vanishing components are square blocks that lie along the diagonal, as indicated by the red dotted.

Working within this subspace for fixed $a$, we can then find a new orthonormal basis which are eigenstates of $B$, call them $|a, b, m\rangle$, which all have eigenvalue $a$ under $A$ and various eigenvalues $b$ under $B$. We then repeat this process until we have explore all eigenvalues $a$. In the end, we end up with a basis which is simultaneously eigenvectors of both $A$ and $B$, so both $A$ and $B$ will end up looking like (3.55). The process can be repeated as many times as you want. In summary, given a set of commuting observables, a basis can be found such that every basis vector is simultaneously eigenstates of each of these observables. In general, if this process is repeated enough times, we will end up with enough operators such that the eigenstates are completely labeled by their eigenvalues. When this happens, we call such a set of observables a
complete set of commuting observables (CSCO’s for short). So if our operators are \( \{ A, B, \ldots, Z \} \), our basis will look like \( \{|a, b, \ldots, z\}\) , and will satisfy equations like

\[
A |a, b, \ldots, z\rangle = a |a, b, \ldots, z\rangle ,
B |a, b, \ldots, z\rangle = b |a, b, \ldots, z\rangle ,
\vdots
Z |a, b, \ldots, z\rangle = z |a, b, \ldots, z\rangle .
\]

Please note that in the basis \( \{|a, b, \ldots, z\}\) which are eigenvalues of \( \{ A, B, \ldots, Z \} \) , all of these matrices are diagonal. This process of finding such a basis is, therefore, commonly called diagonalizing the operators.

I. Finding Eigenvectors and Eigenvalues

Much of the work of the rest of this course will involve finding eigenvectors and eigenvalues. When we have only formulas for our operators, such as the Hamiltonian, this will prove difficult, but if we have a matrix form, it is straightforward. Suppose you are given a matrix representation of an operator \( A \) and asked to find the eigenvectors and eigenvalues of it. In other words, you are attempting to solve \( A |v\rangle = a |v\rangle \) , or

\[
(A - a \mathbf{1}) |v\rangle = 0 ,
\]

where we have explicitly included the identity operator \( \mathbf{1} \) for clarity. If we are working in an \( N \)-dimensional vector space, this is \( N \) equations in the unknown vector \( |v\rangle \), which has \( N \) components, so in general, there is a unique solution. This solution is the trivial solution, \( |v\rangle = 0 \), which is unacceptable. It must follow that in (3.56), the solutions are redundant, so the equations are not independent. In general, the matrix equations (3.56) can have non-trivial solutions only if the determinant of the matrix vanish. This allows us to find the eigenvalues of \( A \) by solving the characteristic equation, given by

\[
\det (A - a \mathbf{1}) = 0 .
\]

For a general matrix \( A \) of size \( N \times N \), equation (3.57) will be an \( N \)’th degree polynomial, which in general will have \( N \) roots, though some of them may not be distinct.\(^1\) Once you have the eigenvalues \( a \) from (3.57), you can plug them back into \( A |v\rangle = a |v\rangle \) to find \( |v\rangle \).

This vector will only be defined to (at best) an overall constant, which is determined by the demand that we want our states to be orthonormal, so \( \langle v | v \rangle = 1 \).

A few quick tricks that help make finding eigenvectors easier are in order. If a matrix is diagonal, then it is already diagonalized, so the basis we are using will form eigenvectors. If it is block diagonal, you can break it into pieces and solve each of the

\(^1\) In general, an \( N \)’th order polynomial has \( N \) complex roots, but for Hermitian matrices, as we have already stated, the eigenvalues will always be real, so (3.57) will yield only real roots.
pieces separately. If a matrix has an overall constant multiplied out front, then ignore the constant; the resulting eigenvectors will be correct, but the eigenvalues must be multiplied by that constant.

Let’s do a simple example to illustrate how this works. Suppose we were asked to diagonalize the matrix

$$\mathbf{H} = \begin{pmatrix} a & 0 & 0 \\
0 & b & 0 \\
0 & 0 & c \end{pmatrix},$$

where $a$, $b$ and $c$ are arbitrary real numbers. The first thing to note is that it is block diagonal, as I have illustrated with the red dashed lines I have added to the matrix. This allows us to reduce the problem to two smaller matrices:

$$
\begin{align*}
\mathbf{H}_1 &= \begin{pmatrix} a & 0 \\
0 & b \end{pmatrix} & \text{and} & \mathbf{H}_2 &= \begin{pmatrix} 0 & c \\
c & 0 \end{pmatrix}.
\end{align*}
$$

Now, $\mathbf{H}_1$ is diagonal, because it is non-zero only on its diagonal, and the two eigenvectors are simply trivial vectors in this basis, so we have

$$
|a\rangle = \begin{pmatrix} 1 \\
0 \end{pmatrix} \quad \text{and} \quad |b\rangle = \begin{pmatrix} 0 \\
1 \end{pmatrix}.
$$

(3.58)

These are eigenvectors of $\mathbf{H}_1$ with eigenvalues $a$ and $b$ respectively. Furthermore, they are already orthonormal. For $\mathbf{H}_2$, we will have to work harder.\(^1\) Let the eigenvalue be $\lambda$, then use equation (3.57) to find the eigenvalues:

$$
0 = \det \begin{pmatrix} 0 & c \\
c & 0 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\
0 & 1 \end{pmatrix} = \det \begin{pmatrix} -\lambda & c \\
c & -\lambda \end{pmatrix} = \lambda^2 - c^2.
$$

This equation has solutions $\lambda = \pm c$. For each of these eigenvalues, we give arbitrary names to the components of the eigenvector, so first let

$$
|c\rangle = \begin{pmatrix} \alpha \\
\beta \end{pmatrix}.
$$

(3.59)

Substitute this into (3.56) to yield

$$
\begin{pmatrix} 0 & c \\
c & 0 \end{pmatrix} \begin{pmatrix} \alpha \\
\beta \end{pmatrix} - c \begin{pmatrix} \alpha \\
\beta \end{pmatrix} = 0.
$$

Multiplying this out, we get the equation $\alpha = \beta$ twice, so our eigenvector is

\(^1\) At this point, it would not be a bad idea to remove the common factor of $c$ from $\mathbf{H}_2$, but the matrix is simple enough that this precaution is hardly necessary.
This is an eigenvector for any value of $\alpha$, but we would normally want to pick it to be normalized, so we want

$$1 = \langle c | c \rangle = (\alpha^* \alpha) \left( \begin{array}{c} \alpha \\ \alpha \end{array} \right) = 2|\alpha|^2.$$  \hspace{1cm} (3.60)

This determines the magnitude of the complex number $\alpha$, but not the phase, which is arbitrary. We’ll just pick $\alpha = 1/\sqrt{2}$, which then gives us the explicit form of the eigenvector. Repeating steps (3.59) through (3.60) for the remaining eigenvector $|-c\rangle$, we ultimately find the two eigenvectors of $H_2$ are

$$|c\rangle = \frac{1}{\sqrt{2}} \left( \begin{array}{c} 1 \\ 1 \end{array} \right) \quad \text{and} \quad |-c\rangle = \frac{1}{\sqrt{2}} \left( \begin{array}{c} 1 \\ -1 \end{array} \right).$$  \hspace{1cm} (3.61)

Between (3.58) and (3.61) we now have the eigenvectors of the two smaller blocks of $H_1$ and $H_2$, and we now simply put them together to get the eigenvectors of the original $H$

$$|a\rangle = \left( \begin{array}{c} 1 \\ 0 \\ 0 \end{array} \right), \quad |b\rangle = \left( \begin{array}{c} 0 \\ 1 \\ 0 \end{array} \right), \quad |c\rangle = \frac{1}{\sqrt{2}} \left( \begin{array}{c} 0 \\ 0 \\ 1 \end{array} \right), \quad |-c\rangle = \frac{1}{\sqrt{2}} \left( \begin{array}{c} 0 \\ 0 \\ -1 \end{array} \right).$$  \hspace{1cm} (3.62)

It is commonly the case that once you have found the eigenvalues and eigenvectors of a matrix in the original basis, you want to change basis, and write all operators, kets, and bras in the new basis. This requires merely that you construct the matrix $V$ given in (3.53). This is easy. The bras $\langle \phi_i |$ are the original basis, and the kets $|\phi'_i\rangle$ are the eigenvectors, so in matrix notation, we will have

$$V = \begin{pmatrix} |v_1\rangle & |v_2\rangle & \cdots \end{pmatrix},$$

Where $V$ is a square matrix, and each of the $|v_i\rangle$’s is a column matrix, collectively arranged into a square matrix. For example, if we want to diagonalize the matrix $H$, we can immediately deduce the matrix $V$ from the eigenvectors (3.62):

$$V = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & 0 \\ 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}.$$

You then convert any ket, bra, or operator to the new basis using (3.54). In particular, the matrix $H$ in the new basis will turn out to be diagonal, as it must, since the new basis was chosen to make sure $H$ would be diagonal:

52  © 2011, Eric D. Carlson
\[ H' = V^\dagger HV = \text{diag} \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & -c \end{pmatrix}. \]

The values on the diagonal are simply the corresponding eigenvalues for the four eigenvectors. Note that if our goal was to determine \( H' \), the matrix \( H \) in the new basis, we need only find the eigenvalues, which we can find from the characteristic equation (3.57), rather than facing the far more daunting task of finding all the eigenvectors.

Problems for Chapter 3

1. Prove Schwartz’s inequality, \((\phi, \psi)(\psi, \phi) \leq (\phi, \phi)(\psi, \psi)\). You may prove it however you want; however, here is one way to prove it. Expand out the inner product of \( a\phi + b\psi \) with itself, which must be positive, where \( a \) and \( b \) are arbitrary complex numbers. Then substitute in \( a = (\phi, \psi) \) and \( b = - (\phi, \phi) \). Simplify, and you should have the desired result.

2. Our goal in this problem is to develop an orthonormal basis for polynomial functions on the interval \([-1,1]\), with inner product defined by \( \langle f | g \rangle = \int_{-1}^{1} f^*(x) g(x) dx \).

Consider the basis function \( \phi_n \), for \( n = 0, 1, 2, \ldots \), defined by \( \phi_n(x) = x^n \).

(a) Find the inner product \( \langle \phi_n | \phi_m \rangle \) for arbitrary \( n, m \), and then use (3.25) to produce a set of orthogonal states \( \phi'_n \) for \( n \) up to 4.

(b) Now produce a set of orthonormal states \( \phi''_n \) using (3.26) for \( n \) up to 4.

(c) Compare the resulting polynomials with Legendre polynomials. How are they related?

3. Prove the following identities about the operators \( A, B, \) and \( C \):

(a) Commutators of products:

(b) The Jacobi identities:

(c) Show that if \( A \) and \( B \) are Hermitian, then \( i[A, B] \) is also Hermitian.
4. Define the angular momentum operators \( \mathbf{L} = \mathbf{R} \times \mathbf{P} \), or in components

\[
L_x = YP_z - ZP_y, \quad L_y = ZP_x - XP_z, \quad L_z = XP_y - YP_x.
\]

(a) With the help of problem 2(a), work out the six commutators \([L_z, \mathbf{R}]\) and \([L_z, \mathbf{P}]\).

(b) With the help of problems 2(a) and 3(a), work out the commutators \([L_z, L_x]\) and \([L_z, L_y]\).

5. Prove the parity operator \( \Pi \), defined by (3.40), is both Hermitian and unitary.

6. Consider the Hermitian matrix:

\[
H = E_0 \begin{pmatrix} 0 & 3i & 0 \\ -3i & 8 & 0 \\ 0 & 0 & 8 \end{pmatrix}.
\]

(a) Find all three eigenvalues and eigenvectors of \( H \).

(b) Construct the unitary matrix \( V \) which diagonalizes \( H \). Check explicitly that \( V^*V = 1 \) and \( V^*HV = H' \) is real and diagonal.
IV. The Postulates of Quantum Mechanics

Within physics, no subject is more controversial than the postulates of quantum mechanics. By and large, all (or most) physicists agree on how to perform calculations in quantum mechanics, but they disagree on exactly what the right way to formulate the postulates is. In these lectures, we will give what is termed the Copenhagen interpretation, but this is not to suggest that this is the only or even the preferred interpretation.

A. Revisiting Representations

In the previous chapter, we talked about observables, operators whose basis sets are complete, such that any wave function can be written in terms of them. It’s time to revisit some of our examples.

Two sets of operators we discussed first were the position and momentum operators $R$ and $P$, defined by

\[
\langle r | R | \psi \rangle = r \langle r | \psi \rangle, \tag{4.1a}
\]

\[
\langle r | P | \psi \rangle = -i\hbar \nabla \langle r | \psi \rangle. \tag{4.1b}
\]

If you take the Hermitian conjugate of (4.1a), it is easy to see that $|r\rangle$ is an eigenstate of all three operators $R$ with eigenvalues given by $r$. If, on the other hand, you substitute the explicit form (3.19) in (4.1b), you will find $|k\rangle$ is an eigenstate of $P$, so we have

\[
R |r\rangle = r |r\rangle, \quad P |k\rangle = \hbar k |k\rangle.
\]

Each of $\{|r\rangle\}$ and $\{|k\rangle\}$ form a complete basis, and since $R$ and $P$ are Hermitian operators, it follows, by definition, that they are observables. This correspondence of observables and things we can measure will be one of our postulates.

Furthermore, to calculate the probability of getting a particular value of the position or the momentum, we needed to calculate $|\psi (r)|^2 = |\langle r | \psi \rangle|^2$ or $|\psi (k)|^2 = |\langle k | \psi \rangle|^2$ respectively. This suggests that we need to develop a basis set for our operators, and calculate the magnitude squared of the inner product of these basis sets with our state vector to get the probability. We also noted that after a measurement of, say, the position, the particle will be at the position we measured it, i.e., the wave function must change to become an eigenstate of the operator we are measuring. Finally, when we are not performing measurements, we expect the wave function to evolve according to Schrödinger’s equation, or something similar.

We are now prepared to state the postulates of quantum mechanics.
B. The Postulates of Quantum Mechanics

Postulate 1: The state of a quantum mechanical system at time \( t \) can be described as a normalized ket \( |\Psi(t)\rangle \) in a complex vector space with a positive definite inner product.

Note that we have avoided specifying exactly which vector space we are in. For example, for a single spinless particle in 3 dimensions, we use the set of square integrable wave functions, but we would need more (or less) complicated vector spaces if we had particles with spin, or multiple particles, or lots of other possibilities. The stuff about a “positive definite inner product” just means that \( \langle \psi | \psi \rangle \geq 0 \), with equality only if \( \psi = 0 \).

The next statement we need to make is that the wave function evolves according to the Schrödinger equation.

Postulate 2: When you do not perform a measurement, the state vector evolves according to

\[
\frac{i\hbar}{dt} |\Psi(t)\rangle = H(t) |\Psi(t)\rangle,
\]

where \( H(t) \) is an observable.

We will associate \( H(t) \) with the energy of the system, but I don’t see such an assumption as a necessary part of the postulates. We will call this operator the Hamiltonian. Note we have not assumed any particular form for \( H \); doing so would of necessity require that we decide what vector space we are working in. Furthermore, there will be situations where our Hamiltonian will look radically different than cases we have discussed so far. The necessity of making the Hamiltonian Hermitian will be demonstrated in the next section. Stating that this is how it evolves when you aren’t looking at it is necessary when you compare this with postulate 5.

Postulate 3: For any quantity that one might measure there is a corresponding observable \( A \), and the results of that measurement can only be one of the eigenvalues \( \alpha \) of \( A \).

For example, if you measure the \( x \)-position \( X \) of a particle, you can get any number, because \( X \) takes on a continuum of eigenvalues. Later we will encounter examples where only discrete values are possible.
Postulate 4: Let \( \{a,n\} \) be an orthonormal basis of the observable \( A \), with \( A|a,n\rangle = a|a,n\rangle \), and let \( \Psi(t) \) be the state vector at time \( t \). Then the probability of getting the result \( a \) at time \( t \) will be given by

\[
P(a) = \sum_n \langle a,n|\Psi(t)\rangle^2.
\] (4.3)

This is akin to saying the probability density of finding a particle in one dimension at a point \( x \) is given by \( \|x|\Psi\|^2 \). The sum on \( n \) is necessary because there may be multiple basis states with the same eigenvalue \( a \).

We have not yet said that if you perform a measurement, and you get a specific value, the wave function afterwards must have that eigenvalue. This is assured by insisting that the wave function be “reduced” by extracting only those portions that have this eigenvalue.

Postulate 5: If the results of a measurement of the observable \( A \) at time \( t \) yields the result \( a \), the state vector immediately afterwards will be given by

\[
|\Psi(t^+)\rangle = \frac{1}{\sqrt{P(a)}} \sum_n |a,n\rangle \langle a,n|\Psi(t)\rangle.
\] (4.4)

As will be discussed later, this change in the wave function is very different from the Schrödinger evolution described by postulate 2. The superscript + is intended to denote the wave function just after the measurement, which is assumed to take zero time. It is easily demonstrated that if you measure \( A \) again, you will end up getting the result \( a \) again. Indeed, your wave function won’t change due to the second measurement.

The postulates, as presented here, presume that the eigenstates of our operator are discrete, rather than continuous, and that the other parameters \( n \) that describes our basis are discrete. Either or both may be false. If the other parameter \( n \) is continuous, this doesn’t really present much of a problem. You simply define your basis functions in a suitable way; for example, define them as \( |a,\nu\rangle \), and the sums over \( n \) in (4.3) and (4.4) will be converted to integrals over \( \nu \). If it bothers you that the form of the postulates depends on whether one is using a discrete or continuum basis, suffice it to say a common notation can be developed which will allow one to work with the two cases as if they were identical.\(^1\)

If the observable we are measuring has a continuum of eigenvalues, then we need to rethink a couple of the postulates. Suppose, for example, that \( A \) has a continuum of eigenvalues \( \alpha \), with basis states \( |\alpha,\nu\rangle \). These states are assumed to be normalized as

\[
\langle \alpha',\nu'|\alpha,\nu\rangle = \delta(\alpha'-\alpha)\delta(\nu'-\nu).
\]

Then the probability of getting exactly a particular value \( \alpha \) will automatically be zero, because with an infinite number of closely spaced possibilities the probability of getting

\(^1\) For a more extensive discussion of the mathematical details, consult Ballantine.
any one of them vanishes. However, we can still calculate the probability of getting $\alpha$ within a specific range. This modifies postulate 4.

**Postulate 4b:** Let $\{\alpha, \nu\}$ be an orthonormal basis of the observable $A$, and let $|\Psi(t)\rangle$ be the state vector at time $t$. Then the probability of getting the result $\alpha$ in the range $\alpha_1 < \alpha < \alpha_2$ at time $t$ will be given by

$$P(\alpha_1 < \alpha < \alpha_2) = \int_{\alpha_1}^{\alpha_2} d\alpha \int d\nu |\langle\alpha, \nu|\Psi(t)\rangle|^2.$$  \hspace{1cm} (4.5)

This works well for the fourth postulate, but what do we do with the last postulate? Suppose we measure the observable $A$ and discover that the value is $\alpha$. The problem is that the probability of getting this value is zero, because the integral (4.5) vanishes if we set the two limits on $\alpha$ equal. The key to resolving this problem is to recognize that *all* measurements of physical quantities represented by a continuous variable have errors.\(^1\) Thus, in fact, any measurement of a continuous variable doesn’t tell you $\alpha$, you only learn that $\alpha$ lies in some range, say $\alpha_1 < \alpha < \alpha_2$.

**Postulate 5b:** If the results of a measurement of the observable $A$ at time $t$ yields the continuous result $\alpha$ measured to lie in the range $\alpha_1 < \alpha < \alpha_2$, the state vector immediately afterwards will be given by

$$|\Psi(t^+)\rangle = \frac{1}{\sqrt{P(\alpha_1 < \alpha < \alpha_2)}} \int_{\alpha_1}^{\alpha_2} d\alpha \int d\nu |\langle\alpha, \nu|\Psi(t)\rangle|.$$  

Though we ended up with two versions of the fourth and fifth postulates, depending on the situation, we will actually always use the first version when performing a general proof. Repeating the proof with the second version tends to be tedious and provide no great insight.

For a single particle we already know a prescription for finding the Hamiltonian. Specifically, write out the energy as a function of the position $r$ and the momentum $p$, and then promote these two classical quantities to the quantum mechanical operator $R$ and $P$ respectively. For example,

$$E = \frac{1}{2m} p^2 + V(r, t) \rightarrow H = \frac{1}{2m} P^2 + V(R, t).$$

It should be emphasized that there is no necessity of describing a prescription for turning a classical theory into a quantum mechanical theory. Classical and quantum theories yield experimentally different descriptions of the universe, and the only relevant question is whether the observations will agree with one or the other theory. Hence the real question should not be, how do you find the Hamiltonian, but rather, is there a Hamiltonian that we can find that matches the universe we see.

\(^1\) Note that this is not necessarily true of discrete variables. I have two children, exactly two, with no error.
It should be noted that the postulates of quantum mechanics may look radically different depending on who you ask. The main problems seem to do with postulates 3, 4 and 5, which have to do with measurement. In particular, the concept of “measurement” is never exactly defined, though we have a sense of taking information “out there” in the universe and getting the information into our measuring devices (or brains, or published papers, or so on). Note that these postulates involve probability, rather than the cold certainty of Newtonian mechanics. We also seem to have two different rules regarding how the state vector changes, postulates 2 and 5. For these reasons, many have tried to rewrite these postulates in a variety of forms, and with varying success.

C. Consistency of the Postulates

The postulates contain a variety of statements that might or might not be consistent with each other. For example, postulate 1 states that the state vector must be normalized, but postulates 2 and 5 describe how the state vector changes without mentioning normalization. Will these two postulates be consistent with normalization? For example, consider Schrödinger’s equation. It is helpful to take the Hermitian conjugate of (4.2), which yields the equation

$$-i\hbar \frac{d}{dt} \langle \Psi(t) | = \langle \Psi(t) | H^* (t), \quad (4.6)$$

where we have, for the moment, deliberately forgotten that the Hamiltonian is Hermitian. Now (4.3) and (4.6) together allow us to calculate the time derivative of the norm squared of the wave function.

$$\frac{d}{dt} \langle \Psi(t) | \Psi(t) \rangle = \left[ \frac{d}{dt} \langle \Psi(t) | \right] \langle \Psi(t) | + \langle \Psi(t) | \frac{d}{dt} \langle t | \Psi(t) \rangle$$

$$= -i\hbar \langle \Psi(t) | H^* (t) \rangle \langle \Psi(t) | + i\hbar \langle \Psi(t) | H(t) \rangle \langle \Psi(t) |$$

$$= i\hbar \langle \Psi(t) | [H(t) - H^* (t)] \rangle \langle \Psi(t) |.$$ 

Remembering that the Hamiltonian is Hermitian, \( \langle \Psi(t) | \Psi(t) \rangle \) is constant, and therefore will remain normalized.

Similarly, if we perform a measurement, then (4.4) describes how the wave function changes. But is it normalized? We use (4.3) as well to find

$$\langle \Psi(t^+) | \Psi(t^+) \rangle = \frac{1}{P(a)} \left[ \sum_n |a, n\angle \langle a, n | \Psi(t) \rangle \right|^2$$

$$= \frac{1}{P(a)} \sum_n \sum_m \langle \Psi(t) | a, m \rangle \langle a, m | a, n \angle \langle a, n | \Psi(t) \rangle$$

$$= \frac{1}{P(a)} \sum_n \sum_m \langle \Psi(t) | a, m \rangle \delta_{nm} \langle a, n | \Psi(t) \rangle = \frac{1}{P(a)} \sum_n \langle a, n | \Psi(t) \rangle^2.$$
\[ \langle \Psi (r^*) | \Psi (r^*) \rangle = \frac{P(a)}{P(a)} = 1. \]

Hence these rules are consistent.

We also have a statement that the outcome has a certain probability. The probability of a particular event must always be a positive number, and the sum of all probabilities of every possible outcome must always equal 1. It is obvious from (4.3) that probabilities are positive, but do they add to one? This is not hard to prove using the completeness relation:

\[ \sum_a P(a) = \sum_a \sum_n |\langle a,n | \Psi (t) \rangle|^2 = \sum_{a,n} \langle \Psi (t) | a,n \rangle \langle a,n | \Psi (t) \rangle = \langle \Psi (t) | \Psi (t) \rangle = 1. \]

Another concern has to do with our choice of basis. If our observable \( A \) has only non-degenerate eigenvalues, then our eigenstates \( |a\rangle \) are defined up to a phase, but what if we have multiple states \( |a,n\rangle \) which share the same eigenvalue \( a \)? The specific choice of eigenstates appears in postulates 4 and 5. For example, let’s imagine that we have two choices of eigenvectors \( |a,n\rangle \), and \( |a,n'\rangle \). If we calculate the probability in the primed basis, then postulate 4 will look like

\[ P'(a) = \sum_{a'} \langle \Psi (t) | a,n' \rangle \langle a,n' | \Psi (t) \rangle. \]

We now use the completeness relation (3.41) twice to rewrite this as

\[ P'(a) = \sum_{a_1,n_1,a_2,n_2} \sum_{a',n'} \langle \Psi (t) | a_1,n_1 \rangle \langle a_1,n_1 | a_2,n_2 \rangle \langle a_2,n_2 | a' \rangle \langle a' | n' \rangle \langle n' | a_2,n_2 \rangle \langle a_2,n_2 | \Psi (t) \rangle, \]

where \( |a_1,n_1\rangle \) represents the unprimed basis. We now change the sums as follows. First, we note that since eigenstates with different eigenvalues are orthogonal, the sums on \( a_1 \) and \( a_2 \) will vanish except when \( a_1 = a_2 = a \), so we have

\[ P'(a) = \sum_{n_1} \sum_{n_2} \langle \Psi (t) | a_1,n_1 \rangle \langle a_1,n_1 | a_2,n_2 \rangle \langle a_2,n_2 | a_2,n_2 \rangle \langle a_2,n_2 | \Psi (t) \rangle. \]

We now do the same trick in reverse. For the sum over the states \( |a,n'\rangle \), note that we can add additional values of \( a \), which we call \( a' \), with impunity, since they will not contribute to the sum. So we have

\[ P'(a) = \sum_{n_1} \sum_{n_2} \sum_{a',n'} \langle \Psi (t) | a_1,n_1 \rangle \langle a_1,n_1 | a',n' \rangle \langle a',n' | a_2,n_2 \rangle \langle a_2,n_2 | \Psi (t) \rangle. \]

We then use completeness of the basis \( |a',n'\rangle \) and orthonormality of the basis \( |a,n_1\rangle \):

\[ P'(a) = \sum_{n_1} \sum_{n_2} \langle \Psi (t) | a_1,n_1 \rangle \langle a_1,n_1 | a_2,n_2 \rangle \langle a_2,n_2 | \Psi (t) \rangle = \sum_{n_1} \sum_{n_2} \langle \Psi (t) | a_1,n_1 \rangle \delta_{n_1,n_2} \langle a_2,n_2 | \Psi (t) \rangle \]

\[ = \sum_n \langle \Psi (t) | a,n \rangle \langle a,n | \Psi (t) \rangle = \sum_{n} |\langle a,n | \Psi (t) \rangle|^2 = P(a). \]
Hence the probability is independent of the choice of basis.

Similarly, we can show that (4.4) doesn’t depend on basis choice:

\[
\left| \Psi'(t') \right\rangle = \frac{1}{\sqrt{P'(a)}} \sum_{n} |a, n'\rangle \langle a, n| \Psi(t) \]

\[
= \frac{1}{\sqrt{P(a)}} \sum_{a_{1}, a_{2}, n_{1}, n_{2}} |a_{1}, n_{1}\rangle \langle a_{1}, n_{1}| a_{2}, n_{2}\rangle \langle a_{2}, n_{2}| \Psi(t) \]

\[
= \frac{1}{\sqrt{P(a)}} \sum_{a_{1}, a_{2}, n_{1}, n_{2}} |a_{1}, n_{1}\rangle \langle a_{1}, n_{1}| a_{2}', n_{2}\rangle \langle a_{2}', n_{2}| a_{2}, n_{2}\rangle \Psi(t) \]

\[
= \frac{1}{\sqrt{P(a)}} \sum_{n_{1}, n_{2}} \delta_{n_{1}, n_{2}} |a, n_{2}\rangle \Psi(t) = \frac{1}{\sqrt{P(a)}} \sum_{n} |a, n\rangle \langle a, n| \Psi(t) \]

\[
= |\Psi(t')\rangle.
\]

Thus the laws of quantum mechanics as laid out in section B are self-consistent, if rather complex and unpalatable. The primary flaw remaining in this formulation of quantum mechanics is the vague concept of measurement, and the distasteful necessity of using probabilities. These are problems we will return to later.

D. Measurements and Reduction of the Wave Function

One of the surprises about quantum mechanics is the way the wave function apparently changes suddenly when you perform a measurement. To illustrate some of the complexities involved, suppose we have a simple two-dimensional spin system. For the moment, consider a system which has two operators which we can measure, which I will call \(S_{x}\) and \(S_{z}\).\(^{1}\) In the two-dimensional space, we will assume these take the matrix form

\[
S_{x} = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad S_{z} = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.
\]

These are both Hermitian, and, as we will soon demonstrate, observables.

The basis states for \(S_{z}\) are trivial because the matrix is already diagonalized. The eigenvalues are \(\pm \frac{1}{2} \hbar\) and the corresponding eigenvectors are

\[
|+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |-\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.
\]

\(^{1}\) This is not actually an entirely hypothetical situation. The operators are two of the three spin operators \(S\) for a spin-\(\frac{1}{2}\) particle such as an electron.
The basis states for \( S_x \) were found in (3.61). The eigenvalues are again \( \pm \frac{1}{2} \hbar \) and the corresponding eigenvectors are

\[
\begin{align*}
|+,x\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
|-,x\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}
\end{align*}
\]

Now, suppose we start with an initial state

\[
|\Psi(t=0)\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |+,z\rangle.
\] (4.8)

We will now proceed to perform a series of measurements on the state (4.8), in sufficiently rapid succession that there is no time for the state to change due to Schrödinger’s equation.

Suppose we measure \( S_z \). \( S_z \) has eigenvalues \( \pm \frac{1}{2} \hbar \), so by postulate 3 the only result we can get is one of these two values. According to postulate 4 the probability of getting each of these values is given by

\[
P_z(\pm \frac{1}{2} \hbar) = \langle +,z | \Psi \rangle^2 = \left| \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right|^2 = 1,
\ P_z(-\frac{1}{2} \hbar) = \langle -,z | \Psi \rangle^2 = \left| \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right|^2 = 0.
\]

So the result is unambiguously that we will get the result \( + \frac{1}{2} \hbar \). After the measurement, the wave function, by postulate 5, will be given by

\[
|\Psi\rangle = \frac{1}{\sqrt{P_+}}|+,z\rangle \langle +,z | \Psi \rangle = |+,z\rangle,
\]

Where we have abbreviated \( P(\pm \frac{1}{2} \hbar) \) as \( P_+ \). It is completely unchanged. You can repeat the measurement as many times as you want, without changing the outcome.

Now suppose you measure \( S_x \) instead. According to postulate 4, the probability of getting different outcomes will be

\[
P_x = \langle +,x | \Psi \rangle^2 = \left| \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right|^2 = \frac{1}{2},
\ P_x = \langle -,x | \Psi \rangle^2 = \left| \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right|^2 = \frac{1}{2}.
\]

This time we can’t predict which outcome will occur. Furthermore, before we move on to postulate 5, we need to know which of the two outcomes actually occurs. Depending on which case we have, after the measurement the state will be

\[
|\Psi_+\rangle = \frac{1}{\sqrt{P_+}}|+,x\rangle \langle +,x | \Psi \rangle = \frac{\sqrt{1/2}}{\sqrt{1/2}}|+,x\rangle = |+,x\rangle,
\]

\[
|\Psi_-\rangle = \frac{1}{\sqrt{P_-}}|-,x\rangle \langle -,x | \Psi \rangle = \frac{\sqrt{1/2}}{\sqrt{1/2}}|-,x\rangle = |-,x\rangle.
\]

So we see in this case the wave function \( has \) changed as a result of the measurement. If you now perform further measurements of \( S_x \), it is not hard to show that you will continue
to get the same result again; i.e., the result is no longer probabilistic, instead it will
always come out to the same \(S_x\) value it did before. Let’s say, for example, that we got
the minus value, so the wave function is now \(|\Psi\rangle = \{-, x\rangle\).

Now switch back and measure \(S_z\) again. The probabilities for each of the possible
outcomes will now be

\[
P_+ = |\langle +, z |\Psi\rangle|^2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{2},
\]

\[
P_0 = |\langle -, z |\Psi\rangle|^2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{2}.
\]

Recall that initially, this same measurement of \(S_z\) produced the result \(+ \frac{1}{2} \hbar\) with
probability 1, but now the probability is split evenly between the possible outcomes
\(\pm \frac{1}{2} \hbar\). Performing measurement \(S_z\) changed the wave function, and as a result, the
probabilities changed for the measurement \(S_z\). It can then be shown that after this
measurement is performed, depending on the outcome, the wave function will take the
form \(|\Psi\rangle = |+, z\rangle\) or \(|\Psi\rangle = |-, z\rangle\).

Note that after every measurement the resulting wave function was an eigenstate
of the measurement just performed as demanded by postulate 5; indeed, in almost every
case it ended up in a simple basis vector. This does not generally happen. It happened
here because when applying postulate 5 we only had one state with each eigenvalue, and
hence the sum had only a single term. Since the final state is guaranteed to be normalized
it must be only that one basis function, times an arbitrary phase. Without going through
the argument in detail, it can be shown that there is no physical difference between a state
\(|\Psi\rangle\) and the same state multiplied by a phase \(e^{i\theta}|\Psi\rangle\). Hence when there is only one
basis vector for each eigenvalue there is no harm in stating that the state after a
measurement yielding a specific eigenvalue will be the corresponding basis vector.
However, if there are multiple basis vectors with the same eigenvalue, then we will
generally not get a single basis vector, and you must use (4.4) to calculate the new
quantum state \(|\Psi\rangle\).

Suppose we start over and consider only doing two measurements, \(S_x\) and \(S_z\) of the
initial wave function (4.8). If you do \(S_z\) first you will definitely get \(+\), but if you do \(S_x\)
and then \(S_z\) the result of the \(S_z\) measurement will be uncertain. In other words, the order
you perform measurements makes a difference. This happens because the two operators
\(S_z\) and \(S_x\) do not commute. It is not hard to show that if they do commute, then the order
does not matter, so the order in which you perform two measurements matters if the
corresponding operators do not commute. Note that in a similar manner, the order of
multiplication of two operators makes a difference if they do not commute (by
definition). At the risk of confusing you by saying something that sounds like a truism,
order matters if order matters.1

You might wonder what happens if two observables are measured simultaneously.
The short answer is that this must be impossible if the observables do not commute. If
two observables do actually commute, there is no reason in principle you can’t measure
them at the same time, and we say that the observables are simultaneous observables.

---

1 To make this clearer: Order matters when performing measurements if order matters when multiplying.
E. Expectation Values and Uncertainties

In chapter 1 we talked about the expectation value and uncertainty of the position and momentum. We want to generalize these to arbitrary operators. Let \( A \) be any operator. We’ll define the expectation value \( \bar{A} \) of \( A \) as the average value that we get for \( A \) if we measure it for some wave function \( |\psi\rangle \). Two notations are used, and we will use both of them, and we have

\[
\bar{A} = \langle A \rangle = \sum_a a P(a),
\]

(4.9)

where \( a \) represents the various possible results of the measurement, and \( P(a) \) the corresponding probabilities. With the help of postulate 4, we can rewrite this as

\[
\bar{A} = \sum_a a \sum_n |a, n \rangle \langle a, n| \psi\rangle^2 = \sum_a \sum_n a |\psi \rangle \langle a, n| \langle a, n| \psi\rangle.
\]

(4.10)

Now, \( |a, n \rangle \) is an eigenstate of \( A \) with \( A |a, n \rangle = a |a, n \rangle \), so we can use this to rewrite (4.10), then use completeness to show

\[
\bar{A} = \langle A \rangle = \sum_a \sum_n \langle \psi | A | a, n \rangle \langle a, n | \psi \rangle,
\]

\[
\bar{A} = \langle A \rangle = \langle \psi | A | \psi \rangle.
\]

(4.11)

Note that (4.11) allows us to find the expectation value of \( A \) without first writing it in terms of \( A \)'s basis vectors. Hence in chapter 1 when we were finding expectation values of \( p \) it was not necessary to find the Fourier transform of \( \tilde{\psi}(k) \); we could have worked directly with \( \psi(x) \) and just used the operator form of \( P \).

Although many measurements of similarly constructed wave functions \( |\psi\rangle \) will yield some average value, there will be some uncertainty in this value defined by

\[
(\Delta A)^2 \equiv \langle (A - \bar{A})^2 \rangle.
\]

A more useful version of this expression can be found by squaring out the expression, and realizing that any constant inside an expectation value can be taken outside, and the expectation value of any constant is just that constant. So we have

\[
(\Delta A)^2 = \langle A^2 - 2A\bar{A} + \bar{A}^2 \rangle = \langle A^2 \rangle - 2\bar{A} \langle A \rangle + \bar{A}^2 = \langle A^2 \rangle - 2\bar{A}^2 + \bar{A}^2,
\]

(4.12)

\[
(\Delta A)^2 = \langle A^2 \rangle - \langle A \rangle^2.
\]

When discussing uncertainties in chapter 2 we stated without proof that there was a limit on the product of the uncertainty of the position and momentum of the particle. We will now prove not only this, but a similar relationship between any two observables that do not commute. Let \( A \) and \( B \) be any two observables. Consider the ket defined by

\[
[\alpha(A - \bar{A}) + i\beta(B - \bar{B})] |\psi\rangle,
\]

where \( \alpha \) and \( \beta \) are complex numbers, and \( \bar{A} \) and \( \bar{B} \) are the expectation values of \( A \) and \( B \) respectively.
where $|\psi\rangle$ represents the state of our system at some time, and $\alpha$ and $\beta$ are real numbers to be specified later. All kets have a non-negative norm-squared, so we have

$$\left| \left\langle \alpha (A - \bar{A}) + i \beta (B - \bar{B}) \right| \psi \right|^2 \geq 0,$$

$$\langle \psi \left| \left[ \alpha (A - \bar{A}) - i \beta (B - \bar{B}) \right] \left[ \alpha (A - \bar{A}) + i \beta (B - \bar{B}) \right] \psi \rangle \geq 0,$$

$$\alpha^2 \langle \psi \left| (A - \bar{A})^2 \right| \psi \rangle + \beta^2 \langle \psi \left| (B - \bar{B})^2 \right| \psi \rangle + i \alpha \beta \langle \psi \left| \left( (A - \bar{A}) \cdot (B - \bar{B}) \right) \psi \right| \geq 0.$$

In the last term, notice the presence of a commutator. Since constants commute with anything, the last term simplifies a lot and we have

$$\alpha^2 (\Delta A)^2 + \beta^2 (\Delta B)^2 + \alpha \beta \langle \psi \left| [A, B] \right| \psi \rangle \geq 0.$$

We now make the substitution $\alpha = \Delta B$ and $\beta = \mp \Delta A$ to yield two inequalities:

$$2 (\Delta A)^2 (\Delta B)^2 \mp (\Delta A)(\Delta B) \langle \psi \left| i [A, B] \right| \psi \rangle \geq 0.$$

We divide this equation by $2 (\Delta A)(\Delta B)$, a positive number$^1$, and discover

$$(\Delta A)(\Delta B) \geq \frac{1}{2} \langle \psi \left| i [A, B] \right| \psi \rangle. \quad (4.13)$$

The two signs in (4.13) represent two true statements; that is, (4.13) is true with both signs, but whichever one is positive gives the stronger statement, so we have

$$(\Delta A)(\Delta B) \geq \frac{1}{2} \left| \langle i [A, B] \rangle \right|. \quad (4.14)$$

This is the generalized uncertainty principle. For example, let $A$ be any of our position operators, and $B$ any of our momentum operators, then we have

$$(\Delta R_i)(\Delta P_j) \geq \frac{1}{2} \left| \langle i [R_i, P_j] \rangle \right| = \frac{1}{2} \left| \langle i (\hbar \delta_{ij} - \hat{\delta}_{ij} \hat{\delta}_{ij}) \rangle \right| = \frac{1}{2} \hat{\delta}_{ij} \hbar = \frac{1}{2} \delta_{ij} \hbar.$$

In one dimension we normally write this as $(\Delta x)(\Delta p) \geq \frac{1}{2} \hbar$, but we are now using operator notation, where the position $x$ is associated with the operator $R_x = X$ and momentum $p$ is associated with the operator $P_x$. Note, however, that (4.14) applies to any two operators, though if those operators commute it is a trivial relation.

---

$^1$ You might be concerned that we can have $(\Delta A)(\Delta B) = 0$, invalidating our proof. Though this is true, one can show this only occurs when $|\psi\rangle$ is an eigenstate of one of our two operators, and in this case, the right hand side of (4.13) vanishes, so that our uncertainty relation becomes the trivial relationship $0 \geq 0$. 

© 2011, Eric D. Carlson
F. Evolution of Expectation Values

Let $A$ be any observable, and we allow for the possibility that it might explicitly depend on time. How will its expectation value change with time due to evolution under Schrödinger’s equation? With the help of Schrödinger’s equation (4.2) and its Hermitian conjugate (4.6), we find

$$\frac{d}{dt} \langle A \rangle = \frac{d}{dt} \langle \Psi | A | \Psi \rangle = \left( \frac{d}{dt} \langle \Psi | A | \Psi \rangle \right) + \langle \Psi | \left( \frac{\partial A}{\partial t} \right) | \Psi \rangle + \langle \Psi | A \left( \frac{d}{dt} | \Psi \rangle \right)$$

$$= i \left[ \langle \Psi | H(t) | A \rangle | \Psi \rangle - \frac{i}{\hbar} \langle \Psi | A H(t) | \Psi \rangle + \langle \Psi | \left( \frac{\partial A}{\partial t} \right) | \Psi \rangle \right],$$

$$\frac{d}{dt} \langle A \rangle = i \left[ \left[ H(t), A \right] \right] \langle \Psi | \frac{\partial A}{\partial t} | \Psi \rangle. \quad (4.15)$$

In particular, for any observable $A$ that is time independent and commutes with the Hamiltonian, $\langle A \rangle$ will not change. In particular, if the Hamiltonian itself is time independent, it will obviously commute with itself, and we have

$$\frac{d}{dt} \langle H \rangle = 0.$$

Let’s apply this in a particular case. Suppose that we are dealing with a Hamiltonian of the form

$$H = \frac{p^2}{2m} + V(R).$$

We would like to know how the expectation value of the six operators $R$ and $P$ change as a function of time. We have

$$\frac{d}{dt} \langle R \rangle = \frac{i}{\hbar} \left\langle \left[ \frac{p^2}{2m} + V(R), R \right] \right\rangle = \frac{i}{2m\hbar} \left\langle \frac{p^2}{2m} R \right\rangle = \frac{i}{2m\hbar} \left\langle -2i \hbar P \right\rangle = \frac{1}{m} \langle P \rangle, \quad (4.16)$$

where we have used (3.48) to simplify the commutator. For momentum we have

$$\frac{d}{dt} \langle P \rangle = i \left\langle \left[ \frac{p^2}{2m} + V(R), P \right] \right\rangle = \frac{i}{\hbar} \left\langle V(R), P \right\rangle. \quad (4.17)$$

To work out this commutator, it is easiest to imagine it acting on an arbitrary wave function:

$$\left[ V(R), P \right] \psi(r) = V(R)P \psi(r) - PV(R)\psi(r) = -i\hbar \left( V(r) \nabla \psi(r) - \psi(r) \nabla V(r) \right)$$

$$= i\hbar \left[ \nabla V(R) \right] \psi(r).$$

Substituting this into (4.17), and putting it together with (4.16), we have Ehrenfest’s Theorem:

$$\frac{d}{dt} \langle R \rangle = \frac{1}{m} \langle P \rangle$$

and

$$\frac{d}{dt} \langle P \rangle = -\langle \nabla V(R) \rangle. \quad (4.18)$$
These equations are easy to understand classically. The first says that the rate of change of position is momentum over mass, which makes perfect sense. The second says that the rate of change of momentum is equal to minus the gradient of the potential, which is the force. However, in quantum mechanics, a particle does not have a definite position, or a definite momentum, so we have to take expectation values of the position, momentum, and force.

Let’s solve a simple problem using Ehrenfest’s theorem. Suppose we have a particle in a one-dimensional harmonic oscillator, with energy given by

\[ E = \frac{p^2}{2m} + \frac{1}{2} kx^2. \]

We change this to the quantum Hamiltonian

\[ H = \frac{p^2}{2m} + \frac{1}{2} kX^2. \]

How will the expectation value of the position and momentum change over time? Using (4.18a) and (4.18b), these will evolve according to

\[ \frac{d}{dt} \langle X \rangle = \frac{1}{m} \langle P \rangle \quad \text{and} \quad \frac{d}{dt} \langle P \rangle = -k \langle X \rangle. \]  

Substituting the first equation into the second yields

\[ \frac{d^2}{dt^2} \langle X \rangle = -\frac{k}{m} \langle X \rangle. \]

This is identical with the classical equation of motion, and has general solution

\[ \langle X \rangle = A \cos(\omega t + \phi), \]

where \( A \) and \( \phi \) are arbitrary constants and \( \omega = \sqrt{k/m} \) is the classical angular frequency. Hence in this case we were able to solve (4.19) in a straightforward manner. This situation is unusual, however, because we normally do not have \( \langle \nabla V(R) \rangle = \nabla \langle V(R) \rangle \), which made this easy.

G. The Time-Independent Schrödinger Equation

In chapter two we made considerable progress whenever the Hamiltonian was independent of time. Can we do so in general? In this case, Schrödinger’s equation is

\[ i\hbar \frac{d}{dt} \left| \Psi(t) \right\rangle = H \left| \Psi(t) \right\rangle. \]

We conjecture that there will be solutions of the form

---

1 Please note that problems like this would commonly be given simply as \( H = \frac{p^2}{2m} + \frac{1}{2} kx^2 \). Depending on the notation of the author, it must be understood by the reader that \( p \) and \( x \) are to be promoted to operators (which we call \( P \) and \( X \)), but many authors incautiously do not distinguish between operators and their corresponding classical quantities. I will attempt to avoid this confusion, at the risk of making readers confused when they confront such equations from other sources.
\[
\left| \Psi(t) \right\rangle = \left| \psi \right\rangle \phi(t). \tag{4.21}
\]

Substituting this into (4.20), dividing by \( \phi(t) \), and rearranging, we find
\[
H \left| \psi \right\rangle = \left| \psi \right\rangle \left[ i\hbar \frac{1}{\phi(t)} \frac{d}{dt} \phi(t) \right].
\]

We therefore see that \( \left| \psi \right\rangle \) is an eigenvector of \( H \), and since \( H \) is Hermitian, it will have real eigenvalues. Calling the state \( \left| \psi_n \right\rangle \) and the corresponding eigenvalue \( E_n \), which we identify with the energy, we have
\[
H \left| \psi_n \right\rangle = E_n \left| \psi_n \right\rangle, \tag{4.22}
\]
the \textit{time-independent Schrödinger Equation}, and
\[
i\hbar \frac{1}{\phi(t)} \frac{d}{dt} \phi(t) = E_n.
\]

Up to a constant of integration, which ultimately becomes a multiplicative constant, we have \( \phi(t) = e^{-iE_nt/\hbar} \). We substitute this into (4.21) and then take linear combinations of those solutions to find the most general solution of (4.20):
\[
\left| \Psi(t) \right\rangle = \sum_n c_n \left| \psi_n \right\rangle e^{-iE_nt/\hbar}. \tag{4.23}
\]

Because we can choose the eigenstates of \( H \) to form a complete, orthonormal basis, we choose to do so, so we have
\[
\left\langle \psi_n \right| \left| \psi_m \right\rangle = \delta_{nm}. \tag{4.24}
\]

Then if we know the wave function at a particular time, say at \( t = 0 \), then we can find the constants \( c_n \) by letting \( \left\langle \psi_n \right| \) act on (4.23) which yields
\[
c_n = \left\langle \psi_n \right| \left| \Psi(t = 0) \right\rangle. \tag{4.25}
\]

Formally, we now have a complete solution of any time-independent problem. We find the eigenstates and eigenvalues of \( H \) defined by (4.22), we choose them to be orthonormal as specified in (4.24), then for an arbitrary initial state we use (4.25) to get the constants \( c_n \), and then we have the general solution (4.23).

As an illustration, consider a particle of mass \( m \) in a one-dimensional square well. The eigenstates were already found in chapter 2 section E, where we found orthonormal eigenstates and eigenvalues of the Hamiltonian given by (2.37), repeated here for clarity:
\[
\left\langle x \right| \psi_n \right\rangle = \begin{cases} \sqrt{2/a} \sin \left( \frac{\pi nx}{a} \right) & \text{for } 0 < x < a, \\ 0 & \text{otherwise}, \end{cases} \quad \text{and } E_n = \frac{\pi^2 \hbar^2 n^2}{2ma^2}. \tag{4.26}
\]

Suppose the initial wave function is given by
\[ \langle x | \Psi(t = 0) \rangle = \begin{cases} \sqrt{\frac{3}{a^3}} (a - |a - 2x|) & \text{for } 0 < x < a, \\ 0 & \text{otherwise.} \end{cases} \quad (4.27) \]

This function is illustrated in Fig. 4-1. It is straightforward to verify that this wave function is normalized. We now find the constants \( c_n \) with the help of (4.25), so we have

\[
c_n = \langle \Psi(t = 0) | \psi_n \rangle = \int_0^a \psi_n^*(x) \Psi(x, 0) \, dx = \frac{\sqrt{6}}{a^2} \int_0^a \left(a - |a - 2x|\right) \sin \left(\frac{n\pi x}{a}\right) \, dx. \]

For those who enjoy doing such integrals by hand, I will leave you to it, but this is a bit easier to do with the help of Maple.

> assume(n::integer, a > 0);
> sqrt(6)/a^2*integrate((a-abs(a-2*x))*sin(Pi*n*x/a), x=0..a);

\[
c_n = \frac{4\sqrt{6} \sin \left(\frac{1}{2} \pi n\right)}{\pi^2 n^2}. \quad (4.28)\]

It is easy to show that \( \sin \left(\frac{1}{2} \pi n\right) \) vanishes for even \( n \) and alternates with values \( \pm 1 \) for odd \( n \). We can then substitute (4.26) and (4.28) into (4.25) to get the wave function at all times in the allowed region.

\[
\langle x | \Psi(t) \rangle = \sum_{n \text{ odd}} (-1)^{(n-1)/2} \frac{8\sqrt{3}}{\pi^2 n^2} \frac{\sin \left(\frac{n\pi x}{a}\right)}{\sqrt{a}} \exp \left(-i \frac{\pi^2 hn^2}{2ma^2} t\right). \quad (4.29)
\]

It may be a bit disappointing that (4.29) is not in closed form (since it is an infinite sum), but as a practical matter, a computer can numerically sum (4.29) quite quickly. The sum would be faster still if we had picked a smoother function at \( t = 0 \).

One small comment seems in order. In classical physics the absolute energy of a system cannot be determined, only differences in energy. Does this apply in quantum physics as well? Suppose we have two Hamiltonians that differ by a constant, so

\[
H' = H + \Delta E. \quad (4.30)
\]

The two Hamiltonians will have identical eigenvectors, but their eigenvalues will differ by the constant \( \Delta E \), so \( E'_n = E_n + \Delta E \). For an arbitrary initial state, the constants \( c_n \) as given by (4.25) will be identical, but the time evolution (4.23) will be slightly different:

\[
|\Psi'(t)\rangle = \sum_n c_n |\psi_n\rangle e^{-iE_n t/h} = \sum_n c_n |\psi_n\rangle e^{-iE_n t/h} e^{-i\Delta E t/h} = |\Psi(t)\rangle e^{-i\Delta E t/h}.
\]

Thus the wave functions of the two Hamiltonians will differ at any time only by a phase. As we have described before, such a phase difference makes no difference physically, so in fact a constant added to the Hamiltonian has no detectable effect. The two Hamiltonians in (4.30) should be treated as identical.
Problems for Chapter 4

1. It is sometimes said that a watched kettle never boils. In some sense, this is true in quantum mechanics. Consider a quantum system where the state space is two dimensional, with basis states $\{|0\rangle, |1\rangle\}$, the former representing the kettle in the “not boiled” state, the latter the “boiled” state. In terms of these, the Hamiltonian is given by

$$H = \hbar \omega \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$ 

(a) At $t = 0$, the quantum state is given by $|\Psi(t = 0)\rangle = |0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Solve Schrödinger’s equation given the initial conditions, and determine the ket of the kettle at later times.

(b) At time $t = \Delta t$, an observer checks whether the kettle has boiled yet. That is, he measures the quantum system using the boiled operator $B$, defined by $B|0\rangle = 0|0\rangle$ and $B|1\rangle = 1|1\rangle$. What is the probability $P_1$ that the kettle has boiled at this time (i.e., that measuring $B$ yields the eigenvalue 1)? If the kettle is boiled at this time, the total time $T = \Delta t$ is recorded.

(c) If the kettle is not boiled, what is the quantum state immediately after the measurement has been made?

(d) After a second interval of $\Delta t$, the kettle is measured again to see if it has boiled. What is the probability $P_2$ that it is not boiled the first time, and it is boiled the second? If this occurs, the total time $T = 2\Delta t$ is recorded.

(e) The process is repeated until the kettle has actually boiled. What is the general formula $P_n$ that it first boils on the $n$'th measurement? Write a formula for the average time $\langle T \rangle = \langle n\Delta t \rangle$ that it takes for the kettle to boil. The formula below may be helpful.

(f) Demonstrate that in the limit $\Delta t \to 0$, it takes forever for the kettle to boil.

(g) Determine (numerically or otherwise) the optimal time $\Delta t$ so that $\langle T \rangle = \langle n\Delta t \rangle$ will be minimized, so the kettle boils as quickly as possible.

Helpful formula: $1 + 2x + 3x^2 + 4x^3 + \cdots = \frac{d}{dx} \left(1 + x + x^2 + x^3 + x^4 + \cdots \right) = \frac{1}{(1-x)^2}$.

2. The commutation relations of the angular momentum operator $L_z$ with the momentum operator $P$ were to be worked out in chapter 3, problem 3.

(a) Using these commutation relations, derive two non-trivial uncertainty relationships.

(b) Show that if you are in an eigenstate of any observable, the uncertainty in that observable is zero.

(c) Show that if you are in an eigenstate of $L_z$ then you must have $\langle P_x \rangle = \langle P_y \rangle = 0$. 

© 2011, Eric D. Carlson
3. We will eventually discover that particles have spin, which is described by three operators \( S = (S_x, S_y, S_z) \), with commutation relations

\[
\begin{align*}
[S_x, S_y] &= i\hbar S_z, \\
[S_y, S_z] &= i\hbar S_x, \\
[S_z, S_x] &= i\hbar S_y.
\end{align*}
\]

A particle in a magnetic field of magnitude \( B \) pointing in the \( z \)-direction will have Hamiltonian \( H = -\mu BS_z \) where \( \mu \) and \( B \) are constants.

(a) Derive formulas for the time derivative of all three components of \( \langle S \rangle \).

(b) At time \( t = 0 \), the expectation values of \( \langle S \rangle \) are given by

\[
\langle S_x \rangle_{t=0} = a, \quad \langle S_y \rangle_{t=0} = 0, \quad \langle S_z \rangle_{t=0} = b.
\]

Determine the expectation value \( \langle S \rangle \) at later times.

4. A particle of mass \( m \) lies in a one-dimensional infinite square well in the region \([0, a]\).

At \( t = 0 \), the wave function in the allowed region is

\[
\langle x \mid \Psi (t = 0) \rangle = \sqrt{30/\alpha^5} \left( ax - x^2 \right).
\]

(a) Write the wave function in the form \( \mid \Psi (t = 0) \rangle = \sum_n c_n \phi_n \) where \( \phi_n \) are the energy eigenstates. Check that the wave function is properly normalized, both in the original coordinate basis, and in the new basis, either analytically or numerically.

(b) Find the expectation value of the Hamiltonain \( H \) for this wave function at \( t = 0 \), both in the original coordinate basis and the eigenstate basis. Check that they are the same, either analytically or numerically.

(c) Write the wave function \( \mid \Psi (t) \rangle \) at all times.
V. The Harmonic Oscillator

It has been said that physics is a series of courses on the harmonic oscillator, and quantum mechanics is no exception. Not only is it one of the few quantum mechanics problems we can solve exactly, it has great practical application. For example, consider any one-dimensional quantum mechanical system. In the neighborhood of any point \( x_0 \), the potential \( V(x) \) can be expanded in a Taylor series:

\[
V(x) \approx V(x_0) + V'(x_0)(x-x_0) + \frac{1}{2} V''(x_0)(x-x_0)^2 + \frac{1}{6} V'''(x_0)(x-x_0)^3 + \cdots.
\]

Often it is the case that a particle will classically be trapped in the neighborhood of the minimum of the potential, and if this is true quantum mechanically as well, we can choose \( x_0 \) to be this minimum point. In this case, the first derivative of \( V \) will vanish, and expanding to only the second derivative, we have

\[
V(x) \approx V(x_0) + \frac{1}{2} V''(x_0)(x-x_0)^2.
\]

As discussed in chapter four, the constant term plays no physical role, and hence the only relevant term is the quadratic one. We can redefine our origin to make \( x_0 = 0 \), as we will do below.

Not only is the harmonic oscillator useful in one dimension, but the coupled harmonic oscillator can be used in three dimensions, or for multiple particles in three dimensions, or even for an infinite number of degrees of freedom, as we will ultimately discover when we quantize the electromagnetic field. Thus the harmonic oscillator is also the starting point for quantum field theory. But for now, let us start with the simplest case, a single particle in one dimension.

A. The 1D Harmonic Oscillator

The harmonic oscillator Hamiltonian can be written in the form

\[
H = \frac{1}{2m} P^2 + \frac{1}{2} kX^2,
\]

where \( m \) is the mass of the particle and \( k \) is the spring constant. In analogy with classical mechanics, or inspired by section 4F, it makes sense to define the classical angular frequency \( \omega = \sqrt{k/m} \), in terms of which

\[
H = \frac{1}{2m} P^2 + \frac{1}{2} \omega^2 X^2.
\]  

We now wish to find the eigenstates and eigenvalues of this Hamiltonian.

Examining (5.1), we immediately note that at large values of the eigenvalues of \( X \), the potential becomes infinite, and for any finite energy \( E \), it follows that \( E < V(\pm\infty) \). This guarantees that we will only be talking about bound states.
As a practical matter, when faced with a difficult problem it is often helpful to “scale out” all the dimensionful variables in the problem. In this problem we have the constants \( m \), \( \omega \), and once we start using quantum mechanics, \( \hbar \). Since \( X \) has units of distance it is helpful to define a new operator \( \hat{X} \) that is proportional to \( X \) but has been divided by a constant with units of length. It is not hard to see that \( \hbar/m\omega \) has units of length squared, so \( X\sqrt{m\omega/\hbar} \) will be dimensionless. Similarly, we can show that \( \hbar m\omega \) has units of momentum squared, so \( P\sqrt{m\omega/\hbar} \) will also be dimensionless. Beyond this a bit of inspiration is in order. We define the lowering and raising operators as

\[
a = X \sqrt{\frac{m\omega}{2\hbar}} + i \frac{P}{\sqrt{2m\omega\hbar}}, \quad a' = X \sqrt{\frac{m\omega}{2\hbar}} - i \frac{P}{\sqrt{2m\omega\hbar}}.
\] (5.2)

We will often want the inverse versions of these relations, which are

\[
X = \sqrt{\frac{\hbar}{2m\omega}} (a + a'), \quad P = i \sqrt{\frac{\hbar m\omega}{2}} (a' - a).
\] (5.3)

These two new operators have commutation relations

\[
[a, a^\dagger] = \left[ X \sqrt{\frac{m\omega}{2\hbar}} + i \frac{P}{\sqrt{2m\omega\hbar}}, X \sqrt{\frac{m\omega}{2\hbar}} - i \frac{P}{\sqrt{2m\omega\hbar}} \right] = \sqrt{\frac{m\omega}{2\hbar}} \left[ [X, -iP] + [iP, X] \right] = \frac{(-i)(i\hbar) + i(-i\hbar)}{2\hbar} = 1,
\]

or, summarizing,

\[
[a, a^\dagger] = 1, \quad [a', a^\dagger] = [a, a] = 0.
\] (5.4)

Substituting (5.3) into our Hamiltonian (5.1) we find

\[
H = \frac{1}{2m} \frac{\hbar m\omega}{2} \left[ (a' - a)^2 + \frac{\hbar}{2m\omega} \right] + \frac{\hbar}{2m\omega} \frac{\hbar}{2m\omega} \left( a + a' \right)^2 = \frac{\hbar}{4} \hbar \omega \left[ (a' + a)^2 - (a' - a)^2 \right]
\]

\[
= \frac{1}{2} \hbar \omega (a^2 a + aa') = \frac{1}{2} \hbar \omega \left( [a, a^\dagger] + 2aa' \right),
\]

\[
H = \hbar \omega \left( a^2 a + \frac{1}{2} \right).
\] (5.5)

Finding the eigenstates and eigenvectors of \( H \) has been reduced to finding those of the Hermitian operator \( a'a \).

Suppose we find a normalized eigenstate of \( a'a \), which we name \( |n\rangle \), so that

\[
a^\dagger a |n\rangle = n |n\rangle.
\]

Now, consider the state \( a |n\rangle \). It is an eigenstate of \( a'a \) also:

\[
a^\dagger a (a |n\rangle) = \left( [a', a'] + aa' \right) a |n\rangle = (-a + aa'a) |n\rangle = (n - 1) (a |n\rangle).
\]
Its eigenvalue is one less than that of \( |n\rangle \), which is why \( a \) is called the lowering operator. It is not a normalized eigenstate, because
\[
\|a|n\rangle\|^2 = \langle n | a^+ a | n \rangle = n \langle n | n \rangle = n. 
\] (5.6)
We note immediately from (5.6) that we must have \( n \geq 0 \). For \( n > 0 \), we then identify a new normalized eigenstate of \( a^+ a \) given by
\[
|n-1\rangle = \frac{1}{\sqrt{n}} a |n\rangle. 
\] (5.7)
The procedure (5.7) can be repeated, yielding new states \( |n-2\rangle, |n-3\rangle \), etc. However, it cannot yield states with negative values of \( n \), because of (5.6). It follows that the procedure must eventually stop, which can only happen when we get an eigenstate \( n = 0 \). This occurs only if we start with a non-negative integer.

With the help of (5.7), we can similarly show that
\[
a^+ |n-1\rangle = \frac{1}{\sqrt{n}} a^+ a |n\rangle = \sqrt{n} |n\rangle. 
\] (5.8)
Summarizing (5.7) and (5.8), we have
\[
a|n\rangle = \sqrt{n} |n-1\rangle, \quad a^+ |n\rangle = \sqrt{n+1} |n+1\rangle. 
\] (5.9)
Thus we can raise or lower from any eigenstate to any other using (5.9) repeatedly. For example, from the lowest energy state \( n = 0 \), we can find any other by using
\[
|n\rangle = \frac{1}{\sqrt{n!}} (a^+)^n |0\rangle. 
\] (5.10)
The energy eigenvalues are given by (5.5), and we have
\[
E_n = h \omega (n + \frac{1}{2}). 
\] (5.11)
The lowering operator must vanish when acting on the ground state \( n = 0 \), so
\[
a |0\rangle = 0. 
\] (5.12)

If we actually want a coordinate representation of the wave function, it is easiest to first find \( \psi_0(x) \equiv \langle x | 0 \rangle \). Writing (5.12) out explicitly, we have
\[
\left( x \sqrt{\frac{m \omega}{2 \hbar}} + i \frac{P}{\sqrt{2 m \omega \hbar}} \right) |0\rangle = 0, 
\]
\[
\left( x \sqrt{\frac{m \omega}{2 \hbar}} + i \frac{h d}{\sqrt{2 m \omega \hbar}} \right) \psi_0(x) = 0, 
\]
\[
\frac{d}{dx} \psi_0(x) = -\frac{m \omega}{\hbar} x \psi_0(x),
\]
\[
\frac{d\psi_0(x)}{\psi_0(x)} = -\frac{m\omega}{\hbar} x \, dx.
\]

Integrating both sides and then exponentiating both sides of this equation, we find

\[
\psi_0(x) = C \exp\left(-\frac{m\omega x^2}{2\hbar}\right).
\]

The constant \(C\) came from the constant of integration. Since we want our eigenstates to be normalized, we have

\[
1 = \int_{-\infty}^{\infty} |\psi_0(x)|^2 \, dx
\]

\[
= |C|^2 \int_{-\infty}^{\infty} \exp\left(-\frac{m\omega x^2}{\hbar}\right) \, dx
\]

\[
= |C|^2 \frac{\pi \hbar}{m\omega}.
\]

If we choose \(C\) real and positive then

\[
\psi_0(x) = \left(\frac{m\omega}{\pi \hbar}\right)^{1/4} \exp\left(-\frac{m\omega x^2}{2\hbar}\right).
\]

We can then use (5.10) to find any of the other wave functions explicitly.

\[
\psi_n(x) = \frac{1}{\sqrt{n!}} \left(\frac{m\omega}{\pi \hbar}\right)^{1/4} \left(x \sqrt{\frac{m\omega}{2\hbar}} - \frac{\hbar}{2m\omega} \frac{d}{dx}\right)^n \exp\left(-\frac{m\omega x^2}{2\hbar}\right).
\]  \hspace{1cm} (5.13)

Explicit graphs of the first few wave functions given by (5.13) are shown in Fig. 5-1. The general trend is that as \(n\) increases, the wave functions spread out farther and farther from the origin. Notice also that the wave functions are even functions when \(n\) is even, and odd when \(n\) is odd.

**B. Working with the Harmonic Oscillator and Coherent States**

Having found the eigenstates of the harmonic oscillator Hamiltonian, it is straightforward to find a general solution to Schrödinger’s equation. Using (4.23) and (5.11), we see that the general solution is

\[
|\Psi(t)\rangle = \sum_n c_n |n\rangle e^{-i(n+\frac{1}{2})\omega t}.
\]  \hspace{1cm} (5.14)

If we want expectation values of operators such as \(P\) or \(X\) for such a state, we could, in principle, find explicit expressions for the wave functions using (5.13) and then compute them in a straightforward way. Though this will work, it is tedious and difficult. It is far
easier to use the expressions such as (5.3) and (5.9) together with orthonormality to quickly evaluate the expectation values of operators. For example, suppose we are in an eigenstate of the Hamiltonian and we were asked to compute the expectation values of $P$ and $P^2$. We would simply compute:

$$\langle n | P | n \rangle = i \sqrt{\frac{\hbar}{2}} \hbar \omega \langle n | (a^+ - a) | n \rangle = i \sqrt{\frac{\hbar}{2}} \hbar \omega \langle n | (\sqrt{n+1} | n+1 \rangle - \sqrt{n} | n-1 \rangle \rangle$$

$$= i \sqrt{\frac{\hbar}{2}} \hbar \omega (0 - 0) = 0,$$

$$\langle n | P^2 | n \rangle = \| P | n \| ^2 = \| i \sqrt{\frac{\hbar}{2}} \hbar \omega (a^+ - a) | n \rangle \| ^2 = \frac{\hbar}{2} \hbar \omega \| \sqrt{n+1} | n+1 \rangle - \sqrt{n} | n-1 \rangle \| ^2$$

$$= \frac{\hbar}{2} \hbar \omega (n + 1 + n) = \hbar \omega (n + \frac{1}{2}).$$

Interestingly, note that the expectation value for the momentum vanishes for every eigenstate. Also note that the kinetic energy term in the Hamiltonian, $P^2/2m$, will have an expectation value of $\frac{\hbar}{2} \hbar \omega (n + \frac{1}{2})$, exactly half of the energy $E_n$ of this state. The other half of the energy is in the potential term.

One surprise is that $\langle P \rangle$ vanishes for all these states; it is easily demonstrated that $\langle X \rangle$ vanishes as well. This is true for energy eigenstates (also called stationary states) even when you include the time-dependence (5.14). This is surprising because we might expect that when we put a lot of energy into the system it would behave more like a classical harmonic oscillator. Furthermore, for the energy eigenstates it is easy to show that the uncertainty in the position and momentum grow as $n$ grows, so the wave is in no sense located at a particular place or a particular momentum.

The key, of course, is to not use stationary states, but instead include states that have mixtures of different eigenvectors. The simplest example would be

$$| \Psi (t) \rangle = \frac{1}{\sqrt{2}} \{ | n - 1 \rangle e^{-i(n-\frac{1}{2}) \omega t} + | n \rangle e^{-i(n+\frac{1}{2}) \omega t} \} = \frac{1}{\sqrt{2}} e^{-i \omega t} \{ | n - 1 \rangle e^{i \omega t/2} + | n \rangle e^{-i \omega t/2} \}.$$  

This is an exact solution of Schrödinger’s time-dependent equation. The expectation value of $P$, for example, will be

$$\langle \Psi (t) | P | \Psi (t) \rangle = \frac{1}{2} e^{i \omega t} \left( \langle n-1 | e^{i \omega t/2} + \langle n | e^{-i \omega t/2} \rangle P (| n-1 \rangle e^{-i \omega t/2} + | n \rangle e^{i \omega t/2}) \right)$$

$$= \frac{1}{2} i \sqrt{\frac{\hbar}{2}} \hbar \omega \langle n-1 | e^{-i \omega t/2} + \langle n | e^{i \omega t/2} \rangle$$

$$\cdot \left[ \left( \sqrt{n} | n \rangle - \sqrt{n-1} | n-2 \rangle \right) e^{i \omega t/2} + \left( \sqrt{n+1} | n+1 \rangle - \sqrt{n} | n-1 \rangle \right) e^{-i \omega t/2} \right]$$

$$= \frac{1}{2} i \sqrt{\frac{\hbar}{2}} \hbar \omega \left( \sqrt{n} e^{i \omega t} - \sqrt{n} e^{-i \omega t} \right) = \frac{1}{2} i \sqrt{\frac{\hbar}{2}} \hbar \omega \left[ 2i \sin (\omega t) \right]$$

$$= -\sqrt{\frac{\hbar}{2}} \hbar \omega \sin (\omega t).$$

It can be shown that $\langle X \rangle$ oscillates as well. This bears a greater resemblance to a classical harmonic oscillator, though the uncertainties of the two operators $P$ and $X$ can be shown to still be unacceptably large.
You can do far better if, instead of working with just one or two eigenstates of the Hamiltonian, you work with many, even an infinite number. Perhaps most interesting is to work with coherent states, quantum states \( |z\rangle \) for any complex number \( z \) defined by

\[
|z\rangle = e^{-\frac{\hbar}{2} I} \sum_{n=0}^{\infty} \frac{1}{\sqrt{n!}} z^n |n\rangle.
\]

These states are defined to make them eigenstates of the annihilation operator \( a \):

\[
a|z\rangle = e^{-\frac{\hbar}{2} I} \sum_{n=0}^{\infty} \frac{1}{\sqrt{n!}} z^n a |n\rangle = e^{-\frac{\hbar}{2} I} \sum_{n=0}^{\infty} \frac{1}{\sqrt{n!}} z^n \sqrt{n} |n-1\rangle = e^{-\frac{\hbar}{2} I} \sum_{n=0}^{\infty} \frac{1}{\sqrt{(n-1)!}} z^n |n\rangle.
\]

\[
e^{-\frac{\hbar}{2} I} \sum_{n=0}^{\infty} \frac{1}{\sqrt{n!}} z^{n+1} |n\rangle = z \left( e^{-\frac{\hbar}{2} I} \sum_{n=0}^{\infty} \frac{1}{\sqrt{n!}} z^n |n\rangle \right),
\]

\[
a|z\rangle = z |z\rangle.
\]

(5.15)

Since \( a \) is not a Hermitian operator the eigenstates \( |z\rangle \) will not generally have real eigenvalues, nor will they be orthogonal. It is not hard to show they are normal, \( \langle z | z \rangle = 1 \).

We will also find useful the Hermitian conjugate of (5.15), which is

\[
\langle z | a^* \rangle = \langle z | z^* \rangle.
\]

(5.16)

Let’s suppose that at \( t = 0 \) we are in a coherent state, so

\[
|\Psi(t = 0)\rangle = |z_0\rangle = e^{-\frac{\hbar}{2} I} \sum_{n=0}^{\infty} \frac{1}{\sqrt{n!}} z^n_0 |n\rangle.
\]

We know that (5.14) gives the general solution of Schrödinger’s equation, so

\[
|\Psi(t)\rangle = e^{-\frac{\hbar}{2} I} \sum_{n=0}^{\infty} \frac{1}{\sqrt{n!}} z^n_0 |n\rangle \exp(-i\omega(n + \frac{1}{2}) t) = e^{-i\omega t/2} e^{-\frac{\hbar}{2} I} \sum_{n=0}^{\infty} \frac{1}{\sqrt{n!}} (z_0 e^{-i\omega t})^n |n\rangle.
\]

Since \( |z_0 e^{-i\omega t}| = |z_0| \), this state is also a coherent state, apart from the phase factor out front, so we have

\[
|\Psi(t)\rangle = e^{-i\omega t/2} |z_0 e^{-i\omega t}\rangle.
\]

It can be shown that this wave function has a much smaller uncertainty in its position or momentum than previous wave functions we have considered. It is also easy to see using (5.15) and (5.16) that expectation values such as the momentum and position oscillate as we would expect. For example,

\[
\langle \Psi(t) | P | \Psi(t) \rangle = e^{i\omega t/2} \left\langle z_0 e^{-i\omega t} | P | z_0 e^{-i\omega t} \right\rangle e^{-i\omega t/2} = i \sqrt{\frac{\hbar m}{2}} \omega \left\langle z_0 e^{-i\omega t} \right\rangle \left( \langle a^* - a\rangle \right) |z_0 e^{-i\omega t}\rangle
\]

\[
= i \sqrt{\frac{\hbar m}{2}} \omega \left( z_0 e^{-i\omega t} \right) \left[ \left( z_0 e^{-i\omega t} \right)^* - \left( z_0 e^{-i\omega t} \right) \right] |z_0 e^{-i\omega t}\rangle
\]

\[
= i \sqrt{\frac{\hbar m}{2}} \omega (2i) \text{Im} \left( z_0 e^{-i\omega t} \right) = \sqrt{2\hbar m \omega} \text{Im} \left( z_0 e^{-i\omega t} \right).
\]

We cannot simplify this further, since \( z_0 \) might have real and imaginary parts.
C. Multiple Particles and Harmonic Oscillators

We now wish to consider more complicated systems of multiple harmonic oscillators, but first we must consider what our vector space will look like with multiple particles. A fuller discussion of this topic will come in a later chapter, but for now, suppose we have multiple particles in one dimension. According to our postulates, the state of the quantum system at fixed time must be described by a single vector in our vector space. Instead of describing such a system as a function of a single position $\psi(x)$, it must be a function of all the particle’s positions, $\psi(x_1, x_2, \ldots, x_N)$, where $N$ is the number of particles. The classical description of the energy of the particles, $E = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + V(x_1, x_2, \ldots, x_N)$, will then become the quantum Hamiltonian

$$H = \sum_{i=1}^{N} \frac{1}{2m_i} p_i^2 + V(X_1, X_2, \ldots, X_N).$$

The position and momentum operators are most easily defined by their action on a wave function:

$$X_i \psi(x_1, x_2, \ldots, x_N) = x_i \psi(x_1, x_2, \ldots, x_N),$$

$$P_i \psi(x_1, x_2, \ldots, x_N) = -i \hbar \frac{\partial}{\partial x_i} \psi(x_1, x_2, \ldots, x_N).$$

These will have commutation relations

$$[X_i, P_j] = i \hbar \delta_{ij}, \quad [X_i, X_j] = [P_i, P_j] = 0.$$

We are now prepared to write down the problem of multiple uncoupled harmonic oscillators. To simplify, assume all the particles have the same mass, so then we have

$$H = \sum_{i=1}^{N} \frac{1}{2m} p_i^2 + \frac{1}{2} \sum_{i=1}^{N} k_i x_i^2. \quad (5.17)$$

We will solve this by the same techniques we did the single harmonic oscillator. First we define $\omega_i = \sqrt{k_i / m}$, and then define a set of raising and lowering operators by

$$a_i = X_i \sqrt{\frac{m \omega_i}{2 \hbar}}, \quad a_i^\dagger = X_i \sqrt{\frac{m \omega_i}{2 \hbar}}, \quad a_i^\dagger = X_i \sqrt{\frac{m \omega_i}{2 \hbar}} - i \frac{p_i}{\sqrt{2 m \omega_i \hbar}}. \quad (5.18)$$

These have commutation relations

$$[a_i, a_j^\dagger] = \delta_{ij}, \quad [a_i^\dagger, a_j] = [a_i, a_j] = 0.$$

The Hamiltonian in terms of these is given by

$$H = \sum_{i=1}^{N} \hbar \omega_i \left( a_i^\dagger a_i + \frac{1}{2} \right). \quad (5.19)$$
The operators \( a_i^\dagger a_i \) commute with each other, and hence they can all be diagonalized simultaneously. Hence we can choose our basis states to be simultaneously eigenstates of all of the \( a_i^\dagger a_i \)'s with eigenvalues \( n_i \), and we write a general basis state as

\[
|n_1, n_2, \ldots, n_N\rangle. \tag{5.20}
\]

We can now go through the exact same arguments as before to prove that each of the \( n_i \)'s will be non-negative integers. We can also show that the action of a single raising or lowering operator on (5.20) is to increase or decrease one of the eigenvalues by one, so we have

\[
a_i |n_1, n_2, \ldots, n_N\rangle = \sqrt{n_i} |n_1, n_2, \ldots, n_i - 1, \ldots, n_N\rangle,
\]

\[
a_i^\dagger |n_1, n_2, \ldots, n_N\rangle = \sqrt{n_i + 1} |n_1, n_2, \ldots, n_i + 1, \ldots, n_N\rangle.
\]

The eigenvalues of the Hamiltonian are simply given by

\[
E_{n_1, n_2, \ldots, n_N} = \sum_{i=1}^{N} \hbar \omega_i (n_i + \frac{1}{2}). \tag{5.21}
\]

With considerable effort we can write down explicit wave functions starting from the demand that \( a_i |0, \ldots, 0\rangle = 0 \). For the ground state,

\[
\psi_{0, \ldots, 0}(x_1, \ldots, x_N) = \prod_{i=1}^{N} \left( \frac{m \omega_i}{\pi \hbar} \right)^{1/4} \exp \left( -\frac{m \omega_i x_i^2}{2 \hbar} \right). \tag{5.22}
\]

Other explicit wave functions can be found by acting on (5.22) with the various raising operators. In fact, such wave functions are so unwieldy to work with that we will avoid them whenever possible, simply writing our quantum states in the form \( |n_1, n_2, \ldots, n_N\rangle \).

We have succeeded in solving the uncoupled harmonic oscillator. What do we do if our harmonic oscillator is coupled? Consider the Hamiltonian

\[
H = \sum_{i=1}^{N} \frac{1}{2m} p_i^2 + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} k_{ij} x_i x_j. \tag{5.23}
\]

We can always write the coupling terms \( k_{ij} \) in such a way that it is symmetric, so \( k_{ij} = k_{ji} \). How would we solve this problem if we were doing the problem classically? The answer is to change coordinates to new coordinates such that the interactions become diagonal.

Think of the spring constants \( k_{ij} \) as an \( N \times N \) matrix \( K \), defined by

\[
K = \begin{pmatrix}
k_{11} & k_{12} & \cdots & k_{1N} \\
k_{21} & k_{22} & \cdots & k_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
k_{N1} & k_{N2} & \cdots & k_{NN}
\end{pmatrix}.
\]

Since it is real and symmetric, it is Hermitian. As discussed in chapter 3, section I, any such matrix can be diagonalized by multiplying by an appropriate unitary matrix \( V \) on the right and its Hermitian conjugate \( V^\dagger \) on the left. In this case, however, since our matrix

© 2011, Eric D. Carlson
is real, the eigenvectors of $K$ can be chosen real, so that $V$ will be a real matrix, so we have $V^\dagger = V^T$, and with the appropriate choice of $V$,

$$K' = V^T K V = \begin{pmatrix} k'_1 & 0 & \cdots & 0 \\ 0 & k'_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & k'_N \end{pmatrix},$$

where the numbers $k'_1, k'_2, \ldots, k'_N$ are the eigenvalues of $K$. The matrix $V$ is unitary, or, since it is real, orthogonal, which means $V^T V = V V^T = I$.

Now, define new operators $X'_i$ and $P'_i$ related to the old ones by

$$X'_i = \sum_{j=1}^N V_{ij} X_j', \quad P'_i = \sum_{j=1}^N V_{ij} P_j'.$$

(5.24)

Since the inverse of $V$ is $V^T$ we can easily find the inverse relations:

$$X'_i = \sum_{j=1}^N V_{ji} X_j, \quad P'_i = \sum_{j=1}^N V_{ji} P_j.$$

Using (5.24) we can easily work out the commutators of the new operators

$$[X'_i, P'_j] = \left[\sum_{k=1}^N V_{ki} X_k, \sum_{l=1}^N V_{lj} P_l\right] = \sum_{k=1}^N \sum_{l=1}^N V_{ki} V_{lj} [X_k, P_l] = i\hbar \sum_{k=1}^N \sum_{l=1}^N V_{ki} V_{lj} \delta_{kl} = i\hbar \sum_{k=1}^N V_{ki} V_{kj}$$

$$= i\hbar (V^T V)_{ij} = i\hbar \delta_{ij},$$

$$[X'_i, X'_j] = [P'_i, P'_j] = 0.$$

So the new operators have the same commutation relations as the old ones. We now rewrite the Hamiltonian (5.23) in terms of these operators:

$$H = \frac{1}{2m} \sum_{i=1}^N \sum_{k=1}^N \sum_{l=1}^N V_{ki} V_{kl} P'_i P'_l + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N \sum_{l=1}^N k'_i V_{ki} V_{jl} X'_i X'_j$$

$$= \frac{1}{2m} \sum_{k=1}^N \sum_{i=1}^N \sum_{j=1}^N P'_i P'_j \left(V^T V\right)_{ij} + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N X'_i X'_j \sum_{k=1}^N \sum_{l=1}^N \left(V^T V\right)_{kl}$$

$$= \frac{1}{2m} \sum_{k=1}^N \sum_{i=1}^N \sum_{j=1}^N P'_i P'_j \delta_{kl} + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N X'_i X'_j k'_i \delta_{kl} = \frac{1}{2} \sum_{i=1}^N P'_i^2 + \frac{1}{2} \sum_{i=1}^N k'_i X'_i^2.$$

This reduces the Hamiltonian to the uncoupled one, (5.17). The Hamiltonian ultimately takes the form (5.19) and the energy values (5.21), where

$$\omega_i = \sqrt{k'_i/m}.$$

Note that to find expressions for the frequencies, and hence the Hamiltonian or the energies, all we need is the eigenvalues $k'_i$ of the matrix $k$. It is not necessary to actually work out the matrices $V$.  

80  © 2011, Eric D. Carlson
D. The Complex Harmonic Oscillator

It is sometimes helpful to adopt a complex notation when describing two (or more) harmonic oscillators. Let’s back up and explain this in classical terms. Suppose we had a classical energy for a complex position \( z \) given by

\[
E = m \left( \dot{z}^2 + \omega^2 z^2 \right). \tag{5.25}
\]

where \( \dot{z} \) just denotes the time derivative of \( z \). There is nothing particularly unusual about this formula, you just need to think of \( z \) as a way of combining two real variables, \( x \) and \( y \), into a single complex variable according to

\[
z = \frac{1}{\sqrt{2}} (x + iy).
\]

Substituting this into (5.25) we see that the energy can be written in terms of the real positions as

\[
E = \frac{1}{2} m \left( x^2 + y^2 \right) + \frac{1}{2} m \omega^2 \left( x^2 + y^2 \right) = \frac{1}{2m} \left( p_x^2 + p_y^2 \right) + \frac{1}{2} m \omega^2 \left( x^2 + y^2 \right). \tag{5.26}
\]

We then replace \( E \) by the Hamiltonian and the four quantities \( p_x, p_y, x \) and \( y \) by their corresponding operators to obtain

\[
H = \frac{1}{2m} \left( p_x^2 + p_y^2 \right) + \frac{1}{2} m \omega^2 \left( x^2 + y^2 \right). \tag{5.27}
\]

We have, of course, already solved this equation in section C. We define two lowering operators \( a_x \) and \( a_y \), with commutation relations

\[
[a_x, a_y^\dagger] = [a_y, a_x^\dagger] = 1, \quad \text{all others vanish.}
\]

We then write the Hamiltonian using (5.19).

\[
H = \hbar \omega \left( a_x^\dagger a_x + a_y^\dagger a_y + 1 \right). \tag{5.28}
\]

We now define two new creation and annihilation operators as

\[
an_x = \frac{1}{\sqrt{2}} (a_x \pm i a_y), \quad an_y = \frac{1}{\sqrt{2}} (a_y \pm i a_x^\dagger).
\]

These operators can be shown to have commutation relations

\[
[a_x, a_y^\dagger] = [a_y, a_x^\dagger] = 1, \quad \text{all others vanish} \tag{5.29}
\]

The Hamiltonian (5.27) can, with a bit of work, be rewritten in terms of these operators

\[
H = \hbar \omega \left[ \frac{1}{2} \left( a_x^\dagger + a_x \right) (a_x + a_x^\dagger) + \frac{1}{2} \left( a_y^\dagger - a_y \right) (a_y - a_y^\dagger) + 1 \right],
\]

\[
H = \hbar \omega \left( a_x^\dagger a_x + a_y^\dagger a_y + 1 \right). \tag{5.29}
\]

We will later need to reinterpret \( z \) and \( \dot{z} \) as operators, and we will want to do so in terms of our new raising and lowering operators \( a_x \) and \( a_y\). We make the quantum identification
\[ z = \frac{1}{\sqrt{2}}(x + iy) \quad \rightarrow \quad \frac{X + iY}{\sqrt{2}} = \frac{1}{\sqrt{2}} \sqrt{\frac{\hbar}{2m\omega}} (a_x^+ + a_y^+ + ia_y - ia_x) = \sqrt{\frac{\hbar}{2m\omega}} (a_x + a_y^+). \]

A similar computation will show

\[ \dot{z} = \frac{1}{\sqrt{2}}(\dot{x} + i\dot{y}) = \frac{p_x + i p_y}{m\sqrt{2}} \quad \rightarrow \quad \frac{i}{m\sqrt{2}} \sqrt{\frac{\hbar \omega m}{2}} (a_x^+ - a_x + ia_x - ia_x^+) = i \sqrt{\frac{\hbar \omega}{2m}} (a_x^+ - a_x). \]

Not surprisingly, if you take the complex conjugate of a classical quantity, it gets mapped to the Hermitian conjugate when you make the quantum identification, so we have

\[
\begin{align*}
    z &\rightarrow \sqrt{\frac{\hbar}{2m\omega}} (a_x + a_x^+), & z^* &\rightarrow \sqrt{\frac{\hbar}{2m\omega}} (a_x - a_x^+), \\
    \dot{z} &\rightarrow i \sqrt{\frac{\hbar \omega}{2m}} (a_x^+ - a_x), & \dot{z}^* &\rightarrow i \sqrt{\frac{\hbar \omega}{2m}} (a_x - a_x^+).
\end{align*}
\] (5.30)

Thus, if we are faced with an energy expression of the form (5.25), we will immediately convert it to the form (5.29), assume the commutation relations (5.28) apply, and for any quantities where we need to make quantum operators out of classical quantities, we will use the conversion formulas (5.30). This will prove very handy when we quantize the electromagnetic field in a later chapter.
Problems for Chapter 5

1. The Lennard-Jones 6-12 potential is commonly used as a model to describe the potential of an atom in the neighborhood of another atom. Classically, the energy is given by \( E = \frac{1}{2} m x^2 + 4\epsilon \left( \frac{\sigma}{x} \right)^{12} - \left( \frac{\sigma}{x} \right)^{6} \).

   (a) Find the minimum \( x_{\text{min}} \) of this potential, and expand the potential to quadratic order in \( (x - x_{\text{min}}) \).

   (b) Determine the classical frequency \( \omega \) and calculate the quantum mechanical minimum energy, as a function of the parameters.

2. At \( t = 0 \) a single particle is placed in a harmonic oscillator \( H = \frac{P^2}{2m} + \frac{1}{2} m \omega^2 X^2 \) in the superposition state \( |\Psi(t=0)\rangle = \frac{1}{\sqrt{2}} |1\rangle + \frac{1}{\sqrt{2}} |2\rangle \), that is, in a superposition of the first and second excited states.

   (a) What is the wave function \( |\Psi(t)\rangle \) at subsequent times?

   (b) What are the expectation values \( \langle X \rangle \) and \( \langle P \rangle \) at all times?

3. In class we assumed that the coupled harmonic oscillators all had the same mass. Consider now the case where the oscillators have different masses, so that \( H = \sum_i \frac{P_i^2}{2m_i} + \frac{1}{2} \sum_{ij} k_{ij} X_i X_j \).

   (a) Rescale the variables \( P_i \) and \( X_i \) to new variables \( \hat{P}_i \) and \( \hat{X}_i \) with the usual commutation relations: \( \left[ \hat{X}_i, \hat{P}_j \right] = i\hbar \delta_{ij} \) such that in terms of the new variables, the Hamiltonian is given by \( H = \sum_i \frac{\hat{P}_i^2}{2m} + \frac{1}{2} \sum_{ij} k'_{ij} \hat{X}_i \hat{X}_j \), where \( m \) is an arbitrary mass that you can choose.

   (b) Find an expression for \( k'_{ij} \) in terms of the original variables \( k_{ij} \) and \( m_i \), and explain in words how to obtain the eigenvalues of the original Hamiltonian.

   (c) A system of two particles in one dimension has Hamiltonian

   \[ H = \frac{P_x^2}{2m} + \frac{P_z^2}{2(m/4)} + \frac{1}{2} m \omega^2 \left( 5X_1^2 + 2X_2^2 + 2X_1X_2 \right). \]

   Find the eigenvalues \( E_{ij} \) of this Hamiltonian.

4. A particle of mass \( m \) is in a one-dimensional harmonic oscillator with angular frequency \( \omega \). If the particle is in the coherent state \( |z\rangle \), find the uncertainties \( \Delta X \) and \( \Delta P \), and check that they satisfy the uncertainty relation \( \Delta X \Delta P \geq \frac{1}{2} \hbar \).
VI. Symmetry

A. The Translation Operator

In the previous chapter, we solved the Harmonic oscillator. Let’s consider the simplest of all possible variations on this problem, adding a linear term to the Hamiltonian.

\[ H' = \frac{1}{2m} P^2 + \frac{1}{2} kX^2 + bX. \]  

(6.1)

It is easy to see that this Hamiltonian can be rewritten as

\[ H' = \frac{1}{2m} P^2 + \frac{1}{2} k \left( X + \frac{b}{k} \right)^2 - \frac{b^2}{2k}. \]  

(6.2)

It is intuitively obvious that this is just a Harmonic oscillator again, except the energies have been shifted by an amount \(-\frac{b^2}{2k}\), and the center of symmetry has been shifted to \(x = -\frac{b}{k}\). It would be nice to have a notation that can take this into account.

Let us define, in one dimension, the translation operator \(T(a)\) as

\[ T(a) = x + a. \]  

(6.3)

Taking the Hermitian conjugate of (6.3) tells us that

\[ T(a) = x + a. \]  

(6.4)

Combining (6.3) and (6.4), we see that

\[ \langle x | T^\dagger (a) | x' \rangle = \langle x + a | x' + a \rangle = \delta (x + a - x' - a) = \delta (x - x') = \langle x | x' \rangle. \]

Since the states \( |x\rangle \) form a complete basis, it follows that \(T(a)\) is unitary,

\[ T^\dagger (a) T(a) = 1. \]  

(6.5)

Some other useful identities will come in hand later. From (6.3), it is easily proven that

\[ T(0) = 1, \]  

(6.6a)

\[ T(a) T(b) = T(a + b), \]  

(6.6b)

\[ T(-a) = T^{-1}(a) = T^\dagger (a). \]  

(6.6c)

The final of this trio of identities follows from (6.6a) and (6.5). The effect on wave functions of the operator \(T(a)\), for example, can then easily be worked out

\[ \langle x | T(a) | \psi \rangle = \langle x | T^\dagger (-a) | \psi \rangle = \langle x - a | \psi \rangle = \psi (x - a). \]
A little thought will convince you that the effect of \( T(a) \) is to shift the wave function to the right by an amount \( +a \), as illustrated in Fig. 6-1.

It is similarly easy to show the “effect” of the translation operator on the position and momentum operators.

\[
\langle x \mid PT(a) \mid \psi \rangle = -i\hbar \frac{d}{dx} \langle x \mid T(a) \mid \psi \rangle = -i\hbar \frac{d}{dx} \psi(x-a) = -i\hbar \psi'(x-a) = -i\hbar \langle x-a \mid \psi' \rangle
\]

\[
= \langle x-a \mid P \mid \psi \rangle = \langle x \mid T(a)P \mid \psi \rangle,
\]

\[
XT(a)\langle x \rangle = X\langle x+a \rangle = (x+a)\langle x+a \rangle = (x+a)T(a)\langle x \rangle = T(a)\langle X+a \rangle \langle x \rangle.
\]

It is clear that momentum commutes with the translation operator, but position does not.

\[
PT(a) = T(a)P,
\]

\[
XT(a) = T(a)(X+a).
\]

We can multiply these on the left by \( T^\dagger(a) \) to rewrite these as

\[
T^\dagger(a)PT(a) = P, \tag{6.7a}
\]

\[
T^\dagger(a)XT(a) = X+a. \tag{6.7b}
\]

The effect of more complicated combinations of \( X \) and \( P \) are easily worked out using these two expressions; for example,

\[
T^\dagger(a)X^2T(a) = T^\dagger(a)XT(a)T^\dagger(a)XT(a) = (X+a)^2,
\]

or to generalize

\[
T^\dagger(a)V(X)T(a) = V(X+a).
\]

Now that we’ve defined and worked out the properties of the translation operator, we can use it to quickly find eigenstates of our Hamiltonian (6.1) (which is the same as (6.2)). Define the alternative Hamiltonian

\[
H = T^\dagger(-b/k)HT(-b/k) = \frac{1}{2m}P^2 + \frac{1}{2}kX^2 - \frac{b^2}{2k}.
\]

We already know everything about this Hamiltonian; in particular, it has energy eigenstates \( |n \rangle \) satisfying \( H |n \rangle = E_n |n \rangle \), where

\[
E_n = \hbar \sqrt{k/m} \left(n + \frac{1}{2}\right) - b^2/2k. \tag{6.8}
\]
Now consider the expression
\[ H' \left[ T \left( -b/k \right) |n \rangle \right] = T \left( -b/k \right) H' \left( -b/k \right) T \left( -b/k \right) |n \rangle = T \left( -b/k \right) |H |n \rangle \]
\[ = E_n \left[ T \left( -b/k \right) |n \rangle \right]. \]

Hence we see that the eigenstates of \( H' \) are given by
\[ |n' \rangle = T \left( -b/k \right) |n \rangle, \]
and the energies are given by (6.8). Furthermore, since the translation operator is unitary, the eigenstates (6.9) will automatically be normalized. In the coordinate basis, the wave functions will just be given by
\[ \langle x | n' \rangle = \langle x | T \left( -b/k \right) |n \rangle = \langle x + b/k |n \rangle = \psi_n (x + b/k). \]

Hence we see that we can immediately find the eigenstates and eigenvalues of any Hamiltonian that is related to another by translation.

The generalization to three dimensions is straightforward. We can translate in any direction by any amount, so we have translation operators \( T (a) \) which act on the position and momentum operators according to
\[ T^\dagger (a) P T (a) = P \quad \text{and} \quad T^\dagger (a) R T (a) = R + a. \]

The translation operators are still unitary, and the combinations of two translations is still a translation, so \( T^\dagger (a) T (a) = 1 \) and \( T (a) T (b) = T (a + b) \).

B. Rotation Operators

It is similarly helpful to define the idea of rotation. Rotations are linear transformations of the coordinates \( r \). Thinking of \( r \) as a column matrix, we have
\[ r \rightarrow \mathcal{R} r, \]
where (in three dimensions) \( \mathcal{R} \) is a \( 3 \times 3 \) real matrix. It is, however, not an arbitrary matrix, but rather one which preserves the distance, so we want
\[ r^T r = r^2 = (\mathcal{R} r)^2 = (\mathcal{R} r)^T (\mathcal{R} r) = r^T \mathcal{R}^T \mathcal{R} r. \]

The only way we can assure this is satisfied is if we demand that
\[ \mathcal{R}^T \mathcal{R} = 1, \quad \text{(6.10)} \]
i.e., that \( \mathcal{R} \) is an orthogonal matrix, which is the same as saying it is real and unitary.

Taking the determinant of (6.10) and recalling that the determinant of the transpose of a matrix is identical to the determinant of the matrix, we see that
\[ \left[ \det (\mathcal{R}) \right]^2 = 1, \quad \text{so} \quad \det (\mathcal{R}) = \pm 1. \quad \text{(6.11)} \]
We can divide all possible rotations into two categories. *Proper* rotations have \( \det(\mathcal{R}) = 1 \), and can actually be physically performed, such as rotation about an axis. *Improper* rotations have \( \det(\mathcal{R}) = -1 \) and cannot be physically performed, such as reflection in a mirror. We will focus for the moment on proper rotations, those that can be performed.

In one dimension, proper rotations are impossible, because there aren’t enough directions to rotate around. In two dimensions, there is only one direction you can perform a rotation, so all rotations can be described by a single parameter \( \theta \). The rotation matrix takes the form

\[
\mathcal{R}(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.
\]

In three dimensions all possible proper rotations are described in terms of an axis of rotation and an angle, so we write \( \mathcal{R} = \mathcal{R}(\hat{r}, \theta) \).\(^1\) For example, if we pick one of the three coordinate axes, the corresponding formulas for rotations are

\[
\begin{align*}
\mathcal{R}(\hat{x}, \theta) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix}, \\
\mathcal{R}(\hat{y}, \theta) &= \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix}, \\
\mathcal{R}(\hat{z}, \theta) &= \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.
\end{align*}
\] (6.12)

We expect that there should be an operator, which we call the rotation operator, corresponding to rotations, which we define as

\[
\mathcal{R}(\mathcal{R})|r\rangle \equiv |\mathcal{R}r\rangle.
\] (6.13)

The following identities are not difficult to prove and will be useful:

\[
\begin{align*}
R(1) &= 1, \\
R(\mathcal{R}_1)R(\mathcal{R}_2) &= R(\mathcal{R}_1\mathcal{R}_2), \\
\langle r | R^\dagger(\mathcal{R}) &= \langle\mathcal{R}r |, \\
R^\dagger(\mathcal{R})R(\mathcal{R}) &= 1, \\
R^\dagger(\mathcal{R}) &= R(\mathcal{R}^\dagger).
\end{align*}
\] (6.14a-d)

For example, (6.14d) can be proven from (6.13) and (6.14c) directly, with the help of (A.41b) and (6.11), but it is easier to simply think of it intuitively: rotating a wave function must not change its norm. Indeed, this will be a general property we demand of any operator that implements a symmetry: it must always be unitary.

\(^1\) In more than three dimensions, rotations are harder to describe. We won’t go there.
Using (6.14c) and (6.14e) with (6.13), it is not hard to see what effect rotations have on a wave function:

$$\langle r | R(\mathcal{R}) | \psi \rangle = \langle r | R^T(\mathcal{R}^T) | \psi \rangle = \langle \mathcal{R}^T r | \psi \rangle = \psi (\mathcal{R}^T r).$$

It is also not too hard to figure out its effect on the position operator. We have

$$\mathcal{R} R(\mathcal{R}) | r \rangle = R | \mathcal{R} r \rangle = (\mathcal{R} r) | \mathcal{R} r \rangle = \mathcal{R} R(\mathcal{R}) | r \rangle = R(\mathcal{R}) \mathcal{R} R | r \rangle.$$  

Multiplying this on the left by $R^T(\mathcal{R})$, this implies

$$R^T(\mathcal{R}) \mathcal{R} R(\mathcal{R}) = \mathcal{R} R.$$

For momentum, it is a bit harder. A good starting point is to figure out what happens to the momentum eigenstates $| k \rangle$ under an arbitrary rotation. It is not hard to see that

$$\langle r | R(\mathcal{R}) | k \rangle = \langle r | R^T(\mathcal{R}^T) | k \rangle = \langle \mathcal{R}^T r | k \rangle = (2\pi)^{-3/2} \exp\left[i (\mathcal{R}^T r) \cdot k\right] = (2\pi)^{-3/2} \exp\left[i r \cdot (\mathcal{R} k)\right] = \langle r | \mathcal{R} k \rangle,$$

so

$$R(\mathcal{R}) | k \rangle = | \mathcal{R} k \rangle.$$

Then we find

$$\mathcal{P} R(\mathcal{R}) | k \rangle = \mathcal{P} | \mathcal{R} k \rangle = (\hbar \mathcal{R} k) | \mathcal{R} k \rangle = (\hbar \mathcal{R} k) R(\mathcal{R}) | k \rangle = R(\mathcal{R}) \mathcal{R} \mathcal{P} | k \rangle.$$  

We therefore have $\mathcal{P} R(\mathcal{R}) = R(\mathcal{R}) \mathcal{R} \mathcal{P}$, or multiplying by $R^T(\mathcal{R})$ on the left,

$$R^T(\mathcal{R}) \mathcal{P} R(\mathcal{R}) = \mathcal{R} \mathcal{P}.$$  

Indeed, relations (6.15) or (6.16) might be thought of as the definition of a vector operator: if it acts like $\mathcal{P}$ or $\mathcal{R}$ when you perform a proper rotation, then it is a vector.

Just as translations can be used to quickly solve problems that are related by a translation, we can use rotation to relate similar problems. As an example, imagine a particle in two dimensions whose Hamiltonian is given by

$$H = \frac{\mathcal{P}^2}{2m} + \mathcal{V}(\mathcal{R}), \quad \mathcal{V}(x, y) = \begin{cases} 0 & \text{if } |x| + |y| < a, \\ \infty & \text{otherwise}. \end{cases}$$

The region defined by this potential is sketched in Fig. 6-2.

It is obvious that this problem can be simplified if we rotate it by 45 degrees, so we let $\mathcal{R} = \mathcal{R}(45^\circ)$, and define a new Hamiltonian

$$H' = R^T(\mathcal{R}) H R(\mathcal{R}) = \frac{1}{2m} R^T(\mathcal{R}) \mathcal{P}^2 R(\mathcal{R}) + R^T(\mathcal{R}) \mathcal{V}(\mathcal{R}) R(\mathcal{R}) = \frac{1}{2m} (\mathcal{R} \mathcal{P})^2 + \mathcal{V}(\mathcal{R} \mathcal{R}).$$
It isn’t hard to see that \((\mathcal{R} \mathbf{P})^2 = \mathbf{P}^2\), and with some work we can find the new Hamiltonian:

\[
H' = \frac{1}{2m} \mathbf{P}^2 + V' (\mathbf{R}), \quad V'(x, y) = \begin{cases} 
0 & \text{if } |x|, |y| < a/\sqrt{2}, \\
\infty & \text{otherwise}.
\end{cases}
\]

This is now just a 2d infinite square well. If we want to shift it to make the problem easier, we can define

\[
H'' = T^\dagger \left( -\frac{1}{\mathcal{R}} a (\hat{x} + \hat{y}) \right) H' T \left( -\frac{1}{\mathcal{R}} a (\hat{x} + \hat{y}) \right) = \frac{1}{2m} \mathbf{P}^2 + V' (\mathbf{R} - \frac{1}{\mathcal{R}} a (\hat{x} + \hat{y})),
\]

\[
H'' = \frac{1}{2m} \mathbf{P}^2 + V'' (\mathbf{R}), \quad V''(x, y) = \begin{cases} 
0 & \text{if } 0 < x, y < a/\sqrt{2}, \\
\infty & \text{otherwise}.
\end{cases}
\]

This problem is easily solved by the separation of variables. The resulting wave functions and energies are

\[
\langle x, y \mid \psi_{n,p}'' \rangle = \frac{\sqrt{2}}{a} \sin \left( \frac{\pi nx}{a\sqrt{2}} \right) \sin \left( \frac{\pi py}{a\sqrt{2}} \right), \quad E_{n,p} = \frac{\pi^2 \hbar^2}{4ma^2} \left( n^2 + p^2 \right).
\]

We can then find the wave functions for the original problem simply by performing the two transformations; first, the translation to get it back into the correct form for \(H'\), then the rotation for \(H\), so we have

\[
\ket{\psi_{n,p}} = R (\mathcal{R}) T \left( -\frac{1}{\mathcal{R}} a (\hat{x} + \hat{y}) \right) \ket{\psi_{n,p}''}.
\]

The energies are the same for \(H\) and \(H''\).

Up to now we have restricted ourselves to proper rotations; it is now time to briefly consider improper rotations. The simplest example of this is the parity transformation, defined by\(^1\)

\[
\Pi \ket{\mathbf{r}} \equiv R (-1) \ket{\mathbf{r}} = \ket{-\mathbf{r}}.
\]

Applying (6.15) and (6.16) to this operator, we see that

\[
\Pi \mathbf{R} \Pi^\dagger = -\mathbf{R}, \quad \Pi \mathbf{P} \Pi^\dagger = -\mathbf{P}.
\]

You might think this is true of all vector operators, but that is not quite the case. Consider in particular the angular momentum operator \(\mathbf{L} = \mathbf{R} \times \mathbf{P}\). Then we have

\[
\Pi \mathbf{L} \Pi^\dagger = \Pi \mathbf{R} \times \Pi \mathbf{P} = \Pi \mathbf{R} \Pi^\dagger \times \Pi \mathbf{P} \Pi^\dagger = (-\mathbf{R}) \times (-\mathbf{P}) = \mathbf{L}.
\]

So angular momentum actually commutes with parity. We sometimes refer to any vector operator satisfying equations like (6.18) as a \textit{true vector} or just \textit{vector} operator, and anything satisfying (6.19) as a \textit{pseudovector} operator.

\(^1\) Note that this is an improper rotation only in an odd number of dimensions. In even number of dimensions, (6.17) describes a proper rotation. Under such circumstances, one normally would redefine parity to be any specific improper rotation, and many of the subsequent formulas will not work.
C. Parity and the Finite Square Well

Look for a moment at the eigenstate wave functions illustrated in Fig. 5-1. It is clear that for \( n = 0 \) and \( 2 \), the wave function is even, \( \psi_n(-x) = \psi_n(x) \), while for \( n = 1 \) and \( 3 \), the wave function is odd, \( \psi_n(-x) = -\psi_n(x) \). These results are exact, and persist to higher \( n \) values. Is this a coincidence, having to do with the simple nature of the harmonic oscillator problem? It turns out it is not. It comes about because the Hamiltonian commutes with the parity operator.

To demonstrate this, consider

\[
\Pi H \Pi^\dagger = \frac{1}{2m} \Pi P^2 \Pi^\dagger + \frac{1}{2} m \omega^2 \Pi L \Pi^\dagger L \Pi^\dagger
\]

Multiplying this by \( \Pi \) on the right, this becomes \( \Pi H = H \Pi \), so they commute. In general, when two operators commute, they can be simultaneously diagonalized. Thus our eigenstates \( |n\rangle \) can be chosen (and were chosen) such that

\[
H |n\rangle = E_n |n\rangle, \quad \Pi |n\rangle = \lambda_n |n\rangle.
\]

This property can be generalized to any Hamiltonian that commutes with parity. If the Hamiltonian takes the standard form (3.38), then we will find that

\[
\Pi H \Pi^\dagger = \frac{1}{2m} P^2 + V(-R) = H \quad \text{if and only if} \quad V(-R) = V(R).
\]

Whenever this is satisfied we can choose the eigenstates of \( H \) to also be eigenstates of \( \Pi \).

What are the possible eigenvalues of \( \Pi \)? If we have an eigenstate \( |\psi\rangle \) with eigenvalue \( \lambda \), then we have

\[
\Pi |\psi\rangle = \lambda |\psi\rangle, \quad \langle \psi | \Pi^\dagger = \langle \psi | \lambda^*.
\]

Multiplying these and remembering that \( \Pi \) is unitary, we have

\[
\langle \psi | \psi \rangle = \langle \psi | \Pi^\dagger \Pi |\psi\rangle = \langle \psi | \lambda^* \lambda |\psi\rangle = \langle \psi | |\lambda|^2 \rangle.
\]

or, more simply,

\[
|\lambda| = 1 \quad \Leftrightarrow \quad \lambda = e^{i\theta} \quad (6.20)
\]

Hence the eigenvalue must be a complex number of magnitude one, a pure phase. Indeed, the arguments leading to (6.20) apply to any symmetry, not just parity. However, in the case of parity, we have an additional constraint, that parity taken twice yields the identity, so we have
\[ |\psi\rangle = \Pi^2 |\psi\rangle = \lambda^2 |\psi\rangle, \]

from which we conclude
\[ \lambda = \pm 1. \quad (6.21) \]

Hence if we choose eigenstates of parity, they must be either even or odd, exactly what we found before.

The restriction (6.21) does not necessarily apply if there is a symmetry other than parity. Let \( S \) be an arbitrary symmetry operator (translation, rotation, or even parity). \( S \) need not be Hermitian, but it is required to be unitary. If \( S \) commutes with the Hamiltonian, then they can be simultaneously diagonalized. Furthermore, a simple proof will demonstrate that two states with different eigenvalues will necessarily be orthogonal.

Let \( |\psi_1\rangle, |\psi_2\rangle \) be two eigenstates of \( S \) with distinct eigenvalues \( \lambda_1, \lambda_2 \). Both numbers will be of magnitude 1. The inner product of these two kets will be
\[ \langle \psi_1 | \psi_2 \rangle = \langle \psi_1 | S^\dagger S | \psi_2 \rangle = \lambda_1^* \lambda_2 \langle \psi_1 | \psi_2 \rangle = e^{i\theta} \langle \psi_1 | \psi_2 \rangle. \]

which can only be satisfied if \( \lambda_1 = \lambda_2 \) (which is not allowed) or \( \langle \psi_1 | \psi_2 \rangle = 0 \). Hence we can choose orthonormal basis states that are eigenstates of both \( H \) and \( S \).

Let’s do a sample problem and see how symmetry helps us solve it. Consider the 1D finite square well, which has Hamiltonian

\[ H = \frac{1}{2m} \hat{p}^2 + V(X), \quad V(X) = \begin{cases} 0 & \text{if } |X| < a, \\ V_0 & \text{if } |X| > a. \end{cases} \]

This potential is sketched in Fig. 6-3. Our goal is to find the bound states of the Hamiltonian. We’ll search for solutions where \( 0 < E < V_0 \).

If we divide the wave into three regions, as sketched in Fig. 6-3, then in regions I and III, we are trying to solve the equation
\[ (E - V_0) \psi(x) = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x). \]

Keeping in mind that \( E - V_0 \) is negative, the solutions of this equation are
\[ \psi(x) = e^{\pm \beta x} \quad \text{where} \quad \beta = \frac{\sqrt{2m(V_0 - E)}}{\hbar}. \quad (6.22) \]

In region II, we are trying to solve
\[ E \psi(x) = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x). \]
which has solutions\(^1\)

\[\psi(x) = \sin(kx) \quad \text{or} \quad \psi(x) = \cos(kx), \quad \text{where} \quad k = \frac{\sqrt{2mE}}{\hbar}. \quad (6.23)\]

Keeping in mind, as before, that we only want solutions that do not blow up at infinity, we therefore have wave functions:

\[\psi_I(x) = Ae^{\beta x}, \quad \psi_{II}(x) = B \cos(kx) + C \sin(kx), \quad \psi_{III}(x) = De^{-\beta x}.\]

We could now proceed as before, by matching the wave function and its derivative at both boundaries. But there is an easier way! We note that parity is a symmetry of our Hamiltonian, and therefore our eigenstates can be chosen to be eigenstates of parity as well. Hence all our solutions will be either even or odd functions.

The parity symmetry will map region I to region III, and vice versa, while mapping region II onto itself. Hence parity will relate the coefficients \(A\) and \(D\), and restrict \(B\) and/or \(C\). For example, suppose that we demand that we have an even eigenstate, so \(\psi(-x) = +\psi(x)\). This will then demand that \(A = D\), so that regions I and III will match up, and \(C = 0\), so that it will be even in region II. Now consider the boundary conditions at \(x = a\). Demanding that the two functions and their derivatives match at this boundary (since the potential is finite), we have

\[\psi_{II}(a) = \psi_{III}(a) \quad \Rightarrow \quad B \cos(ka) = D \exp(-\beta a),\]

\[\psi'_{II}(a) = \psi'_{III}(a) \quad \Rightarrow \quad -kB \sin(ka) = -\beta D \exp(-\beta a).\]

Dividing the second of these by the first, we have

\[k \tan(ka) = \beta.\]

Using equations (6.22) and (6.23), with a bit of algebra, we can rewrite this as

\[\tan(ka) = \frac{2mV_0}{\hbar^2 k^2} - 1 \quad (6.24)\]

The right hand side of (6.24) is a function that falls from infinity to 0, while \(\tan(ka)\) rises from 0 to infinity, so there must be at least one solution of this equation. Once we have \(k\), we can then use (6.23) to get the energy \(E = \hbar^2 k^2 / 2m\).

---

\(^1\) We choose the real solutions, instead of the complex solutions \(e^{\pm ikx}\), here because it is simpler, but you would be forced to switch to sines and cosines in a moment anyway.
We must now repeat the steps to find solutions for the odd functions. Skipping directly to the answer, we find

\[ -\cot (ka) = \sqrt{\frac{2mV_0}{\hbar^2 k^2} - 1}. \] (6.25)

which gives us another set of energies. Plots of the left and right sides of (6.24) and (6.25) for \( a\sqrt{2mV_0} = \frac{1}{2} \pi \hbar \) can be found in Fig. 6-4. The resulting energies are plotted in Fig. 6-5 as a function of \( V_0 \). In the limit of infinite \( V_0 \), the solutions of (6.24) and (6.25) yield \( ka = \frac{1}{2} \pi n \), the same answer we would get for an infinite square well.

D. Discrete Rotations

Let’s try applying the techniques we learned to parity in situations where there are more complicated symmetries. For example, consider the following two-dimensional Hamiltonian:

\[ H = \frac{\mathbf{p}^2}{2m} + V (\mathbf{R}), \quad V (x, y) = \begin{cases} 0 & \text{if } |x|,|y| < a, \\ \infty & \text{otherwise}. \end{cases} \] (6.26)

This square infinite square well is pretty easily solved exactly, but we would like to use parity to restrict some of the solutions. Consider rotation by \( \frac{1}{2} \pi \), given by the matrix

\[ \mathcal{R} \left( \frac{\pi}{2} \right) = \begin{pmatrix} \cos \frac{\pi}{2} & -\sin \frac{\pi}{2} \\ \sin \frac{\pi}{2} & \cos \frac{\pi}{2} \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \text{so } \mathcal{R} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix}. \] (6.27)

To check that this commutes with the Hamiltonian, all we have to check is if \( V (\mathcal{R} \mathbf{R}) = V (\mathbf{R}) \), or \( V (-Y, X) = V (X, Y) \), which is pretty easy to see. It follows that we can choose our basis states as eigenstates of \( \mathcal{R} \), so we have \( \mathcal{R} |\psi_m\rangle = \lambda_m |\psi_m\rangle \).

When we were working with parity, we pointed out that since \( \Pi^2 = 1 \), the eigenvalues of parity satisfied \( \lambda^2 = 1 \). Can any similar restriction be placed here? Noting that four rotations by \( \frac{1}{2} \pi \) constitutes a single rotation by \( 2\pi \), which is the same as no rotation, we conclude that \( \mathcal{R}^4 (\mathcal{R}) = \mathcal{R} (\mathcal{R}^4) = \mathcal{R} (1) = 1 \), and therefore

\[ |\psi_m\rangle = \left[ \mathcal{R} (\mathcal{R}) \right]^4 |\psi_m\rangle = \lambda_m^4 |\psi_m\rangle. \]"
This implies $\lambda^4_m = 1$, so $\lambda_m$ must be a complex fourth root of 1. For any integer $n$, the $n$ solutions of $z^n = 1$ are of the form $z = e^{\frac{2\pi i n}{n}}$, for $k = 0, 1, 2, \ldots, n-1$, so for $n = 4$ we have

$$\lambda = 1, e^{\pi i/2}, e^{\pi i}, e^{3\pi i/2}, \text{ or } \lambda = 1, i, -1, -i. \quad (6.28)$$

The eigenstates of (6.26), therefore, will fall into four categories, depending on which of the four eigenvalues in (6.28) occur.

The Hamiltonian (6.26) isn’t that hard to solve, and we can actually find eigenstates which satisfy the identities given above. For example, some of the eigenstates are

$$\psi_i(x, y) = \frac{1}{a} \cos (\pi x/a) \cos (\pi y/a),$$

$$\psi_{-i}(x, y) = \frac{1}{a} \sin (\pi x/a) \sin (\pi y/a),$$

$$\psi_{\pm i}(x, y) = \frac{1}{a\sqrt{2}} \left[ \cos (\pi x/2a) \sin (\pi y/a) \pm i \sin (\pi x/a) \cos (\pi y/2a) \right].$$

The energies of these four states are $\pi^2/4ma^2$, $\pi^2/ma^2$, $5\pi^2/8ma^2$, and $5\pi^2/8ma^2$ respectively. Note that the last pair of wave functions are complex conjugates of each other and their energies are the same. This is not surprising. It is obvious that the Hamiltonian (6.26) is real. If we take the complex conjugate $\psi^*_m$ of any eigenstate wave function $\psi_m$ we are assured of getting another eigenstate. Furthermore, since there is nothing “complex” about the rotation (6.27), we are assured of getting an eigenstate of the symmetry with the same energy but a different eigenvalue. In other words, if $S$ is any symmetry operator, then

$$H |\psi_m\rangle = E_m |\psi_m\rangle, \quad S |\psi_m\rangle = \lambda_m |\psi_m\rangle \quad \Rightarrow \quad H |\psi^*_m\rangle = E_m |\psi^*_m\rangle, \quad S |\psi^*_m\rangle = \lambda^*_m |\psi^*_m\rangle.$$

In summary, any quantum state with a complex eigenvalue $\lambda$ under a symmetry operator will be paired with a degenerate state with eigenvalue $\lambda^*$ under the same symmetry. This is the first example of how symmetry causes degeneracies; we will encounter others (mostly having to do with angular momentum). Energy degeneracies are a common feature of symmetries.

It should be noted that not only is rotation by $\frac{\pi}{2}$ a symmetry of (6.26), so also is rotation by $\pi$ and $\frac{3\pi}{2}$. All of these rotations commute, so it is possible to simultaneously diagonalize $H$, $R(\mathcal{R}(\frac{\pi}{2}))$, $R(\mathcal{R}(\pi))$ and $R(\mathcal{R}(\frac{3\pi}{2}))$. Do we gain anything by labeling our states by their eigenvalues as well? No; for example,

$$R(\mathcal{R}(\pi)) |\psi_m\rangle = R \left[ \mathcal{R} \left( \frac{\pi}{2} \right) \right]^2 |\psi_m\rangle = \left[ R \left( \mathcal{R} \left( \frac{3\pi}{2} \right) \right) \right]^2 |\psi_m\rangle = \lambda^2_m |\psi_m\rangle.$$

Hence the eigenvalue under this larger rotation is defined in terms of its eigenvalue under the smaller rotation. In contrast, the eigenvalue under a $\frac{\pi}{2}$ rotation cannot be uniquely deduced from the eigenvalue under a $\pi$ rotation (because every number has two complex
square roots). Hence when labeling states, we should describe how it transforms under
the smallest possible non-zero rotation.

There are additional, improper rotation symmetries of (6.26); for example,
reflection across the $y$-axis. This reflection symmetry, however, does not commute with
the $\frac{1}{2}\pi$ rotation symmetry, and hence they cannot be simultaneously diagonalized.

Studying combinations of rotations (proper and improper) in three dimensions can be
quite tricky and interesting, but we will skip such a discussion in this text, except in the
case of complete spherical symmetry in three dimensions, which will be discussed in
detail in the next chapter.

E. Discrete Translations and the Kronig-Penney Model

Consider now a Hamiltonian that is invariant under translation. In one dimension,
this means

$$T(a)HT^\dagger(a) = H.$$  \hfill (6.29)

Then we can diagonalize $H$ and $T(a)$ simultaneously. Our eigenstates of $H$ will have
eigenvalue $T(a)|\psi\rangle = e^{i\theta}|\psi\rangle$. Unlike the case of finite rotations, there is no restriction on
this phase, because we don’t get back to the original state after a finite number of
translations. Although there is nothing wrong with labeling them this way, traditionally
we let $ka = -\theta$, and call the combination $\hbar k$ the crystal momentum. It follows that

$$\psi(x-a) = \langle x-a|\psi\rangle = \langle x|T(a)|\psi\rangle = \langle x|e^{-ika}|\psi\rangle = e^{-ika}\psi(x).$$  \hfill (6.30)

By letting $x \to x + a$, it is easy to show that a similar relation occurs when we replace $a$
by its negative, and by using (6.30) repeatedly we can generalize it to the statement that
for any integer $n$,

$$\psi(x + na) = e^{i\hbar an}\psi(x).$$  \hfill (6.31)

Equation (6.31) is very useful in practice because it means that the wave function needs
only to be calculated in a single region of the periodic potential, say $[0,a)$, and it will be
known everywhere. It should also be recognized that since (6.31) is unchanged if we add
$2\pi$ to $ka$, the quantity $ka$ is ambiguous by this amount. Traditionally this ambiguity is
avoided by demanding $-\pi < ka \leq \pi$.

Periodic potentials have great practical value, because solid state physics consists
almost entirely of the study of crystals, consisting of sextillions of identical units repeated
in a three-dimensional periodic structure. Crystals are not truly periodic, because they
have boundaries, but it is reasonable to expect that in the interior of bulk substances
boundary effects will be small.

To illustrate some of the generic features of such periodic structures, consider the
Kronig Penney model,\textsuperscript{1} which we will define in terms of the Hamiltonian

\textsuperscript{1}The Kronig-Penney model is more general than the version here. It can have attractive delta functions or
barriers of finite thickness and height.
\[ H = \frac{p^2}{2m} + \sum_{n=-\infty}^{\infty} \lambda \delta (X - na), \]

where we will choose \( \lambda > 0 \). We have illustrated the potential in Fig. 6-6. There are an infinite number of regions, but fortunately they are all related by the translation symmetry.

Let’s find the energy eigenstates of the Kronig-Penney model. Let the energy be \( E \), then in the interior of region I there is no potential and the solutions will take the form
\[
\psi_I(x) = Ae^{i\beta x} + Be^{-i\beta x}, \tag{6.32}
\]
where \( E = \hbar^2 \beta^2 / 2m \). In region II the solutions will take the same form, but we can simply use the crystal momentum to see that
\[
\psi_{II}(x + a) = Ae^{i\beta x + i\kappa a} + Be^{i\kappa a - i\beta x}. \tag{6.33}
\]

We now need to work on boundary conditions. There are boundary conditions at each delta function, but they are all related by the translation symmetry, so we pick the boundary \( x = a \). Schrödinger’s equation written out is
\[
E \psi(x) = -\frac{\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} + \sum_{n=-\infty}^{\infty} \lambda \delta (x - na) \psi(x)
\]
Integrating this across the boundary at \( x = a \), only the \( n = 1 \) contribution to the potential will contribute, and we find
\[
0 = -\frac{\hbar^2}{2m} \left[ \psi_{II}'(a) - \psi_I'(a) \right] + \lambda \psi(a) \tag{6.34}
\]
Since the first derivative has a finite discontinuity, the wave function must be continuous, so we have
\[
\psi_{II}(a) = \psi_I(a). \tag{6.35}
\]
Substituting in our explicit forms (6.32) and (6.33) into (6.35) and (6.34), we find
\[
(A + B)e^{i\kappa a} = Ae^{i\beta a} + Be^{-i\beta a},
\]
\[
\frac{1}{2} i \hbar^2 \beta (Ae^{i\kappa a} - Be^{i\kappa a} - Ae^{-i\beta a} + Be^{-i\beta a}) = \lambda m (A + B)e^{i\kappa a}.
\]
Each of these equations can be turned into an expression for the ratio \( A/B \):
\[
\frac{A}{B} = \frac{e^{i\kappa a} - e^{-i\beta a}}{e^{i\beta a} - e^{i\kappa a}} \quad \text{and} \quad \frac{A}{B} = \frac{\frac{1}{2} i \hbar^2 \beta (e^{i\kappa a} - e^{-i\beta a}) + \lambda me^{i\kappa a}}{\frac{1}{2} i \hbar^2 \beta (e^{i\kappa a} - e^{i\beta a}) - \lambda me^{i\kappa a}}.
\]
Equating and cross multiplying, this becomes
\[
\frac{1}{2}i\hbar^2 \beta (e^{2ika} - e^{i\pi - i\beta a} - e^{i\lambda a + i\beta a} + 1) - \lambda me^{2ika} + \lambda me^{i\lambda a + i\beta a} - \lambda me^{2ika}.
\]

Gathering like terms and multiplying by \(e^{-ika}\), this becomes
\[
\begin{align*}
\frac{i}{2}\hbar^2 \beta (e^{ika} - e^{-i\beta a} - e^{i\lambda a} + 1 + e^{i\lambda a}) - \lambda m e^{ika} + \lambda me^{i\lambda a} = & \ \lambda m e^{ika} - \lambda me^{-i\beta a}, \\
\hbar^2 \beta [\cos(\lambda a) - \cos(\beta a)] = & \ \lambda m \sin(\beta a), \\
\cos(ka) - \cos(\beta a) = & \ \frac{\lambda m a \sin(\beta a)}{\hbar^2}.
\end{align*}
\]

The equation (6.36) will have an infinite number of solutions for each value of \(k\), which can then be substituted into \(E = \hbar^2 \beta^2 / 2m\) to get the energies. A few limiting cases will help us qualitatively understand the result. When \(\lambda = 0\), the various solutions \(\beta, a\) will be given by the solutions of \(\cos(ka) = \cos(\beta a)\), which we label by \(n = 1, 2, 3, \ldots\), and are

\[
\beta, a = \begin{cases} 
\pi(n-1) + |ka| & \text{n odd,} \\
\pi n - |ka| & \text{n even.}
\end{cases}
\]

(6.37)

In the limit of infinite \(\lambda\), the solutions are just \(\beta, a = n\pi\), which is always higher than (6.37). For intermediate values of \(\lambda\), the energies will be somewhere in between.

A plot of the typical results can be seen in Fig. 6-6. Note that for \(\lambda = 0\) the curves are just a wrapped up parabola, as they must be for a free particle. For \(\lambda = \infty\), the energies are identical to those for the infinite square well, since the infinite barriers basically have divided the \(x\)-axis into an infinite number of separate square wells. Note that the energies are the same for \(ka\) and \(-ka\); this happens because these pairs of solutions are just complex conjugates of each other; indeed, normally a plot like Fig. 6-6 would include only half of the region and assume this symmetry.

Especially interesting is to notice that for all values of the parameter greater than zero there are regions of allowed and forbidden energy. In the
tight binding (large \( \lambda \)) limit, only discrete energy level are possible. This is comparable to a lattice of atoms where the electrons are held very tightly by each atom. In the nearly free (small \( \lambda \)) limit, the energy takes on nearly all values of energy, with only small gaps between.

F. Continuous Symmetries and Generators

Up until now, we have discussed symmetry assuming that the allowed symmetry transformations are discrete, so that only large rotations or translations are possible. Consider now the possibility of small translations; for example, suppose that in one dimension, we find that the Hamiltonian commutes with translations by any amount, so that (6.29) applies for all values of \( a \). According to the prescription given in the previous sections, we should concentrate on the smallest possible non-zero translation \( a \), but when \( a \) can take on any real value, there is no such thing as a smallest non-zero translation. What are we to do? Imagine expanding out the operator \( T(a) \) in a Taylor series for small \( a \), so that

\[
T(a) = T(0) + aT'(0) + \frac{1}{2}a^2T''(0) + \cdots. \tag{6.38}
\]

Of course, \( T(0) = 1 \), but let’s focus on the derivative. Let us define an operator \( \tilde{P} \) by

\[
\tilde{P} \equiv i\hbar T'(0) = i\hbar \frac{d}{da} T(a) \bigg|_{a=0}
\]

The reason for choosing the peculiar factors in front will be apparent momentarily. For small \( a \), we see from (6.38) that we have

\[
T(a) = 1 - \frac{ia\tilde{P}}{\hbar} + O(a^2). \tag{6.39}
\]

We want \( T(a) \) to be unitary, which to order \( a \) implies

\[
1 = T^\dagger(a)T(a) = \left[1 + \frac{ia\tilde{P}}{\hbar}\right]\left[1 - \frac{ia\tilde{P}}{\hbar}\right] + O(a^2) = 1 + \frac{ia}{\hbar}(\tilde{P}^\dagger - \tilde{P}) + O(a^2).
\]

This tells us \( \tilde{P} \) is Hermitian. And what is this operator \( \tilde{P} \)? Let (6.39) act on an arbitrary state and work in a coordinate representation, working only to order \( a \), keeping in mind that \( a \) is small.

\[
\langle x | T(a) | \psi \rangle = \langle x | \left[1 - \frac{ia\tilde{P}}{\hbar} + O(a^2)\right] | \psi \rangle,
\]

\[
\langle x - a | \psi \rangle = \langle x | \psi \rangle - \frac{ia}{\hbar} \langle x | \tilde{P} | \psi \rangle + O(a^2),
\]

\[
\psi(x-a) = \psi(x) - \frac{ia}{\hbar} \langle x | \tilde{P} | \psi \rangle + O(a^2),
\]
\[ \langle x | \hat{P} | \psi \rangle = -i \hbar \lim_{a \to 0} \frac{\psi(x) - \psi(x-a)}{a} = -i \hbar \frac{d}{dx} \psi(x) = \langle x | P | \psi \rangle. \]

In other words, \( \hat{P} \) is none other than the momentum operator! We say that the
momentum operator \( P \) is the generator of translations.

If we know how to use \( P \) to get small translations, can we get large translations as
well? A large translation is simply a combination of many small translations. Let’s work
out the effects of a large translation \( T(a) \) by dividing it into \( N \) small translations of size
\( a/N \). Since the combination of two translations is a bigger one, \( T(a) T(b) = T(a+b) \),
we have

\[
T(a) = \left[ T\left(\frac{a}{N}\right)\right]^N = \left[ 1 - \frac{iaP}{N\hbar} + O\left(\frac{a^2}{N^2}\right) \right]^N = \exp\left(-\frac{iaP}{N\hbar}\right) + O\left(\frac{a^2}{N^2}\right).
\]

If we now take the limit of \( N \) going to infinity we have

\[ T(a) = \exp(-iaP/\hbar). \] (6.40)

This equation is true whether or not translation is a symmetry of the Hamiltonian.

Not surprisingly, if we are in three dimensions, there are three generators \( P \) of
translations in each of the three dimensions. These are the three momentum operators,
and the equivalent of (6.40) is simply

\[ T(a) = \exp(-ia \cdot P/\hbar). \] (6.41)

Indeed, since the various translation operators commute with each other, it is easy to
prove that the momentum operators also commute with each other, if we didn’t already
know that.

Are there similar generators for rotations? Consider first rotations in two
dimensions, which take the form

\[
R(\theta) = \begin{pmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{pmatrix} = \begin{pmatrix} 1 & -\theta \\
\theta & 1 \end{pmatrix} + O(\theta^2).
\]

In a manner similar to (6.39), we expand the rotation operator out to leading order in \( \theta \),
defining a new operator which we call \( L_z \) and write

\[ R\left(R(\theta)\right) = 1 - i\theta L_z/\hbar + O(\theta^2). \]

We can recognize \( L_z \) in a manner similar to before by considering small rotations
\( \langle r | R (\mathcal{R}(\theta)) | \psi \rangle = \langle r | (1 - i\theta L_z / \hbar + O(\theta^2)) | \psi \rangle, \)

\( \langle R^T (\theta) r | \psi \rangle = \langle r | \psi \rangle - i\theta \langle r | L_z | \psi \rangle / \hbar + O(\theta^2), \)

\( \langle r | L_z | \psi \rangle \approx i\hbar \frac{\psi(R^T(\theta)r) - \psi(r)}{\theta} \approx i\hbar \frac{\psi(x + \theta y, y - \theta x) - \psi(x, y)}{\theta}, \)

\( \langle r | L_z | \psi \rangle = i\hbar \lim_{\theta \to 0} \frac{\psi(x + \theta y, y - \theta x) - \psi(x, y)}{\theta} = i\hbar \left[ \frac{\partial \psi}{\partial x} - x \frac{\partial \psi}{\partial y} \right] = x \langle r | P_y | \psi \rangle - y \langle r | P_x | \psi \rangle = \langle r | (XP_y - YP_x) | \psi \rangle. \)

Thus we recognize that \( L_z \) is nothing more than angular momentum (around the z-axis, a peculiar way of describing things in two dimensions). ¹

Can large rotations be built up out of small ones? The computation is identical to that leading to (6.40), so with no further work we simply give the answer.

\( R(\mathcal{R}(\theta)) = \exp(-i\theta L_z / \hbar). \)  \hspace{1cm} (6.42)

In three dimensions, it works out that there are three independent axes you can rotate around, and consequently there are three linearly independent generators. These are the three angular momentum operators \( L \), defined by

\( L = R \times P. \)  \hspace{1cm} (6.43)

A rotation by an arbitrary amount about an arbitrary axis \( \hat{r} \) is then given by

\( R(\mathcal{R}(\hat{r}, \theta)) = \exp(-i\theta \hat{r} \cdot L / \hbar). \)  \hspace{1cm} (6.44)

The angular momentum operators \( L \) do not commute with each other, which you can deduce directly from (6.43), or by noting that rotations around different axes do not commute.

G. Solving Problems with Continuous Symmetries

Just as discrete symmetries helped us solve problems, we can use continuous symmetries to solve problems as well. Suppose, for example, we were faced with trying to find the eigenstates in three dimensions of a Hamiltonian of the form

\( H = \frac{1}{2m} P^2 + V(X, Y). \)  \hspace{1cm} (6.45)

Note that \( Z \) does not appear in the potential. This means that the Hamiltonian is translation invariant in the z-direction, and therefore it will commute with \( P_z \). Hence we will be looking for states which are eigenstates of both \( H \) and \( P_z \), say

---

¹ We’ve actually made a subtle but important misidentification here, which will be rectified in the next chapter. The implicit assumption here is that the state space has a basis \( \{|r\} \), which will not work for multiple particles, or even single particles with spin.
\[ P_z |\psi\rangle = \hbar k_z |\psi\rangle. \]

In a coordinate basis, this tells us
\[-i\hbar \frac{\partial}{\partial z} \psi(x, y, z) = \hbar k\psi(x, y, z).\]

This differential equation is easy to solve, we find \(\psi(x, y, z) \propto e^{ikz}\), but the proportionality constant may depend on \(x\) and \(y\). If we call this proportionality “constant” \(\phi(x, y)\), then we have
\[\psi(x, y, z) = \phi(x, y)e^{ikz}.\]

If you substitute this wave function into Schrödinger’s equation, using the Hamiltonian (6.45), you will find
\[\left( E - \hbar^2 k_z^2 \right) \phi(x, y) = -\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \phi(x, y) + V(x, y) \phi(x, y).\]

Thus we have quickly reduced our original three-dimensional problem to a two-dimensional problem.

In the most extreme case, imagine we were faced for the first time with solving the free particle problem, where there is no potential at all. Such a potential is independent of translations in all three dimensions, and therefore the Hamiltonian will commute with all three components of \(\mathbf{P}\). We therefore demand they satisfy \(P_z |\psi\rangle = \hbar k |\psi\rangle\), and we quickly find the eigenstates \(\psi(\mathbf{r}) = \exp(ik \cdot \mathbf{r})\).

Let’s deal with a more difficult problem, say one where we have a rotational symmetry. Suppose we have a Hamiltonian in two (or more) dimensions that commutes with the operator \(L_z\), so we have
\[H = \frac{1}{2\mu} \mathbf{P}^2 + V(|\mathbf{R}|) = \frac{1}{2\mu} \mathbf{P}^2 + V(\sqrt{X^2 + Y^2}),\]  

(6.46)

where we have changed the mass to \(\mu\) to avoid confusion with the quantum number \(m\) we are about to introduce. When you perform a rotation about the origin the potential remains unchanged, and therefore the Hamiltonian must commute with \(L_z\). Therefore, we can choose our eigenstates to be both eigenvectors of \(H\) and of \(L_z\). Since \(L_z\) has the same units as \(\hbar\), it makes sense to define the eigenvalues as
\[L_z |\psi\rangle = \hbar m |\psi\rangle,\]

(6.47)

with \(m\) an as yet undetermined dimensionless real constant.

Since the potential depends only on the distance from the origin, it makes sense to switch to polar coordinates, which are related to Cartesian coordinates by eq. (A.14), reproduced here:

---

1 Or, if you prefer, you can verify directly that \([L_z, X^2 + Y^2] = 0\).
\[ x = \rho \cos \phi, \quad y = \rho \sin \phi. \]

It will prove very helpful to figure out the following partial derivative in the new coordinates:

\[
\frac{\partial}{\partial \phi} = \frac{\partial x}{\partial \phi} \frac{\partial}{\partial x} + \frac{\partial y}{\partial \phi} \frac{\partial}{\partial y} = -\rho \sin \phi \frac{\partial}{\partial x} + \rho \cos \phi \frac{\partial}{\partial y} = x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x}.
\]

Multiplying this by \(-i\hbar\), and letting it act on an arbitrary wave function \(\psi\), we have

\[
-i\hbar \frac{\partial \psi}{\partial \phi} = -x i \hbar \frac{\partial \psi}{\partial y} + y i \hbar \frac{\partial \psi}{\partial x} = (XP_x - YP_y)\psi = L_z \psi.
\]

If we write our wave function in terms of polar coordinates (6.47) will then become

\[
-i\hbar \frac{\partial \psi}{\partial \phi}(\rho, \phi) = \hbar m \psi(\rho, \phi).
\]

It is easy to see that the solution must be proportional to \(e^{im\phi}\), but with a proportionality constant that might depend on \(\rho\). We call this proportionality “constant” \(R(\rho)\), so our wave function is

\[
\psi(\rho, \phi) = R(\rho)e^{im\phi}. \quad (6.48)
\]

In Cartesian coordinates, the variables \(x\) and \(y\) are unambiguous and unrestricted, but in polar coordinates the coordinate \(\rho\) is restricted to be positive, and the angle \(\phi\) is restricted to the range \([0, 2\pi]\). But the two extremes of the range of \(\phi\) are physically identical; there is no difference between \(\phi = 0\) and \(\phi = 2\pi\). So we must have

\[
\psi(\rho, 0) = \psi(\rho, 2\pi) \Rightarrow R(\rho) = R(\rho)e^{2\pi mi} \Rightarrow m \text{ is an integer.}
\]

The eigenvalue \(m\) is restricted to be any integer. It remains only to find the function \(R(\rho)\).

We have assured ourselves that the wave function is an eigenstate of \(L_z\), but not the Hamiltonian. Writing Schrödinger’s equation using the Hamiltonian (6.46), we have

\[
E \psi(r) = -\frac{\hbar^2}{2\mu} \nabla^2 \psi(r) + V(r)\psi(r).
\]

Substituting in the explicit form (6.48) and rewriting the Laplacian in polar coordinates using (A.17c), we have

\[
ER(\rho)e^{im\phi} = -\frac{\hbar^2}{2\mu} \left\{ \frac{1}{\rho} \frac{\partial}{\partial \rho} \left[ \rho \frac{\partial}{\partial \rho} \left[ R(\rho)e^{im\phi} \right] \right] + \frac{1}{\rho^2} \frac{\partial^2}{\partial \phi^2} \left[ R(\rho)e^{im\phi} \right] \right\} + V(\rho)R(\rho)e^{im\phi}.
\]

Expanding and simplifying, this becomes

\[
ER(\rho) = -\frac{\hbar^2}{2\mu} \left\{ \frac{d^2}{d\rho^2} R(\rho) + \frac{1}{\rho} \frac{d}{d\rho} R(\rho) - \frac{m^2}{\rho^2} R(\rho) \right\} + V(\rho)R(\rho). \quad (6.49)
\]
Note that we have managed to reduce the problem of solving a two-dimensional problem to a single differential equation, a vast improvement. Computers can efficiently solve equations like (6.49) numerically, though in a few cases one can solve such equations analytically.

As a concrete example, consider the circular infinite square well, defined by
\[
V(\rho) = \begin{cases} 
0 & \text{if } \rho \leq a, \\
\infty & \text{if } \rho > a.
\end{cases}
\]
The wave function must vanish for \( \rho > a \), but for \( \rho < a \), equation (6.49) can be rewritten
\[
\frac{d^2}{d\rho^2} R(\rho) + \frac{1}{\rho} \frac{d}{d\rho} R(\rho) - \frac{m^2}{\rho^2} R(\rho) = -\frac{2\mu E}{\hbar^2} R(\rho). \tag{6.50}
\]
This equation can be simplified further by defining a new scaled variable \( x \), given by
\[
x = \rho \sqrt{2\mu E / \hbar}.
\]
In terms of \( x \) equation (6.50) is
\[
\frac{d^2}{dx^2} R + \frac{1}{x} \frac{d}{dx} R + \left(1 - \frac{m^2}{x^2}\right) R = 0. \tag{6.51}
\]
Equation (6.51) is called Bessel’s Equation. Because it is a second order differential equation, it has two linearly independent solutions, which are normally named \( J_m(x) \) and \( N_m(x) \) (which are called Bessel functions of the first and second kind); however, \( N_m(x) \) diverges at small \( x \), and hence provides unacceptable solutions. Many sources can give you details of the solutions of Bessel’s Equation; I looked them up in Wikipedia. Maple can plot them for you with the “BesselJ” function. The first four are plotted in Fig. 6-7. As you can see, all of these functions oscillate while their amplitude damps away.

\textbf{Figure 6-7}: The Bessel functions \( J_m(x) \) for \( m = 0, 1, 2, \) and 3 respectively (red, navy, green, brown). Eventually, all Bessel functions oscillate while their amplitude damps away.
The proportionality constant can be determined by demanding the wave function be normalized.

For our function to be continuous at the boundary we need $R$ to vanish at $x = a$, so we have $J_m(a\sqrt{2\mu E}/\hbar) = 0$. Let us call the $k$'th root of $J_m(x)$ by the symbol $j_{mk}$. Then we have $a\sqrt{2\mu E}/\hbar = j_{mk}$. Solving for the energy $E$, we have

$$E_{mk} = \frac{\hbar^2 j_{mk}^2}{2\mu a^2}.$$  

The constants $j_{mk}$ cannot be written in any simple way. Again, Maple is capable of finding them with the “BesselJZeros” function. The energies generally increase as $m$ and $k$ increase, so the lowest energies come from small $m$ and small $k$.

We have completed our discussion of continuous rotations in two dimensions, where there is only one axis you can rotate around. In three dimensions, there are three axes you can rotate around. If only one axis is an axis of rotational symmetry, the analysis is largely identical to what we have just completed. If it is the $z$-axis, then again we will diagonalize $L_z$ and $H$ simultaneously (and if it isn’t the $z$-axis, I recommend changing axes and make it the $z$-axis).

What if we have two axes that can be rotated around, say the $x$- and $y$- axes? Then the Hamiltonian will commute with both $L_x$ and $L_y$,

$$[H, L_x] = [H, L_y] = 0.$$  

Indeed, it is not hard to show that if both of these vanish, that since $[L_x, L_y] = i\hbar L_z$, it also follows that $[H, L_z] = 0$. However, even though $H$ commutes with all three angular momentum operators, these operators do not commute with each other, and hence cannot be simultaneously diagonalized. Solving problems where this occurs is a bit trickier than any of the cases we have handled so far, and leads us into a general discussion of angular momentum and spin.

**Problems for Chapter 6**

1. For the finite square well in section C, we showed that (6.24) is satisfied for the even wave functions. Repeat this derivation for the odd wave functions; i.e., derive (6.25).
2. A particle of mass \( m \) in two dimensions is governed by the Hamiltonian

\[
H = \left( P_x^2 + P_y^2 \right) / 2m + \frac{1}{4} \alpha \left( X^2 + Y^2 \right)^2 + \frac{1}{4} \gamma \left( X^3 - 3XY^2 \right)
\]

(a) Show that the Hamiltonian is invariant under the transformation \( R \left( \mathcal{R} \left( 120^\circ \right) \right) \).

(b) Classify the states according to their eigenvalues under \( R \left( \mathcal{R} \left( 120^\circ \right) \right) \). What eigenvalues are possible?

(c) Suppose that \( \psi (x, y) \) is an eigenstate of \( H \) and \( R \left( \mathcal{R} \left( 120^\circ \right) \right) \) with eigenvalues \( E \) and \( \lambda \) respectively. Show that \( \psi^* (x, y) \) is also an eigenstate of \( H \) and \( R \left( \mathcal{R} \left( 120^\circ \right) \right) \), and determine its eigenvalues. (\( E \) is, of course, real).

(d) Careful measurements of the Hamiltonian discovers that the system has some non-degenerate eigenstates (like the ground state), and some states that are two-fold degenerate (two eigenstates with exactly the same eigenvalue). Explain why these degeneracies are occurring.

3. A particle of mass \( m \) and energy \( E \) in two dimensions is incident on a plane step function given by

\[
V (X, Y) = \begin{cases} 
0 & \text{if } X < 0, \\
V_0 & \text{if } X > 0.
\end{cases}
\]

The incoming wave has wave function \( \psi_{\text{in}} (x, y) = e^{i(k_x x + k_y y)} \) for \( x < 0 \).

(a) Write the Hamiltonian. Determine the energy \( E \) for the incident wave. Convince yourself that the Hamiltonian has a translation symmetry, and therefore that the transmitted and reflected wave will share something in common with the incident wave (they are all eigenstates of what operator?).

(b) Write the general form of the reflected and transmitted wave. Use Schrödinger's equation to solve for the values of the unknown parts of the momentum for each of these waves (assume \( k_x^2 \hbar^2 / 2m > V_0 \)).

(c) Assume the wave function and its derivative are continuous across the boundary \( x = 0 \). Find the amplitudes for the transmitted and reflected waves, and find the probability \( R \) of the wave being reflected.

4. A particle of mass \( M \) in three dimensions has potential \( V (X, Y, Z) = \frac{1}{4} A \left( X^2 + Y^2 \right)^2 \).

(a) Show that this Hamiltonian has two continuous symmetries, and that they commute. Call the corresponding eigenvalues \( m \) and \( k \). Are there any restrictions on \( k \) and \( m \)?

(b) What would be an appropriate set of coordinates for writing the eigenstates of this Hamiltonian? Write the eigenstates as a product of three functions (which I call \( Z, R, \) and \( \Phi \)), and give me the explicit form of two of these functions. Find a differential equation for the third, which may depend on \( k \) and \( m \).
VII. Angular Momentum

A. Angular Momentum Commutation Relations

It is easy to work out the commutation relations of the three angular momentum operators from their form, but I would like to work them out using the fact that they are the generators of rotations. Let us recall that the operator that performs a rotation around a specific axis is given by (6.44), \( R(\hat{\mathbf{r}}, \theta) = \exp(-i \theta \hat{\mathbf{r}} \cdot \mathbf{L}/\hbar) \). Consider the product of small rotations around the \( x \)- and \( y \)-axis. We will make both rotations by the same amount \( \theta \), and we will uniformly expand everything to second order in \( \theta \). First we note that

\[
R(\hat{\mathbf{r}}, \theta) = \exp \left( -i \frac{\theta L_z}{\hbar} \right) \exp \left( -i \frac{\theta L_y}{\hbar} \right)
= \left( 1 - i \frac{\theta L_z}{\hbar} - \frac{\theta^2}{2\hbar^2} L_z^2 \right) \left( 1 - i \frac{\theta L_y}{\hbar} - \frac{\theta^2}{2\hbar^2} L_y^2 \right) + O(\theta^3)
= 1 - i \frac{\theta}{\hbar} (L_z + L_y) - \frac{\theta^2}{2\hbar^2} (L_z^2 + 2L_z L_y + L_y^2) + O(\theta^3)
= 1 - i \frac{\theta}{\hbar} (L_z + L_y) - \frac{\theta^2}{2\hbar^2} (L_z + L_y)^2 - \frac{\theta^2}{2\hbar^2} [L_z, L_y] + O(\theta^3),
\]

Now, consider performing the rotations in the same order, but in the opposite direction, so we let \( \theta \to -\theta \) in (7.1):

\[
R(\hat{\mathbf{r}}, -\theta) R(\hat{\mathbf{r}}, \theta) = \exp \left( -i \frac{\theta}{\hbar} (L_z + L_y) \right) - \frac{\theta^2}{2\hbar^2} [L_z, L_y] + O(\theta^3). \tag{7.2}
\]

Now multiply (7.1) and (7.2), again keeping the computation to \( O(\theta^2) \). The exponentials will multiply to give 1. The commutation terms are multiplied by expressions which to leading order are 1. So we have

\[
R(\hat{\mathbf{r}}, \theta) R(\hat{\mathbf{r}}, -\theta) R(\hat{\mathbf{r}}, \theta) R(\hat{\mathbf{r}}, -\theta) = 1 - \frac{\theta^2}{\hbar^2} [L_z, L_y] + O(\theta^3). \tag{7.3}
\]

We now recall that the product of such operators can be computed using (6.14b), \( R(\mathcal{R}_1) R(\mathcal{R}_2) = R(\mathcal{R}_1 \mathcal{R}_2) \). We can therefore rewrite (7.3) as

\[
R(\hat{\mathbf{r}}, \theta) R(\hat{\mathbf{r}}, \theta) R(\hat{\mathbf{r}}, -\theta) R(\hat{\mathbf{r}}, -\theta) = 1 - \frac{\theta^2}{\hbar^2} [L_z, L_y] + O(\theta^3). \tag{7.4}
\]

Now, we can figure out the combination of various rotations with the help of the explicit forms for rotations given by (6.12). To second order, for example,
\[ \mathcal{R}(\mathbf{x}, \theta) \mathcal{R}(\mathbf{y}, \theta) = \left( \begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 - \frac{1}{2} \theta^2 & -\theta \\ 0 & \theta & 1 - \frac{1}{2} \theta^2 \end{array} \right) \left( \begin{array}{ccc} 1 - \frac{1}{2} \theta^2 & 0 & \theta \\ -\theta & 1 - \frac{1}{2} \theta^2 & 0 \\ 0 & 0 & 1 - \frac{1}{2} \theta^2 \end{array} \right) + O(\theta^3) \]

Substituting \( \theta \to -\theta \) and multiplying the new expression on the right, this is

\[ \mathcal{R}(\mathbf{x}, \theta) \mathcal{R}(\mathbf{y}, \theta) \mathcal{R}(\mathbf{x}, -\theta) \mathcal{R}(\mathbf{y}, -\theta) = \left( \begin{array}{ccc} 1 - \frac{1}{2} \theta^2 & 0 & \theta \\ \theta^2 & 1 - \frac{1}{2} \theta^2 & -\theta \\ -\theta & \theta & 1 - \theta^2 \end{array} \right) \left( \begin{array}{ccc} 1 - \frac{1}{2} \theta^2 & 0 & -\theta \\ -\theta & 1 - \frac{1}{2} \theta^2 & \theta \\ \theta & -\theta & 1 - \theta^2 \end{array} \right) + O(\theta^3) \]

\[ = \left( \begin{array}{ccc} 1 & -\theta^2 & 0 \\ \theta^2 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right) + O(\theta^3). \]

\[ \mathcal{R}(\mathbf{x}, \theta) \mathcal{R}(\mathbf{y}, \theta) \mathcal{R}(\mathbf{x}, -\theta) \mathcal{R}(\mathbf{y}, -\theta) = \mathcal{R}(\mathbf{z}, \theta^2) + O(\theta^3). \]

Substituting this into (7.4) we have

\[ R\left( \mathcal{R}(\mathbf{z}, \theta^2) \right) = 1 - \frac{\theta^2}{\hbar^2} [L_x, L_y] + O(\theta^3). \]  \hspace{1cm} (7.5)

Using (6.44), this is the same as

\[ 1 - \frac{i \theta^2}{\hbar^2} L_z + O(\theta^3) = 1 - \frac{\theta^2}{\hbar^2} [L_x, L_y] + O(\theta^3). \]

Demanding that the \( O(\theta^2) \) terms match, this implies \([L_x, L_y] = i\hbar L_z\). We can similarly derive \([L_y, L_z] = i\hbar L_x\) and \([L_z, L_x] = i\hbar L_y\), so in summary

\[ [L_i, L_j] = i\hbar \sum_k \varepsilon_{ijk} L_k. \]

where we have written the commutation relations for the three angular momentum operators in terms of the Levi-Civita symbol, \( \varepsilon_{ijk} \), defined by

\[ \varepsilon_{xyz} = \varepsilon_{yxz} = \varepsilon_{zxy} = -\varepsilon_{xyz} = -\varepsilon_{yzx} = -\varepsilon_{zxy} = 1, \quad \text{all others vanish}. \]  \hspace{1cm} (7.6)

The point of this laborious exercise is that the commutation relations are a direct result of the fact that the rotations themselves do not commute. The left side of (7.3) is not the identity operator simply because the order in which you do rotations makes a difference.
B. Generalized Angular Momentum

Let us define *angular momentum-like operators* $\mathbf{J}$ as any three Hermitian operators that satisfy the commutation relations

$$\left[ J_i, J_j \right] = i\hbar \sum_k \epsilon_{ijk} J_k .$$

(7.7)

These formulas are sometimes summarized as $\mathbf{J} \times \mathbf{J} = i\hbar \mathbf{J}$, which, if you write it out, is the same thing. Unfortunately, no two of these operators commute with each other, so we can diagonalize no more than one of them at a time. Our goal is to work out a basis such that we have an explicit representation of these matrices. Let us define three new operators by the definition

$$J^2 \equiv J_x^2 + J_y^2 + J_z^2, \quad J_\pm \equiv J_x \pm iJ_y .$$

(7.8)

Clearly, $J^2$ is Hermitian and $J_\pm$ are not, because of the $i$. The inverse relations for $J_\pm$ will prove useful as well:

$$J_x = \frac{1}{2}(J_+ + J_-), \quad J_y = \frac{1}{2}i(J_+ - J_-) .$$

(7.9)

The following identities are straightforward to prove (see problem 1):

$$J^2 J_\pm = J_\pm J^2, \quad [ J^2, J_\pm ] = 0 , \quad [ J_\pm, J_z ] = \pm \hbar J_\pm , \quad J^2 = J_+ J_- + J_z^2 \pm \hbar J_\pm .$$

(7.10a-d)

Though the three $J$’s don’t commute with each other, we now have an additional operator that does commute with them all. Let us pick our basis states to be eigenstates of both $J^2$ and $J_z$. We will call such states $| j, m \rangle$ with

$$J^2 | j, m \rangle = \hbar^2 \left( j^2 + j \right) | j, m \rangle , \quad J_z | j, m \rangle = \hbar m | j, m \rangle .$$

(7.11a-b)

We have factored out some sensible factors such that $j^2 + j$ and $m$ will be dimensionless. The remaining factor of $j^2 + j$ looks mysterious, but for now just treat it as an odd way of labeling the eigenvalue. It is easy to see that $J^2$ will have non-negative eigenvalues, so we can always pick a $j$, which we will choose to be non-negative, such that this works.

Let $J_\pm$ act on any eigenstate $| j, m \rangle$, then it is not hard to show that we get a new eigenstate of both operators:

$$J^2 \left( J_\pm | j, m \rangle \right) = J_\pm J^2 | j, m \rangle = J_\pm \hbar^2 \left( j^2 + j \right) | j, m \rangle = \hbar^2 \left( j^2 + j \right) \left( J_\pm | j, m \rangle \right) , \quad J_z \left( J_\pm | j, m \rangle \right) = \left( [ J_z, J_\pm ] + J_\pm J_z \right) | j, m \rangle = \left( \pm \hbar J_\pm + J_z \hbar m \right) | j, m \rangle = (m \pm 1) \hbar \left( J_\pm | j, m \rangle \right) .$$


Assuming \( J_z |j, m\rangle \) does not vanish, this implies that \( J_z |j, m\rangle \) is a new eigenstate of both \( J^2 \) and \( J_z \), though the eigenvalue of the latter has been increased or decreased by one. So we write

\[
J_z |j, m\rangle = C |j, m\pm 1\rangle.
\]

We can determine the constant \( C \) by demanding that the new state be normalized as well, so that

\[
C |j, m\pm 1\rangle = J_z |j, m\rangle, \quad \|C |j, m\pm 1\rangle\|^2 = \|J_z |j, m\rangle\|^2,
\]

\[
\langle j, m\pm 1 | C^* C | j, m\pm 1 \rangle = \langle j, m | J_z J_z | j, m \rangle,
\]

\[
|C|^2 = \langle j, m | (J^2 - J_z^2 + \hbar J_z) | j, m \rangle = \hbar^2 \langle j, m | (j^2 + j - m^2 \mp m) | j, m \rangle = \hbar^2 (j^2 + j - m^2 \mp m).
\]

where we used the identity (7.10d). We have an arbitrary phase choice to make, and as usual we demand that \( C \) be chosen real and positive, so we have

\[
J_z |j, m\rangle = \hbar \sqrt{j^2 + j - m^2 \mp m} | j, m\pm 1\rangle. \tag{7.12}
\]

When \( j^2 + j - m^2 \pm m \) vanishes, then \( J_z |j, m\rangle = 0 \). The expression \( j^2 + j - m^2 \pm m \) should never be negative. This implies two inequalities, which we collapse into one:

\[
j^2 + j - m^2 \mp m \geq 0,
\]

\[
j^2 + j - |m|^2 - |m| \geq 0,
\]

\[
(j + |m| + 1)(j - |m|) \geq 0,
\]

\[
j \geq |m|. \tag{7.13}
\]

where at the last step we used the fact that \( j + |m| + 1 \) must be positive.

Now, according to (7.12), we can start with a state \( |j, m\rangle \) and then produce the states \( |j, m\pm 1\rangle, |j, m\pm 2\rangle \), etc. Eventually \( m \) will become either so positive or so negative that (7.13) will be violated, or the raising and lowering will at some stage give zero. You only get zero, as is easily demonstrated, if \( j = |m| \), so the process of raising and lowering must terminate when this equality is satisfied. We conclude that \( j \) exceeds \( m \) by an integer, and \( m \) exceeds \(-j\) by an integer as well. Another way to put this is that \( m \) must take on one of the values

\[
m = j, j-1, j-2, \ldots, -j. \tag{7.14}
\]

We note that this implies that the difference between \( j \) and \(-j\) must be an integer, which means that \( 2j \) is an integer, and since we don’t allow it to be negative, we have

\[
j = 0, 1, \frac{3}{2}, \ldots \tag{7.15}
\]
We commonly simply state that \( j \) must be an integer or half-integer.

We have successfully related all the states with fixed \( j \) and various \( m \) via the raising and lowering operators \( J_{\pm} \). We have labeled the states by \( |j, m\rangle \). However, depending on the specific problem, there may be other quantum numbers involved, and in general our quantum states may look more like \( |n, j, m\rangle \). Also note that we have not connected states with one value of \( j \) with any others; therefore it may be that all states have the same value of \( j \) and we will label our states purely as \( |m\rangle \).

It will sometimes be useful to have specific formulas for the operators \( J \) in the basis \( |j, m\rangle \). It is traditional to order these states from the highest value to the lowest of \( m \). Then we can deduce the matrix elements of \( J_z \) and \( J_{\pm} \) from (7.11b) and (7.12) respectively, and then we can find \( J_x \) and \( J_y \) from (7.9). For example, if \( j = 1 \), we would have

\[
J_z = \hbar \begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & -1
\end{pmatrix}, \quad J_+ = \hbar \begin{pmatrix}
0 & \sqrt{2} & 0 \\
0 & 0 & \sqrt{2} \\
0 & 0 & 0
\end{pmatrix}, \quad J_- = \hbar \begin{pmatrix}
0 & 0 & 0 \\
\sqrt{2} & 0 & 0 \\
\sqrt{2} & 0 & 0
\end{pmatrix},
\]

\[
J_x = \hbar \begin{pmatrix}
0 & \frac{\sqrt{2}}{2} & 0 \\
\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\
0 & \frac{\sqrt{2}}{2} & 0
\end{pmatrix}, \quad J_y = \hbar \begin{pmatrix}
0 & -i \frac{\sqrt{2}}{2} & 0 \\
i \frac{\sqrt{2}}{2} & 0 & -i \frac{\sqrt{2}}{2} \\
0 & i \frac{\sqrt{2}}{2} & 0
\end{pmatrix}.
\]

(7.16)

It is then easy to check that \( J^2 = 2\hbar^2 \), as it must. Of special interest is the \( j = \frac{1}{2} \) representation. We find \( J = \frac{\pi}{2} \hbar \sigma \), where \( \sigma \) are the Pauli matrices, given by

\[
\sigma_x = \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}, \quad \sigma_y = \begin{pmatrix}
0 & -i \\
i & 0
\end{pmatrix}, \quad \sigma_z = \begin{pmatrix}
1 & 0 \\
0 & -1
\end{pmatrix}.
\]

(7.17)

These matrices satisfy the multiplication rules

\[
\sigma_i \sigma_j = 1 \delta_{ij} + i \sum_k \epsilon_{ijk} \sigma_k.
\]

(7.18)

C. Spherically Symmetric Problems

Suppose we have a Hamiltonian in three dimensions of the form

\[
H = \frac{1}{2\mu} \mathbf{P}^2 + V(|\mathbf{R}|) = \frac{1}{2\mu} \mathbf{P}^2 + V(\sqrt{\mathbf{R}^2}),
\]

so that the potential depends only on the distance from the origin. We have renamed the mass \( \mu \) to avoid confusion with the quantum number \( m \). Since all rotation operators commute with both \( \mathbf{P}^2 \) and \( \mathbf{R}^2 \), it follows that all rotations commute with \( H \), and therefore
\[ [H, L] = 0 \]. Hence we can describe our states as eigenstates of \( H, L^2, \) and \( L_z \) simultaneously and write our states as \( |n, l, m\rangle \), where

\[
H |n, l, m\rangle = E_n |n, l, m\rangle,  \\
L^2 |n, l, m\rangle = \hbar^2 (l^2 + l) |n, l, m\rangle,  \\
L_z |n, l, m\rangle = \hbar m |n, l, m\rangle.  
\]  
(7.19)

Schrödinger’s time-independent equation will take the form

\[
E\psi (r) = -\frac{\hbar^2}{2\mu} \nabla^2 \psi (r) + V(|r|)\psi (r).  
\]  
(7.20)

Because of the spherically symmetric nature of the problem, it seems very sensible to switch to spherical coordinates, defined by the relations

\[
x = r \sin \theta \cos \phi, \quad y = r \sin \theta \sin \phi, \quad z = r \cos \theta.  
\]

It will prove necessary to rewrite our angular momentum operators in terms of derivatives of the various spherical coordinates. We start by noting that

\[
\nabla = \frac{\partial}{\partial \phi} \cos \phi + \frac{\partial}{\partial \theta} \sin \phi \cot \theta = -r \sin \theta \sin \phi \frac{\partial}{\partial x} + r \sin \theta \cos \phi \frac{\partial}{\partial y} = -r \cos \theta \sin \phi \frac{\partial}{\partial x} + r \cos \phi \frac{\partial}{\partial y},  \\
\nabla \cdot \nabla = \frac{\partial^2}{\partial \phi^2} - \frac{\sin \phi}{\sin \theta} \frac{\partial}{\partial \phi} \sin \phi + \frac{\cos \phi}{\cos \theta} \frac{\partial}{\partial \phi} \cos \phi = r \cos \theta \sin \phi \frac{\partial}{\partial x} + r \cos \phi \frac{\partial}{\partial y} - r \sin \theta \frac{\partial}{\partial z}.  
\]

Now we take some specific linear combinations of these two operators:

\[
\cos \phi \frac{\partial}{\partial \theta} - \sin \phi \cot \theta \frac{\partial}{\partial \phi} = r \cos \theta \cos^2 \phi \frac{\partial}{\partial x} + r \cos \theta \sin \phi \cos \phi \frac{\partial}{\partial y} - r \sin \phi \cos \theta \frac{\partial}{\partial z}  \\
\quad + r \cos \theta \sin^2 \phi \frac{\partial}{\partial x} - r \cos \phi \sin \phi \frac{\partial}{\partial y}  \\
\quad = r \cos \theta \frac{\partial}{\partial x} - r \sin \theta \cos \phi \frac{\partial}{\partial z} = z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z},  \\
- \sin \phi \frac{\partial}{\partial \theta} - \cos \phi \cot \theta \frac{\partial}{\partial \phi} = -r \cos \theta \cos \phi \sin \phi \frac{\partial}{\partial x} - r \cos \theta \sin^2 \phi \frac{\partial}{\partial y} + r \sin \theta \sin \phi \frac{\partial}{\partial z}  \\
\quad + r \cos \theta \sin \phi \cos \phi \frac{\partial}{\partial x} - r \cos \phi \sin \phi \frac{\partial}{\partial y}  \\
\quad = -r \cos \theta \frac{\partial}{\partial y} + r \sin \theta \sin \phi \frac{\partial}{\partial z} = y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y},  \\
\frac{\partial}{\partial \phi} = -r \sin \theta \sin \phi \frac{\partial}{\partial x} + r \sin \theta \cos \phi \frac{\partial}{\partial y} = x \frac{\partial}{\partial z} - y \frac{\partial}{\partial z}.  
\]

Now multiply these expressions by \(-i\hbar\) and let them act on an arbitrary wave function \(\psi\).

The resulting expressions are simply angular momentum operators.
\[ i\hbar \left( \sin \phi \frac{\partial}{\partial \theta} \cos \phi \cot \theta \frac{\partial}{\partial \phi} \right) = i\hbar \frac{\partial}{\partial y} - i\hbar \frac{\partial}{\partial z} - = YP_y - ZP_z = L_y, \]
\[ i\hbar \left( \sin \phi \cot \theta \frac{\partial}{\partial \phi} - \cos \phi \frac{\partial}{\partial \theta} \right) = i\hbar \frac{\partial}{\partial z} - i\hbar \frac{\partial}{\partial x} - = ZP_z - XP_x = L_x, \]
\[ -i\hbar \frac{\partial}{\partial \phi} = i\hbar \frac{\partial}{\partial x} - i\hbar \frac{\partial}{\partial y} = XP_\phi - YP_\phi = L_\phi. \]

So, in summary,
\[ L_x = i\hbar \sin \phi \frac{\partial}{\partial \theta} + i\hbar \cos \phi \cot \theta \frac{\partial}{\partial \phi}, \quad L_y = -i\hbar \cos \phi \frac{\partial}{\partial \theta} + i\hbar \sin \phi \cot \theta \frac{\partial}{\partial \phi}, \quad L_z = -i\hbar \frac{\partial}{\partial \phi}. \]

(7.21a)

(7.21b)

It is also helpful to derive some simple expressions for the “raising” and “lowering” operators \( L_\pm \):
\[ L_\pm = L_x \pm iL_y = i\hbar \left( \sin \phi \pm i \cos \phi \right) \frac{\partial}{\partial \theta} + i\hbar \left( \cos \phi \pm i \sin \phi \right) \cot \theta \frac{\partial}{\partial \phi}. \]

(7.22)

The combination \( L^2 \) can be worked out with some effort starting with (7.21).
\[ L^2 \psi = -\frac{\hbar^2}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \psi}{\partial \theta} \right) - \frac{\hbar^2}{\sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2}. \]

(7.23)

With the help of (A.22d) it is easy to rewrite Schrödinger’s equation (7.20) in terms of angular momentum operators and radial derivatives.
\[ E\psi = -\frac{\hbar^2}{2\mu} \left[ \frac{1}{r} \frac{\partial^2 (r\psi)}{\partial r^2} \right] + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} + V(r)\psi, \]
\[ E\psi = -\frac{\hbar^2}{2\mu r^2} \frac{\partial^2 \psi}{\partial r^2} + \frac{1}{2\mu r^2} L^2 \psi + V(r)\psi. \]

(7.24)

Now, we know that we will be seeking eigenstates of \( L^2 \) and \( L_z \), and we see from (7.23) and (7.21b) that this provides two constraints on angular derivatives. It therefore seems sensible to guess that the eigenvalues of these two operators will provide all the information to determine the angular dependence of the wave functions. We therefore write
\[ \psi (r) = \psi (r, \theta, \phi) = R (r) Y (\theta, \phi). \]

(7.25)

We will demand that this wave function be an eigenstate of our three operators. From (7.19), together with the observation that \( r \) does not appear in \( L_z \) or \( L^2 \), we see that
We now note that the radial functions $R(r)$ cancel out, and we are left with equations describing only the angular part of the wave function:

\[ L_2 \psi = L_2 \left[ R(r)Y(\theta, \phi) \right] = R(r) L_2 Y(\theta, \phi) = m h R(r) Y(\theta, \phi), \]

\[ L^2 \psi = L^2 \left[ R(r)Y(\theta, \phi) \right] = R(r) L^2 Y(\theta, \phi) = \hbar^2 \left( l^2 + l \right) R(r) Y(\theta, \phi). \]

Let us set aside temporarily the goal of finding the angular wave function and instead work on the radial wave function. Substituting (7.25) into (7.24) and using (7.26b), we quickly find

\[ E R(r) Y(\theta, \phi) = -\frac{\hbar^2}{2\mu r} \frac{d^2}{dr^2} \left[ r R(r) Y(\theta, \phi) \right] + \frac{1}{2\mu r^2} R(r) L^2 Y(\theta, \phi) + V(r) R(r) Y(\theta, \phi), \]

\[ ER(r) = -\frac{\hbar^2}{2\mu r} \frac{d^2}{dr^2} \left[ r R(r) \right] + \frac{l^2 + l}{2\mu r^2} \hbar^2 R(r) + V(r) R(r). \]  

The original problem has been reduced to that of solving a simple uncoupled differential equation. Note that $m$ has disappeared from the problem. This implies that states with different $m$ values but the same $l$ value will be degenerate and have the same radial wave function. Note that we can choose $R(r)$ to be a real function.

Assuming we can solve (7.27) (which we certainly can, at least numerically) we can then normalize it by demanding that

\[ 1 = \iiint |\psi(r)|^2 d^3 r = \int_0^\infty R^2(r) r^2 dr \int |Y(\theta, \phi)|^2 d\Omega, \]  

where we have introduced the shorthand notation

\[ \int f(\theta, \phi) d\Omega \equiv \int_0^{2\pi} d\phi \int_0^\pi \sin \theta d\theta f(\theta, \phi) \equiv \int_0^{2\pi} d\phi \int_0^1 d\cos \theta f(\theta, \phi). \]

We call $d\Omega$ the differential solid angle, ($\Omega$ is itself the solid angle). We have free choice in how we divide the two factors in (7.28) so long as we assure their product is one, but it seems easiest to just demand

\[ \int_0^\infty R^2(r) r^2 dr = 1, \]  

\[ \int |Y(\theta, \phi)|^2 d\Omega = 1. \]  

We thus need to solve (7.27) and normalize the solutions using (7.29a), and we have our energies and our wave functions. It remains to work out explicit forms for $R(r)$, which depend on the potential, and to find the angular wave functions $Y(\theta, \phi)$, which do not.
D. Spherical Harmonics

We now will derive the spherical harmonics, denoted \( Y^m_l(\theta, \phi) \), which are solutions to (7.26a) and (7.26b) subject to the normalization condition (7.29b).

We first note that the solution to (7.26a) is easy to deduce. Using (7.21b), we see that \( \partial Y(\theta, \phi)/\partial \phi = imY(\theta, \phi) \), from which we conclude

\[
Y^m_l(\theta, \phi) = f(\theta) e^{im\phi}.
\] (7.30)

In an argument similar to before, we argue that since \( \phi = 0 \) and \( \phi = 2\pi \) are identical, the wave function must be equal at these two places, \( Y^m_l(\theta, 0) = Y^m_l(\theta, 2\pi) \). This will only occur if \( m \) is an integer, and since \( m \) runs from \(-l\) to \(+l\) in integer steps, we conclude that \textbf{both} \( l \) and \( m \) must be integers.

To find \( Y^{-l}_l(\theta, \phi) \) it is easiest to start by working out the special case \( Y^{-l}_l(\theta, \phi) \). In this case, we cannot lower the index \( m \), since it is already at its minimum \(-l\), so we conclude that \( L Y^{-l}_l(\theta, \phi) = 0 \); in other words,

\[
he^{-i\phi} \left\{ -\frac{\partial}{\partial \theta} \left[ f(\theta) e^{-i\phi} \right] + i \cot \theta \frac{\partial}{\partial \phi} \left[ f(\theta) e^{-i\phi} \right] \right\} = 0.
\]

If you expand this out and simplify, it can be rewritten as

\[
\frac{df(\theta)}{f(\theta)} = l \cot \theta d\theta = l \frac{d\sin \theta}{\sin \theta}.
\]

Integrating, exponentiating, and substituting back into (7.30), we find

\[
Y^{-l}_l(\theta, \phi) = N \sin^l \theta e^{-i\phi}.
\]

The normalization constant \( N \) can be determined (up to phase) using (7.29b):

\[
1 = \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi |Y^{-l}_l(\theta, \phi)|^2 = 2\pi |N|^2 \int_0^\pi \sin^{2l+1} \theta d\theta = |N|^2 \frac{4\pi \cdot 2^{2l} (l)!^2}{(2l+1)!}.
\]

The last integral can be demonstrated by integrating by parts and using recursion. If we choose \( N \) to be real and positive, we find

\[
Y^{-l}_l(\theta, \phi) = \frac{1}{2^l l! \sqrt{4\pi}} \sin^l \theta e^{-i\phi}.
\] (7.31)

Now that we have one spherical harmonic, it is straightforward to find the rest. Applying (7.12) and (7.22), we can raise the index \( m \), since

\[
Y^{m+1}_l(\theta, \phi) = \frac{L^m_l(\theta, \phi)}{h(\sqrt{l^2 + l - m^2 - m})} = \frac{e^{i\phi}}{\sqrt{(l-m)(l+m+1)}} \left( \frac{\partial}{\partial \theta} + i \cot \theta \frac{\partial}{\partial \phi} \right) Y^m_l(\theta, \phi).
\]

Applying this operation repeatedlyly to (7.31), we find
We have derived an expression for all of the spherical harmonics, if a rather complicated and impractical one. Several of the spherical harmonics are given at the end of this chapter, or they can be computed in the online Maple document titled “Hydrogen” I have posted on the website.

Some properties of the spherical harmonics are in order. First of all, it is not hard to see that if \( Y_l^m (\theta, \phi) \) is an eigenstate of \( L^2 \) and \( L_z \), so also will \( Y_l^m (\theta, \phi) \), though the \( L_z \) eigenvalue will be \(-m\) instead of \( m\). This implies that \( Y_l^m (\theta, \phi) \) must be virtually the same as \( Y_l^{-m} (\theta, \phi) \), and we find in fact

\[
Y_l^m (\theta, \phi)^* = (-1)^m Y_l^{-m} (\theta, \phi).
\]

This can save half the work of getting the \( Y_l^m (\theta, \phi) \)’s by recursion.

The parity operation takes \( r \to -r \), or, in terms of angles, \( (\theta, \phi) \to (\pi - \theta, \phi + \pi) \). It is not hard to show using (7.31) then that

\[
\Pi Y_l^{-l} (\theta, \phi) = Y_l^{-l} (\pi - \theta, \phi + \pi) = (-1)^l Y_l^{-l} (\theta, \phi).
\]

It is then easy to show that since parity commutes with the angular momentum operators, we can use (7.32) to show that this must work for all values of \( m \):

\[
\Pi Y_l^m (\theta, \phi) = (-1)^l Y_l^m (\theta, \phi).
\]  

(7.33)

Because any two \( Y_l^m (\theta, \phi) \)’s have different eigenvalues under \( L^2 \) or \( L_z \), they must be orthogonal, which implies

\[
\int Y_l^m (\theta, \phi)^* Y_{l'}^{m'} (\theta, \phi) d\Omega = \delta_{mm'}.
\]

(7.34)

We should be able to write any wave function in terms of spherical harmonics, so for any angular function,

\[
f (\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_{lm} Y_l^m (\theta, \phi)
\]

(7.35)

for an appropriate choice of \( c_{lm} \). It is not hard to get the \( c_{lm} \)’s with the help of (7.34), which tells us

\[
c_{lm} = \int Y_l^m (\theta, \phi)^* f (\theta, \phi) d\Omega.
\]

Substituting this back into (7.35) yields a complenlteness relation, which looks like

\[
\sum_{l=0}^{\infty} \sum_{m=-l}^{l} Y_l^m (\theta, \phi)^* Y_{l'}^{m'} (\theta', \phi') = \delta (\phi - \phi') \delta (\cos \theta - \cos \theta')
\]

\[
= \csc \theta \delta (\phi - \phi') \delta (\theta - \theta').
\]

(7.36)
The extra factor of \( \csc \theta \) may seem odd, but it is merely to counteract the factor of \( \sin \theta \) in the solid angle differential \( d\Omega = \sin \theta d\theta d\phi \).

We have now nearly solved an arbitrary spherically symmetric problem. We start by picking values of \( l \) and \( m \), which tells us the angular dependence. We then solve the radial equation (7.27), whose solution will be labeled \( R_m(r) \), normalize it using (7.29a), and our final wave function will be

\[
\psi(r, \theta, \phi) = R(r)Y_l^m(\theta, \phi).
\]  

(7.37)

E. The Hydrogen Atom

As a specific example, consider the hydrogen atom, consisting of two particles (a nucleus and an electron) bound together by electric attraction. The Hamiltonian is

\[
H_{\text{tot}} = \frac{p_1^2}{2m_1} + \frac{p_2^2}{2m_2} - \frac{ke^2}{|R_1 - R_2|},
\]

where \( e \) is the charge on the nucleus (-\( e \) on the electron) and \( ke \) is Coulomb’s constant. At first glance, it is not obvious how to proceed. If we were working with this classically, we would first note that since there are no external forces, total momentum would be conserved, and we would expect the center of mass position to move linearly at all times. We would then rewrite the problem in terms of the relative coordinate. Inspired by these classical ideas, we define four new operators

\[
\begin{align*}
P_{\text{cm}} &= p_1 + p_2, \\
R_{\text{cm}} &= \frac{m_1R_1 + m_2R_2}{m_1 + m_2}, \\
P &= \frac{m_2p_1 - m_1p_2}{m_1 + m_2}, \\
R &= R_1 - R_2.
\end{align*}
\]

(7.38)

Three of these seem obvious choices, since \( P_{\text{cm}}, R_{\text{cm}}, \) and \( R \) are the total momentum, center of mass position, and relative position respectively. The relations were chosen to make sure that we will have simple commutation relations between our operators, which are

\[ [R_{\text{cm},i}, P_{\text{cm},j}] = [R_i, P_j] = i\hbar \delta_{ij}, \quad \text{all other commutators vanish.} \]

It is a straightforward matter to rewrite the Hamiltonian in terms of these new variables. We find

\[
H_{\text{tot}} = \frac{p_{\text{cm}}^2}{2M} + \frac{p^2}{2\mu} - \frac{ke^2}{|R|},
\]

where

\[\text{(7.39)}\]

\[\text{(7.40)}\]

\[\text{(7.41)}\]
\[ M = m_1 + m_2, \quad \mu = \frac{m_1 m_2}{m_1 + m_2} = \left( m_1^{-1} + m_2^{-1} \right)^{-1}. \] (7.39)

\( M \) is the total mass of the system, and \( \mu \) is the so-called reduced mass of the system. For an electron orbiting a nucleus, \( \mu \) is just a fraction of a percent below the electron’s mass.

To make things clearer, we divide \( H_{\text{tot}} \) into two pieces, given by

\[ H_{\text{tot}} = H_{\text{cm}} + H, \] (7.40a)

\[ H_{\text{cm}} = \frac{\mathbf{p}_{\text{cm}}^2}{2M}, \] (7.40b)

\[ H = \frac{\mathbf{p}^2}{2\mu} - \frac{k e^2}{|\mathbf{R}|}. \] (7.40c)

The first Hamiltonian is concerned only with the center of mass of the system, and the second only with the relative motion of the two particles. It is clear that the problem has neatly divided itself into two problems which can be solved separately. The solution to the first is plane waves. If we write our wave functions, for example, in terms of the center of mass coordinate \( r_{\text{cm}} \) and the relative coordinate \( r \), the wave function will factor and we have

\[ \Psi(r_{\text{cm}}, r) \propto \exp \left( i \mathbf{k}_{\text{cm}} \cdot r_{\text{cm}} \right) \psi(r). \]

If we substitute this into the Hamiltonian (7.40) and use Schrödinger’s time independent equation, we will then find

\[ E_{\text{tot}} \psi(r) = \frac{\hbar^2 k_{\text{cm}}^2}{2M} \psi(r) + \frac{\mathbf{p}^2}{2\mu} \psi(r) - \frac{k e^2}{|\mathbf{R}|} \psi(r). \]

The first term on the right is just the kinetic energy of the whole atom. The remaining terms tell us that \( \psi(r) \) is an eigenstate of \( H \). If we call its eigenvalue \( E \), then we have

\[ E_{\text{tot}} = \frac{\hbar^2 k_{\text{cm}}^2}{2M} + E, \quad H |\psi\rangle = E |\psi\rangle, \]

where \( H \) is given by (7.40c).

We have thus reduced the original problem with two particles into one involving only one. To solve this, we first note that the Hamiltonian (7.40c) is rotationally invariant, and thus we can immediately write down most of the answer. Our wave functions are given by (7.37), \( \psi(r, \theta, \phi) = R(r) Y_l^m(\theta, \phi) \). The radial wave function will satisfy (7.27), so

\[ E R(r) = -\frac{\hbar^2}{2\mu} \frac{1}{r} \frac{d^2}{dr^2} \left[ r R(r) \right] + \frac{l^2 + l + 1}{2\mu r^2} \hbar^2 R(r) - \frac{k e^2}{r} R(r). \] (7.41)

Our goal now is to find all bound state solutions of (7.41).

We begin by studying its overall behavior, first at very large \( r \), and then at very small \( r \). We note that at large \( r \), most of the terms drop off as powers of \( r \). Keeping only those terms that don’t, we have

© 2011, Eric D. Carlson 117
large $r$: \[ ER(r) \approx -\frac{\hbar^2}{2\mu} \frac{d^2 R(r)}{dr^2}. \]

If we now define the constant $a$ by
\[ E = -\frac{\hbar^2}{2\mu a^2}, \quad (7.42) \]
then at large $r$, $R^*(r) \approx R(r) a^{-2}$, a differential equation with two linearly independent solutions, $R(r) = e^{\pm r/a}$. The solution $e^{r/a}$ will diverge at infinity, which is unacceptable, so we keep only the minus sign.

Of course, $R(r) = e^{-r/a}$ is only a crude approximation, but we define a new function $f(r)$ as
\[ R(r) = e^{-r/a} f(r) \quad (7.43) \]
and hope that $f(r)$ will have less severe $r$-dependence than $R$. Note that (7.43) contains no assumptions at all; we can make this change of functions without restricting in any way the solutions we are considering. We substitute (7.43) and (7.42) into (7.41) and simplify to find
\[ \frac{1}{r} \frac{d}{dr} \left[ r f(r) \right] - \frac{2}{ar} \frac{d}{dr} \left[ r f(r) \right] - \frac{l^2 + l}{r^2} f(r) + \frac{2\mu k_e^2}{\hbar^2 r} f(r) = 0. \quad (7.44) \]

A reasonable second step would be to deduce the small $r$ behavior. If we assume that for small $r$ we have $f(r) \sim r^\beta$, then substituting this into (7.44), we would approximate
\[ \text{small } r: \quad \frac{1}{r} \beta (\beta + 1) r^{\beta - 1} - \frac{2}{ar} (\beta + 1) r^\beta = \frac{l^2 + l}{r^2} r^\beta + \frac{2\mu k_e^2}{\hbar^2 r} r^\beta \approx 0. \]

For small $r$ the important terms will be those with the lowest power of $r$, the first and third terms, and, dropping the other terms, we therefore find $\beta^2 + \beta = l^2 + l$. This equation has two solutions, $\beta = l$ and $\beta = -l - 1$, but since $l$ is zero or positive, if we have $f(r) \sim r^{-l-1}$, our radial wave function (7.43) will diverge at the origin, which is unacceptable.\(^1\) We therefore demand that for small $r$, $f(r) \sim r^l$.

We now imagine expanding $f$ in a Taylor series about the origin, so that
\[ f(r) = \sum_{i=0}^{\infty} f_i r^i. \quad (7.45) \]

\(^1\) You might think it is okay for the wave function to weakly diverge; for example, the normalization integral includes an extra factor of $r^2$, which will therefore converge provided $\beta > -\frac{1}{2}$. However, the kinetic term is more problematic, and will lead to divergences unless $\beta > -\frac{1}{2}$, so $\beta = -l - 1$ does not work for any $l$.\]
Because we have already argued that \( f(r) \sim r^l \) for small \( r \), the first \( l - 1 \) terms in this sum must vanish, and the sum actually starts at \( i = l \).

We now substitute (7.45) into (7.44), taking the derivatives term by term, and recall that our sums start at \( l \).

\[
\sum_{i=l}^{\infty} (i^2 + i) f_{i} r^{i-2} - \frac{2}{a} \sum_{i=l}^{\infty} (i+1) f_{i} r^{i-1} - (l^2 + l) \sum_{i=l}^{\infty} f_{i} r^{i-2} + \frac{2\mu k e^2}{\hbar^2} \sum_{i=l}^{\infty} f_{i} r^{i-1} = 0,
\]

\[
\sum_{i=l+1}^{\infty} (i^2 + i - l^2 - l) f_{i} r^{i-2} = \sum_{i=l+1}^{\infty} \left[ (i+1) \left( \frac{2}{a} - \frac{2}{a_0} \right) \right] f_{i-1} r^{i-2},
\]

(7.46)

where we have defined the Bohr radius:

\[
a_0 = \frac{\hbar^2}{\mu k e^2}.
\]

(7.47)

In (7.46) note that the first term on the left side vanishes, so we might as well start the sum at \( i = l + 1 \). On the right side, redefine \( i \to i - 1 \). We then have

\[
\sum_{i=l+1}^{\infty} (i^2 + i - l^2 - l) f_i r^{i-2} = \sum_{i=l+1}^{\infty} \left( \frac{2i}{a} - \frac{2}{a_0} \right) f_{i-1} r^{i-2}.
\]

In this equation both sides are given as a power series in \( r \); the only way they can match is if the coefficients of every power of \( r \) match, so that

\[
f_i = \frac{2i/a - 2/a_0}{i^2 + i - l^2 - l} f_{i-1}.
\]

(7.48)

It looks like we can simply start from any negative energy \( E \), derive \( a \) from (7.42), and then use the relation (7.48) to get an infinite series of coefficients which are then substituted into (7.45) and (7.43) to find a valid eigenstate of the Hamiltonian. This turns out not to be correct. Consider the expression (7.48) as \( i \) becomes very large. Ignoring all the smaller terms, we would have \( f_i \sim (2i/a) f_{i-1}/i! \), and so substituting this into (7.45) and then into (7.43), we would then have for large \( r \) (where the large \( i \)-values are the most important),

\[
R(r) \sim e^{-r/a} \sum_i f_i r^i \sim e^{-r/a} \sum_i \left( \frac{2r}{a} \right)^i i! = e^{-r/a} e^{2r/a} = e^{r/a}.
\]

As already explained, this is unacceptable. We only want wave functions that can be normalized, and this one cannot be. How can we avoid this happening? Well, the derivation assumed that no term in the series is zero. If at any point we find a value, say \( i = n \), such that \( f_n = 0 \), then (7.48) will assure that all subsequent terms also vanish, and the disaster will not actually happen. We therefore demand that they vanish at some point, such that \( f_n = 0 \), while all previous ones were not, so that we have from (7.48)
\[ 0 = f_n = \frac{2n/a - 2/a_0}{n^2 + n - l^2 - l} f_{n-1} \Rightarrow \frac{2n}{a} - \frac{2}{a_0} = 0, \]
\[ a = na_0. \quad (7.49) \]

We can substitute (7.49) into (7.48), and also put (7.45) into (7.43), and finally copy our normalization condition (7.29a) to get all our equations for the radial wave functions together.

\[ R_{nl}(r) = e^{-r/na} \sum_{i=l}^{a-1} f_i r^i, \quad (7.50a) \]

\[ f_i = \frac{2(i-n)}{na_0(i^2 + i - l^2 - l)} f_{i-1}, \quad (7.50b) \]

\[ \int_0^\infty R_{nl}^2(r) r^2 dr = 1, \quad (7.50c) \]

\[ \psi_{nlm}(r) = R_{nl}(r) Y_l^m(\theta, \phi), \quad (7.50d) \]

where the constant \( a_0 \) was given in (7.47). Several of these wave functions are given at the end of this chapter, or they can be computed in the online Maple document titled “Hydrogen.” The phases are assumed to be real, and the overall sign is normally chosen so that \( f_i \) is positive. We can also get the energy from (7.42), (7.49) and (7.47), and we find

\[ E_n = -\frac{\hbar^2}{2\mu n^2 a_0^2} = -\frac{\mu k_e^2 e^4}{2\hbar^2 n^2}. \quad (7.51) \]

We note first that the internal energy depends only on the quantum number \( n \), which is called the principal quantum number. In most other problems it would depend on both \( n \) and \( l \). Another more useful version of (7.51) is

\[ E_n = -\frac{\alpha^2(\mu c^3)}{2n^2} = -\frac{13.6 \text{ eV}}{n^2}, \quad (7.52) \]

where \( \alpha \) is a dimensionless constant known as the fine structure constant, given by

\[ \alpha \equiv \frac{k_e e^2}{\hbar c} \approx \frac{1}{137.04}. \]

We have thus found a complete description of all hydrogen wave functions and their energies. The quantum numbers \( n, l \) and \( m \) are chosen according to the constraints

\[ n = 1, 2, 3, \ldots, \]
\[ l = 0, 1, 2, \ldots, n - 1, \]
\[ m = -l, -l+1, \ldots, +l. \quad (7.53) \]

\[ ^1 \text{One advantage of (7.52) is that it is true in any system of units. Expressions like (7.51), as well as the} \]
\[ \text{definition of the fine structure constant, will depend on units, though the value of the fine structure constant} \]
\[ \text{will always be } \alpha = 1/137. \]
The wave function is the product of the radial wave functions and the spherical harmonics. The energies are determined by \( n \) alone, in this approximation, and are given by (7.52). We expect that the energy would not depend on \( m \), since the various states with different \( m \) values are related by the generators of rotational symmetry, but the independence of the quantum number \( l \) is a surprising result, and we might designate such a case an accidental degeneracy.\(^1\) It is interesting to figure out how many different states have the same energy. Looking at our constraints in (7.53), we see that for any given value of \( l \) there are \( 2l + 1 \) possible values of \( m \). Adding this up for all allowed \( l \) for a fixed \( n \) tells you the degeneracy of the states with energy \( E_n \):

\[
d(E_n) = \sum_{l=0}^{n-1} (2l+1) = \sum_{l=0}^{n-1} \left[(l+1)^2 - l^2\right] = n^2 - 0^2 = n^2.
\]

Later, when we learn about spin, we will realize that we have missed an additional degree of freedom, and the correct degeneracy is \( d(E_n) = 2n^2 \).

It is worth contemplating the various equations and their validity. For example, we have ignored relativistic effects. Do we expect there to be any? A naive calculation can help get us the answer. Classically the velocity of a particle can be found using \( E = \frac{1}{2}mv^2 \), so if we equate this to the magnitude of the binding energy (a vague guess) (7.52), we can solve for the velocity, and we find \( v \sim c \), or \( v/c \sim \alpha = 1/137 \). Since relativistic effects tend to be of order \( (v/c)^2 \), such effects will tend to be down by four or five orders of magnitude from the size of the formula (7.52). However, spectroscopy is a very mature experimental field, and ultimately we will consider these small but measurable relativistic effects.

How about the size of the nucleus? We have treated the nucleus as a point source of electric potential, but it has a finite size, and if the electron is within this size, we no longer can treat it as a point source. But the characteristic size of an atom is around \( na_0 \), and in ordinary atoms we can estimate \( \mu \approx m_e \), and we find \( a_0 = 53.0 \text{ pm} \), while a hydrogen nuclear size is around \( R = 2 \text{ fm} \) or so, so this seems like a safe approximation. Hence we conclude that (7.52) will provide an excellent approximation to the energies of the various levels in hydrogen. Note that the mass appearing in (7.52) is the reduced mass, a combination of the electron and nuclear masses given by (7.39). The shift between, say, \(^2\text{H} \) and \(^1\text{H} \), the two naturally occurring isotopes of hydrogen, are small but easily measurable in the laboratory. Indeed, in some astrophysical settings, it is possible to distinguish between these two isotopes and determine the isotopic composition of, for example, clouds of otherwise invisible hydrogen that produce absorption lines between us and quasars.

Can the formulas we’ve derived be applied to anything other than an electron of charge \( -e \) circling a nucleus of charge \( e^+ \)? The answer, with a bit of modification, is yes.

---

\(^1\) Actually, it isn’t entirely accidental. The degeneracy ultimately can be traced to the existence of more quantities that commute with the Hamiltonian, the so called Runge-Lenz vector operators.

\(^2\) The electron, in contrast, has no size that has been experimentally observed.

© 2011, Eric D. Carlson
Our equations can be applied not just to hydrogen, but to any electromagnetically bound system of two particles. For example, the hydrogen nucleus can be replaced by some other nucleus with charge $+Ze$, which would make our system a He$^{+}$ ion, or Li$^{+2}$, Be$^{+3}$, etc. The effect is that any factor of $e^2$ would become a factor of $Ze^2$. In particular, the binding energy and characteristic radius would change to

$$E_n^Z = -\frac{Z^2\alpha^2 (\mu e^2)}{2n^2} = -\frac{(13.6 \text{ eV}) Z^2}{n^2}$$

and

$$a_n^Z = \frac{a_0}{Z}.$$  

The decrease of the typical radius would drop the electron to less than a pm from the nucleus at large $Z$, still substantially larger than the size of the nucleus, but no longer irrelevant. The characteristic velocity of an electron would change considerably, so we would have for the ground state $v/c \sim \alpha Z$, which suggests that as we move from hydrogen to uranium ($Z = 92$), relativistic effects would gradually change from a tiny but measurable correction to a large effect.

It is possible to replace the proton with other exotic particles, such as the anti-muon or the anti-electron, to form muonium and positronium respectively. The anti-muon is an unstable particle with a typical lifespan of $2\mu$s, long enough for it to find an electron and bind with it. Because the anti-muon, like the muon, is some 207 times heavier than the electron, the reduced mass of muonium is still very close to that of the electron, and the size of the resulting atom $a_0$ and the binding energies $E_n$ will be nearly identical. In contrast, positronium has an anti-electron, also called a positron, with a mass equal to that of the electron. The positron is stable, but in the proximity of an electron it can annihilate with it to produce two or more photons. The reduced mass is then $\mu = \frac{1}{2}m_e$, which means the binding energies will be half that of hydrogen. The characteristic size $a_0$ of positronium will be twice as large as that of hydrogen, but if we draw a classical sort of picture of the size of positronium, as shown in Fig. 7-1, we see that it still has the same diameter as that of hydrogen.

It is also possible to replace the electron with various negatively charged particles. Bound states have been formed by replacing the electron by a muon, the $\pi^-$ and $K^-$ mesons, the anti-proton, and the $\Sigma^-$ baryon. With the exception of the muon, these other particles

**Figure 7-1**: Comparison of hydrogen (left) and positronium (right). Even though the separation between the two components is twice as big for positronium, the diameter of the structure is the same.
are all strongly interacting, which tends to cause large changes in the energy levels, so
formulas such as (7.52) become rather unreliable. The muon, on the other hand, is long-
lived, and can easily be captured to bind to a proton or indeed any nucleus. Indeed, the
large mass means that the energy levels (7.52) will be much larger for such muonic
atoms. The muon will move with the same approximate speed, so relativistic effects are
no less nor more important than for the electron. However, the characteristic radius will
be of order

\[ a_0^{\mu Z} = \frac{a_0}{Z} \left( \frac{m_e}{m_\mu} \right) \approx \frac{255 \text{ fm}}{Z}. \]

For small \( Z \), this is still much larger than the nuclear size, but as we increase \( Z \), it will
become comparable. Indeed, the resulting shifts in energies can be measured for muonic
atoms, and this can be one of the best tools for accurately measuring the approximate size
of the nucleus. Note also that since the muon is so close to the nucleus, any electrons in
the atom will have little effect on the muon, so we can treat the system as hydrogen-like,
even though there are multiple particles in the atom.

As a final comment, it is possible to replace both the nucleus and the electron, though the only case I am aware of is anti-hydrogen, which has an anti-proton and an
anti-electron. The energy levels, wave functions, etc., are all identical to that of
hydrogen.
G. Tables of Spherical Harmonics and Radial Wave Functions

\[ R_{nlm}(r, \theta, \phi) = R_n^l(r) Y_{l}^{m}(\theta, \phi) \]

Spherical Harmonics for \( l = 0, 1, 2, 3, 4: \)

\[ Y_0^0(\theta, \phi) = \frac{1}{\sqrt{\pi}} \]
\[ Y_0^2(\theta, \phi) = \frac{\sqrt{5}}{4\sqrt{\pi}} (3 \cos^2 \theta - 1) \]
\[ Y_1^0(\theta, \phi) = \frac{\sqrt{3}}{2\sqrt{\pi}} \cos \theta \]
\[ Y_2^\pm(\theta, \phi) = \mp \frac{\sqrt{15} e^{\pm i\phi}}{2\sqrt{2}\pi} \sin \theta \cos \theta \]
\[ Y_1^\pm(\theta, \phi) = \mp \frac{\sqrt{3} e^{\pm i\phi}}{2\sqrt{2}\pi} \sin \theta \]
\[ Y_2^0(\theta, \phi) = \frac{\sqrt{7}}{4\sqrt{\pi}} (5 \cos^3 \theta - 3 \cos \theta) \]
\[ Y_3^\pm(\theta, \phi) = \mp \frac{\sqrt{35} e^{\pm i\phi}}{8\sqrt{\pi}} \sin^3 \theta \cos \theta \]
\[ Y_3^\pm(\theta, \phi) = \mp \frac{\sqrt{35} e^{\pm i\phi}}{8\sqrt{\pi}} \sin^3 \theta \cos \theta \]
\[ Y_3^\pm(\theta, \phi) = \mp \frac{\sqrt{35} e^{\pm i\phi}}{8\sqrt{\pi}} \sin^3 \theta \cos \theta \]
\[ Y_4^\pm(\theta, \phi) = \mp \frac{3\sqrt{35} e^{\pm i\phi}}{16\sqrt{\pi}} \sin^4 \theta = Y_4^\pm(\theta, \phi) \]

Hydrogen Radial Wave Functions for \( n = 1, 2, 3, 4, 5: \) Where \( a_0 = \hbar^2/\mu k_e e^2 \)

\[ R_{10}(r) = \frac{2e^{-r/a_0}}{\sqrt{2}a_0} \]
\[ R_{20}(r) = \frac{e^{-r/2a_0}}{\sqrt{2}a_0^2} \left(1 - \frac{r}{2a_0}\right) \]
\[ R_{21}(r) = \frac{re^{-r/2a_0}}{\sqrt{2}a_0^3} \]
\[ R_{22}(r) = \frac{2\sqrt{2}r^2e^{-r/3a_0}}{81\sqrt{15}a_0^7} \]
\[ R_{23}(r) = \frac{2\sqrt{2}r^2e^{-r/4a_0}}{140625\sqrt{35}a_0^9} \]
\[ R_{30}(r) = \frac{2e^{-r/3a_0}}{3\sqrt{3}a_0^3} \left(1 - \frac{2r}{3a_0} + \frac{2r^2}{27a_0^2}\right) \]
\[ R_{31}(r) = \frac{4\sqrt{2}re^{-r/3a_0}}{27\sqrt{3}a_0^5} \left(1 - \frac{r}{6a_0}\right) \]
\[ R_{32}(r) = \frac{4\sqrt{2}re^{-r/3a_0}}{81\sqrt{15}a_0^7} \]
\[ R_{33}(r) = \frac{4\sqrt{2}re^{-r/4a_0}}{140625\sqrt{35}a_0^9} \]
\[ R_{40}(r) = \frac{e^{-r/4a_0}}{4\sqrt{a_0^3}} \left(1 - \frac{3r}{4a_0} + \frac{r^2}{8a_0^2} - \frac{r^3}{192a_0^3}\right) \]
\[ R_{41}(r) = \frac{\sqrt{5}re^{-r/4a_0}}{16\sqrt{3}a_0^5} \left(1 - \frac{r}{4a_0} + \frac{r^2}{80a_0^2}\right) \]
\[ R_{42}(r) = \frac{r^2e^{-r/4a_0}}{64\sqrt{5}a_0^7} \left(1 - \frac{r}{12a_0}\right) \]
\[ R_{43}(r) = \frac{r^2e^{-r/4a_0}}{768\sqrt{35}a_0^9} \]
\[ R_{44}(r) = \frac{8\sqrt{2}r^4e^{-r/5a_0}}{9375\sqrt{35}a_0^9} \left(1 - \frac{r}{20a_0}\right) \]
\[ R_{50}(r) = \frac{2e^{-r/5a_0}}{5\sqrt{5}a_0^9} \left(1 - \frac{4r}{5a_0} + \frac{4r^2}{25a_0^2} - \frac{4r^3}{375a_0^3} + \frac{2r^4}{9375a_0^4}\right) \]
Problems for Chapter 7

1. Derive the identities (7.10b), (7.10c), and (7.10d) using only the commutation relations (7.7).

2. For $j = 2$, we will work out the explicit form for all of the matrices $J$.
   (a) Write out the expression for $J_z$ and $J_\pm$ as an appropriately sized matrix.
   (b) Write out $J_x$ and $J_y$.
   (c) Check explicitly that $J^2 = J_z^2 + J_y^2 + J_x^2$ is a constant matrix with the appropriate value.

3. Show that the Pauli matrices, given by (7.17), satisfy
   (a) $\left(\sigma \cdot \hat{\mathbf{r}}\right)^2 = 1$ for any unit vector $\hat{\mathbf{r}}$, and
   (b) $\exp\left(-\frac{i}{2} \theta \sigma \cdot \hat{\mathbf{r}}\right) = \cos\left(\frac{1}{2} \theta\right) - i \sin\left(\frac{1}{2} \theta\right) \left(\sigma \cdot \hat{\mathbf{r}}\right)$.

4. Two particles have Hamiltonian $H_{\text{tot}} = P_1^2/2m_1 + P_2^2/2m_2 + V\left(|\mathbf{R}_1 - \mathbf{R}_2|\right)$, where the $\mathbf{P}$'s and $\mathbf{R}$'s have commutation relations
   \[
   \left[R_{ai}, P_{bj}\right] = i\hbar \delta_{ab} \delta_{ij}, \quad \left[R_{ai}, R_{bj}\right] = \left[P_{ai}, P_{bj}\right] = 0.
   \]
   (a) Show that if we define the four new vector operators $\{\mathbf{R}_{cm}, \mathbf{P}_{cm}, \mathbf{R}, \mathbf{P}\}$ as given in (7.38), they satisfy the commutation relations
   \[
   \left[R_{cmi}, P_{cmj}\right] = \left[R_i, P_j\right] = i\hbar \delta_{ij}, \quad \left[R_{cmi}, P_j\right] = \left[R_i, P_{cmj}\right] = 0.
   \]
   (b) Show that the Hamiltonian can then be written $H_{\text{tot}} = P_{cm}^2/2M + P^2/2\mu + V\left(|\mathbf{R}|\right)$, where $M$ and $\mu$ are given by (7.39).

5. It is often important to find expectations values of operators like $R_i$, which when acting on a wave function $\psi$ yields one of the quantities $\{x, y, z\}$.
   (a) Write each of the quantities $\{x, y, z\}$ in spherical coordinates, and then show how each of them can be written as $r$ times some linear combination of spherical harmonics. I recommend against trying to “derive” them, just try looking for expressions similar to what you want.
   (b) Show that the six quantities $\{x^2, y^2, z^2, xy, xz, yz\}$ can similarly be written as $r^2$ times various combinations of spherical harmonics. There should not be any products or powers of spherical harmonics, so you can’t derive them from part (a).
6. Consider the spherical harmonic oscillator, \( H = \mathbf{P}^2/2m + \frac{1}{2} m \omega^2 \mathbf{R}^2 \). This potential is most easily solved by separation of variables, but it is instructive to take advantage of the spherical symmetry to find solutions.

(a) Factor eigenstates of this Hamiltonian into the form \( \psi(r, \theta, \phi) = R(r) Y^m_l(\theta, \phi) \). Find a differential equation satisfied by the radial wave function \( R(r) \).

(b) At large \( r \), which term besides the derivative term dominates? Show that for large \( r \), we can satisfy the differential equation to leading order if \( R(r) \sim \exp(-\frac{1}{2} A r^2) \), and determine the factor \( A \) that will make this work.

(c) Write the radial wave function in the form \( R(r) = f(r) \exp(-\frac{1}{2} A r^2) \), and show that \( f \) must satisfy
\[
\frac{2mE}{\hbar^2} f = -\frac{1}{r} \frac{d}{dr} \left( r f' \right) + 2A \frac{d}{dr} \left( r f' \right) + A f + \frac{l^2 + l}{r^2} f.
\]

(d) Assume that at small \( r \), the wave function goes like \( f(r) \sim r^k \). What value of \( k \) will make this equation work?

(e) Assume that the radial wave function can be written as a power series, similar to what we did in class, \( f(r) = \sum_{i=0}^{n} f_i r^i \). Substitute this into the differential equation for \( f \), and thereby discover a recursion relation on the \( f_i \)'s. Unlike the recursion relationship we found, you will get a recursion relationship relating \( f_i \) to \( f_{i+2} \). Hence the series actually requires only odd powers of \( r \) or even powers of \( r \), not both.

(f) Assume, as in class, that the series terminates, so that \( f_n \) is the last term, and hence that \( f_{n+2} \) vanishes. Find a condition for the energy \( E \) in terms of \( n \).

(g) Given \( n \), which describes the energy of the system, what restrictions are there on \( l \), the total angular momentum quantum number?
VIII. Spin and Adding Angular Momentum

A. Rotations revisited

Up until now, we have been making a simplifying assumption, and it is time to reexamine this assumption and see if it is justified. In chapter 1, section B, just before we derived Schrödinger’s equation, we explicitly stated that our wave function $\Psi$ has no components. It was pointed out that other examples of fields, such as the electric or magnetic fields, do have components. We then went on to compound our error in chapter 6, when discussing symmetry, by stating that the effects of rotation on a wave function are simply to rotate the coordinates, so that $R(\mathcal{R})\psi(r) = \psi(\mathcal{R}^T r)$. What could seem more reasonable? After all, this was derived simply by assuming that the states $|r\rangle$ form a basis for our wave functions, and that $R(\mathcal{R})|r\rangle = |\mathcal{R}r\rangle$.

Let’s look at a simple example from electromagnetism and see if these assumptions are justified. Consider the magnetic field caused by a current-carrying wire coming out of the plane of view, as illustrated in Fig. 8-1. The problem is clearly rotationally invariant about the wire, and not surprisingly, the resulting magnetic fields are rotationally invariant as well. You can just see that they are. However, suppose we perform a 90-degree counter-clockwise rotation of these magnetic fields, and define the naively rotated magnetic field $B'(r)$ by the equation

$$B'(r) \equiv R(\mathcal{R})B(r) = B(\mathcal{R}^T r). \quad (8.1)$$

The problem is that when we perform a rotation we must not only rotate the coordinate $r$, we must rotate the magnetic field as well. The correct formula is

$$B'(r) \equiv R(\mathcal{R})B(r) = \mathcal{R}B(\mathcal{R}^T r). \quad (8.2)$$

Which of the two formulas, (8.1) or (8.2), is appropriate to describe the rotation of a quantum mechanical wave function? Perhaps neither. There are other possibilities, even for classical fields. We must consider all possibilities. In general, if the wave function $\psi(r)$ has multiple components, like the magnetic field, we must consider the possibility that

$$R(\mathcal{R})\psi(r) = D(\mathcal{R})\psi(\mathcal{R}^T r), \quad (8.3)$$

Figure 8-1: The magnetic field (green arrows) from a current-carrying wire are clearly invariant under rotation. But if a counterclockwise rotation by 90 degrees is performed (large black arrow), and the magnetic field is only moved by this amount, it will change the field (blue arrow). The field itself must also be rotated the same way (small black arrow) to produce an invariant magnetic field (red arrow).
where \( D(\mathcal{R}) \) is an as-yet unspecified finite dimensional matrix. The only constraint we want to have is that the product of two rotation operators should be the same as the combination of rotations, equation (7.28b), which we apply here to an arbitrary wave function

\[
R(\mathcal{R}_1)R(\mathcal{R}_2)\psi(r) = R(\mathcal{R}_1^\mathcal{R}_2)\psi(r),
\]

\[
R(\mathcal{R}_1)[D(\mathcal{R}_2)]\psi(\mathcal{R}_2^T r) = D(\mathcal{R}_1^\mathcal{R}_2)\psi\left(\mathcal{R}_2^T\mathcal{R}_1^T r\right).
\]

\[
D(\mathcal{R}_1)D(\mathcal{R}_2)\psi(\mathcal{R}_2^T\mathcal{R}_1^T r) = D(\mathcal{R}_1^\mathcal{R}_2)\psi(\mathcal{R}_2^T\mathcal{R}_1^T r),
\]

\[
D(\mathcal{R}_1)D(\mathcal{R}_2) = D(\mathcal{R}_1^\mathcal{R}_2).
\]

(8.4)

Our goal now is to work out all possible matrices \( D(\mathcal{R}) \) that satisfy (8.4).

We will proceed with the same technique we used in the last two chapters when studying continuous symmetries. Consider very small rotations. Naturally, when we perform no rotation at all, we will have \( D(1) = 1 \). For small rotations, we define the spin matrices \( S \) in terms of the derivative of \( D \) at small angles, so we have for small angles

\[
D(\mathcal{R}(\hat{\mathcal{R}},\theta)) = 1 - \frac{i\theta}{\hbar} \cdot S + \mathcal{O}(\theta^2).
\]

In a manner identical to what we did before, we discover

\[
D(\mathcal{R}(\hat{\mathcal{R}},\theta)) = \exp(-i\hat{\mathcal{R}} \cdot S/\hbar).
\]

(8.5)

In chapter 7, section A, we demonstrated using only that \( R(\mathcal{R}_1)R(\mathcal{R}_2) = R(\mathcal{R}_1^\mathcal{R}_2) \) and \( R(\mathcal{R}(\hat{\mathcal{R}},\theta)) = \exp(-i\hat{\mathcal{R}} \cdot L/\hbar) \) that the three angular momentum operators \( L \) satisfy the appropriate commutation relations. We can copy that proof exactly to demonstrate using only (8.4) and (8.5) that the spin matrices \( S \) satisfy exactly the same identities, so

\[
[S_i, S_j] = i\hbar \sum_k \epsilon_{ijk} S_k.
\]

In chapter 7, section B, we worked out all possible matrix representations that satisfy such commutation relations. We found that the “states” can be labeled by a pair of numbers, which we now call \( |s, m_s\rangle \)

\[
S_z |s, m_s\rangle = \hbar m_s |s, m_s\rangle.
\]

Then the matrices \( S \) work out to be \( 2s + 1 \) dimensional matrices, so our wave function \( \psi(r) \) has \( 2s + 1 \) components. We could label them by their eigenvalues; for example, as

\[
^1 \text{The second label is called} \ m_s \text{so that we may distinguish it from other angular momentum labels; for example,} \ m_l \text{will be the labels for the eigenstates of} \ L_z.
\]
ψ_{s,m_s}(\mathbf{r})$; however, all components have the same $s$ value, and hence we simply drop the $s$ index and write $ψ_{m_s}(\mathbf{r})$.

When discussing arbitrary angular momentum $\mathbf{J}$, we argued that the total angular momentum quantum number $j$ must be integer or half integer, but when we specialized to the operators $\mathbf{L}$, we argued ultimately that only integers work. Which is true of $s$, the quantum number corresponding to the operator $2S$? Suppose, for example, we chose $s = \frac{1}{2}$, then consider making a rotation by $2\pi$ in any direction. It is not hard to show using (8.5) and the specific forms of the Pauli matrices (7.17) that for an arbitrary angle $\theta$ we have

$$D(\mathbf{R}(\hat{\mathbf{r}},\theta)) = \exp(-i\theta \hat{\mathbf{r}} \cdot \mathbf{S}/\hbar) = \exp\left(-\frac{i}{2} \theta \hat{\mathbf{r}} \cdot \mathbf{σ}ight) = \cos\left(\frac{i}{2} \theta\right) - i \sin\left(\frac{i}{2} \theta\right) (\hat{\mathbf{r}} \cdot \mathbf{σ}).$$

Then for a $2\pi$ rotation we would have $D(\mathbf{R}(\hat{\mathbf{r}},\theta)) = -1$, which implies that if we rotate a system by $2\pi$, the wave function could undergo the transformation $|\Psi⟩ → -|\Psi⟩$. This seems wrong, so we might believe that once again we should always choose $s$ to be an integer. However, the wave functions $|\Psi⟩$ and $-|\Psi⟩$ differ only by a phase, and hence are physically identical. So in fact we have no reason to prefer one to the other. Hence it is an experimental question what the total spin $s$ of any particular particle is. Indeed, it is experimentally observed that many particles, including the proton, neutron, and electron, have $s = \frac{1}{2}$, and hence $S^2 = \frac{1}{4}\hbar^2$. In general, the spin $s$ of a particle can take on any integer or half-integer value.

To summarize, in coordinate notation our wave function will look like

$$ψ(\mathbf{r}) = \begin{cases} ψ_s(\mathbf{r}) \\ ψ_{s-1}(\mathbf{r}) \\ \vdots \\ ψ_{-s}(\mathbf{r}) \end{cases}.$$

We would like to rewrite this in our now more comfortable bra and ket notation. We now recognize that the kets $|\mathbf{r}\rangle$ no longer form an adequate basis for our state space, because we must specify not only the location where we are evaluating the wave function, but also which component we are talking about, so we rewrite this basis as $|\mathbf{r},m_s⟩$, where $m_s$ is a discrete basis taking on only the $2s + 1$ values from $-s$ to $+s$. This basis is defined as eigenstates of $\mathbf{R}$ and $S_z$, so we have

$$\mathbf{R}|\mathbf{r},m_s⟩ = |\mathbf{r},m_s⟩ \text{ and } S_z|\mathbf{r},m_s⟩ = \hbar m_s |\mathbf{r},m_s⟩.$$  \hspace{1cm} (8.6)

We can write the effect of the other spin operators on this basis with the help of (7.11) and (7.12):

$$S_\pm |\mathbf{r},m_s⟩ = \hbar \sqrt{s^2 + m_s^2 + m_s} |\mathbf{r},m_s \pm 1⟩,$$

$$S_+ = \frac{1}{2}(S_+ + S_-), \quad S_- = \frac{1}{2}i(S_+ - S_-), \quad S^2 = \frac{\hbar^2}{4}(s^2 + s).$$  \hspace{1cm} (8.7)
A couple of notational comments are in order. In the special case $s = \frac{1}{2}$, by far the most common case we will be considering, the basis states should be logically labeled $|r, \pm \frac{1}{2}\rangle$, but to save time, this is normally abbreviated to $|r, \pm \rangle$. Sometimes we aren’t even interested in the spatial variation, in which case the basis just becomes $|\pm \rangle$. In this basis, the spin operators are given by $J = \frac{1}{2} \hbar \sigma$ and (7.17), which means that

$$S_z |\pm \rangle = \frac{\hbar}{2} |\mp \rangle, \quad S_+ |\pm \rangle = \pm \frac{i \hbar}{2} |\mp \rangle, \quad S_- |\pm \rangle = \pm \frac{i \hbar}{2} |\mp \rangle. \tag{8.8}$$

Of course, it was an arbitrary choice to diagonalize the matrix $S_z$. To designate the fact that we have made such a choice, it is common to write a subscript on the basis, so we call our basis states $|\pm_+ \rangle$ and $|\pm_- \rangle$. This is especially helpful when working with more than one basis. For example, one can show that the eigenstates of $S_x$ and $S_y$, which we would call $|\pm_x \rangle$ and $|\pm_y \rangle$ can be written in terms of the eigenstates of $S_z$ as

$$|\pm_x \rangle = \frac{1}{\sqrt{2}} (|+_z \rangle \pm |-_z \rangle) \quad \text{and} \quad |\pm_y \rangle = \frac{1}{\sqrt{2}} (|+_z \rangle \pm i |-_z \rangle).$$

It is easy to see from (8.6) and (8.7) that operators like the position operator commute with the spin operator. In fact, any operator that concerns itself only with coordinates will naturally commute with spin, so we have, in particular,

$$[\mathbf{R}, \mathbf{S}] = [V(\mathbf{R}), \mathbf{S}] = [\mathbf{P}, \mathbf{S}] = [\mathbf{L}, \mathbf{S}] = 0.$$

It is tempting to conclude that the Hamiltonian will commute with spin as well. For example, consider again the hydrogen atom, with Hamiltonian given by (7.40c). This Hamiltonian does commute with all components of $\mathbf{S}$. Hence when we diagonalize $H$, we can choose eigenstates that are eigenstates of the five commuting operators $\{H, \mathbf{L}^2, \mathbf{L}_z, \mathbf{S}^2, S_z\}$. Since we are dealing with an electron, we simply have $\mathbf{S}^2 = \frac{1}{2} \hbar^2$, so there is no point labeling our quantum states this way, and we write them as $|n, l, m_s, m_s \rangle$. The energies will be given by (7.51), and the only effect of the new index $m_s$ is that the number of quantum states with any given value of $n$ has doubled from $n^2$ to $2n^2$, corresponding to the two possible values $m_s = \pm \frac{1}{2}$.

Although this works for the Hamiltonian (7.40c), it will not necessarily work in general. It is possible for the spin of the electron $\mathbf{S}$ to play a role in the Hamiltonian, and then spin will not necessarily commute with $H$, and this simple analysis will not generalize. Indeed, real hydrogen has small interactions between the spin and the orbit (called spin-orbit coupling), which means small additional terms in the Hamiltonian beyond (7.40c), and these do indeed ruin the simple analysis we have just completed. This brings us logically to our next subject.

B. Total Angular Momentum and Addition of Angular Momentum

Until this chapter we have termed $\mathbf{L}$ the angular momentum, but I have tried to avoid using this term in this chapter, because we want to reinvestigate this nomenclature.
Our starting point is equation (8.3). Rewriting this in our more familiar ket notation, we could write this as

$$R\left(\mathcal{R}(\hat{r},\theta)\right)|\psi\rangle = D\left(\mathcal{R}(\hat{r},\theta)\right)\exp\left(-i\hat{r}\cdot\mathbf{L}/\hbar\right)|\psi\rangle,$$

(8.9)

because we know that the factor $\exp\left(-i\hat{r}\cdot\mathbf{L}/\hbar\right)$ rotates the coordinate $\mathbf{r}$, and $D\left(\mathcal{R}(\hat{r},\theta)\right)$ rotates the components of the wave function. Substituting (8.5) and dropping the common ket $|\psi\rangle$, we see that

$$R\left(\mathcal{R}(\hat{r},\theta)\right) = \exp\left(-i\hat{r}\cdot\mathbf{S}/\hbar\right)\exp\left(-i\hat{r}\cdot\mathbf{L}/\hbar\right) = \exp\left[-i\hat{r}\cdot(\mathbf{S} + \mathbf{L})/\hbar\right],$$

$$R\left(\mathcal{R}(\hat{r},\theta)\right) = \exp\left(-i\hat{r}\cdot\mathbf{J}/\hbar\right),$$

(8.10)

where we have used the fact that $\mathbf{S}$ and $\mathbf{L}$ commute to simplify our expression, and we define the total angular momentum operator $\mathbf{J}$ as

$$\mathbf{J} \equiv \mathbf{L} + \mathbf{S}.\quad (8.11)$$

It is clear from (8.10) that neither $\mathbf{S}$ nor $\mathbf{L}$ is the actual generator of rotations, but rather $\mathbf{J}$. The operator $\mathbf{L}$ is the orbital angular momentum operator. The physical interpretation is easy to understand: $\mathbf{L}$ measures the angular momentum of one object circling another, while $\mathbf{S}$ measures the internal angular momentum of the object itself, as illustrated in Fig. 8-2.

Which of these operators would we expect to commute with the Hamiltonian, and hence be conserved? For the specific Hamiltonian we have already discussed, they all do. But in general we would expect the laws of physics to remain invariant when we rotate the whole system, and rotations are generated by $\mathbf{J}$. Hence when angular momentum commutes with the Hamiltonian, we expect it to be $\mathbf{J}$ that commutes, not $\mathbf{L}$ or $\mathbf{S}$.

We now turn our discussion to the addition of angular momentum, but to keep our discussion as general as possible, we will generalize our nomenclature. Suppose we have two sets of angular momentum operators, $\mathbf{J}_1$ and $\mathbf{J}_2$, which each satisfy appropriate commutation relations, but also commute with each other, so

$$\left[J_{1i}, J_{1j}\right] = i\hbar \sum_k \epsilon_{ijk} J_{1k}, \quad \left[J_{2i}, J_{2j}\right] = i\hbar \sum_k \epsilon_{ijk} J_{2k}, \quad \left[J_{1i}, J_{2j}\right] = 0.$$

It is then trivial to see that the four operators $\left\{J_{1z}, J_{2z}, J_{1\perp}, J_{2\perp}\right\}$ will all commute with each other. We therefore can choose basis states $|j_1, j_2, m_1 m_2\rangle$ which are eigenstates of these operators, having eigenvalues

\hspace{1in}

1 Note that the relationship $\exp(A)\exp(B) = \exp(A + B)$ works only for operators that commute.
The eigenvalues $m_1$ and $m_2$ will take on values $m_a = j_a, j_a - 1, \ldots, -j_a$, so for fixed $j_1$ and $j_2$, this represents $(2j_1 + 1)(2j_2 + 1)$ basis states, and this is the dimensionality of our vector space.

Now define the total angular momentum $J$ by $J = J_1 + J_2$. It is then easy to see that $J$ will also be an angular momentum operator. Because our Hamiltonian $H$ might not commute with $J_1$ or $J_2$, it might be a good idea to change basis so we have eigenstates of some of the $J$ operators. It is not hard to show that the four operators

$$\{J_1^z, J_2^z, J_1^2, J_2^2, J_z\}$$

all commute with each other, and we therefore wish to change bases to a basis $|j_1,j_2; jm\rangle$, eigenvectors of these operators, which we will abbreviate as $|jm\rangle$, where the two new labels are defined by

$$J^2 |jm\rangle = \hbar^2 (j^2 + j) |jm\rangle, \quad J_z |jm\rangle = \hbar m |jm\rangle.$$

Given a set of $(2j_1 + 1)(2j_2 + 1)$ basis states $|j_1,j_2; m_1m_2\rangle$, what will be the allowed eigenvalues of the new basis states $|jm\rangle$? For fixed $j$, we know that $m$ must take on the values $m = j, j-1, \ldots, -j$. This will be the key to figuring out the answer.

Let us first note that the original basis states $|j_1,j_2; m_1m_2\rangle$ are also eigenstates of $J_z$, as we can see:

$$J_z |j_1,j_2; m_1m_2\rangle = (J_{1z} + J_{2z}) |j_1,j_2; m_1m_2\rangle = \hbar (m_1 + m_2) |j_1,j_2; m_1m_2\rangle.$$

Since we know the range of values for $m_1$ and $m_2$, it easily follows that we can figure out how many basis vectors have any given eigenvalue $m$ of $J_z$. This turns out to be sufficient to figure out what eigenvalues $j$ and $m$ are possible given only $j_1$ and $j_2$.

Rather than giving a general argument, consider a graphical proof which I will illustrate with the specific example $j_1 = 2$ and $j_2 = 1$. There will be a total of 15 basis vectors $|21; m_1m_2\rangle$, and we have plotted all possible allowed values of $(m_1, m_2)$ in Fig. 8-3. It is then an easy matter to see that $m$ will run from -3 to +3. There is exactly one basis state with $m = \pm 3$, then two with $m = \pm 2$, and three with $m = \pm 1$. Finally, there are three again with $m = 0$, as illustrated in Fig. 8-4.

**Figure 8-3**: The possible eigenvalues of $J_{z1}$ and $J_{z2}$ for $j_1 = 2$ and $j_2 = 1$. The resulting eigenvalues for $J_z$ can be read off of the diagonal (red) lines; for example, there are three basis states with $m = 0$ (the line passing through the origin) and only one for the largest value $m = 3$ (upper-right-most line).
Now, when we change to the new basis $|jm\rangle$, the number of basis states with any given $m$ value must remain the same, because the dimensionality of this subspace with this eigenvalue must be the same. Now, if we have some states with a specific $j$ value, then $m$ will go as high as $m = j$. Since we have no states with $m$ bigger than 3, we can’t allow $j$ bigger than 3, but we must have $j = 3$ for some of our states, so we can get one basis vector with $m = 3$. But if we have $j = 3$, we will have $m$ running from -3 to +3, which corresponds to seven basis states, illustrated by the seven states in the bottom row as illustrated in Fig. 8-4.

Now, this isn’t enough states total, so there must be some more $(j,m)$ values. If we eliminate the first row in Fig. 8-4, we see there are eight basis states left, and the highest remaining $m$-value is $m = +2$. By the same argument as before, this must come from states in the new basis where $j = 2$. Then we will have states with $j = 2$ and $m$ running from -2 to +2, the second row of states in Fig. 8-4. If we remove these, there are three states left, which corresponds to $j = 1$ and $m$ running from -1 to +1.

In summary, when we switch from the original basis to the new basis, if $j_1 = 2$ and $j_2 = 1$, we will get a new basis where $j$ takes on the values 1, 2, and 3, and then $m$ runs from $-j$ to $j$. How does this generalize to other $j_1$ and $j_2$ values? The largest value $m$ takes on (which corresponds to the largest $j$ value) is just $j = j_1 + j_2$. We then get successively smaller values of $j$, decreasing by one, until we reach the level where the columns stop getting taller, as illustrated in Fig. 8-4. This happens, as can be seen in Fig. 8-3, when we reach the corners of our rectangle, which occurs when $m = \pm |j_1 - j_2|$, so our smallest value of $j$ will be $j = |j_1 - j_2|$. In summary, our states $|jm\rangle$ are constrained by

$$j = |j_1 - j_2|, |j_1 - j_2| + 1, \ldots, j_1 + j_2 \quad \text{and} \quad m = j, j - 1, \ldots, -j.$$  \hspace{1cm} (8.12)

It is probably worth checking the dimensionality. In our original basis we had $(2j_1 + 1)(2j_2 + 1)$ basis states, and in our final basis the number of basis states is just $2j + 1$ for each of the allowed $j$ values, which means the dimensionality is

$$\sum_{j=|j_1-j_2|}^{j_1+j_2} \sum_{j=|j_1-j_2|}^{j_1+j_2} [(j+1)^2 - j^2] = (j_1 + j_2 + 1)^2 - |j_1 - j_2|^2 = (2j_1 + 1)(2j_2 + 1).$$

So, indeed, it works out.
C. Clebsch-Gordan Coefficients

Since the two bases both describe the same state, it will prove helpful to convert from one basis to the other. Since both bases are complete, we can write either basis in terms of the other, so that

\[
|jm\rangle = \sum_{m_1=-j}^{j} \sum_{m_2=-j}^{j} |j,j_z;m,m_2\rangle \langle j,j_z;m,m_2|jm\rangle \quad (8.13a)
\]

\[
|j,j_z;m,m_2\rangle = \sum_{j_1=|j-j_z|}^{j} \sum_{m_1=-j_1}^{j_1} |jm\rangle \langle jm|j,j_z;m,m_2\rangle \quad (8.13b)
\]

The coefficients \(\langle j,j_z;m,m_2|jm\rangle\) are called Clebsch-Gordan coefficients,\(^1\) (CG coefficients for short) and are useful in a variety of settings. I have included a Maple routine called “Clebsch” that computes them on my web page if you ever need them. There are several constraints that they must satisfy to be non-zero:

1. \(-j \leq m \leq j\), \(-j_1 \leq m_1 \leq j_1\), and \(-j_2 \leq m_2 \leq j_2\), by integers;
2. \(j_1 - j_2 \leq j < j_1 + j_2\), by integers; and
3. \(m = m_1 + m_2\).

In the first two conditions, when I use the phrase “by integers,” what I mean, for example, is that \(m\) must differ from \(-j\) and \(j\) by an integer amount. Condition (3) comes about because both bases are eigenstates of \(J_z\), and the \(J_z\) eigenvalue must match.

If you want to work them out by hand, the process is tedious, but not particularly difficult. We start by noting that \(J_z |jj\rangle = 0\), because you cannot raise this index any more. Taking the Hermitian conjugate of this relation, we find

\[
\langle jj |J_+|j,j_z;m,m_2\rangle = \langle jj |(J_+ + J_-)|j,j_z;m,m_2\rangle = \hbar \sqrt{j_1^2 + j_1 - m_1^2 + m_1} \langle jj |j,j_z;m_1-1,m_2\rangle + \hbar \sqrt{j_2^2 + j_2 - m_2^2 + m_2} \langle jj |j,j_z;m_1,m_2-1\rangle,
\]

\[
\langle jj |j,j_z;m_1-1,m_2\rangle = -\sqrt{\frac{j_2^2 + j_2 - m_2^2 + m_2}{j_1^2 + j_1 - m_1^2 + m_1}} \langle jj |j,j_z;m_1,m_2-1\rangle. \quad (8.14)
\]

Keeping in mind that these coefficients are non-zero only when \(j = m_1 + m_2\), it is easy to convince yourself that this allows one to find the ratio of all non-zero coefficients of the form \(\langle jj |j,j_z;m,m_2\rangle\). If we add the normalization condition (8.13a), we see that

\[
1 = \langle jj |jj\rangle = \sum_{m,m_2} \langle jj |j,j_z;m,m_2\rangle \langle j,j_z;m,m_2|jj\rangle = \sum_m |\langle jj |j,j_z;m_1,j - m_1\rangle|^2. \quad (8.15)
\]

This together with (8.14), completely determines \(\langle jj |j,j_z;m,m_2\rangle\) up to an overall phase, which is normally chosen so that \(\langle jj |j,j_z;j_1,j-j_1\rangle > 0\).

---

\(^1\) The order in which the arguments appear, as well as the way they are written, will differ slightly from source to source, and there may even be different phase conventions. As long as one applies their notation and phase conventions consistently, such details will never make a difference.
To determine the CG coefficients for other values of \( m \), we start by noting that

\[
\langle j, m-1|j_1, j_2; m, m_2\rangle = \langle jm|J_z|j, j_1, j_2; m, m_2\rangle = \frac{jm|J_z = j_1, j_2; m, m_2\rangle}{\hbar \sqrt{j^2 + j - m^2 - m}}.
\]

Using this, we have

\[
\langle j, m-1|j_1, j_2; m, m_2\rangle = \sqrt{\frac{j^2 + j_1^2 - m_1^2}{j^2 + j - m^2 + m}} \langle jm|j_1, j_2; m_1 + 1, m_2\rangle + \sqrt{\frac{j^2 + j_2^2 - m_2^2}{j^2 + j - m^2 + m}} \langle jm|j_1, j_2; m_1, m_2 + 1\rangle.
\]

(8.16)

In short, once we have worked out \( \langle jj|j_1, j_2; m, m_2\rangle \), we can work out \( \langle j, j-1|j_1, j_2; m, m_2\rangle \), then \( \langle j, j-2|j_1, j_2; m, m_2\rangle \), etc., until we have all of the CG coefficients.

The Clebsch-Gordan coefficients, with the standard choices of phases, are all real, and also satisfy a variety of identities, some of which I list here:

\[
\langle j_2; j_1; m_2, m_1|jm\rangle = (-1)^{j_2, j_1 - j} \langle j_1, j_2; m_1, m_2|jm\rangle,
\]

(8.17a)

\[
\langle j_1, j_2; -m_1, -m_2|jm\rangle = (-1)^{j_1, j_2 - j} \langle j_1, j_2; m_1, m_2|jm\rangle,
\]

(8.17b)

\[
\langle 00|jm\rangle = \delta_{j0}\delta_{m0} = 1.
\]

(8.17c)

Except for the phase, which must be worked out, (8.17a) basically points out that writing \( J = J_1 + J_2 \) is the same as \( J = J_2 + J_1 \). Equation (8.17c) is just the recognition that if \( J_i = 0 \) for either of the \( J_i \)'s, then adding angular momentum is trivial. Most tables will only list the coefficients for \( j_1 \geq j_2 > 0 \) and \( m \geq 0 \); you must then use the relations (8.17) to get the rest. It is also easy to show with the help of (8.17a) that when you combine two equal angular momenta, \( j_1 = j_2 \), the CG coefficients are symmetric under interchange \( m_1 \leftrightarrow m_2 \) when \( 2j_1 - j \) is even, and antisymmetric when \( 2j_1 - j \) is odd.

The CG coefficients for \( j_2 = \frac{1}{2} \) are especially useful, since it is common that we are adding spin \( \frac{1}{2} \) to an arbitrary spin. In this case we have \( j = j_1 \pm \frac{1}{2} \), and \( m = \pm \frac{1}{2} \). The non-vanishing CG coefficients turn out to be:

\[
\langle j_1, \frac{1}{2}; m-\frac{1}{2}, \frac{1}{2}|j_1, \frac{1}{2}, m\rangle = \langle j_1, \frac{1}{2}; m+\frac{1}{2}, -\frac{1}{2}|j_1, -\frac{1}{2}, m\rangle = \frac{j_1 + \frac{1}{2} + m}{2j_1 + 1},
\]

(8.18)

To try to clarify what is going on, let’s assume we’re adding angular momentum \( L \) to spin \( S \). We’ll drop the spin label \( s = j_2 = \frac{1}{2} \) that appears in the original basis, and use (8.18) and (8.13a) to write our new basis in terms of the old. Finally, we’ll abbreviate \( m_2 = m_s = \pm \frac{1}{2} \) as \( \pm \), while clarifying \( j_1 = l \) and \( m = m_j \). We have
\[ |l \pm \frac{1}{2}, m_j \rangle = \sqrt{\frac{l + \frac{1}{2} \pm m_j}{2l+1}} |l, m_j + \frac{1}{2}, \mp \rangle \pm \sqrt{\frac{l + \frac{1}{2} \pm m_j}{2l+1}} |l, m_j - \frac{1}{2}, \mp \rangle. \tag{8.19} \]

Note that since \( j \) is positive, we are forced to take the upper sign if \( l = 0 \).

For example, in section A we labeled our states of hydrogen \( |n, l, m, m_s \rangle \). We could instead label them \( |n, l, j, m_j \rangle \), where the last two indices refer to the eigenvalues under \( J^2 \) and \( J_z \). Our new states in terms of our old would be

\[ |n, l, j = l \pm \frac{1}{2}, m_j \rangle = \sqrt{\frac{l + \frac{1}{2} \pm m_j}{2l+1}} |n, l, m_j + \frac{1}{2}, - \rangle \pm \sqrt{\frac{l + \frac{1}{2} \pm m_j}{2l+1}} |n, l, m_j - \frac{1}{2}, + \rangle. \tag{8.20} \]

The states with different \( m_j \) values would be expected to have exactly the same energy, since they are related by the rotation group, even if we add relativistic correction and so on, so the basis states (8.20) are better to work with than our original basis.

Because the states are exactly degenerate, it is common to refer to the levels of hydrogen or other atoms only by giving the eigenvalues of \( n, l, \) and \( j \), without giving \( m_j \). A sensible way to do so would be to write something like \( (n, l, j) \), but that would be too straightforward, so the standard nomenclature has been to denote \( n \) by an integer, \( l \) by a letter, and \( j \) by a subscript. The letters corresponding to \( l = 0, 1, 2, 3, 4 \) etc. are \( s, p, d, f, g \), with the letters continuing alphabetically thereafter, but skipping \( j \).

So a typical state might be called \( 4d_5/2 \), for example.

As a final example, consider the case of adding two spin-\( \frac{1}{2} \) states, which we can quickly figure out from (8.19). We suppress \( l = \frac{1}{2} \) and abbreviate \( m_l \) as \( \pm \). The total angular momentum will be 1 or 0, and the four states work out to

\[ |1, +1 \rangle = |++ \rangle, \quad |1, 0 \rangle = \frac{1}{\sqrt{2}} (|+- \rangle + |-- \rangle), \quad |1, -1 \rangle = |-- \rangle, \quad |0, 0 \rangle = \frac{1}{\sqrt{2}} (|+- \rangle - |-- \rangle). \tag{8.21} \]

D. Scalars, Vectors, Tensors

Consider, for the moment, an operator \( S \) that is unchanged under rotations; that is,

\[ R^\dagger (\mathcal{R}) S R (\mathcal{R}) = S. \]

Such an operator is called a scalar operator. Using (8.10), it is easy to see that this implies \( [\mathbf{J}, S] = 0 \). We want to work out a similar relationship for vector operators. We start with a generalized version of (6.15) or (6.16). We define a vector operator \( \mathbf{V} \) as a set of three operators \( \mathbf{V} = (V_x, V_y, V_z) \) which satisfy \( R^\dagger (\mathcal{R}) \mathbf{V} R (\mathcal{R}) = \mathcal{R} \mathbf{V} \). In components, this equation looks like

\footnote{This system breaks down at \( l = 12 \), which should correspond to \( p \) again, but I suspect these are normally handled by stopping the silly letter notation and denoting them by \( l \) again.}
\[ R^l (\mathcal{R}) V_j R (\mathcal{R}) = \sum_k \mathcal{R}_{jk} V_k. \]  

(8.22)

Let's figure out what commutation relations this must satisfy with the generators \( \mathbf{J} \) of rotations. To do so, consider a small rotation. The rotation matrix \( \mathcal{R}_{ij} \) for small angles can be easily worked out with the help of (6.12). To leading order it is the identity matrix, and to first order in rotations we have

\[ \mathcal{R}_{jk}(\hat{\mathbf{r}}, \theta) = \delta_{jk} - \theta \sum_i \hat{\mathbf{r}}_i \varepsilon_{ijk}. \]

Then using (8.10) in equation (8.22), and expanding to leading order in \( \theta \), we have

\[ (1 + i \theta \hat{\mathbf{r}} \cdot \mathbf{J}/\hbar) V_j (1 - i \theta \hat{\mathbf{r}} \cdot \mathbf{J}/\hbar) = \sum_k \left( \delta_{jk} - \theta \sum_i \hat{\mathbf{r}}_i \varepsilon_{ijk} \right) V_k, \]

\[ V_j + \frac{i \theta}{\hbar} \left[ \hat{\mathbf{r}} \cdot \mathbf{J}, V_j \right] = V_j - \theta \sum_{i,k} \hat{\mathbf{r}}_i \varepsilon_{ijk} V_k, \]

\[ \sum_i \hat{\mathbf{r}}_i [J_i, V_j] = i\hbar \sum_{i,k} \hat{\mathbf{r}}_i \varepsilon_{ijk} V_k. \]

The only way this sum can be true for any direction \( \hat{\mathbf{r}} \) is if the expression multiplying it is identical on the two sides, so

\[ [J_i, V_j] = i\hbar \sum_k \varepsilon_{ijk} V_k \]  

(8.23)

It is possible to generalize beyond vectors to tensors. A tensor operator is defined by

\[ R^l (\mathcal{R}) T_{jk} R (\mathcal{R}) = \sum_{l,m} \mathcal{R}_{jl} \mathcal{R}_{km} T_{lm}. \]

Finding a tensor operator is easy. We simply multiply any two vector operators; for example, \( T_{jk} = P_{j} S_{k} \). It is then a straightforward manner to figure out the commutation relations of the angular momentum operators with a tensor; the result works out to

\[ [J_i, T_{jk}] = i\hbar \sum_m \left( \varepsilon_{ijm} T_{mk} + \varepsilon_{kim} T_{jm} \right). \]

We can continue and define objects with more than two indices. These generalized operators are all called “tensors”, and we refer to the rank of the tensor as the number of indices. Then scalar and vector operators are tensor operators of rank zero and one respectively, and the tensor operators \( T_{jk} \) is of rank two.

Now, back when you first learned about vectors, you learned about two ways of multiplying vectors, the dot product and the cross-product. Let \( \mathbf{V} \) and \( \mathbf{W} \) be two vector operators, then it is easy to demonstrate that \( \mathbf{V} \cdot \mathbf{W} \) and \( \mathbf{V} \times \mathbf{W} \) are new scalar and vector operators respectively. However, if you look at the tensor operator \( T_{jk} = V_j W_k \), it is clearly true that the dot and cross product are “contained” in this tensor product. Rather than working with a general tensor \( T_{jk} \), we would like to come up with a notation that allows us to avoid the scalar and vector parts (except when we want them) and work...
directly with those parts that are not the dot or cross product. The key to doing so is to define spherical tensors.

We will illustrate first the concept of a spherical tensor by using the example of a vector operator. Let \( V \) be any vector operator, then define three operators \( V_q \) for \( q = -1, 0, 1 \) by the relations

\[
V_0 \equiv V_x, \quad V_{\pm 1} \equiv \frac{1}{\sqrt{2}} (\mp V_x - i V_y).
\]

Obviously, \( V_q \) contains the same three operators as \( V \). The commutation relations of these three operators with \( J \) can be shown to be

\[
\left[ J, V_q \right] = \sum_{q' = -1}^{1} V_{q'} \langle 1q' \mid J \mid 1q \rangle.
\]

where \( \langle 1q' \mid J \mid 1q \rangle \) are the matrix elements of \( J \) in the case \( j = 1 \), given by (7.16). We call this a spherical tensor of rank 1. We now generalize the concept to a spherical tensor of rank \( k \), which consists of \( 2k + 1 \) operators \( T_q^{(k)} \) where \( q \) runs from \(-k\) to \( k \), defined by

\[
\left[ J, T_q^{(k)} \right] = \sum_{q' = -k}^{k} T_q^{(k)} \langle kq' \mid J \mid kq \rangle.
\]

For example, a scalar operator \( S \) is just a spherical tensor of rank 0, since \( \langle 00 \mid J \mid 00 \rangle = 0 \) and \( S \) commutes with \( J \). This equation is perhaps clearer if we write it out explicitly for the operators \( J_z \) and \( J_{\pm} \), for which we have

\[
\left[ J_z, T_q^{(k)} \right] = \hbar q T_q^{(k)}, \quad \left[ J_\pm, T_q^{(k)} \right] = \hbar \sqrt{k^2 + k - q^2 + q} T_q^{(k)}.
\]

It turns out that it isn’t too hard to build arbitrary rank spherical tensors out of vector operators; indeed, we can build spherical tensor operators out of any other order tensors. Let \( V_{q_1}^{(k_1)} \) and \( W_{q_2}^{(k_2)} \) be spherical tensors of rank \( k_1 \) and \( k_2 \) respectively. Then we can build a new spherical tensor \( T_q^{(k)} \) by taking appropriate linear combinations of \( V_{q_1}^{(k_1)} \) and \( W_{q_2}^{(k_2)} \). Specifically, let

\[
T_q^{(k)} = \sum_{q_1, q_2} V_{q_1}^{(k_1)} W_{q_2}^{(k_2)} \langle k, k_2, q_1, q_2 \mid kq \rangle,
\]

where the last factor is the CG coefficients. Then we can work out the commutator of \( J \) with \( T_q^{(k)} \):

\[
\left[ J, T_q^{(k)} \right] = \sum_{q_1, q_2} \left[ J, V_{q_1}^{(k_1)} \right] W_{q_2}^{(k_2)} \langle k, k_2, q_1, q_2 \mid kq \rangle
\]

\[
= \sum_{q_1, q_2} \left( \left[ J, V_{q_1}^{(k_1)} \right] W_{q_2}^{(k_2)} + V_{q_1}^{(k_1)} \left[ J, W_{q_2}^{(k_2)} \right] \right) \langle k, k_2, q_1, q_2 \mid kq \rangle.
\]
\[
[J, T_q^{(k)}] = \sum_{q_1} \sum_{q_2} \left( \sum_{q_1'} V_{q_1 q_1'}^{(k)} \langle k, q_1' | J | k, q_1 \rangle W^{(k_2)}_{q_2} + \sum_{q_2'} V_{q_2 q_2'}^{(k_2)} \langle k_2, q_2' | J | k_2, q_2 \rangle \right) \langle k, k_2 ; q_1, q_2 | kq \rangle
\]

\[
= \sum_{q_1} \sum_{q_2} \sum_{q_1'} V_{q_1 q_1'}^{(k)} W^{(k_2)}_{q_2} \left( \langle k, q_1' | k_2, q_2 \rangle \langle k, k_2 ; q_1, q_2 | kq \rangle + \langle k, q_1' | k_2, q_2 \rangle \langle k, k_2 ; q_1, q_2 | kq \rangle \right)
\]

\[
= \sum_{q_1} \sum_{q_2} \sum_{q_1'} V_{q_1 q_1'}^{(k)} W^{(k_2)}_{q_2} \left( \langle k, k_2 ; q_1, q_2 | J_1 | k, k_2 ; q_1, q_2 \rangle + \langle k, k_2 ; q_1, q_2 | J_2 | k, k_2 ; q_1, q_2 \rangle \right) \langle k, k_2 ; q_1, q_2 | kq \rangle
\]

\[
= \sum_{q_1} \sum_{q_2} V_{q_1 q_2}^{(k_2)} \langle k, k_2 ; q_1, q_2 | (J_1 + J_2) | kq \rangle = \sum_{q_1} \sum_{q_2} V_{q_1 q_2}^{(k_2)} \langle k, k_2 ; q_1, q_2 | J | kq \rangle.
\]

We now insert a complete set of states \( |k'q'\rangle \langle k'q'| \), but noting that \( J \) does not change the total quantum number \( k \), only the terms with \( k' = k \) will contribute. So we have

\[
[J, T_q^{(k)}] = \sum_{q_1} \sum_{q_2} \sum_{q_1'} V_{q_1 q_1'}^{(k)} W^{(k_2)}_{q_2} \langle k, k_2 ; q_1, q_2 | kq' \rangle \langle kq' | J | kq \rangle = \sum_{q_1} \sum_{q_2} T^{(k)}_{q_1 q_2} \langle kq' | J | kq \rangle.
\]

Hence we have shown that we can build new spherical tensors out of old ones. However, this technique only works if the CG coefficients don't vanish, which demands

\[
|k_1 - k_2| \leq k \leq k_1 + k_2.
\]

In particular, if we combine two vector operators, we can produce spherical tensors of rank 0, 1, or 2. The number of components that result are 1, 3, or 5, and correspond to the dot product, cross-product, and a third part which could be called the traceless symmetric part. For example, the scalar part is just

\[
T^{(0)}_0 = \sum_{q_1, q_2} V_{q_1 q_2}^{(1)} W_{q_2}^{(1)} \langle 11 ; q_1, q_2 | 00 \rangle = \frac{1}{\sqrt{3}} \left( V_{+1} W_{+1} + V_{-1} W_{+1} - V_0 W_0 \right)
\]

\[
= \frac{1}{\sqrt{3}} \left[ \frac{1}{2} \left( -V_x + i V_y \right) \left( W_x + i W_y \right) + \frac{1}{2} \left( V_x + i V_y \right) \left( -W_x + i W_y \right) - V_z W_z \right]
\]

\[
= -\frac{1}{\sqrt{3}} \left( V_x W_x + V_y W_y + V_z W_z \right) = -\frac{1}{\sqrt{3}} \mathbf{V} \cdot \mathbf{W}.
\]

Spherical tensors come up most often when calculating matrix elements. Any combination of vector operators can be turned into sums of spherical tensors, and hence if we can find the matrix elements of these spherical tensors, we are in good shape to find any matrix element we might be interested in. This brings us naturally to the Wigner-Eckart theorem, the subject of the next section.
E. The Wigner-Eckart Theorem

Suppose for the moment we are working in a basis \( |\alpha j m\rangle \) that are eigenstates of \( \mathbf{J}^2 \) and \( J_z \), where \( \mathbf{J} \) is an angular-momentum-like operator, with \( j \) and \( m \) the corresponding eigenvalues, and \( \alpha \) corresponding to any other labels of our basis vectors. We would like to know the matrix elements of some spherical tensor \( T_q^{(k)} \) between these states, so our goal is to find

\[
\langle \alpha j m | T_q^{(k)} | \alpha' j' m' \rangle. \tag{8.27}
\]

Because the \( T_q^{(k)} \)'s are spherical tensors, they satisfy (8.26). Using (8.26a), we see that

\[
\langle \alpha j m | \left( J_z T_q^{(k)} - T_q^{(k)} J_z \right) | \alpha' j' m' \rangle = \langle \alpha j m | \left[ J_z, T_q^{(k)} \right] | \alpha' j' m' \rangle,
\]

\[
(\hbar m - \hbar m')\langle \alpha j m | T_q^{(k)} | \alpha' j' m' \rangle = \hbar q \langle \alpha j m | T_q^{(k)} | \alpha' j' m' \rangle.
\]

We now see that the matrix element (8.27) is non-vanishing only when

\[
\langle \alpha j m | T_q^{(k)} | \alpha' j' m' \rangle \neq 0 \implies m = m' + q. \tag{8.28}
\]

We will now demonstrate that the matrix elements with \( m = j \) are all related. We use the fact that we cannot increase \( m \) any further, so \( \langle \alpha jj | J_+ = 0 \rangle \), so we have

\[
0 = \langle \alpha jj | J_q T_q^{(k)} | \alpha' j' m' \rangle = \langle \alpha jj | \left[ J_z, T_q^{(k)} \right] | \alpha' j' m' \rangle + \langle \alpha jj | T_q^{(k)} J_- | \alpha' j' m' \rangle
\]

\[
= \hbar \left( \sqrt{k^2 + k - q^2 + q} \langle \alpha jj | T_q^{(k)} |-j' m' \rangle + \sqrt{j^2 + j' - m'^2 + m} \langle \alpha jj | T_q^{(k)} | \alpha' j', m' - 1 \rangle \right),
\]

\[
\langle \alpha jj | T_q^{(k)} |-j' m' \rangle = -\frac{\sqrt{j^2 + j' - m'^2 + m}}{\sqrt{k^2 + k - q^2 + q}} \langle \alpha jj | T_q^{(k)} | \alpha' j', m' - 1 \rangle. \tag{8.29}
\]

Since each side is non-zero only if \( j = q + m' - 1 \), these relate all the non-vanishing expressions with \( m = j \).

To find relations for other values of \( m \), consider the following commutator:

\[
\langle \alpha j m | \left[ J_+, T_q^{(k)} \right] | \alpha' j' m' \rangle = \langle \alpha j m | \left( J_+ T_q^{(k)} - T_q^{(k)} J_+ \right) | \alpha' j' m' \rangle,
\]

\[
\hbar \sqrt{k^2 + k - q^2 - q} \langle \alpha j m | T_q^{(k)} |-j' m' \rangle = \hbar \sqrt{j^2 + j' - m'^2 - m} \langle \alpha j, m - 1 | T_q^{(k)} | \alpha' j', m' + 1 \rangle
\]

\[
-\hbar \sqrt{j^2 + j' - m'^2 - m} \langle \alpha j m | T_q^{(k)} | \alpha' j', m' + 1 \rangle,
\]

\[
\langle \alpha j, m - 1 | T_q^{(k)} | \alpha' j' m' \rangle = \sqrt{k^2 + k - q^2 - q \langle \alpha j m | T_q^{(k)} |-j' m' \rangle}
\]

\[
\sqrt{j^2 + j' - m'^2 - m} \langle \alpha j m | T_q^{(k)} | \alpha' j', m' + 1 \rangle \tag{8.30}
\]

Between (8.29), (8.30), and the constraint (8.28), we can demonstrate that we can relate all the non-zero matrix elements of the form (8.27).
If you compare the various equations relating these matrix elements, they are virtually identical to the relations for the CG-matrix elements. If you make the association $T^{(k)}_q |\alpha' j' m'\rangle \rightarrow |j' k; m' q\rangle$, and drop the index $\alpha$, then the equations (8.29) and (8.30) will become identical with (8.14) and (8.16) respectively, with the variable names relabeled, so $|j' k; m' q\rangle \leftrightarrow |j_i; j'; m; m_2\rangle$. Indeed, the two equations (8.14) and (8.16), together with the normalization condition (8.15) completely determines all the CG coefficients. The conclusion is that the matrix elements (8.27) must be identical with the CG coefficients, except for normalization. Hence they are proportional, so

$$\langle \alpha jm | T^{(k)}_q | \alpha' j' m'\rangle \propto \langle jm | j' k; m' q\rangle,$$

where the proportionality constant is independent of $m$, $q$, and $m'$. The proportionality constant can be written in different ways, but it turns out to make some sense to take a factor of $1/\sqrt{2j+1}$ out and write

$$\langle \alpha jm | T^{(k)}_q | \alpha' j' m'\rangle = \frac{\langle \alpha | T^{(k)}_q | \alpha' j' \rangle}{\sqrt{2j+1}} \langle jm | j' k; m' q\rangle. \quad (8.31)$$

The last factor is not a quantity which is somehow computed separately, but simply a symbol representing the unknown proportionality constant, and the notation simply signifies that this quantity is independent of $m$, $q$, and $m'$, though it will generally depend on $j$ and $j'$, and probably any other labels $\alpha$ and $\alpha'$ for our states, as well, of course, on which operator $T^{(k)}$ we are using. Equation (8.30) is the Wigner-Eckart Theorem; it tells us that all of these $(2j+1)(2j'+1)(2k+1)$ matrix elements can be computed if we know any one of them. Any non-zero $\langle jm | k' q; m' \rangle$ can be used to find the reduced matrix element $\langle \alpha j | T^{(k)}_q | \alpha' j' \rangle$, and then if we know the CG coefficients, we can immediately write down the rest of them. The choice of the factor of $1/\sqrt{2j+1}$ is somewhat arbitrary; however, when you include it, you can show that if $T^{(k)}_0$ is Hermitian (as it usually will be), then these reduced matrix elements will satisfy

$$\langle \alpha j | T^{(k)}_q | \alpha' j' \rangle = \langle \alpha' j' | T^{(k)}_q | \alpha j \rangle^*,$$

and hence look a great deal like ordinary matrix elements.

---

1 Most of this discussion has been taken from Sakurai; it should be noted that his equation (3.10.31) mistakenly has the wrong factor in the denominator. Also, this relationship is usually rewritten by changing $\langle jm | j' k; m' q\rangle \rightarrow \langle j' k; m' q | jm\rangle$; valid since the CG matrix elements are chosen real.
F. Products of Spherical Harmonics

As an application of the Wigner-Eckart Theorem, consider the product of two spherical harmonics, \( Y_{l_1}^{m_1}(\theta, \phi) \) and \( Y_{l_2}^{m_2}(\theta, \phi) \). Since this is a function only of angles, it must be writeable in terms of other spherical harmonics, so we have

\[
Y_{l_1}^{m_1}(\theta, \phi) Y_{l_2}^{m_2}(\theta, \phi) = \sum_{l,m} c_{lm} Y_l^m(\theta, \phi).
\]  

(8.32)

The constants \( c_{lm} \) can then be obtained by taking advantage of the orthonormality of the spherical harmonics, so

\[
c_{lm} = \int Y_{l_1}^{m_1}(\theta, \phi)^* Y_{l_2}^{m_2}(\theta, \phi) d\Omega.
\]

(8.33)

Integrals such as (8.33) commonly arise, for example, when calculating atomic transition rates.

How can we see this is an application of the Wigner-Eckart theorem? We rewrite the integral (8.33) as

\[
c_{lm} = \int Y_{l_1}^{m_1}(\theta, \phi)^* Y_{l_2}^{m_2}(\theta, \phi) d\Omega = \langle lm | Y_{l_2}^{m_2}(\Theta, \Phi) | l_1 m_1 \rangle.
\]

(8.34)

where \( \Theta \) and \( \Phi \) are the operators that multiply the wave function by \( \theta \) and \( \phi \) respectively. It is then a straightforward matter to show that \( Y_{l_2}^{m_2}(\Theta, \Phi) \) is a spherical tensor operator of rank \( L \), since it satisfies the commutation relations (8.26) (treating \( L \) as \( J \) here, since there is no spin involved). Then we can immediately apply the Wigner-Eckart theorem to (8.34) to obtain

\[
c_{lm} = \langle lm | Y_{l_2}^{m_2} | l_1 m_1 \rangle = \frac{\langle l | Y_{l_2}^{m_2} | l_1 \rangle}{\sqrt{2l+1}} \langle lm | l_2 | m_1 m_2 \rangle.
\]

(8.35)

It remains only to calculate the reduced matrix element \( \langle l | Y_{l_2}^{m_2} | l_1 \rangle \).

A trick, together with a little knowledge of the spherical harmonics, will allow us to finish the computation. First, use (8.35) to find the \( c_{lm} \)’s, which you then substitute back into (8.32). We have

\[
Y_{l_2}^{m_2}(\theta, \phi) Y_{l_1}^{m_1}(\theta, \phi) = \sum_{l,m} \frac{\langle l | Y_{l_2}^{m_2} | l_1 \rangle}{\sqrt{2l+1}} \langle lm | l_2 | m_1 m_2 \rangle Y_l^m(\theta, \phi).
\]

Now multiply this by the matrix elements \( \langle l_1 l_2 | m_1 m_2 | l' m' \rangle \) and sum over \( m_1 \) and \( m_2 \). The result is

\[
\sum_{m_1, m_2} \langle l_1 l_2 | m_1 m_2 | l' m' \rangle Y_{l_1}^{m_1}(\theta, \phi) Y_{l_2}^{m_2}(\theta, \phi)
\]

\[
= \sum_{m_1, m_2} \sum_{l,m} \frac{\langle l | Y_{l_2}^{m_2} | l_1 \rangle}{\sqrt{2l+1}} \langle lm | l_2 | m_1 m_2 \rangle \langle l_1 l_2 | m_1 m_2 | l' m' \rangle Y_l^m(\theta, \phi).
\]
On the right we now have a complete sum over all the states $|l_1 l_2; m_1 m_2\rangle$, which lets us collapse the sum over $m_1$ and $m_2$ into

$$
\sum_{m_1, m_2} \langle l_{12}; m_1 m_2 | l' m' \rangle Y_i^{m_1} (\theta, \phi) Y_i^{m_2} (\theta, \phi) = \sum_{l, m} \frac{\langle l \parallel Y_{l_1} \parallel l \rangle}{\sqrt{2l + 1}} \langle l m | l' m' \rangle Y_i^m (\theta, \phi)
$$

$$
= \frac{\langle l' \parallel Y_{l_1} \parallel l \rangle}{\sqrt{2l' + 1}} Y_{l'}^m (\theta, \phi).
$$

Dropping the now superfluous primes, we have

$$
\langle l \parallel Y_{l_1} \parallel l \rangle Y_i^m (\theta, \phi) = \sqrt{2l + 1} \sum_{m_1, m_2} \langle l_{12}; m_1 m_2 | l m \rangle Y_i^{m_1} (\theta, \phi) Y_i^{m_2} (\theta, \phi). \tag{8.36}
$$

We now evaluate both sides of this equation at $\theta = 0$. It turns out there is a very simple formula for the spherical harmonics at this point, which is

$$
Y_i^m (0, \phi) = \sqrt{\frac{2l + 1}{4\pi}} \delta_{m,0}. \tag{8.37}
$$

So in the double sum in (8.36), only a single term contributes. We find

$$
\langle l \parallel Y_{l_1} \parallel l \rangle = \sqrt{2l + 1} \langle l_{12}; 00 | l0 \rangle \sqrt{\frac{2l + 1}{4\pi}} \sqrt{\frac{2l + 1}{4\pi}},
$$

$$
\langle l \parallel Y_{l_1} \parallel l \rangle = \langle l_{12}; 00 | l0 \rangle \sqrt{\frac{(2l + 1)(2l + 1)}{4\pi}}.
$$

Substituting this back into (8.34), we have

$$
\int Y_i^m (\theta, \phi)^* Y_i^{m_1} (\theta, \phi) Y_i^{m_2} (\theta, \phi) d\Omega = \langle l m | l_{12}; m_1 m_2 \rangle \langle l_{12}; 00 | l0 \rangle \sqrt{\frac{(2l + 1)(2l + 1)}{4\pi (2l + 1)}}. \tag{8.38}
$$

Equation (8.38) has considerable practical value. It is worth noting under what conditions the results are non-zero. The resulting matrix elements will be non-zero only if $m_1 + m_2 = m$, $|l_1 - l_2| \leq l \leq l_1 + l_2$, and $l_1 + l_2 - l$ is even. The first two are obvious since otherwise the CG coefficients $\langle l m | l_{12}; m_1 m_2 \rangle$ will automatically vanish, the last one can be obtained by noting from (8.17b) that if $l_1 + l_2 - l$ is odd, then $\langle l_{12}; 00 | l0 \rangle = 0$. It can also be proven by parity arguments, since the spherical harmonics are multiplied by $(-1)^l$ under parity.

---

1 Note that the matrix element $\langle l m | l_{12}; m_1 m_2 \rangle$ is often written as $\langle l_{12}; m_1 m_2 | l m \rangle$ in this expression. This is permissible since it is real.
Problems for Chapter 8

1. In section A, we were searching for matrices $D(\mathcal{R})$ which satisfy
   
   $$D(\mathcal{R}_1)D(\mathcal{R}_2) = D(\mathcal{R}_1\mathcal{R}_2).$$ 

   One easy way to make this equation work out is to define
   
   $$D(\mathcal{R}) = \mathcal{R}.$$ 

   Our goal in this problem is to identify the spin.

   (a) Using the equations (6.12) and the definition of the spin matrices
   
   $$D(\mathcal{R}(\hat{r},\theta)) = 1 - i\theta \hat{r} \cdot \mathbf{S}/\hbar + O(\theta^2),$$

   work out the three spin matrices $\mathbf{S}$.

   (b) Find the eigenvalues of $S_z$. This should be enough for you to conjecture what
   value of $s$ this representation corresponds to.

   (c) Check explicitly that $S^2$ is a constant matrix with the appropriate value.

2. Suppose that $\mathbf{L}$ and $\mathbf{S}$ are two sets of angular momentum-like operators that commute
   with each other, so that $[L_i, S_j] = 0$. In this problem, you may assume from the
   commutation relations that it follows that $[\mathbf{L}^2, \mathbf{L}] = 0 = [\mathbf{S}^2, \mathbf{S}]$.

   (a) Define $\mathbf{J} = \mathbf{L} + \mathbf{S}$. Show that $\mathbf{J}$ is also an angular momentum-like operator. It
   follows automatically that $[\mathbf{J}^2, \mathbf{J}] = 0$.

   (b) Show that $[\mathbf{L}^2, \mathbf{J}] = 0 = [\mathbf{S}^2, \mathbf{J}]$.

   (c) Convince yourself (and me) that the four operators $\mathbf{J}^2$, $\mathbf{L}^2$, $\mathbf{S}^2$, and $J_z$ all
   commute with each other (this is six commutators in all).

   (d) Convince yourself that $L_z$ and $S_z$ do not commute with $\mathbf{J}^2$.

3. Suppose we have a particle with spin $\frac{1}{2}$. Let’s define the spin operator $S_\theta$ as

   $$S_\theta = \frac{1}{2} \hbar(\cos \theta \sigma_z + \sin \theta \sigma_x)$$

   In other words, we are measuring the spin of the particle along an axis that is at an
   angle $\theta$ compared to the $z$-axis

   (a) Verify that

   $$|+\theta\rangle = \cos \left(\frac{1}{2} \theta\right) |+z\rangle + \sin \left(\frac{1}{2} \theta\right) |-_z\rangle \quad \text{and} \quad |-_\theta\rangle = -\sin \left(\frac{1}{2} \theta\right) |+z\rangle + \cos \left(\frac{1}{2} \theta\right) |-_z\rangle$$

   are normalized eigenvectors of $S_\theta$, and determine their eigenvalues. In other
   words, you have to demonstrate 3 things, (i) they are normalized, (ii) they are
   eigenvectors, (iii) determine their eigenvalues.

   (b) Suppose a particle is initially in the state $|+_z\rangle$. If a subsequent measurement at
   angle $\theta$ is done, what is the probability the result will come out positive? What is
   the state immediately after the measurement is done?

   (c) After the measurement at angle $\theta$ yields a positive result, another measurement is
   done, this time at angle $\theta'$ is performed. What is the probability this time that the
   result comes out positive?
4. It is common to have to calculate matrix elements of the form
\[ \langle n', l', m'_l, m'_m | \mathbf{L} \cdot \mathbf{S} | n, l, m_l, m_m \rangle, \]
where \( \mathbf{L} \) and \( \mathbf{S} \) are the orbital and spin angular momenta respectively, and \( l, m_l, \) and \( m_s \) are quantum numbers corresponding to the operators \( L^2, L_z, \) and \( S_z, \) respectively (\( n \) represents some sort of radial quantum number).

(a) Show that \( \mathbf{L} \cdot \mathbf{S} \) can be written in a simple way in terms of \( L^2, S^2, \) and \( J^2 = (\mathbf{L} + \mathbf{S})^2. \) You may assume any commutation relations that you know are true about \( \mathbf{L} \) and \( \mathbf{S}, \) or that you proved in a previous problem.

(b) Show that the operator \( \mathbf{L} \cdot \mathbf{S} \) commutes with \( L^2, S^2, \) and \( J^2. \)

(c) A more intelligent basis to use would be eigenstates of \( L^2, S^2, J^2, \) and \( J_z, \) so our states would look like \( |n, l, j, m_j \rangle \) (the constant \( s \) is implied). Assuming our states are orthonormal, work out a simple formula for
\[ \langle n', l', j', m'_j | \mathbf{L} \cdot \mathbf{S} | n, l, j, m_j \rangle \]

(d) For arbitrary \( l = 0, 1, 2, \ldots \) and \( s = \frac{1}{2}, \) what are the possible values of \( j? \) Work out all matrix elements of the form above in this case.

5. Suppose we have two particles with spin \( \frac{1}{2}, \) and they are in the total spin-0 state,
\[ |00 \rangle = \frac{1}{\sqrt{2}} \left( |+z \rangle - |-z \rangle \right), \]
where we have introduced the subscripts \( z \) to make clear that the states are eigenstates of the two operators \( S^z_{1z} \) and \( S^z_{2z} \). Let’s define the spin operators \( S^z_{1\theta} \) and \( S^z_{2\theta} \) as \( S^z_{i\theta} = \frac{1}{2} \hbar (\cos \theta \sigma_i + \sin \theta \sigma_i), \) where \( i = 1, 2 \) simply denotes which of the two spins we are measuring.

(a) Show that the state can be written in the form
\[ |00 \rangle = \frac{1}{\sqrt{2}} \left( |+\theta \rangle - |-\theta \rangle \right), \]
where these states are eigenstates of the two operators \( S^z_{i\theta} \).

(b) Suppose you measure \( S^z_{1\theta} \), the spin of particle 1 at an angle \( \theta \). What is the probability that you get each of the possible results? In each case, what is the wave function afterwards?

(c) After completing the measurement in part (b), you perform a second measurement, at a different angle, on the other spin, \( S^z_{2\theta}. \) Show in each case that the probability of the second measurement yielding the same result as the first is
\[ P(m_1 = m_2) = \sin^2 \left[ \frac{1}{2} (\theta' - \theta) \right] \]

(d) Spin one is going to be measured on one of the two axes \( a \) and \( c \) sketched at right. Spin two is going to be measured on one of the two axes \( b \) and \( d \) sketched at right. What is the probability, as a percent, in all four combinations, that the spins match?
6. Suppose that $U$ and $V$ are vector operators each satisfying (8.23). Show that $S = U \cdot V$ is a scalar operator and $W = U \times V$ is a vector operator.

7. Let $V$ be any vector operator satisfying (8.23), and define $V_q$ for $q = -1, 0, 1$ as given by (8.24). Show that this results in a tensor of rank 1; that is, that (8.26) is satisfied by $V_q$ with $k = 1$.

8. Our goal in this problem is to find every non-vanishing matrix element for Hydrogen of the form $\langle 41m | R | 42m' \rangle$; that is, all matrix elements between 4d and 4p states.

(a) Find the matrix element $\langle 410 | R_0 | 420 \rangle$. It may be helpful to use the Maple routines that I have put online that allow you to calculate the radial integrals efficiently.

(b) Find the reduced matrix element $\langle 41 \| R \| 42 \rangle$

(c) Find all non-zero components of $\langle 41m | R_q | 42m' \rangle$. There should be nine non-zero ones (one of which you have from part (a)).

(d) To show that you understand how to do it, find $\langle 410 | X | 421 \rangle$.

9. The quadrupole operators are spherical tensors of rank 2; that is, a spherical tensor with $k = 2$. Its components are:

$$
T_{22}^{(2)} = \frac{1}{2} (X \pm i Y)^2, \quad T_{21}^{(2)} = \mp XZ - i YZ, \quad T_0^{(2)} = \sqrt{\frac{3}{2}} \left( 2Z^2 - X^2 - Y^2 \right)
$$

(a) Show that these operators either commute or anti-commute with parity, $\Pi$.

(b) To calculate electric quadrupole radiation, it is necessary to calculate matrix elements of the form $\langle \alpha m | T_q^{(2)} | \alpha' l m' \rangle$. Based on the Wigner Eckart theorem, what constraints can we put on $m'$, $m$, and $q$? What constraints can we put on $l$ and $l'$?

(c) Based on parity, what constraints can we put on $l$ and $l'$?

(d) Given $l'$, what values of $l$ are acceptable? List all acceptable values of $l$ for $l' = 0, 1, 2, 3, 4, 5$. 

146 © 2011, Eric D. Carlson
IX. Electromagnetic Forces

A. Electric Forces

In electromagnetism, electric fields cause a force \( \mathbf{F}(\mathbf{r}, t) = -e\mathbf{E}(\mathbf{r}, t) \) on an electron. For us to use our conventions, we must write this force as the gradient of a potential, \( \mathbf{F}(\mathbf{r}, t) = -\nabla V(\mathbf{r}, t) \). Can we make this work in quantum mechanics?

Our starting point is Faraday’s Law, one of Maxwell’s equations, which says

\[
\nabla \times \mathbf{E}(\mathbf{r}, t) + \frac{\partial}{\partial t} \mathbf{B}(\mathbf{r}, t) = 0. \tag{9.1}
\]

Let us, for now, consider only the case in which there are no magnetic fields, then (9.1) tells us that the electric field has no curl. It is a general theorem that functions that have no curl can be written as the gradient of a potential, so we can write

\[
\mathbf{E}(\mathbf{r}, t) = -\nabla U(\mathbf{r}, t). \tag{9.2}
\]

It follows immediately that we can write our potential \( V(\mathbf{r}, t) = -eU(\mathbf{r}, t) \), and we have

\[
H = \frac{1}{2m} \mathbf{p}^2 - eU(\mathbf{r}, t). \tag{9.2}
\]

For a particle has some other charge \( q \), we simply replace \( e \to -q \) in (9.2) and in all subsequent formulas.

Is the choice of the scalar potential \( U(\mathbf{r}, t) \) unique? The electric field \( \mathbf{E} \) is physically measurable, but \( U \) is not. Suppose that two different potential \( U_1 \) and \( U_2 \) produced the same electric field \( \mathbf{E} \), then it is not clear how we would determine which one of them is the “real” potential. So suppose we have

\[
\mathbf{E}(\mathbf{r}, t) = -\nabla U(\mathbf{r}, t) = -\nabla U'(\mathbf{r}, t). \tag{9.3}
\]

Then we easily see that

\[
0 = \nabla [U'(\mathbf{r}, t) - U(\mathbf{r}, t)].
\]

The only functions that have no gradient are functions that are independent of space. Hence \( U \) and \( U' \) can only differ by a function of time. For reasons that will soon become clear, let us define \( \chi(t) \) to be the function whose derivative is the difference between the two potentials, so

\[
U'(\mathbf{r}, t) = U(\mathbf{r}, t) - \frac{d\chi(t)}{dt}. \tag{9.3}
\]

\[1 \] Recall, here and throughout this chapter and subsequent chapters we will use the SI system of units.
There is no way, at least classically, to distinguish between these two potentials. Let’s look now at Schrödinger’s equation. Imagine we are given an electric field, and we start by deriving the scalar potential \( U(r, t) \). We now find a solution of Schrödinger’s equation using this potential, so we have

\[
\frac{i\hbar}{\partial t} \Psi(r, t) = -\frac{\hbar^2}{2m} \nabla^2 \Psi(r, t) - eU(r, t) \Psi(r, t)
\] (9.4)

Meanwhile, the physicist in the office next door is working on the same problem, and has found another potential \( U' \), related by (9.3) to \( U \). It is obvious looking at (9.4) that our wave function will not satisfy (9.4) if we simply change \( U \rightarrow U' \). But we expect, physically, for there to be no difference between these two potentials, since the electric field is the same. Hence there must be some simple way to find a new solution to (9.4). The secret is to modify our wave function so that (9.4) will still be satisfied. We want to do this by making a change to the wave function that depends on time, and not space, so that when we take the derivative we get an extra term that corresponds exactly to the extra term that comes from the scalar potential.

A bit of trial and error will lead you ultimately to the correct formula. Let us define

\[
\Psi'(r, t) = \Psi(r, t) \exp \left[ -\frac{ie}{\hbar} \chi(t) \right].
\] (9.5)

Let’s take the time derivative of this expression and see what we get.

\[
\frac{i\hbar}{\partial t} \Psi'(r, t) = \frac{i\hbar}{\partial t} \left[ \Psi(r, t) e^{-ie\chi(t)/\hbar} \right] = \frac{i\hbar}{\partial t} \Psi(r, t) e^{-ie\chi(t)/\hbar} + e\Psi(r, t) \frac{d\chi(t)}{dt} e^{-ie\chi(t)/\hbar}
\]

\[
= -\frac{\hbar^2}{2m} \nabla^2 \Psi(r, t) - eU(r, t) \Psi(r, t) + e\frac{d\chi(t)}{dt} \Psi(r, t) e^{-ie\chi(t)/\hbar}
\]

\[
= -\frac{\hbar^2}{2m} \nabla^2 \left[ \Psi(r, t) e^{-ie\chi(t)/\hbar} \right] + \left[ -eU(r, t) + e\frac{d\chi(t)}{dt} \right] \Psi(r, t) e^{-ie\chi(t)/\hbar},
\]

\[
\frac{i\hbar}{\partial t} \Psi'(r, t) = \frac{\hbar^2}{2m} \nabla^2 \Psi'(r, t) - eU'(r, t) \Psi'(r, t).
\]

In other words, if we perform the simultaneous change (9.3) and (9.5) Schrödinger’s equation will remain satisfied. Note that we have only multiplied the wave function by a phase in (9.5), so if we measured, for example, the position of the particle, the probability density \( |\Psi(r, t)|^2 \) would remain the same. Other operators, like the momentum \( \mathbf{P} \), will similarly be unaffected, since the phase change is independent of position. The “gauge transformation” (9.3) would be the same if we were working with particles with other charges, but (9.5) depends on the charge of the particle, with \( e \rightarrow -q \) for other charges.
B. Electromagnetic Fields and Forces

In the previous section, we took advantage of the absence of magnetic fields to write the electric field in terms of a gradient of a potential. As soon as we allow magnetic fields, a series of problems seem to arise. Obviously, we can no longer write our electric field so simply, but the magnetic fields themselves also present an obstacle. The Lorentz force on an electron in the presence of electric and magnetic fields is given by

$$ \mathbf{F}(\mathbf{r}, t) = -e \left[ \mathbf{E}(\mathbf{r}, t) + \mathbf{v} \times \mathbf{B}(\mathbf{r}, t) \right]. $$

The magnetic term does not resemble the gradient of a potential, so it is not immediately obvious how to proceed.

With the electric field, we started by writing the electric field as a gradient of a scalar potential. For the magnetic field, we recall that there are no magnetic monopoles for magnetic fields, so that we have

$$ \nabla \cdot \mathbf{B}(\mathbf{r}, t) = 0. $$

It is a general theorem that any vector function with zero divergence can be written as the curl of some other vector function. We therefore write

$$ \mathbf{B}(\mathbf{r}, t) = \nabla \times \mathbf{A}(\mathbf{r}, t), \tag{9.6} $$

where $\mathbf{A}(\mathbf{r}, t)$ is the vector potential.

Starting with the vector potential, it is not difficult to figure out how to write the electric field in terms of potentials. Substituting (9.6) into (9.1), we find

$$ 0 = \nabla \times \mathbf{E}(\mathbf{r}, t) + \frac{\partial}{\partial t} \nabla \times \mathbf{A}(\mathbf{r}, t) = \nabla \times \left[ \mathbf{E}(\mathbf{r}, t) + \frac{\partial}{\partial t} \mathbf{A}(\mathbf{r}, t) \right]. $$

Since the curl of the expression in square brackets is zero, it must be writeable in terms of some gradient. As before, we call this function $U$, so we have

$$ \mathbf{E}(\mathbf{r}, t) = -\frac{\partial}{\partial t} \mathbf{A}(\mathbf{r}, t) - \nabla U(\mathbf{r}, t). \tag{9.7} $$

Equations (9.6) and (9.7) then give the electric and magnetic fields in terms of the vector and scalar potentials $\mathbf{A}$ and $U$.

Not surprisingly, the choice of $\mathbf{A}$ and $U$ is not unique. For example, if we have two different vector potentials, $\mathbf{A}$ and $\mathbf{A}'$, that both yield the same magnetic field, it is easy to see from (9.6) that the curl of their difference must vanish, and therefore their difference can be written as the gradient of some function:

$$ \mathbf{A}'(\mathbf{r}, t) = \mathbf{A}(\mathbf{r}, t) + \nabla \chi(\mathbf{r}, t). \tag{9.8} $$

If we want to keep the electric field the same at the same time it will be necessary to modify the scalar potential $U$ as well, and we see from (9.7) that we want to have
\[ E(r,t) = -\frac{\partial}{\partial t} A(r,t) - \nabla U(r,t) = -\frac{\partial}{\partial t} A'(r,t) - \nabla U'(r,t), \]
\[ 0 = \nabla \left[ U'(r,t) - U(r,t) \right] + \frac{\partial}{\partial t} \left[ A'(r,t) - A(r,t) \right] \]
\[ = \nabla \left[ U'(r,t) - U(r,t) + \frac{\partial}{\partial t} \chi(r,t) \right], \]
\[ U'(r,t) = U(r,t) - \frac{\partial}{\partial t} \chi(r,t). \quad (9.9) \]

At the moment we don’t know what Schrödinger’s equation will look like, but we suspect it will look something like
\[ i\hbar \frac{\partial \Psi}{\partial t} = H\Psi = \frac{1}{2m} P^2 \Psi - eU\Psi. \quad (9.10) \]

Whatever it really looks like, we want to make sure that if we make the transformations (9.8) and (9.9), our wave function will still satisfy it. Of course, it is likely that our wave function will also have to make some sort of transformation, and looking at (9.5) it is easy to guess what that transformation will be. We now go ahead and define a \textit{gauge transformation} as the three equations (9.8), (9.9), and a generalized (9.5), summarized here for clarity:

\[ U'(r,t) = U(r,t) - \frac{\partial}{\partial t} \chi(r,t), \quad (9.11a) \]
\[ A'(r,t) = A(r,t) + \nabla \chi(r,t) \quad (9.11b) \]
\[ \Psi'(r,t) = \Psi(r,t) \exp \left[ -\frac{i e}{\hbar} \chi(r,t) \right]. \quad (9.11c) \]

What will Schrödinger’s equation look like? Clearly, (9.10) is not going to work. The basic problem is that (9.10) has the operator \( P \) in it, and if you perform the gauge transformation, the derivative \( P \) will act on not only the wave function but the extra phase factor as well. So we have
\[ P\Psi'(r,t) = -i\hbar \nabla \left\{ \Psi(r,t) \exp \left[ -\frac{i e}{\hbar} \chi(r,t) \right] \right\} \]
\[ = -i\hbar \left[ \nabla \Psi(r,t) \right] \exp \left[ -\frac{i e}{\hbar} \chi(r,t) \right] - e\Psi(r,t) \exp \left[ -\frac{i e}{\hbar} \chi(r,t) \right] \nabla \chi(r,t) \]
\[ = \exp \left[ -\frac{i e}{\hbar} \chi(r,t) \right] \left[ P - e \nabla \chi(r,t) \right] \Psi(r,t) \]
\[ = \exp \left[ -\frac{i e}{\hbar} \chi(r,t) \right] \left[ P - eA'(r,t) + eA(r,t) \right] \Psi(r,t), \]
\[ \left[ P + eA'(r,t) \right] \Psi'(r,t) = \exp \left[ -\frac{i e}{\hbar} \chi(r,t) \right] \left[ P + eA(r,t) \right] \Psi(r,t). \quad (9.12) \]

We now make an inspired choice. We define the \textit{kinematic momentum} operator
where the generalization \( e \rightarrow -q \) can be made for other charged particles as needed. Note that \( \pi \) depends on the choice of gauge, so if we changed \( A \) to \( A' \) we would have to change \( \pi \) to \( \pi' \). In terms of this new operator, (9.12) is just

\[
\pi'\Psi'(r,t) = \exp\left[ -\frac{ie}{\hbar} \chi(r,t) \right] \pi\Psi(r,t).
\]

and we can then easily proceed to prove

\[
\pi'^2\Psi'(r,t) = \exp\left[ -\frac{ie}{\hbar} \chi(r,t) \right] \pi^2\Psi(r,t).
\]

We are now ready to conjecture what Schrödinger’s equation is. We speculate that it is

\[
i\hbar \frac{\partial \Psi'(r,t)}{\partial t} = \frac{1}{2m} \pi'^2\Psi - eU\Psi = \frac{1}{2m} (\hbar \nabla + eA)^2 \Psi - eU\Psi.
\]  

(9.14)

At the moment, (9.14) is little more than an educated guess. Let’s first check that it is, in fact, invariant under a gauge transformation. We find

\[
i\hbar \frac{\partial \Psi'(r,t)}{\partial t} = i\hbar \left\{ \frac{\partial \Psi'(r,t)}{\partial t} \exp\left[-\frac{ie}{\hbar} \chi(r,t) \right] \right\}
\]

\[
= i\hbar \left\{ \frac{\partial \Psi'(r,t)}{\partial t} \right\} \exp\left[-\frac{ie}{\hbar} \chi(r,t) \right] + e\Psi'(r,t) \exp\left[-\frac{ie}{\hbar} \chi(r,t) \right] \frac{\partial \chi(r,t)}{\partial t}
\]

\[
= \left[ \frac{\pi'^2}{2m} \Psi'(r,t) - eU(r,t)\Psi'(r,t) + e \frac{\partial \chi(r,t)}{\partial t} \Psi'(r,t) \right] \exp\left[-\frac{ie}{\hbar} \chi(r,t) \right]
\]

\[
= \frac{\pi'^2}{2m} \Psi'(r,t) - e \left[U(r,t) - \frac{\partial \chi(r,t)}{\partial t} \right] \Psi'(r,t),
\]

\[
i\hbar \frac{\partial \Psi'(r,t)}{\partial t} = \frac{1}{2m} \pi'^2\Psi'(r,t) - eU'(r,t)\Psi'(r,t).
\]

So indeed, (9.14) is a gauge invariant Schrödinger’s equation, and our Hamiltonian is

\[
H = \frac{1}{2m} \pi^2 - eU.
\]  

(9.15)

In order to solve Schrödinger’s equation in the presence of electric or especially magnetic fields, it is necessary first to choose the explicit form of the vector potential \( A \) and scalar potential \( U \), and these two potentials are not uniquely defined. Deciding which of several (infinitely many) related potentials you choose is called a gauge choice. But even in ordinary physics, we often need to make arbitrary choices, such as coordinate axes. It is possible, with great effort, to create a formalism that does not refer to a specific set of coordinates in classical physics, but it is easier to not use such formalisms. It is similarly possible in quantum mechanics to formulate equations that are independent of gauge.
Making a gauge choice is a bit of an art. Like making a coordinate choice, the problem can be easier if you pick the right choice, and virtually intractable if you do not. It is generally a good idea to make a gauge choice that respects whatever symmetries may be present in the problem. For example, if the problem is invariant under translation in the $x$-direction, you should probably choose $U$ and $A$ to be independent of $x$. But such simple instructions do not always work. When in doubt, it is sometimes best to try two different gauges and see which one is most productive.

C. Working with Kinematic Momentum

It should be noted that we expect all experimental results to be independent of our choice of gauge, so that the results are unchanged when we perform a gauge transformation (9.11). For example, suppose we measure the momentum $P$. It is not hard to show that if you measure the expectation value of $P$, you will get different results in the two gauge choices. In contrast, if you measure the expectation value of $\pi$, you find

$$\langle \pi \rangle = \int d^3r \, \Psi^* (r, t) \, \pi \Psi (r, t) = \int d^3r \, \Psi^* (r, t) \, e^{i \omega x (r, t) / \hbar} \, e^{-i \omega x (r, t) / \hbar} \, \pi \Psi (r, t)$$

$$= \int d^3r \, \Psi^* (r, t) \, \pi \Psi (r, t) = \langle \pi \rangle.$$

Hence if you measure $\pi$, the result will be independent of your choice of gauge. This suggests that $\pi$ might be physically measurable; in contrast, $P$ must not be. In other words, **only gauge invariant operators can be measurable.**

It will prove necessary to have on hand commutation relations of the six operators $R$ and $\pi$. These work out to

\[ [R_i, R_j] = 0 \]
\[ [R_i, \pi_j] = [R_i, P_j + e A_j (R, t)] = [R_i, P_j] = i \hbar \delta_{ij} \]
\[ [\pi_i, \pi_j] = [P_i + e A_i (R, t), P_j + e A_j (R, t)] = e [P_i, A_j (R, t)] - e [P_j, A_i (R, t)] = -i e h \left( \frac{\partial}{\partial R_i} A_j (R, t) - \frac{\partial}{\partial R_j} A_i (R, t) \right), \]

\[ [\pi_i, \pi_j] = -i e h \sum_k \epsilon_{ijk} B_k (R, t) \]

where $B_k$ is the magnetic field.

We would like to now turn back to Ehrenfest’s theorem, equations (4.18), which describe how the expectation values of the momentum and position of a particle change with time. These were calculated using (4.15), the general formula for the time evolution of an operator’s expectation value. Applying this to the position $R$, we see that

\[ \frac{d}{dt} \langle R_i \rangle = \frac{i}{\hbar} \langle [H, R_i] \rangle = \frac{i}{\hbar} \frac{1}{2m} \left( [\pi^2, R_i] \right) - \frac{i e}{\hbar} \left( [U (R, t), R_i] \right), \]
\[
\frac{d}{dt} \langle R_i \rangle = \frac{i}{2m} \sum_j \left\langle \left[ \pi_j, [\pi_j, R_i] \right] \right\rangle + \left\langle \left[ \pi_j, R_i \right] \pi_j \right\rangle = -\frac{i(\hbar)}{2m} \sum_j \delta_{ij} \left[ \langle \pi_i \rangle + \langle \pi_i \rangle \right],
\]
\[
\frac{d}{dt} \langle R_i \rangle = \frac{\langle \pi_i \rangle}{m}.
\] (9.17)

Hence \( \pi \) is acting like momentum, since it tells you how fast the centroid of the wave packet is moving. Similarly,

\[
\frac{d}{dt} \langle \pi_i \rangle = \frac{i}{\hbar} \langle [H, \pi_i] \rangle = \frac{i}{\hbar} \frac{1}{2m} \left\langle \left[ \pi^2 + \pi_i \right] \right\rangle - \frac{ie}{\hbar} \left\langle \left[ U(R, t), \pi_i \right] \right\rangle - \frac{e}{\hbar} \frac{\partial}{\partial t} \langle \pi_i \rangle
\]

\[
= \frac{i}{2m} \sum_j \left\langle \left[ \pi_j, \pi_i \right] \right\rangle + \left\langle \left[ \pi_j, \pi_i \right] \pi_j \right\rangle - \frac{ie}{\hbar} \frac{\partial}{\partial R_i} U(R, t) + e \left\langle \frac{\partial}{\partial t} A_i(R, t) \right\rangle
\]

\[
= -\frac{i(\hbar)}{2m} \sum_j \sum_k \varepsilon_{jik} \left[ \langle \pi_j B_k + B_k \pi_i \rangle - e \langle E_i(R, t) \rangle \right],
\]

\[
\frac{d}{dt} \langle \pi_i \rangle = -\frac{e}{2m} \sum_j \sum_k \varepsilon_{jik} \left[ \langle \pi_j B_k - B_k \pi_i \rangle - e \langle E_i \rangle \right].
\] (9.18)

where in the last step we have taken advantage of the anti-symmetry of \( \varepsilon_{ijk} \) to rewrite a couple of terms to make it more obvious what is going on. We recognize the combination of \( \varepsilon_{ijk} \) with two vectors as a cross product, and so we write together now our equations (9.17) and (9.18) as

\[
\frac{d}{dt} \langle R \rangle = \frac{1}{m} \langle \pi \rangle,
\] (9.19a)

\[
\frac{d}{dt} \langle \pi \rangle = -\frac{e}{2m} \langle \pi \times B - B \times \pi \rangle - e \langle E \rangle.
\] (9.19b)

Equation (9.19b) is comparable to the classical Lorentz force \( dp/dt = F = -e(v \times B + E) \). Hence we see that the first terms in (9.19b) correspond to the magnetic force, and the last to the electric force.

D. Magnetic Dipole Moments

Consider, classically, a small spinning charged sphere with uniform mass density and charge density. Such an object has an angular momentum which can be computed by integrating the angular momentum over the sphere (see Fig. 9-1). Similarly, since the charge is moving, it has a current density associated with its rotation, which makes it act like a current loop. It is not difficult to show that there is a simple relationship between the angular momentum and the magnetic dipole moment, namely \( \mathbf{u} = qS/2m \), where \( S \) is the spin (internal angular momentum), \( m \) is the

Figure 9-1: A naive model of a spinning charged elementary particle gives a simple prediction for a relation between the spin and magnetic dipole moment.
mass, \( q \) is the charge. Of course, there is no reason to expect this estimate to be realistic, but it is not implausible that there will be such a magnetic dipole moment. In fact, using the Dirac equation, a relativistic version of Schrödinger’s equation, we can demonstrate that we should actually have exactly twice this for an electron. We will write \( \mu = gqS/2m \), where we expect approximately \( g = 2 \) for an electron. For other particles, such as the proton, the situation is more complicated, primarily because the proton is actually a combination of three quarks with different charges that are spinning in different directions (see Fig. 9-2). As a consequence, the relationship between charge and spin is not so simple, and experimentally \( g_p = 5.5857 \). To keep our discussion as general as possible, we will treat \( g \) as an unknown quantity and keep it in our equations.

Classically, a magnetic dipole in a magnetic field has a potential energy \( V = -\mu \cdot B \), and there is no reason to imagine that this doesn’t apply to an electron. So our Hamiltonian (9.15) must be modified accordingly:

\[
H = \frac{\pi^2}{2m} - eU + \frac{ge}{2m} B \cdot S = \frac{1}{2m} \left[ P + eA(R,t) \right]^2 - eU(R,t) + \frac{ge}{2m} B(R,t) \cdot S. \tag{9.20}
\]

The new term is manifestly gauge invariant, because \( B \) is gauge invariant. For particles with a different charge, we simply change \( e \rightarrow -q \).

The new term is important experimentally since it allows us an easy way to measure the spin of the particle. Rather than discussing this quantum mechanically, it is easier to think about the problem classically. When an electron enters a region of magnetic field, it acquires an additional energy \( geB(R,t) \cdot S/2m \). If the field is non-uniform, the particle will begin to accelerate in the direction which decreases its energy. The force will be given by

\[
F = -\nabla \left[ \frac{ge}{2m} B(R) \cdot S \right] = \frac{ge\hbar}{4m} \nabla B(R).
\]

You can create a magnetic field with a strong gradient by making the two poles of the magnet of different sizes, so the field is concentrated near one of them and spreads out towards the other. Such a technique of measuring spin is generically called a Stern-Gerlach experiment, after the original experiment of this type (which was performed with silver atoms, not electrons).\(^1\)

E. Simple Problems with Magnetic Fields

Let’s do a simple problem to understand how to use such a Hamiltonian. Consider a charged spin-$\frac{1}{2}$ particle in a uniform magnetic field $\mathbf{B} = B\hat{k}$. Classically, such a particle will move at an arbitrary speed in the $z$-direction, while moving simultaneously in circles in the $xy$-plane, and finally, the spin will precess like a top around the $z$-axis. Let’s study its behavior quantum mechanically. The first step is to write the magnetic field as the curl of some vector potential $\mathbf{A}$. The magnetic field has lots of symmetries, such as rotational symmetry around the $z$-axis, translational symmetry in all three directions, etc., but it turns out it is impossible to pick a vector potential that respects all of these symmetries. The $z$-component of magnetic field is given by $B_z = \partial A_y/\partial x - \partial A_x/\partial y$, which suggests we need $\mathbf{A}$ to point in either the $x$- or $y$-direction. There are several possible choices, but one that works is

$$\mathbf{A}(\mathbf{r}) = -By\hat{x}.$$  

This choice sacrifices both translation in the $y$-direction and rotational invariance. Our Hamiltonian is then

$$H = \frac{1}{2m}[\mathbf{p} - e\mathbf{B}\hat{y}]^2 + \frac{ge}{2m}BS_z = \frac{1}{2m}\left[ P_x^2 + P_y^2 + \left( P_z - eBY \right)^2 + geBS_z \right].$$

The first step when dealing with a new Hamiltonian is to attempt to find operators that commute with the Hamiltonian, and each other. It isn’t hard to see that the four operators $\{H, P_x, P_y, S_z\}$ all commute with each other, and hence we can imagine that the eigenstates can be labeled $|n, k_z, k_s, m_s\rangle$ such that

$$H |n, k_z, k_s, m_s\rangle = E |n, k_z, k_s, m_s\rangle,$$  \hspace{1cm} (9.21a)

$$P_x |n, k_z, k_s, m_s\rangle = \hbar k_s |n, k_z, k_s, m_s\rangle,$$  \hspace{1cm} (9.21b)

$$P_y |n, k_z, k_s, m_s\rangle = \hbar k_s |n, k_z, k_s, m_s\rangle,$$  \hspace{1cm} (9.21c)

$$S_z |n, k_z, k_s, m_s\rangle = \hbar m_s |n, k_z, k_s, m_s\rangle.$$  \hspace{1cm} (9.21d)

We are trying to solve the time-independent Schrödinger equation, which is just (9.21a). We use the explicit form of $H$ and take advantage of (9.21b,c,d) to write

$$E |n, k_z, k_s, m_s\rangle = H |n, k_z, k_s, m_s\rangle = \frac{1}{2m}\left[ P_x^2 + P_y^2 + \left( P_z - eBY \right)^2 + geBS_z \right]|n, k_z, k_s, m_s\rangle$$

$$= \frac{1}{2m}\left[ \hbar^2 k_s^2 + P_y^2 + \left( \hbar k_s - eBY \right)^2 + geBh\hbar m_s \right]|n, k_z, k_s, m_s\rangle,$$

$$\left( E - \frac{\hbar^2 k_s^2}{2m} - \frac{geB\hbar m_s}{2m} \right)|n, k_z, k_s, m_s\rangle = \left[ \frac{1}{2m} P_x^2 + \frac{e^2B^2}{2m} \left( Y - \frac{\hbar k_s}{eB} \right)^2 \right]|n, k_z, k_s, m_s\rangle,$$

The right hand side is nothing more than a shifted harmonic oscillator. To make this more obvious, define $\omega_b = eB/m$ and $Y' = Y - \hbar k_s/eB$. Then in terms of this new constant and variable, we are trying to solve
\[
\left( E - \frac{\hbar^2 k_z^2}{2m} - \frac{1}{2} g \hbar \omega_B m_s \right) |n, k_z, k_x, m_s\rangle = \left[ \frac{1}{2m} P_y^2 + \frac{1}{2} m \omega_B^2 Y'^2 \right] |n, k_z, k_x, m_s\rangle ,
\]

We know the eigenvalues of the Harmonic oscillator, which are \( \hbar \omega_B (n + \frac{1}{2}) \), which allows us to find the energies, which are

\[
E = \frac{\hbar^2 k_z^2}{2m} + \hbar \omega_B \left( n + \frac{1}{2} + \frac{1}{2} g m_s \right).
\]  

(9.22)

The terms in (9.22) are relatively easy to understand. The first term is just the kinetic energy in the \( z \)-direction, which is decoupled from all other motion. The term given by \( \hbar \omega_B \left( n + \frac{1}{2} \right) \) represents its motion in the \( xy \)-plane, and the last term corresponds to its precession.

Recall that a naive relativistic quantum mechanical model predicts for particles like the electron \( g = 2 \). Since \( m_s = \pm \frac{1}{2} \), this means that the expression in parentheses in (9.22) will either take the value \( n \) or \( n + 1 \). This suggests that an accidental degeneracy will occur, so we have two degenerate states, one where there is more kinetic motion in the \( xy \)-plane, and the spin is anti-aligned with the magnetic field, and the other where the spin is aligned but there is less kinetic energy. A more thorough treatment of the problem using quantum electrodynamics (QED for short) actually predicts a slight difference of \( g \) from 2 (called the anomalous magnetic moment), and this implies that the energies will be very close, but not exactly degenerate. The slight difference can be measured with high precision by stimulating the electrons at frequency \( \frac{1}{2} \omega_B (g - 2) \), where there will be resonant absorption. This measurement can be performed with unbelievable accuracy, yielding the experimental value \( g = 2.0023193043622(15) \), where the 15 denotes the error in the last two digits. This factor can be computed in quantum electrodynamics to nearly as many digits, and the resulting agreement represents the most accurate test of any theory.

As a second example, let us consider a hydrogen atom in a “strong” magnetic field, where we will carefully define strong later. If it is an external magnetic field, it is likely to be virtually uniform over the scale of the atom, so we treat it as constant, and let \( \hat{B} = B \hat{z} \) as before. It is tempting to immediately write down the vector potential \( \mathbf{A}(r) = -B \hat{y} \hat{x} \) as before, but in this case this is not the easiest way to proceed. This vector potential has translational symmetry in the \( x \)- and \( z \)-directions, but the electric potential from the nucleus lacks these symmetries, so this is not likely to be a good way to proceed. Instead we will place the nucleus at the origin, and then use a vector potential which preserves the rotational symmetry around the \( z \)-axis, namely

\[
\mathbf{A}(r) = \frac{1}{2} B (x \hat{y} - y \hat{x})
\]

The Hamiltonian is then

\[
H = \frac{1}{2m} \left[ \mathbf{P} + \frac{1}{2} e B \hat{y} \mathbf{X} - \frac{1}{2} e B \hat{y} \mathbf{Y} \right] - \frac{k_e^2}{|\mathbf{R}|} - \frac{ge}{2m} BS_z,
\]
\[ H = \frac{1}{2m} \left[ (P_x - \frac{1}{2} eBY)^2 + (P_y + \frac{1}{2} eBX)^2 + P_z^2 + geBS_z \right] - \frac{k_r e^2}{|R|} \]
\[ = \frac{1}{2m} \left[ P_x^2 + P_y^2 + P_z^2 + eB \left( XP_y - YP_x \right) + geBS_z \right] - \frac{k_r e^2}{|R|} + \frac{e^2 B^2}{8m} \left( X^2 + Y^2 \right), \]
\[ H = \frac{1}{2m} \left[ P_x^2 + P_y^2 + P_z^2 + eB \left( L_z + gS_z \right) \right] - \frac{k_r e^2}{|R|} + \frac{e^2 B^2}{8m} \left( X^2 + Y^2 \right). \]

Consider for the moment the last term, which looks like a harmonic oscillator helping to keep the electron near the atom. Putting in units, this works out to
\[ \frac{e^2 B^2}{8\mu} \left( X^2 + Y^2 \right) = \left( \frac{B}{T} \right)^2 \left( \frac{X^2 + Y^2}{0.1 \text{ nm}} \right)^2 \left( 2.20 \times 10^{-10} \text{ eV} \right). \]

Since the scale of an atom tends to be less than 0.1 nm, and it is difficult to produce magnetic fields stronger than a few tens of tesla, the contribution of this last term tends to be irrelevant. Hence when we say a strong magnetic field, we actually do not mean too strong. We therefore neglect the last term compared to the others. We therefore have
\[ H = H_0 + \frac{eB}{2m} \left( L_z + gS_z \right), \]

where \( H_0 \) is the Hamiltonian independent of the magnetic field. In chapter seven we and found the eigenstates of \( H_0 \), and after adding spin in chapter eight, we found that we could describe our states as \( |n, l, m_i, m_s \rangle \). Fortunately, this means these states are still eigenstates of the full Hamiltonian (9.23), so we can immediately identify the corresponding energies.

\[ E \left| n, l, m_i, m_s \right\rangle = H \left| n, l, m_i, m_s \right\rangle = \left[ H_0 + \frac{eB}{2m} \left( L_z + gS_z \right) \right] \left| n, l, m_i, m_s \right\rangle = \left[ -\frac{\alpha^2 mc^2}{2n^2} + \frac{eB\hbar}{2m} \left( m_i + gm_s \right) \right] \left| n, l, m_i, m_s \right\rangle, \]
\[ E = -\frac{\alpha^2 mc^2}{2n^2} + \frac{eB\hbar}{2m} \left( m_i + gm_s \right). \]  

The eigenstates of the Hamiltonian have been split, and now depend on the quantities \( m_i \) and \( m_s \). Indeed, the letter \( m \) is used for these two quantities because they are split by magnetic fields, and they are called the magnetic quantum numbers.

To obtain (9.24), we had to assume the magnetic field is not too strong, but when did we assume that it was strong at all? The answer is that there are small corrections to the energies due to relativistic effects, and these mean that the Hamiltonian, even in the absence of magnetic fields, does not commute with \( L_z \) or \( S_z \). It is possible to analyze the full effects of both magnetic fields and these spin-orbit interactions, but this requires perturbation theory, a subject of a later chapter. In the strong field limit, when (9.24) applies, the effect is called the Paschen-Back effect.
In quantum mechanics, magnetic fields are introduced into the Hamiltonian via
the vector potential, which cannot, however, be directly observed. Are there any
quantum mechanical effects that can indirectly measure the vector potential? Most
interesting would be an experiment that can detect the vector potential in a situation
where there is no magnetic field at all present.

Suppose we have a region of space that contains a vector potential \( \mathbf{A} \neq 0 \), but \( \mathbf{B} = 0 \). Then we have \( 0 = \mathbf{B} = \nabla \times \mathbf{A} \). It is a general theorem that a field whose curl vanishes
everywhere can be written as the gradient of some other field, so say \( \mathbf{A} = -\nabla \chi \). Then if
we perform a gauge transformation \( \mathbf{A} \rightarrow \mathbf{A} + \nabla \chi \), the vector potential goes away. Since
our theory is supposed to be gauge invariant, this seems to suggest we can never see the
effects of the vector potential itself.

There is a flaw in this argument. The
assumption was that the magnetic field vanished
everywhere. By previous arguments, you can show
that you can eliminate the vector potential in any
simply-connected region where the magnetic field
vanishes.\(^1\) However, consider the case illustrated in
Fig. 9-4, where the magnetic field exists in only a
small region, but there is a net magnetic flux through
that region. Such a field, for example, is created
when you have a long, thin solenoid with current
flowing in the direction of the solid arrow. Then
consider a loop surrounding the region with magnetic
field, as illustrated in Fig. 9-4. Let’s calculate the
magnetic flux through this loop, which is given by

\[
\Phi_C = \oint_C \mathbf{B} \cdot \mathbf{n} \, ds = \oint_S \nabla \times \mathbf{A} \cdot \mathbf{n} \, da = \oint_C \mathbf{A} \cdot ds
\]

where \( C \) is the dashed curve. It is always possible to
change gauge such that \( \mathbf{A} \) vanishes in just one region,
say in the region above the
solenoid, but
not everywhere.

This
leads to some
interesting
consequences.
For example,
consider a
double slit-style
experiment of

\(^1\) A simply connected region is one where any closed loop can be shrunk to a point.
the type illustrated in Fig. 9-5. A beam of electrons passes through a narrow slit, then propagates towards a second barrier with two slits, before proceeding to a screen where the resulting position of the electron is detected. Let’s first analyze it without the solenoid. Assuming we have an initial wave function with energy $E = \hbar^2 k^2/2m$, the wave will propagate from the initial point to the two slit region looking like $\psi(r) \sim \psi_0 e^{ikr}$ where $r$ is the distance from the first slit. Thus when it arrives at the second pair of slits, the wave will be proportional to $\psi(r) \sim \psi_0 e^{ikL_1}$ and $\psi_0 e^{ikL_2}$ respectively. It then propagates forward from these two points, producing a final wave of the form

$$\psi_1 \sim \psi_0 \exp(ikL_1) \quad \text{and} \quad \psi_2 \sim \psi_0 \exp(ikL_2).$$

where $L_1$ and $L_2$ are the total distances traveled between the original source and the screen. The probability density on the screen is given by

$$\left|\psi_1\right|^2 \sim \left|\psi_0\right|^2 \exp(ikL_1) + \exp(ikL_2) = 4\left|\psi_0\right|^2 \cos^2\left[\frac{1}{2}k(L_1 - L_2)\right].$$

The path length difference $L_1 - L_2$ will be different for different locations on the screen, resulting in the familiar interference pattern.

Now imagine adding a solenoid behind the second screen, so the two paths leading to the screen pass around the solenoid. This introduces a vector potential $A(r)$ that makes the problem harder. Fortunately, since there is no magnetic field outside the solenoid, the vector potential has no curl, and can be written as the gradient of some scalar potential, provided we stick to one path or the other. In other words, we can write

$$A(r) = \begin{cases} -\nabla \chi_1(r) & \text{on path } 1, \\ -\nabla \chi_2(r) & \text{on path } 2. \end{cases}$$

It is important to note that the two functions $\chi_1(r)$ and $\chi_2(r)$ are not the same functions. We now perform one of two gauge transformations:

$$A(r) \rightarrow A'_1(r) = A(r) + \nabla \chi_1(r), \quad A(r) \rightarrow A'_2(r) = A(r) + \nabla \chi_2(r) \quad (9.26)$$

If we perform the first gauge transformation, for example, then $A$ will vanish everywhere on path 1, and we will get the exact same solutions as before, and a similar argument applies to path 2, so we get relations that look like

$$\psi'_1 \sim \psi_0 \exp(ikL_1), \quad \text{and} \quad \psi'_2 \sim \psi_0 \exp(ikL_2)$$

These two wave functions are not in the same gauge, however, and we must convert them back to our original gauge before we attempt to add them. This is done using equation (9.11c), and we conclude

$$\psi_1 \sim \psi_0 \exp\left(ikL_1 + \frac{ie}{\hbar} \chi_1\right), \quad \text{and} \quad \psi_2 \sim \psi_0 \exp\left(ikL_2 + \frac{ie}{\hbar} \chi_2\right).$$

The probability density at the screen will therefore be
\[ |\psi|^2 \sim |\psi_0|^2 \exp \left( ikL_1 + \frac{ie}{\hbar} \chi_1 \right) + \exp \left( ikL_2 + \frac{ie}{\hbar} \chi_2 \right) \]
\[ = 4 |\psi_0|^2 \cos^2 \left[ \frac{k(L_1 - L_2)}{2} + \frac{e(\chi_1 - \chi_2)}{2\hbar} \right], \tag{9.27} \]

where the gauge functions \( \chi_1 \) and \( \chi_2 \) are evaluated at the screen.

We can find a simple expression that eliminates the gauge functions from (9.27). With the help of (9.26), we see that
\[
\int \mathbf{A} \cdot ds = -\int \nabla \chi_1 \cdot ds = -\chi_1, \quad \int \mathbf{A} \cdot ds = -\int \nabla \chi_2 \cdot ds = -\chi_2. \tag{9.28} \]

Subtracting the first from the second, we see that
\[ \chi_1 - \chi_2 = \int_{\text{path 2}} \mathbf{A} \cdot ds - \int_{\text{path 1}} \mathbf{A} \cdot ds = \oint_{\text{c}} \mathbf{A} \cdot ds = \Phi_B. \]

Substituting this back into (9.27), we see that
\[ |\psi|^2 \sim 4 |\psi_0|^2 \cos^2 \left[ \frac{k\Delta L}{2} + \frac{e\Phi_B}{2\hbar} \right]. \]

Thus, quantum mechanically, we can use interference to detect the presence of magnetic flux through the solenoid, even though our electrons never pass through the region of magnetic field! This remarkable effect is called the Aharanov-Bohm effect.

Interestingly, this effect only occurs if the total flux \( e\Phi_B \) is not a multiple of \( 2\pi\hbar = \hbar \). Imagine for the moment a very thin, very long solenoid (or bar magnet), as illustrated in Fig. 9-6, with a magnetic flux \( \Phi_B \) passing through it. If we make the radius of the solenoid small enough, then classically we can’t see the magnetic field passing along its length. But we can see the magnetic fields escaping from the ends. If the other end is far enough away (or infinitely far away), it will look to us like a magnetic monopole, with magnetic fields emanating from the end.

However, no matter how small the radius, we could still detect the solenoid by an Aharonov-Bohm type effect, provided \( e\Phi_B \neq n\hbar \). If we expect magnetic monopoles to exist, therefore, we want the solenoids to be invisible, and therefore we conclude that
\[ \Phi = \frac{n\hbar}{e}, \]

where \( n \) is an arbitrary integer. This is called the Dirac quantization condition. In fact, many extensions of the standard model predict the existence of such monopoles, and they typically have such magnetic charges.

Figure 9-6: A long thin solenoid (blue line) with one end infinitely far away looks like a magnetic monopole, with magnetic fields emanating from the visible end (green lines).
Problems for Chapter 9

1. This problem has nothing to do with quantum mechanics. In the presence of a charged plasma, it is possible to create electromagnetic waves that are longitudinal, having electric polarization parallel to the direction of propagation, so it might take the form \( \mathbf{E}(\mathbf{r}, t) = E_0 \hat{z} \cos(kz - \omega t) \), but with no magnetic field, \( \mathbf{B}(\mathbf{r}, t) = 0 \).

   (a) Show that this electric field (and lack of magnetic field) can be written purely in terms of a vector potential \( \mathbf{A}_1(\mathbf{r}, t) \), without the use of the scalar potential, so that \( U_1(\mathbf{r}, t) = 0 \).

   (b) Show that the same electric fields (and lack of magnetic field) can also be written purely in terms of a scalar potential \( U_2(\mathbf{r}, t) \), with no vector potential, so that \( \mathbf{A}_2(\mathbf{r}, t) = 0 \).

   (c) Show that these two sets of potential, \( (\mathbf{A}_1, U_1) \) and \( (\mathbf{A}_2, U_2) \), are related by a gauge transformation, and determine explicitly the form of the gauge function \( \chi(\mathbf{r}, t) \) that relates them.

2. In chapter two, we defined the probability density \( \rho \) and probability current \( \mathbf{j} \) as

   \[
   \rho = \Psi^* \Psi \quad \text{and} \quad \mathbf{j} = \hbar \left(-i \Psi^* \nabla \Psi + i \Psi \nabla \Psi^* \right)/2m = \left( \Psi^* \mathbf{P} \Psi - \Psi \mathbf{P} \Psi^* \right)/2m,
   \]

   and then derived the conservation of probability formula \( \partial \rho/\partial t + \nabla \cdot \mathbf{j} = 0 \) from Schrödinger’s equation (2.1b). However, Schrödinger’s equation has just changed into (9.15), and our proof is no longer valid.

   (a) Define the modified probability current \( \mathbf{j}' \) the same as \( \mathbf{j} \), but replacing \( \mathbf{P} \rightarrow \mathbf{\pi} \) defined by \( \mathbf{\pi} \Psi = (\mathbf{P} + e \mathbf{A}) \Psi \) and \( \mathbf{\pi} \Psi^* = (\mathbf{P} - e \mathbf{A}) \Psi^* \). Find an expression for \( \mathbf{j}' \).

   (b) Which of the quantities \( \rho \), \( \mathbf{j} \), and \( \mathbf{j}' \) are gauge invariant?

   (c) Demonstrate that only one of \( \partial \rho/\partial t + \nabla \cdot \mathbf{j} = 0 \) and \( \partial \rho/\partial t + \nabla \cdot \mathbf{j}' = 0 \) is still valid, using the modified Schrödinger’s equation (9.15).
3. Suppose an electron lies in a region with electric and magnetic fields: \( \mathbf{B} = B\hat{z} \) and \( \mathbf{E} = \left( \frac{m\omega^2}{e} \right) \hat{x} \).

(a) Find the electric potential \( U(x) \) such that \( \mathbf{E} = -\nabla U(x) \) that could lead to this electric field.

(b) The magnetic field is independent of translations in all three dimensions. However, the electrostatic potential is independent of translations in only two of those dimensions. Find a vector potential \( \mathbf{A} \) with \( \mathbf{B} = \nabla \times \mathbf{A} \) which has translation symmetry in the same two directions.

(c) Write out the Hamiltonian for this system. Eliminate \( B \) in terms of the cyclotron frequency \( \omega_c = eB/m \). What two translation operators commute with this Hamiltonian? What spin operator commutes with this Hamiltonian?

(d) Write your wave function in the form \( \psi(r) = X(x)Y(y)Z(z)|m_x\rangle \). Based on some of the operators you worked out in part (c), deduce the form of two of the unknown functions.

(e) Replace the various operators by their eigenvalues in the Hamiltonian. The non-constant terms should be identifiable as a shifted harmonic oscillator.

(f) Make a simple coordinate replacement that shifts it back. If your formulas match mine up to now, they should look like:

\[
X = X' - \frac{\hbar k_x\omega_c}{m\left(\omega_c^2 + \omega_0^2\right)}.
\]

(g) Find the energies of the Hamiltonian.

(h) Check that they give sensible answers in the two limits when there is no electric field (pure Landau levels) or no magnetic fields (pure harmonic oscillator plus \( y \)- and \( z \)-motion).
X. Multiple Particles

A. Tensor Products of Vector Spaces

Before we get to a discussion of multiple particles, it is probably worthwhile to
discuss tensor product spaces. Consider any two vector spaces $\mathcal{V}$ and $\mathcal{W}$ of
dimensionality $N$ and $M$. Let $\{|\phi_i\rangle\}$ and $\{|\xi_j\rangle\}$ be orthonormal bases for $\mathcal{V}$ and $\mathcal{W}$ respectively. In general, an arbitrary vector in $\mathcal{V}$ and $\mathcal{W}$ can then be described as

$$|v\rangle = \sum_{i=1}^{N} v_i |\phi_i\rangle \quad \text{and} \quad |w\rangle = \sum_{j=1}^{M} w_j |\xi_j\rangle.$$ 

We now define the tensor product space $\mathcal{T} = \mathcal{V} \otimes \mathcal{W}$ as the vector space with basis $\{|\phi_i\rangle \otimes |\xi_j\rangle\}$. An arbitrary vector in this tensor product space will take the form

$$|t\rangle = \sum_{i,j} t_{ij} |\phi_i\rangle \otimes |\xi_j\rangle.$$ 

This is a vector space of dimension $NM$. Because the notation is rather cumbersome, the
basis is often simply abbreviated as $\{|\phi_i, \xi_j\rangle\}$. It should be noted that we should not think
of $t_{ij}$ as a matrix, but as a column vector with $NM$ components. The indices $i$ and $j$
collectively refer to which component we are looking at of this column vector.

We need to define an inner product for this space, for which it suffices to define it
for the basis states, for which we assume $\langle \phi_i, \xi_j | \phi_k, \xi_l \rangle \equiv \delta_{ik} \delta_{jl}$. Then for any two vectors

$$|t\rangle = \sum_{i,j} t_{ij} |\phi_i\rangle \otimes |\xi_j\rangle$$ \quad \text{and} \quad $$|s\rangle = \sum_{i,j} s_{ij} |\phi_i\rangle \otimes |\xi_j\rangle$$, the inner product will be

$$\langle t | s \rangle = \sum_{i,j} t^*_{ij} s_{ij}.$$ 

Let $v$ and $w$ be two vectors in the spaces $\mathcal{V}$ and $\mathcal{W}$, then we define the tensor
product (also called an outer product) $v \otimes w$ as

$$v \otimes w = \sum_i v_i |\phi_i\rangle \otimes \sum_j w_j |\xi_j\rangle \equiv \sum_{i,j} v_i w_j |\phi_i, \xi_j\rangle.$$ 

However, realize that most vectors in the tensor space cannot be written this way.

We actually have already encountered several examples of tensor product spaces, we
simply didn’t call them such at the time. A general operator $A$ is an element of the
tensor product space $\mathcal{H} \otimes \mathcal{H}'$, because it can be written in terms of the general basis
states $|\phi_i\rangle \langle \phi_j|$. When we studied spin, our state space was promoted from simple Hilbert
space $\mathcal{H}$ to the product of Hilbert space and spin space $\mathcal{H} \otimes \mathcal{S}$ with basis $|r, m_s\rangle$. And
when we added angular momentum, we were working with the tensor space $\mathcal{J}_1 \otimes \mathcal{J}_2$.

However, the most important example of tensor product spaces we have encountered
occurred when we talked about multiple particles.
B. Multiple Particles

Up until now, with rare exceptions, we have only discussed the quantum mechanics of a single particle. Suppose we have two particles, 1 and 2, each of which would individually be described by a vector in a state space $\mathcal{H}_1$ and $\mathcal{H}_2$. Then when both particles are present, the state space will be described by the vector space $\mathcal{H}_1 \otimes \mathcal{H}_2$. Similarly, if you have $N$ particles, the state space would be given by $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_N$. For example, if you have two particles with spin, so that the basis for each particle would take the form $|11, m_r\rangle$ and $|22, m_r\rangle$, then the basis for two particles would take the form $|11, m_1, m_2\rangle$ or $|22, m_1, m_2\rangle$. If the two particles are spin-$1/2$, for example, we might write this as $|12, \pm, \pm\rangle$. In old-fashioned wave function notation, the wave function would have four components, and we would write

$$\Psi_{m_1, m_2}(r_1, r_2) = \langle r_1, r_2, m_1, m_2 | \psi \rangle.$$ 

The square of this amplitude would have units of $|\Psi|^2 \sim L^{-6}$, and tells you the probability density that the first particle is in spin state $m_1$ and at location $r_1$, and the second in spin state $m_2$ and at location $r_2$. If we want to know the probability that the first particle is in region $R_1$ and the second in region $R_2$, for example, and we don’t care about spin, it would be given by

$$P(\mathbf{r}_1 \in R_1, \mathbf{r}_2 \in R_2) = \int_{R_1} d^3 \mathbf{r}_1 \int_{R_2} d^3 \mathbf{r}_2 \sum_{m_1} \sum_{m_2} |\psi_{m_1, m_2}(\mathbf{r}_1, \mathbf{r}_2)|^2.$$ 

In the absence of magnetic fields or spins a typical Hamiltonian will take the form

$$H = \sum_{i} \frac{1}{2m_i} \mathbf{p}_i^2 + V(\mathbf{R}_i).$$ 

The operators $\mathbf{P}_i$ and $\mathbf{R}_i$ are defined by

$$\langle \mathbf{r}_1, \ldots, \mathbf{r}_N | \mathbf{R}_i | \psi \rangle = \mathbf{r}_i \langle \mathbf{r}_1, \ldots, \mathbf{r}_N | \psi \rangle, \quad \langle \mathbf{r}_1, \ldots, \mathbf{r}_N | \mathbf{P}_i | \psi \rangle = -i\hbar \nabla_i \langle \mathbf{r}_1, \ldots, \mathbf{r}_N | \psi \rangle,$$

where $\nabla_i$ means we are taking derivatives only with respect to one coordinate. The potential term $V(\mathbf{R}_i)$ will generally be a single function of all of the positions, and cannot be simply described. This is what makes multiple particles so difficult. In general the masses $M_i$ will all be different, and the potential $V(\mathbf{R}_i)$ will often be asymmetric, since the interactions of one particle will differ from that of another particle. However, it will sometimes be the case that two particles will be exactly the same type of particle, or at least have identical interactions.
C. Symmetry and Interchangeable Particles

It is common that the Hamiltonian describing the interaction of two or more particles is unchanged under the interchange of those particles. For example, suppose we have an electron and a positron (anti-electron) interacting via only Coulomb interactions, so that the Hamiltonian takes the form

\[ H = \frac{1}{2m} \mathbf{p}_1^2 + \frac{1}{2m_e} \mathbf{p}_2^2 - \frac{k_e e^2}{|\mathbf{r}_1 - \mathbf{r}_2|}. \]  

(10.1)

It is clear that this Hamiltonian is unchanged if we interchange the two particles. We define the permutation operator \( P(1 \leftrightarrow 2) \) as the operator which interchanges these two particles. For example, if we have a basis state \( |a, b\rangle \), this operator interchanges the two basis vectors, so

\[ P(1 \leftrightarrow 2)|a, b\rangle = |b, a\rangle. \]

The expression \( P(1 \leftrightarrow 2) \) should be interpreted as “the argument in position 1 goes to position 2, and vice versa.” If we have three or more particles, more complicated interchanges are possible, for example, with three,

\[ P(1 \to 2 \to 3 \to 1)|a, b, c\rangle = |c, a, b\rangle. \]

It isn’t too difficult to figure out how to multiply these operators. For example,

\[ P(i \leftrightarrow j)P(j \leftrightarrow k) = P(i \to j \to k \to i), \]

provided \( i, j, \) and \( k \) are all distinct. How can we see this? Well, start with \( i \). We see that \( P(j \leftrightarrow k) \) leaves \( i \) unmoved, but then \( P(i \leftrightarrow j) \) switches it to \( j \). If we start with \( j \), then \( P(j \leftrightarrow k) \) switches it to \( k \) and then \( P(i \leftrightarrow j) \) leaves it unchanged, and finally if we start with \( k \) you see that \( P(j \leftrightarrow k) \) switches it to \( j \) and then \( P(i \leftrightarrow j) \) switches it to \( i \). In general, if you have \( N \) particles, there are exactly \( N! \) different permutation operators, including the trivial permutation \( P() = 1 \).

If our permutation operators do not commute with the Hamiltonian they are not really very interesting. On the other hand, when they do, we can use this fact to help us diagonalize the Hamiltonian. It is easily demonstrated that permutations are unitary operation (they don’t change the norm), so their eigenvalues are of magnitude unity. When we have two particles, there is only one non-trivial permutation, \( P(1 \leftrightarrow 2) \), so we classify our states by their eigenvalue under this switching:

\[ P(1 \leftrightarrow 2)|\psi\rangle = \lambda_{12}|\psi\rangle. \]

But obviously, if you pair switch twice, you would get back to your initial state, so

\[ |\psi\rangle = \left[ P(1 \leftrightarrow 2) \right]^2 |\psi\rangle = \lambda_{12}^2 |\psi\rangle, \quad \lambda_{12} = \pm 1. \]
Hence we should be able to organize our eigenstates of the Hamiltonian into two types, one where the wave function is symmetric (+) under this pair switching, and one where it is anti-symmetric (−). Indeed, the two lowest states of positronium are nearly degenerate, the symmetric ortho-positronium and anti-symmetric para-positronium.¹

Now let’s think about three particles. It might be tempting to think that since there are three possible pair switchings \( P(i \leftrightarrow j) \), we could classify states by their eigenvalues \( \lambda_{12} \), \( \lambda_{13} \) and \( \lambda_{23} \), but this supposition is doubly wrong. First, because the various permutation operators do not generally commute, there is no guarantee that we can find eigenstates of all three pair-switching operators. Furthermore, in cases where states can be found that are simultaneously eigenstates of all pair-switchings, it is not very hard to show that all pair-switching operators \( P(i \leftrightarrow j) \) have exactly the same eigenvalue. Hence they are all plus one, or all minus one. In general, whenever you have more than two particles, there are three possibilities:

(a) The state can be completely symmetric, in that it is unchanged under any pair switching;
(b) The state can be completely anti-symmetric, in that it acquires a minus sign under any pair switching; or
(c) The state can be of mixed symmetry, which is a state that is not an eigenstate of all pair switchings.

The third case is mathematically the most interesting, but we will focus almost exclusively on the first two.

It is pretty easy to see that any permutation can be built up out of pair-switchings, and as a consequence in the first two cases it is obvious that the state will be an eigenstate of all permutation operators. If it has eigenvalue +1 under pair switchings, then under an arbitrary permutation \( P \) it will have eigenvalue +1, so

\[
P|\psi\rangle = |\psi\rangle, \quad \text{all permutations } P. \tag{10.2}
\]

Any type of particle that satisfies (10.2) is deemed a boson. In contrast, if a particle has eigenvalue -1 under pair switchings, then for an arbitrary permutation \( P \) we will have

\[
P|\psi\rangle = \eta_p |\psi\rangle, \quad \text{all permutations } P. \tag{10.3}
\]

where \( \eta_p = +1 \) for any permutation that can be built out of an even number of pair switchings, and \( \eta_p = -1 \) for any odd number of pair switchings. It is not obvious but well-known that any permutation is uniquely even or odd; if it can be achieved in an even (odd) number of pair-switchings, then it cannot be achieved in an odd (even) number of pair-switchings. Any particle satisfying (10.3) is deemed a fermion.

¹ The spatial part of the wave function is identical in each case, just the way the spins are combined that makes them different. The two spin-½ particles combine in the symmetrical way to make spin-1 ortho-positronium, and in the anti-symmetrical way to make spin-0 para-positronium.
D. Identical Particles

The situation gets more interesting if the two particles are not just interchangeable in the Hamiltonian, but actually identical particles. The distinction is subtle, and we aren’t really prepared to define it, but an electron and positron are not really identical, since the electron has negative charge and the positron positive. But if we had two electrons then the only change to (10.1) would be a change in sign of the potential term which would leave the Hamiltonian still symmetric. The interesting and surprising fact is that the two electrons would always be in an antisymmetric state. In other words, it is an experimental fact that any two electrons will have a state vector \( |\psi\rangle \) such that

\[
P(1 \leftrightarrow 2) |\psi\rangle = - |\psi\rangle.
\]

Now comes a deep result that can only be proven with considerable work using relativistic quantum mechanics, or particle theory, the spin-statistics theorem. **Particles with half-integer spin are always fermions, and have anti-symmetric state vectors, and particles with integer spin are always bosons, and have symmetric state vectors.** Hence electrons, protons, and neutrons are all fermions, and have anti-symmetric state vectors.

It is helpful to define the symmetric and anti-symmetric operators \( S \) and \( A \) respectively. These are defined for states with \( N \) identical particles, as

\[
S \equiv \frac{1}{N!} \sum_P P \quad \text{and} \quad A \equiv \frac{1}{N!} \sum_{\eta_P} \eta_P P.
\]

The sums are taken over all \( N! \) permutations. For example, suppose we have a state involving three particles which are in three states which we label 1, 2, and 3, so \( |\psi\rangle = |123\rangle \), then

\[
S |123\rangle = \frac{1}{6} \left( |123\rangle + |231\rangle + |312\rangle + |132\rangle + |213\rangle + |321\rangle \right),
\]

\[
A |123\rangle = \frac{1}{6} \left( |123\rangle + |231\rangle + |312\rangle - |132\rangle - |213\rangle - |321\rangle \right).
\]

Both of these operators are projection operators, which means they are pulling out some subspace of the total space, and have the properties \( S^2 = S \) and \( A^2 = A \). These two spaces are orthogonal to each other (if there is more than one particle), so \( SA = AS = 0 \).

Indeed, since the quantum state is always symmetric or anti-symmetric, it is obvious that our state space is not really \( \mathcal{V}_N = \mathcal{V}_1 \otimes \mathcal{V}_1 \otimes \cdots \otimes \mathcal{V}_1 \), where \( N \) is the number of particles and \( \mathcal{V}_1 \) is the state space for one particle, but actually

\[
\mathcal{V}_N = S \left[ \mathcal{V}_1 \otimes \mathcal{V}_1 \otimes \cdots \otimes \mathcal{V}_1 \right] \quad \text{for bosons (with integer spin), or} \\
\mathcal{V}_N = A \left[ \mathcal{V}_1 \otimes \mathcal{V}_1 \otimes \cdots \otimes \mathcal{V}_1 \right] \quad \text{for fermions (with half-integer spin)}.
\]

Where the operators \( S \) and \( A \) are understood to project out the appropriate subspaces.

If we have multiple types of particles there is no necessity to symmetrize or anti-symmetrize different types of particles, so these ideas are easily generalizable to more
complex systems. For example, if we have three bosons of type 1, two fermions of type 2, and another boson of type 3, we might write the state space as

\[ V = S[V_1 \otimes V_2 \otimes V_3] \otimes A[V_4 \otimes V_5] \otimes V_6. \]

There is no necessity to symmetrize the last boson, since there is only one of them.

It will prove helpful to have basis states for any system of bosons or fermions. If \( \{ | \phi_i \rangle \} \) forms an orthonormal basis of the vector space \( V_i \) for one particle, then for \( N \) particles, we can always write any state in the form

\[ | \psi \rangle = \sum_{i_1, i_2, \ldots, i_N} c_{i_1, i_2, \ldots, i_N} | \phi_{i_1}, \phi_{i_2}, \ldots, \phi_{i_N} \rangle. \]

If the particles are bosons, then

\[ | \psi \rangle = S| \psi \rangle = \sum_{i_1, i_2, \ldots, i_N} c_{i_1, i_2, \ldots, i_N} S| \phi_{i_1}, \phi_{i_2}, \ldots, \phi_{i_N} \rangle. \]

However, because the symmetrization operator doesn’t care about the order of the arguments inside the ket, we are adding the same kets multiple times to make our wave function. We can therefore drop the redundant terms in the sum, which can be achieved by restricting the sum to terms in some non-trivial way; for example,

\[ | \psi \rangle = S| \psi \rangle = \sum_{i_1, i_2, \ldots, i_N} c_{i_1, i_2, \ldots, i_N} S| \phi_{i_1}, \phi_{i_2}, \ldots, \phi_{i_N} \rangle. \]

Indeed, it is pretty easy to see that \( S| \phi_{i_1}, \phi_{i_2}, \ldots, \phi_{i_N} \rangle \) with the restriction \( i_1 \leq i_2 \leq \cdots \leq i_N \) forms an orthogonal basis, but not orthonormal. If you want the basis orthonormal, you need to take

\[ \left\{ \frac{N!}{n_1! n_2! \cdots n_k!} S| \phi_{i_1}, \phi_{i_2}, \ldots, \phi_{i_N} \rangle : i_1 \leq i_2 \leq \cdots \leq i_N \right\}, \]

where \( n_i \) is the number of particles in an identical state of type 1, \( n_2 \) of type 2, etc. It is not hard to show that if the dimensionality of the one-particle state space is \( d \), then the dimensionality of the \( N \)-particle state is

\[ \dim\left( S[V \otimes V \otimes \cdots \otimes V] \right) = \binom{d + N - 1}{N} = \frac{(d + N - 1)!}{N!(d - 1)!}. \]

In a similar manner, we can show that for fermions we can write

\[ | \psi \rangle = A| \psi \rangle = \sum_{i_1, i_2, \ldots, i_N} c_{i_1, i_2, \ldots, i_N} A| \phi_{i_1}, \phi_{i_2}, \ldots, \phi_{i_N} \rangle. \]

Once again, since any permutation of the arguments inside the ket produces either the identical state or a minus sign when anti-symmetrized, there’s no point in keeping this redundant basis. Furthermore, in this case, if any two of the arguments of the ket match, the anti-symmetrization will cause the expression \( A| \phi_{i_1}, \phi_{i_2}, \ldots, \phi_{i_N} \rangle \) to vanish, so we further restrict our sum to
\[ | \psi \rangle = \sum_{i_1 < i_2 < \cdots < i_N} c_{i_1, i_2, \ldots, i_N} \mathcal{A} | \phi_{i_1}, \phi_{i_2}, \ldots, \phi_{i_N} \rangle. \]  

(10.7)

If we need an orthonormal basis, we can use

\[ \left\{ \sqrt{N!} \mathcal{A} | \phi_{i_1}, \phi_{i_2}, \ldots, \phi_{i_N} \rangle : i_1 < i_2 < \cdots < i_N \right\}. \]  

(10.8)

No factors of \( n_i! \) are required in the denominator, since it is forbidden to put two particles in the same quantum state. The dimensionality of the full space is

\[ \text{dim} (\mathcal{A} [\mathcal{V} \otimes \mathcal{V} \otimes \cdots \otimes \mathcal{V}]) = \binom{d}{N} = \frac{d!}{N!(d-N)!}. \]  

(10.9)

Since we are usually working in an infinite-dimensional space, the equations (10.6) and (10.9) are rarely helpful.

Equation (10.8) contains the essence of the Pauli Exclusion Principle, usually stated as something like \textbf{two fermions cannot occupy the same quantum state}. However, it should be understood that in most cases, there are many terms in expression (10.7), and there is no sense in which one can keep track of individual electrons and unambiguously state which quantum states \textit{are} occupied.

The choice of basis \( \left\{ | \phi_i \rangle \right\} \) for single particles will determine the ease with which sums like (10.7) are an efficient way to describe multi-particle states. A poor choice will result in many comparable terms that need to be added together, a more sophisticated choice may mean that (10.7) has only a few important terms in it. Because we have so much freedom in the single-particle basis \( \left\{ | \phi_i \rangle \right\} \), it is often possible to find many-body basis states like (10.5) or (10.8) which resemble actual eigenstates of the many-body system, at least qualitatively.

If we actually need the wave function corresponding to the basis state of the many particle system for a set of fermions, this works out to

\[ \phi_{i_1, i_2, \ldots, i_N} (r_1, r_2, \ldots, r_N) = \frac{1}{\sqrt{N!}} \det \begin{pmatrix} \phi_{i_1} (r_1) & \phi_{i_2} (r_2) & \cdots & \phi_{i_N} (r_N) \\ \phi_{i_1} (r_1) & \phi_{i_2} (r_2) & \cdots & \phi_{i_N} (r_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{i_1} (r_1) & \phi_{i_2} (r_2) & \cdots & \phi_{i_N} (r_N) \end{pmatrix}. \]

It should be understood that the arguments \( r_i \) might stand not only for position, but also other properties, typically the spin component of the quantum state. This method of combining single particle wave functions to make a multi-particle wave function is called a Slater Determinant.

It is sometimes helpful \textit{not} to work with fundamental particles, but with composite particles. For example, an \(^1\text{H}\) atom consist of a proton and an electron, each of which is a fermion. We might describe a single hydrogen atom, therefore, in terms of basis states of protons and electrons, something like

\[ | ^1\text{H}, k \rangle = \sum_{i,j} c^k_{ij} | \phi_i, \xi_j \rangle. \]
where \( k \) represents any quantum numbers describing the combined system (such as the momentum of the hydrogen atom, for example). A system containing two hydrogen atoms would therefore look something like

\[
|^{1}H,^{1}H, l\rangle = \sum_{k,k'} d_{kk'}^{l} |^{1}H, k\rangle \otimes |^{1}H, k'\rangle = \sum_{k,k'} \sum_{i,j} c_{ij}^{k} c_{ij}^{k'} \phi_{i}, \xi_{j}, \phi_{l}, \xi_{j}
\]

However, since both the proton and the electrons are fermions, it follows that the coefficient of \( |\phi_{i}, \xi_{j}, \phi_{l}, \xi_{j}\rangle \) and \( |\phi_{l}, \xi_{j}, \phi_{i}, \xi_{j}\rangle \) must differ by two factors of -1, i.e. they must be identical. This tells us

\[
\sum_{k,k'} d_{kk'}^{l} c_{ij}^{k} c_{ij}^{k'} = \sum_{k,k'} d_{kk'}^{l} c_{ij}^{k} c_{ij}^{k'} = \sum_{k,k'} d_{kk}^{l} c_{ij}^{k} c_{ij}^{k'},
\]

\[
d_{kk}^{l} = d_{kk}^{l}.
\]

In other words, an \(^{1}H\) atom is a boson, even though it is built out of two fermions. This principle can easily be generalized to any number of particles; namely, **any composite particle containing an even number of fermions is a boson, and containing an odd number of fermions is a fermion.** The number of bosons in the composite particle is irrelevant, so, for example, a composite particle made of three bosons is a boson, since it contains an even number (zero) of fermions.

Note that the spin-statistics rule still works, even for composite particles. The internal angular momentum of a composite particle comes from adding the spins of all its constituents, together with any angular momentum representing the constituents orbiting each other. Every fermion will contribute a half-integer multiple of \( \hbar \) to the total spin, while the bosons and all the orbiting motion contributes only integer multiples of \( \hbar \), and it is easy to show that when you add an even (odd) number of angular momenta which are half-integers, together with any number of integer angular momenta, the result is always an integer (half-integer).

E. Non-Interacting Particles and Fermi Degeneracy Pressure

Suppose we have \( N \) non-interacting particles with Hamiltonian

\[
H = \sum_{i=1}^{N} \hat{H}_{ii},
\]

where \( \hat{H}_{ii} \) acts only on one particle at a time, for example, \( \hat{H}_{ii} = p_{i}^{2}/2m + V(R_{i}) \). Then if we let the orthonormal eigenstates of the one-particle Hamiltonians be \( \{|\phi_{i}\rangle\} \) with energies \( \hat{H}_{ii} |\phi_{i}\rangle = E_{i} |\phi_{i}\rangle \), then it is easy to see that the states \( \{|\phi_{1}, \phi_{2}, \ldots, \phi_{N}\rangle\} \) will be eigenstates of the Hamiltonian with energies

\[
E_{1} + E_{2} + \cdots + E_{N} \rangle |\phi_{1}, \phi_{2}, \ldots, \phi_{N}\rangle.
\]

Of course, if we have bosons or fermions, the states (10.10) are not generally allowed, because they are not symmetric nor anti-symmetric. We must therefore
symmetrize or anti-symmetrize them, using (10.5) or (10.8). If the one-particle states are ordered in increasing energy \( \{ |\phi_1\rangle, |\phi_2\rangle, \ldots \} \), then the ground state for the \( N \)-particle state will be

\[
|\psi_{N_0}^B\rangle = |\phi_1, \phi_2, \ldots, \phi_1\rangle \quad \text{(10.11)}
\]

for bosons and

\[
|\psi_{N_0}^F\rangle = \sqrt{N!} A |\phi_1, \phi_2, \ldots, \phi_N\rangle \quad \text{(10.12)}
\]

for fermions. In the thermodynamic limit, \( N \to \infty \), these states represent only the \( T \to 0 \) ultracold limit. Nonetheless, it is possible to get a large fraction of bosonic particles in a system into the same quantum state, producing what is known as a Bose-Einstein condensate, as suggested by (10.11). For fermions, it is often the case that the vast majority of low-lying energy states are occupied, with a relatively sharp threshold above which most states are unoccupied. The ground state energy in each case is, of course,

\[
E_{B_0} = NE_1, \quad E_{F_0} = \sum_{i=1}^{N} E_i.
\]

Let’s work out a particularly interesting case. Suppose we have a large number of non-interacting spin-1/2 particles trapped in a cubical infinite square well of volume \( V = L^3 \). The one-particle Hamiltonian is given by

\[
H_{ii} = \frac{1}{2m} p_i^2 + V(R_i), \quad V(R_i) = \begin{cases} 0 & \text{if } 0 < X_i, Y_i, Z_i < L, \\ \infty & \text{otherwise.} \end{cases}
\]

The eigenstates of the one particle problem are well known. The states are labeled \( |n_x, n_y, n_z, m_s\rangle \) where \( n_x, n_y, \) and \( n_z \) are positive integers, while \( m_s = \pm \frac{1}{2} \), and the wave functions look like

\[
\langle x, y, z, m_s | n_x, n_y, n_z, m_s \rangle = \frac{8}{L^3} \sin \left( \frac{\pi n_x x}{L} \right) \sin \left( \frac{\pi n_y y}{L} \right) \sin \left( \frac{\pi n_z z}{L} \right) \delta_{m,m_s},
\]

and have energy

\[
E_{n_x, n_y, n_z, m_s} = \frac{\pi^2 \hbar^2}{2mL^2} \left( n_x^2 + n_y^2 + n_z^2 \right).
\]

We’d like now to find the ground state energy of the \( N \)-body state. To do this, we simply sum up all the values of the energy, starting with the lowest one, until we get to the correct total number of particles. Let’s define the Fermi energy \( E_F \) as the energy of the \( N' \)th lowest energy state, and define the quantity \( R_F^2 \) by the formula

\[
R_F = \frac{\sqrt{2mE_F L}}{\pi \hbar}.
\]
We also commonly define the Fermi momentum as \( p_F = \pi \hbar R_F / L \). To find the energy of the ground state energy for the \( N \)-particle state, we simply sum

\[
E = \frac{\pi^2 \hbar^2}{2mL^2} \sum_{n_x,n_y,n_z}^{n_x^2 + n_y^2 + n_z^2 \leq R_F^2} \left( n_x^2 + n_y^2 + n_z^2 \right) = \frac{\pi^2 \hbar^2}{mL^2} \sum_{n_x,n_y,n_z}^{n_x^2 + n_y^2 + n_z^2 \leq R_F^2} \left( n_x^2 + n_y^2 + n_z^2 \right).
\]

The upper limit on the sum is a bit trickier than it looks, because often the highest energy state will be degenerate, and we want to only include enough states to end up with exactly \( N \) total particles, or terms in the sum. In the thermodynamic limit, \( N \to \infty \), this is irrelevant.

For general \( N \), getting (10.13) exactly is very difficult, but in the limit of large \( N \), we can get a pretty good idea of what is going on, as illustrated in Fig. 10-1. As \( N \) gets large, we can approximate the number of lattice points \( n = (n_x,n_y,n_z) \) that are closer than \( R_F \) from the origin as the volume of one-eighth of a sphere of radius \( R_F \). Since there are also two spin states, we then estimate

\[
N \approx 2 \cdot \frac{1}{8} \cdot \left( \frac{4}{3} \pi R_F^3 \right) = \frac{1}{4} \pi R_F^3,
\]

\[
R_F = \left( \frac{3}{\pi} N \right)^{1/3}.
\]

Similarly, we can estimate the ground state energy in (10.13) by replacing the sums with integrals

\[
E = \frac{\pi^2 \hbar^2}{mL^2} \iiint_{n^2 \leq R_F^2} \left( n_x^2 + n_y^2 + n_z^2 \right) d^3 n
\]

Converting this integral into spherical coordinates, and keeping in mind that only positive values of all three components of \( n \) are allowed, this can be rewritten as

\[
E = \frac{\pi^2 \hbar^2}{mL^2} \iiint_{n^2 \leq R_F^2} \left( n_x^2 + n_y^2 + n_z^2 \right) d^3 n = \frac{\pi^2 \hbar^2}{mL^2} \int_0^{R_F} \int_0^{\pi} \sin \theta d\theta \int_0^{\pi/2} \sin \phi d\phi \int_0^{\pi/2} n^2 n^2 dn = \frac{\pi^3 \hbar^2}{2mL^2} \frac{R_F^5}{5}.
\]

Substituting in (10.14), this becomes

\[
E = \frac{\pi^3 \hbar^2}{10mL^2} \left( \frac{3N}{\pi} \right)^{5/3} = \frac{3\hbar^2}{10m} \left( \frac{3\pi^2 N}{V} \right)^{5/3} N.
\]

That the energy goes to infinity as one takes the number of particles to infinity is hardly surprising. It is probably more instructive to work instead with the average energy per particle, and write it in terms of the number density of particles, \( n = N/V \). We start by writing the Fermi momentum and Fermi energy as
\[ p_F = \frac{\pi \hbar R_F}{L} = \frac{\pi \hbar}{L} \left( \frac{3}{\pi} N \right)^{1/3} = \hbar \left( \frac{3\pi^2 N}{V} \right)^{1/3} = \hbar \left( 3\pi^2 n \right)^{1/3}, \quad (10.16a) \]

\[ E_F = \frac{p_F^2}{2m} = \frac{\hbar^2}{2m} \left( 3\pi^2 n \right)^{2/3}. \quad (10.16b) \]

The energy (10.15) then can be rewritten as

\[ E = \frac{1}{2} n E_F. \]

Even more interesting than the energy is the Fermi degeneracy pressure, given by

\[ P_F = -\frac{\partial E}{\partial V} = \frac{\hbar^2}{5m} \left( 3\pi^2 \right)^{2/3} \left( \frac{N}{V} \right)^{5/3} = \frac{\hbar^2}{5m} \left( 3\pi^2 \right)^{2/3} n^{5/3} = \frac{2}{5} nE_F. \quad (10.17) \]

Note that this pressure is for non-interacting fermions and exists at zero temperature. Note also that this pressure is inversely proportional to mass, so for ordinary matter, the electron contribution will vastly exceed any nuclear contribution.

Although (10.17) was derived for a cubical box, it can be shown that the same result works for any rectilinear box, and furthermore, it works for any shape that can be built out of such boxes. Indeed, one can argue that the result has nothing to do with the shape of the container; indeed, one can even change the boundary conditions. Equation (10.17) describes a macroscopic property of fermions with a fixed density in the ground state: the Pauli Exclusion Principle causes there to be an associated pressure.

Equation (10.17) can give a qualitative understanding of why metals are difficult to compress. In metals, one or more electron from each atom is so far from the atom that in bulk material, they are free to roam the crystal and behave a bit like free particles with no interactions. At room temperature the metal is near to its ground state, so (10.12) does a reasonable job of describing the ground state of these conduction electrons. The large degeneracy pressure is countered by the attraction of the electrons to the positively charged nuclei they circulate among. But if you attempt to squeeze the metal, you will increase that pressure, and the metal will resist compression. In general, however, you can get little more than qualitative understanding from (10.17).

In contrast, there are many astrophysical situations where (10.17) is valuable quantitatively. The cores of many types of stars are so highly compressed that the degeneracy pressure completely overwhelms ordinary gas pressure. For example, white dwarf stars are burned-out stars consisting of a mixture of \(^{12}\)C and \(^{16}\)O atoms held together by gravity. The density is so immense that the nuclear attractions are nearly irrelevant, and all of the electrons, even the innermost electrons, are effectively free. The star escapes collapse only because the enormous force of the electron degeneracy counteracts the equally powerful force of gravity.

High mass white dwarfs are actually unstable, and the reason is that the Fermi momentum has risen so high that the Fermi momentum, (10.16a), corresponds to a relativistic velocity. Equation (10.16b) is then no longer valid, and the degeneracy pressure now rises as \(n^{4/3}\), no longer sufficient to overcome the enormous gravitational attraction of the collapsing star. In a type II supernova, the core of a giant star (which at this stage is composed mostly of \(^{56}\)Fe) begins to catastrophically collapse, the electron Fermi energy rising faster and faster (as illustrated in Fig. 10-2). Eventually, new nuclear
processes take over, where the iron nuclei first dissolve into their constituent protons and neutrons, and the protons and electrons combine to make neutrons. Eventually (meaning less than a second later) the collapse becomes so severe that the ball of neutrons starts to feel neutron degeneracy pressure, and the star slams to a halt. The resulting energy literally explodes the stellar remnant still surrounding the core, destroying it, and the remaining ball of neutrons gradually cools, forming what is called a neutron star, consisting of almost pure neutrons.

F. Atoms

We now consider a different type of system, one consisting only of a few electrons, namely an isolate atom. A neutral atom consists of a nucleus of charge $Z e$ together with $Z$ electrons each with charge $-e$. To simplify, let’s treat the nucleus as a stationary source, then the Hamiltonian is

$$H = \sum_{i=1}^{Z} \left( \frac{1}{2m} \mathbf{p}_i^2 - k_e Z e^2 \right) + \sum_{1 \leq i < j \leq Z} \frac{k_e e^2}{|\mathbf{r}_i - \mathbf{r}_j|}.$$  \hfill (10.18)

Looking at (10.18) it is not clear how to proceed; indeed, getting accurate approximate solutions for (10.18) is very difficult. As a first attempt, we will note that the first sum in (10.18) contains $Z$ potential terms each of which is proportional to $k_e Ze^2$, whereas the second sum contains $\frac{1}{2}(Z^2 - Z)$ terms, each of which is proportional to $k_e e^2$. We might therefore conjecture that the first potential term is about twice as important as the second. As a very crude approximation, which I deem the ultra-naive atomic model, we ignore the second term and keep only the first. In this case we have non-interacting electrons, and to find the ground state of an atom we simply insert electrons into the various states $|n,l,m_l,m_s\rangle$ starting at the lowest $n$ and proceeding upwards. Since there are $2n^2$ electrons in each shell, we would expect particularly stable (non-reactive) atoms.
whenever we finish filling a shell, which would occur at \( Z = 2, 10, 28, \) and 60, (He, Ne, Ni, and Nd). Though helium and neon are quite stable, nickel and neodymium are not.

We need a more sophisticated picture of what is going on, which takes into account, at least qualitatively, interactions between electrons. Although it is true that the lowest energy states (1s electrons) will essentially see primarily a naked nucleus, these electrons are so close to the atom compared to others (recall the scale factor \( a = a_0 n^2 \)) that all other electrons will effectively see a shielded nucleus of charge \( Z - 2 \), rather than the full charge. This means that the outer electrons are more loosely bound than our naive model would predict. However, whenever an electron manages to stray inside the innermost electrons, it will see the more attractive unshielded nucleus, and will hence have lower energy. Recalling that the wave function at small \( r \) is proportional to \( l^r \), it is clear that the smaller \( l \) values will be more likely to penetrate this inner shell of electrons. Hence, we anticipate that the 2s electrons (\( n = 2, l = 0 \)) will have lower energy than the 2p electrons (\( n = 2, l = 1 \)), even though our naive picture says they have the same energy.

Indeed, this argument generalizes to higher \( n \) as well. As we go out to each shell, we would expect the energy of the various \( nl \) values to be non-degenerate for different \( l \) values, so that in general \( E_{nl} < E_{n'l'} \) if \( l < l' \). In other words, there is an energy penalty for increasing \( l \), not just \( n \). As a crude approximation, we find that the energy increases not just as \( n \) increases, but as \( n + l \), so that, for example the 4s states (\( n + l = 4 \)) have lower energy than the 3d states (\( n + l = 5 \)). However, when \( n + l \) values are tied, \( n \) is the more important number, so 3p is lower energy than 4s, even though both have \( n + l = 4 \).

These simple rules allow us to understand qualitatively most of the details of the order in which electrons are introduced into neutral atoms. If we make a chart of all possible values of \( n \) and \( l \), we can read from the table the order in which electrons are added to an atom, as illustrated in Fig. 10-3. If we denote how many electrons we have room for in each subshell by a superscript, so 2p\(^6\) denotes that six electrons can fit in the 2p-subshell, then the electrons will be introduced into atoms in the order

1s\(^2\)2s\(^2\)2p\(^6\)3s\(^2\)3p\(^4\)4s\(^2\)3d\(^10\)4p\(^6\)5s\(^2\)4d\(^10\)5p\(^2\)6s\(^2\)4f\(^4\)5d\(^10\)6p\(^6\)7s\(^2\)5f\(^{14}\)6d\(^10\)7p\(^6\) . . .

For example, to find the electrons that appear in sulfur, \( Z = 16 \), we just take the first 16 electrons, so \( 1s^2 2s^2 2p^6 3s^2 3p^4 \). If you follow this rule blindly, you will find that of the first 110 elements, it gets 19 of them wrong by one electron, and three wrong by two electrons. Not bad for such a naive picture.

This order of filling also gives a qualitative understanding of the periodic table of the elements (see Fig. 10-4). For example, any atom that ends in the combination \( ns^1 \) has a single electron farther out than the others, and this electron will be very likely to be shared with other atoms. These alkali metals lie in the first column of the periodic table, and the alkali earth metals end in \( ns^2 \), to its immediate right. The atoms on the right side of the table are having their p-subshells filled, and the transition metals are having their d-subshells filled, while the rare earths are adding electrons in their f-subshells. The

© 2011, Eric D. Carlson
noble gases have a completely filled np-subshell (1s for helium) and adding an electron would require it to go into the higher level (n+1)s state, so they wish to neither give up nor gain an electron, making them particularly stable. We never get to the n = 8 levels, which is why there is no 8'th row of the table, and we don’t get as far as the g-subshells either, which should start as we pass element 120 (as of 2011, the heaviest element is 118).

Does it matter which orbitals (m_l-values) and spin states (m_s-values) we put the electrons in as we fill the various levels? For the inner electrons, it does not, because these inner electrons are so tightly bound that all these states must be filled, but for the outer electrons, the repulsion between the electrons means that the electrons want to avoid each other. Even though spin states allow two electrons to occupy the same orbital (have the same m_l-value), the electrons, which repel each other, want to be in different orbitals as much as possible, so electrons will go into different orbitals. This effect can be large enough that an electron might shift to a slightly higher subshell to avoid having two in the same orbital, especially if two subshells are very close. For example, although our rules state that 5s subshell is lower energy than the 4d, they are actually very close, and several elements (Nb, Mo, Tc, Ru, Rh, Ag) shift an extra electron up to the 4d level so there can be only one in the 5s shell.

You might think that spin states are irrelevant, since our Hamiltonian has included no spin interactions, but this is not the case. Consider, for example, carbon, which has six electrons in the configuration 1s^22s^22p^2. The 1s and 2s electrons will, of course, have opposite spin, but what about the 2p electrons? Let’s focus on just these two electrons. Since we are combining two spin-1/2 electrons, the total spin can be 0 or 1. Suppose we write our wave function in the form...
\[ \psi_i (\mathbf{r}_1, \mathbf{r}_2) |10\rangle = \psi_i (\mathbf{r}_1, \mathbf{r}_2) \frac{1}{\sqrt{2}} (|+\rangle + |\rangle + |\rangle - |\rangle - |\rangle), \]
\[ \psi_0 (\mathbf{r}_1, \mathbf{r}_2) |00\rangle = \psi_0 (\mathbf{r}_1, \mathbf{r}_2) \frac{1}{\sqrt{2}} (|+\rangle - |\rangle - |\rangle + |\rangle - |\rangle). \]

where for ease of comparison we have put both states in the total spin \( m_s = 0 \) state. However, remember that these two electrons are fermions, and this wave function must be antisymmetric, which means that if you interchange the two particles (position and spin), you must get minus. So we have

\[ \psi_i (\mathbf{r}_2, \mathbf{r}_1) = -\psi_i (\mathbf{r}_1, \mathbf{r}_2), \]
\[ \psi_0 (\mathbf{r}_2, \mathbf{r}_1) = +\psi_0 (\mathbf{r}_1, \mathbf{r}_2). \]

In other words, for spin 1 we need a spatial wave function that is anti-symmetric, but for spin 0 we need a spatial wave function that is symmetric.

Which of these is likely to have lower energy? Remember that electrons repel each other, so they do not like to be near each other. Taking the extreme case where the electrons are at the same location, we can see from (10.18a) that for spin 1, we have \( \psi_i (\mathbf{r}, \mathbf{r}) = 0 \), whereas no such restriction occurs for the spatially symmetric spin 0 case; indeed, symmetric wave functions tend to have an increased likelihood that the particles will be near each other. In conclusion, in order to keep the particles apart, they will prefer an anti-symmetric spatial wave function, which causes them to have a symmetric spin state, so the electrons will be in a spin 1 state. This is a simple example of ferromagnetism, where electrons tend to align, not due to some spin-related force, but rather due to electron repulsion and the necessity of anti-symmetrizing the quantum state.

How does this generalize to more complicated atoms? Whenever you put electrons into a subshell the spatial wave function will tend to be as anti-symmetric as possible, which can be achieved so long as you don’t put any two electrons in the same orbital (which means the subshell is no more than half full). The spin state will then be completely symmetric, which can be achieved, for example, by putting the electrons in the state \( |+++\rangle \), a quantum state with spin \( s = \frac{1}{2} n \), where \( n \) is the number of electrons in the subshell. As you pass the halfway mark, the electrons have to be loaded into some of the already singly occupied orbitals, which forces them, by the Pauli exclusion principle, to cancel the spin of the electron already there, and the spin starts to decrease. When the subshell is full, the spins all cancel out, and you end up with the total spin of the atom being zero.

Actually solving the problem is quite complicated, and finding interactions, or chemical bonds, between atoms is even harder. Interactions between electrons mean that it is impossible to simply determine which of two closely spaced subshells has the lower energy, which is why there are exceptions to the rules stated above. But now we are wandering into the area of quantum chemistry, which is beyond the scope of this course.
Problems for Chapter 10

1. Two non-identical spinless particles are in the asymmetric infinite square well, defined by

\[ H = \frac{p_1^2 + p_2^2}{2m} + V(X_1, X_2) \]

where

\[ V(X_1, X_2) = \begin{cases} 0 & \text{if } 0 < X_1 < X_2 < a \\ \infty & \text{otherwise} \end{cases} \]

The ground state wave function in the allowed region \(0 < x_1 < x_2 < a\) is given by

\[ \psi(x_1, x_2) = N \sin \left( \frac{\pi x_1}{a} \right) \sin \left( \frac{\pi x_2}{a} \right) \left[ \cos \left( \frac{\pi x_1}{a} \right) - \cos \left( \frac{\pi x_2}{a} \right) \right] \] .

(a) Check that this wave function is a solution of Schrödinger’s equation, and determine the ground state energy. Also check that it is continuous.
(b) Find the normalization constant \(N\).
(c) The position of the two particles is measured in this state. What is the probability that \(x_1 \leq \frac{1}{2} a\), \(x_1 \geq \frac{1}{2} a\), and \(x_2 \leq \frac{1}{2} a\), and that both are true?

2. A system of more than two particles is an eigenstate of every pair switching operator; that is,

\[ (\mathcal{P}_{ij}) \psi = \lambda_{ij} \psi \]

for every \(i \neq j\).
(a) For \(i, j, k\) all different, simplify the product \(\mathcal{P}(i \leftrightarrow j) \mathcal{P}(i \leftrightarrow k) \mathcal{P}(i \leftrightarrow j)\).
(b) Demonstrate that \(\lambda_{ik} = \lambda_{jk}\) for any \(i, j, k\) all different.
(c) Argue that for any pair \(\lambda_{ij}\) and \(\lambda_{kl}\), \(\lambda_{ij} = \lambda_{kl}\), whether there are any matches or not. Hence there is only one common eigenvalue for all pair switchings.

3. Three indistinguishable particles have Hamiltonian \(H = \sum_{i=1}^{3} \left( \frac{p_i^2}{2m} + \frac{1}{2} m \omega^2 X_i^2 \right)\).
(a) If the particles are not identical, find the three lowest energies, give me a list of all quantum states having those energies, and tell me the degeneracy (number of states with that energy) for each.
(b) Repeat part (a) if the three particles are identical bosons. I still want three energies, but I am not guaranteeing that the energies will be the same.
(c) Repeat part (a) if the three particles are identical fermions.

4. Look up and/or calculate the number density of atoms in copper. Assume there is one conduction electron per atom.
(a) What is the Fermi energy (in eV) and Fermi degeneracy pressure (in GPa) for electrons at this density?
(b) The bulk modulus is defined as \(K = -V \left( \frac{\partial P}{\partial V} \right)\). Find a formula for the bulk modulus due to electron degeneracy pressure. Use this to estimate the bulk modulus of copper (in GPa). Look it up somewhere for copper and compare the result.
5. A white dwarf is a spherical star consisting of a mix of $^{12}$C atoms (6 electrons, atomic mass 12 g/mol) and $^{16}$O atoms (8 electrons, atomic mass 16 g/mol). It is under such incredible pressure that the electrons are essentially a free Fermi gas. Similarly, a neutron star consists of almost pure neutrons (atomic mass 1 g/mol). We will approximate each of these objects as uniform spheres of mass $M$ and radius $R$. They are held together by gravity, which is opposed by degeneracy pressure.

(a) Write a formula applicable in either case for the degeneracy pressure $P_F$ in terms of the mass $M$, the mass of the particle causing the degeneracy pressure $m$, the mass per particle causing degeneracy pressure $\mu$, and the radius $R$. (Note: for the neutron star, $\mu$ and $m$ are the same thing, but for the white dwarf, they are very different, because most of the mass is not in the form of electrons).

(b) Multiply the degeneracy pressure by the cross-sectional area of the star sliced in half. Equate this to the gravitational attraction, which you estimate as follows: pretend each half of the star has mass $M/2$, and are separated by a distance $R$. Find an equation for $R$ in terms of $M$ (and other constants).

(c) What is $\mu$ for a $^{12}$C or $^{16}$O white dwarf? Find the numerical value of (b) in km for a white dwarf or neutron star of mass equal to the Sun ($M = 1.989 \times 10^{30}$ kg).

(d) Find a formula for the Fermi energy $E_F$ in terms of the mass $M$ (and other constants).

(e) When the Fermi energy reaches $\frac{1}{2}mc^2$, the approximations break down, and the star undergoes catastrophic collapse. Estimate this mass $M$ for each of the two cases, in solar masses, where this collapse occurs. The actual values are 1.4 solar masses (white dwarf) and around 2.5 solar masses (neutron star).

6. Two spin-1/2 non-interacting fermions are in the $n = 0$ and $n = 1$ state of a 1D harmonic oscillator; $H = \sum_{i=1}^{2} \left( p_i^2 / 2m + \frac{1}{2} m \omega^2 x_i^2 \right)$. Naively, we might write the quantum state as $|0, 1\rangle$, but this is wrong because (i) it’s not antisymmetrized, and (ii) we didn’t include spin states.

(a) Taking into account spin states, write four correctly antisymmetrized wave functions, written so they are eigenstates of $S^2$ and $S_z$, where $S$ is the total spin of the two electrons. You can write your states in whatever basis is most convenient; I recommend $|n_1 n_2; m_1 m_2\rangle = |n_1\rangle \otimes |n_2\rangle$ and/or $|n_1 n_2; m_1 m_2\rangle = |n_1\rangle \otimes |\pm\rangle$.

(b) For each of the four states, calculate $\langle \psi | (X_1 - X_2)^2 | \psi \rangle$

(c) Suppose the particles had an additional repulsive force between them, so the fermions prefer to be far apart. Which of the two states would have lower energy?
XI. Interpretation of Quantum Mechanics

Even though quantum mechanics has been around for more than a century, there is still considerable dispute about exactly what it means. Indeed, there are different ways of reformulating it that give (or appear to give) the exact same results. We devote a significant portion of this chapter to discussion about what quantum mechanics might or might not mean. Fortunately, most of us can agree how to perform actual quantum computations, whether or not we agree on what it means. For this reason, I will try to intersperse some useful tools for not only thinking about quantum mechanics, but actually performing useful computations. Hence the first seven sections of this chapter have little about philosophy, and much more about tools for computation.

A. The Time Evolution Operator

Suppose we know the quantum state at some time $t_0$ and wish it at another time $t$. Then there should exist a mapping from state space to itself that finds the new state vector as a function of the old,

$$|\Psi(t)\rangle = U(t,t_0)|\Psi(t_0)\rangle. \quad (11.1)$$

This operator is called the *time evolution operator*. Among its properties must be

$$U(t_0,t_0)=1, \quad (11.2a)$$

$$U^\dagger(t,t_0)U(t,t_0)=1, \quad (11.2b)$$

$$U(t_2,t_1)U(t_1,t_0)=U(t_2,t_0). \quad (11.2c)$$

Equation (11.2b) is true because we want our state vector to remain normalized, and (11.2a) and (11.2c) can easily be derived from (11.1). If we apply Schrödinger’s time-dependant equation to (11.1), we find one additional property

$$i\hbar \frac{\partial}{\partial t} |\Psi(t)\rangle = H |\Psi(t)\rangle,$$

$$i\hbar \frac{\partial}{\partial t} U(t,t_0)|\Psi(t_0)\rangle = HU(t,t_0)|\Psi(t_0)\rangle,$$

$$i\hbar \frac{\partial}{\partial t} U(t,t_0) = HU(t,t_0). \quad (11.3)$$

Equations (11.2a) and (11.3) together determine $U(t,t_0)$. If $H$ is constant, we can demonstrate that

$$U(t,t_0) = \exp \left[-\frac{i}{\hbar} H (t-t_0) \right]. \quad (11.4)$$

If we have a complete set of eigenstates of the Hamiltonian $|\phi_n\rangle$ with energy $E_n$ we can rewrite (11.4) in the form
\begin{equation}
U(t,t_0) = \sum_n \exp \left[ -\frac{i}{\hbar} H(t-t_0) \right] |\phi_n\rangle \langle \phi_n| = \sum_n |\phi_n\rangle \exp \left[ -\frac{i}{\hbar} E_n(t-t_0) \right] \langle \phi_n|.
\end{equation}

If $H$ is time dependant, but commutes with itself at different times, (11.4) needs to be modified slightly:

\begin{equation}
U(t,t_0) = \exp \left[ -i\hbar^{-1} \int_{t_0}^t H(t_1) dt_1 \right].
\end{equation}

When the Hamiltonian does not commute with itself at different times, it is possible to show that a general expression for $U$ is

\begin{equation}
U(t,t_0) = 1 + \sum_{N=1}^\infty (i\hbar)^{-N} \int_{t_0}^t dt_1 H(t_1) \int_{t_0}^{t_1} dt_2 H(t_2) \cdots \int_{t_0}^{t_{N-1}} dt_N H(t_N).
\end{equation}

An expression nearly identical to this will be demonstrated in the chapter on time-dependant perturbation theory, so we will not pursue it now.

There is an additional property of $U$ that I want to emphasize because it will prove crucial later on. Recall that Schrödinger’s equation is linear, so that if $|\psi_1(t)\rangle$ and $|\psi_2(t)\rangle$ satisfy it, so does $c_1 |\psi_1(t)\rangle + c_2 |\psi_2(t)\rangle$. It logically follows that if you start in the state $c_1 |\psi_1(t_0)\rangle + c_2 |\psi_2(t_0)\rangle$, you end up in the corresponding state later, so in other words

\begin{equation}
U(t,t_0) \left[ c_1 |\psi_1(t_0)\rangle + c_2 |\psi_2(t_0)\rangle \right] = c_1 |\psi_1(t)\rangle + c_2 |\psi_2(t)\rangle,
\end{equation}

\begin{equation}
U(t,t_0) \left[ c_1 |\psi_1(t_0)\rangle + c_2 |\psi_2(t_0)\rangle \right] = c_1 U(t,t_0) |\psi_1(t_0)\rangle + c_2 U(t,t_0) |\psi_2(t_0)\rangle.
\end{equation}

In other words, $U$ is linear.

Sometimes the simple properties of $U$ can make a problem trivial, or make it trivial to prove that something is impossible. For example, suppose you have a complicated wave function at $t_0$ in a 1D Harmonic oscillator potential. You want to know the probability at some later time $t$ that it is in, let’s say, the ground state. The computation you need to do is:

\begin{equation}
P_0(t) = \left| \langle 0 | U(t,t_0) | \Psi(t_0) \rangle \right|^2 = \left| \langle 0 | \exp \left[ -\frac{i}{\hbar} H(t-t_0) \right] | \Psi(t_0) \rangle \right|^2.
\end{equation}

At first blush, this looks pretty complicated, since our wave function might have a very complicated time evolution. But the Hamiltonian operator can be thought of as operating on the right or the left, and on the left it is acting on an eigenstate of the Hamiltonian with energy $\frac{1}{2} \hbar \omega$, so this becomes simply

\begin{equation}
P_0(t) = \left| \langle 0 | \exp \left[ -\frac{i}{\hbar} \omega (t-t_0) \right] | \Psi(t_0) \rangle \right|^2 = \left| \langle 0 | \Psi(t_0) \rangle \right|^2.
\end{equation}

So the problem is really pretty trivial; you perform a single integration and you are done.

Let’s take what might seem like an easy problem. We are presented with a single electron, and told it is in the $|+\rangle$ state or the $|-\rangle$ state, but we don’t know which one. We
are asked to turn it into pure $|+\rangle$ state. We are allowed to apply any Hamiltonian with any time dependence we want. Can we do it? In other words, we’d like to find a time evolution operator such that

$$U(t,t_0)|+\rangle = |+\rangle \quad \text{and} \quad U(t,t_0)|-\rangle = |+\rangle.$$  

Now, take the inner product of these two expressions with each other. We find

$$1 = \langle +|+\rangle = \langle -|U^\dagger(t,t_0)U(t,t_0)|+\rangle = \langle -|+\rangle = 0,$$

where we have used the unitarity of $U$. Obviously this is nonsense. Indeed, you can see by such an argument that no matter what the Hamiltonian, two states that are initially orthogonal will always evolve into states that are orthogonal.

It is worth mentioning this now to make it very clear: time evolution by the operator $U$ is, at least in principle, reversible. Given the final state $|\Psi(t)\rangle$, we can determine the initial state $|\Psi(t_0)\rangle$ simply by multiplying by $U^\dagger(t,t_0)$. This clearly suggests that the time evolution operator is not responsible for the collapse of the wave function, which is an irreversible process. Time evolution is governed by $U$ only when you are not performing measurements.\(^1\)

Naively this seems to suggest that if you are given an electron in an unknown spin state you can never change it to what you want. This is not the case. The secret is to introduce a second particle, so your quantum state initially is $|+,a_0\rangle$ or $|-,a_0\rangle$, where $a_0$ represents the quantum state of some other particle(s). We can then let the two interact in such a way that the states evolve according to

$$U(t,t_0)|+,a_0\rangle = |+,a_+\rangle \quad \text{and} \quad U(t,t_0)|-,a_0\rangle = |+,a_-\rangle.$$  

The electron is now in a $+$ state, and, provided the states $|a_\pm\rangle$ are orthogonal, our previous disproof of such an interaction fails.

B. The Propagator

Assume for the moment that we have a single spinless particle, so that $|r\rangle$ forms a basis for our state space. Suppose we are given the initial wave function $\Psi(r_0,t_0)$ and are asked to get the final wave function $\Psi(r,t)$. The answer is

$$\Psi(r,t) = \langle r|\Psi(t)\rangle = \langle r|U(t,t_0)|\Psi(t_0)\rangle = \iiint (r|U(t,t_0)|r_0)\langle r_0|\Psi(t_0)\rangle d^3r_0. \quad (11.6)$$

We define the propagator, also known as the kernel, as

$$K(r,t;r_0,t_0) = \langle r|U(t,t_0)|r_0\rangle. \quad (11.7)$$

\(^1\) Wording can get very tricky here. I am generally following the viewpoint of the Copenhagen interpretation. In other interpretations, such as many worlds, the evolution of the universe is always governed by the time evolution operator.
Then the wave function at any time $t$ will be given by
\[
\Psi(r,t) = \iiint K(r,t;r_0,t_0) \Psi(r_0,t_0) d^3r_0. \tag{11.8}
\]

Finding the propagator can be fairly difficult in practice. If the two times are set equal, it is pretty easy to see from (11.6) that it must be simply a delta function
\[
K(r,t;r',t) = \delta^3(r-r') \tag{11.9}
\]

Now, suppose our Hamiltonian takes the form $H = \mathbf{P}^2/2m + V(R,t)$. By applying Schrödinger’s equation to (11.8), we see that
\[
\frac{i\hbar}{\partial t} \Psi(r,t) = \left[ -\frac{\hbar^2}{2m} \nabla^2 + V(r,t) \right] \Psi(r,t),
\]
\[
\frac{i\hbar}{\partial t} \iiint K(r,t;r',t') \Psi(r',t') d^3r' = \left[ -\frac{\hbar^2}{2m} \nabla^2 + V(r,t) \right] \iiint K(r,t;r',t') \Psi(r',t') d^3r' = 0,
\]
\[
\frac{i\hbar}{\partial t} K(r,t;r',t') = -\frac{\hbar^2}{2m} \nabla^2 + V(r,t) K(r,t;r',t'),
\tag{11.10}
\]

where at the last step we argue that since the integral vanishes for arbitrary $\Psi(r',t')$, what is multiplying it must vanish. The boundary condition (11.9) and the differential equation (11.10) determine the propagator.

Actually finding the propagator by the use of (11.10) is pretty hard. It is often easier to simply use (11.7) together with (11.5) to see that
\[
K(r,t;r',t') = \sum_n \langle r | \phi_n \rangle \exp \left[ -\frac{i}{\hbar} E_n (t-t') \right] \langle \phi_n | r' \rangle
\]

Let’s work this out in the special case of a free particle of mass $m$ in one dimension. In this case, our energy eigenstates are momentum eigenstates $|k\rangle$ with wave functions given by (3.19) (modified for one dimension) and energy $E = \hbar^2 k^2 / 2m$. We therefore find the propagator
\[
K(x,t;x',t') = \int dk \langle x | k \rangle \exp \left[ -i \frac{\hbar^2 k^2 (t-t')}{2m\hbar} \right] \langle k | x' \rangle
\]
\[
= \int \frac{dk}{2\pi} \exp \left[ -i \frac{\hbar k^2 (t-t')}{2m} + ik(x'-x) \right] = \frac{1}{2\pi} \sqrt{\frac{\pi 2m}{ih(t-t')}} \exp \left[ \frac{2m^2 (x'-x)^2}{4i\hbar(t-t')} \right],
\]
\[
K(x,t;x',t') = \frac{m}{2\pi i\hbar(t-t')} \exp \left[ \frac{im(x-x')^2}{2\hbar(t-t')} \right]. \tag{11.11}
\]
In three dimensions, the free propagator would have the first factor cubed, and the exponential would change from \((x - x')^2\) to \((r - r')^2\).

We have thus far discussed only a single spinless particle. If we have a particle with spin, or if we have two particles, then (11.7) would have to be modified to

\[ K_{nn'}(r, t; r', t') = \langle r, m| U(t, t')| r', m' \rangle, \quad \text{or} \quad K(r_1, r_2; r'_1, r'_2, t, t') = \langle r_1, r_2| U(t, t')| r'_1, r'_2 \rangle, \]

where \(m\) and \(m'\) are spin indices. We can then get the wave function at later times from an initial time using generalized versions of (11.8):

\[
\Psi_m(r, t) = \sum_{m'} \int K_{nn'}(r, t; r', t') \Psi_{m'}(r', t') d^3r',
\]

\[
\Psi(r_1, r_2, t) = \int \int K(r_1, r_2; r'_1, r'_2, t, t') \Psi(r'_1, r'_2, t') d^3r_1 d^3r_2.
\]

It is, of course, possible to expand this to any number of particles with arbitrary spin.

C. The Feynman Path Integral Formalism

We now discuss a rather interesting way of looking at things, normally attributed to Feynman. For definiteness, consider a single particle in one-dimension with Hamiltonian \(H = P^2/2m + V(X, t)\). We start by trying to find the propagator from an initial time \(t_0\) to a very short time thereafter \(t_1\). We expect that over such a short time period, the wave function would hardly change at all, so that the propagator will look like a Dirac delta function. We'd also expect the potential not to change much. Hence in the last term of (11.10), we approximate \(V(x, t)\) as \(V(x_0, t_0)\) which allows us to cleverly write this equation as

\[
\left[ i\hbar \frac{\partial}{\partial t} + \frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} - V(x_0, t_0) \right] K(x, t; x_0, t_0) = 0,
\]

\[
\left[ i\hbar \frac{\partial}{\partial t} + \frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \right] \left\{ \exp \left[ \frac{i}{\hbar} V(x_0, t_0)(t - t_0) \right] \right\} K(x, t; x_0, t_0) = 0.
\]

This equation is identical to the free particle problem, whose solution is given by (11.11), so changing \(t\) to \(t_1\) and \(x\) to \(x_1\), we have

\[
K(x_1, t_1; x_0, t_0) \exp \left[ \frac{i}{\hbar} V(x_0, t_0)(t_1 - t_0) \right] = \sqrt{\frac{m}{2\pi i\hbar(t_1 - t_0)}} \exp \left[ \frac{im(x_1 - x_0)^2}{2\hbar(t_1 - t_0)} \right],
\]

\[
K(x_1, t_1; x_0, t_0) = \sqrt{\frac{m}{2\pi i\hbar(t_1 - t_0)}} \exp \left[ \frac{i}{\hbar} \frac{m(x_1 - x_0)^2}{2(t_1 - t_0)} - V(x_0, t_0)(t_1 - t_0) \right].
\]

Abbreviating the time difference \(\Delta t \equiv t_1 - t_0\), we have
\[
\Psi(x, t) = \sqrt{\frac{m}{2\pi i\hbar\Delta t}} \int dx_0 \exp \left\{ \frac{i}{\hbar} \int_0^t \left[ \frac{m}{2} \left( \frac{x - x_0}{\Delta t} \right)^2 - V(x_0, t_0) \right] dt \right\} \Psi(x_0, t_0).
\]

Now, let’s try to propagate the wave forward another step in time, to \(t_2\), and to keep things simple, we’ll assume a common time step so

\[
\Psi(x, t_2) = \frac{m}{2\pi i\hbar\Delta t} \int dx_1 \int dx_0 \exp \left\{ \frac{i}{\hbar} \int_0^{t_2} \left[ \frac{m}{2} \left( \frac{x_2 - x_1}{\Delta t} \right)^2 - V(x_1, t_0) + \frac{m}{2} \left( \frac{x_1 - x_0}{\Delta t} \right)^2 - V(x_0, t_0) \right] dt \right\} \Psi(x_0, t_0).
\]

Let’s generalize to many steps. Provided each step is small enough, we have

\[
\Psi(x_N, t_N) = \left( \frac{m}{2\pi i\hbar\Delta t} \right)^{N/2} \int dx_{N-1} \cdots \int dx_0 \exp \left\{ \frac{i}{\hbar} \int_0^{t_N} \sum_{i=1}^{N} \left[ \frac{m}{2} \left( \frac{x_i - x_{i-1}}{\Delta t} \right)^2 - V(x_{i-1}, t_{i-1}) \right] dt \right\} \Psi(x_0, t_0).
\]

By comparing this with (11.8), and rearranging a bit, we see that the propagator is just

\[
K(x_N, t_N; x_0, t_0) = \left( \frac{m}{2\pi i\hbar\Delta t} \right)^{N/2} \int dx_{N-1} \cdots \int dx_0 \exp \left\{ \frac{i}{\hbar} \sum_{i=1}^{N} \Delta t \left[ \frac{m}{2} \left( \frac{x_i - x_{i-1}}{\Delta t} \right)^2 - V(x_{i-1}, t_{i-1}) \right] \right\}.
\]

(11.12)

Now, we expect (11.12) to be a good approximation only if each step is infinitesimal. In this limit, we are performing integrations over \(x\) at every intermediate time, and as a consequence we are, in a sense, integrating over all possible functions \(x(t)\), as illustrated in Fig. 11-1. The paths are constrained, however, at the endpoints, where \(x(t_0) = x_0\) and \(x(t_N) = x_N\). We now define a path integral as

\[
\int_{x(t_0)}^{x(t_N)} \mathcal{D}[x(t)] \equiv \lim_{N \to \infty} \left( \frac{m}{2\pi i\hbar\Delta t} \right)^{N/2} \int dx_{N-1} \cdots \int dx_1.
\]

(11.13)

In the limit of small step sizes we see that the quantity in round brackets in (11.12) is just the time derivative of \(x\), so we have

\[
\int_{x(t_0)}^{x(t_N)} \mathcal{D}[x(t)] \equiv \lim_{N \to \infty} \left( \frac{m}{2\pi i\hbar\Delta t} \right)^{N/2} \int dx_{N-1} \cdots \int dx_1.
\]
\[
\sum_{i=1}^{N} \Delta t \left[ \frac{m}{2} \left( \frac{x_i - x_{i-1}}{\Delta t} \right)^2 - V(x_{i-1}, t_{i-1}) \right] = \sum_{i=1}^{N} \Delta t \left\{ \frac{1}{2} m \left[ \dot{x}(t_i) \right]^2 - V(x(t_{i-1}), t_{i-1}) \right\}
\]

\[
= \int_{t_0}^{t_N} dt \left\{ \frac{1}{2} m \left[ \dot{x}(t) \right]^2 - V(x(t), t) \right\}
\]

\[
= \int_{t_0}^{t_N} dt L \left[ \dot{x}(t), x(t) \right] = S\left[ x(t) \right],
\]

where we have recognized the limit of the sum as an integral, and then realized that the expression in curly brackets is nothing more than the classical Lagrangian, whose integral is the classical action. Substituting this into (11.12), we have

\[
K(x_N, t_N; x_0, t_0) = \int_{t_0}^{t_N} D[x(t)] \exp \left\{ i \hbar^{-1} S \left[ x(t) \right] \right\}.
\]

This expression is certainly compact, but unfortunately it is rarely useful in performing actual computations.\(^\ast\) It does, however, allow us to immediately write down an alternative form of the second of our quantum mechanical postulates, which would now look something like

**Postulate 2:** Between measurements, the state vector changes according to

\[
\left| \Psi(t) \right\rangle = \int_{t_0}^{t} D[x(t)] \exp \left\{ i \hbar^{-1} S \left[ x(t) \right] \right\} \left| \Psi(t_0) \right\rangle,
\]

where \(S\left[ x(t) \right]\) is the classical action associated with the path \(x(t)\).

I have not attempted to make the notation completely clear or generalizable, because I have no intention of using the second postulate in this form. Indeed, it can be shown to be identical with the second postulate we have used up to now, so it should not be considered a separate interpretation of quantum mechanics.

Why would anyone consider this statement superior to the original version of the postulate? Well, it isn’t mathematically simpler, since the path integral is such a beastly thing to define. However, it has a variety of advantages. First of all, classical physics normally starts from the Lagrangian and the action, and then works for a while before defining the Hamiltonian. This formulation allows us to jump to the equivalent of Schrödinger’s equation in a single step. Furthermore, if you consider relativity, the Hamiltonian turns out not to be Lorentz invariant, and if you use the Hamiltonian formulation of quantum mechanics you will end up with a lot of non-Lorentz invariant formulas. In contrast, the action is Lorentz invariant, so this version of the second postulate more naturally allows one to proceed to finding relativistic quantum mechanical theories.

\(^\ast\) You can actually solve the harmonic oscillator using this method, though it is cumbersome. In quantum chromodynamics, the quantum mechanical theory of the strong force, perturbative and other approaches tend to fail, and path integrals are sometimes used for brute-force computations. Such techniques take phalanxes of high-speed computers working in parallel, and only in the last decade or two have they been able to provide any useful data.
There is also a very nice connection between (11.14) and classical mechanics. Consider the fact that in terms of any ordinary dimensions, Planck’s constant is very tiny, and hence the argument of the exponential will be a large imaginary number. If you change the path $x(t)$ even a tiny bit the action will generally change, and therefore the exponential will suddenly change its phase. Hence when you integrate over all paths, most paths will have contributions that are nearly cancelled by nearby paths whose phase is rapidly varying. The only exception is paths such that tiny changes in the path produce no change in the action; to be precise, paths such that

$$\delta S[x(t)] \bigg/ \delta x(t) = 0. \quad (11.15)$$

Indeed, a common method for estimating integrals of the form (11.14) is called the stationary phase approximation, where you assume it is dominated by paths near to (11.15). But we recognize (11.15) from classical mechanics as simply the classical path. Hence to leading order, in the limit where Planck’s constant is small, the path that has the largest quantum mechanical contribution is none other than the classical path. But in quantum mechanics, the particle takes all paths.

D. The Heisenberg Picture

Suppose we have a quantum system initially in the state vector $|\Psi(t_0)\rangle$ and are performing a measurement corresponding to an operator $A$ for a quantum system with at a later time $t$. Then the expectation value of the operator is given by

$$\langle A \rangle \equiv \langle \Psi(t)|A|\Psi(t)\rangle = \langle \Psi(t_0)|U^\dagger(t,t_0)AU(t,t_0)|\Psi(t_0)\rangle. \quad (11.16)$$

Up to now we have regarded the state vector as changing, while the operator $A$ stays “constant.” However, looking at (11.16), isn’t it equally consistent to regard the state vector as constant, while the operator changes? Let’s define the Heisenberg version of the operator $A$ as

$$A_H(t) \equiv U^\dagger(t,t_0)A_HU(t,t_0),$$

where we have used subscripts to denote the operator in the usual Schrödinger picture and the new Heisenberg picture. This will apply even to the Hamiltonian,

$$H_H(t) = U^\dagger(t,t_0)H_S(t)U(t,t_0),$$

though if $H$ contains no explicit time dependence, then $U$ commutes with $H$ and therefore $H_H = H_S$. We will also define the Heisenberg state vector as

$$|\Psi_H\rangle \equiv |\Psi_S(t_0)\rangle. \quad (11.17)$$

Then (11.17) can be written in either the Heisenberg or Schrödinger picture.

$$\langle A \rangle \equiv \langle \Psi_S(t)|A_S|\Psi_S(t)\rangle = \langle \Psi_H|A_H(t)|\Psi_H\rangle. \quad (11.18)$$
The difference is that in the Heisenberg picture it is not the state vector but the operators that change with time.

Can we find a simple formula for the time evolution of $A$ in the Heisenberg picture? First take the Hermitian conjugate of (11.3), which is

$$
-i\hbar \frac{\partial}{\partial t} U^\dagger (t, t_0) = U^\dagger (t, t_0) H_S.
$$

We then use this and (11.3) to find the derivative of $A$:

$$
\frac{d}{dt} A_H(t) = \left[ \frac{\partial}{\partial t} U^\dagger (t, t_0) \right] A_S U(t, t_0) + U^\dagger (t, t_0) A_S \left[ \frac{\partial}{\partial t} U(t, t_0) \right] 
$$

$$
= \frac{i}{\hbar} U^\dagger (t, t_0) H_S A_S U(t, t_0) - \frac{i}{\hbar} U^\dagger (t, t_0) A_S H_S U(t, t_0) 
$$

$$
= \frac{i}{\hbar} \left\{ U^\dagger (t, t_0) H_S U(t, t_0) U^\dagger (t, t_0) A_S U(t, t_0) \right\} = \frac{i}{\hbar} \left[ H_H A(t)_H - A(t)_H H_H \right],
$$

$$
\frac{d}{dt} A(t)_H = \frac{i}{\hbar} [H_H, A(t)_H].
$$

Indeed, in the Heisenberg picture, we would have to rewrite the second postulate something like this:

**Postulate 2: All observables $A$ evolve according to**

$$
\frac{d}{dt} A(t) = \frac{i}{\hbar} [H, A(t)], \tag{11.19}
$$

**where $H$ is another observable.**

If you are working in the Heisenberg picture you should rewrite the other postulates as well, removing references to $t$ in the state vector and adding them to the observables. Indeed, the Heisenberg picture was developed before the Schrödinger picture. We can see from our derivations that it is equivalent to it, but there is no way to tell which of the two representations is “correct.” The Heisenberg picture is generally harder to deal with; for example, the expectation value of most operators will change with time because the operators themselves are changing with time. Furthermore, it is easier for me to imagine a description where the thing being observed (the state vector) changes with time, while the methods of observation (observables) are constant, but both are consistent.

Technically (11.19) is correct only for operators that contain no explicit time dependence; if there is explicit time dependence there will be an additional term $\partial A/\partial t$ on the right hand side. In particular note that if the Hamiltonian has no explicit time dependence, (11.19) shows that it will have no time dependence at all.

Let’s do a simple example to illustrate how things work in the Heisenberg picture. Consider a single free particle in one dimension, with Hamiltonian $H = p^2/2m$. Note first that the momentum operator commutes with this Hamiltonian, and therefore by
(11.19) it will be constant, so that \( P(t) = P(0) \). In contrast, the position operator does not commute with the Hamiltonian, and we have
\[
\frac{d}{dt} X = \frac{i}{\hbar} [H, X] = \frac{i}{\hbar} \left[ \frac{1}{2m} P^2, X \right] = \frac{i}{2m\hbar}(P[P, X] + [P, X]P) = \frac{1}{m} P,
\]
with solution
\[
X(t) = X(0) + \frac{1}{m} P(0)t.
\]
This equation is identical with the classical equation of motion. Note that not only is the position operator at different times different, it doesn’t even commute with itself:
\[
\left[ X(t), X(0) \right] = \left[ X(0) + \frac{1}{m} P(0)t, X(0) \right] = \frac{t}{m} \left[ P(0), X(0) \right] = -\frac{i\hbar t}{m}.
\]
We know that any operators that don’t commute lead to an uncertainty principle, which in this case would take the form
\[
\left[ \Delta x(t) \right] \left[ \Delta x(0) \right] \geq \frac{\hbar t}{2m}.
\]
Though this could be proven directly with the Schrödinger picture, it is easiest to prove in the Heisenberg picture.

Yet another variation is the so-called interaction picture. Suppose the Hamiltonian consists of two parts, \( H_0 \) and \( H_1 \), and to simplify we’ll assume \( H_0 \) has no time dependence, so
\[
H = H_0 + H_1(t).
\]
Then it is possible to assume that the state vector evolves due to \( H_1 \), while the operators change due to \( H_0 \); in other words
\[
i\hbar \frac{d}{dt} \left| \Psi(t) \right\rangle = H_1(t) \left| \Psi(t) \right\rangle, \quad \text{and} \quad \frac{d}{dt} A(t) = \frac{i}{\hbar} \left[ H_0, A(t) \right].
\]  
(11.21)
You can build up a perfectly consistent view of quantum mechanics treating (11.21) and as postulates. Clearly, it is not intellectually satisfying to do so, so why would anyone ascribe to such an approach? The answer is that it often has great practical value. Suppose \( H_0 \) is large, and easy to solve and understand, while \( H_1 \) (the interaction) is smaller, and perhaps more confusing. Then it is sometimes helpful to think of quantum states as unchanging due to \( H_0 \), while the interactions cause the state to evolve. This is effectively what we will do when we do time-dependant perturbation theory, though we won’t actually work in this formalism. Furthermore, in relativistic quantum mechanics, also called quantum field theory, it is not the position of a particle \( R \) that plays the role of an observable, but a field at a point, something like \( E(r) \). In the Schrödinger picture, this operator depends on space, but not time. By switching to the interaction picture, our field operators will depend on time as well, \( E(r, t) \), which makes it easier to make the theory Lorentz invariant.
E. The Trace

Before moving on to the next subject it is helpful to define the trace of an operator. Let $\mathcal{V}$ be a vector space with orthonormal basis $\{|\phi_i\rangle\}$, then the trace of any operator $A$ is defined as

$$\text{Tr}(A) \equiv \sum_i \langle \phi_i | A | \phi_i \rangle = \sum_i A_{ii}.$$ 

In matrix notation, the trace is just the sum of the diagonal elements. The trace may look like it depends on the choice of basis, but it does not, as is easily demonstrated:

$$\text{Tr}'(A) = \sum_i \langle \phi_i | A | \phi_i \rangle = \sum_i \sum_j \langle \phi_i | A | \phi_j \rangle \langle \phi_j | \phi_i \rangle = \sum_i \sum_j \langle \phi_j | A | \phi_i \rangle = \sum_j \langle \phi_j | A | \phi_j \rangle = \text{Tr}(A).$$

A few properties of the trace are in order. If $A$ and $B$ are two operators, then it is easy to see that

$$\text{Tr}(AB) = \sum_i \langle \phi_i | AB | \phi_i \rangle = \sum_i \sum_j \langle \phi_i | A | \phi_j \rangle \langle \phi_j | B | \phi_i \rangle = \sum_i \sum_j \langle \phi_j | B | \phi_i \rangle \langle \phi_i | A | \phi_j \rangle = \sum_j \langle \phi_j | BA | \phi_j \rangle = \text{Tr}(BA).$$

It is important to realize that the trace is not invariant under arbitrary permutations of the operators (like the determinant), but only under cyclic permutations, so for example, if $A$, $B$, and $C$ are all operators then in general

$$\text{Tr}(ABC) = \text{Tr}(BCA) = \text{Tr}(CAB) \neq \text{Tr}(ACB) = \text{Tr}(CBA) = \text{Tr}(BAC).$$

A general trace reduces an operator to a pure number, but when your vector space is a tensor product space, it is sometimes useful to take only a partial trace. For example, if our vector space is $\mathcal{V} \otimes \mathcal{W}$, with basis states $\{|v_j, w_j\rangle\}$, then we can perform a trace on only one of the two vector spaces as follows:

$$\text{Tr}_v(A) = \sum_i \sum_j \sum_k |w_j\rangle \langle v_j, w_j | A | v_i, w_i \rangle \langle w_k |.$$ 

This cumbersome expression is probably clearer if we write it in components:

$$\text{Tr}_v(A)_{ij} = \sum_k A_{ki,j}.$$ 

Such traces will prove useful, for example, when we have two particles, one of which we wish to ignore. The trace on $\mathcal{V}$ reduces the operator $A$ from an operator acting on $\mathcal{V} \otimes \mathcal{W}$ to one acting only on $\mathcal{W}$. 
F. The State Operator, or Density Matrix

One of the aspects of quantum mechanics that many find distasteful is the inherent role of probability. Probability, of course, is not new to quantum mechanics. For example, when we pull a card off the top of a deck, we might say there’s a 25 percent chance that it is a heart, but this does not imply that we believe the cards are in a superposition of quantum states, which collapses into a single suit the moment we observe the card. Instead, we are expressing only our ignorance of the situation. Most physicists are not bothered by this role of probability. But even though we have discussed extensively quantum probability, even including it in our postulates, we have not discussed this more benign notion of probability resulting from ignorance.

Suppose we do not know the initial quantum state, but instead only know that it is in one of several possible states, with some distribution of probabilities. We might describe such a system by a list of states and probabilities, \( \{ (\Psi_i(t)), f_i \} \), where \( \Psi_i(t) \) represent a list of possible quantum states (normalized, but not necessarily orthogonal), and \( f_i \) represents the probabilities of each state. The probabilities will all be positive, and their sum will be one:

\[
f_i \geq 0 \quad \text{and} \quad \sum_i f_i = 1.
\]

We now define the state operator, also known as the density matrix, as

\[
\rho(t) = \sum_i f_i \langle \Psi_i(t) | \Psi_i(t) \rangle.
\]  

(11.22)

What properties will this state operator have? Well, consider first the trace

\[
\text{Tr}(\rho) = \sum_j \langle \phi_j | \sum_i f_i \langle \Psi_i | \Psi_i \rangle | \phi_j \rangle = \sum_i f_i \sum_j \langle \Psi_i | \phi_j \rangle \langle \phi_j | \Psi_i \rangle = \sum_i f_i \langle \Psi_i | \Psi_i \rangle = \sum_i f_i.
\]  

(11.23)

Note also that its expectation value with any state vector is positive

\[
\langle \psi | \rho | \psi \rangle = \sum_i f_i \langle \psi | \Psi_i \rangle \langle \Psi_i | \psi \rangle = \sum_i f_i \left| \langle \psi | \Psi_i \rangle \right|^2 \geq 0
\]  

(11.24)

In other words, the density matrix is positive semi-definite and has trace one. It is also apparent from (11.22) that it is Hermitian.

Like any Hermitian operator, the state operator will have a set of orthonormal eigenvectors \( \{ | \phi_i \rangle \} \) and eigenvalues \( \{ \rho_i \} \) which satisfy

\[
\rho | \phi_i \rangle = \rho_i | \phi_i \rangle
\]

When working in this basis, constraints (11.24) and (11.23) become

\[
\rho_i \geq 0 \quad \text{and} \quad \sum_i \rho_i = 1.
\]

In terms of this basis, of course, the state operator is
Comparing (11.25) with (11.22), we see that we can treat the state operator as if it is made up of a set of orthonormal states $|\phi_i\rangle$, with probabilities $\rho_i$. We can also see that any state operator satisfying (11.23) and (11.24) can be written in the form (11.25), which is the same form as (11.22).

One way to satisfy (11.25) is to have all of the probability in a single state, so $\rho_i = \delta_{in}$ for some $n$. In such a circumstance, we say that we have a pure state. It isn’t hard to show the only way this can be achieved is to have only a single state vector $|\Psi_i(t)\rangle$ contributing in the sum (11.23) (up to an irrelevant overall phase), so we don’t actually have any ignorance of the initial state vector. Whenever this is not the case, we say that we have a mixed state.

Given the state operator in matrix form it is easy to check if it is a pure state or not. Squaring (11.25), we see that

$$\rho^2 = \sum_i \rho_i |\phi_i\rangle \langle \phi_i| \sum_j \rho_j |\phi_j\rangle \langle \phi_j| = \sum_i \sum_j \rho_i \rho_j |\phi_i\rangle \langle \phi_i| |\phi_j\rangle \langle \phi_j| = \sum_i \rho_i^2 |\phi_i\rangle \langle \phi_i|$$

This will be the same as $\rho$ if all the eigenvalues are 0 or 1, which can happen only for a pure state. Hence $\rho^2 = \rho$ if and only if $\rho$ is a pure state. If you take the trace of $\rho^2$ the result will be

$$\text{Tr}(\rho^2) = \sum_i \rho_i^2 \leq \sum_i \rho_i = 1.$$  

The inequality occurs because $0 \leq \rho_i \leq 1$, and the inequality becomes an equality only if all the eigenvalues are 0 or 1, which again implies a pure state, so

$$\text{Tr}(\rho^2) = 1 \iff \rho \text{ is a pure state}.$$  

You can also take the logarithm of a positive definite matrix. One measure of how mixed a mixed state is is the quantum mechanical entropy, defined as

$$S \equiv -k_B \text{Tr}(\rho \ln \rho) = -k_B \sum_i \rho_i \ln \rho_i$$

where $k_B$ is Boltzmann’s constant. Entropy vanishes for pure states and is positive for mixed states.

The evolution of the state operator can be computed with the help of Schrödinger’s equation:

$$\frac{d\rho}{dt} = \sum_i \left\{ f_i \frac{d}{dt} |\Psi_i\rangle \langle \Psi_i| + f_i |\Psi_i\rangle \langle \Psi_i| \frac{d}{dt} |\Psi_i\rangle \right\}$$

$$= \frac{1}{i\hbar} \sum_i \left\{ f_i H |\Psi_i\rangle \langle \Psi_i| - f_i |\Psi_i\rangle \langle \Psi_i| H \right\} = \frac{1}{i\hbar} (H \rho - \rho H),$$

$$\frac{d}{dt} \rho(t) = \frac{1}{i\hbar} [H, \rho(t)].$$  

(11.26)
It is important not to confuse the (Schrödinger picture) state operator Schrödinger equation (11.26) with the Heisenberg evolution of observables (11.19). They look very similar, and this is not a coincidence, but there is a sign difference. Note that if we know the state operator at one time (11.26) allows us to compute it at any other time; we do not need to know the individual components $|\Psi_i\rangle$ and their corresponding probabilities $f_i$ to compute the evolution of the state operator.

Now, suppose we perform a measurement corresponding to an observable $A$ with complete orthonormal basis $\{a, n\}$, where $a$ corresponds to the eigenvalue under the operator $A$. What is the probability of obtaining the result $a$? There are two kinds of probability to deal with, the probability that we are in a given state $|\Psi\rangle$ to begin with, and the probability that such a state yields the designated eigenvalue. The combined probability is

$$P(a) = \sum_i f_i P_i(a) = \sum_i f_i \sum_n \langle a, n | \Psi_i \rangle \langle \Psi_i | a, n \rangle,$$

$$P(a) = \sum_n \langle a, n | \rho | a, n \rangle.$$

The expectation value of the operator is even easier to compute:

$$\langle A \rangle = \sum_i f_i \langle A \rangle_i = \sum_i f_i \langle \Psi_i | A | \Psi_i \rangle = \sum_i f_i \sum_j \langle \Psi_i | A | \phi_j \rangle \langle \phi_j | \Psi_i \rangle = \sum_i \sum_j f_i \langle \phi_j | \Psi_i \rangle \langle \Psi_i | A | \phi_j \rangle = \sum_j \langle \phi_j | \rho A | \phi_j \rangle = \text{Tr} (\rho A).$$

Hence we don’t need to know the individual probabilities and state vectors, all we need is the state operator.

The point I am making here is that if you know the state operator at one time, you can find it at all later times. Furthermore, you can predict the probability of obtaining different outcomes at future times. This suggests that it is possible to rewrite the postulates of quantum mechanics entirely in terms of the state operator. We illustrate with the first postulate:

**Postulate 1:** The state of a quantum mechanical system at time $t$ can be described as a positive semi-definite Hermitian operator $\rho(t)$ with trace one acting on a complex vector space with a positive definite inner product.

If you are going to change this postulate you will have to change the others as well. The second postulate (Schrödinger’s equation) will get replaced by equation (11.26), and the fourth will be replaced by (11.27). The third postulate requires no modifications. The fifth postulate, which describes what happens immediately after a measurement, needs to be rewritten as well. Deriving the corresponding formula would require a bit of discussion of probability theory, but I would rather simply give the relevant postulate.
**Postulate 5:** If the results of a measurement of the observable \( A \) at time \( t \) yields the result \( a \), the state operator immediately afterwards will be given by

\[
\rho(t^{'}) = \frac{1}{P(a)} \sum_{n,n'} |a,n\rangle \langle a,n| \rho(t)|a,n'\rangle \langle a,n'|.
\]  

(11.28)

The approach of working with only state operators has several advantages: it allows one to simultaneously deal with both quantum and classical indeterminacy in the same language; the irrelevant overall phase of a state vector does not appear in the state operator; and adding a constant to the Hamiltonian obviously has no effect, since it gives no contribution to (11.26). But the use of matrices when column vectors suffice adds a level of complexity that I would prefer to avoid, and I also feel it is advantageous to distinguish between things that are unknowable and things that are simply unknown.

It is also possible to combine some of these ideas; for example, we can come up with a Heisenberg-state operator approach where the particles are described by static state operators and it is other operators that change over time. Then the state operator changes only due to collapse of the state operator, as given by (11.28).

It should be noted that under time evolution, pure states remain pure, and mixed states remain mixed. To demonstrate this, simply look at the time derivative of the trace of \( \rho^2 \):

\[
\frac{d}{dt} \text{Tr}(\rho^2) = \frac{1}{i\hbar} \text{Tr}([H, \rho] \rho + \rho[H, \rho]) = \frac{1}{i\hbar} \text{Tr}(H \rho^2 - \rho^2 H) = 0.
\]

The same is true of other traces of functions of the state operator, including the entropy. Thus pure states remain pure, and mixed states remain mixed, at least under ordinary evolution, though measurement can, in principle, return a mixed state to a pure state. This might make one wonder how a mixed state can ever be produced. The answer will be explored in a later section of this chapter.
G. Working with the State Operator

Whether or not you believe the postulates should be written in terms of the state operator, it comes in handy in a great many circumstances. For example, suppose I have a single spin-1/2 particle, and I am only interested in its spin. Suppose the particle is in an eigenstate of $S_z$, but we don’t know which one, so it is half the time in the state $|+\rangle$ and half the time in the state $|-\rangle$. Then the state operator is

$$\rho = \frac{i}{2}(|+\rangle\langle +| + |-\rangle\langle -|) = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (11.29)$$

Suppose instead we put it in an eigenstate of

$$S_\theta = \cos \theta S_z + \sin \theta S_x \quad (11.30)$$

which has eigenvectors

$$|+\rangle = \cos \left(\frac{\theta}{2}\right)|+\rangle + \sin \left(\frac{\theta}{2}\right)|-\rangle \quad \text{and} \quad |\neg\rangle = -\sin \left(\frac{\theta}{2}\right)|+\rangle + \cos \left(\frac{\theta}{2}\right)|-\rangle. \quad (11.31)$$

It is then not too hard to show that the state operator is still

$$\rho = \frac{i}{2}(|+\rangle\langle +| + |\neg\rangle\langle \neg|)$$

$$= \frac{1}{2} \begin{bmatrix} \cos \left(\frac{\theta}{2}\right) & \cos \left(\frac{\theta}{2}\right) \\ \sin \left(\frac{\theta}{2}\right) & -\sin \left(\frac{\theta}{2}\right) \end{bmatrix} \begin{bmatrix} \cos \left(\frac{\theta}{2}\right) & \sin \left(\frac{\theta}{2}\right) \\ -\sin \left(\frac{\theta}{2}\right) & \cos \left(\frac{\theta}{2}\right) \end{bmatrix} + \frac{1}{2} \begin{bmatrix} -\sin \left(\frac{\theta}{2}\right) & \cos \left(\frac{\theta}{2}\right) \\ \cos \left(\frac{\theta}{2}\right) & \sin \left(\frac{\theta}{2}\right) \end{bmatrix}$$

$$= \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (11.32)$$

Since (11.29) and (11.32) are identical the two situations are physically indistinguishable. Thus it is not important to distinguish between them, and in either case we simply say that the particle is unpolarized. If we need to calculate the probability of an unpolarized electron interacting in a particular way, we don’t have to consider electrons in all possible initial states, instead we can simply pick a basis (such as eigenstates of $S_z$) and then average the results for each of the two spin states.

Consider the following commonplace situation. Suppose we have a source of particles which produces particles in a well-defined quantum state $|\varphi_0\rangle$, but at an unknown time $t_0$. In other words,

$$|\Psi(t_0)\rangle = |\varphi_0\rangle$$

We wish to find the state vector at an arbitrary time. Formally, we can do so with the help of (11.1) and (11.5), rewriting it in terms of energy eigenstates:

$$|\Psi(t)\rangle = U(t,t_0)|\Psi(t_0)\rangle = \sum_n |\varphi_n\rangle \exp \left[ -\frac{i}{\hbar} E_n (t-t_0) \right] \langle \varphi_n | \Psi_0 \rangle.$$

If we know the exact time $t_0$ when the wave is produced, we can now write the state operator as
\[ \rho(t) = \langle \Psi(t) | \Psi(t) \rangle = \sum_n \sum_m \langle \phi_n | \Psi_0 \rangle \langle \Psi_0 | \phi_m \rangle \exp \left[ \frac{i}{\hbar} (E_m - E_n)(t - t_0) \right] |\phi_n \rangle \langle \phi_m |. \]

However, the initial time \( t_0 \) is assumed to be completely unknown. We therefore average it over a long time interval of length \( T \), running from \(-\frac{1}{2}T\) to \(+\frac{1}{2}T\), and average over this time interval.

\[ \rho(t) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} dt_0 \sum_n \sum_m \langle \phi_n | \Psi_0 \rangle \langle \Psi_0 | \phi_m \rangle \exp \left[ \frac{i}{\hbar} (E_m - E_n)(t - t_0) \right] |\phi_n \rangle \langle \phi_m |. \]  \hspace{1cm} (11.33)

Focus for the moment on the limit of the integral of \( t_0 \). We can perform it to find

\[ \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} \exp \left[ -\frac{i}{\hbar} (E_m - E_n)t_0 \right] dt_0 = \lim_{T \to \infty} \frac{\sin \left[ \frac{T}{\hbar} (E_m - E_n)T / \hbar \right]}{\frac{T}{\hbar} (E_m - E_n)T / \hbar}. \]  \hspace{1cm} (11.34)

Since the sine function is limited to never be bigger than 1, and the denominator blows up at infinity, this limit is zero unless \( E_m - E_n \) vanishes. When this expression vanishes, we see that we have been not very careful with the integration, but it is obvious from the left-hand side of (11.34) that the limit is just 1. In other words,

\[ \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} \exp \left[ -\frac{i}{\hbar} (E_m - E_n)t_0 \right] dt_0 = \delta_{E_mE_n}. \]

If we assume that the energy levels are non-degenerate, this is the same as \( \delta_{nm} \).

Substituting this into (11.33), we see that

\[ \rho(t) = \sum_n \sum_m \delta_{nm} \langle \phi_n | \Psi_0 \rangle \langle \Psi_0 | \phi_m \rangle \exp \left[ \frac{i}{\hbar} (E_m - E_n)t \right] |\phi_n \rangle \langle \phi_m | = \sum_n |\phi_n \rangle \langle \phi_n | \langle \Psi_0 | \Psi_0 \rangle. \]

In other words, no matter what the actual state vector that is produced by our source, we can treat it as if it is in an unknown energy eigenstate, where the probability of it being in such an energy eigenstate is just given by \( |\langle \phi_n | \Psi_0 \rangle|^2 \). For example, if we are studying scattering, we can treat the incoming wave as if it is a monochromatic plane wave, even if it is not. We simply have to sum over all possible energy values in the end with the probability corresponding to the overlap of our initial state vector with the energy eigenvectors. Hence if we have a source of waves, all we need to know is the spectrum of those waves, the set of energy values, not the details of how those waves are produced.
H. Separability

One of the odder features of quantum mechanics is the fifth postulate, or collapse of the state vector. According to this postulate, when any property of a particle is measured, the state vector changes instantly and everywhere. As a simple example, consider a state vector describing the spin state of two electrons, initially in a spin-0 state:

$$|00\rangle = \frac{1}{\sqrt{2}} (|+\rangle - |\rangle - +\rangle).$$

(11.35)

This is an example of an entangled state, because it cannot be written as the tensor product of the spins of the two particles separately. Now imagine separating the two electrons by a large distance, and then measuring the spin of one or the other, as sketched in Fig. 11-2. If we measure the spin of the first particle, the resulting spin state will be given by

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

(11.36)

depending on whether the outcome is positive or negative. Note that the second particle’s spin has instantly been modified based on the outcome of the first particle’s spin measurement. This apparently violates relativity, which demands that information can never travel faster than light. It is possible in principle to measure the spin of the second particle nearly instantaneously, before information could possibly travel the intervening distance, even at light speed. Experimentally, the two electrons always come out with opposite spins, which seems to suggest that information is being transmitted instantaneously, or at least faster than light. This apparent paradox was first discussed by Einstein, Podolsky, and Rosen, and is called the EPR paradox. It might seem possible to somehow use this in a practical way to, say, communicate faster than light, but this is not the case.

Let’s try to demonstrate this in general. Suppose our system is separable, meaning it consists of two particles, or two sets of particles, sufficiently distant from each other that there is no interaction between them. In other words, our state space is a vector product space $\mathcal{V}_1 \otimes \mathcal{V}_2$, and our Hamiltonian will consist of two pieces

$$H = H_1 + H_2,$$

where $H_1$ and $H_2$ affect only the vector space $\mathcal{V}_1$ and $\mathcal{V}_2$ respectively. Let the basis states for this vector space be $\{|\phi_i, \chi_j\rangle\}$, where $|\phi_i\rangle$ and $|\chi_j\rangle$ are basis states for $\mathcal{V}_1$ and $\mathcal{V}_2$.
respectively. To be as general as possible, we will imagine that we describe the initial state as a density operator with components
\[ \rho_{ij,kl} = \langle \phi_i, \chi_j | \rho | \phi_k, \chi_l \rangle. \]

Now, suppose we measure some operator \( A \) which acts only on the space \( \lambda' \). In other words, in components,
\[ A_{ij,kl} = \langle \phi_i, \chi_j | A | \phi_k, \chi_l \rangle = \langle \phi_i | A | \phi_k \rangle \langle \chi_j | \chi_l \rangle = A_{ik} \delta_{jl}. \]

For example, we can look at the expectation value
\[
\langle A \rangle = \text{Tr}(A \rho) = \sum_{ij} \langle \phi_i, \chi_j | A \rho | \phi_i, \chi_j \rangle = \sum_{ijkl} \langle \phi_i, \chi_j | A | \phi_k, \chi_l \rangle \langle \phi_i, \chi_j | \rho | \phi_k, \chi_l \rangle
\]
\[
= \sum_{ijkl} A_{ij,kl} \rho_{kl,ij} = \sum_{ik} A_{ik} \sum_{jl} \rho_{ij,kl}.
\]

We recognize the inner sum as a partial trace, i.e., define
\[ \rho_i = \text{Tr}_2(\rho), \]
then we simply have
\[
\langle A \rangle = \text{Tr}_1(A \rho_i).
\]

where the trace is now only going over the smaller space \( \lambda' \). Hence in this simple case we can ignore the second particle (or group of particles) entirely, and work only with the first particle.

Now, suppose instead that a distant observer measures the second set of particles, using some operator \( B \), and determines some aspect of these other particles. The outcome of their measurement, we will assume, is not communicated to us by conventional means, but their measurement will inevitably cause a collapse of the state vector. Wouldn’t we notice this effect? To make things simple, assume that the basis states \( \{ | \chi_j \rangle \} \) are eigenstates of \( B \) with eigenvalues \( b_j \). Then the probability that the distant observer measures the result as \( b_j \) is
\[
P(b_j) = \sum_i \langle \phi_i, \chi_j | \rho | \phi_i, \chi_j \rangle = \sum_i \rho_{ij,ij}.
\]

Afterwards, if the result of the measurement is \( b_j \), the state operator is
\[
\rho^+_{b_j} = \frac{1}{P(b_j)} \sum_{i,k} | \phi_i, \chi_j \rangle \langle \phi_i, \chi_j | \rho | \phi_k, \chi_j \rangle | \phi_k, \chi_j \rangle \langle \phi_k, \chi_j |.
\]

However, if we do not know the result of a measurement, but only that a measurement occurs, we must average all possible outcomes for the state operator, weighted by the corresponding probability, so we have
\[
\rho^+ = \sum_j P(b_j) \rho^+_{b_j} = \sum_j \sum_{i,k} | \phi_i, \chi_j \rangle \langle \phi_i, \chi_j | \rho | \phi_k, \chi_j \rangle | \phi_k, \chi_j \rangle \langle \phi_k, \chi_j |,
\]
or in components,

\[ \rho_{ij,kl}^* = \delta_{ji} \rho_{ij,kj} \cdot \]

If we now take a partial trace after the measurement on the second particle, we find

\[ (\rho^*)_{ik} = \sum_j \rho_{ij,kj}^* = \sum_j \delta_{ji} \rho_{ij,kj} = \sum_j \rho_{ij,kj} = (\rho_1)_{ik} , \]

i.e., the measurement on the second particle has no influence whatsoever on the reduced state operator \( \rho_1 \). We can once again disregard the second particle if we are going to focus on the first.

To illustrate these ideas more concretely, if we start in the pure state described by (11.35), and work in a basis \( \{ |++\rangle, |+-\rangle, |-+\rangle, |--\rangle \} \), the initial density matrix will be

\[
\rho = |00\rangle \langle 00| = \frac{1}{2} \begin{pmatrix}
0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix} = \frac{1}{2} \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 \\
0 & -1 & 1 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}.
\]

After measuring the spin of the second particle, the state vector will be in one of the two states (11.36), with equal probability, and hence the density matrix after the measurement will be

\[
\rho^* = \frac{1}{2} |++\rangle \langle ++| + \frac{1}{2} |+-\rangle \langle +-| + \frac{1}{2} |-_+\rangle \langle _-+| = \frac{1}{2} \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}.
\]

The reduced density matrix will be given by tracing over just the second particle. Before the measurement, it is given by

\[
\rho_1 = \frac{1}{2} \begin{pmatrix}
\text{Tr} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} & \text{Tr} \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix} \\
\text{Tr} \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} & \text{Tr} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}
\end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
\]

The reduced density matrix after the measurement can be found by taking a similar partial trace of (11.37), and will be identical to (11.38). So we see that the measurement on the distant particle 2 has no effect on the reduced density matrix.

A final concern might be that the evolution of the state operator due to (11.26) might somehow allow faster-than-light communication. The expectation value of the operator \( A \) will evolve according to
However, since $A$ and $H_2$ are acting on separate spaces, they commute, and hence $H_2$ causes no change in the expectation values. It can further be shown that $H_2$ causes no change in the reduced state operator $\rho_1$, so once again it is impossible for it to have any effect. In summary, in a separable system, so long as we measure and interact with only one component of the system, we can work with the reduced state operator involving only that component, by taking the trace over the other component. This allows one, for example, to ignore the large (infinite?) number of particles an electron may have interacted with in the past when focusing on that one particle. We also see clearly that using entangled states in no way provides a mechanism for instantaneous (faster than light) communication. It is possible, of course, that some other mechanism might provide such a means, such as a non-local term in the Hamiltonian, but quantum mechanics per se does not demand it.

A comment is in order here: when you reduce a system from two components to one the entropy of the system may well change. In particular, if you start with a pure state, then if you lose track of one or more particles after an interaction, you may end up with a mixed state. As an example, if you have two electrons in a spin-0 state (a pure state), and one of them escapes your experiment, you will end up in a completely unpolarized state (a mixed state), if you ignore the lost particle.

I. Hidden Variables and Bell’s Inequality

It is worth noting how different the fourth and fifth postulates concerning measurement are from the second postulate, Schrödinger’s equation. Schrödinger’s equation describes the evolution of the state vector which is continuous, predictable and reversible. It is a differential equation, and hence continuous. It is predictable, in that there is no probability. And as argued at the start of this chapter, the time evolution operator is unitary, which means it has an inverse, so one can always reconstruct the initial state if you know the final state. In contrast, the fourth postulate specifically mentions probability, and the fifth describes a discontinuous and irreversible process. For example, suppose a spin-1/2 particle is initially in one of the two eigenstates of $S_z$, that is, in one of the two states

$$|+,z\rangle = \frac{1}{\sqrt{2}} (|+\rangle + |\rangle) \quad \text{or} \quad |-,z\rangle = \frac{1}{\sqrt{2}} (|+\rangle - |\rangle). \quad (11.38)$$

If we measure $S_z$, we will now get a plus result half of the time and a minus result the other half. The fourth postulate states that we can only predict this in a probabilistic way. Suppose for example that the result is now plus, then the state vector after the measurement will be $|+\rangle$. Indeed, there is no way to deduce from the state vector afterwards which of the two states (11.38) was the initial state.
Although we argued in section H that we cannot transmit any useful information via quantum mechanics, it appears that quantum mechanical information is transmitted instantaneously. Measuring the spin of one electron in an EPR pair instantly changes the other. Of course, this instantaneous transmission of information is demanded only provided the Copenhagen interpretation is the correct one, but there are many alternative possibilities. Let’s discuss the concept of hidden variables. A hidden variable is some parameter in an experiment which is unknown to the experimenter, and perhaps unknowable, which determines the outcome. As a trivial example, suppose we take two cards, one red and one black, mix them up, place them in two opaque envelopes, and then separate them by a large distance. Initially, we have no idea which card is red and which is black. However, as soon as we open one envelope, we know not only the color of that card, but the color of the other as well. There is no relativistic time delay, and no spooky action at a distance. Could this be all that is occurring in quantum mechanics?

I would now like to demonstrate that this is not all that is going on. Consider the following two assumptions, known as the philosophy of local hidden variables:

1. The outcome of a particular experiment is definite; i.e., if we could actually know all the hidden variables (and we had the correct theory), we could predict the outcome; and
2. It is experimentally possible to separate two particles such that they cannot influence each other (separability). In particular, you can separate them by performing the measurements at widely separated points at simultaneous (or nearly simultaneous) times, so that information cannot be transmitted from one to the other by the limitations of speed of light (Einstein separability).

In particular, it should be noted that these are in contradiction with the Copenhagen interpretation, since measurement has probabilistic outcomes, and collapse of the state vector is a non-local event.

I would now like to demonstrate that for the simple experiment sketched in Fig. 11-2, the predictions of quantum mechanics disagree with these assumptions. Consider measuring the spin of particle 1 along one of two axes, marked A and C, and the spin of the second along one of two other axes, marked B and D, as illustrated in Fig. 11-3. We would like to discuss the probability that the results of these measurements, which we name $a$, $b$, $c$, and $d$, will be unequal, for example, $P(a \neq b)$ represents the probability that, when you set your Stern-Gerlach apparatus up such that you measure spin one on axis A and spin two on axis B, you get + for one and – for the other. According to the philosophy of local hidden variables, we can discuss these probabilities even if we don’t actually perform the measurements. Furthermore, the outcome of the measurement along one axis, say A, cannot affect the outcome of B, since the two particles are physically separated.

Figure 11-3: The first spin is measured along axis A or C, and the second along B or D. The resulting probabilities predicted by quantum mechanics contradict the philosophy of local hidden variables.
Note that it is obvious that
\[
(a = b \text{ and } b = c \text{ and } c = d) \implies a = d.
\]
Logically, this statement can be reversed to conclude that
\[
a \neq d \implies (a \neq b \text{ or } b \neq c \text{ or } c \neq d).
\]
Thus in any circumstance where \(a \neq d\), at least one of the other three statements must be true. In terms of probabilities, this leads to an inequality called Bell’s Inequality, named after John S. Bell:\(^1\)
\[
P(a \neq d) \leq P(a \neq b) + P(b \neq c) + P(c \neq d).
\] (11.39)
This prediction follows simply from the nature of the experiment and the assumptions of local hidden variables. No assumptions about the nature of the experiment or the method of measurement are involved. Nonetheless, this result disagrees with the predictions of quantum mechanics.

It is easy to see from (11.35) or (11.36) that if you start with a pair of electrons in a spin-0 state, and you measure the spin of each along the same axis, say, \(S_z\), the resulting spins will always come out to be opposite. Suppose instead that you measure the spin of the first along the \(z\)-axis and the spin of the second along some other axis, \(S_\theta\), measured at an angle \(\theta\) compared to the first axis, as illustrated by equation (11.30). The eigenvectors of \(S_\theta\) are given in (11.31). Then it is not hard to show that the probability of getting a different result for the two spins is given by
\[
P(a \neq b) = \cos^2 \left( \frac{\pi}{2} \theta \right).
\] (11.40)
where \(\theta\) is the angle between the two measurements. Although the argument has been made assuming one measurement is along the \(z\)-axis, it is actually pretty easy to show that it is valid for any angle; all that matters is the relative angle between the two measurements.

We can use (11.40) to calculate the probability that the outcome of any pair of measurements will be unequal. Because of the angles involved, we see that
\[
P(a \neq b) = P(b \neq c) = P(c \neq d) = \cos^2 \left( \frac{\pi}{4} 135^\circ \right) = 0.1464,
\]
\[
P(a \neq d) = \cos^2 \left( \frac{\pi}{4} 45^\circ \right) = 0.8536.
\]
We now test this prediction against Bell’s inequality, (11.39) which would then demand
\[
0.8536 \leq 0.1464 + 0.1464 + 0.1464 + 0.1464 = 0.4391.
\]
This inequality is clearly false, and hence quantum mechanics yields predictions that violate the philosophy of local hidden variables.

Of course, just because local hidden variable disagrees with quantum mechanics doesn’t mean it is incorrect. The relevant question is whether Bell’s inequality, (11.39),

---

\(^1\) The presentation of Bell’s inequality is different than the one I usually see, but logically is a bit simpler. I suppose I could call it Carlson’s inequality, but it is such a similar argument to the one I usually read that I don’t deserve any credit for it.
is violated experimentally. Although this experiment with spins has never, to my
knowledge, been done, similar experiments involving photons have shown with a high
level of confidence that Bell’s inequality is violated.

If you wish to cling to the hidden variables philosophy, there are many possible
ways of rescuing the viewpoint. It is quite possible that the probabilistic outcome of a
particular experiment is not due to some sort of hidden variable associated with the
particle, but with the measuring apparatus. Hence, in a sense we can’t ask what the
outcome of C would have been unless the experiment is set up to measure C. One could
imagine, for example, that some sort of interaction between the measuring device and the
source causes the electrons to leave the source in a different way, depending on what
angle the detector for particle one is set up for. If the experiment is set up well in
advance, this transfer of information from the Stern-Gerlach device to the source can
occur slower than the speed of light.

One could imagine attempting to get around this problem by making sure to
adjust the measuring devices only at the last moment, while the electrons are en route to
the devices. It is not physically feasible to rotate the massive Stern-Gerlach devices in
that short a time, but if you instead measure the polarization of photons, it is possible to
adjust the light en route based on last-moment adjustments, for example, via a computer
controlled system which chooses which axis based on a random number generator
implemented via opto-electronics. Such “delayed choice” experiments have been done,
in limited ways, and again confirm the predictions of quantum mechanics. A die-hard fan
might point out that computer random number generators, despite the word “random,” are
actually deterministic, and hence, in a sense, the choice is made long before the
experiment is begun. At least as a gedanken experiment, one could then imagine
replacing the computer with a graduate student with fast reactions, or perhaps slightly
more realistically, make the experiment astronomical in scale. This leads into questions
of whether graduate students (or people in general) actually have free will, which is
beyond the scope of this discussion.

For hardcore fans of local hidden variables, it should be noted that there is a
loophole. All real experiments involve a certain fraction of particles that escape the
detector, and hence are lost. For individual particles, such as electrons or photons, this
loss rate is rather high. Experimentally, you can show that the probability of getting
equal or unequal results on two spins remains pretty much constant as you increase the
efficiency of your detectors. It is normally assumed that these correlated probabilities
would stay the same, even if you had perfect detectors. Indeed, though showing
experimentally that Bell’s inequality is violated requires high efficiency detectors, it does
not require perfect detectors. However, the actual efficiency to formally demonstrate
such a violation is very high, significantly higher than we can currently achieve
experimentally. It is logically and philosophically possible that Bell’s inequality is not
experimentally violated. If such is the case, then the measured disagreement rates
\[ P(a \neq b) \]
must change in a statistically significant way as we improve our measurements.

Though it is possible to imagine such a result, it would mean a complete rewriting of
much (or all) of what we understand about physics. Every experiment would suddenly
become suspect, and all of the successes of quantum mechanics might be simply artifacts
of our imperfect measuring devices. Most physicists do not take such speculations
seriously.
If you look through the postulates as presented in chapter two, it is notable that three of the five postulates concern measurement. Most physicists would have no particular trouble with the first two postulates. True, some prefer the Heisenberg picture to the Schrödinger, which requires rewriting the postulates a bit, or the path integral formulation, or even working with the state operator instead of state vectors.

But the real problems come when we discuss measurement. There is nothing particularly troubling about the third postulate, which merely states that measurements correspond to Hermitian operators. It is the last two postulates – describing the probability of various measurement outcomes, and how the state vector changes when you perform that measurement – that cause concern.

What actually occurs during a measurement? As a theorist, I generally don’t want to know how measurements are performed, but experimentally, measurements always involve interactions between specific particles. To detect an electron, for example, one could measure the electric force as it passes near another charged particle, or to detect a photon, one could use a photomultiplier. If you ask an experimenter how it works, he will not say something like, “well, it just collapses the state vector.” Instead there will be an explanation of how in a photomultiplier, a photon is absorbed by a piece of metal, causing an electron to be released via the photoelectric effect, which then moves to another piece of metal, knocking loose still more electrons, etc., finally creating such a cascade that we can simply measure the macroscopic current, and our computer increases its count of photons by one. The point is that there is nothing magical about detection; it is merely a special type of interaction, or another term in the Hamiltonian. The only thing that distinguishes measurement from some other type of interaction is that ultimately the effects become large, large enough for us to notice it even with our unaided senses.

If interactions represent nothing more than terms in the Hamiltonian, then we should be able to describe its effects in terms of some sort of time evolution operator. In order to do so, we must treat the measuring apparatus in a quantum mechanical way as well. For example, suppose you wish to measure the spin $S_z$ of a single electron, whose initial spin state is completely unknown. Imagine a measuring device $M$ with three mutually orthogonal states, $|M_0\rangle$, $|M_+\rangle$ and $|M_-\rangle$, the first representing the initial state of the measuring device, before a measurement has occurred, and the others representing the measuring device after performing a measurement and obtaining the results + or – respectively. The initial state of the system, therefore, is

$$|\Psi(t_0)\rangle = |\chi\rangle \otimes |M_0\rangle = |\chi, M_0\rangle,$$

where $|\chi\rangle$ is the electron’s spin state. If we perform a measurement on either of the states $|\pm\rangle$, we expect the interaction to lead to a time evolution operator such that the measuring device changes its quantum state, so

$$U(t_1,t_0)|+, M_0\rangle = |+, M_+\rangle,$$

$$U(t_1,t_0)|-, M_0\rangle = |-, M_-\rangle.$$
Thus far, we have only used Schrödinger’s equation to change the state vector of the system.

What will happen if we start in a mixed state, namely,

$$|\Psi(t_0)\rangle = |+,\rangle \otimes |M_0\rangle = \frac{1}{\sqrt{2}} (|+,M_0\rangle + |-,M_0\rangle)$$  \hspace{1cm} (11.41)$$

Then the time evolution operator $U(t_1,t_0)$, because it is linear, must evolve this state to

$$|\Psi(t_1)\rangle = U(t_1,t_0)|\Psi(t_0)\rangle = \frac{1}{\sqrt{2}} (|+,M_+\rangle + |-,M_-\rangle).$$  \hspace{1cm} (11.42)$$

Thus far we have not yet collapsed the state vector. Should we? That depends on when we think a measurement occurs. If we think of measurement as something that happens just before the interaction with the measuring device, then we have done our analysis incorrectly. We should split our state vector into two cases, each with 50% probability, before the interaction with the measuring device, and analyze it something like this:

$$\frac{1}{\sqrt{2}} (|+,M_0\rangle + |-,M_0\rangle) \rightarrow |+,M_+\rangle, \quad (50\%),$$

$$|-,M_0\rangle \rightarrow |-,M_-\rangle, \quad (50\%).$$

If, on the other hand, we think measurement occurs after the interaction, then our analysis looks like this

$$\frac{1}{\sqrt{2}} (|+,M_0\rangle + |-,M_0\rangle) \rightarrow \frac{1}{\sqrt{2}} (|+,M_+\rangle + |-,M_-\rangle) \rightarrow |+,M_+\rangle, \quad (50\%),$$

$$|-,M_-\rangle, \quad (50%).$$  \hspace{1cm} (11.43)$$

Note that in each case, we see that the result is the same.

Is there any way experimentally to determine when exactly the collapse of the state vector occurs? To do so, we would have to intervene in the intermediate state, and determine if the state vector is a superposition, or in one of the two states. For example, suppose we measure the expectation value of $S_x$, the spin along the $x$-axis. The initial state (11.41) can easily be shown to have a non-zero expectation value for $S_x$:

$$\langle \Psi(t_0)|S_x| \Psi(t_0)\rangle = \frac{1}{2} (\langle +,M_0|+,-,M_0) S_x |+,M_0\rangle + |-,M_0\rangle)$$

$$= \frac{1}{2} \cdot \frac{1}{2} \hbar \left[ (+,M_0|+,-,M_0)(|+,M_0\rangle + |-,M_0\rangle) \right] = \frac{1}{2} \hbar .$$

Once the state vector of the electron has collapsed, however, it will be in one of the two eigenstates $|\pm\rangle$, each of which has $\langle S_x \rangle = 0$, and so this seems like a way we could tell if the state vector has collapsed. On the other hand, using (11.42), we see that

$$\langle \Psi(t_1)|S_x| \Psi(t_1)\rangle = \frac{1}{2} (\langle +,M_+|+,-,M_-\rangle S_x |+,M_+\rangle + |-,M_-\rangle)$$

$$= \frac{1}{2} \cdot \frac{1}{2} \hbar \left[ (+,M_+|+,-,M_-)(|+,M_+\rangle + |-,M_-\rangle) \right] = 0.$$  \hspace{1cm} (11.43)$$

In other words, the expectation value of $S_x$ will vanish whether or not the state vector actually collapses. This occurs simply because the measuring device ends up in an entangled state with the electron.
Of course, we know that the state vector must *eventually* collapse, since obviously in the end we will observe the electron to be either spin up or spin down, not both. To make things more interesting, let’s imagine the measuring device $M$ feeds its information to a computer $C$, which is then read by an experimenter $E$, as illustrated in Fig. 11-4. We might initially describe the state as

$$|\Psi(t_0)\rangle = |\psi_s\rangle \otimes |M_0\rangle \otimes |C_0\rangle \otimes |E_0\rangle = |\psi_s, M_0, C_0, E_0\rangle.$$  

We could then imagine a process where there is a series of interactions, where first the measuring device interacts with the electron, which then transmits the information to the computer, whose data is read out by the experimenter. If we start in a pure spin up or spin down state, this state will evolve according to the steps

$$|+, M_0, C_0, E_0\rangle \rightarrow |+, M_+, C_0, E_0\rangle \rightarrow |+, M_+, C_+, E_0\rangle \rightarrow |+, M_+, C_+, E_+\rangle,$$

$$|-, M_0, C_0, E_0\rangle \rightarrow |-, M_-, C_0, E_0\rangle \rightarrow |-, M_-, C_-, E_0\rangle \rightarrow |-, M_-, C_-, E_-\rangle.$$  

If we instead start with a spin state $|\psi_s\rangle = |+,\rangle$, then at each step we would end up with a superposition, not just of the spin state, but also of the measuring device, computer, and then the experimenter. No matter at which step the collapse of the state vector is assumed to occur, in the end we find out that half the time the experimenter ends up learning the electron is spin up, and half the time spin down. In summary, the predictions of quantum mechanics do not change as you change the moment of collapse.

It is worth adding one caveat to this: you mustn’t collapse the state vector too early. For example, it is possible to interact a single photon with a second photon such that information about the first photon becomes entangled with the second; in essence, the second photon now contains a “measurement” of the first photon. However, it is possible to interact the measuring photon with the initial photon again, in such a way that it becomes disentangled. Basically, this is just saying that if the “measurement” is sufficiently small, it is possible to reverse the process (since the time-evolution operator is always reversible) and return to the original, pre-measured state. Since this can be demonstrated experimentally, it follows that such small-scale measurements must *not* collapse the state vector.

---

1 Things that are obvious are not necessarily true. As we will argue in the next section, it is possible to imagine a wave function that never collapses.
Ballantine argues that the state vector must ultimately collapse, since at some stage the effects are macroscopic. This argument is incorrect. Let’s imagine, for example, that the computer is in a superposition of two states, and the collapse happens the moment the experimenter looks at the computer screen. Then even though the computer is in a superposition, the experimenter will not notice.

Let’s take this to the penultimate extreme. Suppose we treat the experimenter as nothing more than a sophisticated computer, and claim that even the experimenter doesn’t collapse the state vector. Initially, the electron and everything is in the state

\[ \ket{\psi(0)} = \frac{1}{\sqrt{2}} (\ket{+, M_0, C_0, E_0} + \ket{-, M_0, C_0, E_0}) \]

After all the interactions, the state is

\[ \ket{\psi(t)} = \frac{1}{\sqrt{2}} (\ket{+, M_+ C_+ E_+} + \ket{-, M_- C_- E_-}) \]  

We now see that the experimenter is in a superposition. Now when another physicist, say a theorist \( T \), asks the experimenter what the result of the experiment is, the whole thing will collapse, placing the experimenter (finally) in a state which is no longer a superposition.

There seems to be an obvious flaw in this argument. An experimenter is not like a mindless computer; surely she would notice she was in a superposition. We can simply ask her and find out whether she was in the state (11.44). To consider this possibility, we expand our description of the experiment somewhat, namely, because she is not only aware of the spin of the electron, \( |E_+\rangle \), we imagine she might also be aware of whether it is in a definite, non-superposition state \( |D_+\rangle \), or an indefinite, superposition state \( |D_+\rangle \).

For example, suppose a spin + electron is fed into the system, then the initial state would be

\[ \ket{\psi(t_0)} = |+\rangle \otimes (M_0 \otimes C_0 \otimes E_0 \otimes D_0) \]

We now let the measuring device interact, the computer receive the data, and the experimenter see the result:

\[ \ket{\psi(t_1)} = |+\rangle \otimes (M_+ \otimes C_0 \otimes E_0 \otimes D_0) \]
\[ \ket{\psi(t_2)} = |+\rangle \otimes (M_+ \otimes C_+ \otimes E_0 \otimes D_0) \]
\[ \ket{\psi(t_3)} = |+\rangle \otimes (M_+ \otimes C_+ \otimes E_+ \otimes D_0) \]  

(11.45)

The experimenter now notices that her awareness of the spin state is not a superposition, and therefore concludes it is in a definite state, not a superposition.

\[ \ket{\psi(t_4)} = |+\rangle \otimes (M_+ \otimes C_+ \otimes E_+ \otimes D_0) \]  

(11.46)

In a similar way, suppose we start with a spin down electron, then after reading the result off a computer screen, the quantum state would be

\[ \ket{\psi(t_5)} = |-\rangle \otimes (M_- \otimes C_- \otimes E_- \otimes D_0) \]  

(11.47)
Again, her awareness is not in a superposition, and hence she would conclude she is in a definite state:

$$\Psi(t_4) = |>\rangle \otimes |M\rangle \otimes |C\rangle \otimes |E, D\rangle.$$ (11.48)

Now, suppose instead we had started in a mixed state. Then the quantum state after the experimenter observes the computer will be

$$\Psi(t_4) = \frac{1}{\sqrt{2}} \left( |+\rangle \otimes |M\rangle \otimes |C\rangle \otimes |E, D\rangle + |>\rangle \otimes |M\rangle \otimes |C\rangle \otimes |E, D\rangle \right) \quad (11.49)$$

And now, at long last, we are prepared to reach a truly startling conclusion. We know that upon introspection, the experimenter’s internal awareness will evolve from (11.45) to (11.46). We also know that it will evolve from (11.47) to (11.48). Because the time evolution operator is linear, it inevitably follows that (11.49), upon introspection, will change into

$$\Psi(t_3) = \frac{1}{\sqrt{2}} \left( |+\rangle \otimes |M\rangle \otimes |C\rangle \otimes |E, D\rangle + |>\rangle \otimes |M\rangle \otimes |C\rangle \otimes |E, D\rangle \right)$$

Note that both portions of the state vector are in the state $|D\rangle$; that is, the experimenter will incorrectly conclude that the state vector is not in a superposition, even though it is.

Hence the theorist $T$, when he asks whether the experimenter has measured a definite state or a superposition, will always conclude that it is a definite state, even though no collapse has occurred. In this picture (and the Many Worlds Interpretation), the universe will often consist of superpositions of state vectors that are very different, not simply a single electron spin being up or down, but where major events, even astronomical events, are not definite until they are observed.

This allows one to develop a philosophy of quantum mechanics I call solipsism, where you take the attitude that you alone are capable of collapsing the state vector. Particles, measuring devices, computers, and experimenters may be in a superposition up to the moment you interrogate them. One can end up in ridiculous situations, such as the famous Schrödinger’s cat, where a cat is in a superposition of being alive and dead, until the moment you observe it. This viewpoint is self-consistent, and has no experimental contradictions that I am aware of, but seems remarkably egotistical. In the next section, we take this philosophy one step farther and discuss the Many Worlds Interpretation.
The Many Worlds Interpretation, or MWI, is based on the assumption that state vectors never collapse. No one, including myself, is capable of collapsing the state vector. As such, the postulates of quantum mechanics are reduced to just two, reproduced here for completeness:

**Postulate 1:** The state of a quantum mechanical system at time \( t \) can be described as a normalized ket \( \Psi(t) \) in a complex vector space with a positive definite inner product.

**Postulate 2:** The state vector evolves according to

\[
i \hbar \frac{\partial}{\partial t} \Psi(t) = H(t) \Psi(t),
\]

where \( H(t) \) is an observable.

And that’s it! Indeed, postulate 2 is simpler than before, since we no longer need to restrict Schrödinger’s equation to situations where measurement does not occur. Measurement is interaction, governed by the Hamiltonian, and hence does not require a separate postulate.

It is worth mentioning that the two postulates listed above can be changed to the Heisenberg picture, or by using path integrals, or using the state operator instead. I personally prefer Schrödinger’s equation, but don’t really have any strong feelings one way or the other.

The third postulate, namely, that measurements correspond to operators, is no longer needed, because all measurement is governed by Hamiltonians, and since the Hamiltonian is itself an observable, it isn’t necessary to make such a separate assumption. The fifth postulate, concerning collapse of the state vector, is no longer necessary. This does lead to a weird picture of the universe – one in which many possibilities coexist, in some of which this document doesn’t exist, because the author was killed in a car accident, or was never born, or perhaps life itself never came into existence on Earth! The possibility of “many worlds” coexisting seems absurd, but is taken seriously by many physicists.¹

This leaves only the fourth postulate. According to the Copenhagen interpretation, if we measure \( S_z \) for an electron in the spin state \( |\pm, \rangle \), we will get the result + or the result -, with a 50% probability of each. According to MWI, we will get both results, with a state vector that is equal parts of each, though we will (incorrectly) believe we only got one or the other. These results seem similar – but wait, there is a specific statement of probability in Copenhagen. This seems a superiority of Copenhagen over MWI – it predicts a particular probability. If MWI cannot reproduce this success, we must conclude that MWI is wrong, or at least inferior to Copenhagen.

¹ Including me.
Well, experimentally, if we actually have only one electron, giving a probabilistic answer to an experimental question is not really very helpful. If we measure the one electron, and the spin comes out +, have we verified our hypothesis or not? Probabilities are not statements about individual experiments, but rather about a series of repeated experiments in the limit that the experiment is repeated indefinitely. Let’s imagine we have a series of $N$ electrons, each one initially in the state $|+\rangle$. Suppose we measure $S_z$ for each of these, and then average the results. According to the Copenhagen interpretation of quantum mechanics, if you measure $S_z$ of one of these, you will get the result $\frac{1}{\sqrt{2}} \hbar$ half the time and $-\frac{1}{\sqrt{2}} \hbar$ the other half, and over a long series of experiments they will average to zero. Does MWI make the same prediction?

The initial quantum states for all the electrons considered together will be

$$|\psi_N(t_0)\rangle = |+\rangle \otimes |+\rangle \otimes |+\rangle \otimes |+\rangle \otimes \cdots \otimes |+\rangle$$

It is easy to see that

$$S_z |+\rangle = \frac{1}{\sqrt{2}} S_z (|+\rangle + |-\rangle) = \frac{1}{\sqrt{2}} \hbar (|+\rangle - |-\rangle) = \frac{\hbar}{2} |-\rangle$$

Now, we want to measure the average spin $S_z$, which we define as

$$\overline{S}_z = \frac{1}{N} \sum_{i=1}^{N} S_{zi}$$

Now, if we were to perform a measurement of this average spin, the expectation value of the result would be

$$\langle \psi_N | \overline{S}_z | \psi_N \rangle = \frac{1}{N} \sum_{i=1}^{N} \left( \langle +_x | \otimes \langle +_x | \otimes \cdots \otimes \langle +_x | \right) S_{zi} \left( |+_x \rangle \otimes |+_x \rangle \otimes \cdots \otimes |+_x \rangle \right)$$

$$= \frac{\hbar}{2N} \left( \langle +_x | \otimes \langle +_x | \otimes \cdots \otimes \langle +_x | \right)$$

$$= \frac{\hbar}{2N} (0 + 0 + \cdots + 0) = 0.$$

Now, consider also the expectation value of the square of this operator

$$\langle \psi_N | \overline{S}_z^2 | \psi_N \rangle = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \langle +_x | \otimes \langle +_x | \otimes \cdots \otimes \langle +_x | \right) S_{zi} S_{zj} \left( |+_x \rangle \otimes |+_x \rangle \otimes \cdots \otimes |+_x \rangle \right)$$

$$= \frac{\hbar^2}{4N^2} \left( \langle +_x | \otimes \langle +_x | \otimes \cdots \otimes \langle +_x | \right)$$

$$= \frac{\hbar^2}{4N^2} \left( \langle +_x | \otimes \langle +_x | \otimes \cdots \otimes \langle +_x | \right) . \quad (11.50)$$
Now, in the sums in (11.50), there are $N$ terms where the minus sign appears in exactly the same factor, and $N^2 - N$ terms where they are in different factors. All of the former yield 1, and all of the latter yield zero, so that
\[
\langle \psi_N | \overline{S_z}^2 | \psi_N \rangle = \frac{\hbar^2}{4N^2} N = \frac{\hbar^2}{4N}.
\]

We now calculate the uncertainty in this operator, given by
\[
\left( \Delta \overline{S_z} \right)^2 = \langle \psi_N | \overline{S_z}^2 | \psi_N \rangle - \langle \psi_N | \overline{S_z} | \psi_N \rangle^2 = \frac{\hbar^2}{4N}.
\] (11.51)

For any finite $N$, this uncertainty is non-zero, and hence the average value of $\overline{S_z}$ cannot be predicted. However, in the limit $N \to \infty$, (11.51) vanishes. **Hence MWI, in the limit of infinitely repeating the experiment, makes the same prediction as Copenhagen.** Indeed, Copenhagen also makes only a prediction in the limit of infinite repetitions. We can only make predictions for a finite number of repetition insofar as we expect a large number of repetitions to resemble an infinite number of repetitions. But reproducing the results of Copenhagen came at a heavy price — we had to keep track of all the trials simultaneously, resulting in accounting nightmares like (11.50).

Working with the MWI interpretation is inconvenient, to say the least. It is far easier to say that when you measure the spin of an electron, the result comes out with one possibility or the other, and then trace through each possibility separately, than it is to imagine gradually larger and larger parts of the universe ending up in ridiculous superpositions. For this reason, many (most?) physicists reject MWI as absurd, but there is no clear experimental evidence that it is incorrect. The universe looks exactly as we would expect it to look if MWI is correct, and parsimony, or Ockham’s Razor\(^1\), prefers MWI over alternative approaches, such as Copenhagen. Given a choice between the simplicity of postulates for MWI, vs. the ambiguity of “measurement” discussed in the Copenhagen interpretation, I much prefer the former, but reasonable and very intelligent people have disagreed with me.

Because it is easier computationally (and it doesn’t make my head hurt so much), I will return to the Copenhagen interpretation. Mostly we will be solving Schrödinger’s equation, which is more or less the same in every interpretation. But when we need to discuss probabilities and measurements, we will work in Copenhagen.

---

\(^1\) Ockham’s razor, named for the fourteenth century Franciscan friar and logician William of Ockham, states that entities should not be multiplied without necessity. I interpret this as: given a choice between competing explanations, one should choose the one with fewer (or simpler) hypotheses.
Problems for Chapter 11

1. An electron in an unknown spin state $|a\rangle$ is brought into proximity with a second electron in a known spin state $|b\rangle$. We wish to make the spin of the second electron match the first. A quantum Xerox device will copy it onto the second spin, so

$U_{\text{Xerox}} |a,b\rangle = |a,a\rangle$. A quantum teleporter will swap the two spin states, so

$U_{\text{Teleport}} |a,b\rangle = |b,a\rangle$.

(a) By considering the three initial spin states $|a\rangle = |+\rangle$, $|a\rangle = |\rangle$, and $|a\rangle = \frac{1}{\sqrt{2}}(|+\rangle+|\rangle)$, show that the quantum Xerox device is impossible.

(b) By considering the same three initial states, show that the same problem does not apparently occur for the quantum teleport device.

2. At $t = 0$, the wave function of a free particle of mass $m$ is given by

$\Psi(x,t=0) = \left(A/\pi^4\right)^{1/2} \exp\left[-\frac{1}{2} A(x-x_0)^2\right]$. Find $\Psi(x,t)$, using the propagator.

3. This problem should be worked entirely in the Heisenberg formulation of quantum mechanics. A particle lies in the one-dimensional harmonic oscillator potential, so $H = P^2/2m + \frac{1}{2} m \omega^2 X^2$.

(a) Work out $dX/dt$ and $dP/dt$.

(b) Define the operators $a(t) = \sqrt{m\omega/2\hbar}X(t) + iP(t)/\sqrt{2\hbar m}$ and its Hermitian conjugate $a^\dagger(t)$. Show that these satisfy equations $da(t)/dt \propto a(t)$ and $d a(t)/dt \propto a(t)$, and determine the proportionality constant in each case.

(c) Solve the differential equations for $a(t)$ and $a^\dagger(t)$ in terms of $a(0)$ and $a^\dagger(0)$.

As a check, confirm that the Hamiltonian $H = \hbar \omega \left[a^\dagger(t)a(t) + \frac{1}{2}\right]$, is independent of time.

(d) Rewrite $X(t)$ and $P(t)$ in terms of $a(t)$ and $a^\dagger(t)$, and rewrite $a(0)$ and $a^\dagger(0)$ in terms of $X(0)$ and $P(0)$, so that $X(t)$ and $P(t)$ depend only on $X(0)$ and $P(0)$. You may find the identities below useful.

$X(t) = \sqrt{\hbar/2m\omega}\left[a(t) + a^\dagger(t)\right]$ and $P(t) = i\sqrt{\hbar m \omega/2}\left[a^\dagger(t) - a(t)\right]$.

As a check, you should find $X(T) = X(0)$, if $T$ is the classical period.

(e) Suppose the quantum state (which is independent of time) is chosen to be an eigenstate of $X(t)$ or $P(t)$. You may find the identities below useful.

$X(t) = \sqrt{\hbar/2m\omega}\left[a(t) + a^\dagger(t)\right]$ and $P(t) = i\sqrt{\hbar m \omega/2}\left[a^\dagger(t) - a(t)\right]$.

As a check, you should find $X(T) = X(0)$, if $T$ is the classical period.

(f) Suppose the quantum state (which is independent of time) is chosen to be an eigenstate of $X(t)$ or $P(t)$, and determine its eigenvalue.
4. An electron is in the + state when measured in the direction $S_\theta = S_z \cos \theta + S_x \sin \theta$, so that $S_\theta |+\rangle = +\frac{1}{2} \hbar |+\rangle$. However, the angle $\theta$ is uncertain. In each part, it is probably a good idea to check at each step that the trace comes out correctly.

(a) Suppose the angle is $\theta = \pm \frac{1}{2} \pi$, with equal probability for each angle. What is the state operator in the conventional $|\pm\rangle$ basis?

(b) Suppose the angle $\theta$ is randomly distributed in the range $-\frac{1}{2} \pi < \theta < \frac{1}{2} \pi$, with all angles equally likely. What is the state operator in the conventional $|\pm\rangle$ basis?

(c) Suppose the angle $\theta$ is randomly distributed in the range $-\pi < \theta < \pi$, with all angles equally likely. What is the state operator in the conventional $|\pm\rangle$ basis?

(d) In each of the cases listed above, what is the expectation value of $S_z$?

5. A general Hermitian operator in a two-dimensional system, such as the state vector for the spin of a spin-$1/2$ particle, takes the form $\rho = \frac{1}{2} (a \mathbf{1} + \mathbf{r} \cdot \sigma)$, where $\sigma$ are the Pauli matrices, $\mathbf{1}$ is the unit matrix, and $\mathbf{r}$ is an arbitrary three-dimensional vector.

(a) Find the eigenvalues of this matrix in general. It may be helpful to recall that $\mathbf{r} \cdot \sigma = \mathbf{r}^2$.

(b) What restrictions can be placed on $a$ and $\mathbf{r}$ if this represents a state operator?

(c) Show that all four components of $a$ and $\mathbf{r}$ are determined if we know every component of the expectation value of the spin $\langle \mathbf{S} \rangle$.

(c) Under what constraints will this density matrix be a pure state?

(d) A particle with this density matrix is under the influence of a Hamiltonian $H = \frac{1}{2} \hbar \omega \sigma_z$. Find formulas for $d\mathbf{r}/dt$ and $da/dt$, one of which must be trivial.

6. There is another version of the disproof of the “local hidden variables” hypothesis that does not require a discussion of probabilities. Consider a system consisting of three spins in the state $\psi = \frac{1}{\sqrt{2}} (|++ \rangle - |-- \rangle )$. Each of these spins will be measured on either the $x$-axis or the $y$-axis, that is, we will be measuring one of each of the following pairs: $\{(\sigma_{x1}, \sigma_{y1}),(\sigma_{x2}, \sigma_{y2}),(\sigma_{x3}, \sigma_{y3})\}$. The measurements will yield three eigenvalues, which will be one each from the pairs $(x_1, y_1)$, $(x_2, y_2)$, $(x_3, y_3)$. Each of these eigenvalues can take only the values $\pm 1$.

(a) Consider each of the operators $B_1 = \sigma_{x1} \sigma_{y2} \sigma_{y3}$, $B_2 = \sigma_{y1} \sigma_{x2} \sigma_{y3}$, $B_3 = \sigma_{y1} \sigma_{y2} \sigma_{x3}$.

Show that $\psi$ is an eigenstate of each of these operators, and calculate the eigenvalue.

(b) According to quantum mechanics, suppose you happened to measure one of the three combinations $B_1$, $B_2$, or $B_3$. What is the prediction for the product of the results of those measurements, $x_1 y_2 y_3$, $y_1 x_2 y_3$, or $y_1 y_2 x_3$?

© 2011, Eric D. Carlson
(c) According to the philosophy of hidden variables, the values for the products you found in part (b) must be true, even if you don’t make those exact measurements. Based on the three formulas you found in part (b), make a prediction for the product \( x_1x_2x_3 \) (hint: consider the product of the three possibilities in part (b)). *This has nothing to do with quantum mechanics!*

(d) According to the philosophy of hidden variables, the product \( x_1x_2x_3 \) must be what you would get if you performed all three measurements in the x-direction, and then multiplied them. Hence, the operator

\[
A = \sigma_{x1} \sigma_{x2} \sigma_{x3}
\]

will yield what value if you were to measure it, according to the philosophy of hidden variables?

(e) Show that \( |\psi\rangle \) is, in fact, an eigenstate of \( A \), but it has the wrong eigenvalue according to the prediction of part (d).

7. In lecture I showed how MWI can account for the results when you measure a series of electrons with spin states \( |+\rangle \). In this problem, we will instead measure the spin of electrons with spin state \( |\theta\rangle \equiv \cos \left( \frac{\theta}{2} \right) |+\rangle + \sin \left( \frac{\theta}{2} \right) |-\rangle \).

(a) According to the Copenhagen interpretation, if we measure the spin \( S_z \) of this particle, what results might we get, and what are the probabilities of each of these results? What is the average value we would get for this result?

(b) Now we start working on the Multiple Worlds calculation. Define the state of \( N \) particles as \( |\Psi_N\rangle = |\theta\rangle \otimes |\theta\rangle \otimes \cdots \otimes |\theta\rangle \). Define the average spin operator

\[
\overline{S}_z = \frac{1}{N} \sum_{i=1}^{N} S_{zi}
\]

Write out explicitly \( \overline{S}_z |\Psi_N\rangle \). You may find it useful to use the state \( |-\theta\rangle \equiv \cos \left( \frac{\theta}{2} \right) |+\rangle - \sin \left( \frac{\theta}{2} \right) |-\rangle \) in your answer.

(c) Work out the expectation values \( \langle \psi | \overline{S}_z | \psi \rangle \) and \( \langle \psi | \overline{S}_z^2 | \psi \rangle \). I recommend you first write out expressions for the quantities \( \langle \theta | \theta \rangle \), \( \langle \theta | -\theta \rangle \), \( \langle -\theta | \theta \rangle \) and \( \langle -\theta | -\theta \rangle \).

(d) Show that \( \left( \Delta \overline{S}_z \right)^2 = \hbar^2 \sin^2 \theta / 4N \) and show that therefore, in the limit \( N \to \infty \), the state vector becomes an eigenstate of \( \overline{S}_z \) with the eigenvalue found in part (a).
XII. Approximate Methods for Time Independent Systems

Up to now, we have treated nearly every problem we have encountered by using exact methods. In the real world, we almost never are capable of finding such exact results. Fortunately, in physics, unlike mathematics, it is common acceptable to find an approximate answer. Physicists are interesting in comparing their theories with experimentally measurable quantities, and for this reason, approximation methods are often as good, and sometimes better, than exact results. Different techniques are useful in different circumstances, and we will focus on a variety of techniques to give a flavor of the techniques that will prove most useful.

A. The Variational Approach

The variational approach is ideal for approximating the ground-state wave function and energy. It is widely used for many complex systems, since it can be used for almost any problem. It is based on the idea that the wave function with the minimum energy is the ground state. Suppose we are given the Hamiltonian $H$. Let us imagine that we have already found all of the eigenstates of this Hamiltonian, which we label $\{ |\phi_n \rangle, n = 1, 2, \ldots \}$, with $n = 1$ corresponding to the ground state. These states will have energies

$$H |\phi_n \rangle = E_n |\phi_n \rangle,$$

where the energies satisfy

$$E_1 \leq E_2 \leq E_3 \leq \cdots.$$

Now, consider any wave function $|\psi \rangle$. Any such wave function can be written in terms of the eigenstates, so we write

$$|\psi \rangle = \sum_n c_n |\phi_n \rangle.$$  \hspace{1cm} (12.1)

This wave function may or may not be normalized, since

$$\langle \psi | \psi \rangle = \left( \sum_m c^*_m \langle \phi_m | \right) \left( \sum_n c_n |\phi_n \rangle \right) = \sum_m \sum_n c^*_m c_n \langle \phi_m |\phi_n \rangle = \sum_n |c_n|^2.$$

Now, consider the expectation value of the Hamiltonian for the state (12.1):

$$\langle \psi | H | \psi \rangle = \left( \sum_m c^*_m \langle \phi_m | \right) H \left( \sum_n c_n |\phi_n \rangle \right) = \sum_m c^*_m \sum_n c_n \langle \phi_m | H | \phi_n \rangle = \sum_m \sum_n c^*_m c_n E_n \langle \phi_m | \phi_n \rangle$$

$$= \sum_n |c_n|^2 E_n = \sum_n |c_n|^2 E_n + \sum_n |c_n|^2 (E_n - E_1),$$

$$\langle \psi | H | \psi \rangle = \langle \psi | \psi \rangle E_1 + \sum_n |c_n|^2 (E_n - E_1).$$ \hspace{1cm} (12.2)
In the final expression in (12.2), we note that $E_n - E_i \geq 0$, so we have

$$\frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle} \geq E_i.$$  \hspace{1cm} (12.3)

This provides us with a strict upper bound on the ground-state energy. It is clear from (12.2) that this inequality will become an equality if all of the $c_n$'s vanish save for $c_1$, i.e., if we manage to pick a wave function proportional to the ground state energy. It is also clear that if we have a bit of the other wave functions mixed in, the error in the energy will be quadratic in the quantity $c_n$ of the other state vectors mixed in, and therefore our approximation of the ground state energy will tend to be pretty good if $c_n$ is small for $n > 1$.

We can improve this considerably by considering not just a single wave function $|\psi\rangle$, but rather a large number of them, and then pick the one that has the lowest expectation value (12.3). This is done not by picking a number of dissimilar functions, but rather by picking functions that are described by one or more parameters. We define a series of states $|\psi(a)\rangle$, where $a$ represents one or more parameters, and for each of them calculate the expectation value of the ground-state energy,

$$E(a) = \frac{\langle \psi(a) | H | \psi(a) \rangle}{\langle \psi(a) | \psi(a) \rangle}. \hspace{1cm} (12.4)$$

The result will be a (usually complicated) function of the parameter(s) $a$. The ground state energy is guaranteed to be smaller than all of these. We wish to find the minimum value of this function, which can be found by demanding that the function be stationary with respect to small changes in the parameters $a$:

$$\frac{\partial}{\partial a_i} E(a) \bigg|_{a_{\text{min}}} = 0.$$  

We then substitute this minimum point $a_{\text{min}}$ back into our function to estimate the ground state energy

$$E_i \approx E(a_{\text{min}}). \hspace{1cm} (12.5)$$

This approximation is guaranteed to be higher than the actual ground state energy. If you pick your wave function well, you will generally find that the resulting approximation is very good. Furthermore, you can obtain an approximate ground-state wave function as well, which is simply

$$|\psi_i\rangle \approx |\psi_i'\rangle = \frac{1}{\sqrt{\langle \psi(a_{\text{min}}) | \psi(a_{\text{min}}) \rangle}} |\psi(a_{\text{min}})\rangle, \hspace{1cm} (12.6)$$

where the denominator is required to get the wave function properly normalized. However, this wave function tends to be less-well approximated by (12.6) than the energy is by (12.5), because the errors in the wave function tend to be linear, while the errors in the energy are quadratic.
Let us illustrate how we can use the variational method to estimate the ground state energy of Helium. Helium has charge \( Z = 2 \), with Hamiltonian given by

\[
H = \frac{1}{2m} \mathbf{P}_1^2 + \frac{1}{2m} \mathbf{P}_2^2 - \frac{2ke^2}{|\mathbf{R}_1|} - \frac{2ke^2}{|\mathbf{R}_2|} + \frac{ke^2}{|\mathbf{R}_1 - \mathbf{R}_2|}.
\]  

(12.7)

The two electrons must be placed in an anti-symmetric state. To a crude approximation, we would expect the two electrons to both be in the 1s state, but with an anti-symmetric spin state, so we might write the quantum state as

\[
|\psi\rangle \approx |\psi_r, 00\rangle = |\psi_r\rangle \otimes \frac{1}{\sqrt{2}}(|+\rangle - |\rangle - +\rangle),
\]

where \( |\psi_r\rangle \) represents the space part of the state. The spin part is completely anti-symmetric, so the spatial part will be symmetric. If we ignore the interactions between the two electrons, the wave function should then take the form

\[
\psi(r_1, r_2) = \exp\left(-\frac{Zr_1}{a_0}\right)\exp\left(-\frac{Zr_2}{a_0}\right),
\]

(12.8)

where \( a_0 \) is the Bohr radius given by \( a_0 = \frac{\hbar^2}{mke^2} \) and \( Z = 2 \) for helium. We have deliberately not normalized the wave function, since the variational method doesn't require us to do so. The wave function (12.8) won't be exactly right, because the electrons will repel each other. Each electron will spend a lot of time near the nucleus, effectively canceling out part of the charge, and hence effectively lowering the nuclear charge. This suggests that instead of \( Z = 2 \), the best approximation might be to use \( Z = 1 \), or perhaps even better, some value in between. Instead of guessing this value, let us make \( Z \) a variational parameter, and simply find which value of \( Z \) minimizes the resulting energy.

To find the energy, we must find \( \langle \psi | \psi \rangle \) and \( \langle \psi | H | \psi \rangle \), where \( H \) is given by (12.7). We will work out each piece separately, and then assemble it to find \( E(Z) \). We find

\[
\langle \psi | \psi \rangle = \int d^3r_1 \int d^3r_2 e^{-2Zr_1//a_0} e^{-2Zr_2/a_0} = \left[ \int d\Omega \int_0^\infty r^2 dr e^{-2Zr/a_0} \right]^2 = \left[ 4\pi \left(\frac{a_0}{2Z}\right)^3 \right]^2 = \frac{\pi^2 a_0^6}{Z^6},
\]

(12.9a)

\[
\langle \psi | \mathbf{P}_1^2 | \psi \rangle = \| \mathbf{P}_1 | \psi \| \|^2 = \int d^3r_1 \int d^3r_2 \left(-i\hbar \nabla_1 e^{-Zr_1/a_0} e^{-Zr_2/a_0} \right)^2 = \int d^3r_1 \int d^3r_2 \frac{i\hbar Z}{a_0} \left(\frac{r_1}{a_0} e^{-Zr_1/a_0} e^{-Zr_2/a_0} \right)^2 = \frac{\hbar^2 Z^2}{a_0^2} \langle \psi | \psi \rangle,
\]

\[
\langle \psi | \mathbf{P}_2^2 | \psi \rangle = \langle \psi | \mathbf{P}_2^2 | \psi \rangle = \frac{\pi^2 a_0^4 \hbar^2}{Z^4},
\]

(12.9b)

\[
\langle \psi | \frac{1}{|\mathbf{R}_1|} | \psi \rangle = \int d^3r_1 \int d^3r_2 e^{-2Zr_1/a_0} \int d^3r_2 e^{-2Zr_2/a_0} = \frac{\pi a_0^3}{Z^3} \int d\Omega \int d\mathbf{r} e^{-2Zr/|a_0|} = \frac{4\pi a_0^2}{Z^3} \left(\frac{a_0}{2Z}\right)^2,
\]

© 2011, Eric D. Carlson 217
The remaining term is the most difficult to calculate. We start by writing down the integral in a straightforward way.

\[ \langle \psi \mid \frac{1}{|\mathbf{R}_1 - \mathbf{R}_2|} \mid \psi \rangle = \int d^3 \mathbf{r}_1 e^{-2Z_1/a_0} \int d^3 \mathbf{r}_2 e^{-2Z_2/a_0} \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \]

(12.10)

The tricky part is the distance between the two points, which, by the law of cosines, is given by

\[ |\mathbf{r}_1 - \mathbf{r}_2|^2 = r_1^2 + r_2^2 - 2r_1 r_2 \cos \theta_{\text{rel}}. \]

where \( \theta_{\text{rel}} \) is the angle between the two directions. The key to simplifying (12.10) is to change the way you define one of the angular integrals. It is, of course, always arbitrary how you pick the z-axis, and there is no reason to have to pick the same z-axis for each of the vectors \( \mathbf{r}_1 \) and \( \mathbf{r}_2 \). We can, for example, define the direction \( \mathbf{r}_1 \) with respect to some fixed z-axis, and then define the angles for \( \mathbf{r}_2 \) as if \( \mathbf{r}_1 \) were the z-axis for it. In other words, instead of defining the spherical angles \((\theta_2, \phi_2)\) relative to some fixed axis, use the angles \((\theta_{\text{rel}}, \phi_{\text{rel}})\) relative to the direction \( \mathbf{r}_1 \), as sketched in Fig. 12-1. Then (12.10) becomes

\[ \langle \psi \mid \frac{1}{|\mathbf{R}_1 - \mathbf{R}_2|} \mid \psi \rangle = \int d\Omega_1 \int_0^\infty r_1^2 dr_1 e^{-2Z_1/a_0} \int d\Omega_2 \int_0^\infty r_2^2 dr_2 e^{-2Z_2/a_0} \left( r_1^2 + r_2^2 - 2r_1 r_2 \cos \theta_{\text{rel}} \right)^{-1/2}. \]

We can now perform both \( d\Omega \) integrals, and the \( d\Omega_{\text{rel}} = d\theta_{\text{rel}} d\cos \theta_{\text{rel}} \) integral can be split into a trivial and a non-trivial part to yield

\[ \langle \psi \mid \frac{1}{|\mathbf{R}_1 - \mathbf{R}_2|} \mid \psi \rangle = 8\pi^2 \int_0^\infty r_1^2 dr_1 e^{-2Z_1/a_0} \int_0^\infty r_2^2 dr_2 e^{-2Z_2/a_0} \left( r_1^2 + r_2^2 - 2r_1 r_2 \cos \theta_{\text{rel}} \right)^{-1/2} d \cos \theta_{\text{rel}} \]

\[ = 8\pi^2 \int_0^\infty r_1^2 dr_1 e^{-2Z_1/a_0} \int_0^\infty r_2^2 dr_2 e^{-2Z_2/a_0} \left( -\frac{1}{r_1 r_2} \right) \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos \theta_{\text{rel}}} \]

\[ \langle \psi \mid \frac{1}{|\mathbf{R}_1 - \mathbf{R}_2|} \mid \psi \rangle = 8\pi^2 \int_0^\infty r_1 dr_1 e^{-2Z_1/a_0} \int_0^\infty r_2 dr_2 e^{-2Z_2/a_0} \left[ (r_1 + r_2) - |r_1 - r_2| \right]. \]

The integral now gets confusing, because of the absolute value. The easiest way to complete it is to take advantage of the fact that it is symmetric under interchange of \( r_1 \) and \( r_2 \). So we can restrict to the region \( r_1 < r_2 \) and simply double the result:
We now find the energy \( E(Z) \) by using (12.4) with (12.8), substituting (12.11) and the three equations (12.9) as needed:

\[
E(Z) = \frac{\langle \psi(Z) | H | \psi(Z) \rangle}{\langle \psi(Z) | \psi(Z) \rangle} = \frac{\langle \psi \left( \left( \mathbf{P}_1^2 + \mathbf{P}_2^2 \right) \right) | \psi \rangle}{2m \langle \psi | \psi \rangle} + \frac{k_e e^2}{\langle \psi | \psi \rangle} \langle \psi \left( \left( \frac{-2}{|\mathbf{R}_1|} + \frac{-2}{|\mathbf{R}_2|} + \frac{1}{|\mathbf{R}_1 - \mathbf{R}_2|} \right) \right) | \psi \rangle
\]

\[
= \frac{Z^6}{\pi^2 a_0^6} \left[ \frac{2\alpha_0 e^2 h^2}{2mZ^4} + k_e e^2 \left[ -\frac{4\alpha_0^2 e^2}{Z^2} + \frac{5\pi^2 a_0^3}{8Z^2} \right] \right] = \frac{\hbar^2 h^2}{m a_0^2} - \frac{27k_e e^2 Z}{8a_0^3} = k_e e^2 \left( Z^2 - \frac{27}{8} \right).
\]

To find the minimum of this energy, we set the derivative equal to zero, and find

\[
0 = \frac{d}{dZ} E(Z) \bigg|_{Z_{\min}} = \frac{k_e e^2}{a_0} \left( 2Z_{\min} - \frac{27}{8} \right).
\]

We therefore find \( Z_{\min} = \frac{27}{16} = 1.6875 \), which means that effectively the nuclear charge has been slightly reduced by the screening from the other electron. The estimate of the ground state energy is then

\[
E_1 = E \left( Z_{\min} \right) = \frac{k_e e^2}{a_0} \left( \frac{27}{16} \right) \left( \frac{27}{16} \right) = \frac{27k_e e^2}{a_0} = -77.48 \text{ eV}.
\]

The correct experimental value is -79.01 eV, an error of about two percent. Not bad, for such a crude approximation! The approximated ground state wave function is just given by (12.8), normalized by multiplying it by \( Z/\pi a_0^3 \), where \( Z = 1.6875 \).

As we have demonstrated, the variational method is quite successful at finding the ground state, but can it be used for higher energy states? The answer is yes. Suppose we have a complicated Hamiltonian, and we have first found the approximate normalized ground state vector \( |\psi_1 \rangle \). We now create a parameterized set of approximate first excited states \( |\psi_2(\mathbf{b}) \rangle \). We then remove any vestige of our estimated ground state by defining

\[
|\psi_2(\mathbf{b}) \rangle = |\psi_2(\mathbf{b}) \rangle - \langle \psi_1 | \psi_2(\mathbf{b}) \rangle |\psi_1 \rangle.
\]  

(12.12)
As is easily demonstrated, this state is guaranteed to have no overlap with \(|\psi_1'\rangle\), so 
\[ \langle \psi_1' | \psi_2'(b) \rangle = 0. \]
Since we expect \(|\psi_1'\rangle\) to be close to the ground state, this should mean that \(|\psi_2'(b)\rangle\) has almost none of the ground state in it. We then find the function

\[
E_2(b) = \frac{\langle \psi_2'(b) | H | \psi_2'(b) \rangle}{\langle \psi_2'(b) | \psi_2'(b) \rangle}.
\]
If we minimize this function, we will have a good estimate of \(E_2\), the first excited state energy

\[
E_2 \approx E_2(\text{b_{min}}).
\] (12.13)

This procedure can be repeated indefinitely. However, it should be recognized that, in general, (12.13) is not necessarily a lower bound on the first excited state energy. We have been careful, in (12.12), to subtract our approximation of the ground state from our trial state \(|\psi_2(b)\rangle\), but there is no guarantee that we have removed all of the exact ground state. Since there might be a small amount of the true ground state in \(|\psi_2(b)\rangle\), the resulting function \(E_2(b)\) might occasionally dip below the actual first excited state energy. There are exceptions, where such a statement can be made, however. For example, suppose you were asked to find the ground and first excited state for a particle in a one-dimensional symmetric potential \(V(x) = V(-x)\). We know that the ground state of such a potential will be symmetric, and the first excited state anti-symmetric. If we pick for our trial wave function \(|\psi_2(b)\rangle\) only odd functions, \[ \langle -x | \psi_2(b) \rangle = -\langle x | \psi_2(b) \rangle, \]
then we know that it can have no mixture of the even ground state. Hence we are guaranteed to obtain a lower bound on the first excited state energy.

In general, the variational method works best for the ground state, and its accuracy degrades as you increase to each of the excited states. This is because the errors accumulate as you move up each level, eventually rendering your energy estimates poor and your wave functions poorer. A nice complement to this method is the WKB method, which works well on excited states, but rather more poorly on the ground state.
B. The WKB Method

The WKB method was actually developed before Schrödinger’s equation, and works well in the classical limit, when you have much more energy than the separation between quantum states. Although it can be used in a variety of situations, it is most easily understood and applied to one-dimensional problems.

We start by looking at Schrödinger’s equation,

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

Then if you define

$$k^2(x) = \frac{2m}{\hbar^2} \left[ E - V(x) \right],$$

then (12.14) can be rewritten

$$\frac{d^2}{dx^2} \psi(x) = -k^2(x)\psi(x).$$

If the potential $V(x)$ is slowly varying, then $k(x)$ will be nearly constant, and hence the solution to (12.16) will look something like $\psi(x) \sim e^{\pm ikx}$. This suggests writing $\psi(x)$ as the product of an amplitude and a phase, where we would expect the amplitude to be slowly varying. We write

$$\psi(x) = A(x)e^{\pm i\phi(x)},$$

where $A(x)$ and $\phi(x)$ are real functions. If $\psi(x)$ is a solution of (12.16), so is $\psi^*(x)$, so we expect this to work for both signs of the phase. Substituting this into (12.16), we see that

$$-k^2 A e^{\pm i\phi} = \frac{d^2}{dx^2} \left( A e^{\pm i\phi} \right) = \left( A'' \pm 2iA'\phi' - A\phi'^2 \pm iA\phi^* \right)e^{\pm i\phi},$$

where primes denote derivatives with respect to $x$, and the dependence of $A$ and $\phi$ on $x$ is implicit. Cancelling the common phase and equating the real and imaginary parts we see that

$$-Ak^2 = A'' - A\phi'^2,$$

$$0 = 2A'\phi' + A\phi^*.$$  \hspace{1cm} (12.18a)

If we define $W(x) = \phi'(x)$, then we can multiply (12.18b) by $A$ to yield

$$\frac{d}{dx} \left[ A^2(x)W(x) \right] = 0,$$

which implies $A(x) \propto 1/\sqrt{W(x)}$. Substituting this expression into (12.16) and (12.18a) yields
\[ \psi(x) = \frac{N}{\sqrt{W(x)}} \exp \left[ -i \int W(x') dx' \right], \quad (12.19a) \]

\[ W^2 = k^2 - \frac{3W'^2}{4W^2} - \frac{W''}{W}. \quad (12.19b) \]

Now, we anticipate that we are in the classical large energy limit, so we expect \( k \) to be large. To leading order we would then anticipate that the largest term on the right side of (12.19b) would be \( k^2 \), so that \( W(x) = k(x) \). This is the zeroth order WKB approximation.\(^1\) This yields the leading order WKB wave function, which is

\[ \psi(x) = \frac{N}{\sqrt{k(x)}} \exp \left[ \pm i \int k(x') dx' \right], \quad (12.20) \]

with \( k(x) \) given by (12.15). This approximation works well only when \( k \) is large, which corresponds to high energy.

We would like to use (12.20) to find the energy of bound states quickly and efficiently. Consider first an infinite not-square well, with a potential that is infinite except in the region \( a < x < b \), but has some non-trivial potential in between, as illustrated in Fig. 12-2. Since we are dealing with bound states, we want real functions, so instead of using the form (12.20), we can take sums and differences to produce sines and cosines. We rewrite (12.20), therefore, in the form

\[ \psi(x) \approx \frac{N}{\sqrt{k(x)}} \sin \left[ \int_a^x k(x') dx' + \gamma \right], \quad (12.21) \]

where we have changed our indefinite integral into a definite one by picking the arbitrary lower bound \( a \), and including an explicit constant of integration \( \gamma \). Despite appearances, equation (12.21) contains both the possibility of a sine or a cosine, because the constant of integration can be chosen to be \( \gamma = \frac{1}{2} \pi \). However, we know that the function must vanish at the boundaries, and to make this work at \( x = a \), we simply set \( \gamma = 0 \). However, it must also vanish at \( x = b \), which can only be achieved if the argument of the sine vanishes there, which implies

\[ \int_a^b k(x) dx = (n+1) \pi, \quad (12.22) \]

where \( n \) is a non-negative integer. This is then the (approximate) quantization condition for the infinite non-square well illustrated in Fig. 12-2.

---

\(^1\) To obtain higher orders, you repeatedly substitute the expression for \( W \) in (12.19b) into itself. We will not go beyond leading order.
Consider now a situation like that illustrated in Fig. 12-3, where there are finite gaps at the boundaries, and in particular assume that the difference $E-V(x)$ changes sign abruptly at the boundary, while keeping the same magnitude $\Delta$. We can use our previous solution (12.21) as long as we are within the “classically allowed” region $a < x < b$, but what do we do outside the classically allowed region? In these regions in analogy with (12.15), we define

$$\alpha^2(x) = \frac{2m}{\hbar^2} [V(x) - E].$$

Then we can follow an argument similar to before and find, for example,

$$\psi(x) = \frac{1}{\sqrt{\alpha(x)}} \begin{cases} N_< \exp \left[ \int_a^x \alpha(x') dx' \right] & \text{for } x < a, \\ N_> \exp \left[ -\int_x^b \alpha(x') dx' \right] & \text{for } x > b. \end{cases}$$

(12.23)

The sign inside the exponential is chosen to make sure the wave function vanishes as $x \to \pm \infty$.

Because the gap is symmetric, by design, we will have $k(a) = \alpha(a) = k$, where we will, for the moment, treat $k$ as a constant. Then the wave function in the neighborhood of the boundary $x = a$ will look like

$$\psi(x) = \frac{1}{\sqrt{k}} \begin{cases} N_\times \exp \left[k(x-a)\right] & x < a, \\ N_\times \sin \left[k(x-a) + \gamma\right] & x > a, \end{cases}$$

where $\gamma$ is still an unknown constant of integration. These two functions must be equal and have equal derivatives at the boundary, which implies

$$\frac{N_<}{\sqrt{k}} = \frac{N_>}{\sqrt{k}} \sin \gamma \text{ and } N_< \sqrt{k} = N_> \sqrt{k} \cos \gamma.$$ 

The ratio of these equations implies $\tan \gamma = 1$, which we satisfy by choosing

$$\gamma = \frac{1}{4} \pi.$$ 

(12.24)

A similar analysis can be performed at the boundary $x = b$. The result is

$$\tan \left[ \int_a^b k(x) dx + \gamma \right] = -1,$$

$$\int_a^b k(x) dx + \gamma = (n + \frac{3}{4}) \pi,$$ 

(12.25)
where \( n \) is a non-negative integer. Combining (12.24) and (12.25), we obtain the Bohr-Sommerfeld quantization condition,

\[
\int_{a}^{b} k(x) \, dx = \left(n + \frac{1}{2}\right) \pi .
\tag{12.26}
\]

The difference between (12.22) and (12.26) can easily be understood qualitatively. When the boundaries are infinite, as they are in Fig. 12-2, the wave function must vanish at the boundary, whereas with the finite symmetrical boundaries, the wave function should not vanish, but instead must simply be diminishing so that it can transition into an exponential decay after crossing into the forbidden zone. This “softening” of the boundary means we don’t need quite as big a phase change between the two ends.

Consider now a third case, where the potential smoothly crosses the value \( E \), so that the function moves steadily from classically forbidden to classically allowed and back again, as illustrated in Fig. 12-4. The problem is that in the neighborhood of these points, we really can’t use the WKB approach at all, because the difference between the potential and the energy is small, guaranteeing that approximate wave functions like (12.20) and (12.23) will fail (note that these functions diverge at the classical turning points \( x = a \) and \( x = b \)). To do this computation properly, we would have to have three types of regions; regions where \( E \gg V(x) \), regions where \( E \ll V(x) \), and where \( E \approx V(x) \). In the classical limit, the transition region will be very small, and we can approximate \( V(x) \) as linear in this regions. We can solve the problem exactly\(^1\) in these regions; the result is rather complicated combinations of Bessel functions. We then match the boundary conditions with the other two regions. Without going into detail, I will simply state that the wave function (12.21) with the phase shift (12.24) is still valid, at least away from the turning points, and most importantly, (12.26) is still valid. In summary, our wave function and quantization condition will take the form

\[
\int_{a}^{b} \sqrt{2m \left[E - V(x)\right]} \, dx = \left(n + \frac{1}{2}\right) \pi \hbar ,
\tag{12.27a}
\]

\[
\psi(x) = \frac{N}{\left(2m \left[E - V(x)\right]\right)^{1/4}} \sin \left\{ \frac{1}{\hbar} \int_{a}^{x} \sqrt{2m \left[E - V(x')\right]} \, dx' + \frac{\pi}{4} \right\} ,
\tag{12.27b}
\]

where we have used (12.15) to eliminate our function \( k(x) \), and have arbitrarily removed an irrelevant factor of \( \sqrt{\hbar} \) from the (unknown) normalization. The turning points are the places where \( V(a) = V(b) = E \). As we shall see, (12.27a) allows us to efficiently estimate the energy of all of the states.

\(^1\) See L.I. Schiff, “Quantum Mechanics,” (c) 1968 McGraw-Hill for most of the details.
Consider the probability density associated with the approximation (12.27b)

$$\left| \psi(x) \right|^2 = \frac{N^2}{\sqrt{2m[E-V(x)]}} \sin^2 \left\{ \frac{1}{\hbar} \int_x^a \sqrt{2m[E-V(x)]} dx' + \frac{\pi}{4} \right\}.$$  

In the classical limit $\hbar \to 0$, the oscillations become so rapid that the sine function changes too fast to be physically measurable, and we can effectively replace $\sin^2$ with its average value, $\frac{1}{2}$.

$$\left| \psi(x) \right|^2 \approx \frac{\frac{1}{2} N^2}{\sqrt{2m[E-V(x)]}}.$$  

(12.28)

The denominator is simply the classical momentum, which is proportional to the velocity, so the probability density is inversely proportional to the velocity. Classically, this makes a lot of sense. Imagine a particle moving classically in the potential sketched in Fig. 12-4, which moves from $a$ to $b$ and back again. If we observe the particle at a random time, then it is unlikely we will find it at a point where it is moving swiftly, because it will move through such positions very rapidly, and classically the probability of finding the particle at any given point will be inversely proportional to the velocity. Hence (12.28) agrees with the classically expected probability distribution. Hence we can see how quantum and classical predictions correspond well, at least when the energy $E$ is much larger than the potential $V(x)$.

Let’s apply our quantization rule (12.27a) to a particle of mass $m$ contained in the harmonic oscillator potential, $V(x) = \frac{1}{2} m \omega^2 x^2$. The first step is to find the turning points, defined as the points where $V(x) = E$. Solving this equation is straightforward:

$$a, b = \pm \frac{1}{\omega} \sqrt{\frac{2E}{m}}.$$  

We then substitute these limits in the quantization rule (12.27a), which tells us

$$\left( n + \frac{1}{2} \right) \pi \hbar = \int_{-\sqrt{2E/m}/\omega}^{\sqrt{2E/m}/\omega} dx \sqrt{2mE-m^2 \omega^2 x^2} = \cdots = \frac{\pi E}{\omega}.$$  

Solving for the energy, we have $E = \hbar \omega \left( n + \frac{1}{2} \right)$, which is in fact exactly right. This is, however, a coincidence that does not apply in general.

Although the WKB approach works best for one-dimensional problems, it can be applied in more complicated cases as well. One particularly simple and interesting case is a particle in a spherically symmetric potential. If we factor our wave function into radial and angular parts, as we did in chapter 7, the radial part must satisfy the differential equation (7.27), which we repeat here while multiplying by a factor of $r$.

$$E \left[ rR(r) \right] = \frac{-\hbar^2}{2m} \frac{d^2}{dr^2} \left[ rR(r) \right] + \left[ \frac{l^2 + l}{2 \mu r^2} \hbar^2 + V(r) \right] \left[ rR(r) \right].$$  

(12.29)
This is identical to the one-dimensional Schrödinger equation if we treat $rR(r)$ as the wave function and replace the potential $V(r)$ with $V_{\text{eff}}(r) = V(r) + \hbar^2 \left( l^2 + l \right)/2mr^2$. The quantization condition (12.27a) then becomes

$$\int_a^b dr \sqrt{2mE - 2mV(r) - \frac{l^2 + l}{r^2} \hbar^2} = (n + \frac{1}{2}) \pi \hbar.$$  

The turning points are generally the places where the radical vanishes, though some care must be taken, as we shall see.

Let’s use these formulas to get an approximation for the hydrogen atom, with potential $V(r) = -k_e e^2/r$, so we have

$$\int_a^b dr \sqrt{2mE + 2mk_e e^2 \frac{l^2 + l}{r^2} \hbar^2} = \begin{cases} 
(n + \frac{1}{2}) \pi \hbar & l \neq 0, \\
(n + \frac{1}{2}) \pi \hbar & l = 0.
\end{cases}  \tag{12.30}$$

A word is in order about the turning points, as well as the right hand side of this equation. The upper limit $b$ always comes by equating the expression in the radical to zero. The lower limit also comes from setting this expression to zero provided $l \neq 0$. For $l = 0$, there will only be one place where the expression vanishes. However, the combination $rR(r)$ appearing in (12.29) must vanish at $r = 0$, and therefore we take, in this case, the lower bound to be $a = 0$. However, because the combination $rR(r)$ vanishes, there is no phase shift $\gamma$, similar to the boundary conditions applied in Fig. 12-2, and therefore when you use (12.25), you end up with a factor of $n + \frac{1}{2}$, rather than the usual $n + \frac{1}{4}$.

With considerable work, the turning points can be found and the integrals in (12.30) completed. The turning points are given by

$$a, b = \frac{-k_e e^2 \pm \sqrt{k_e^2 e^4 + 2Eh^2 \left( l^2 + l \right)/m}}{2E}.$$  

and the integral (12.30) works out to

$$\int_a^b dr \sqrt{2mE + 2mk_e e^2 \frac{l^2 + l}{r^2} \hbar^2} = \pi \left( k_e e^2 \sqrt{-\frac{m}{2E} - \hbar \sqrt{l^2 + l}} \right).  \tag{12.31}$$

Equating (12.30) and (12.31), and solving for $E$, we find

$$E_{nl} = -\frac{k_e^2 e^4 m}{2\hbar^2 \left( n + \frac{1}{2} + \frac{1}{4} \delta_{l0} + \sqrt{l^2 + l} \right)^2},  \tag{12.32}$$

where $n$ is any non-negative integer. The correct formula is

$$E_{nl} = -\frac{k_e^2 e^4 m}{2\hbar^2 \left( n + l + 1 \right)^2}.  \tag{12.33}$$
The difference between (12.33) and our conventional answer (7.51) occurs because we have labeled the lowest energy state for a given \( l \) as \( n = 0 \), rather than the usual notation \( n = l + 1 \). Comparison of (12.32) and (12.33) reveals that the formula is not very good for the lowest states, but considerably better as we increase \( l \) or \( n \).

C. Perturbation Theory

Perturbation theory is one of the most powerful and useful techniques in quantum mechanics, and indeed all of physics. Perturbation theory is most useful when you can divide your problem into two pieces, the larger of which you can solve exactly. Assume you have a Hamiltonian of the form

\[
H = H_0 + W ,
\]

where \( W \) is given by

\[
W = \lambda \hat{W} ,
\]

where we will treat \( \lambda \) as a small parameter. The orthonormal eigenstates \( |\phi_p\rangle \) and eigenvalues \( \varepsilon_p \) of \( H_0 \) are presumed known, so that

\[
H_0 |\phi_p\rangle = \varepsilon_p |\phi_p\rangle \quad \text{and} \quad \langle \phi_p | \phi_q \rangle = \delta_{pq} .
\]

(12.34)

Since \( H \) and \( H_0 \) are close, it seems reasonable to assume that the eigenstates \( |\psi_p\rangle \) and eigenenergies \( E_p \) of \( H \) will be near to the eigenstates and eigenenergies of \( H_0 \). We anticipate that we can therefore expand \( E_p \) and \( |\psi_p\rangle \) in powers of \( \lambda \), so that

\[
E_p = \varepsilon_p + \lambda \varepsilon'_p + \lambda^2 \varepsilon''_p + \cdots ,
\]

(12.35a)

\[
|\psi_p\rangle = |\phi_p\rangle + \lambda |\phi'_p\rangle + \lambda^2 |\phi''_p\rangle + \cdots .
\]

(12.35b)

We want to make sure that \( |\psi_p\rangle \) is an eigenstate of \( H \), so we need

\[
E_p |\psi_p\rangle = H |\psi_p\rangle = \left( H_0 + \lambda \hat{W} \right) |\psi_p\rangle .
\]

(12.36)

This alone is not sufficient to define \( |\psi_p\rangle \), since we can multiply \( |\psi_p\rangle \) by any complex constant and it will still satisfy (12.36). The easiest way to proceed is to choose this complex constant such that \( \langle \phi_p | \psi_p \rangle = 1 \) to all orders in perturbation theory. Substituting in (12.35b), we see that this implies

\[
\langle \phi_p | \phi'_p \rangle = \langle \phi_p | \phi''_p \rangle = \langle \phi_p | \phi'''_p \rangle = \cdots = 0 .
\]

(12.37)

This definition does have the disadvantage that \( |\psi_p\rangle \) will not be normalized, though this problem only appears at second order in perturbation theory. It has no effect on our finding the eigenvalues of the full Hamiltonian.
We now proceed to find the energies $E_p$ to third order in $\lambda$ and eigenstates $|\psi_p\rangle$ to second order in $\lambda$. Substituting (12.35a) and (12.35b) into Schrödinger’s equation (12.36) and expanding to third order in $\lambda$, we have

$$
\left(\varepsilon_p + \lambda \varepsilon_p' + \lambda^2 \varepsilon_p'' + \lambda^3 \varepsilon_p''' + \cdots\right) \left(\phi_p + \lambda \phi_p' + \lambda^2 \phi_p'' + \lambda^3 \phi_p''' + \cdots\right) = \left(H_0 + \lambda \hat{W}\right) \left(\phi_p + \lambda \phi_p' + \lambda^2 \phi_p'' + \lambda^3 \phi_p''' + \cdots\right),
$$

where we have used (12.34) on the leading term. We now match powers of $\lambda$. Ignoring the trivial leading term, this tells us

$$
\varepsilon_p \phi_p + \lambda \left(\varepsilon_p' \phi_p + \varepsilon_p \phi_p'\right) + \lambda^2 \left(\varepsilon_p'' \phi_p + \varepsilon_p' \phi_p'' + \varepsilon_p \phi_p''\right) + \lambda^3 \left(\varepsilon_p''' \phi_p + \varepsilon_p'' \phi_p''' + \varepsilon_p' \phi_p''' + \varepsilon_p \phi_p'''\right) + O(\lambda^4) \quad (12.38a)
$$

$$
\varepsilon_p' \phi_p' + \varepsilon_p \phi_p'' + \varepsilon_p' \phi_p''' + \varepsilon_p \phi_p''' = \hat{W} \phi_p' + H_0 \phi_p' \quad (12.38b)
$$

$$
\varepsilon_p'' \phi_p'' + \varepsilon_p' \phi_p''' + \varepsilon_p \phi_p'' + \varepsilon_p' \phi_p''' + \varepsilon_p \phi_p''' = \hat{W} \phi_p'' + H_0 \phi_p'' \quad (12.38c)
$$

Of course, these formulas can be continued indefinitely as needed.

We now proceed order by order as follows. We first let the bra $\langle \phi_p \rangle$ act on both sides of these equations, which allows us to get the corrections to the energies. Then we let $\langle \phi_q \rangle$ act on the left side for $q \neq p$. We then construct the state vectors order by order.

When $\langle \phi_p \rangle$ acts on the left, only the lead term will survive because of (12.37), so these equations become

$$
\varepsilon_p = \langle \phi_p \ | \hat{W} \phi_p \rangle + \langle \phi_p \ | H_0 \phi_p \rangle = \langle \phi_p \ | \hat{W} \phi_p \rangle + \varepsilon_p \langle \phi_p \ | \phi_p' \rangle = \langle \phi_p \ | \hat{W} \phi_p \rangle \quad (12.39a)
$$

$$
\varepsilon_p' = \langle \phi_p' \ | \hat{W} \phi_p' \rangle + \langle \phi_p' \ | H_0 \phi_p' \rangle = \langle \phi_p' \ | \hat{W} \phi_p' \rangle + \varepsilon_p' \langle \phi_p' \ | \phi_p'' \rangle = \langle \phi_p' \ | \hat{W} \phi_p' \rangle \quad (12.39b)
$$

$$
\varepsilon_p'' = \langle \phi_p'' \ | \hat{W} \phi_p'' \rangle + \langle \phi_p'' \ | H_0 \phi_p'' \rangle = \langle \phi_p'' \ | \hat{W} \phi_p'' \rangle + \varepsilon_p'' \langle \phi_p'' \ | \phi_p''' \rangle = \langle \phi_p'' \ | \hat{W} \phi_p'' \rangle \quad (12.39c)
$$

The pattern is obvious. We now let $\langle \phi_q \rangle$ for $q \neq p$ act on the left of (12.38a) and (12.38b), but this time the first term on the left is the only one guaranteed to vanish. So we have

$$
\varepsilon_p \langle \phi_q \ | \phi_p' \rangle = \langle \phi_q \ | \hat{W} \phi_p' \rangle + \langle \phi_q \ | H_0 \phi_p' \rangle = \langle \phi_q \ | \hat{W} \phi_p' \rangle + \varepsilon_q \langle \phi_q \ | \phi_p' \rangle, \\
\varepsilon_p' \langle \phi_q \ | \phi_p'' \rangle + \varepsilon_p \langle \phi_q \ | \phi_p''' \rangle = \langle \phi_q \ | \hat{W} \phi_p'' \rangle + \langle \phi_q \ | H_0 \phi_p'' \rangle = \langle \phi_q \ | \hat{W} \phi_p'' \rangle + \varepsilon_q \langle \phi_q \ | \phi_p'' \rangle, \\
\varepsilon_p'' \langle \phi_q \ | \phi_p''' \rangle + \varepsilon_p' \langle \phi_q \ | \phi_p'' \rangle + \varepsilon_p \langle \phi_q \ | \phi_p''' \rangle = \langle \phi_q \ | \hat{W} \phi_p''' \rangle + \langle \phi_q \ | H_0 \phi_p''' \rangle = \langle \phi_q \ | \hat{W} \phi_p''' \rangle + \varepsilon_q \langle \phi_q \ | \phi_p''' \rangle.
$$

We rearrange these equations to find for $q \neq p$

$$
\langle \phi_q \ | \phi_p' \rangle = \frac{\langle \phi_q \ | \hat{W} \phi_p' \rangle}{\varepsilon_p - \varepsilon_q} \quad (12.40a)
$$
\[
\langle \phi_q | \phi_p^* \rangle = \sum_{q} \langle \phi_q | \phi_q | \phi_p^* \rangle = \sum_{q \neq p} \frac{\langle \phi_q | \hat{W} | \phi_p \rangle}{\epsilon_p - \epsilon_q}.
\]

We now use completeness with (12.40a) and (12.37) to get the first order correction to the state vector \( |\phi_p^* \rangle \):

\[
|\phi_p^* \rangle = \sum_q |\phi_q \rangle \langle \phi_q | \phi_p^* \rangle = \sum_{q \neq p} \frac{\langle \phi_q | \hat{W} | \phi_p \rangle}{\epsilon_p - \epsilon_q}.
\]

We already have the first order correction to the energy (12.39a). We now substitute (12.41) into (12.39b) to get the second order correction to the energy:

\[
\epsilon_p^* = \frac{\langle \phi_p | \hat{W} | \phi_p^* \rangle}{\epsilon_p - \epsilon_q} = \sum_{q \neq p} \frac{\langle \phi_q | \hat{W} | \phi_q \rangle \langle \phi_q | \hat{W} | \phi_p \rangle}{\epsilon_p - \epsilon_q} = \sum_{q \neq p} \frac{| \langle \phi_q | \hat{W} | \phi_p \rangle |^2}{\epsilon_p - \epsilon_q}.
\]

We also use (12.41), (12.39a), and (12.40a) in (12.40b) for \( q \neq p \) to obtain

\[
\langle \phi_q | \phi_p^* \rangle = \frac{\langle \phi_q | \hat{W} | \phi_p^* \rangle}{\epsilon_p - \epsilon_q} = \sum_{n \neq p} \frac{\langle \phi_q | \hat{W} | \phi_n \rangle \langle \phi_n | \hat{W} | \phi_p \rangle}{(\epsilon_p - \epsilon_q)(\epsilon_p - \epsilon_n)} - \frac{\langle \phi_q | \hat{W} | \phi_p \rangle \langle \phi_p | \hat{W} | \phi_p \rangle}{(\epsilon_p - \epsilon_q)^2}.
\]

Again, we use completeness and (12.37) to get the second order correction to the state vector:

\[
|\phi_p^* \rangle = \sum_q |\phi_q \rangle \langle \phi_q | \phi_p^* \rangle = \sum_q |\phi_q \rangle \left[ \sum_{n \neq p} \frac{\langle \phi_q | \hat{W} | \phi_n \rangle \langle \phi_n | \hat{W} | \phi_p \rangle}{(\epsilon_p - \epsilon_q)(\epsilon_p - \epsilon_n)} - \frac{\langle \phi_q | \hat{W} | \phi_p \rangle \langle \phi_p | \hat{W} | \phi_p \rangle}{(\epsilon_p - \epsilon_q)^2} \right].
\]

We can then substitute this in (12.39c) to get the third order correction to the energy:

\[
\epsilon_p^* = \sum_{q \neq p} \langle \phi_p | \hat{W} | \phi_p^* \rangle \left[ \sum_{n \neq p} \frac{\langle \phi_q | \hat{W} | \phi_n \rangle \langle \phi_n | \hat{W} | \phi_p \rangle}{(\epsilon_p - \epsilon_q)(\epsilon_p - \epsilon_n)} - \frac{\langle \phi_q | \hat{W} | \phi_p \rangle \langle \phi_p | \hat{W} | \phi_p \rangle}{(\epsilon_p - \epsilon_q)^2} \right].
\]

The procedure can be repeated to whatever order is desired. We can now substitute (12.41), (12.42), (12.43), and (12.44) into (12.35a) and (12.35b), replacing \( \hat{W} \) by \( W \) to make things simpler. We then have

\[
E_p = \epsilon_p + \sum_{q \neq p} \frac{| \langle \phi_q | W | \phi_p \rangle |^2}{\epsilon_p - \epsilon_q} + \sum_{q \neq p} \langle \phi_p | W | \phi_q \rangle \times \left[ \sum_{n \neq p} \frac{\langle \phi_q | W | \phi_n \rangle \langle \phi_n | W | \phi_p \rangle}{(\epsilon_p - \epsilon_q)(\epsilon_p - \epsilon_n)} - \frac{\langle \phi_q | W | \phi_p \rangle \langle \phi_p | W | \phi_p \rangle}{(\epsilon_p - \epsilon_q)^2} \right],
\]

(12.45a)
\[ |\psi_p\rangle = |\phi_p\rangle + \sum_{q \neq p} |\phi_q\rangle \left[ \frac{\langle \phi_q | W | \phi_p \rangle}{\epsilon_p - \epsilon_q} + \sum_{n \neq p} \frac{\langle \phi_q | W | \phi_n \rangle \langle \phi_n | W | \phi_p \rangle}{(\epsilon_p - \epsilon_n)(\epsilon_p - \epsilon_q)} - \frac{\langle \phi_p | W | \phi_q \rangle \langle \phi_q | W | \phi_p \rangle}{(\epsilon_p - \epsilon_q)^2} \right]. \] 

(12.45b)

We keep only the first two terms if we want to do first order perturbation theory. Note that (12.45b) is *not* normalized, though it is normalized to first order.

Before proceeding, let us consider the validity of our approximation. When doing perturbation theory, a good technique is to do the calculation to the highest order you can conveniently calculate, then see whether each term in the series is smaller than the previous one. Consider, for example, the energy. Let’s compare the second correction to the first. Let’s assume that the state \( |\phi_q\rangle \) nearest in energy to the state we are interested in \( |\phi_p\rangle \) is separated by an energy gap \( \Delta \) from it, so that

\[ |\epsilon_p - \epsilon_q| \geq \Delta \quad \text{for} \quad q \neq p. \]

Then the denominators in the second correction in (12.45a) will always be bigger than \( \Delta \), which gives an upper bound on this term of

\[ \sum_{q \neq p} \frac{\langle \phi_q | W | \phi_p \rangle}{|\epsilon_p - \epsilon_q|} \leq \sum_{q \neq p} \frac{\langle \phi_q | W | \phi_p \rangle^2}{\Delta} \leq \sum_{q} \frac{\langle \phi_p | W | \phi_q \rangle \langle \phi_q | W | \phi_p \rangle}{\Delta}, \]

\[ \sum_{q \neq p} \frac{\langle \phi_q | W | \phi_p \rangle^2}{|\epsilon_p - \epsilon_q|} \leq \frac{\langle \phi_p | W^2 | \phi_p \rangle}{\Delta}. \]

The expression on the left is the second order correction to the energy, the expression on the right resembles the first order correction, but with an extra factor of \( W \) in the numerator and a factor of \( \Delta \) in the denominator. In other words, perturbation theory works well when the size of the perturbation is small compared to the splitting \( \Delta \) between levels. Provided \( \Delta \) is non-zero, and the parameter \( \lambda \) is small enough, we can always be sure perturbation theory will work.

Let’s do a simple example to see how to use equations (12.45a) and (12.45b). Consider a Hamiltonian

\[ H = \frac{1}{2m} P^2 + \frac{1}{2} m \omega^2 X^2 + \lambda X^4, \]

where \( \lambda \) is assumed small. The first two terms we name \( H_0 \) and recognize it as the harmonic oscillator, with quantum states \( |n\rangle \) with energies \( \epsilon_n = \hbar \omega (n + \frac{1}{2}) \). Our perturbation is \( W = \lambda X^4 \). Let’s find the second order correction to the ground state energy and first order correction to the wave function. Taking advantage of raising and lowering operators, we find
The energy is then
\[
E_0 = \varepsilon_0 + \langle 0 \vert W \vert 0 \rangle + \sum_{q=0}^{\infty} \frac{|q \vert W \vert 0 \rangle|^2}{\varepsilon_0 - \varepsilon_q} = 0 \frac{3\lambda h^2}{4m^2\omega^2} + \frac{\langle 2 \vert W \vert 0 \rangle^2}{-2\hbar\omega} + \frac{\langle 4 \vert W \vert 0 \rangle^2}{-4\hbar\omega} \\
= \frac{\hbar\omega}{2} + \frac{3\lambda h^2}{4m^2\omega^2} + \left( \frac{\lambda h^2}{4m^2\omega^2} \right)^2 \left( \frac{72}{2\hbar\omega} - \frac{24}{4\hbar\omega} \right) = \frac{\hbar\omega}{2} + \frac{3\lambda h^2}{4m^2\omega^2} - \frac{21\lambda^2 h^4}{8m^4\omega^5}.
\]

The ground state wave function, to first order, is
\[
\vert \psi_0 \rangle = \vert 0 \rangle + \sum_{q=0}^{\infty} \frac{|q \vert W \vert 0 \rangle}{\varepsilon_0 - \varepsilon_q} = \vert 0 \rangle - \frac{\lambda h}{8m^2\omega^3} \left( 6\sqrt{2} \vert 2 \rangle + \sqrt{6} \vert 4 \rangle \right).
\]

D. Degenerate Perturbation Theory

When we look at equations (12.45a) and (12.45b), a problem quickly comes to mind. What do we do if two or more unperturbed states have the same unperturbed energy? If \( \varepsilon_p = \varepsilon_q \), then we have vanishing denominators, and (12.45a) and (12.45b) will yield nonsensical results. Hence it is clear perturbation theory, as described in the previous section, will fail. Of course, it is always possible we will get lucky, and the numerators \( \langle \phi_q \vert W \vert \phi_p \rangle \) will vanish. If we could somehow assure that this will vanish, then in fact all the previous calculations will work just fine, and we can proceed as usual. This is the key to degenerate perturbation theory.

Suppose we have a set of states \( \{ \vert \phi_1 \rangle, \vert \phi_2 \rangle, \ldots, \vert \phi_g \rangle \} \), all of which have the same unperturbed energy, so,
\[
H_0 \vert \phi_i \rangle = \varepsilon \vert \phi_i \rangle, \quad i = 1, 2, \ldots, g.
\]

Then the same will be true of any linear combinations of the eigenstates \( \vert \phi_i \rangle \). Our goal, therefore, is to find orthonormal states \( \vert \phi'_i \rangle \) which are linear combinations of the \( \vert \phi_i \rangle \)'s, but such that \( \langle \phi'_i \vert W \vert \phi'_i \rangle \) vanishes when \( i \neq j \). To do this, let’s define the matrix \( \tilde{W} \), a finite square matrix of size \( g \times g \), with matrix elements \( \tilde{W}_{ij} = \langle \phi'_i \vert W \vert \phi'_j \rangle \), so
We first find the orthonormal eigenvectors \( \mathbf{v} \) of \( \tilde{W} \), so that
\[
\tilde{W}\mathbf{v}_i = w_i \mathbf{v}_i, \quad \text{and} \quad \mathbf{v}_j^\dagger \mathbf{v}_j = \delta_{ij} \quad \text{for} \quad i, j = 1, \ldots, g.
\]

Then we can define a new set of eigenstates
\[
|\phi_i\rangle = \sum_{j=1}^g (\mathbf{v}_j)_j |\phi_j\rangle
\]
where \((\mathbf{v}_j)_j\) is the \(j\)'th component \(\mathbf{v}_j\). Then it isn’t hard to show that
\[
\langle \phi_j | W | \phi_i \rangle = w_i \delta_{ij} \quad \text{and} \quad \langle \phi_j | \phi_i \rangle = \delta_{ij}.
\]

In other words, \( W \) is diagonal in the new basis when restricted to the degenerate states with energy \( \varepsilon_i = \varepsilon \). We then proceed with perturbation theory in the usual way, and our difficulties have disappeared!

Note that this procedure is completed before we begin doing perturbation theory. Indeed, the orthonormal states \( |\phi_i\rangle \) are independent of how weak the perturbation \( W \) is; they are, in a sense, zeroth order in our small parameter \( \lambda \). This may seem surprising, but in the absence of any perturbation, the eigenstates \( |\phi_i\rangle \) are ambiguous. Only the perturbation breaks the degeneracy, and determines which are your actual eigenstates.

After switching to the new states \( |\phi_i\rangle \), equations (12.45a) and (12.45b) become
\[
E_p = \varepsilon_p + \langle \phi_p^\dagger | W | \phi_p^\dagger \rangle + \sum_{e_q \neq e_p} \frac{\langle \phi_q^\dagger | W | \phi_p^\dagger \rangle^2}{e_p - e_q} + \sum_{e_a \neq e_p} \langle \phi_p^\dagger | W | \phi_a^\dagger \rangle \times
\]
\[
\times \left[ \sum_{e_q \neq e_p} \frac{\langle \phi_q^\dagger | W | \phi_a^\dagger \rangle \langle \phi_a^\dagger | W | \phi_p^\dagger \rangle}{(e_p - e_q)(e_p - e_a)} \right] - \langle \phi_p^\dagger | W | \phi_p^\dagger \rangle \langle \phi_p^\dagger | W | \phi_p^\dagger \rangle,
\]
\[\text{(12.46a)}\]
\[
|\psi_p\rangle = |\phi_p^\dagger\rangle + \sum_{e_q \neq e_p} |\phi_q^\dagger\rangle \left[ \frac{\langle \phi_q^\dagger | W | \phi_p^\dagger \rangle}{e_p - e_q} + \sum_{e_a \neq e_p} \frac{\langle \phi_p^\dagger | W | \phi_a^\dagger \rangle \langle \phi_a^\dagger | W | \phi_q^\dagger \rangle}{(e_p - e_a)(e_p - e_q)} \right] - \langle \phi_p^\dagger | W | \phi_p^\dagger \rangle \langle \phi_p^\dagger | W | \phi_p^\dagger \rangle \langle \phi_p^\dagger | W | \phi_p^\dagger \rangle,
\]
\[\text{(12.46b)}\]

The sums have been reduced to only include terms with non-vanishing denominators.

Note that the first order correction to the energies, \( \langle \phi_p^\dagger | W | \phi_p^\dagger \rangle \), are simply the eigenvalues \( w_p \) of the matrix \( \tilde{W} \).
As an illustration of how this is implemented, suppose we have a particle of mass $m$ in a perturbed two-dimensional infinite square well, with potential

$$V(x, y) = \begin{cases} \lambda \delta(x - \frac{1}{2}a) \delta(y - \frac{1}{2}a) & \text{if } 0 < x, y < a, \\ \infty & \text{otherwise}, \end{cases}$$

where $\lambda$ is assumed to be small. Then the unperturbed wave functions look like

$$\psi_{pq}(x, y) = \frac{2}{a} \sin \left(\frac{\pi px}{a}\right) \sin \left(\frac{\pi qy}{a}\right),$$

with energies

$$\varepsilon_{pq} = \frac{\pi^2 h^2}{2ma^2} (p^2 + q^2).$$

The ground states $|pq\rangle = |11\rangle$ is non-degenerate, and we need not change basis for this state. However, the first excited states $|pq\rangle = |12\rangle$ and $|pq\rangle = |21\rangle$ are degenerate, so to leading order our eigenstates will be determined by the splitting. We first need to find the matrix $\tilde{W}$, which is given by

$$\tilde{W} = \begin{pmatrix} \langle 12 | W | 12 \rangle & \langle 21 | W | 12 \rangle \\ \langle 12 | W | 21 \rangle & \langle 21 | W | 21 \rangle \end{pmatrix} = \frac{4\lambda}{a^2} \begin{pmatrix} \sin^4 \left(\frac{1}{2} \pi\right) & \sin^2 \left(\frac{1}{3} \pi\right) \sin \left(\frac{2}{3} \pi\right) \sin \left(\frac{1}{3} \pi\right) \\ \sin^2 \left(\frac{1}{3} \pi\right) \sin \left(\frac{2}{3} \pi\right) \sin \left(\frac{1}{3} \pi\right) & \sin^2 \left(\frac{2}{3} \pi\right) \sin^2 \left(\frac{1}{3} \pi\right) \end{pmatrix} = \frac{\lambda}{4a^2} \begin{pmatrix} 9 & 3\sqrt{3} \\ 3\sqrt{3} & 3 \end{pmatrix}.$$

The eigenvectors of $\tilde{W}$ are

$$v_1 = \frac{1}{2} \begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix}, \text{ and } v_2 = \frac{1}{2} \begin{pmatrix} -1 \\ \sqrt{3} \end{pmatrix}.$$

with eigenvalues $w_a = 3\lambda/a^2$ and $w_b = 0$ respectively. The eigenstates to zeroth order and the energies to first order will be given by

$$|\psi_1\rangle = \frac{1}{2} \left(\sqrt{3} |12\rangle + |21\rangle\right), \quad E_1 = \frac{5\pi^2 h^2}{2ma^2} + \frac{3\lambda}{a^2},$$

$$|\psi_2\rangle = \frac{1}{2} \left(-|12\rangle + \sqrt{3} |21\rangle\right), \quad E_2 = \frac{5\pi^2 h^2}{2ma^2}.$$
Problems for Chapter 12

1. A particle of mass \( m \) is contained in the potential \( V(x) = \frac{1}{2} \lambda x^4 \).
   (a) What symmetry does this potential have? What is the likely eigenvalue of the ground state under this symmetry? What is the likely eigenvalue of the first excited state under this symmetry?
   (b) Consider the trial wave function \( \psi(x) = \exp\left(-\frac{1}{2} A x^2\right) \). Does this have the right symmetry properties for the ground state, as given in part (a)? Calculate the normalized expectation value of the Hamiltonian using this wave function.
   (c) Minimize the expectation value and find an estimate of the ground state energy. Can we conclude that the ground state energy is lower than this value?
   (d) Construct a one-parameter trial wave function with the appropriate symmetry properties for the first excited state. Calculate the normalized expectation value of the Hamiltonian using this wave function.
   (e) Minimize the expectation value and find an estimate of the ground state energy. Can we conclude that the first excited state energy is lower than this value?

2. Joe Ignorant is unaware that the hydrogen atom has a known, exact solution, so he attempts to use the variational principle on it. The trial wave function he decides to try is \( \psi(r) = e^{-A r^2} \), where \( A \) is a variational parameter. Using this trial wave function, estimate the ground state energy. Compare the result with the exact result (7.51) with \( n = 1 \).

3. A particle lies in one dimension with Hamiltonian \( H = \frac{p^2}{2m} + F|X| \). Using the WKB method, our goal is to find the eigenenergies of this Hamiltonian.
   (a) For energy \( E \), find the classical turning points \( a \) and \( b \).
   (b) Perform the integral required by the WKB method.
   (c) Solve the resulting equation for \( E_n \).

4. We never completed a discussion of how to normalize the WKB wave function, given by eq. (12.27b).
   (a) Treating the average value of \( \sin^2 \rightarrow \frac{1}{2} \), and including only the wave function in the classically allowed region \( a < x < b \), write an integral equation for \( N \).
   (b) We are now going to find a simple formula for \( N \) in terms of the classical period \( T \), the time it takes for the particle to get from point \( a \) to \( b \) and back again. As a first step, find an expression for the velocity \( v(x) = dx/dt \) as a function of position. This is purely a classical problem!
   (c) Use the equation in part (b) to get an integral expression for the time it takes to go from \( a \) to \( b \). Double it to get an expression for \( T \).
   (d) Relate your answers in parts (a) and (c) to get a NON-integral relationship between the normalization \( N \) and the classical period \( T \).
5. A particle of mass $m$ lies in one dimension with Hamiltonian near the origin of

$$H = \frac{P^2}{2m} + \frac{1}{2} m \omega^2 X^2 + \gamma X^3$$

where $\gamma$ is very small.

(a) Find the energy of all quantum states to second order in $\gamma$.
(b) Find the eigenstates of $H$ to first order in $\gamma$.
(c) Find the expectation value $\langle \psi_p | X | \psi_p \rangle$ for the eigenstates $| \psi_p \rangle$ to first order in $\gamma$.

6. A particle of mass $m$ lies in a one-dimensional infinite slightly tilted square well

$$V(x) = \begin{cases} bx & \text{if } 0 < x < a \\ \infty & \text{otherwise} \end{cases}$$

where $b$ is small.

(a) What are the normalized eigenstates and energies if $b = 0$?
(b) Find the eigenstates and energies to first order in $b$.
(c) Find the ground state energy to second order in $b$. If you can’t do the sums exactly, do them numerically.

7. A particle of mass $m$ lies in a two-dimensional harmonic oscillator plus a perturbation

$$H = \frac{P_x^2}{2m} + \frac{P_y^2}{2m} + \frac{1}{2} m \omega^2 (X^2 + Y^2) + \gamma X^3 P_y$$

where $\gamma$ is very small.

(a) What are the eigenstates and eigenenergies of this in the limit $\gamma = 0$?
(b) Find the ground state and energy to first order and second order in $\gamma$ respectively.
(c) Find the first excited states and energy to zeroth and first order in $\gamma$ respectively.

8. A hydrogen atom in some combination of the $n = 2$ states is placed in an electric field which adds a perturbation $W = \lambda (X^2 - Y^2)$ where $\lambda$ is small. Ignore any spin-orbit or hyperfine splitting of the hydrogen atom; i.e., treat all $n = 2$ states of hydrogen as perfectly degenerate before $W$ is included.

(a) Find all non-vanishing matrix elements $\langle 2l'm' | W | 2lm \rangle$ for this interaction.
(b) Find the perturbed eigenstates and eigenenergies of the $n = 2$ states to zeroth and first order in $\lambda$ respectively.
XIII. Applications of Approximation Methods

An electron in an isolated hydrogen-like atom, has approximate Hamiltonian

$$H_0 = \frac{1}{2m_e} \mathbf{p}^2 - \frac{k_e e^2 Z}{|\mathbf{R}|},$$

(13.1)

where $m_e$ is the mass of the electron. The eigenstates of (13.1), ignoring spin, are described by three quantum numbers, $|n,l,m\rangle$, and have energy

$$\varepsilon_n = -\frac{k_e^2 e^4 m_e Z^2}{2\hbar^2 n^2}.$$ 

This formula is not the true energy, because:

- The nucleus will recoil, or move in response to the electron;
- The nucleus was treated as infinitesimal, when in fact it is finite
- The electron is moving fast enough that there are relativistic effects
- The nucleus has spin as well, which may react with the spin of the electron

The first of these has already been treated carefully in the section on the hydrogen atom; it can be included by replacing the electron mass $m$ with the reduced mass

$$\mu^{-1} = m_e^{-1} + m_N^{-1},$$

where $m_N$ is the nuclear mass. We now consider each of the others in turn.

A. Finite Nuclear Size

The nucleus is not in fact an infinitesimal object, but instead an object of finite radius $R$. We will treat the nucleus as a uniformly charged sphere of total charge $Ze$. The total charge inside a sphere of radius $r$ will therefore be

$$Q(r) = \begin{cases} 
Ze & r > R, \\
Zer^3/R^3 & r < R.
\end{cases}$$

Gauss’s Law can then be used to find the electric field everywhere

$$E(r) = \begin{cases} 
\frac{kZe}{r^2} & r > R, \\
\frac{kZer^3}{R^3} & r < R.
\end{cases}$$

The potential energy is the negative of the integral of the force on the particle used to bring it in from infinity:

![Figure 13-1: The electric potential an electron feels (red curve) from a nucleus of radius $R$ with a uniform charge density (green curve).](image)
This potential is illustrated in Fig. 13-1. Our Hamiltonian, therefore, should be

$$H = \frac{1}{2m} \mathbf{p}^2 - V(|\mathbf{R}|)$$

where $V$ is given by (13.3). We expect that the eigenstates and eigenvalues we found before will be very close to the correct ones, because the nucleus is tiny compared with a typical atom. We therefore rewrite this as

$$H = H_0 + W_{\text{nuc}}(|\mathbf{R}|)$$

where $H_0$ is given by (13.1), with energy (13.2). $W_{\text{nuc}}$ represents the difference between the exact potential (13.3) and the naïve potential $-k_e Z e^2/r$ for a pointlike nucleus. This difference is non-zero only inside the nucleus, where it is given by

$$W_{\text{nuc}}(r) = k_e Z e^2 \quad \text{for} \quad r < R. \quad (13.4)$$

To leading order in perturbation theory, the shift in energy to the various states should be

$$\delta E_{nlm} = \langle nlm|W(r)|nlm\rangle = \int d^3r \left| \psi_{nlm}(r) \right|^2 W_{\text{nuc}}(r) \quad (13.5)$$

A word of caution is in order here. The hydrogen atom has degenerate eigenstates, and therefore, shouldn’t we be more careful? Isn’t this a case where degenerate perturbation theory is called for? In fact, because the perturbation depends only on $r$, it commutes with $L^2$ and $L_z$, and therefore it cannot connect matrix elements with different $l$’s or $m$’s. Therefore, the states are already diagonalized, and we need not change basis before proceeding.

Keeping in mind the small size of the nucleus, the perturbation (13.5) is non-zero only very near the origin, and in this region the wave function $\psi(r)$ should change very little. It therefore makes sense to approximate the shift (13.5) by approximating $\psi(r) = \psi(0)$, which vanishes unless $l = m = 0$, so we have

$$\delta E_{n00} = \left| \psi_{n00}(0) \right|^2 \int d^3r W_{\text{nuc}}(r) = \left| \psi_{n00}(0) \right|^2 \frac{4\pi k_e Z e^2}{3} \left[ \frac{1}{r} + \frac{r^2}{2R^3} - \frac{3}{2R} \right] r^2 dr ,$$

$$\delta E_{n00} = \frac{2\pi}{3} k_e Z e^2 R^3 \left| \psi_{n00}(0) \right|^2 .$$

For example, for the 1s-state in a hydrogen-like atom, we have

$$\delta E_{1s} = \frac{2\pi}{3} k_e Z e^2 Z R^2 \left| \psi_{100}(0) \right|^2 = \frac{2\pi k_e e^2 Z R^3}{5} \frac{Z^3}{\pi a_0^3} = \frac{2ke^2Z^4 R^2}{5a_0^5} .$$

For hydrogen, this is smaller than the binding energy of the 1s state by a factor of $4R^2/5a_0^3$, or about $10^{-10}$, which is smaller than most of the other perturbations we will consider. The shift in energy is difficult to see in any case, since usually differences in
energy are easier to measure than a shift in a specific level. In contrast, the shift in the 2s state is easier to measure, since it is (nearly) degenerate with the 2p state, even though the shift itself is eight times smaller. For heavier hydrogen-like atoms, the shift increases by a factor of \( Z^4 \), and is further enhanced by the larger nuclear radius \( R \), but it is still a small effect (also, the binding energy of the electron has increased by \( Z^2 \) as well), though noticeable. If the electron is replaced by the much heavier muon \((m_\mu/m_e \approx 207)\), the radius \( a_0 \) decreases by the ratio of masses, enhancing this effect enormously, and allowing one to effectively measure the radius \( R \) of the nucleus.

B. Spin-Orbit Coupling

In addition to finite nuclear effects, which tend to be quite small, there are significant relativistic effects. An electron in the ground state of a hydrogen-like atom has a typical velocity of about \( v/c \sim Z\alpha \), where \( \alpha \approx \frac{1}{137} \) is the fine structure constant. A proper treatment of relativistic effects requires the Dirac Equation, the subject of a later chapter, and we will not go into it here. The one that interests us at the moment is the spin-orbit coupling, the only one of these interactions that depends on spin.

Spin-orbit coupling can be seen as an effect of interactions between the magnetic dipole moment of the spinning electron as it moves in the presence of an electric field. As can be shown from relativity, a moving object in an electric field experiences, in its own frame, a magnetic field, which to leading order at small velocities \( v \) is given by

\[
B = -\frac{1}{c^2} v \times E = \frac{1}{mc^2} p \times \nabla U(r),
\]

where \( U(r) \) is the electrostatic potential. A magnetic dipole \( \mu \) in a magnetic field experiences an energy shift

\[
\delta E = -B \cdot \mu = -\frac{1}{mc^2} [p \times \nabla U(r)] \cdot \mu.
\]

If we then substitute the magnetic moment \( \mu = ggS/2m \) of the electron from section 9D, we would anticipate an energy shift of the form

\[
\delta E = -\frac{gq}{2mc^2} [p \times \nabla U(r)] \cdot S = -\frac{g}{2mc^2} [p \times \nabla V_e(r)] \cdot S.
\]

where \( V_e(r) = qU(r) \) is the Coulomb potential experienced by the electron. Since \( V_e(r) \) is spherically symmetric, the resulting classical energy shift in the electron would be

\[
\delta E = -\frac{dV_e(r)}{dr} \frac{g}{2mc^2} [p \times \hat{r}] \cdot S = \frac{g}{2m^2c^2} \frac{1}{r} \frac{dV_e(r)}{dr} \mathbf{L} \cdot \mathbf{S}, \tag{13.6}
\]

where \( \mathbf{L} = r \times p \) is the angular momentum. Now, (13.6) was derived working in the rest frame of the electron. Do we get the same answer if we work in the rest frame of the nucleus? In this frame, there is no magnetic field, and hence the energy “shift” due to the
magnetic moment of the electron is simply $\delta E = 0$. Which equation, (13.6) or $\delta E = 0$, is the correct one to use? The answer cannot be found without a relativistic theory of the electron, but it turns out that the right approach is to take the average of these two. We will do so, and now reinterpret $\delta E$ as a perturbation $W_{SO}$ on the Hamiltonian.

$$W_{SO} = \frac{g}{4m^2c^2} \frac{1}{r} \frac{dV_e(r)}{dr} \mathbf{L} \cdot \mathbf{S}, \quad (13.7)$$

where $g = 2.00$ for an electron.

What will be the effect of $W$ on the quantum states of hydrogen-like atoms? Before proceeding, we must expand our basis states $|n,l,m\rangle$ which normally describe hydrogen to include the spin state, so our states now look like $|n,l,m_s,m_l\rangle$. However, in hydrogen we have degeneracy between various states with the same principle quantum number $n$. According to degenerate perturbation theory, we must now calculate all the matrix elements of the form $\langle n,l',m_l',m_s'|W|n,l,m_s,m_l\rangle$. For arbitrary $n$, this looks like quite a task, since there are $2n^2$ total states, so we would end up having to then diagonalize $2n^2 \times 2n^2$ matrix.

Fortunately, things are not as bleak as they seem. Our perturbation $W$ commutes with $\mathbf{L}^2$, and therefore it cannot connect states with different $l$ values. This makes things easier, but still a bit unwieldy. Our interaction will connect states with different $m_l$ and $m_s$ values, which is still more work than we want to do.

Fortunately, we can do better. The problem is that we originally chose eigenstates defined by the four commuting operators $\{H_0, \mathbf{L}^2, \mathbf{L}_z, S_z\}$. When the only terms in the Hamiltonian were all in $H_0$, this made sense, since they all commute with $H_0$. But rotation of only the wave function or only the spin is no longer a symmetry of the total Hamiltonian. The new term (13.7) remains unchanged only if you rotate the wave function and the spin together. This suggests working with $J = \mathbf{L} + \mathbf{S}$, the total angular momentum, rather than with $\mathbf{L}$ and $\mathbf{S}$ separately.

Instead of using the eigenstates we have worked with up to now, let's use eigenstates defined by the operators $\{H_0, \mathbf{L}^2, \mathbf{J}_z\}$ with states $|n,l,j,m_j\rangle$. Because the perturbation (13.7) commutes with all of these operations except for $H_0$, the only degenerate states $W$ connects are states with themselves. We can also use a standard trick we learned long ago to deal with the spins and angular momentum appearing in (13.7), so we have (taking advantage of the fact that $\mathbf{S}^2 = \frac{1}{2}\hbar^2$ and $g = 2$):

$$\langle n,l,j,m_j|W_{SO}|n,l,j,m_j\rangle = \frac{1}{2m^2c^2} \langle n,l,j,m_j| \frac{1}{r} \frac{dV_e(r)}{dr} \mathbf{L} \cdot \mathbf{S} |n,l,j,m_j\rangle$$

$$= \frac{1}{4m^2c^2} \langle n,l,j,m_j| \frac{1}{r} \frac{dV_e(r)}{dr} \left(\mathbf{J}^2 - \mathbf{L}^2 - \mathbf{S}^2\right) |n,l,j,m_j\rangle,$$

$$\langle n,l,j,m_j|W_{SO}|n,l,j,m_j\rangle = \frac{\hbar^2}{4m^2c^2} \left(\mathbf{j}^2 + \mathbf{l}^2 - \mathbf{l} \cdot \mathbf{j} - \frac{1}{4}\right) \langle n,l,j,m_j| \frac{1}{r} \frac{dV_e(r)}{dr} |n,l,j,m_j\rangle.$$
There is no angular dependence in the matrix element, so we can reduce this to a simple radial integral. Note that the result is independent of $m_j$; this is inevitable since $J_\pm$ commutes with the Hamiltonian. Now, what is most interesting, is the difference between energy levels and since the only possible values for $j$ are $j = l \pm \frac{1}{2}$, we find

$$E_{n,l,j+\frac{1}{2}} - E_{n,l,j-\frac{1}{2}} = \frac{\hbar^2}{8m^2c^2} \left( 2l + 1 \right) \int_0^\infty r^2 dr \left[ \frac{1}{r} \frac{dV_\phi(r)}{dr} \right] R_{nl}^2(r).$$ (13.8)

Of course, this only makes sense for $l > 0$; for $l = 0$ there cannot be such a splitting, since only $j = \frac{1}{2}$ is allowed. It is now a straightforward matter to substitute the explicit form of the Coulomb potential $V_\phi(r) = -\frac{k_e e^2}{r}$ into (13.7) and calculate, for example, the splitting between the $2p_{3/2}$ and $2p_{1/2}$ states of hydrogen.

It is useful to get some idea of the magnitude of (13.8). For hydrogen, the factor of $1/r$ and the derivative each tend to bring in a factor of order $0\alpha^{-1}$, and the average value of the Coulomb potential is of order the binding energy of hydrogen. Ignoring numerical factors, therefore, this splitting is of the order

$$\Delta E_{SO} \sim \frac{\hbar^2}{m^2c^2\alpha^2} E = \frac{k_e^2 e^4}{\hbar^2 c^2} E = \alpha^2 E ,$$

where $\alpha = \frac{1}{137}$ is the fine structure constant. One might be tempted to naively calculate other splittings, such as between $2s_{1/2}$ and $2p_{1/2}$. This turns out to be a mistake. In addition to the spin-orbit coupling, there are other relativistic effects; for example, there is an additional term in the Hamiltonian, $-\mathbf{P}\cdot\mathbf{A}/8m^3c^2$, which comes from the difference between the relativistic and non-relativistic formulas for kinetic energy. The contribution of this term turns out to be exactly the same magnitude as the contribution from $W_{SO}$. It is, however, independent of spin, and hence formulas like (13.8) will still be valid. However, if we are truly working with a hydrogen-like atom, we should use the full Dirac equation, which can be exactly solved in this case, and therefore equation (13.8) is not really necessary anyway.

The real value of an equation like (13.8) is not for hydrogen-like atoms, but rather for alkali metals. For example, sodium has eleven electrons, which (in the ground state) have electron configuration 1s$^2$ 2s$^2$ 2p$^6$ 3s$^1$. Because of the screening of the inner electrons, the outermost electron effectively sees a nucleus with charge +1, except for the relatively rare fraction of the wave function which penetrates inside the inner “core” electrons. Hence sodium looks a great deal like hydrogen, except for a significant variation in the Coulomb potential (which is why we wrote the Coulomb potential in the general form $V_\phi(r)$). The variation in the potential causes the 3s (actually, 3s$_{1/2}$) electron to be about 2 eV lower in energy than the 3p states. Because of the spin-orbit coupling, the 3p states are split, such that the 3p$_{1/2}$ state has slightly less energy than the 3p$_{3/2}$. When an electron moves from the 3p to the 3s state, a single photon will be emitted, but the exact wavelength of that photon will depend on the spin state the electron started in. The spectrum of sodium has a very distinctive yellow appearance, which is made of these two lines at 588.9950 nm and 589.5924 nm. The spectral line from the shorter of these two wavelengths has an intensity twice as strong as the other, because
there are twice as many spin states for the $3p_{3/2}$ electrons as the $3p_{1/2}$ electrons, and these states become populated in excited atoms in equal proportion.

C. Hyperfine Splitting

As already discussed, since the laws of physics are apparently invariant under rotation, one would anticipate that the full Hamiltonian $H$ commutes with the total angular momentum $J$, and hence atoms which differ only by their $m_J$ value should be truly degenerate. This argument fails for the simple reason that the $J$ is not truly the total angular momentum. In addition to the orbital angular momentum of the electron $L$ about the nucleus and the electrons spin $S$, the nucleus itself will often have a spin $I$. Like the electron, this means that the nucleus (which, for hydrogen is normally just a proton) may have a magnetic moment $\mu_p$, which will take the form

$$\mu_p = \frac{g_p e}{2m_p} I,$$

where $g_p = 5.585$ for the proton. The heavy proton mass $m_p$ in the denominator assures that this effect will be small. To understand the effects of this magnetic moment on the energy levels, we will treat the nucleus as a uniformly magnetized sphere of radius $a$, having magnetization

$$M = \frac{\mu_p}{\frac{3}{2} \pi a^3}.$$

The resulting magnetic fields can be found with some difficulty; the details are beyond us but can be found in standard textbooks. The vector potential is given by

$$A(r) = \frac{\mu_0}{4\pi r_\varnothing} \left( \mu_p \times r \right),$$

where $\mu_0 = 4\pi \times 10^{-7} \text{ N}/\text{A}^2$ is the permittivity of free space and $r_\varnothing$ is the larger of $r$ and $a$. The resulting magnetic field is given by

$$B = \nabla \times A = \frac{\mu_0}{4\pi} \begin{cases} 2\mu_p / a^3 & r < a, \\ \frac{3\mathbf{r} \cdot \mu_p}{3} - \frac{\mu_p}{r^3} & r > a. \end{cases}$$

These magnetic fields are sketched in Fig. 13-2 above.

---

1 See, for example, Jackson, Classical Electrodynamics, Third Edition, pp. 198-200 (© 1999, Wiley)
Because the nucleus is so small, we would like to take the limit \( a \to 0 \), which suggests we should always use the lower of these two equations, but this is not correct. If we consider the magnetic field integrated over the volume of the nucleus itself, we find

\[
\int B \, d^3r = \frac{\mu_0}{4\pi} \frac{2\mu_p}{a^3} \frac{4\pi a^3}{3} = 2\mu_0 \mu_p,
\]

which tells us that we cannot ignore the contribution from inside the nucleus. We can take this limit, however, by simply including a pointlike addition to the magnetic field, so we have

\[
B = \frac{\mu_0}{4\pi r^3} \left[ 3\hat{r} \left( \hat{r} \cdot \mu_p \right) - \mu_p \right] + 2\mu_0 \mu_p \delta^3(r).
\]

(13.11)

In contrast, the vector potential gets no significant contribution from small \( r \), so we can always use the formula for \( r > a \) in (13.10).

We now need to add these contributions to the Hamiltonian. We first expand our Hamiltonian (9.20) (approximating \( g = 2 \) for the electron) as

\[
H = \frac{1}{2m} \left[ \mathbf{P} + e\mathbf{A} \right]^2 - \frac{k_e e^2}{|\mathbf{R}|} + \frac{e}{m} \mathbf{B} \cdot \mathbf{S} = \frac{\mathbf{P}^2}{2m} - \frac{k_e e^2}{|\mathbf{R}|} + \frac{e}{2m} (\mathbf{P} \cdot \mathbf{A} + \mathbf{A} \cdot \mathbf{P}) + \frac{e^2}{2m} \mathbf{A}^2 + \frac{e}{m} \mathbf{B} \cdot \mathbf{S}.
\]

We recognize the first two terms as \( H_0 \), the unperturbed Hamiltonian. We ignore the \( \mathbf{A}^2 \) term, since we assume the magnetic field of the nucleus is small. We substitute (13.11) and the \( r > a \) formula from (13.10) into this equation. The vector potential has no divergence, so \( \mathbf{P} \cdot \mathbf{A} = \mathbf{A} \cdot \mathbf{P} \), and we find

\[
W_{\text{HF}} = \frac{e\mu_0}{m 4\pi} \left\{ \frac{1}{R^3} \left( \mu_p \times \mathbf{R} \right) \cdot \mathbf{P} + \frac{1}{R^3} \left[ 3 \left( \hat{\mathbf{R}} \cdot \mathbf{S} \right) \left( \hat{\mathbf{R}} \cdot \mu_p \right) - \mu_p \cdot \mathbf{S} \right] + \frac{8\pi}{3} \delta^3(\mathbf{R}) \mu_p \cdot \mathbf{S} \right\}.
\]

If we now substitute in the specific form (13.9) for the magnetic moment of the proton, we have

\[
W_{\text{HF}} = \frac{\mu_0 g_p e^2}{8\pi m m_p} \left\{ \frac{1}{R^4} \mathbf{I} \cdot \mathbf{L} + \frac{1}{R^4} \left[ 3 \left( \hat{\mathbf{R}} \cdot \mathbf{S} \right) \left( \hat{\mathbf{R}} \cdot \mathbf{I} \right) - \mathbf{I} \cdot \mathbf{S} \right] + \frac{8\pi}{3} \delta^3(\mathbf{R}) \mathbf{I} \cdot \mathbf{S} \right\}.
\]

(13.12)

where we have simplified the first term using \( (\mathbf{I} \times \mathbf{R}) \cdot \mathbf{P} = (\mathbf{R} \times \mathbf{P}) \cdot \mathbf{I} = \mathbf{L} \cdot \mathbf{I} \).

If we look at the form of (13.12), it is clear that the last term will contribute only for \( s \)-waves, since the wave function vanishes at the origin for \( l \neq 0 \). In contrast, the first term vanishes for \( s \)-waves, because of the angular momentum operator. What is, perhaps, less obvious is that the middle term also vanishes for \( s \)-waves. We can rewrite this middle term as

\[
\frac{1}{R^4} \left[ 3 \left( \hat{\mathbf{R}} \cdot \mathbf{S} \right) \left( \hat{\mathbf{R}} \cdot \mathbf{I} \right) - \mathbf{I} \cdot \mathbf{S} \right] = \frac{1}{R^4} \sum_{j,k} \left( 3 \hat{R}_j \hat{R}_k - \delta_{jk} \right) S_j I_k,
\]

and when this is integrated for an \( s \)-wave, we get expressions like \( \int \left( 3 \hat{R}_j \hat{R}_k - \delta_{jk} \right) d\Omega \), which can be shown to vanish. Hence, when considering first-order perturbation theory, we need only include the last term for \( s \)-waves, and the other terms for all other modes.
A complete basis for the bound states of hydrogen, ignoring perturbations, would look like \( |n, l, m_s, m_i, m_f\rangle \), where the last quantum number describes the eigenstate of the proton’s spin under \( I_z \). As already described, the spin-orbit coupling causes a splitting of states with the same \( n \) and \( l \), which suggests that a better basis would be \( |n, l, j, m_j, m_f\rangle \).

Until we include the hyperfine splitting, the states with different values of \( m_j \) and \( m_f \) must be degenerate, but the hyperfine interaction (13.12) will cause splittings. Again, we have to deal with degenerate perturbation theory, and again, the logical way to proceed is to change basis. The full Hamiltonian, including (13.12), does not commute with \( L, S, \) or \( I \), nor even the combination \( J = L + S \), since that is not truly the total angular momentum. The total angular momentum is

\[
F \equiv J + I = L + S + I,
\]

and \( F \) must commute with the full Hamiltonian, even if we have missed other perturbations. We therefore change basis to \( |n, l, j, f, m_f\rangle \), where the new quantum numbers are defined by their eigenvalues

\[
\begin{align*}
F^2 |n, l, j, f, m_f\rangle &= \hbar^2 \left( f^2 + f \right) |n, l, j, f, m_f\rangle, \\
F_z |n, l, j, f, m_f\rangle &= \hbar m_f |n, l, j, f, m_f\rangle.
\end{align*}
\]

The new interactions cannot connect states with different values of \( m_F \), and therefore they are much simpler in this new basis.

Let’s complete the computation for the \( s \)-wave states of hydrogen. These states have \( l = 0 \), and when combined with the spin of the electron, they must have \( j = \frac{1}{2} \). The proton is also spin \( i = \frac{1}{2} \), so the total spin \( f \) must take on values from 0 to 1, with \( m_f \) running from \(-f\) to \( f \). Keeping in mind that only the last term of (13.12) can contribute to the 1s state, we have

\[
\langle n, 0, \frac{1}{2}, f', m_f' | W_{\text{HF}} | n, 0, \frac{1}{2}, f, m_f \rangle = \frac{\mu_0 g_p e^2}{3 m m_p} \delta^3(R) I \cdot S | n, 0, \frac{1}{2}, f, m_f \rangle.
\]

The delta function simply pulls out the wave function at the origin, and recalling that since we are in a 1s state, \( J = S \), we can use our standard trick for addition of angular momentum to write

\[
\begin{align*}
 I \cdot S | \frac{1}{2}, f, m_f \rangle &= I \cdot J | \frac{1}{2}, f, m_f \rangle = \frac{1}{2} \left[ (J + I)^2 - J^2 - I^2 \right] | \frac{1}{2}, f, m_f \rangle \\
&= \frac{1}{2} \hbar^2 \left( f^2 + f - \frac{3}{2} \right) | \frac{1}{2}, f, m_f \rangle.
\end{align*}
\]

This guarantees that the perturbation will be diagonal in this basis, and we find the hyperfine splitting is

\[
E_{f=1} - E_{f=0} = \frac{\mu_0 g_p e^2 \hbar^2}{3 m m_p} | \psi_{n00} (0) |^2.
\]
The permeability of free space is related to Coulomb’s constant by the relation
\[ \mu_0 c^2 = 4\pi k_e, \]
and substituting in the explicit form of the wave function for the 1s state, we find
\[ \Delta E_{HF} = \frac{4\pi g_\rho k_e e^2}{3\pi c^2 m_p a_0^3} = \frac{8g_\rho m}{3m_p} |E_{1s}| \alpha^2. \]
This splitting is down by about a factor of \( m/m_p \) compared to the spin-orbit coupling, or about a factor of \( 10^{-3} \). It is, however, the only thing that distinguishes between the 1s-states of hydrogen, and determines that the true ground state has total angular momentum zero. Emission lines caused by this low-energy transition are responsible for the famous 21 cm line that is used to map cool hydrogen clouds in astronomy.

D. The Zeeman Effect

Up to now we have considered only the effects of an atom by itself, but let’s now consider the effect of external influences on an atom. In the presence of an external magnetic field in the \( z \)-direction, assuming these fields are not exceptionally strong, there will be a perturbation in the Hamiltonian
\[ W_B = \frac{eB}{2m}(L_z + gS_z), \quad (13.14) \]
as can be seen from eq. (9.23). It should be noted that although (13.14) was derived for a single electron, the same derivation will hold in a many-electron atom, but now \( L_z \) and \( S_z \) will simply represent the sums of the individual angular momenta and spin respectively. If we ignore the spin-orbit coupling, then we can simultaneously diagonalize \( W_B \) and \( H_0 \), as we did in chapter 9. We now wish to consider the weak field limit, where we assume \( W_B \) is smaller than the spin-orbit coupling. We will, however, continue to ignore the much smaller hyperfine interaction.

Ignoring the spin of the nucleus, the spin-orbit coupling and other relativistic effects causes atoms to be split into energy eigenstates \( |n,l,s,j,m_j⟩ \), where the energy, for a general atom (not necessarily hydrogen), depends on the first four indices, \( E_{nlsj} \). Note the inclusion of the spin eigenvalue \( s \), appropriate since for a more complicated atom the total spin state of the electrons can be complicated, and not necessarily \( \frac{1}{2} \). The index \( n \) does not necessarily represent principal quantum number, since we may not be discussing hydrogen. The unperturbed Hamiltonian \( H_0 \) includes the spin-orbit coupling and any other internal interactions of the atom. It must commute with \( J_z = L_z + S_z \), and by inspection our perturbation (13.14) also commutes with \( J_z \) as well. Hence \( m_j \) will still be a good quantum number when we include the perturbation, and when we compute matrix elements of \( W_B \), we need only include states with the same \( m_j \), and hence degenerate perturbation theory is unnecessary. The shift in energy due to \( W_B \) is then simply...
These matrix elements can be computed by introducing a complete set of intermediate states $|n, l, s, m_j, m_j\rangle$, so that our energy becomes

$$\Delta E_B = \frac{eB}{2m} \sum_{m_i, m_j} \langle n, l, s, j, m_j | (L_z + gS_z) | n, l, s, m_j, m_j\rangle \langle n, l, s, m_i, m_i | n, l, s, j, m_j \rangle$$

$$= \frac{eBh}{2m} \sum_{m_i, m_j} (m_i + gm_j) \langle n, l, s, j, m_j | n, l, s, m_i, m_i \rangle \langle n, l, s, m_i, m_i | n, l, s, j, m_j \rangle.$$  

The remaining matrix elements are nothing more than Clebsch-Gordan coefficients

$$\Delta E_B (m_j) = \frac{eBh}{2m} \sum_{m_i} (m_i + gm_j) \langle ls; m_i, m_i | jm_j \rangle^2.$$  

Although this contains an apparent double sum, in fact the only non-vanishing terms occur when $m_i + m_s = m_j$. As a trivial example, if the electrons have no total spin, $s = 0$, then $m_s = 0, m_l = m_j$, and this equation reduces to

$$\Delta E_B (m_j)_{s=0} = \frac{eBh}{2m} m_j.$$  

This divides the previously degenerate state into $2j + 1$ no longer degenerate states, where those aligned with the magnetic field have more positive energy.

What do we do if the magnetic field is large enough that the Zeeman effect is comparable to the spin-orbit coupling? The answer is that both the external magnetic effects $W_B$ and the spin orbit coupling $W_{SO}$ must be treated simultaneously, and we must work with the perturbation $W = W_{SO} + W_B$. If we work in the basis $|n, l, s, j, m_j\rangle$, our perturbation $W$ will not connect states with different $m_j$, since $J_z$ commutes with $W$, but it will connect states with different $j$, since the magnetic part $W_B$ does not commute with $J^2$. Hence work must actually be done to diagonalize the $\hat{W}$ matrix. One exception is the so-called “stretched” states, where $j = l + s$ and $m_j = \pm j$, since the states in this case are eigenstates of all of $W$. But we will not consider these cases further.

E. The Van Der Waals Interaction

As an example of the application of second order perturbation theory, consider a pair of atoms separated by a displacement $\mathbf{a} = a\hat{z}$. For definiteness, we will use a pair of hydrogen atoms separated along the $z$-axis. If we let $\mathbf{R}_1$ and $-\mathbf{R}_2$ represent the separation of the first electron from the first nucleus and second electron from the second nucleus, then the Hamiltonian is

$$H = \frac{\mathbf{p}_1^2 + \mathbf{p}_2^2}{2m} - k_e \frac{e^2}{|\mathbf{R}_1|} - k_e \frac{e^2}{|\mathbf{R}_2|} + \left( \frac{k_e^2}{|\mathbf{a}|} - \frac{k_e^2}{|\mathbf{a} + \mathbf{R}_1|} - \frac{k_e^2}{|\mathbf{a} + \mathbf{R}_2|} + \frac{k_e^2}{|\mathbf{a} + \mathbf{R}_1 + \mathbf{R}_2|} \right).$$
The various distances are illustrated in Fig. 13-3. We are treating the two nuclei as being at fixed points, while the electrons will be treated quantum mechanically. We will be interested in the large behavior, when the separation \( a \) is much bigger than the typical atomic size \( a_0 \).

Before proceeding, we should probably discuss a little the approximations that will go into our computation. We generally expect that one electron will stay near one nucleus, and the other near the other nucleus. The wave functions will fall exponentially with distance, such that there is effectively no overlap between the two electron’s wave functions. Anti-symmetrizing the wave function will have no effect on where you are likely to find an electron. One can show that in this case, even though the electrons are identical, we can describe them in terms of “the electron near nucleus one” and “the electron near nucleus two,” and this will make them effectively distinguishable particles. An exponentially small error will ensue, and hence any energy we compute will be slightly wrong. But this error will vanish as the separation grows.

The unperturbed Hamiltonian is just
\[
H_0 = \frac{P_1^2 + P_2^2}{2m} - \frac{k_e e^2}{|\mathbf{R}_1|} - \frac{k_e e^2}{|\mathbf{R}_2|},
\]
and the perturbation is
\[
W = \frac{k_e e^2}{|\mathbf{a}|} - \frac{k_e e^2}{|\mathbf{a} + \mathbf{R}_1|} - \frac{k_e e^2}{|\mathbf{a} + \mathbf{R}_2|} + \frac{k_e e^2}{|\mathbf{a} + \mathbf{R}_1 + \mathbf{R}_2|}.
\] (13.15)

The remaining terms will be treated as a perturbation; however, before we evaluate them, we would like to approximate them. We expect that \( \mathbf{R}_i \) will be small compared to \( \mathbf{a} \), so we can expand the distances to leading order in \( \mathbf{R}_i/|\mathbf{a}| \). Choosing \( \mathbf{a} = a\hat{z} \), and expanding to second order in \( \mathbf{R}_i \), you can see that
\[
\frac{1}{|\mathbf{a} + \mathbf{R}|} = \left[ (a + Z)^2 + X^2 + Y^2 \right]^{-\frac{1}{2}} = \frac{1}{a} \left[ 1 + \frac{2Z}{a} + \frac{\mathbf{R}^2}{a^2} \right]^{-\frac{1}{2}} = \frac{1}{a} \left[ 1 - \frac{Z}{a} - \frac{\mathbf{R}^2}{2a^2} + \frac{3Z^2}{2a^2} \right] + \cdots.
\]

Substituting this into (13.15) and simplifying, we find
\[
W = \frac{k_e e^2}{a} (1 - 1 - 1 + 1) + \frac{k_e e^2}{a^2} \left[ -Z_1 - Z_2 + (Z_1 + Z_2) \right]
\]
\[
+ \frac{k_e e^2}{2a^3} \left[ (\mathbf{R}_1^2 - 3Z_1^2) + (\mathbf{R}_2^2 - 3Z_2^2) + 3(Z_1 + Z_2)^2 - (\mathbf{R}_1 + \mathbf{R}_2)^2 \right]
\]
\[
= \frac{1}{2} k_e e^2 a^{-3} (6Z_1 Z_2 - 2\mathbf{R}_1 \cdot \mathbf{R}_2) = k_e e^2 a^{-3} \left( 2Z_1 Z_2 - X_1 X_2 - Y_1 Y_2 \right).
\]
We now wish to perform second-order perturbation theory on the ground state energy. The unperturbed wave functions look like \( |nlm; n'l'm'\rangle \), and have unperturbed energies \( \epsilon_n + \epsilon_{n'} \). The first order perturbation to the ground state looks like

\[
\langle 100; 100 | W | 100; 100 \rangle = k_e e^2 a^3 \left[ 2 (\langle 100 | Z | 100 \rangle)^2 - (\langle 100 | X | 100 \rangle)^2 - (\langle 100 | Y | 100 \rangle)^2 \right].
\]

However, all of these matrix elements vanish, because matrix elements of the form \( \langle nlm | R | n'l'm' \rangle \) vanish unless \( |I - I'| = 1 \), as proven in chapter 7. Hence we must go to second order, and we have

\[
E_g = 2\epsilon_1 + \sum_{nlm, n'l'm'} \left( \frac{\langle nlm; n'l'm' | W | 100; 100 \rangle}{2\epsilon_1 - \epsilon_n - \epsilon_{n'}} \right)^2 \tag{13.16}
\]

Only \( I = I' = 1 \) contributes in this sum. This expression is unfortunately very difficult to calculate, but we can get useful lower and upper bounds on its value. Because we are working with the ground state, the denominator is always negative, and therefore any finite sum will give us an overestimate of the energy. Let us consider only the lowest possible energy, which occurs for \( n = n' = 2 \). Consider, for example, the \( Z_1 Z_2 \) contribution, which, since it commutes with \( L_z \), is non-vanishing only when \( m = m' = 0 \).

Indeed, for this term in the sum, it is the only contribution, and we find

\[
\langle 210; 210 | W | 100; 100 \rangle = 2k_e e^2 a^3 \langle 210; 210 | Z_1 Z_2 | 100; 100 \rangle = 2k_e e^2 a^3 \langle 210 | Z | 100 \rangle^2 = \cdots = \left( \frac{8}{9} \right)^5 a_0^2 2k_e e^2 a^3.
\]

This term alone turns out to contribute two-thirds of the \( n = n' = 2 \) contribution to (13.16), so that the contribution from \( n = n' = 2 \) yields the upper bound

\[
E_g \leq 2\epsilon_1 + \frac{2k_e e^2 a_0^4 (4 + 1 + 1)(8)^{10}}{a^6 (2\epsilon_1 - 2\epsilon_2)} = -\frac{k_e e^2}{a_0} - \frac{6k_e e^2 a_0^4}{a^6 (k_e e^2/a_0 - k_e e^2/4a_0)} \left( \frac{8}{9} \right)^{10} \tag{13.17}
\]

\[
= \frac{k_e e^2}{a_0} - \frac{8k_e e^2 a_0^4}{a^6} \left( \frac{8}{9} \right)^{10} = \frac{k_e e^2}{a_0} - 2.46k_e^2 a_0^5.
\]

We can also use (13.16) to derive a lower limit on the ground state energy. When deriving (13.17), we effectively used (13.16), but treated the energy denominators as if they were all infinite except for the lowest non-vanishing energy state \( n = n' = 2 \). Of course, they are all higher in energy, so for all the other states, we know that

\[
\epsilon_n + \epsilon_{n'} - 2\epsilon_1 \geq 2\epsilon_2 - 2\epsilon_1. \tag{13.18}
\]

If we decrease all the denominators to this limiting value, we will be increasing the size of each of these (negative) terms, and therefore obtain a lower bound on the energies. In fact, for the bound states, the gap between the first excited state that contributes and all the other contributing bound states is not much bigger, so replacing (13.18) with an equality will lead not only to a lower bound, but probably a pretty good estimate.

Substituting in (13.16), we obtain the upper bound
\[ E_g \geq 2e_1 + \sum_{nlm,n'l'm'} \frac{|\langle nlm; n'l'm'|W|100;100\rangle|^2}{2e_1 - 2e_2} \]
\[ = -\frac{ke^2}{a_0} + \sum_{nlm,n'l'm'} \frac{\langle 100;100|W|nlm; n'l'm'\rangle \langle nlm; n'l'm'|W|100;100\rangle}{k_e^2/a_0 - k_e^2/4a_0}, \]
\[ E_g \geq -\frac{ke^2}{a_0} - \frac{4a_0}{3k_e^2} \langle 100;100|W^2|100;100\rangle, \] (13.19)

where in the last step, we used completeness of the intermediate states to perform the sum. We now need to calculate
\[ \langle 100;100|W^2|100;100\rangle = \left(ke^2a^{-3}\right)^2 \langle 100;100|(2Z_1Z_2 - X_1X_2 - X_2Y_2)^2|100;100\rangle. \]

Because of the high symmetry of the 1s state, cross terms like \(X_1X_2Z_1Z_2\) will have zero expectation value. Hence all the cross terms will vanish, and the matrix elements we will need will be
\[ \langle 100;100|W^2|100;100\rangle = k_e^2 e^4 a^{-6} \langle 100;100|\left(4Z_1^2Z_2^2 + X_1^2X_2^2 + Y_1^2Y_2^2\right)|100;100\rangle \]
\[ = k_e^2 e^4 a^{-6} \left(4\langle 100|Z^2|100\rangle^2 + \langle 100|X^2|100\rangle^2 + \langle 100|Y^2|100\rangle^2\right). \] (13.20)

Each of the matrix elements can then be calculated, and once again, due to symmetry, we expect them to be all the same, so we have
\[ \langle 100|X^2|100\rangle = \langle 100|Y^2|100\rangle = \langle 100|Z^2|100\rangle = \frac{1}{3} \langle 100|R^2|100\rangle \]
\[ = \frac{1}{3} \int d\Omega |Y_1^0(\theta,\phi)|^2 \int_0^\infty r^2 dr r^2 R_{10}^2(r) = \frac{4}{3} a_0^{-3} \int_0^\infty r^4 e^{-2r/a_0} dr = a_0^2. \]

Substituting this into (13.20) and then (13.19) yields
\[ E_g \geq -\frac{ke^2}{a_0} - \frac{8k_e^2 a_0^5}{a^6}. \] (13.21)

Comparing this with (13.17), we see that
\[ E_g \approx -\frac{ke^2}{a_0} - \frac{\gamma k_e^2 a_0^5}{a^6}, \]

where \(2.64 < \gamma < 8\). We anticipate that the rather crude estimate (13.17) would not be as good as (13.21), and indeed, a far more sophisticated analysis shows that \(\gamma = 6.50\). This new term represents an attractive potential between the two atoms for \(a \gg a_0\). This new interaction is called the \textit{van der Waals} interaction, and experimentally exhibits the behavior of neutral atoms separated by large distances.

Of course, we have only demonstrated this \(a^6\) behavior for hydrogen, but similar arguments will apply in many other cases. In general, the matrix elements \(\langle \phi_k|W|\phi_g\rangle\) will vanish, since the ground state tends to have a definite value of \(l\), which means that we always end up going to second order. The second order contributions will always
look like $\frac{\left|\langle \phi' | W | \phi \rangle\right|^2}{\Delta \varepsilon}$, will be attractive, and will come out like $-\gamma k_e e^2 a_0^5 / a_0^6$, though the constant $\gamma$ will depend on the type of atom involved. If the first excited state is far above the ground state in energy, such as occurs for a noble gas like helium, the energy denominator will tend to be large and the van der Waals force small, whereas more reactive atoms will tend to have smaller energy denominators, and feel a stronger long-range force.

Problems for Chapter 13

1. In section A we worked out the effect of the perturbation of the finite nuclear size assuming the nucleus is a uniform sphere with radius $R$.
   (a) Redo the calculation assuming instead that the nucleus is a spherical shell of radius $R$ with total charge $e$.
   (b) Now, assume that the nucleus consists of two shells of radii $R_1$ and $R_2$, with charges $q_1$ and $q_2$ with $e = q_1 + q_2$. What is the shift due to the finite nuclear size in this case?
   (c) Generalize part (b) to the case of $N$ charges of magnitude $q_i$ at radii $R_i$. Now generalize to the case of continuous charge distribution $q(r)$ which is normalized so $\int q(r) dr = e$. I don't need a formal proof, but if you did part (b) correctly, the generalization should be obvious.
   (d) For a uniform sphere of charge $q$ and radius $R$, what is the charge density $\rho$? What is the charge between radius $r$ and $r + dr$? Use this to find the energy shift for a uniform sphere. Check it against the formula derived in class, to make sure you didn't make a mistake.

2. Complete the computation of the spin-orbit splitting for hydrogen for the 2$p$, 3$p$, and 3$d$ states of hydrogen. Write your answers as multiples of $\alpha^2 |E_n|$, where $\alpha$ is the fine structure constant and $|E_n|$ is the unperturbed binding energy of this state.

3. Prove, as asserted in section C, that $\int \left( 3 \hat{r}_j \hat{r}_k - \delta_{jk} \right) d\Omega = 0$. This is actually nine formulas, but only six of them are independent.

4. We would like to know the hyperfine splitting for ordinary hydrogen $^1$H and heavy hydrogen $^2$H.
   (a) For $^1$H, the electron orbits a spin- $\frac{1}{2}$ ($i = \frac{1}{2}$) proton with $g_p = 5.5857$. Find the frequency of radio waves emitted from hydrogen, and compare with the experimental value of 1420.4 MHz.
   (b) For $^2$H, the electron orbits a spin-1 ($i = 1$) deuteron (bound state of proton and neutron), with magnetic moment given by $\mathbf{\mu} = \left( g_d e / 2m_p \right) \mathbf{I}$, where $g_d = 0.85743$. Find a formula for the splitting of the energy in this case (you will have to redo some of the computations), and then find the hyperfine frequency for deuterium.
5. In section D, we worked out the Zeeman (weak magnetic field) effect only in the case where the total spin of the atom is 0. Work out the splitting for all values of \( j \) and \( m_j \) in the case of spin-1/2. Formulas for the relevant Clebsch-Gordan coefficients can be found in eq. (8.18).

6. In section E we worked out an estimate of the energy between a pair of neutral hydrogen atoms. Consider now the case of a naked proton and a single electron bound to a second proton, as illustrated below.

(a) Write the Hamiltonian describing the interaction between these three particles. The protons will both be treated as fixed.
(b) Divide the Hamiltonian into \( H_0 \) and a perturbation \( W \), where the perturbation terms are small in the limit of large \( a \). How do you describe the quantum states and their energies in the limit of infinite \( a \)?
(c) Let \( a = a \hat{z} \). Expand \( W \) to only linear order in the displacement \( \mathbf{R} \).
(d) Find the leading non-vanishing perturbative correction to the ground state energy \( E_g \). The term should have a sum over, say, \( n, l, \) and \( m \). What values of \( l \) and \( m \) contribute?
(e) Find an upper bound on the energy of the ground state at long range by including only the leading term in the sum.
(f) Find a lower bound on the energy of the ground state at long range by including all the terms in the sum, but underestimating the energy denominators.
XIV. Scattering

Let us now turn our attention to the related topic of scattering. In a scattering experiment, some sort of projectile is fired at a target, and the resulting outgoing particles are observed, as sketched in Fig. 14-1. The properties of the target, the projectiles, and their interaction are then related to the rates of incoming and scattered particles. The experimental results are normally given in terms of cross-section.

A. Cross-Section and Differential Cross-Section

In general, many particles will be fired at a target, and only a small fraction of them will actually scatter. Generally the particles are sufficiently spread out that only one at a time strikes the target. It makes sense that the rate at which collisions occur $\Gamma$ would be proportional to the number density $n_{in}$ of the particles. It also might be sensible to imagine that it will be proportional to the difference in speed $|v|$ of the projectiles compared to the target, since the faster they are moving compared to the target, the more chances they will have to impact the target. Then the rate of collision is

$$\Gamma = \sigma n_{in} |v|, \quad (14.1)$$

where $\sigma$ is the cross-section, which has units of area. Classically, it is easy to understand the concept of cross section: if the projectiles are presumed to be infinitesimal, then the cross-section is just the area of the target as viewed by the impinging particles. Quantum mechanically, however, the cross section will usually depend on energy. Formula (14.1) has been given in the greatest generality possible, but in most circumstances the relative velocity $|v|$ is simply the velocity of the projectiles, since the targets are stationary. However, (14.1) is the definition of cross-section, and therefore by definition correct, where $|v|$ is just the relative speed as measured in the observer’s coordinate system. For example, relativistic particles colliding head on (a common experimental situation) would have $|v| \approx 2c$.

For a stationary target (which is all we will be considering), it is more useful to think in terms of flux, which is simply the product of number density times velocity. $\Phi = n |v|$. The incident flux appears explicitly in equation (14.1). The rate $\Gamma$ is normally determined by setting up detectors far from the target (which is typically an atom or molecule, or sometimes even smaller!) and counting the number of particles striking our detectors. Mathematically, this means
Substituting this into (14.1), we see that the cross-section is given by

\[ \sigma = \frac{1}{\Phi_{in}} \lim_{R \to \infty} \left[ R^2 \Phi_{out} \hat{R} d\Omega \right]. \]  

(14.2)

It is often the case that we wish to not perform the angular integration appearing in (14.2), either because our detectors do not cover all possible angles, or because we record the outgoing angles, and do not wish to discard this data. We therefore define the differential cross-section as the same formula, but not integrated, so

\[ \frac{d\sigma}{d\Omega} = \frac{1}{\Phi_{in}} \lim_{R \to \infty} \left[ R^2 \Phi_{out} \hat{R} d\Omega \right], \]

(14.3a)

\[ \sigma = \int \frac{d\sigma}{d\Omega} d\Omega. \]

(14.3b)

If your measurement only covers certain angles, you can restrict the range of integration in (14.3b) appropriately. If your detector has less than perfect efficiency, you can add an efficiency correction to the integral (14.3b). Experimentally, you also commonly use many identical targets, instead of just one, and then you can simply multiply by the number of targets to get an effective cross-section; however, sometimes there can be coherence between scattering from separate targets, complicating the computation.

Some complications arise when you consider which frame of reference you will be working in. Although in particle physics, it is common to collide particles head on, it is overall at least as common to collide particles with a stationary target. It is, however, usually easier to do your calculations in the center of mass frame, a frame where the projectile and target have total momentum zero, as illustrated in Fig. 14-2. Now, suppose in the lab frame the velocity of the projectiles of mass \( m \) is \( \mathbf{v} \) in the direction \( \mathbf{v} = v \hat{z} \) and the velocity of the target \( M \) is zero, then the center of mass frame will have a velocity

\[ \mathbf{u}_{cm} = \frac{\gamma}{1 + \gamma} \mathbf{v} = \frac{\gamma v}{1 + \gamma} \hat{z}, \quad \text{where} \quad \gamma = \frac{m}{M}. \]

The velocity of the projectile in this frame will be

\[ \mathbf{v}_{cm} = \mathbf{v} - \mathbf{u}_{cm} = \frac{v}{1 + \gamma} \hat{z}. \]

Figure 14-2: Comparison of lab (top) and center of mass (bottom) frame pictures of the same scattering event. The two angles \( \theta \) and \( \theta_L \) will in general be different.
Now, imagine that the particle scatters into an angle \((\theta, \phi)\) in this frame. By conservation of energy, the final speed will be the same, so the final velocity will be
\[
v'_{cm} = \frac{v}{1 + \gamma} (\sin \theta \cos \phi \hat{x} + \sin \theta \sin \phi \hat{y} + \cos \theta \hat{z}).
\]
If we now go back to the lab frame, the scattered velocity measured in the lab will be
\[
v' = v'_{cm} + u_{cm} = \frac{v}{1 + \gamma} \left[ \sin \theta \cos \phi \hat{x} + \sin \theta \sin \phi \hat{y} + (\cos \theta + \gamma) \hat{z} \right].
\tag{14.4}
\]
It is easy to see from (14.4) that the azimuthal angle \(\phi\), which can be found from the ratio of the \(x\)- and \(y\)-components of the velocity, are the same in the lab and center of mass frame, so \(\phi_L = \phi\). However, the polar angle \(\theta\), which is found from the ratio of the \(xy\)-components and the \(z\)-component, will be given in the lab frame by
\[
\tan \theta_L = \frac{\sin \theta}{\gamma + \cos \theta}, \quad \text{or equivalently,} \quad \cos \theta_L = \frac{\cos \theta + \gamma}{\sqrt{1 + 2\gamma \cos \theta + \gamma^2}}.
\tag{14.5}
\]
The rate at which particles scatter into a given angle, as defined by (14.1), must however be the same for both, since time runs at the same rate in both frames,\(^1\) so that we must have \(\int (d\sigma/d\Omega)_L d\Omega_L = \int (d\sigma/d\Omega) d\Omega\). Indeed, these must match not only if we integrate over all angles, but indeed also if we integrate over restricted angles, provided we match the integration range according to equation (14.5). It follows that
\[
\left(\frac{d\sigma}{d\Omega}\right)_L d\Omega_L = \left(\frac{d\sigma}{d\Omega}\right) d\Omega.
\]
Since solid angle is given by \(d\Omega = d\phi d\cos \theta\), it follows that
\[
\left(\frac{d\sigma}{d\Omega}\right)_L = \left(\frac{d\sigma}{d\Omega}\right) \left(\frac{d\cos \theta_L}{d\cos \theta}\right)^{-1} = \left(\frac{d\sigma}{d\Omega}\right) \left(1 + 2\gamma \cos \theta + \gamma^2\right)^{-\frac{3}{2}}.
\tag{14.6}
\]
This allows us to relate the \(cm\) frame differential cross-section to the center of mass frame cross-section. However, the total cross-section is identical in the two cases, \(\sigma = \sigma_L\).

In quantum mechanics it is complicated to consider many particles at once, so to simplify, we will study a single particle scattering off of a potential. What do we use for the flux? Flux is particles per unit volume times velocity. For one particle, the particles per unit volume is just the probability density \(\rho = \Psi^* \Psi\), which integrates to one for a normalized wave function, and the classical analog of \(\rho \vec{v}\) is just the probability current \(j\). This suggests that we make the replacement \(\Phi_{in} = |j_{in}|\) and \(\Phi_{out} = |j_{out}|\) in (14.3a).

We now start to tackle the problem quantum mechanically. As argued in chapter 11 section G, one can generally treat an arbitrary initial state as if it is an eigenstate of

\(^1\) This argument gets more complicated when relativity is considered. However, it is still true that the total cross-section remains the same.
energy, but whose energy is unknown. Therefore we need only study the case where we have an energy eigenstate (and, in fact, most experiments try to measure that initial energy as well as possible). We therefore are trying to solve Schrödinger’s time-independent equation, which takes the form

\[ E\psi(r) = -\frac{\hbar^2}{2\mu} \nabla^2 \psi(r) + V(r)\psi(r), \]  

(14.7)

where \( V(r) \) is the potential from our target. We first define \( k \) and \( U(r) \) by

\[ k^2 = \frac{2\mu}{\hbar^2} E \quad \text{and} \quad U(r) = \frac{2\mu}{\hbar^2} V(r), \]  

(14.8)

then (14.7) becomes

\[ (k^2 + \nabla^2)\psi(r) = U(r)\psi(r). \]  

(14.9)

We will think of \( U \) as being non-zero only in the immediate vicinity of the target, so that at large radii, \( U(r) \) will be treated as zero. In fact, our results will be valid provided \( U(r) \) falls off fast enough; one can show that it is only necessary that it fall faster than \( r^{-1} \).

In general, there will be an infinite number of linearly independent solutions of (14.9), but we will only be interested in particular ones. Typically, the incoming wave will be coming from a specific direction (normally chosen to be \( -\hat{z} \)). These incoming waves will take the form \( e^{ikz} \), an exact solution of (14.9) in the absence of the potential. We also expect there to be outgoing waves. To understand these, let us write (14.9) in spherical coordinates, ignoring the potential term, and keep only to order \( r^{-1} \), so we have

\[ k^2\psi(r) + \frac{1}{r} \frac{\partial}{\partial r} \left[r^2 \frac{\partial \psi(r)}{\partial r}\right] = 0, \]

which has solutions

\[ \psi(r) \approx f_k(\theta, \phi) e^{\pm ikr}. \]

However, we are only interested in outgoing waves, representing the scattered waves, which have the plus sign. Adding to this the incoming wave, we anticipate at large \( r \) we will have

\[ \lim_{r \to \infty} \psi(r) = e^{ikr} + f_k(\theta, \phi) e^{ikr}, \]  

(14.10)

where the first term represents the incoming wave, and the second the scattered wave.

We then calculate the two fluxes using \( \Phi = |\mathbf{j}| \), working out the flux for the incoming wave in Cartesian coordinates and the outgoing in spherical coordinates. For the latter we keep only the leading order term, which is the radial part.
If we now substitute these into (14.3a), we find

\[ \frac{d\sigma}{d\Omega} = |f_k(\theta, \phi)|^2. \] (14.11)

Hence the problem of finding the cross section is reduced to that of finding the functions \( f_k(\theta, \phi) \), defined by the large \( r \) limit of the wave function, (14.10).

One detail that should be mentioned: we never correctly normalized our wave function, and indeed, plane waves cannot be normalized. Fortunately, any normalization factor that we multiply the wave function by will not contribute, because it will cancel between the numerator and the denominator, so this is not a concern.

B. The Born Approximation

We now proceed to attempt to find the solutions of Schrödinger’s equation (14.9), starting from a plane wave initial state. Let’s imagine writing our total wave function in the form

\[ \psi(r) = e^{ikz} + \phi(r). \] (14.12)

Then substituting this into the left side of (14.9), and recalling that the first term is an exact solution in the absence of a potential, we have

\[ (k^2 + \nabla^2)\phi(r) = U(r)\psi(r). \] (14.13)

We’ll now imagine that the right hand side of (14.13) is known and attempt to solve for the left side. We will use the method of Green’s functions. Imagine, at first, that the right side of (14.13) were a delta function centered at the origin. Then we would have to solve the equation

\[ (k^2 + \nabla^2)G(r) = \delta^3(r). \] (14.14)

To solve this, we first note that the problem is spherically symmetric, so we anticipate the solution will be spherically symmetrical: \( G(r) = G(r) \). Then, working in spherical coordinates, (14.14) becomes

\[ k^2G(r) + \frac{1}{r} \frac{\partial^2}{\partial r^2}\left[rG(r)\right] = \delta^3(r). \]

Away from the origin, the solution to this equation is \( G(r) \propto e^{ikr}/r \). But we are interested in an outgoing wave, not an incoming, so we pick the plus sign, and we assume
To determine the constant $\alpha$, substitute this formula into equation (14.14) and integrate over a small spherical volume of radius $R$. We have

$$1 = \int_v d^3r \left( \nabla^2 + k^2 \right) G(r) = \int_v d^3r \nabla \left[ \nabla G(r) \right] + k^2 \int_v d^3r G(r)$$

$$= \int_s (dA) \hat{r} \cdot \nabla G(r) + k^2 4\pi \alpha \int_0^R r^2 dr \ r^{-1} e^{ikr} = 4\pi \alpha \left[ ika e^{-ika} - e^{-ika} - ika e^{ika} + e^{ika} - 1 \right]$$

$$= -4\pi \alpha .$$

Substituting this into (14.15) and then (14.14), we have

$$\left( k^2 + \nabla^2 \right) \left( -\frac{e^{ikr}}{4\pi r} \right) = \delta^3(r).$$

We now have a solution to (14.13) if the right hand side is a delta-function centered at the origin. Obviously, this problem is translation invariant, so we can make the source centered anywhere if we replace $r$ with $r - r'$:

$$\left( k^2 + \nabla^2 \right) \left( -\frac{e^{ikr - r'}}{4\pi |r - r'|} \right) = \delta^3(r - r'),$$

(14.16)

where the Laplacian $\nabla^2$ is understood to be with respect to $r$, not $r'$. Now that we have the solution to (14.13) for a point source at an arbitrary location, we need only use superposition to find the solution for an arbitrary source. We multiply (14.16) by $U(r') \psi(r')$ on the right, and then integrate over $r'$ to obtain

$$\int d^3r' \left( k^2 + \nabla^2 \right) \left( -\frac{e^{ikr - r'}}{4\pi |r - r'|} \right) U(r') \psi(r') = \int d^3r' \delta^3(r - r') U(r') \psi(r') = U(r) \psi(r),$$

Comparison with (14.13) makes it clear that we have found an integral solution to (14.13); namely, $\phi(r)$ is simply the integral given above, and substituting this into (14.12) then tells us

$$\psi(r) = e^{ikz} - \int d^3r' \frac{U(r') e^{ikr - r'}}{4\pi |r - r'|} \psi(r').$$

(14.17)

Naively, it looks like we haven’t accomplished much, since eq. (14.17) tells us that we can find the wave function if we know the wave function. But this appearance is deceptive. Imagine substituting the whole expression on the right side of (14.17) into itself, so that we obtain

$$\psi(r) = e^{ikz} - \int d^3r' \frac{U(r') e^{ikr - r'}}{4\pi |r - r'|} e^{ikz} + \int d^3r' \frac{U(r') e^{ikr - r'}}{4\pi |r - r'|} \int d^3r'' \frac{U(r') e^{ikr - r'} e^{ikr'' - r'}}{4\pi |r' - r''|} \psi(r').$$

This process can be repeated as many times as necessary, for example
\[
\psi (\mathbf{r}) = e^{\mathbf{i}kz} - \int d^3 \mathbf{r}' \frac{U(\mathbf{r}') e^{\mathbf{i}k|\mathbf{r}'-\mathbf{r}|}}{4\pi |\mathbf{r}-\mathbf{r}'|} e^{\mathbf{i}kz} + \int d^3 \mathbf{r}' \frac{U(\mathbf{r}') e^{\mathbf{i}k|\mathbf{r}'-\mathbf{r}|}}{4\pi |\mathbf{r}-\mathbf{r}'|} \int d^3 \mathbf{r}'' \frac{U(\mathbf{r}'') e^{\mathbf{i}k|\mathbf{r}''-\mathbf{r}|}}{4\pi |\mathbf{r}'-\mathbf{r}''|} e^{\mathbf{i}kz} + \cdots
\]  

(14.18)

It is clear we are obtaining a perturbation series in the potential \(U(\mathbf{r})\). Any finite version of this infinite sum is called the Born approximation, and indeed, if you keep only the first integral, it is called the first Born approximation or, confusingly, sometimes just the Born approximation. This series has a simple interpretation: the leading term represents the non-scattered particles, the first integral represents particles that scatter once, the next term represents those that scatter twice, and so on, as illustrated in Fig. 14-3.

![Figure 14-3: Graphical illustration of eq. (14.18), the Born approximation. The projectile scatters repeatedly from the target. The first Born approximation is illustrated by the second figure, where all the scattering comes from a single scatter.](image)

Let’s focus on the first Born approximation. We are not really interested in the wave function everywhere, we want to know the cross-section, which requires only that we know the wave function at large \(r\). In the limit of large \(r\), we have

\[
|r - r'| = \sqrt{(r - r')^2} = \sqrt{r^2 - 2r \cdot r' + r'^2} \approx r \left(1 - \frac{r \cdot r'}{r^2}\right) = r - \hat{r} \cdot r'.
\]

Substituting this into (14.18), and keeping only the leading integral (first Born approximation), and ignoring any terms that are suppressed by extra powers of \(r^{-1}\), we have

\[
\lim_{r \to \infty} \psi(\mathbf{r}) = e^{\mathbf{i}kz} - \int d^3 \mathbf{r}' \frac{U(\mathbf{r}') e^{\mathbf{i}(k - \hat{k})r}}{4\pi r} e^{\mathbf{i}kz} = e^{\mathbf{i}kz} - \frac{\mathbf{e}^{\mathbf{i}k \hat{r}}}{r} \int d^3 \mathbf{r}' U(\mathbf{r}') e^{\mathbf{i}k_z - \mathbf{i}k \hat{r} \cdot \mathbf{r}'}.\]

Comparison with (14.10) tells us that

\[
f_s(\hat{r}) = -\frac{\int d^3 \mathbf{r}' U(\mathbf{r}') e^{\mathbf{i}k_z - \mathbf{i}k \hat{r} \cdot \mathbf{r}'}}{4\pi}.
\]

Using (14.8) and (14.11), then, we find the differential cross-section is

\[
\frac{d\sigma}{d\Omega} = \frac{\mu^2}{4\pi^2 \hbar^2} \left| \int d^3 \mathbf{r} V(\mathbf{r}) e^{-iK \cdot r} \right|^2,
\]

(14.19)

where \(K\) is the difference in wave number between the scattered and initial waves, \(K = \mathbf{k'} - \mathbf{k}\), where \(\mathbf{k} = k\hat{z}\) is the incoming wave number and \(\mathbf{k'}\) the final wave number, so

\[
K = k (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta - 1),
\]
where it is understood that \((\theta, \phi)\) describe the directions of the outgoing waves, and have nothing to do with the variable \(r\) integrated over in (14.19). Thus the problem of finding the differential cross-section has been reduced to simply that of finding the Fourier transform of the potential.

Let’s illustrate the use of (14.19) for, the Yukawa potential, given by

\[
V(\mathbf{r}) = \frac{\beta}{r} e^{-\lambda r}. \quad (14.20)
\]

The first step is to find the Fourier transform of this potential, which is given by

\[
\int d^3 \mathbf{r} V(\mathbf{r}) e^{-i \mathbf{K} \cdot \mathbf{r}} = \beta \int d^3 \mathbf{r} r^{-1} e^{-i \mathbf{K} \cdot \mathbf{r}} e^{-\lambda r}. \quad (14.20)
\]

In this expression, we can pick our \(z\)-axis for the integration to be any axis we want, so to simplify, pick it in the direction of \(\mathbf{K}\). Then the Fourier transform becomes

\[
\int d^3 \mathbf{r} V(\mathbf{r}) e^{-i \mathbf{K} \cdot \mathbf{r}} = \beta \int d^3 \mathbf{r} r^{-1} e^{-i \mathbf{K} \cdot \mathbf{r}} e^{-\lambda r} \frac{1}{-i \mathbf{K} \cdot \mathbf{r}} e^{-\lambda r} = 4\pi \beta \frac{1}{\lambda^2 + \mathbf{K}^2}.
\]

It is easy to see that \(\mathbf{K}^2 = (\mathbf{k} - \mathbf{k})^2 = 2k^2(1 - \cos \theta), \) so we have

\[
\int d^3 \mathbf{r} V(\mathbf{r}) e^{-i \mathbf{K} \cdot \mathbf{r}} = 4\pi \beta \frac{1}{\lambda^2 + 2k^2(1 - \cos \theta)} \int d^3 \mathbf{r} r^{-1} e^{-i \mathbf{K} \cdot \mathbf{r}} e^{-\lambda r} = \frac{4\pi \beta}{\lambda^2 + 2k^2(1 - \cos \theta)}. \quad (14.21)
\]

The total cross section is given by

\[
\sigma = \int d\Omega \frac{d\sigma}{d\Omega} = \frac{8\pi \mu^2}{\hbar^4} \int_{-1}^{1} (1 - \lambda^2 + 2k^2 - 2k^2 \cos \theta)^2 \frac{d\cos \theta}{(1 - \lambda^2 + 2k^2 \cos \theta)^2} = \frac{4\pi \mu^2}{\hbar^4 k^2} \left(1 - \frac{1}{\lambda^2 + 4k^2}\right),
\]

\[
\sigma = \frac{16\pi \mu^2}{\hbar^4 \lambda^2 (\lambda^2 + 4k^2)}. \quad (14.21)
\]

Of particular interest to us is the case of Coulomb scattering, for which

\[
V_c(\mathbf{r}) = k_c q q'/r, \quad \text{where } k_c \text{ is Coulomb’s constant.}
\]

This is a special case of (14.20), so we immediately find the differential cross-section from (14.21). We have

\[
\frac{d\sigma_c}{d\Omega} = \frac{\mu^2 k_c^2 q^2 q'^2}{\hbar^4 k^4 (1 - \cos \theta)^2}.
\]

If we use the formula for momentum, \(\hbar k = \mu \nu\), we can write this in the more convenient form

\[
\frac{d\sigma_c}{d\Omega} = \frac{k_c^2 q^2 q'^2}{\mu^2 \nu^4 (1 - \cos \theta)^2} = \frac{k_c^2 q^2 q'^2}{4\mu^2 \nu^4 \sin^4 \left(\frac{1}{2} \theta\right)}. \quad (14.22)
\]
Note that the velocity in (14.22) is the relative velocity, and \( \mu \) is the reduced mass. The differential cross-section is then given in the center of mass frame. If the mass of the target is not large compared to the projectile, you would then have to use (14.5) and (14.6) to transform it to the lab frame, a messy proposition.

Of some concern might be that we have assumed a potential that falls off as a sufficiently strong power of \( r \) such that our waves can be assumed to asymptotically reach a region where we can ignore the effect of the potential. This turns out not to be the case for the Coulomb potential. The \( r^{-1} \) potential causes the wave to continue oscillating at large \( r \) in such a way that our solutions are not really right, but instead have an additional distance-dependant phase shift. Fortunately, this shift in only the phase of the wave does not affect the probabilities. We have obscured this problem by treating the Coulomb potential as the limiting case of the Yukawa potential, but if you reexamine the steps such as (14.21), you will see that problems arise if \( \lambda = 0 \). Nonetheless, it is known that (14.22) is correct; indeed, it is known to be not only correct in the first Born approximation, but in fact always exact\(^1\).

The total Coulomb cross-section is in fact infinite, which can be seen by integrating (14.22). The problem is small angles, where the differential cross-section blows up. This just represents an effect of the long-range interaction. No matter how far off-axis the incoming projectile is, it will always be deflected a tiny bit by this long-range force. Experimentally, the measured cross-section will always be finite, because there is some minimum angle \( \theta_{\text{min}} \) that can be experimentally distinguished from zero. This effectively introduces a cutoff in our integration, yielding an experimentally finite cross-section.

C. Scattering from a Spherically Symmetric Potential

One of the most common situations we are interested in is scattering from a spherically symmetric potential. In such cases, we know that the solutions of Schrödinger’s equation take on the form of spherical harmonics times radial wave functions, so we suspect we can find solutions of (14.9) of the form

\[
\psi_{lm}(r) = \hat{R}_l(r) Y_l^m(\theta, \phi). \tag{14.23}
\]

If we substitute this into (14.9) and use the fact that the angular part of the spherical Laplacian in spherical coordinates (A.22d) can be written in terms of the angular momentum operator \( \mathbf{L}^2 \) (7.23), we find

\[
k^2 \hat{R}_l(r) Y_l^m(\theta, \phi) + \frac{1}{r} \frac{\partial}{\partial r} \left[ r R_l(r) Y_l^m(\theta, \phi) \right] - \frac{\hbar^2}{\mu^2} R_l(r) Y_l^m(\theta, \phi) = U(r) R_l(r) Y_l^m(\theta, \phi),
\]

\[
\frac{1}{r} \frac{d^2}{dr^2} \left[ r R_l(r) \right] = \left[ \frac{l^2 + 1}{r^2} + U(r) - k^2 \right] R_l(r). \tag{14.24}
\]

\(^1\) Well, exact if we ignore relativistic corrections, radiative corrections (which come from quantizing the electromagnetic field), internal structure of the projectile or target, magnetic interactions, and so on. No realistic calculation is ever really exact.
The differential equation (14.24) must be solved subject to appropriate boundary conditions.

Let us solve (14.24) in two extreme cases, when \( r \) is very small, and when it is very large. For small \( r \), the angular momentum term will tend to dominate, so we have

\[
\frac{1}{r} \frac{d^2}{dr^2} \left[ r R_i(r) \right] \approx \frac{l^2 + l}{r^2} R_i(r).
\]

If we attempt a solution of the form \( R_i(r) \propto r^n \), and substitute this in, we find

\[
\frac{1}{r} \frac{d^2}{dr^2} \left[ r^{n+1} \right] \approx \frac{l^2 + l}{r^2} r^n,
\]

which implies \( n^2 + n = l^2 + l \) with solutions \( n = l \) or \( n = -1 - l \). The latter is generally unacceptable, since it represents a wave that diverges badly at the origin, so we will generally demand that \( R_i(r) \propto r^l \) at small \( l \). Indeed, it can be shown that \( R_i \) tends to be quite small out to a radius of about \( kr \sim l \), and for \( kr \ll l \) the wave function will effectively vanish. This leads to the very useful fact that if the potential \( U(r) \) vanishes (or is very small) beyond some radius \( r_0 \), we can ignore the potential when \( l \) is substantially bigger than \( kr_0 \).

Consider now equation (14.24) in the region where the potential vanishes. Setting \( U = 0 \) and letting \( x = kr \), equation (14.24) becomes

\[
\frac{1}{x} \frac{d^2}{dx^2} \left( R_i \right) - \left( \frac{l^2 + l}{x^2} - 1 \right) R_i = 0.
\]

This equation has two linearly independent solutions, the spherical Bessel functions, called \( j_l(x) \) and \( n_l(x) \),\(^1\) which are given by

\[
j_l(x) \equiv (-x)^l \left( \frac{1}{x} \frac{d}{dx} \right)^l \frac{\sin x}{x} \quad \text{and} \quad n_l(x) \equiv (-x)^l \left( \frac{1}{x} \frac{d}{dx} \right)^l \frac{\cos x}{x}.
\]

The \( j_l \)'s are well behaved at the origin, while the \( n_l \)'s diverge, so if there truly were no potential, then we would only be interested in the \( j_l \)'s. The first few of these are given explicitly by

\[
\begin{align*}
j_0(x) &= \frac{\sin x}{x}, & n_0(x) &= -\frac{\cos x}{x}, \\
j_1(x) &= -\frac{\cos x}{x} + \frac{\sin x}{x^2}, & n_1(x) &= -\frac{\sin x}{x} - \frac{\cos x}{x^2}, \\
j_2(x) &= -\frac{\sin x}{x} - 3 + \frac{3\sin x}{x^2} + \frac{3\sin x}{x^3}, & n_2(x) &= \frac{\cos x}{x} - \frac{3\sin x}{x^2} - \frac{3\cos x}{x^3}.
\end{align*}
\]

Some of these are plotted in Fig. 14-4.

---

\(^1\)Some sources call \( n_l \) by the name \( y_l \). As far as I can tell, they mean exactly the same thing.
These functions are related to the Bessel functions described in chapter 6, section G by the relations

\[ j_l(x) = \sqrt{\frac{\pi}{2x}} J_{l+\frac{1}{2}}(x), \]
\[ n_l(x) = \sqrt{\frac{\pi}{2x}} N_{l+\frac{1}{2}}(x). \]

It is not difficult from eqs. (14.25) to determine the asymptotic behavior of these functions at large \( x \). We want the derivative to act on the trigonometric functions, not the powers of \( x \), to make them as large as possible. The asymptotic behavior is given by

\[ j_l(x) \xrightarrow{\; x \to \infty \;} \frac{\sin \left( x - \frac{1}{2} \pi l \right)}{x} \quad \text{and} \quad n_l(x) \xrightarrow{\; x \to \infty \;} -\frac{\cos \left( x - \frac{1}{2} \pi l \right)}{x}. \]

A general solution of (14.24) in the asymptotic region, then, will be an arbitrary linear combination of these two solutions.

\[ R_l(r) = \alpha j_l(kr) - \beta n_l(kr). \quad (14.26) \]

If we define new constants \( A_l^m \) and \( \delta_l \) by the relations

\[ \alpha = A_l^m \cos \delta_l \quad \text{and} \quad \beta = A_l^m \sin \delta_l, \]
\[ A_l^m = \sqrt{\alpha^2 + \beta^2} \quad \text{and} \quad \delta_l = \tan^{-1}(\beta/\alpha), \quad (14.27a)\]

then (14.26) can be rewritten as

\[ R_l(r) = A_l^m \left[ j_l(kr) \cos(\delta_l) - n_l(kr) \sin(\delta_l) \right]. \quad (14.28) \]

This is, of course, only the solution outside the region where the potential vanishes. Inside, we must explicitly solve equation (14.24), let’s say numerically, subject to the boundary condition that \( R_l(r) \sim r^l \) at small \( r \). When we reach sufficiently large \( r \) that we can ignore the potential, we then match our numerically found radial function into the form (14.28). This determines \( R_l(r) \) up to a normalization constant, which means that the phase shift \( \delta_l \) will be determined. Thus our potential only causes a phase shift. This phase shift vanishes in the absence of the potential.

Assuming we know the phase shifts \( \delta_l \), we can immediately write down the large \( r \) behavior of our wave functions:
\[
\psi (r) = R_i(r) Y^m_l (\theta, \phi) = A^m_l \left[ j_i ( kr ) \cos ( \delta_l ) - n_i ( kr ) \sin ( \delta_l ) \right] Y^m_l (\theta, \phi)
\]

\[
\text{as} \quad r \to \infty \quad \frac{A^m_l}{kr} \left[ \sin (kr - \frac{1}{2} \pi l) \cos (\delta_l) + \cos (kr - \frac{1}{2} \pi l) \sin (\delta_l) \right] Y^m_l (\theta, \phi),
\]

\[
\psi (r) \to A^m_l \frac{\sin (kr - \frac{1}{2} \pi l + \delta_l)}{kr} Y^m_l (\theta, \phi).
\]

(14.29)

We have successfully solved (14.24), but we seem to be no closer to finding the cross-section. To clarify the problem, write out (14.29) in terms of exponentials, so we have

\[
\psi (r) \to \frac{A^m_l}{2ikr} \left[ \exp (ikr - i \frac{1}{2} \pi l + i \delta_l) - \exp (-ikr + i \frac{1}{2} \pi l - i \delta_l) \right] Y^m_l (\theta, \phi).
\]

(14.30)

This clearly consists of two waves, an incoming and an outgoing wave, but the incoming wave is coming from all directions (or at least, most directions), not just the z-direction. The key to getting this to work out is to take linear combinations of (14.30), choosing constants \(A^m_l\) in such a way to turn the second incoming wave into one of the form \(e^{ikr} = \exp (ikr \cos \theta)\). In other words, we will change our wave function to

\[
\psi (r) \to \frac{1}{r \to \infty} \frac{1}{2ikr} \sum A^m_l \left[ i^l e^{ikr e^{i\delta_l}} - i^l e^{-ikr e^{-i\delta_l}} \right] Y^m_l (\theta, \phi).
\]

(14.31)

Our goal is to cleverly choose \(A^m_l\) in such a way that we recognize the sum as a term that goes like \(\exp (ikr \cos \theta)\), added to a term that looks like an outgoing wave. To assist us, we need to write \(\exp (ikr \cos \theta)\) as a sum of spherical harmonics. This can always be done, so we are trying to find constants \(c^m_l\) (functions of \(r\)), such that

\[
\sum c^m_l Y^m_l (\theta, \phi) = \exp (ikr \cos \theta).
\]

(14.32)

To do so, we will use the orthonormality of the \(Y\)'s, so that

\[
c^m_l = \int d \Omega \: Y^m_l (\theta, \phi) \: \exp (ikr \cos \theta).
\]

The \(\phi\) integral is easy to do, since the \(Y^m_l\)'s always go as \(e^{i m \phi}\). The result vanishes unless \(m = 0\), and for \(m = 0\) we have

\[
c^0_l = 2\pi \int_{-1}^{1} d (\cos \theta) \: Y^0_l (\theta, \phi)^* \: \exp (ikr \cos \theta).
\]

We are interested in the large \(r\) behavior, so we integrate this by parts to yield

\[
c^0_l = \frac{2\pi}{ikr} Y^0_l (\theta, \phi)^* \: \exp (ikr \cos \theta) \bigg|_{\theta=0}^{\theta=\pi} - \frac{2\pi}{ikr} \int d (\cos \theta) \exp (ikr \cos \theta) \frac{d}{d(\cos \theta)} Y^0_l (\theta, \phi)^* \]

\[
= \frac{2\pi}{ikr} \left[ e^{ikr} Y^0_l (0, \phi)^* - e^{-ikr} Y^0_l (\pi, \phi)^* \right] + O(r^{-2}).
\]
The upper limit can be found in equation (8.37), and (7.33) helps you get the lower limit, so

$$c_l^0 = \frac{1}{2ikr} \sqrt{4\pi(2l+1)} \left[ e^{ikr} - e^{-ikr}(-1)^l \right].$$

Substituting this back into (14.32), we conclude that, for large $r$,

$$\frac{1}{2ikr} \sum_l \sqrt{4\pi(2l+1)} \left[ e^{ikr} - e^{-ikr}(-1)^l \right] Y_l^0(\theta, \phi) = \exp(ikr \cos \theta) = e^{ikz}. \quad (14.33)$$

Now, our goal was to find coefficients $A_l^m$ such that (14.31) would have the correct asymptotic form at large $r$, so that it would have a term that looks like $e^{ikz}$, and the rest would look like an outgoing wave. We can choose $A_l^m$ to be anything we want, and so we choose it to be

$$A_l^m = \delta_{m0} \sqrt{4\pi(2l+1)} e^{ijl}.$$ 

Substituting this into (14.31) and then using (14.33), we have

$$\psi(r) \to \frac{1}{2ikr} \sum_l \sqrt{4\pi(2l+1)} e^{ijl} \left[ e^{ikr} e^{i\delta_l} - i^{2l} e^{-ikr} e^{-i\delta_l} \right] Y_l^0(\theta, \phi),$$

$$\psi(r) - e^{ikz} \to \frac{1}{2ikr} \sum_l \sqrt{4\pi(2l+1)} \left[ e^{ikr} e^{2i\delta_l} - (-1)^l e^{-ikr} e^{-2i\delta_l} + (-1)^l e^{-ikr} \right] Y_l^0(\theta, \phi)$$

$$= \frac{e^{ikr}}{kr} \sum_l e^{i\delta_l} \sqrt{4\pi(2l+1)} \frac{e^{i\delta_l} - e^{-i\delta_l}}{2i} Y_l^0(\theta, \phi),$$

$$\lim_{r \to \infty} \psi(r) = e^{ikz} + \frac{e^{ikr}}{kr} \sum_l e^{i\delta_l} \sqrt{4\pi(2l+1)} \sin(\delta_l) Y_l^0(\theta, \phi).$$

Comparison with (14.10) tells us that

$$f_k(\theta) = \frac{1}{k} \sum_l e^{i\delta_l} \sqrt{4\pi(2l+1)} \sin(\delta_l) Y_l^0(\theta).$$

I have dropped the irrelevant $\phi$ argument for this case. Our differential cross section is

$$\frac{d\sigma}{d\Omega} = \left| f_k(\theta) \right|^2 = \frac{4\pi}{k^2} \sum_l e^{i\delta_l} \sqrt{2l+1} \sin(\delta_l) Y_l^0(\theta) \right|^2. \quad (14.34)$$

The total cross-section can then be calculated simply by integrating this expression, which is helped by the fact that the $Y_i^m$'s are orthonormal when integrated over angles. We find

$$\sigma = \frac{4\pi}{k^2} \sum_l (2l+1) \sin^2(\delta_l). \quad (14.35)$$

Our procedure is now clear. We start by solving (numerically or analytically) eq. (14.24), subject to the boundary conditions, typically that $R_l(r)$ does not diverge at the...
inner boundary. We then proceed to large enough $r$ that we can neglect the potential, and extract the coefficients $\alpha$ and $\beta$ in (14.26), and then use (14.27b) to get the phase shifts $\delta_l$. As we increase $l$, these will tend to get smaller and smaller, so we only need a finite number of them. We then use (14.34) to get the differential cross-section, or (14.35) to get the total cross section.

Let’s do a sample calculation to see how this works. Consider a particle of mass $m$ colliding with a hard-sphere potential,

$$V(r) = \begin{cases} 0 & \text{if } r > a, \\ \infty & \text{if } r < a. \end{cases}$$

The perturbative approach of the previous section will fail, because the potential is infinite. The scaled potential $U(r)$ will be the same as $V(r)$. Outside the sphere, the potential vanishes, so we will be solving the equation (14.24) with no potential, which becomes Bessel’s equation again, so we have solution (14.26):

$$R_l(r) = \alpha j_l (kr) - \beta n_l (kr).$$

Our boundary condition, however, will be $R_l(a) = 0$, so that

$$\alpha j_l (ka) = \beta n_l (ka).$$

We now need the phase shift, which we get from (14.27b):

$$\delta_l = \tan^{-1} \left( \frac{\beta}{\alpha} \right) = \tan^{-1} \left( \frac{j_l (ka)}{n_l (ka)} \right).$$

To make the problem tractable, we will work in the limit $ak \ll 1$, which means that we can focus primarily on the $l = 0$ contribution to scattering. We see that

$$\delta_0 = \tan^{-1} \frac{j_0 (ka)}{n_0 (ka)} = \tan^{-1} \left( -\frac{\sin(ka)}{\cos(ka)} \right) = -ka.$$  

This gives us the differential cross section from (14.34); in this limit:

$$\frac{d\sigma}{d\Omega} = \frac{4\pi}{k^2} |e^{i\delta_0} \sqrt{2\cdot 0 + 1} \sin(\delta_0) Y_0^0(\theta)|^2 = \frac{1}{k^2} \sin^2(ka) \approx a^2. \quad (14.36)$$

We are not justified in keeping more than the leading order term in the expression $\sin(ka)$, since we have dropped $\delta_l$ and all higher modes. The total cross-section can be found from (14.35) or (14.36) to be $\sigma = 4\pi a^2$. Perhaps surprisingly, the cross-section is in this case the total surface area, which is sometimes explained in that the waves are diffracting around the sphere and hitting it from all sides. To me, this is a numerical coincidence, but perhaps not that surprising; after all, the answer must dimensionally go like $a^2$, there is inevitably a factor of $2\pi$ from the azimuthal angle, and this leaves only the additional factor of 2 as a coincidence.
Problems for Chapter 14

1. This problem has nothing to do with quantum mechanics. A rigid sphere of radius $xa$ is launched at a stationary sphere of radius $a$, with $x < 1$. The two spheres have the same density $\rho$. In the center of mass frame, the differential cross section is given by $d\sigma/d\Omega = \frac{1}{4}a^2(x+1)^2$. This differential cross-section is independent of both the velocity and the angle of deflection.

(a) What is the total cross section measured in the lab frame? (this is easy)

(b) As a function of $x$, for what center of mass frame angles $\theta$ will the lab frame scattering be by an angle greater than 90 degrees, so $\theta_{\text{lab}} > 90^\circ$?

(c) Find a formula for the cross-section, but restricted to angles exceeding 90 degrees in the lab. Sketch its behavior (by hand or with a computer) for the entire range $0 < x < 1$.

(d) For what value of $x$ is the restricted cross section you found in part (c) the largest? You can do this numerically, if necessary.

3. Find a formula for the center of mass angle $\theta$ in terms of the lab angle $\theta_{\text{lab}}$. For what set of masses $m$ and $M$ is there a maximum deflection angle $\theta_{\text{max}}$ by which an object can be deflected, as measured in the lab frame? Find a formula for $\theta_{\text{max}}$ when this applies.

4. A particle of mass $\mu$ impacts a stationary weak potential given by $V(r) = Ae^{-\beta r^2/2}$.

(a) Find the Fourier transform $\int d^3r V(r) e^{-iK \cdot r}$ for this potential for arbitrary $K$.

Note that this potential is not rotationally symmetric, so you do not have freedom to pick $K$ to be any axis you want. I found it easiest to work in Cartesian coordinates.

(b) A particle with wave number $k$ comes in along the z-axis. Find the differential cross section and the total cross section.

(c) The same particle comes in along the x-axis. Find the differential cross section and the total cross section. You might find it useful to rotate the problem to make it easier.
5. In class we found the scattering cross section for Coulomb scattering by a charge \( q' \) from a point charge \( q \) located at the origin, with potential \( V(\mathbf{r}) = k_r q q' / r \). We needed the Fourier transform, which turned out to be \( \int d^3r V(\mathbf{r}) e^{-i \mathbf{k} \cdot \mathbf{r}} = 4\pi k_r q q' / \mathbf{K}^2 \).

(a) Place the source \( q \) not at the origin, but at \( \mathbf{r} = a \). What is the potential now? What is the Fourier transform now? **Hint:** Don’t actually do the work, just shift your integration variable, and use previous work. Convince yourself (and me) that the differential cross-section is unchanged.

(b) Suppose we replaced \( q \) with a series of charges \( q_i \) located at several locations \( a_i \). What would be the Fourier transform now? What if, instead of a series of discrete charges \( q_i \), we had a charge distribution \( \rho(\mathbf{r}) \) spread around in space?

(c) Show that the differential cross-section for a charge \( q' \) scattering off of a charge distribution \( \rho(\mathbf{r}) \) is given by

\[
\frac{d\sigma}{d\Omega} = \frac{4\mu^2 k_r^2 q'^2}{h^4 (\mathbf{K}^2)^2} \left[ \int d^3r \rho(\mathbf{r}) e^{-i \mathbf{k} \cdot \mathbf{r}} \right]^2
\]

where \( \mathbf{K} = \mathbf{k}' - \mathbf{k} \), the change in the wave number.

6. An electron of mass \( m \) scatters from a neutral hydrogen atom in the ground state located at the origin.

(a) What is the charge distribution for a neutral hydrogen atom? Don’t forget the nucleus! What is the Fourier transform of the charge?

(b) Find the differential and total cross section in this case.

7. A particle of mass \( m \) scatters from a potential \( V(\mathbf{r}) = F \delta(\mathbf{r} - a) \), so that the potential exists only at the surface of a thin sphere of radius \( a \).

(a) What equation must the radial wave functions \( R_l(\mathbf{r}) \) satisfy? Solve this equation in the regions \( r < a \) and \( r > a \).

(b) Apply appropriate boundary at \( r = 0 \). Deduce appropriate boundary conditions at \( r = a \).

(c) Assume now that \( ka \ll 1 \), so that the scattering will be dominated by the \( l = 0 \) term. Find a formula for the phase shift \( \delta_0 \). Find the differential cross-section. Check that your formula agrees with the formula found in section C for the hard sphere in the case \( F \to \infty \).

(d) Redo the problem using the first Born approximation. Again assume \( ka \ll 1 \) (effectively, this means \( Ka = 0 \)). Check that the resulting differential cross-section in this case is identical with that found above in the limit \( F \to 0 \).
XV. Time-Dependent Methods

Up to now we have focused almost exclusively on Hamiltonians that have no time-dependence, so \( H(t) = H \). In this chapter we focus on using methods where the Hamiltonian might contain explicit time dependence, so we have

\[
i \hbar \frac{d}{dt} |\Psi(t)\rangle = H(t)|\Psi(t)\rangle.
\]  

(15.1)

In general, we would like to imagine a Hamiltonian which changes from an initial Hamiltonian \( H_I \) to a final Hamiltonian \( H_F \), over the course of a time period \( T \), as illustrated in Fig. 15-1. Usually the initial state will be some eigenstate of the initial Hamiltonian \( |\psi_I\rangle \) and we would like to know the probability that the wave function ends in some other state \( |\psi_F\rangle \), where

\[
H_I |\psi_I\rangle = E_I |\psi_I\rangle \quad \text{and} \quad H_F |\psi_F\rangle = E_F |\psi_F\rangle.
\]

The probability will simply be given by

\[
P(I \rightarrow F) = |\langle \psi_F | \Psi(t) \rangle|^2.
\]

Calculating this “transition probability”, \( P(I \rightarrow F) \) will be our primary goal.

A. The Sudden Approximations

Consider first the situation where the Hamiltonian changes very quickly from its initial state to its final state. At \( t = 0 \), it is in the initial state \( |\Psi(0)\rangle = |\psi_I\rangle \). For short times, we can Taylor expand the wave function around this state, so that

\[
|\Psi(t)\rangle = |\Psi(0)\rangle + t \frac{d}{dt} |\Psi(t)\rangle_{t=0} + \mathcal{O}(T^2) \approx |\psi_I\rangle + \frac{t}{i\hbar} H(0)|\psi_I\rangle.
\]

(15.2)

If the time is short enough, then at \( t = T \), the first term will dominate the second, and we can approximate \( |\Psi(T)\rangle \approx |\psi_I\rangle \); however, this state is no longer an eigenstate of the Hamiltonian. The probability that it ends in a particular new eigenstate is therefore given simply by the overlap, and therefore

\[
P(I \rightarrow F) = |\langle \psi_F | \psi_I \rangle|^2.
\]

(15.3)
This is called the sudden approximation, and applies if the Hamiltonian changes quickly. What do we mean by quickly? From equation (15.2) we see that the corrections will be small if \( TH \psi_i / \hbar \) is small compared to \( \psi_i \). The Hamiltonian typically leads to a factor of the energy of various eigenstates, so we would expect (15.3) to hold provided \( E \Delta T / \hbar \ll 1 \), where \( E_i \) is an eigenvalue of \( H \). This turns out to be unnecessarily restrictive.

Recall that adding a constant to the Hamiltonian never affects anything, so our results can never depend on the actual energies themselves. A more careful analysis will reveal that the restriction really has to do with the difference in energies between eigenstates. Hence we can use (15.3) provided \( T \Delta E \ll \hbar \).

As a simple example of the sudden approximation, imagine a Harmonic oscillator with a time-dependant frequency, so that

\[
H(t) = \frac{1}{2m} p^2 + \frac{1}{2} m [\omega(t)]^2 x^2.
\]

where the frequency increase from \( \omega(0) = \omega_0 \) to \( \omega(T) = 2\omega_0 \) over a short time interval \( T \), as illustrated in Fig. 15-2. What is the probability that the particle in the ground state remains there as the frequency changes?

The ground state wave function is given by

\[
\psi_i(x) = \psi_0(x) = \left( \frac{m \omega_0}{\pi \hbar} \right)^{1/4} \exp \left( -\frac{m \omega_0 x^2}{2 \hbar} \right)
\]

and the final state wave function is identical if we replace \( \omega_0 \) by \( 2\omega_0 \). We therefore have

\[
P(I \rightarrow F) = \left| \int_{-\infty}^{\infty} dx \psi_\ast_f(x) \psi_i(x) \right|^2 = \sqrt{2m \omega_0} \sqrt{m \omega_0} \left[ \int_{-\infty}^{\infty} \exp \left( -\frac{m \omega_0 x^2}{2 \hbar} - \frac{m \omega_0 x^2}{\hbar} \right) dx \right]^2
\]

\[
= \sqrt{2} \frac{m \omega_0}{\pi \hbar} \left( \frac{2\pi \hbar}{3m \omega_0} \right)^2 = \frac{2\sqrt{2}}{3}.
\]

The condition of validity \( T \Delta E \ll \hbar \) is just \( \omega_0 T \ll 1 \).

B. The Adiabatic Approximation

The complementary situation is when the Hamiltonian changes very slowly. At each moment \( 0 < t < T \), imagine finding all the eigenstates \( \psi_i(t) \), satisfying

\[
H(t) \psi_i(t) = E_i(t) \psi_i(t).
\]

They will be orthonormal at all times, so in particular we see that

\[
\int dx \psi_\ast_f(x) \psi_i(x) = \delta_{f,i}.
\]
\[ 0 = \frac{d}{dt} \langle \psi_i | \psi_i \rangle = \langle \dot{\psi}_i | \psi_i \rangle + \langle \psi_i | \dot{\psi}_i \rangle = 2 \text{Re} \left( \langle \psi_i | \dot{\psi}_i \rangle \right), \]

where we have dropped the explicit time-dependence of our eigenstates, and written time derivatives with a dot. The real part of this overlap must vanish, but in fact the imaginary part is undetermined and undeterminable; we can change it simply by redefining the phase of our eigenstates as an arbitrary function of time. Hence we can choose the imaginary part to be anything we want, and to make things simple, we will define our eigenstates in such a way that

\[ \langle \psi_i | \dot{\psi}_i \rangle = 0. \]

Taking the time-derivative of (15.4), we can also demonstrate that the overlap between an eigenstate and the time derivative of any other eigenstate will also be small:

\[ \frac{d}{dt} (H | \psi_i \rangle) = \frac{d}{dt} (E_i | \psi_i \rangle), \]

\[ \dot{H} | \psi_i \rangle + H | \dot{\psi}_i \rangle = \dot{E}_i | \psi_i \rangle + E_i | \dot{\psi}_i \rangle. \]

Let the bra \( \langle \psi_j | \) act on both sides of this expression for some \( j \neq i \):

\[ \langle \psi_j | \dot{H} | \psi_i \rangle + \langle \psi_j | H | \dot{\psi}_i \rangle = \langle \psi_j | \dot{E}_i | \psi_i \rangle + \langle \psi_j | E_i | \dot{\psi}_i \rangle, \]

\[ \langle \psi_j | \dot{H} | \psi_i \rangle + E_j \langle \psi_j | \dot{\psi}_i \rangle = 0 + E_i \langle \psi_j | \dot{\psi}_i \rangle, \]

\[ \langle \psi_j | \dot{\psi}_i \rangle = \frac{1}{E_i - E_j} \langle \psi_j | \dot{H} | \psi_i \rangle \quad \text{for} \ j \neq i. \]  \hspace{1cm} (15.5)

Thus our eigenstates are changing little if the Hamiltonian \( H \) is changing slowly.

If we have eigenstates at all times, we can write our wave function in terms of them, so that

\[ | \Psi(t) \rangle = \sum_i c_i | \psi_i \rangle. \]

where both our coefficient \( c_i \) and our eigenstates \( | \psi_i \rangle \) are implicit functions of time. We now attempt to solve Schrödinger’s equation, (15.1). We have

\[ \frac{i\hbar}{\hbar} \frac{d}{dt} \sum_i c_i | \psi_i \rangle = H \sum_i c_i | \psi_i \rangle, \]

\[ i\hbar \sum_i \left[ \dot{c}_i | \psi_i \rangle + c_i | \dot{\psi}_i \rangle \right] = \sum_i c_i E_i | \psi_i \rangle. \]

Act on the left with \( \langle \psi_j | \) on both sides, using (15.5) and orthonormality.

\[ i\hbar \sum_i \left[ \dot{c}_i \langle \psi_j | \psi_i \rangle + c_i \langle \dot{\psi}_j | \psi_i \rangle \right] = \sum_i c_i E_i \langle \psi_j | \psi_i \rangle, \]

\[ i\hbar \dot{c}_j + i\hbar \sum_i c_i \langle \psi_j | \dot{\psi}_i \rangle = c_j E_j. \]
\[ i \hbar \frac{\partial}{\partial t} c_j = \sum_{i \neq j} \frac{i \hbar c_i}{E_j - E_i} \langle \psi_j | \dot{H} | \psi_i \rangle. \] (15.7)

Now, if the Hamiltonian is changing slowly, the right side will be small, and perhaps we can ignore it. This inspires the introduction of new constants \( b_j \), defined by

\[ b_j(t) \equiv \exp \left[ \frac{i}{\hbar} \int_0^t E_j(t') dt' \right] c_j(t), \] (15.8)

Solving for the \( c_j \)'s, and substituting into (15.7), we see that

\[ \left( i \hbar b_j + E_j b_j - E_j b_j \right) \exp \left[ -\frac{i}{\hbar} \int_0^t E_j(t') dt' \right] = \sum_{i \neq j} \frac{i \hbar c_i}{E_j - E_i} \langle \psi_j | \dot{H} | \psi_i \rangle \exp \left[ -\frac{i}{\hbar} \int_0^t E_i(t') dt' \right], \]

\[ \dot{b}_j = \sum_{i \neq j} \frac{b_j}{E_j - E_i} \langle \psi_j | \dot{H} | \psi_i \rangle \exp \left[ \frac{i}{\hbar} \int_0^t \left( E_j(t') - E_i(t') \right) dt' \right]. \]

We now simply integrate this expression over time to get the answer:

\[ b_j(T) = b_j(0) + \sum_{i \neq j} \int_0^T \frac{b_j(t)}{E_j(t) - E_j(t)} \exp \left[ \frac{i}{\hbar} \int_0^t \left( E_j(t') - E_i(t') \right) dt' \right] dt. \] (15.9)

This rather messy looking expression took a lot of work, but we are about to get a big payoff. First note that the factor \( \langle \psi_j | \dot{H} | \psi_i \rangle \) is small, because the Hamiltonian is changing slowly; however, it is being integrated over a large time \( T \). However, the exponential will be oscillating rapidly, while the other factors, such as \( b_i(t) \), will be changing relatively slowly. Hence, to leading order, the outer integral will vanish, and we can approximate \( b_j(T) = b_j(0) \). Since (15.8) tells us that \( c_j \) and \( b_j \) are the same up to a phase, we conclude that \( |c_j(T)|^2 = |c_j(0)|^2 \). Since it started in the eigenstate \( |\psi_i \rangle \), it ends in the state \( |\psi_i \rangle \) (except for phase), and therefore

\[ P(I \rightarrow F) = \delta_{Fj}. \] (15.10)

You can’t get a much simpler formula than that! It should be noted that this does not mean that the final state is identical with the original. The original and final states are, in general, eigenstates of different Hamiltonians. What has happened is that the eigenstates of the original Hamiltonian have smoothly transformed into the eigenstates of the final Hamiltonian. Hence if the particle starts in the ground state (or first excited state, or second . . .) it will end in the ground state (or first excited state, or second . . .). Equation (15.10) is called the adiabatic approximation.

When is the result (15.10) valid? Look carefully at equation (15.9). Let’s assume that \( \langle \psi_j | \dot{H} | \psi_i \rangle \) is not only small, but changing slowly, so that it is relatively constant on time scales of order \( T \). The exponential factor goes through its whole cycle (adding a total of zero to the integration) over one period, which is given by \( P = 2\pi \hbar / \Delta E \). If this time is much shorter than the characteristic time \( T \) during which the Hamiltonian
changes, then our approximation will be valid. In other words, we need \( T \Delta E \gg \hbar \). Obviously, this is just the complement of the condition for the sudden approximation. As an illustration of the adiabatic principle, consider an example identical to the one illustrated in Fig. 15-2, a harmonic oscillator that doubles its frequency, but this time assume \( T \omega_0 \gg 1 \). Then the probability of the ground state becoming the ground state is just 1.

Using the adiabatic approximation in more complicated situations can be rather complex. It might seem like the \( n \)th energy state always goes into the \( n \)'th energy state, provided the Hamiltonian changes slowly enough, but sometimes this just isn’t true. In general, if you gradually change the Hamiltonian, the various eigenenergies will shift up and down, and sometimes two states will approach each other in energy (a level crossing). The most common situation is that as the states become very close in energy, they will have a great deal of mixing, as is evident from the energy denominators, in equation (12.45b). This causes the two states to never really have the same energy, which generically leads to the kind of energy structure illustrated in Fig. 15-3. In such cases, the relevant characteristic time \( T \) is the time it takes us to move through the large mixing region, and the energy splitting \( \Delta E \) is dominated by this mixing region, where the energy splitting is at a minimum. Indeed, it is even possible for the two states to be completely unmixed, and then there can be a true level crossing, as illustrated in Fig. 15-3. In general, if there is absolutely no mixing between the levels, it is probably because the mixing is prevented by some symmetry. Provided a symmetry is present, the Hamiltonian can never change the corresponding eigenvalues, and the eigenstates will pass through the level crossing, after which the lower energy state will have become the higher, and vice versa.

**Figure 15-3:** Level mixing. Two states have energies that approach each other at some critical time. Generally, mixing will cause the lower energy eigenstate to always remain the lowest energy state (solid lines) so a system starting in state 1 will smoothly evolve to state 1, and similarly, 2 will evolve to 2. However, if there is a symmetry preventing mixing (dashed lines), the states will never mix, and a system beginning in state 1 will smoothly evolve to the higher energy state 2, and vice versa.
C. Time-dependent Perturbation Theory

We now approach the subject of time-dependent perturbation theory. Suppose we have a Hamiltonian that depends on time, but only by a small amount, so that

\[ H = H_0 + W(t), \]

where, in a manner similar to before, we assume \( W(t) \) is small. We will also assume that \( W(t) \) is non-zero only for \( 0 < t < T \). We will assume that we know the exact eigenstates of \( H_0 \). Let these states \( |\phi_n\rangle \) be orthonormal and have energy \( E_n \), so that

\[ H_0 |\phi_n\rangle = E_n |\phi_n\rangle. \]

Any quantum state can be written as a linear combination of these states. In particular, the exact wave function we are interested in can be written as

\[ |\Psi(t)\rangle = \sum_m c_m(t) |\phi_m\rangle, \]

where our goal is to find the functions \( c_m \). We start by substituting these wave functions into Schrödinger’s equation, which tells us

\[
i\hbar \frac{d}{dt} |\Psi\rangle = H |\Psi\rangle = [H_0 + W(t)] |\Psi\rangle,
\]

\[ \sum_m i\hbar \frac{d}{dt} c_m(t) |\phi_m\rangle = \sum_m c_m(t) [E_m + W(t)] |\phi_m\rangle.\]

Let the bra \( \langle \phi_n \rangle \) act on this to yield

\[ i\hbar \frac{d}{dt} c_n(t) = E_n c_n(t) + \sum_m c_m(t) \langle \phi_n | W(t) |\phi_m\rangle. \tag{15.11} \]

Now, since \( W \) is presumed small, we expect that the primary source of change for the coefficient \( c_n \) will be due to the right hand side, which suggests \( c_n(t) \sim \exp(-i E_n t / \hbar) \). We therefore define the quantity \( b_n(t) \) by

\[ c_n(t) = b_n(t) \exp(-i E_n t / \hbar). \tag{15.12} \]

Substituting this into (15.11) and simplifying, we find

\[
[i\hbar \dot{b}_n(t) + E_n] e^{-i E_n t / \hbar} = E_n b_n(t) e^{-i E_n t / \hbar} + \sum_m b_m(t) e^{-i E_m t / \hbar} \langle \phi_n | W(t) |\phi_m\rangle,
\]

\[ i\hbar \dot{b}_n(t) = \sum_m b_m(t) \langle \phi_n | W(t) |\phi_m\rangle \exp(i E_n t / \hbar - i E_m t / \hbar) \tag{15.13} \]

We now define the frequency difference and abbreviate the matrix element as

\[ \omega_{nm} \equiv (E_n - E_m) / \hbar, \]

\[ W_{nm}(t) \equiv \langle \phi_n | W(t) |\phi_m\rangle. \]
Substituting these into (15.13), we have
\[ \dot{b}_n(t) = \left(\frac{i\hbar}{\gamma}\right) \sum_m b_m(t) W_{nm}(t) e^{i\omega t}. \]

We now take this equation, and integrate it over time, to yield
\[ b_n(T) = b_n(0) + \frac{1}{i\hbar} \sum_m \int_0^T dt W_{nm}(t) e^{i\omega t} b_m(t). \]

In general, we want to assume that we are starting in an eigenstate \( |\phi_i\rangle \) of the initial Hamiltonian \( H_0 \), so that \( b_n(0) = c_n(0) = \delta_{ni} \). Then this equation becomes
\[ b_n(T) = \delta_{ni} + \frac{1}{i\hbar} \sum_m \int_0^T dt W_{nm}(t) e^{i\omega t} b_m(t). \]

Now, substitute this equation into itself repeatedly. Each time we do so, we will get a new term, with an extra integral. There will also be an additional sum, but one of the sums will be handled automatically by the Kronecker delta function. We find
\[ b_n(T) = \delta_{ni} + \frac{1}{i\hbar} \sum_m \int_0^T dt W_{nm}(t) e^{i\omega t} b_m(T). \]

It is clear we are producing a perturbation series in \( W \). Since we are primarily interested in the probability of going to a particular final state, let’s focus on \( n = F \). We’ll give this amplitude a special name, \( S_{FI} \), for which we have, to third order in \( W \),
\[ S_{FI} = b_F(T) = \delta_{FI} + \frac{1}{i\hbar} \int_0^T dt W_{FI}(t) e^{i\omega t} + \frac{1}{(i\hbar)^2} \sum_m \int_0^T dt W_{FM}(t) W_{MF}(t') e^{i\omega t} e^{i\omega t'} + \frac{1}{(i\hbar)^3} \sum_m \int_0^T dt W_{FM}(t) W_{MF}(t') W_{MF}(t') e^{i\omega t} e^{i\omega t'} + \cdots. \]

\( S_{FI} \) is often called the scattering matrix, or \( S \)-matrix. It is a unitary matrix representing the amplitude for ending in a particular state starting from an arbitrary state. The leading term is clearly the probability that nothing happens, the first term represents the amplitude that a single transition occurs, etc.

The probability of having a transition, then, is simply given by
\[ P(F \rightarrow F) = \left| \langle \phi_F | \Psi(T) \rangle \right|^2 = |S_{FI}|^2. \]

We are rarely interested in the probability of the state remaining the same, so usually we will drop the first term in (15.14). To leading order, the transition probability is given by

---

1 Much of this discussion is clearer in the interaction picture (see section 11D). In this picture, the states are changing only due to the interaction \( W \), so that the factors \( c_n(t) \) are never introduced, and eq. (15.13) is arrived at almost immediately. The phases in the interaction term simply represent the additional time dependence that appears in operators when relating the interaction and Schrödinger pictures.
\[ P(I \rightarrow F) = \frac{1}{\hbar^2} \int_0^\infty W_{fi}(t) e^{i\omega_{fi}t} dt. \] (15.16)

As a simple illustration, imagine we have a hydrogen atom initially in the ground state \(|100\rangle\) when a passing electromagnetic pulse produces an electric field \(E = E_0 \hat{z} \exp(-t^2/2T^2)\). This pulse is illustrated in Fig. 15-4. We would like to calculate the probability that this pulse causes a transition to a new state \(|nlm\rangle\). An electric field \(E\) can be produced from an electric potential \(U(r)\) with the relation \(E = -\nabla U(r)\). The easiest way to produce a constant field, then, is with a potential \(U = -eU\), which means that an electron with charge \(-e\) would have potential \(V = -eU\), so the perturbation is

\[ W = E_0 eZ \exp \left( -\frac{t^2}{2T^2} \right). \]

Now, in our approximations, we assumed that the perturbation is turned on at 0 and off at \(T\), but clearly this expression never vanishes. Nonetheless, the field dies away quickly enough that there is no harm in integrating over all time. Hence our probability is just

\[ P(I \rightarrow F) = \frac{1}{\hbar^2} \int_{-\infty}^\infty dt \langle nlm | Z | 100 \rangle E_0 e \exp \left( -\frac{t^2}{2T^2} \right) e^{i\omega_{fi}t} \left| \int_{-\infty}^\infty dt \exp \left( \frac{t^2}{2T^2} + i\omega_{fi}t \right) \right|^2. \]

\[ = E_0^2 e^2 \hbar^2 \left| \langle nlm | Z | 100 \rangle \right|^2 \left[ \int_{-\infty}^\infty dt \exp \left( -\frac{t^2}{2T^2} + i\omega_{fi}t \right) \right]^2, \]

\[ P(I \rightarrow F) = 2\pi E_0^2 e^2 \hbar^2 \left| \langle nlm | Z | 100 \rangle \right|^2 T^2 e^{-T^2 \omega_{fi}^2}. \] (15.17)

The matrix elements may be computed easily from the explicit form of the hydrogen wave functions. The operators \(R\) connect only states with \(l\) values that differ by one, so we must have \(l = 1\), and \(Z\) commutes with \(L_z\), so \(m = 0\).

It is interesting to note that the expression (15.17) vanishes both in the limit of small and large \(T\). Why? In the limit of small \(T\), we can evaluate the answer by the sudden approximation, in which the wave function doesn’t change in the short time available, and since the final Hamiltonian is identical with the initial Hamiltonian, the electron is still in the ground state. In the limit of large \(T\), the Hamiltonian is changing adiabatically, and therefore the ground state remains the ground state. Indeed, the maximum transition probability occurs when \(T\omega_{fi} = 1\), or \(T\Delta E = \hbar\), exactly the boundary given between the limiting constraints \(T\Delta E \ll \hbar\) and \(T\Delta E \gg \hbar\).
D. Harmonic Perturbations

Of particular interest, especially when studying the interaction of light with atoms, is perturbations which are harmonic, that is, having a time dependence like a sine or cosine. Suppose our perturbation takes the form

\[ W(t) = W e^{-i\omega t} + W^\dagger e^{i\omega t}. \]  

(15.18)

We will implicitly assume \( \omega > 0 \), so we are definitely not talking about a constant perturbation. The operators \( W \) and \( W^\dagger \) are independent of time. For example, if we pick \( W \) to be a Hermitian (anti-Hermitian) operator, we will simply have an operator with a cosine (sine) time dependence. Of course, to have a truly harmonic source, it should behave this way forever, from minus infinity to plus infinity, so we will be interested primarily in situations where \( T \) is large enough that the perturbation goes through many cycles of oscillation. To help give the “sense” of it being on forever, and to make the math a tad simpler, we will turn it on at \(-\frac{T}{2}\) and off again at \(+\frac{T}{2}\), changing our limits in (15.14).

Let’s calculate the scattering matrix assuming the initial and final states are distinct. We’ll only include the first order perturbation, so that

\[ S_{FI} = \frac{1}{i\hbar} \left[ dt \left[ W_{FI} e^{-i\omega t} + W_{FI}^\dagger e^{i\omega t} \right] e^{i\omega t} \right] = \frac{1}{i\hbar} \left[ W_{FI} e^{i(\omega F_I - \omega) t} + W_{FI}^\dagger e^{i(\omega + \omega F_I) t} \right] \right|_{-\frac{T}{2}}^{\frac{T}{2}}, \]

(15.19)

where \( W_{FI} \equiv \langle \phi_I | W | \phi_F \rangle \). We now wish to make an approximation. Assuming the perturbation \( W \) is small, this expression will be small unless one of the two denominators is small. We will therefore focus on regions where one or the other denominators is small, but not both, since we assumed \( \omega > 0 \). For example, suppose \( \omega F_I > 0 \) and is close to \( \omega \), then the first term will dominate the second. The probability for a transition will then take the form

\[ P(I \rightarrow F) = \frac{4|W_{FI}|^2}{\hbar^2} \sin^2 \left[ \frac{1}{2} (\omega_{FI} - \omega) T \right]. \]

(15.20)

Consider for the moment the last factor. If \( T \) is large but finite, it is a function with a very large peak at \( \omega = \omega_{FI} \) which grows taller and narrower as \( T \) becomes larger as illustrated in Fig. 15-5. This suggests that in the limit of large \( T \), it approaches a

Figure 15-5: A plot of the frequency dependence of the probability of a transition, plotted for \( \omega_{FI} T = 40 \). As \( T \) increases, the peak becomes narrower and taller, increasingly resembling a Dirac delta function.
To determine the strength of the delta function, we simply integrate both sides. The integral on the right is just $A$, and the left side can be done by contour integration (Maple can do it) to yield $\frac{1}{2} \pi T$, so $A = \frac{1}{2} \pi T$. Thus we have

$$P(I \to F) = 2 \pi h^{-2} |W_{fi}|^2 \delta(\omega_{fi} - \omega) T.$$  

A couple of comments are in order here. First of all, the probability is proportional to time, which makes sense. The longer you wait, the more chance you have for a conversion to occur. The probability per unit time is called the decay rate and denoted $\Gamma$. With the help of (A.40), we can easily show that we can include a factor of $\hbar$ inside the delta function if we include a factor of $\hbar$ outside as well, so this can be rewritten, using $\omega_{nm} = (E_n - E_m)/\hbar$, as

$$\Gamma(I \to F) = 2 \pi h^{-1} |W_{fi}|^2 \delta(E_F - E_I - \hbar \omega), \quad E_F > E_I. \quad (15.21)$$

We have completed the case where the final energy is greater than the initial energy. We now want to calculate the other case. Rather than reversing the inequality, let’s keep $E_F > E_I$ but simply reverse the role of the final and initial state, so that $|\phi_F\rangle$ is the initial state, and $|\phi_I\rangle$ the final state. Without going through the details, we will only keep the second term in (15.19), and we convert to a delta function as before. We find

$$\Gamma(F \to I) = 2 \pi h^{-1} |W_{IF}|^2 \delta(E_I - E_F + \hbar \omega), \quad E_I > E_F. \quad (15.22)$$

Note that in (15.21) and (15.22), energy is not conserved. This is to be expected when the Hamiltonian is not time-dependent. The external source of the Hamiltonian is being treated as an inexhaustible source of energy. Clearly, when our quantum state is increasing in energy, from (15.21) we see that it must be absorbing an energy of $\hbar \omega$ from the source of this external perturbation, and from (15.22), it is emitting this amount of energy. When we interpret this external source as an electromagnetic wave, we will simply interpret this as the absorption or emission of one photon of energy. However, to make this rigorous, we will have to quantize the electromagnetic field, a task we wish to put off until a later chapter.

Comparison of (15.21) and (15.22) reveals a surprising fact. The delta functions differ only by a minus sign in their arguments and are therefore identical, and for any operator $W$, $W^\dagger = W^*$, so the matrix elements are identical as well. Hence the two rates are equal. This presents a small paradox. Suppose we place a collection of atoms in a thermal background of electromagnetic energy (black body radiation). According to (15.21) and (15.22), the atom is as likely to transition from, say, a 1s state to a 2p state as vice versa. Hence, given enough time, there will be equal numbers of electrons in the two states (actually, there would be more in the 2p states, since there are three times as many states). We will ultimately realize that (15.21) is basically correct, but (15.22) is
incorrect. The flaw will be corrected when we quantize the electromagnetic field in chapter seventeen, and then include interactions with atoms in chapter eighteen.

How do we interpret the delta functions in (15.21) and (15.22)? If our system has only discrete energy levels, these are a bit difficult to interpret, since a delta function is either zero or infinity. One solution in this case is to revert to forms like (15.20). We will find that if the energy conservation conditions are nearly satisfied, the rate is large but not infinite. In contrast, if they aren’t well matched, the transitions do not grow with time. Roughly speaking, it is easy to show that the rate will be large provided the energy is matched with an approximate error $T \Delta E < \hbar$. This is yet another uncertainty relationship from quantum mechanics, one we will discuss later.

In other situations, the delta function is not really problematic. One that is commonly encountered is a situation where the perturbation is not strictly speaking harmonic, but rather a complicated function of many frequencies. In such circumstances, it is possible to break the perturbation into its Fourier components, and then add up (integrate) over all frequencies. The delta functions will pick out only that Fourier component with the right frequency to cause a transition. For example, in a thermal bath of electromagnetic radiation (black body radiation), there are electromagnetic waves of all frequencies present. An integration over all frequencies is necessary, in which case the delta function will make the computation easier.

Another common situation is when there are a continuum of final states $|\phi_F\rangle$; for example, if the initial state is a bound electron and the final state a free electron. Then the final state eigenstates might be labeled $|k\rangle$, and have energy

$$H_0 |k\rangle = E_k |k\rangle = \frac{\hbar^2 k^2}{2m} |k\rangle.$$  

These states will be normalized, $\langle k'|k\rangle = \delta^3 (k' - k)$. We are not generally interested in the probability into a particular final state $|k\rangle$, but rather the probability of the electron coming free, which will be given by

$$\Gamma (I \to |k\rangle) = \int d^3k \frac{2\pi}{\hbar} |W_{ki}|^2 \delta (E_k - E_i - \hbar \omega) = \frac{2\pi}{\hbar} \int d\Omega \int_0^\infty k^2 dk |W_{ki}|^2 \delta \left( \frac{\hbar^2 k^2}{2m} - E_i - \hbar \omega \right)$$

$$= \frac{2\pi mk^2}{\hbar^2 k} \int d\Omega |W_{ki}|^2 = \frac{2\pi mk}{\hbar^3} \int d\Omega |W_{ki}|^2.$$  

The matrix element $W_{ki}$ will generally depend on direction. It is also possible to find the differential rate $d\Gamma/d\Omega$ simply by not performing the final angular integral.
E. Electromagnetic Waves and the Dipole Approximation

We consider now a special circumstance of considerable practical importance, the effect of electromagnetic plane waves on atoms. The full Hamiltonian will be given by (9.21), repeated here for clarity:

$$ H = \sum_{j=1}^{N} \left\{ \frac{1}{2m} \left[ p_j + e A \left( r_j, t \right) \right]^2 - e U \left( r_j, t \right) + \frac{e}{m} B \left( r_j, t \right) \cdot s_j \right\} + V_{a} \left( r_1, r_2, \ldots \right), $$

where $V_{a} \left( r_1, r_2, \ldots \right)$ represents all the internal interactions of the atom, and $A$ and $B$ represent the effects of the external field, and (for this chapter and beyond) we have approximated $g = 2$ for the electron. We will now define the unperturbed Hamiltonian $H_0$ as

$$ H_0 = \sum_{j=1}^{N} \frac{1}{2m} p_j^2 + V_{a} \left( r_1, r_2, \ldots \right). $$

This Hamiltonian will have atomic eigenstates $|\phi_m\rangle$ with energy $E_m$. The perturbation will be all the remaining terms, but since we only want to keep only leading order in the electromagnetic fields, we will drop the $A^2$ term. So we have

$$ W \left( t \right) = \sum_{j=1}^{N} \left\{ \frac{e}{m} \left[ A \left( r_j, t \right) \cdot p_j + B \left( r_j, t \right) \cdot s_j \right] - e U \left( r_j, t \right) \right\}. \quad (15.23) $$

We now wish to include an electromagnetic wave. There is more than one way (“gauge choice”) to write such a wave, which we will address more carefully in chapter seventeen, but for a specific choice of gauge (“Coulomb Gauge”), it can be written as

$$ U \left( r, t \right) = 0, \quad A \left( r, t \right) = A_0 \left( e^{ik \cdot r - i\omega t} + e^{* -ik \cdot r + i\omega t} \right). \quad (15.24) $$

where $e$ is a unit vector (the “polarization”) normalized so that $e \cdot e^* = 1, \ e \cdot k = 0$, and $A_0$ denotes the amplitude of the wave. The electric and magnetic fields are given by

$$ E = -\nabla U \left( r, t \right) - A \left( r, t \right) = iA_0 \omega \left( e^{ik \cdot r - i\omega t} - e^{* -ik \cdot r + i\omega t} \right), $$
$$ B = \nabla \times A = iA_0 \left[ \left( k \times e \right) e^{ik \cdot r - i\omega t} - \left( k \times e^* \right) e^{* -ik \cdot r + i\omega t} \right]. $$

Substituting into (15.22), we have

$$ W \left( t \right) = \frac{e}{m} A_0 \sum_{j=1}^{N} \left\{ e^{ik \cdot r_j - i\omega t} \left[ e \cdot p_j + i \left( k \times e \right) \cdot s_j \right] + e^{* -ik \cdot r_j + i\omega t} \left[ e^* \cdot p_j - i \left( k \times e^* \right) \cdot s_j \right] \right\}. $$

Comparison with (15.18) then tells us the perturbation with the time part factored out is

---

1 Equation (15.23) seems to contain an error, since we commuted $A$ with $P_i$, not always a valid assumption. This assumption turns out to work in Coulomb gauge, since the choice of $A$ given in (15.24) has no divergence.

2 The polarization vector $e$ may be chosen real, yielding cosine dependence, or pure imaginary, yielding sine dependence. More complicated possibilities exist; for example, one may make one component real and another imaginary, resulting in circular polarization.
\[ W = \frac{e}{m} A_0 \sum_{j=1}^{N} e^{ijR_j} \left[ \mathbf{e} \cdot \mathbf{P}_j + i (\mathbf{k} \times \mathbf{e}) \cdot \mathbf{S}_j \right]. \quad (15.25) \]

We now want to spend a moment looking at (15.25). We will consider the case where we are using electromagnetic waves with energy \( \hbar \omega \) comparable to the energy level splittings of an atom. These energy levels are typically of order \( mc^2 \alpha^2 \), where \( \alpha = \frac{1}{137} \) is the fine structure constant. A wave number \( k \) for such light will be of order \( k \sim mc^2 \alpha^2 / \hbar c = mc \alpha^2 / \hbar \). In contrast, the characteristic size of an atom will tend to be of order \( a_0 = h^2 / k e^2 m = \hbar / ec \). As a consequence, we have \( k \cdot \mathbf{R}_j \sim ka_0 \sim \alpha \). This is a small number, so to leading order, we simply treat it as zero, and drop the phase factor in (15.25). Similarly, if you compare the first and second terms in (15.25), the first term will be of order \( P \sim \hbar / a_0 = mc \), while the second will be of order \( k \mathbf{S} \sim \hbar k \sim mc^2 \).

Hence we can ignore this as well. In summary, to leading order, our perturbation is

\[ W_{E1} = \frac{e}{m} A_0 \sum_{j=1}^{N} \mathbf{e} \cdot \mathbf{P}_j = \frac{e}{m} A \mathbf{e} \cdot \mathbf{P}, \]

where \( \mathbf{P} \) is just the sum of the momenta of all the electrons. This approximation is called the electric dipole approximation, and the \( E1 \) subscript just denotes this fact. The transition rate between two atomic states is then given by (15.21), assuming the final energy is greater than the initial energy.

\[ \Gamma (I \rightarrow F) = \frac{2\pi A_0^2 e^2}{\hbar^2} \left| \mathbf{e} \cdot \langle \phi_e | \mathbf{P} | \phi_f \rangle \right|^2 \delta (\omega_{fi} - \omega). \quad (15.26) \]

A couple of observations allow us to convert this into a more convenient form. The first is to note that the commutator of \( H_0 \) with \( [ \mathbf{R}_0, \mathbf{R}_j ] \) is given by

\[ [H_0, \mathbf{R}] = \left[ \sum_k \frac{1}{2m} \mathbf{P}_k^2 + V_0 (\mathbf{R}_1, \mathbf{R}_2, \ldots), \sum_j \mathbf{R}_j \right] = \frac{1}{2m} \sum_{k,j} [\mathbf{P}_k^2, \mathbf{R}_j] = -\frac{i\hbar}{m} \sum_j \mathbf{P}_j = -\frac{i\hbar}{m} \mathbf{P}. \]

It follows that the matrix elements can be rewritten as

\[ \langle \phi_e | \mathbf{P} | \phi_f \rangle = \frac{im}{\hbar} \langle \phi_e | [H_0, \mathbf{R}] | \phi_f \rangle = \frac{im}{\hbar} \langle \phi_e | (E_f \mathbf{R} - \mathbf{R} E_f) | \phi_f \rangle = im\omega_{fi} \langle \phi_e | \mathbf{P} | \phi_f \rangle. \quad (15.27) \]

It is also useful to rewrite the amplitude in terms of intensity. The intensity of a wave is given by the time-averaged magnitude of the Poynting vector,\(^1\) which is given by

\[ \langle \mathbf{S} \rangle = \frac{1}{\mu_0} \langle \mathbf{E} \times \mathbf{B} \rangle = \frac{A_0^2 \omega}{\mu_0} \left[ \mathbf{e} \times (\mathbf{k} \times \mathbf{e})' + \mathbf{e}' \times (\mathbf{k} \times \mathbf{e}) \right] = \frac{2A_0^2 \omega}{\mu_0} \mathbf{k}. \]

Recalling that for light we have \( \omega = c k \), and finally remembering that \( \mu_0 = 4\pi k_e / c^2 \), we find that we can write the intensity \( \mathcal{I} \), or power per unit area, as

\[ \mathcal{I} = \frac{1}{\mu_0} \mathcal{E} \cdot \mathbf{H} = \frac{A_0^2 \omega}{\mu_0 \sqrt{2}} \mathbf{k}. \]

---

Substituting (15.27) and (15.28) into (15.26) then yields

\[ \Gamma(I \rightarrow F) = \frac{4 \pi^2 \kappa e^2 \omega_f \mathcal{I}}{\hbar c \alpha \omega} |\mathbf{e} \cdot \mathbf{r}_{fi}|^2 \delta(\omega_f - \omega). \]

The matrix element is often abbreviated \( \mathbf{r}_{fi} \equiv \langle \phi_f | \mathbf{R} | \phi_i \rangle \) (the electric dipole matrix element). The delta function assures that \( \omega_f = \omega \), and we can write this more easily with the help of the fine structure constant \( \alpha = k_e e^2 / \hbar c \), so this simplifies to

\[ \Gamma(I \rightarrow F) = 4 \pi^2 \alpha \hbar^{-1} \mathcal{I} |\mathbf{e} \cdot \mathbf{r}_{fi}|^2 \delta(\omega_f - \omega). \]

This yields an infinite rate if the frequency is appropriate. If the incoming wave has a range of frequencies, as all real sources do, then the intensity can be described as an intensity distribution, \( \mathcal{I}(\omega) \) with units of power per unit area per unit angular frequency, so that

\[ \mathcal{I} = \int \mathcal{I}(\omega) d\omega. \]

We therefore modify (15.29) by including an appropriate integral, yielding

\[ \Gamma(I \rightarrow F) = 4 \pi^2 \alpha \hbar^{-1} \mathcal{I}(\omega_f) |\mathbf{e} \cdot \mathbf{r}_{fi}|^2. \]

The rate will be identical for the reverse reaction.

Equation (15.30) is appropriate for an electromagnetic wave with known polarization vector \( \mathbf{e} \). If the polarization is random, we need to average over both polarizations perpendicular to the direction of the wave \( \mathbf{n} \). The easiest way to do this is to sum over all three potential polarizations and subtract the “missing” polarization \( \mathbf{n} \).

The result is

\[ \Gamma_{\text{unpol}} = 2 \pi^2 \alpha \hbar^{-1} \mathcal{I}(\omega_f) \left( |\mathbf{r}_{fi}|^2 - |\mathbf{n} \cdot \mathbf{r}_{fi}|^2 \right). \]

If the incoming wave is coming in random direction, or if the atom is itself oriented in a random direction, we need to average over all angles, which yields

\[ \Gamma_{\text{random}} = \frac{\pi^2}{3} \alpha \hbar^{-1} \mathcal{I}(\omega_f) |\mathbf{r}_{fi}|^2. \]

Let’s try applying our formulas in a simple case. Consider a hydrogen atom in the ground state that is influenced by an electromagnetic field with sufficient frequency so that the electron can be ionized. The initial state is \( |100\rangle \), but what shall we use for the final state? Far from the atom, the electron will be approximately a plane wave, but nearby the wave function is much more complicated. To simplify, assume the frequency

---

1 A more likely experimental quantity than the intensity per unit angular frequency would be intensity per unit frequency \( \mathcal{I}(f) = 2 \pi \mathcal{I}(\omega) \) or intensity per unit wavelength \( \mathcal{I}(\lambda) = \omega^2 \mathcal{I}(\omega) / 2 \pi c \).
is much higher than is necessary to free an electron from hydrogen. In this case the final state will also be nearly a plane wave near the atom, since the kinetic energy is much greater than a typical binding energy. We need to calculate the matrix element

$$\mathbf{r}_{fi} = \langle \phi_f | \mathbf{R} | \phi_i \rangle = \langle \mathbf{k} | \mathbf{R} | 100 \rangle,$$

which with the help of (15.27) we can write as

$$\mathbf{r}_{fi} = \langle \mathbf{k} | \mathbf{R} | 100 \rangle = \frac{1}{i\omega} \langle \mathbf{k} | \mathbf{P} | 100 \rangle = \frac{\hbar \mathbf{k}}{i\omega} \langle \mathbf{k} | 100 \rangle.$$

We now simply put in the specific form of the wave function and take its Fourier transform to give

$$\mathbf{r}_{fi} = \frac{-i\hbar \mathbf{k}}{i\omega} \left( 2\pi \right)^{3/2} \int d^3r e^{-i\mathbf{k} \cdot \mathbf{r}} e^{-\mathbf{r}^2/4a_0^2} = \frac{-i\hbar \mathbf{k}}{i\omega} \left( 2\pi \right)^{3/2} \int d\cos\theta \int_0^\infty r^2 dr e^{-r^2/4a_0^2 - ikr \cos\theta}$$

$$= -\frac{\sqrt{2}\hbar \mathbf{k}}{i\omega} \left( 2\pi \right)^{3/2} \int_0^\infty d\cos\theta \frac{d\cos\theta}{(a_0^{-1} + ik \cos\theta)} = \frac{\hbar \mathbf{k}}{i\omega} \left( a_0^{-1} + ik \cos\theta \right)^{-2} \int_0^\infty d\cos\theta = -\frac{\sqrt{2}\hbar \mathbf{k}}{i\omega} \left( 2\pi \right)^{3/2} \left( a_0^{-1} + ik \right)^{-2} \left( a_0^{-1} - ik \right)^{-2}.$$

Substituting this into (15.29), we have

$$\Gamma(1s \rightarrow \mathbf{k}) = 4\pi^2 \alpha \hbar^{-1} I \left( \frac{8\hbar^3 a_0^3}{\pi^2 m^2 \omega^2} \right) \left( 1 + a_0^2 k^2 \right)^{-4} |\mathbf{\varepsilon} \cdot \mathbf{k}|^2 \delta \left( \omega_{fi} - \omega \right)$$

$$= 32\alpha I a_0^3 \hbar^2 |\mathbf{\varepsilon} \cdot \mathbf{k}|^2 m^{-2} \omega^{-2} \left( 1 + a_0^2 k^2 \right)^{-4} \delta \left( \frac{\hbar^2 k^2}{2m} - E_f - \hbar \omega \right).$$

The dot product is just $\mathbf{\varepsilon} \cdot \mathbf{k} = k \cos\theta$ where $\theta$ is the angle between the polarization of the light and the direction of the final electron. We would then integrate this over all possible final wave numbers $\mathbf{k}$ to yield

$$\Gamma(1s \rightarrow \mathbf{k}) = \frac{32\alpha I a_0^3 \hbar^2 k^2}{m^2 \omega^2 \left( 1 + a_0^2 k^2 \right)^4} \int_0^\infty k^2 dk \int \cos^2\theta d\Omega,$$

$$\frac{d\Gamma}{d\Omega} = \frac{32\alpha I a_0^3 \hbar^3 k^3}{m^2 \omega \left( 1 + a_0^2 k^2 \right)^4} \cos^2\theta,$$

where the final wave number $k$ is defined by $\hbar^2 k^2 / 2m = E_f + \hbar \omega$. The final angular integral, if desired, yields an additional factor $\int d\Omega \cos^2\theta = \frac{1}{4} \pi$. 

© 2011, Eric D. Carlson 281
F. Beyond the Dipole Approximation

Suppose the dipole moment \( \mathbf{r}_F = \langle \phi_F | \mathbf{R} | \phi_i \rangle \) between an initial and final state vanishes. After all, the matrix element will generally vanish unless \( l_F = l_i \pm 1 \), so this will commonly be the case. Does this mean that the decay does not occur? No, it only means that its rate will be suppressed. Let us expand out the expression (15.25), but this time keeping more than just the leading term. We have

\[
W = \frac{\alpha}{m} A_0 \sum_{j=1}^{N} \left[ \epsilon \cdot \mathbf{P}_j + i \left( \mathbf{k} \cdot \mathbf{R}_j \right) \left( \epsilon \cdot \mathbf{P}_j \right) + i \left( \mathbf{k} \times \epsilon \right) \cdot \mathbf{S}_j \right].
\]  

(15.31)

The leading term is responsible for electric dipole transitions. The middle term can be rewritten, with the help of some fancy vector identities, in the form

\[
\left( \mathbf{k} \cdot \mathbf{R}_j \right) \left( \epsilon \cdot \mathbf{P}_j \right) + \left( \mathbf{k} \cdot \mathbf{P}_j \right) \left( \epsilon \cdot \mathbf{R}_j \right) = \frac{i m}{\hbar} \left[ H_0, \left( \mathbf{k} \cdot \mathbf{R}_j \right) \left( \epsilon \cdot \mathbf{R}_j \right) \right].
\]

The latter term contains the angular momentum \( \mathbf{L}_j \) of the \( j \)th electron, and it is not hard to show that the first term can be written as a commutator, in a manner similar to before

\[
\left( \mathbf{k} \cdot \mathbf{R}_j \right) \left( \epsilon \cdot \mathbf{P}_j \right) + \left( \mathbf{k} \cdot \mathbf{P}_j \right) \left( \epsilon \cdot \mathbf{R}_j \right) = \frac{i m}{\hbar} \left[ H_0, \left( \mathbf{k} \cdot \mathbf{R}_j \right) \left( \epsilon \cdot \mathbf{R}_j \right) \right].
\]

With a bit of work, we then find the matrix elements are given by

\[
\langle \phi_F | W | \phi_i \rangle = \frac{i A_0 e}{m} \sum_{j=1}^{N} \frac{i m}{\hbar} \left[ H_0, \left( \mathbf{k} \cdot \mathbf{R}_j \right) \left( \epsilon \cdot \mathbf{R}_j \right) \right] + \frac{i}{2} \left( \mathbf{k} \times \epsilon \right) \cdot \mathbf{L}_j + \left( \mathbf{k} \times \epsilon \right) \cdot \mathbf{S}_j \right] \langle \phi_F | \phi_i \rangle,
\]

\[
W_{F \rightarrow i} = -\frac{i}{2} \omega_{F_i} A_0 e \sum_{j=1}^{N} \langle \phi_F | \left( \mathbf{k} \cdot \mathbf{R}_j \right) \left( \epsilon \cdot \mathbf{R}_j \right) | \phi_i \rangle + \frac{i A_0 e}{m} \left( \mathbf{k} \times \epsilon \right) \cdot \langle \phi_F | \left( \frac{i}{2} \mathbf{L}_j + \mathbf{S}_j \right) | \phi_i \rangle.
\]

The first term is referred to as the electric quadrupole term, and the second is the magnetic dipole term. Both terms commute with parity, unlike the electric dipole term, and therefore they can only connect states with the same parity. Hence the states connected by these terms are always different from the ones connected by the electric dipole term, and there is no interference between the two. The electric quadrupole term is a rank two spherical tensor operator; as such it changes \( l \) by either zero or two, but cannot connect two \( l = 0 \) states. The magnetic dipole term commutes with \( H_0 \), assuming we ignore spin-orbit coupling, hyperfine splitting, and external magnetic fields. Hence it can only cause transitions for states that are split by these smaller effects.

These expressions can then be used to compute rates, in a manner similar to the electric dipole moment. For example, the quadrupole contribution to decay is given by

\[
\Gamma (I \rightarrow F) = \frac{\alpha^2}{\hbar} \sum_{j=1}^{N} \left( \mathbf{k} \cdot \mathbf{R}_j \right)^2 \left( \epsilon \cdot \mathbf{R}_j \right)^2 \left| \langle \phi_F | \phi_i \rangle \right|^2.
\]

This is suppressed compared to the electric dipole rate by a factor \( (\mathbf{k} \cdot \mathbf{R})^2 \), typically of order \( \alpha^2 \).
G. Time-dependent perturbation theory with a constant perturbation

The last type of perturbation we want to consider is one that is constant in time, \( W(t) = W \). It may seem odd to use time-dependent perturbation theory in such a circumstance, but in fact, it is often a very good way of thinking about things. For example, if I scatter a particle off of a potential, we intuitively imagine a particle approaching the potential (initially in a plane wave), interacting with the potential (the interaction) and then leaving the potential again (a plane wave). Although the potential is in fact always constant, it is easier to think of it as a temporary perturbation.

Our starting point will once again be (15.14), but we will make a couple of modifications. First of all, we ultimately want to consider the perturbation as constant, not having time dependence at all, so that it is always on and always will be on. For this reason, we will change the integration to run from \(-\frac{1}{2}T\) to \(\frac{1}{2}T\). Eq. (15.14) then will look like

\[
S_{fl} = \delta_{fl} + \frac{W_{fl}}{i\hbar} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} dt e^{i\omega t} + \sum_{m} \frac{W_{lm}}{(i\hbar)^3} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} dt \int_{-\frac{1}{2}T}^{\frac{1}{2}T} dt' \theta(t-t') e^{i\omega t} e^{i\omega t'}
\]

\[+ \sum_{m} \sum_{n} W_{mn} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} dt \int_{-\frac{1}{2}T}^{\frac{1}{2}T} dt' \theta(t-t')\theta(t'-t'') e^{i\omega t} e^{i\omega t'} e^{i\omega t''} + \cdots , (15.32)\]

where we have used the Heaviside function, defined by

\[
\theta(\tau) \equiv \begin{cases} 1 & \text{if } \tau > 0 , \\ 0 & \text{if } \tau < 0 . \end{cases}
\]

We now wish to prove the identity

\[
\theta(\tau) = \lim_{\epsilon \to 0^+} \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{e^{i\omega \tau} d\omega}{\omega - i\epsilon} . (15.33)
\]

This integral will be evaluated by the method of contour integration.

Contour integration is a method that can be used to perform integrals in the complex plane. As defined, the integral in (15.33) runs from minus infinity to infinity along the real axis. The first step of performing a contour integral is to “close” the contour: to add an additional path from infinity back to minus infinity to make a complete loop. The trick is to choose that path such that it contributes negligibly to the integral. Since we are going to be wandering into values of \( \omega \) that are imaginary, our goal is to pick values of \( \omega \) such that \( e^{i\omega \tau} \) will be very small. This will be the case provided

\[\text{Figure 15-6: Contours for performing the integral (15.33). To the initial path (purple line), we add a half-circle at infinity in the upper half plane for } \tau > 0 \text{ (red semi-circle) or the lower half plane for } \tau < 0 \text{ (blue semi-circle). The orange X is the pole at } \omega = i\epsilon .\]
\(i\omega\tau\) has a large negative real part. If \(\tau > 0\), this is simply achieved by forcing \(\text{Im}(\omega) > 0\) (the “upper half plane”) whereas if \(\tau < 0\), we force \(\text{Im}(\omega) < 0\) (the “lower half plane”). We can therefore add a half loop “at infinity” to make the whole integral a closed loop, as illustrated in Fig. 15-6.

We now employ the magic of contour integration, which says that the result of an integral is independent of the path chosen, provided we do not cross the pole. We can therefore shrink, in each case, the integral to a tiny circle, enclosing the pole in the case \(\tau > 0\), and enclosing nothing for \(\tau < 0\), as illustrated in Fig. 15-7. While we are at it we can take the limit \(\varepsilon \to 0\), which places the pole at the origin. We now have

\[
\lim_{\varepsilon \to 0} \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{e^{i\omega \tau} d\omega}{\omega - i\varepsilon} = \begin{cases} \int_{0}^{\infty} e^{i\omega \tau} d\omega / \omega & \text{if } \tau > 0, \\ \int_{-\infty}^{0} e^{i\omega \tau} d\omega / \omega & \text{if } \tau < 0. \end{cases}
\]

The loop for \(\tau < 0\) can be shrunk to zero with impunity, and we can locate it well away from the pole if we wish, so the integral vanishes. For \(\tau > 0\), we shrink it to a tiny circle of radius \(\delta\). In the limit \(\delta \to 0\) we have \(\omega \approx 0\), and we can approximate \(e^{i\omega \tau} = 1\). However, we dare not make this approximation in the denominator. We parameterize the tiny circle around the origin by an angle \(\phi\) and let \(\omega = \delta e^{i\phi}\), where \(\phi\) runs from 0 to \(2\pi\). Then we have, for \(\tau > 0\),

\[
\lim_{\varepsilon \to 0} \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{e^{i\omega \tau} d\omega}{\omega - i\varepsilon} = \frac{1}{2\pi i} \int_{0}^{2\pi} d(\delta e^{i\phi}) = \frac{1}{2\pi i} \int_{0}^{2\pi} i d\phi = 1,
\]

while it vanishes for \(\tau < 0\), which proves (15.33).

We now wish to use this identity to simplify (15.32). We replace each of the Heaviside functions with the messy integral (15.33), which results in

\[
S_{F_1} = \delta_{F_1} + \frac{W_{F_1}}{i h} \int_{-\frac{T}{4}}^{\frac{T}{4}} dt e^{i\omega t} + \lim_{\varepsilon \to 0} \sum_{m} \frac{W_{F_m}}{(i h)^2} \int_{-\frac{T}{4}}^{\frac{T}{4}} dt \int_{-\frac{T}{4}}^{\frac{T}{4}} dt' e^{i\omega t} e^{i\omega t'} \int_{-\infty}^{\infty} \frac{e^{i\omega (t-t')} d\omega}{2\pi i (\omega - i\varepsilon)} + \sum_{m} \sum_{n} \frac{W_{F_m} W_{m_n}}{(i h)^3} \int_{-\frac{T}{4}}^{\frac{T}{4}} dt \int_{-\frac{T}{4}}^{\frac{T}{4}} dt' e^{i\omega t} e^{i\omega t'} \int_{-\infty}^{\infty} \frac{e^{i\omega (t-t')} d\omega}{2\pi i (\omega - i\varepsilon)} + \cdots
\]

We now proceed to perform all but one of the time integrals. We will do so simply by working in the infinite time limit, so that the integrals turn into trivial expressions like \(\int_{-\infty}^{\infty} e^{i(\omega t - \omega')} dt' = 2\pi \delta(\omega_m - \omega)\). Using this on all the integrals except for \(dt\), we obtain

\[\text{Figure 15-7: The contours after they have been shrunk. The pole has been moved to } \omega = 0.\]
\[
S_{Fl} = \delta_{Fl} + \frac{W_{Fl}}{i\hbar} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} dt e^{i\omega_F t} + \lim_{\epsilon \to 0^+} \left\{ \sum_{m} W_{Fm} W_{ml} \frac{1}{(i\hbar)^2} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} e^{i\omega t} dt \int_{-\infty}^{\infty} \frac{\delta(\omega - \omega_m) d\omega}{i(\omega - i\epsilon)} \right. \\
+ \sum_{m} \sum_{n} W_{Fm} W_{mn} W_{nl} \frac{1}{(i\hbar)^3} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} e^{i\omega t} \delta(\omega_m + \omega' - \omega) d\omega \int_{-\infty}^{\infty} \frac{\delta(\omega_n - \omega') d\omega'}{i(\omega' - i\epsilon)} + \ldots \right\}.
\]

We now use the Dirac delta functions to do all the \(\omega\)-integrations.

\[
S_{Fl} = \delta_{Fl} + \frac{W_{Fl}}{i\hbar} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} dt e^{i\omega_F t} + \lim_{\epsilon \to 0^+} \left\{ \sum_{m} W_{Fm} W_{ml} \frac{1}{(i\hbar)^2} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} e^{i(e_{\omega_m} + e_{\omega'}) t} dt \\
+ \sum_{m} \sum_{n} W_{Fm} W_{mn} W_{nl} \frac{1}{(i\hbar)^3} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} e^{i(e_{\omega_m} + e_{\omega'}) t} dt \right\} + \ldots.
\]

We rewrite \(\hbar \omega_c = E_i - E_f\), and write this in the form

\[
S_{Fl} = \delta_{Fl} + \frac{1}{i\hbar} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} dt e^{i\omega_F t} \left\{ \sum_{m} W_{Fm} W_{ml} \frac{1}{(E_i - E_m + i\epsilon\hbar)} \right. \\
+ \sum_{m} \sum_{n} \frac{W_{Fm} W_{mn} W_{nl}}{(E_i - E_m + i\epsilon\hbar)(E_i - E_n + i\epsilon\hbar)} + \ldots \right\}
\]

We now define the transition matrix \(T_{Fl}\) as

\[
T_{Fl} \equiv W_{Fl} + \lim_{\epsilon \to 0^+} \left\{ \sum_{m} \frac{W_{Fm} W_{ml}}{(E_i - E_m + i\epsilon\hbar)} + \sum_{m} \sum_{n} \frac{W_{Fm} W_{mn} W_{nl}}{(E_i - E_m + i\epsilon\hbar)(E_i - E_n + i\epsilon\hbar)} + \ldots \right\},
\]

where we have changed \(\epsilon\hbar \to \epsilon\) since we are taking the limit \(\epsilon \to 0^+\) anyway. The pattern is clear, at each order in perturbation theory, we get one more factor of \(W\) in the numerator, summed over all possible intermediate states, while in the denominator we always get factors of the initial energy minus any intermediate energy, together with an \(i\epsilon\) term that will bear some discussion later. Then our scattering matrix will be given by

\[
S_{Fl} = \delta_{Fl} + \frac{1}{i\hbar} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} T_{Fl} dt e^{i\omega_F t} = \delta_{Fl} + \frac{1}{i\hbar} T_{Fl} \frac{2\sin(\frac{1}{2}\omega_{Fl}T)}{\omega_{Fl}}.
\]

If the final state is different from our initial state, we ignore the first term and find

\[
P(I \to F) = \frac{4}{\hbar^2} |T_{Fl}|^2 \sin^2\left(\frac{1}{2}\omega_{Fl}T\right).
\]

This expression was encountered around (15.20), where we argued that in the limit of large \(T\), it becomes a delta function. Using our results there, we find
We again let the rate be the probability divided by time, and conventionally place one factor of \( \hbar \) inside the delta-function, to obtain Fermi’s Golden Rule

\[
\Gamma (I \rightarrow F) = 2\pi \hbar^{-1} |T_{FI}|^2 \delta (E_F - E_I).
\]  

(15.35)

This, together with (15.34), allows us to calculate transition rates to arbitrary order.

To understand how to apply Fermi’s Golden rule in a simple situation, consider a plane wave scattering off of a weak potential. We will let the unperturbed Hamiltonian be the free particle Hamiltonian, \( H_0 = P^2 / 2m \), and the potential \( V \) be the perturbation. Then our eigenstates of \( H_0 \) will be plane waves, and to leading order, \( T_{FI} \) will be

\[
T_{FI} = V_{FI} = \langle k_F | V(R) | k_I \rangle = \int \frac{d^3r}{(2\pi)^3} e^{-ik_F \cdot r} V(r) e^{ik_I \cdot r}.
\]

The rate (15.35) is then given by

\[
\Gamma (I \rightarrow F) = \frac{1}{(2\pi)^5 \hbar} \left| \int d^3r e^{-ik_F \cdot r} V(r) \right|^2 \delta \left[ \frac{\hbar^2}{2m} \left( k_F^2 - k_I^2 \right) \right].
\]

There are two problems with this formula: first, neither our initial nor our final state plane waves are normalized properly, and second, the delta function is a bit difficult to interpret. We can get rid of one and one-half of these problems by “summing” over the final state momenta, which then becomes an integral.

\[
\Gamma (I \rightarrow F) = \frac{1}{(2\pi)^5 \hbar} \int d^3k_F \left| \int d^3r e^{-ik_F \cdot r} V(r) \right|^2 \delta \left[ \frac{\hbar^2}{2m} \left( k_F^2 - k_I^2 \right) \right]
\]

\[
= \frac{1}{(2\pi)^5 \hbar} \int_0^\infty k_F^2 \hbar \delta \left[ \frac{\hbar^2}{2m} \left( k_F^2 - k_I^2 \right) \right] \int d\Omega \left| \int d^3r e^{-ik_F \cdot r} V(r) \right|^2,
\]

\[
\frac{d\Gamma}{d\Omega} = \frac{mk}{(2\pi)^3 \hbar} \left| \int d^3r e^{-ik_F \cdot r} V(r) \right|^2.
\]  

(15.36)

The one remaining difficulty is that the incident plane wave cannot be normalized. In a manner similar to last chapter, we note that the incident wave has a probability density \( |\psi|^2 = (2\pi)^3 \) and is moving with classical velocity \( \hbar k/m \), implying a flux of

\[ \Phi = (\hbar k/m)(2\pi)^{-3}. \]

This allows us to convert (15.36) into a differential cross-section.

\[
\frac{d\sigma}{d\Omega} = \frac{1}{\Phi} \frac{d\Gamma}{d\Omega} = \frac{m^2}{4\pi^2 \hbar^4} \left| \int d^3r e^{-ik_F \cdot r} V(r) \right|^2.
\]

This is none other than the first Born approximation, eq. (14.19).
Problems for Chapter 15

1. An isolated tritium (hydrogen) atom $^3$H has its electron in the ground state when it suddenly radioactively decays to $^3$He, (helium) but the nucleus stays in the same place (no recoil). What is the probability that the atoms remains in the ground state? What is the probability that it goes into each of the $n = 2$ states $|2lm\rangle$?

2. A neutral boron atom has a total angular momentum $l = 1$ and spin $s = \frac{1}{2}$. In the absence of a magnetic field, the lowest energy states might be listed as $|l, s, j, m_j\rangle = |1, \frac{1}{2}, j, m_j\rangle$, with the $j = \frac{1}{2}$ state having higher energy. The atom is placed in a region of space where a magnetic field is being turned on in the $+z$ direction. At first, the spin-orbit coupling dominates, but at late times the magnetic interactions dominate.

(a) Which of the nine operators $L$, $S$ and $J$ will commute with the Hamiltonian at all times? Note that the state must remain an eigenstate of this operator at all times.

(b) At strong magnetic fields, the states are dominated by the magnetic field. The eigenstates are approximately $|l, s, m_l, m_s\rangle = |1, \frac{1}{2}, m_l, m_s\rangle$. For each possible value of $m_l = m_s + m_s$, deduce which state has the lower energy. Atoms in strong magnetic fields are discussed in chapter 9, section E.

(c) If we start with a particular value of $|l, s, j, m_j\rangle$ (six cases), calculate which states $|l, s, m_l, m_s\rangle$ it might evolve into, assuming the magnetic field increases (i) adiabatically (slowly) or (ii) suddenly. When relevant, give the corresponding probabilities. The relevant Clebsch-Gordan coefficients are given in eq. (8.18).

3. A particle of mass $m$ is initially in the ground state of a harmonic oscillator with frequency $\omega$. At time $t = 0$, a perturbation is suddenly turned on of the form $W(t) = AXe^{-\alpha t}$. At late times ($t \to \infty$), the quantum state is measured again.

(a) Calculate, to second order in $A$, the amplitude $S_{n0}$ that it ends up in the state $|n\rangle$, for all $n$ (most of them will be zero).

(b) Calculate, to at least second order, the probability that it ends up in the state $|n\rangle$. Check that the sum of the probabilities is 1, to second order in $A$.

4. A particle of mass $m$ is in the ground state $|1\rangle$ of an infinite square well with allowed region $0 < X < a$. To this potential is added a harmonic perturbation $W(t) = AX \cos(\omega t)$, where $A$ is small.

(a) Calculate the transition rate $\Gamma(1 \to n)$ for a transition to another level. Don’t let the presence of a delta function bother you. What angular frequency $\omega$ is necessary to make the transition occur to level $n = 2$?

(b) Now, instead of keeping a constant frequency, the frequency is tuned continuously, so that at $t = 0$ the frequency is 0, and it rises linearly so that at $t = T$...
it has increased to the value \( \omega(T) = 2\pi^2 \hbar / ma^2 \). The tuning is so slow that at any given time, we may treat it as a harmonic source. Argue that only the \( n = 2 \) state can become populated (to leading order in \( A \)). Calculate the probability of a transition using \( P(1 \rightarrow 2) = \int_0^\Gamma T(1 \rightarrow 2)dt \).

5. A hydrogen atom is in the 1s ground state while being bathed in light of sufficient frequency to excite it to the \( n = 2 \) states. The light is traveling in the \( +z \) direction and has circular polarization, \( \epsilon = \frac{1}{\sqrt{2}} (\hat{x} + i\hat{y}) \).

(a) Calculate all relevant dipole moments \( r_{fi} \) for final states \( |2lm\rangle \).

(b) Find a formula for the rate at which the atom makes this transition.

(c) What is the wavelength required for this transition? Assume at this wavelength the power is \( \mathcal{I}(\lambda) = 100 \text{ W/m}^2/\text{nm} \). Find the rate at which the atom converts.

(Note the footnote on p. 280)

6. A hydrogen atom is in interstellar space in the 1s state, but not in the true ground state \( (F = 0) \), but rather in the hyperfine excited state \( (F=1) \), specifically in the state \( |\phi_i\rangle = |n, l, j, F, m_F\rangle = |1, 0, \frac{1}{2}, 1, 0\rangle \). It is going to transition to the true ground state \( |\phi_f\rangle = |n, l, j, F, m_F\rangle = |1, 0, \frac{1}{2}, 0, 0\rangle \) via a magnetic dipole interaction.

(a) Write out the initial and final states in terms of the explicit spin states of the electron and proton \( |\pm, \pm\rangle \). Find all non-zero components of the matrix \( \langle \phi_f | S | \phi_i \rangle \), where \( S \) is the electron spin operator.

(b) Show that the rate for this transition for a wave going in a specific direction with a definite polarization is given by \( \Gamma = 4\pi^2 m^2 \omega^2 c^6 \mathcal{I} \left| (\mathbf{k} \times \epsilon) \cdot \mathbf{S}_{fi} \right|^2 \delta \left( E_f - E_i + \hbar \omega \right) \).

(c) Show that for a wave going in a random direction with random polarization, this simplifies to \( \Gamma (I \rightarrow F) = \frac{\pi}{2} \sqrt{\frac{a}{m}} \hbar c^4 \mathcal{I} \left| S_{fi} \right|^2 \delta \left( E_f - E_i + \hbar \omega \right) \).

(d) For low frequencies, the cosmic microwave background intensity is \( \mathcal{I}(\omega) = k_B T \omega^2 / \pi^2 c^2 \) where \( k_B \) is Boltzman’s constant and \( T \) is the temperature. Integrate the flipping rate over frequency. Find the mean time \( \Gamma^{-1} \) for a hydrogen atom to reverse itself in a background temperature \( T = 2.73 \text{ K} \) for \( \omega_{fi} = 2\pi (1.420 \text{ GHz}) \).

7. A spin ½ particle of mass \( m \) lies in a one-dimensional spin-dependant potential \( H_0 = P^2 / 2m + \frac{1}{2} m \omega_0^2 X^2 \). The potential only affects particles in a spin-up state.

(a) Find the discrete energy eigenstates for spin-up \( |+, \uparrow\rangle \) and the continuum energy eigenstates for spin-down \( |-, \downarrow\rangle \). Also, identify their energies.

(b) At \( t = 0 \), a spin-dependant perturbation of the form \( V = \hbar \gamma \sigma_z \), where \( \sigma_z \) is a Pauli matrix, is turned on. Calculate the rate \( \Gamma \) at which the spin-up ground state “decays” to a continuum state.
XVI. The Dirac Equation

We turn our attention to a discussion of a relativistic theory of the electron, the Dirac Equation. Although the approach we use is now considered obsolete, it does provide important insights into relativistic quantum mechanics, and ultimately it was an important step on the way to a modern theory of particle physics.

As a first step, consider the procedure we used to produce the free particle Hamiltonian. We started with the non-relativistic formula for energy, namely, 

\[ E = \frac{p^2}{2m} \]

multiplied by a wave function \( \Psi(\mathbf{r},t) \) and then made the substitutions 

\[ E \rightarrow i\hbar \frac{\partial}{\partial t} \]

and promoted \( p \) to an operator, \( \mathbf{p} \rightarrow \mathbf{P} \) to produce the free particle Schrödinger equation,

\[ i\hbar \frac{\partial}{\partial t} \Psi = \frac{1}{2m} \mathbf{p}^2 \Psi . \]

Now, the corresponding relativistic equation for energy is 

\[ E^2 = c^2 \mathbf{p}^2 + m^2 c^4 . \]

It is then easy to derive a corresponding Schrödinger-like equation by the same prescription:

\[ \left( i\hbar \frac{\partial}{\partial t} \right)^2 \Psi = c^2 \mathbf{p}^2 \Psi + m^2 c^4 \Psi , \] (16.1)

the Klein-Gordon equation. The problem with (16.1) is that it is second order in time. As such, to predict the wave function \( \Psi(\mathbf{r},t) \) at arbitrary time, we would need to know not just the initial wave-function \( \Psi(\mathbf{r},0) \), but also its first derivative \( \Psi'(\mathbf{r},0) \), contradicting the first postulate of quantum mechanics. We need to find a way to convert (16.1) into a first order differential equation for the wave function.

A. The Dirac Equation

As inspiration for finding the correct equation, let’s back up a bit, and look at our formula for the Hamiltonian in the presence of electromagnetic fields, eq. (9.20), but we will approximate \( g = 2 \). The Schrödinger equation is

\[ i\hbar \frac{\partial}{\partial t} \Psi = \frac{\pi^2}{2m} \Psi - eU \Psi + \frac{e\hbar}{2m} \mathbf{B} \cdot \mathbf{\sigma} \Psi . \] (16.2)

where we have used the explicit form of the spin operator \( \mathbf{S} = \frac{1}{2} \hbar \mathbf{\sigma} \). We wish to rewrite this in a simpler form. We start with identity (7.18) for the Pauli matrices. With the help of (9.16c), it is then not hard to show

\[ (\mathbf{\sigma} \cdot \mathbf{\pi})^2 = \pi^2 + i\sigma_x \left[ \pi_y, \pi_z \right] + i\sigma_y \left[ \pi_z, \pi_x \right] + i\sigma_z \left[ \pi_x, \pi_y \right] = \pi^2 + e\hbar \mathbf{\sigma} \cdot \mathbf{B} . \]

This allows us to write (16.2) in the alternate form

\[ i\hbar \frac{\partial}{\partial t} \Psi = \frac{1}{2m} (\mathbf{\sigma} \cdot \mathbf{\pi})^2 \Psi - eU \Psi . \] (16.3)
Let us, at least temporarily, drop the electromagnetic interactions, in which case (16.3) takes the simpler form

\[ i\hbar \frac{\partial}{\partial t} \Psi = \frac{1}{2m} (\sigma \cdot P)^2 \Psi. \]  

(16.4)

This is, of course, identical with the standard form \( H = P^2/2m \), but the presence of spin will ultimately help us develop the Dirac equation.

Inspired by (16.4), let us consider rewriting (16.1) in the form

\[ \left( i\hbar \frac{\partial}{\partial t} + c\sigma \cdot P \right) \phi = c^2 (\sigma \cdot P)^2 \phi + m^2 c^4 \phi, \]

where we have written the wave function as \( \phi \) for reasons that will soon become apparent. Note that \( \phi \) has two components, as it must for this expression to have any meaning. We now shift the momentum term to the left hand side, and then factor the resulting expression as a difference of squares:

\[ \left( i\hbar \frac{\partial}{\partial t} + c\sigma \cdot P \right) \left( i\hbar \frac{\partial}{\partial t} - c\sigma \cdot P \right) \phi = m^2 c^4 \phi. \]  

(16.5)

We now define the quantity \( \phi^\dagger \) by the relationship

\[ \left( i\hbar \frac{\partial}{\partial t} - c\sigma \cdot P \right) \phi^\dagger = -mc^2 \phi. \]  

(16.6)

Substituting this into (16.5), we see that

\[ \left( i\hbar \frac{\partial}{\partial t} + c\sigma \cdot P \right) \phi^\dagger = -mc^2 \phi. \]  

(16.7)

Equations (16.6) and (16.7) can be written together as a single matrix equation

\[ \begin{bmatrix} i\hbar \frac{\partial}{\partial t} - c \sigma \cdot P & 0 \\ 0 & -\sigma \cdot P \end{bmatrix} \begin{bmatrix} \phi \\ \phi^\dagger \end{bmatrix} = \begin{bmatrix} 0 & -mc^2 \\ -mc^2 & 0 \end{bmatrix} \begin{bmatrix} \phi \\ \phi^\dagger \end{bmatrix}. \]  

(16.8)

Now, up to this point, we have treated \( \phi \) as our wave function, and \( \phi^\dagger \) as a halfway point defined by (16.6). Our result is the coupled first-order differential equations (16.8). Our goal is to find a first-order Schrödinger equation for our wave function. The key is to stop thinking of \( \phi^\dagger \) as being defined by (16.6), but rather to think of it as an independent piece of the wave function, so that the whole wave function is defined by \( \phi \) and \( \phi^\dagger \) together, which we write as

\[ \Psi \equiv \begin{bmatrix} \phi \\ \phi^\dagger \end{bmatrix}. \]

Equation (16.8) will then become our Schrödinger equation, now first order in time,
\[ i \hbar \frac{\partial}{\partial t} \Psi = c \mathbf{a} \cdot \mathbf{P} \Psi + mc^2 \beta \Psi, \]  

(16.9)

where

\[ \mathbf{a} \equiv \begin{pmatrix} \sigma & 0 \\ 0 & -\sigma \end{pmatrix} \text{ and } \beta \equiv \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}. \]  

(16.10)

Equation (16.9) is the (free) Dirac equation, where \( \mathbf{a} \) and \( \beta \) are the Dirac matrices, \( 4 \times 4 \) matrices defined by (16.10). The Hamiltonian is

\[ H = c \mathbf{a} \cdot \mathbf{P} + mc^2 \beta. \]

It will prove useful to have handy the following identities for products of the Dirac matrices:

\[ \alpha_i \alpha_j + \alpha_j \alpha_i = 2 \delta_{ij}, \quad \alpha_i \beta + \beta \alpha_i = 0, \quad \beta^2 = 1. \]  

(16.11)

In other words, all four matrices have a square of one, and they all anti-commute with each other. They are also all Hermitian. There are other ways of writing them. Let \( U \) be any \( 4 \times 4 \) unitary matrix, and define

\[ \Psi' \equiv U \Psi, \quad \mathbf{a}' \equiv U \mathbf{a} U^\dagger, \quad \text{and} \quad \beta' \equiv U \beta U^\dagger. \]

Then by multiplying (16.9) on the left and inserting factors of \( 1 = U^\dagger U \) appropriately, it is easy to show it is equivalent to

\[ i \hbar \frac{\partial}{\partial t} \Psi' = c \mathbf{a}' \cdot \mathbf{P} \Psi' + mc^2 \beta \Psi'. \]

This is just obviously the Dirac equation, but with \( \mathbf{a} \) and \( \beta \) redefined. The new matrices \( \mathbf{a}' \) and \( \beta' \) will still be Hermitian and still satisfy the relations (16.11). In particular, for

\[ U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \]

the Dirac matrices will take the form

\[ \mathbf{a}' = \begin{pmatrix} 0 & \sigma \\ \sigma & 0 \end{pmatrix} \text{ and } \beta' = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \]  

(16.12)

The version (16.12) (the Dirac representation) is preferred by those who do non-relativistic physics, and (16.10) (the Chiral representation) is better for ultrarelativistic particles, when the mass becomes negligible. We will stick with the Chiral representation because it is slightly simpler for our purposes.
B. Solving the Free Dirac Equation

We would like to solve the time-independent Dirac equation, which is simply

\[(c\alpha \cdot \mathbf{P} + mc^2 \beta)\psi = E\psi.\]

We expect to find plane wave solutions of the form \(\psi = u(k)e^{ikr}\), where \(u\) is independent of \(r\). Substituting this expression in, we find

\[\left(c\alpha \cdot \mathbf{P} + mc^2 \beta\right)u(k)e^{ikr} = \begin{pmatrix} c\beta \sigma \cdot \mathbf{k} & -mc^2 \\ -mc^2 & -c\beta \sigma \cdot \mathbf{k} \end{pmatrix}u(k)e^{ikr} = Eu(k)e^{ikr}. \quad (16.13)\]

We are thus trying to solve this eigenvalue equation. As a first step, we consider the \(2 \times 2\) matrix \(\sigma \cdot \mathbf{k}\), which when squared yields \((\sigma \cdot \mathbf{k})^2 = k^2\), and therefore has eigenvalues \(\pm k\). Let us define the two eigenstates of this matrix \(\phi_{\pm}\), so that \((\sigma \cdot \mathbf{k})\phi_{\pm} = \pm k\phi_{\pm}\). In explicit form, if you need them, one can show that if \(k\) is in the spherical direction denoted by \(\theta\) and \(\phi\), we have

\[\phi_+ = \begin{pmatrix} \cos \left(\frac{\phi}{2}\right) \\ \sin \left(\frac{\phi}{2}\right)e^{i\theta} \end{pmatrix} \quad \text{and} \quad \phi_- = \begin{pmatrix} -\sin \left(\frac{\phi}{2}\right) \\ \cos \left(\frac{\phi}{2}\right)e^{i\theta} \end{pmatrix}.\]

Then we will guess that we can find solutions of (16.13) of the form

\[u(k) = \begin{pmatrix} a\phi_+ \\ b\phi_- \end{pmatrix}. \quad (16.14)\]

Substituting this into (16.13), we find

\[E\begin{pmatrix} a\phi_+ \\ b\phi_- \end{pmatrix} = \begin{pmatrix} c\beta \sigma \cdot \mathbf{k} & -mc^2 \\ -mc^2 & -c\beta \sigma \cdot \mathbf{k} \end{pmatrix}\begin{pmatrix} a\phi_+ \\ b\phi_- \end{pmatrix} = \begin{pmatrix} \left(-bmc^2 \pm achk\right)\phi_+ \\ \left(-amc^2 \mp bckh\right)\phi_- \end{pmatrix}.\]

This yields the two equations

\[-\frac{a}{b} = \frac{E \pm ckh}{mc^2} = \frac{mc^2}{E \mp ckh}. \quad (16.15)\]

Cross-multiplying leads to the equation \(E^2 - c^2h^2k^2 = m^2c^4\), exactly what we want. Up to normalization, we can solve (16.15) for \(a\) and \(b\), and then substitute into (16.14), yielding two solutions for each value of \(k\):\(^1\)

\[\psi_{\pm}(r) = u_{\pm}(k)e^{ikr} = \begin{pmatrix} \sqrt{E \pm ckh}\phi_+ \\ -\sqrt{E \mp ckh}\phi_- \end{pmatrix}e^{ikr}, \quad (16.16)\]

---

\(^1\) If you wish to use wave functions of amplitude 1, these should be divided by \(\sqrt{2E}\). For technical reasons this proves to be inconvenient, though it is irrelevant for us.
where \( E = \sqrt{c^2 \hbar^2 k^2 + m^2 c^4} \).

What are these two solutions? Obviously, they have the same momentum \( p = \hbar k \). It turns out that the spin operator in the Dirac equation is given by

\[
S = \frac{1}{2} \hbar \sum, \quad \text{where} \quad \sum = \begin{pmatrix} \sigma & 0 \\ 0 & -\sigma \end{pmatrix}.
\]

(16.17)

If we measure the spin along the direction the particle is traveling we will find

\[
\left( \hat{k} \cdot S \right) \psi_{\pm} = \frac{\pm \hbar}{2} \begin{pmatrix} \hat{k} \cdot \sigma & 0 \\ 0 & -\hat{k} \cdot \sigma \end{pmatrix} \psi_{\pm} = \pm \frac{\hbar}{2} \psi_{\pm}.
\]

Thus it is in an eigenstate of spin in this direction. Of course, linear combinations of the two solutions (16.16) are also allowed. Hence the Dirac equation predicts two positive energy solutions for every value of the wave number \( k \) corresponding to the two spin states.

The eigenvalue equation (16.14) involves a \( 4 \times 4 \) matrix, and hence should have four eigenvectors, not two. This suggests that we have somehow missed two solutions. Indeed, we have, and the problem is they have negative energy. They work out to be

\[
\psi_{\pm}(r) = v_{\pm}(k) e^{-ikr} = \begin{pmatrix} \sqrt{-E \mp \hbar k \phi_+} \\ \sqrt{-E \pm \hbar k \phi_+} \end{pmatrix} e^{-ikr},
\]

where \( E = -\sqrt{c^2 \hbar^2 k^2 + m^2 c^4} \), as can be verified by direct substitution into the time-independent Dirac equation. It has momentum \( -\hbar k \) and spin along the direction \( \hat{k} \) of \( \mp \frac{\hbar}{2} \) (which makes it look like we mislabeled these solutions). But what do these solutions mean?

Dirac came up with the following solution. Dirac reasoned that because these states had negative energy, the ground state would be a state where these particles were present, not just for one momentum \( p \), but for all momenta. He reasoned that since electrons are fermions, they satisfied the Pauli exclusion principle. Hence once you “filled” these states, there was no lower energy state. He suggested that what we normally term the “vacuum” is not truly empty, but rather is filled with electrons sitting in all the negative energy states. We cannot notice this “sea” of states because it is just the normal state of affairs.

\[
\text{Figure 16-1: The Dirac equation has both positive (blue curve) and negative (red) energy solutions. Dirac assumed all the negative energy states were “filled” with an undetectable sea of electrons (black dots). Sufficient energy can move (green arrow) one of these electrons from a negative energy state to a positive energy state, producing a visible hole (open red circle) and an electron (blue dot).}
\]
Does this mean these negative energy states are irrelevant? Not at all. Suppose we applied a perturbation, say an electromagnetic pulse, of sufficient energy to take one of these negative energy states and promote it to a positive energy state. We would instantly “see” an electron that had not been there previously, as illustrated in Fig. 16-1. We would also simultaneously “see” a missing state from the negative energy sea, much as a bubble underwater indicates an absence of water. This “hole,” as he called it, would be perceived as an absence of momentum \(-\hbar \mathbf{k}\), (or in other words, a momentum \(\hbar \mathbf{k}\)), an absence of spin angular momentum \(\hbar \frac{\mathbf{\sigma}}{2}\) in the direction of \(\mathbf{k}\) (spin \(\pm \frac{\hbar}{2}\)), an absence of energy \(-\sqrt{\hbar^2 k^2 + m^2 c^4}\) (or a presence of energy \(E = \sqrt{\hbar^2 k^2 + m^2 c^4}\)) and an absence of charge \(-e\) (or \(q = +e\)). In other words, we would see a particle identical to the electron, with the same mass and spin, except having the opposite charge. Dirac assumed this particle was the proton (and was puzzled why it didn’t have the same mass), but we now recognize it as a separate particle, the positron, also known as the anti-electron. Dirac’s arguments can be generalized to other spin-1/2 particle, and, in fact, we now expect every elementary particle to have an anti-particle, and, indeed, all particles do apparently have anti-particles (though some particles, like the photon, are their own anti-particle).

C. Electromagnetic Interactions and the Hydrogen Atom

We now wish to include electromagnetic interactions in our Hamiltonian. This is an easy matter: we simply change \(P\) to \(\mathbf{\pi} = \mathbf{P} + e \mathbf{A}\) and then add the electrostatic potential energy \(-eU\) to yield

\[
H = c\mathbf{\alpha} \cdot \left[\mathbf{P} + e \mathbf{A}(\mathbf{R}, t)\right] - eU(\mathbf{R}, t) + mc^2 \beta .
\]

Let’s try solving a particularly difficult and interesting problem in this case, namely, hydrogen, with potential \(U = k_e e / |\mathbf{R}|\). Then our Hamiltonian is

\[
H = c\mathbf{\alpha} \cdot \mathbf{P} - \frac{k_e e^2}{|\mathbf{R}|} + mc^2 \beta = c\mathbf{\alpha} \cdot \mathbf{P} - \frac{\alpha \hbar c}{|\mathbf{R}|} + mc^2 \beta ,
\]

where we have used the fine structure constant \(\alpha = k_e e^2 / \hbar c\) (not to be confused with \(\alpha\)) to rewrite the potential term. We will be trying to find eigenstates of this equation, solutions of the time-independent Schrödinger equation

\[
E\psi = \left(c\mathbf{\alpha} \cdot \mathbf{P} - \frac{\alpha \hbar c}{r} + mc^2 \beta\right)\psi ,
\]

\[
\left(E + \frac{\alpha \hbar c}{r}\right)\psi = \left(c\mathbf{\alpha} \cdot \mathbf{P} + mc^2 \beta\right)\psi .
\]

We now let the operator on the right act a second time on both sides to yield

\[
\left(c\mathbf{\alpha} \cdot \mathbf{P} + mc^2 \beta\right)\left(E + \frac{\alpha \hbar c}{r}\right)\psi = \left(c\mathbf{\alpha} \cdot \mathbf{P} + mc^2 \beta\right)^2 \psi .
\]
On the right hand side, we can use the fact that all four matrices $\alpha$ and $\beta$ anti-commute with each other, so the cross terms all vanish, and their squares are all the identity matrix. On the left side, we note that if the factor of $\left(c \alpha \cdot P + mc^2 \beta\right)$ were allowed to act on the wave function, we would only get another factor of $(E + \alpha \hbar c / r)$. However, the derivative $P$ can also act on the potential $\alpha \hbar c / r$, yielding another term. The result is

$$\left(E + \frac{\alpha \hbar c}{r}\right)^2 \psi + i\alpha \hbar^2 c^2 \left(\frac{1}{r^2} (\alpha \cdot \hat{r})\psi = \left(c^2 P^2 + m^2 c^4\right)\psi. \right.$$  

We now notice something interesting: The only matrix in this whole equation is $\alpha$, which is block-diagonal. Therefore, if we write our wave function $\psi$ in terms of two pairs of components, as we originally defined it, we would find that the equations for $\phi$ and $\tilde{\phi}$ completely decouple. Hence we can focus on one at a time. It is not hard to show that the two yield identical equations, so we will focus exclusively on $\phi$, for which our equation is

$$\left(E^2 + \frac{2\alpha \hbar c E}{r} + \frac{\hbar^2 c^2}{r^2} \left[\mathbf{\alpha}^2 + i\alpha (\mathbf{\sigma} \cdot \hat{r})\right]\right)\phi = \left(c^2 P^2 + m^2 c^4\right)\phi. \quad (16.18)$$

Our problem is spherically symmetric, so we should probably switch to spherical coordinates. With the help of (A.22d) and (7.23), it is not hard to show that in spherical coordinates,

$$P^2 \phi = -\frac{\hbar^2}{r} \frac{\partial^2}{\partial r^2} (r \phi) + \frac{L^2}{r^2} \phi.$$

Substituting this into (16.18) and rearranging a bit, we have

$$(E^2 - m^2 c^4) \phi = -\frac{\hbar^2 c^2}{r} \frac{\partial}{\partial r} (r \phi) - \frac{2\alpha \hbar c E}{r} \phi + \frac{\hbar^2 c^2}{r^2} \left(L^2 - \alpha^2 - i\alpha \mathbf{\sigma} \cdot \hat{r}\right) \phi. \quad (16.19)$$

We note that all the angular and spin-dependence of this expression is contained in the last factor, which we define as $A$, so that

$$A \equiv \frac{L^2}{\hbar^2} - \alpha^2 - i\alpha \mathbf{\sigma} \cdot \hat{r}. \quad$$

It makes sense to try to factor $\phi$ into an angular and radial part, just like we did for non-relativistic Hydrogen. We let

$$\phi(r) = R(r) \Upsilon(\theta, \phi),$$

where $R(r)$ is an ordinary function, but $\Upsilon(\theta, \phi)$ has two components, so it has both the angular and spin dependence. Substituting this form into (16.19), we have

---

1 This is why we chose the chiral representation, the decoupling is far less obvious in the Dirac representation.

© 2011, Eric D. Carlson 295
\[
\left(E^2 - m^2c^4\right)R(r)Y(\theta, \phi) = -\frac{\hbar^2c^2}{r} \frac{\partial^2}{\partial r^2} \left[ rR(r) \right] Y(\theta, \phi) - \frac{2\alpha \hbar cE}{r} R(r) Y(\theta, \phi) + \frac{\hbar^2c^2}{r^2} R(r) AY(\theta, \phi).
\]

We’d like our angular functions $Y(\theta, \phi)$ to be eigenstates of $A$. This is the equivalent of finding the spherical harmonics, with the added complication of spin, though our primary interest here is only finding the eigenvalues of $A$.

To find the eigenvalues of $A$, consider first that the original Hamiltonian must be rotationally invariant, and hence commutes with all components of $\mathbf{J}$, and specifically with $J^2$ and $J_z$. It therefore makes sense to try to work with eigenstates of these two operators. We already know that the angular and spin states of a spherically symmetric problem can be written in the basis $|j, m_j\rangle$ (we have suppressed the spin label $s = \frac{1}{2}$). Since $J^2$ and $J_z$ commute with the Hamiltonian, it is hardly surprising that they commute with $A$, and therefore $A$ can only connect states with the same values of $j$ and $m_j$. Thus the only non-zero matrix elements of $A$ in this basis are

\[
\langle j, m_j | A | j', m_{j'} \rangle = \langle l, j, m_j \left| \left( \frac{L^2}{\hbar^2} - \alpha^2 - i\alpha \sigma \cdot \mathbf{r} \right) \right| j', j, m_{j'} \rangle. \tag{16.21}
\]

Since $j = l \pm \frac{1}{2}$, there are only two possible values for $l$, and hence finding the eigenstates and eigenvalues of $A$ is reduced to the problem of diagonalizing a $2 \times 2$ matrix.

The first two terms in (16.21) are obviously diagonal in this basis, so we see that

\[
\langle l, j, m_j \left| \left( \frac{L^2}{\hbar^2} - \alpha^2 \right) \right| l', j, m_{j'} \rangle = \delta_{jj'}(l^2 + l - \alpha^2).
\]

The last term is more problematic. If you rewrite $\sigma \cdot \mathbf{r} = \sigma \cdot \mathbf{r}/r$, and then recall that the operator $R$ (which becomes $\mathbf{r}$) connects only states which differ by one value of $l$, you will realize that the on-diagonal components of this term must vanish, so

\[
\langle l, j, m_j \left| \sigma \cdot \mathbf{r} \right| j, m_j \rangle = 0.
\]

On the other hand, we know that $(\sigma \cdot \mathbf{r})^2 = \mathbf{1}$, the identity matrix. It follows that

\[
1 = \langle l, j, m_j \left| (\sigma \cdot \mathbf{r})(\sigma \cdot \mathbf{r}) \right| l, j, m_j \rangle = \sum_{l' = j \pm \frac{1}{2}} \langle l, j, m_j \left| \sigma \cdot \mathbf{r} \right| l', j, m_{j'} \rangle \langle l', j, m_{j'} \left| \sigma \cdot \mathbf{r} \right| l, j, m_j \rangle
\]

\[
= \sum_{l' = j \pm \frac{1}{2}} \left| \langle l, j, m_j \left| \sigma \cdot \mathbf{r} \right| l', j, m_{j'} \rangle \right|^2,
\]

where we have inserted a complete set of intermediate states. However, there are only two terms in the sum, and one of them is zero, so the other one must be a number of magnitude one.\footnote{Actually, it is one, but we won’t need this fact.} In other words,

\[
\langle j - \frac{1}{2}, j, m_j \left| \sigma \cdot \mathbf{r} \right| j + \frac{1}{2}, j, m_{j'} \rangle = \langle j + \frac{1}{2}, j, m_j \left| \sigma \cdot \mathbf{r} \right| j - \frac{1}{2}, j, m_{j'} \rangle^* = \eta, \text{ where } |\eta|^2 = 1.
\]
We now know every matrix element of \( A \); in the basis where we put \( l = j - \frac{1}{2} \) first and \( l = j + \frac{1}{2} \) second, \( A \) takes the form

\[
A = \begin{pmatrix}
(j - \frac{1}{2})(j + \frac{1}{2}) - \alpha^2 & -i\alpha \eta \\
-i\alpha \eta^* & (j + \frac{1}{2})(j + \frac{3}{2}) - \alpha^2
\end{pmatrix}.
\]

It is a straightforward matter to find the eigenvalues of this matrix, which turn out to be

\[
(j + \frac{1}{2})^2 - \alpha^2 \pm \sqrt{(j + \frac{1}{2})^2 - \alpha^2}.
\]

It is very helpful to define the quantity

\[\lambda_j \equiv \sqrt{(j + \frac{1}{2})^2 - \alpha^2}.\]  \hspace{1cm} (16.22)

Then our eigenvalues are \( \lambda_j^2 \pm \lambda_j \). Hence for every value of \( j \) and \( m_j \) we find two solutions \( \Psi_{j,m_j}^m(\theta, \phi) \) satisfying

\[
AY_{j,m_j}^m(\theta, \phi) = \left( \frac{L^2}{\hbar^2} - \alpha^2 - i\alpha \mathbf{\hat{\sigma}} \cdot \mathbf{\hat{r}} \right)Y_{j,m_j}^m(\theta, \phi) = (\lambda_j^2 \pm \lambda_j)Y_{j,m_j}^m(\theta, \phi).
\]

Substituting this into (16.19), we see that the angular portion cancels out, and we are left with two very similar radial equations,

\[
\left( E^2 - m^2c^4 \right)R(r) = -\frac{\hbar^2 c^2}{r} \frac{d^2}{dr^2} \left[ rR(r) \right] - \frac{2\alpha \hbar cE}{r} R(r) + \frac{\hbar^2 c^2 \left( \lambda_j^2 \pm \lambda_j \right)}{r^2} R(r).
\]

Solving this will prove to be quick, because of its remarkable similarity to the standard hydrogen atom. To make the comparison as obvious as possible, first divide by \( 2E \).

\[
\frac{E^2 - m^2c^4}{2E} R(r) = -\frac{\hbar^2 c^2}{2Er} \frac{d^2}{dr^2} \left[ rR(r) \right] - \frac{\alpha \hbar c}{r} R(r) + \frac{\hbar^2 c^2 \lambda_j^2 \pm \lambda_j}{2Er} R(r). \hspace{1cm} (16.23)
\]

Compare this to the radial equation for hydrogen, eq. (7.41), which we reproduce here, rewritten slightly:

\[
ER(r) = -\frac{\hbar^2}{2\mu r} \frac{d^2}{dr^2} \left[ rR(r) \right] - \frac{\alpha \hbar c}{r} R(r) + \frac{\hbar^2 c^2 l^2 + l}{2\mu r^2} R(r). \hspace{1cm} (16.24)
\]

Comparing (16.23) and (16.24), we see that the latter can be transformed into the former if we identify

\[
E \rightarrow \frac{E^2 - m^2c^4}{2E}, \hspace{0.5cm} \mu \rightarrow \frac{E}{c^2}, \hspace{0.5cm} l^2 + l \rightarrow \lambda_j^2 \pm \lambda_j.
\]

The last transformation must be treated as two separate cases, but it is pretty easy to see that the two cases are

\[
l \rightarrow \lambda_j \quad \text{or} \quad l \rightarrow \lambda_j - 1.
\]
The point is that we can now instantly use every equation we derived from hydrogen, we just need to make the corresponding transformations. For example, our formula for the radial wave function will take the form (similar to (7.50a))

\[ R_{nl}(r) = e^{-r/a} \sum_{p=-\lambda_j}^{\lambda_j-1} f_{p} r^{p} \quad \text{or} \quad R_{nl}(r) = e^{-r/a} \sum_{p=-\lambda_j}^{\lambda_j-1} f_{p} r^{p}, \quad (16.25) \]

where \( a \) is given by (similar to (7.42)):

\[ \frac{E^2 - m^2 c^4}{2E} = -\frac{\hbar^2 c^2}{2Ea^2}, \quad \text{or} \quad a = \frac{\hbar c}{\sqrt{m^2 c^4 - E^2}}. \]

Equations (16.25) require a bit of discussion. Note that the lower limit on the sum is not an integer, and therefore the powers \( p \) of \( r \) are non-integer as well. The upper limit \( \nu \) will therefore also not be an integer, but it will differ from the lower limit by an integer; that is \( \nu = \lambda_j + k \), where \( k \) is a positive integer in the first case, and a non-negative integer in the second case.

The energies are given by an equation analogous to (7.51)

\[ \frac{E^2 - m^2 c^4}{2E} = -\frac{E_k^2 e^4}{2\hbar^2 c^2 \nu^2} = -\frac{E_k^2}{2(\lambda_j + k)^2}, \]

\[ E = mc^2 \left[ 1 + \alpha^2 \left( \lambda_j + k \right)^2 \right]^{\frac{1}{2}}. \quad (16.26) \]

This, together with (16.22), gives us the energies. It should be noted that in (16.26), \( k \) can be zero or a positive integer. When it is zero, we only have one case, but when \( k \) is positive, we will have two distinct cases, corresponding to the two angular functions \( \Upsilon_{j,\pm}^{m_j} (\theta, \phi) \).

If we are dealing with a hydrogen-like atom, rather than true hydrogen, the only change is to substitute \( \alpha \to Z \alpha \). So the more general version of (16.22) and (16.26) is

\[ \lambda_j = \sqrt{\left( j + \frac{1}{2} \right)^2 - Z^2 \alpha^2}, \quad (16.27a) \]

\[ E = mc^2 \left[ 1 + Z^2 \alpha^2 \left( \lambda_j + k \right)^2 \right]^{\frac{1}{2}}. \quad (16.27b) \]

It is helpful to expand these in powers of the fine structure constant to understand what the various choices correspond to. We find, to fourth order in \( \alpha \),

\[ E = mc^2 - \frac{mc^2 \alpha^2 Z^2}{2\left( j + \frac{1}{2} + k \right)^2} + \frac{mc^2 \alpha^4 Z^4}{\left( j + \frac{1}{2} + k \right)^4} \left( \frac{3}{8} - \frac{j + \frac{1}{2} + k}{2j + 1} \right) + \cdots. \]

Looking at these terms order by order, it is easy to identify them. The first term is the rest energy of an electron. The second is the leading order binding energy. Indeed, since \( j \) is half-integer, the combination \( j + \frac{1}{2} + k \) is an integer, which we name \( n \), and the last term is simply the first relativistic correction. We therefore have
\[ E = mc^2 - \frac{mc^2\alpha^2 Z^2}{2n^2} + \frac{mc^2\alpha^4 Z^4}{n^4} \left( \frac{3}{8} - \frac{n}{2j+1} \right) + \cdots, \quad (16.28) \]

with now the restriction that \( n \geq j + \frac{1}{2} \). Indeed, the same restriction occurs for non-relativistic hydrogen. When \( n > j + \frac{1}{2} \), there are two different possibilities with exactly the same energy (corresponding to \( l = j \pm \frac{1}{2} \)), while for \( n = j + \frac{1}{2} \), there is only one (since \( l = j + \frac{1}{2} \) is no longer allowed). The final expression in (16.28) is simply the first relativistic correction. The \( j \)-dependence is reflecting the conventional spin-orbit coupling, though other effects are included as well. Note, as expected, the corrections grow as powers of \( \alpha Z \), which means that larger \( Z \)'s will have much larger effects. In most practical situations, such atoms will in fact be neutral (or nearly neutral), and hence they won’t be hydrogen-like at all.

It is interesting to note that the energy (16.27b) depends only on \( j \) and \( k \), or in conventional notation, \( j \) and \( n \). This means that, for example, the \( 2s_{1/2} \) and \( 2p_{1/2} \) levels are still exactly degenerate, if we ignore hyperfine interactions. This is surprising, because there is no evident symmetry protecting them. There are interactions that split this apparent degeneracy, but they come about by interactions with the quantized electromagnetic field, the subject of the next chapter.

Despite the successes of the Dirac equation, the awkward way it deals with antiparticles and other problems ultimately led to its abandonment as an approach to developing a fully relativistic theory of particles. Due to its complications, we will henceforth ignore it, and use non-relativistic electrons whenever possible. But electromagnetic waves, which always move at the speed of light, must be quantized in a fully relativistic formalism.

Problems for Chapter 16

1. For the free Dirac equation, define the probability density as \( \rho = \psi^* \psi \) and the probability current as \( \mathbf{J} = c \psi^* \alpha \psi \). Show that probability is locally conserved, i.e., that \( \partial \rho / \partial t + \nabla \cdot \mathbf{J} = 0 \). Will this equation still be valid in the presence of electromagnetic effects coming from \( \mathbf{A}(\mathbf{r}, t) \) and \( U(\mathbf{r}, t) \)?

2. In the class notes, we claimed that the spin was defined by \( \mathbf{S} = \frac{1}{2} \hbar \mathbf{\Sigma} \), eq. (16.17). To make sure this is plausible:
   a) Demonstrate that \( \mathbf{S} \) satisfies proper commutations relations \( [S_i, S_j] = \sum_k i \hbar \epsilon_{ijk} S_k \).
   b) Work out the six commutators \( [\mathbf{L}, H] \) and \( [\mathbf{S}, H] \) for the free Dirac Hamiltonian.
   c) Show that \( [\mathbf{J}, H] = 0 \), where \( \mathbf{J} = \mathbf{L} + \mathbf{S} \).
XVII. Quantization of Electromagnetic Fields

When Max Planck first “discovered” quantum mechanics, it was by conjecturing that electromagnetic waves come in quantized packets of energy \( E = h\omega \). And yet here we are, 300 pages of notes later, and we have yet to quantize the electromagnetic field. This is because electromagnetism presents some special challenges. Fortunately, these challenges have been generally overcome, and we have a fully relativistic theory of the electromagnetic field. Ultimately, we will want to interact electromagnetic waves with electrons or other particles, but for now, let us try to quantize the pure EM field.

A. Gauge Choice and Energy

One complication that is especially annoying about electromagnetism is that it will turn out to be necessary to work with the vector and scalar potential, \( A(\mathbf{r}, t) \) and \( U(\mathbf{r}, t) \), not the electric and magnetic fields \( \mathbf{E} \) and \( \mathbf{B} \) directly. As we discussed in chapter 9, we must make a gauge choice; that is, we have to decide which of several possible specific forms for \( A \) and \( U \) we will use to describe the electric and magnetic fields. A gauge transformation is given by (9.11), repeated here:

\[
U'(\mathbf{r}, t) = U(\mathbf{r}, t) - \frac{\partial}{\partial t} \chi(\mathbf{r}, t) \quad \text{and} \quad A'(\mathbf{r}, t) = A(\mathbf{r}, t) + \nabla \chi(\mathbf{r}, t). \tag{17.1}
\]

Though several different gauges are useful for different purposes, we will choose, in this case, to select the Coulomb gauge, defined by the constraint

\[
\nabla \cdot A(\mathbf{r}, t) = 0.
\]

Can such a gauge choice be made? The answer is yes. Suppose, for example, that this were not the case; that is, suppose \( \nabla \cdot A \neq 0 \). Then define \( \chi(\mathbf{r}, t) \) to be the solution of the equation

\[
\nabla^2 \chi(\mathbf{r}, t) = -\nabla \cdot A(\mathbf{r}, t) \tag{17.2}
\]

We know, in general, that such a solution always exists; indeed, it isn’t hard to show that it is explicitly given by

\[
\chi(\mathbf{r}, t) = \int \frac{\nabla \cdot A(\mathbf{r}', t)}{4\pi|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r}'.
\]

Then it follows from (17.1) and (17.2) that \( \nabla \cdot A'(\mathbf{r}, t) = 0 \).

In terms of the vector and scalar potential, the electric and magnetic fields are given by (9.6) and (9.7), repeated here

\[
\begin{align*}
\mathbf{E}(\mathbf{r}, t) &= -\dot{A}(\mathbf{r}, t) - \nabla U(\mathbf{r}, t), \\
\mathbf{B}(\mathbf{r}, t) &= \nabla \times A(\mathbf{r}, t). \tag{17.3}
\end{align*}
\]
The first of Maxwell’s equations (Coulomb’s Law) states that

\[ \rho(r,t) = \varepsilon_0 \nabla \cdot E(r,t) = -\varepsilon_0 \nabla \cdot [\mathbf{A}(r,t) + \nabla U(r,t)] = -\varepsilon_0 \nabla^2 U(r,t). \]

Note there are no time derivatives in this equation, so the scalar potential is determined, in this gauge, by the instantaneous charge density

\[ U(r,t) = \int \frac{\rho(r',t)}{4\pi\varepsilon_0 |r-r'|} d^3r', \]

the same formula we have been using all along. All of the dynamics, and interesting quantum mechanical effects, will come from the vector potential \( A \). In the absence of charges, the scalar potential will be trivial, \( U(r,t) = 0 \).

We now want to write the classical Hamiltonian, which is just the energy formula. The energy density of the electromagnetic fields is

\[ u(r,t) = \frac{1}{2} \varepsilon_0 E^2(r,t) + \frac{1}{2} \mu_0^{-1} B^2(r,t) = \frac{1}{2} \varepsilon_0 \left[ E^2(r,t) + c^2 B^2(r,t) \right]. \]  

(17.4)

so the energy of the fields is given by

\[ E = \frac{1}{2} \varepsilon_0 \left[ E^2(r,t) + c^2 B^2(r,t) \right] d^3r = \frac{1}{2} \varepsilon_0 \int \left\{ \mathbf{A}^2(r,t) + c^2 \left[ \nabla \times \mathbf{A}(r,t) \right]^2 \right\} d^3r. \]  

(17.5)

Though it is not obvious, (17.5) is nothing more than a formula for an infinite number of coupled harmonic oscillators. Think of \( \mathbf{A}(r,t) \) as an infinite number of independently varying “positions”; then the time derivative term is just like a kinetic energy term for a particle. The curl term is a derivative, which we can think about as the difference between \( \mathbf{A}(r,t) \) at “adjacent” points, and hence this term is a coupling term.

B. Fourier Modes and Polarization Vectors

To make our work easier, we proceed to work in finite volume. We will imagine that the universe has volume \( V = L^3 \), possessing periodic boundary conditions in all three directions. Therefore, any function \( \psi(r) \) will have the property

\[ \psi(x + L, y, z) = \psi(x, y + L, z) = \psi(x, y, z + L) = \psi(x, y, z). \]

Any such function can written as a sum of Fourier modes,

\[ \mathbf{A}(r,t) = \sum_k \mathbf{A}_k(t) e^{ikr}. \]  

(17.6)

---

1 In previous chapters, we have avoided using \( \varepsilon_0 \), and preferred Coulomb’s constant \( k_c \). The two are related by \( 4\pi k_c \varepsilon_0 = 1 \).

2 See, for example, Jackson, Classical Electrodynamics, third edition (Wiley, 1999) equations (4.89), p. 166 and (5.148), p. 213
Because of the periodic boundary conditions, our $k$-values in the sum are not continuous, but discrete, so that

$$k = \frac{2\pi}{L} n = \frac{2\pi}{L} (n_x, n_y, n_z).$$

where $n$ is a triplet of integers. The mode functions $A_k(t)$ are further restricted by the demand that they must lead to a real vector potential $A(r,t)$, which requires that $A_{-k}(t) = A_k^*(t)$. Also, because we are working in Coulomb gauge, we must have $\nabla \cdot A = 0$, which implies $k \cdot A_k(t) = 0$.

Substituting (17.6) into (17.5), we see that the energy takes the form

$$E = \frac{1}{2} e_0 \int \left[ \sum_k \dot{A}_k(t) e^{ikr} \right]^2 + c^2 \left[ \sum_k i k \times A_k(t) e^{ikr} \right]^2 \right] d^3 r$$

$$= \frac{1}{2} e_0 \sum_k \sum_k \int \left[ \dot{A}_k(t) \cdot \dot{A}_{-k}(t) + i c^2 [k \times A_k(t)] [k' \times A_{-k}(t)] \right] e^{i(k+k')r} d^3 r.$$

The space integral is easy to do, $\int e^{i(k+k')r} d^3 r = V \delta_{k,-k'}$, which yields

$$E = \frac{1}{2} e_0 V \sum_k \left( \dot{A}_k(t) \cdot \dot{A}_{-k}(t) + c^2 k^2 A_k(t) \cdot A_{-k}(t) \right), \quad (17.7)$$

where we have used $k \cdot A_k(t) = 0$ to simplify the dotted cross products. We now need to take advantage of the restriction $A_{-k}(t) = A_k^*(t)$. This allows us to simplify the terms in (17.7). It is also apparent that the sums in (17.7) contains pairs of identical terms. To avoid confusion later, we will divide the $k$ values in half, treating half of them as positive; for example, we can define $k > 0$ as those values for which the first non-zero component of $k$ is positive.$^1$ Then (17.7) becomes

$$E = e_0 V \sum_{k > 0} \left( \dot{A}_k(t) \cdot \dot{A}_k^*(t) + c^2 k^2 A_k(t) \cdot A_k^*(t) \right). \quad (17.8)$$

We have successfully decoupled the different Fourier modes, but we still have vector quantities to deal with. Because of the restriction $k \cdot A_k(t) = 0$, $A_k(t)$ has only two independent components. We define two orthonormal polarization vectors $\mathbf{e}_{k\sigma}$ where $\sigma = 1, 2$, chosen so that

$$k \cdot \mathbf{e}_{k\sigma} = 0 \quad \text{and} \quad \mathbf{e}_{k\sigma} \cdot \mathbf{e}_{k\tau} = \delta_{\sigma\tau}.$$

We then write our modes $A_k(t)$ as

$$A_k(t) = \sum_{\sigma=1,2} A_{k\sigma}(t) \mathbf{e}_{k\sigma}. \quad (17.9)$$

$^1$ This might leave you concerned about the special case $k = 0$; however, it is easily demonstrated that a gauge change can always eliminate such a space-independent mode.
Substituting this into (17.8), we have

\[ E = \varepsilon_0 V \sum_{k=0}^{\infty} \sum_{\sigma} \left[ A_{k\sigma}(t) \cdot \dot{A}_{k\sigma}^*(t) + c^2 k^2 A_{k\sigma}(t) A_{k\sigma}^*(t) \right]. \] (17.10)

We are now prepared to quantize our theory.

C. Quantizing the Electromagnetic Fields

Equation (17.10) is nothing more than a sum of complex harmonic oscillators. We therefore can take advantage of all of the work of section 5F and quickly quantize the theory. The classical theory we quantized in section 5F had an energy given by (5.25), \[ E = m \left( \dot{z}^2 + \omega^2 z^2 \right). \] Comparison with (17.10) tells us that we can take over all our old formulas if we make the following associations:

\[ m \rightarrow \varepsilon_0 V, \quad \omega \rightarrow \omega_k \equiv ck, \quad z \rightarrow A_{k\sigma}. \]

We then simply copy equations like (5.29), without doing any additional work:

\[ H = \sum_{k>0} \sum_{\sigma} \hbar \omega_k \left( a_{k,\sigma}^\dagger a_{k,\sigma} + a_{-k,\sigma}^\dagger a_{-k,\sigma} + 1 \right). \]

Our notation is getting a bit unwieldy. At the moment, there are three indices on \( a \), one denoting which of the two annihilation operators we are talking about for our complex variables \( A_{k\sigma} \), one denoting the vector \( k \) (which is restricted to be positive), and one denoting the polarization. We can combine the first two into a single index \( k \) which is no longer restricted to be positive by defining \( a_{+,k,\sigma} = a_{k,\sigma} \) and \( a_{-,k,\sigma} = a_{-k,\sigma} \). The commutation relations for these operators is

\[ \left[ a_{k,\sigma}, a_{k',\sigma'}^\dagger \right] = \delta_{kk'} \delta_{\sigma\sigma'}, \quad \left[ a_{k,\sigma}, a_{k',\sigma'} \right] = \left[ a_{k,\sigma}^\dagger, a_{k',\sigma'}^\dagger \right] = 0. \] (17.11)

Then, splitting the constant term in half, we can rewrite the Hamiltonian in the more compact form

\[ H = \sum_{k,\sigma} \hbar \omega_k \left( a_{k,\sigma}^\dagger a_{k,\sigma} + \frac{i}{2} \right) \quad \text{where} \quad \omega_k \equiv ck. \] (17.12)

Note we have abbreviated \( a_{k,\sigma} \) as \( a_{k\sigma} \), dropping the comma. The sum is no longer restricted to positive \( k \)-values.

The analog of equations (5.30) can then be written in terms of creation and annihilation operators

\[ A_{k\sigma} \rightarrow \sqrt{\frac{\hbar}{2\varepsilon_0 V \omega_k}} \left( a_{k\sigma} + a_{k,-k,\sigma}^\dagger \right), \quad \dot{A}_{k\sigma} \rightarrow i \sqrt{\frac{\hbar \omega_k}{2\varepsilon_0 V}} \left( a_{-k,\sigma}^\dagger - a_{k,\sigma} \right). \]

This allows us to write our vector potential with the help of (17.6) and (17.9)

\[ \mathbf{A}(r) = \sum_{k,\sigma} \sqrt{\frac{\hbar}{2\varepsilon_0 V \omega_k}} \left( a_{k\sigma} + a_{-k,\sigma}^\dagger \right) \mathbf{e}_{k\sigma} e^{ik \cdot r}. \]
Our analysis has not produced a manifestly Hermitian expression for $A(r)$, which is surprising, since its classical analog is always real. The complication has to do with the way the sums were temporarily restricted to $k > 0$, with the restriction $A_{-k}(t) = A_k^*(t)$. The effect is that we are forced to choose our polarization vectors such that $\varepsilon_{-k,\sigma}^* = \varepsilon_{k,\sigma}^*$. With this restriction, it is more useful to rewrite the final term by taking $k \to -k$, yielding the manifestly Hermitian expression

$$A(r) = \sum_{k,\sigma} \sqrt{\frac{\hbar}{2\varepsilon_0 V \omega_k}} \left( \varepsilon_{k,\sigma} e^{i k r} a_{k,\sigma} + \varepsilon_{k,\sigma}^* e^{-i k r} a_{k,\sigma}^\dagger \right).$$

(17.13)

We can then take the curl of this expression to get the magnetic field. Furthermore, we can find an expression for the electric field by using our expression for $A_{k,\sigma}$ for the time derivatives. Skipping many steps, the results are

$$E(r) = \sum_{k,\sigma} \sqrt{\frac{\hbar \omega_k}{2\varepsilon_0 V}} \left( \varepsilon_{k,\sigma} e^{i k r} d_{k,\sigma} - \varepsilon_{k,\sigma}^* e^{-i k r} d_{k,\sigma}^\dagger \right),$$

(17.14a)

$$B(r) = \sum_{k,\sigma} \sqrt{\frac{\hbar}{2\varepsilon_0 V \omega_k}} i k \times \left( \varepsilon_{k,\sigma} e^{i k r} d_{k,\sigma} - \varepsilon_{k,\sigma}^* e^{-i k r} d_{k,\sigma}^\dagger \right).$$

(17.14b)

In most circumstances, we will work with real polarizations, in which case (17.13) and (17.14) can be simplified.

Note that the time dependences in (17.13) and (17.14) have been dropped. This is because $A$, $E$, and $B$ are now all to be interpreted as operators, which act on some sort of state vector. In particular, $E(r)$ and $B(r)$ are quantities that can be measured experimentally, whose values will be uncertain, due to the spread in whatever wave function you might be interested in ($A(r)$ is not measurable, since it is not gauge invariant), just as the position $x$ of a particle gets replaced by an operator $X$ when we quantize the harmonic oscillator. The quantum states $|\psi\rangle$ are, however, far more complicated, and therefore require some additional comment.

D. Eigenstates of the Electromagnetic Fields

We focus for the moment on the Hamiltonian, given by (17.12), which is clearly just a sum of harmonic oscillators. We start by guessing that the ground state will be the state $|0\rangle$, the state annihilated by all the lowering operators, so $a_{k,\sigma} |0\rangle = 0$. This state has energy $E_0 = \sum \frac{1}{2} \hbar \omega_k = \infty$, thanks to the infinite number of terms in the sum. One might hope that this might have something to do with the finite volume, and that in the limit $V \to \infty$ this infinity might be spread throughout all of space, thereby rendering a relatively innocuous finite energy density, but a more careful analysis will show that in this limit, the energy density is infinite. How can we deal with this infinite energy?
The answer is deceptively simple. As always, adding a constant to the Hamiltonian has no effect on the physics. We always measure differences in energy, not absolute energy values, and hence we can simply rewrite (17.12) as

$$H = \sum_{k,\sigma} \hbar \omega_k a_{k\sigma}^{\dagger} a_{k\sigma}$$

where $$\omega_k = ck$$.  

(17.15)

The ground state now has energy zero.

An arbitrary eigenstate of $$H$$ could be listed by simply listing the eigenstates of all the number operators $$N_{k\sigma} = a_{k\sigma}^{\dagger} a_{k\sigma}$$, but since we have an infinite number of such operators, such a listing would become unwieldy. We therefore list only those states which have non-zero values for these operators, so we denote our states as

$$|n_1, k_1, \sigma_1; n_2, k_2, \sigma_2; \ldots; n_M, k_M, \sigma_M\rangle,$$

(17.16)

with energy

$$E = \hbar \left( n_1 \omega_1 + n_2 \omega_2 + \cdots + n_M \omega_M \right).$$

We will describe such a quantum state as having $$n_1$$ photons of momentum $$\hbar k_1$$ and polarization $$\sigma_1$$, $$n_2$$ photons of momentum $$\hbar k_2$$ and polarization $$\sigma_2$$, etc. These states can be built up from the ground state (vacuum) in the usual way, so that

$$|n_1, k_1, \sigma_1; n_2, k_2, \sigma_2; \ldots; n_M, k_M, \sigma_M\rangle = \frac{1}{\sqrt{n_1! n_2! \cdots n_M!}} \left( a_{k_1\sigma_1}^{\dagger} \right)^{n_1} \left( a_{k_2\sigma_2}^{\dagger} \right)^{n_2} \cdots \left( a_{k_M\sigma_M}^{\dagger} \right)^{n_M} |0\rangle.$$

(17.16)

The order of the triplets in (17.16) is irrelevant. Creation and annihilation operators act as usual on the states (17.16):

$$a_{k\sigma} |n, k, \sigma; \ldots\rangle = \sqrt{n} |n-1, k, \sigma; \ldots\rangle,$$

$$a_{k\sigma}^{\dagger} |n, k, \sigma; \ldots\rangle = \sqrt{n+1} |n+1, k, \sigma; \ldots\rangle.$$

We will sometimes abbreviate our states still further. For example, if we know we will not be discussing more than one photon at a time, we might write the state as $$|k\sigma\rangle$$.

**E. Momentum of photons**

Classically, a photon of energy $$E = \hbar \omega$$ should have a momentum $$p = E/c = \hbar k$$. Will this work quantum mechanically as well? In electromagnetism, the momentum density of the electromagnetic field is given by

$$\mathbf{g} = \varepsilon_0 \mathbf{E} \times \mathbf{B}.$$ We need to integrate this over our volume $$V$$, using our explicit form of the electric and magnetic fields (17.14a,b):

$$\mathbf{P}_{em} = \varepsilon_0 \int \mathbf{E} \times \mathbf{B} \, d^3 \mathbf{r} = \frac{\hbar c}{2V} \sum_{k,\sigma} \sum_{k',\sigma'} \left[ \left( e^{i\mathbf{k} \cdot \mathbf{r}} \mathbf{e}_{k\sigma} a_{k\sigma}^{\dagger} - e^{-i\mathbf{k} \cdot \mathbf{r}} \mathbf{e}_{k\sigma}^{\dagger} a_{k\sigma} \right) \times \left( e^{i\mathbf{k}' \cdot \mathbf{r}} \mathbf{e}_{k'\sigma'} a_{k'\sigma'}^{\dagger} - e^{-i\mathbf{k}' \cdot \mathbf{r}} \mathbf{e}_{k'\sigma'}^{\dagger} a_{k'\sigma'} \right) \right] d^3 \mathbf{r},$$

where $$\mathbf{e}_{k\sigma} = (\mathbf{e}_{k1}, \mathbf{e}_{k2}, \mathbf{e}_{k3})$$.

---

For each of the first terms \( k a_{k,\sigma} a_{-k,\sigma} \), there will be another term \(-k a_{-k,\sigma} a_{k,\sigma}\) which exactly cancels it. The same applies to the last term. The middle two terms can be combined with the help of (17.11) to yield

\[
\mathbf{P}_{em} = \frac{1}{2} \hbar \sum_{k,\sigma} k \left( 1 + 2 a^\dagger_{k,\sigma} a_{k,\sigma} \right).
\]

The term that is just a sum of \( k \) will now cancel between the positive and negative allowed values of \( k \), and the remaining term simplifies to

\[
\mathbf{P}_{em} = \sum_{k,\sigma} \hbar k a^\dagger_{k,\sigma} a_{k,\sigma},
\]

which we simply interpret as photons having momentum \( \hbar k \), as expected.

F. Taking the infinite volume limit

We will commonly want to take the limit where the volume of the universe is increased to infinity. Generally, our strategy will be to keep the volume finite as long as possible, and then apply the limit only when necessary. Two formulas make this easy.

Consider first the following integral, in both the finite and infinite volume limit.

\[
\int d^3 r e^{ikr} e^{-\Delta k r} = \begin{cases} 
V \delta_{k,k'} & \text{if } V < \infty, \\
\left(2\pi\right)^3 \delta^3(k-k') & \text{if } V = \infty.
\end{cases}
\]

This leads to our first rule: We want the former expression to turn into the latter as \( V \) increases to infinity, or in other words,

\[
\lim_{V \to \infty} \left( V \delta_{k,k'} \right) = \left(2\pi\right)^3 \delta^3(k-k').
\]

The other expression comes from considering carefully any sum over momentum states. Consider first a one-dimensional sum of \( k \)-values. In 1D, in the finite volume limit, \( k \) can only take on the discrete values \( k = n\Delta k \) where \( n \) is an integer, and \( \Delta k = 2\pi/L \). In the large size limit, this spacing becomes very small. Indeed, the one-dimensional integral \( \int f(k)dk \) is defined more or less as
\[
\int f(k) \, dk \equiv \lim_{\Delta k \to 0} \Delta k \sum_n f(n \Delta k) = \lim_{L \to \infty} \frac{2\pi}{L} \sum_k f(k).
\]

In one dimension, therefore, we see that the appropriate limit can be written as

\[
\lim_{L \to \infty} \left[ \frac{1}{L} \sum_k f(k) \right] = \int \frac{dk}{2\pi} f(k).
\]

In three dimensions, this generalizes to

\[
\lim_{V \to \infty} \left[ \frac{1}{V} \sum_k f(k) \right] = \int \frac{d^3k}{(2\pi)^3} f(k).
\]

Indeed, (17.17) and (17.18) are flip sides of each other. If you take (17.17) and substitute it into (17.18), both sides become simply one.

G. The nature of the vacuum

Consider an experiment where we measure the electric and magnetic field of empty space. Surely, we would expect the result to be zero. However, in quantum mechanics, things can get more complicated. What would be the expectation value for these fields in the vacuum? Could they fluctuate from zero? To answer these questions, we would need the expectation values of the fields \( \langle \mathbf{E}(r) \rangle \) and \( \langle \mathbf{B}(r) \rangle \) and their squares \( \langle \mathbf{E}^2(r) \rangle \) and \( \langle \mathbf{B}^2(r) \rangle \). We start by computing \( \mathbf{E}(r)\mathbf{|0} \) and \( \mathbf{B}(r)\mathbf{|0} \), which work out to

\[
\mathbf{E}(r)\mathbf{|0} = -i \sum_{k, \sigma} \sqrt{\frac{\hbar}{2\varepsilon_0 V}} \mathbf{\epsilon}_{k\sigma}^* e^{-ik \cdot r} \mathbf{|1, k, \sigma>},
\]

\[
\mathbf{B}(r)\mathbf{|0} = -i \sum_{k, \sigma} \sqrt{\frac{\hbar}{2\varepsilon_0 V}} \mathbf{\epsilon}_{k\sigma}^* e^{-ik \cdot r} \mathbf{|1, k, \sigma>}.
\]

We then immediately find

\[
\langle 0 | \mathbf{E}(r) | 0 \rangle = \langle 0 | \mathbf{B}(r) | 0 \rangle = 0.
\]

In contrast, we have

\[
\langle 0 | \mathbf{E}^2(r) | 0 \rangle = \langle \mathbf{E}(r) | \mathbf{0} \rangle^2 = \frac{\hbar}{2\varepsilon_0 V} \sum_{k, \sigma} \sum_{k', \sigma'} \epsilon_{k\sigma} \epsilon_{k'\sigma'}^* e^{ik \cdot r} \cdot e^{-ik' \cdot r} \mathbb{I} \mathbf{|1, k, \sigma\rangle \langle 1, k', \sigma'|} e^{i(k-k') \cdot r}
\]

\[
= \frac{\hbar}{2\varepsilon_0 V} \sum_{k, \sigma} \hbar c \sum_k \frac{\hbar c}{\varepsilon_0} \int \frac{d^3k}{(2\pi)^3} k,
\]

\[
\langle 0 | \mathbf{B}^2(r) | 0 \rangle = \langle \mathbf{B}(r) | \mathbf{0} \rangle^2 = \frac{\hbar}{2\varepsilon_0 V} \sum_{k, \sigma} \sum_{k', \sigma'} \epsilon_{k\sigma} \epsilon_{k'\sigma'}^* (\mathbf{k} \times \mathbf{\epsilon}_{k\sigma}) \cdot (\mathbf{k}' \times \mathbf{\epsilon}_{k'\sigma'}) \mathbb{I} \mathbf{|1, k, \sigma\rangle \langle 1, k', \sigma'|} e^{i(k-k') \cdot r}
\]

\[
= \frac{\hbar}{2\varepsilon_0 V} \sum_{k, \sigma} \frac{|\mathbf{k} \times \mathbf{\epsilon}_{k\sigma}|^2}{\epsilon_k} \frac{\hbar}{2\varepsilon_0 V} \sum_{k, \sigma} \frac{k^2}{\epsilon_k} \frac{\hbar}{\varepsilon_0 V} \sum_k \frac{\hbar}{\varepsilon_0 V} \int \frac{d^3k}{(2\pi)^3} k^2,
\]

(17.19)
where at the last step, we went to the infinite volume limit. Surprisingly, even though the average value of the electric field and magnetic fields are finite, the expectation values of the squares are infinite, suggesting that the fields have large fluctuations. In retrospect, we probably should have anticipated this, since these two terms contribute to the energy density, and we already know that is infinite.

Experimentally, how come we don’t notice this infinity? One reason is that it is impossible to measure the electric or magnetic field at exactly one point. Suppose that we use a probe of finite size, which measures not \( \mathbf{E}(\mathbf{r}) \), but rather

\[
E_f(\mathbf{r}) = \int f(s) \mathbf{E}(\mathbf{r} + s) d^3 s, \quad \text{where} \quad \int f(s) d^3 s = 1,
\]

and \( f(s) \) is concentrated in the neighborhood of zero. Then we have

\[
E_f(\mathbf{r})|0\rangle = -i \sum_k \int f(s) d^3 s e^{-i k (r + s)} \mathbf{e}_k \sqrt{\frac{\hbar \omega_k}{2e_0 V}} |1, k, \sigma\rangle.
\]

Performing the \( s \) integral will then yield a Fourier transform of \( f \), which yields a factor of 1 at long wavelength, but it will suppress the short wavelength contribution. Since real probes are always finite in size, this will reduce \( \langle 0 | \mathbf{E}_f^2(\mathbf{r}) | 0 \rangle \) and \( \langle 0 | \mathbf{B}_f^2(\mathbf{r}) | 0 \rangle \) and produce a finite value for the fluctuations. But it is always possible, at least in principle, to measure finite fluctuations in the field due to finite sized probes. Without going into too many details, the infinity also goes away if we average over time; the infinite contribution comes from very fast fluctuating fields, and realistic probes can simply not measure fluctuations that fast.

H. The Casimir effect

Before moving onwards, it is worth commenting again on the infinite energy density of the vacuum, which we can obtain easily from (17.19) substituted in (17.4) to yield

\[
\langle 0 | u(\mathbf{r}) | 0 \rangle = \hbar c \int \frac{d^3 k}{(2 \pi)^3} k. \quad (17.20)
\]

As we stated before, we can argue that since the only thing that matters is differences in energies, and this energy is there even in the vacuum, you can normally ignore this effect. There is, however, one important effect which requires that we consider this more carefully. Eq. (17.20) was obtained by considering the limit of infinite volume. What if we don’t have infinite volume? Consider, for example, a parallel plate capacitor, consisting of two conductors very closely spaced together. We will assume the plates are each of area \( A = L^2 \), where \( L \) is very large compared to the separation \( a \), as illustrated in Fig. 17-1. The area is assumed to be so large that it is effectively infinite, but the separation is small enough that the modes in this direction

---

**Figure 17-1**: A parallel plate capacitor consisting of two conducting plates of area \( A \) separated by a small distance \( a \).
will take on only distinctly discrete values. Therefore the energy density between the plates will not be given by (17.20). We must redo the work of section B, and quantize subject to these restrictions.

Considerable work is required to figure out what modes are allowed subject to our constraints. It is forbidden to have electric fields parallel to a conducting surface, which suggests that we should in general have \( \mathbf{A}_\parallel(x, y, 0, t) = \mathbf{A}_\parallel(x, y, a, t) = 0 \). Since virtually all modes have some component of \( \mathbf{A} \) parallel to the conducting plates, this suggests we write

\[
\mathbf{A}(r, t) = \sum_k \mathbf{A}_k(t) e^{i \mathbf{k} \cdot \mathbf{r}} \frac{\sin(\mathbf{k}_z z)}{\mathbf{k}_z},
\]

(17.21)

The wave numbers are now restricted to be of the form

\[
\mathbf{k} = \left( \frac{2\pi n_x}{L}, \frac{2\pi n_y}{L}, \frac{\pi n_z}{a} \right),
\]

(17.22)

where \( n_x \) and \( n_y \) are arbitrary integers, but \( n_z \) must be a positive integer. However, there is one other type of mode that we have missed in (17.21), which occurs because if \( \mathbf{k} \) is parallel to the capacitor plates, we can choose our polarization \( \hat{z} \), perpendicular to the capacitor plates. This leads to additional terms of the form \( \mathbf{A}_k(t) \hat{z} e^{i \mathbf{k} \cdot \mathbf{r}} \), where \( k_z = 0 \). Hence each value of \( \mathbf{k} \) given in (17.22) with positive \( n_z \) will have two possible polarizations, but there will be a single additional polarization with \( n_z = 0 \).

Now, our Hamiltonian will still be given by (17.12), and therefore the ground state energy will be \( E_0(a) = \frac{1}{2} \hbar c \sum_{k, \sigma} k \). In the usual manner, we can turn the \( k_x \) and \( k_y \) sums into integrals by taking the limit \( L \to \infty \), but the \( k_z \) and polarization sum must be dealt with explicitly and we find

\[
E_0(a) = L^2 \sum_{k_y, \sigma} \int \frac{dk_y}{(2\pi)^\frac{1}{2}} \frac{1}{2} \hbar c k.
\]

Not surprisingly, this number is infinite. However, we are interested in the difference between this value and the empty space vacuum value, which can be found by multiplying the energy density (17.20) by the volume \( L^2 a \), so we have

\[
\Delta E \equiv E_0(a) - E_0 = \frac{1}{2} \hbar c L^2 \sum_{k_y, \sigma} \int \frac{dk_y}{(2\pi)^\frac{1}{2}} \frac{1}{2} k - L^2 a \hbar c \int \frac{d^3k}{(2\pi)^\frac{3}{2}} k.
\]

(17.23)

Now, we have a problem in that both expressions are infinite, and we don’t know how to subtract them. Fortunately, this infinity is not physical, because real conductors do not work at arbitrarily high frequencies. Hence we should put in some sort of cutoff function \( f(k) \) that is one for small \( k \) and vanishes for large \( k \). The exact form used can be shown not to matter, so I will use the simple formula \( f(k) = e^{-\beta k} \), where \( \beta \) is some small number, governing when the conductor effectively becomes transparent. Substituting this into (17.23), we have
\[ \Delta E = \frac{1}{2} \sum_{\sigma} \int \frac{dk_x dk_y}{(2\pi)^2} e^{-\beta k} - L^2 ahc \int \frac{d^3k}{(2\pi)^3} e^{-\beta k}. \]

On the first term, we switch first to cylindrical coordinates \( k_x^2 = k_x^2 + k_y^2 \) and angle \( \phi \), perform the \( \phi \) integral, and then switch to the spherical coordinate \( k^2 = k_x^2 + k_z^2 \). In the latter term, we switch to spherical coordinates and perform all integrals. The result is

\[
\Delta E = \frac{L^2 ahc}{4\pi} \sum_{k_x, \sigma} \int_0^\infty k_x^2 e^{-\beta k_x} dk_x = \frac{L^2 ahc}{4\pi} \int_0^\infty k^2 e^{-\beta k} dk - \frac{3L^2 ahc}{\pi^2 \beta^4}.
\]

It’s time to do the sums on \( k_z \). As argued above, \( k_z = \pi n/a \), where \( n \) is an integer. There will be two polarizations when \( n \) is positive, and one when \( n = 0 \), so we have

\[
\Delta E = \frac{L^2 ahc}{2\pi} \left[ \frac{1}{\beta^3} + \sum_{n=1}^{\infty} \left( \frac{\pi^2 n^2}{\beta a^2} + \frac{2\pi n}{\beta^2 a} + \frac{2}{\beta^3} \right) e^{-\pi n \beta/\alpha} \right] - \frac{3L^2 ahc}{\pi^2 \beta^4}. \quad (17.25)
\]

All the sums can be performed exactly using the identities

\[
\sum_{n=1}^{\infty} x^n = \frac{x}{1-x}, \quad \sum_{n=1}^{\infty} n x^n = \frac{x}{(1-x)^2}, \quad \sum_{n=1}^{\infty} n^2 x^n = \frac{x+x^2}{(1-x)^3},
\]

valid for \( |x| < 1 \). The first is just the geometric series, and the other two can be easily derived by taking the derivative of the previous series with respect to \( x \). Applying these equations to (17.25), we see that

\[
\Delta E = \frac{L^2 ahc}{2\pi} \left[ \frac{\pi^2 (e^{-\pi \beta/\alpha} + e^{-2\pi \beta/\alpha})}{a^2 \beta (1-e^{-\pi \beta/\alpha})^3} + \frac{2\pi e^{-\pi \beta/\alpha}}{a \beta^2 (1-e^{-\pi \beta/\alpha})} + \frac{2e^{-\pi \beta/\alpha}}{\beta^3 (1-e^{-\pi \beta/\alpha})} + \frac{1}{\beta^3} \right] - \frac{3L^2 ahc}{\pi^2 \beta^4},
\]

Now, the actual frequencies at which the cutoff occurs tend to correspond to wavelengths much shorter than the experimental separation we can achieve, so \( \beta \ll a \). We therefore expand \( \Delta E \) in this limit. Rewrite our expression in terms of \( w = \pi \beta/2a \).

\[
\Delta E = \frac{\pi^2 L^2 ahc}{16a^3} \left( \frac{\cosh w}{w^3 \sinh^2 w} + \frac{1}{w^2 \sinh^2 w} + \frac{\cosh w}{w^3 \sinh w} - \frac{3}{w^4} \right).
\]

We now expand in powers of \( w \). We’ll let Maple do the work for us:

\[
\text{> series(cosh(w)/w^3*sinh(w)^3 + 1/w^2*2/sinh(w)^2 + cosh(w)/w^3/sinh(w) - 3/w^4,w,11);}
\]

\[
\Delta E = \frac{\pi^2 L^2 ahc}{16a^3} \left( -\frac{4}{45} + \frac{4}{315} w^2 - \frac{1}{315} w^4 + \cdots \right) = -\frac{\pi^2 L^2 ahc}{720a^3} \left[ 1 + O(w^2) \right].
\]
We see that in the limit \( w \to 0 \) the energy difference is finite and non-zero. Not surprisingly, the energy is proportional to the area, so we actually have a formula for the energy per unit area. Note that the result is negative, so there is less energy as the plates draw closer. This means there is an attractive force between them, or rather, a force per unit area (pressure) given by

\[
P = -\frac{1}{L^2} \frac{d}{da}(\Delta E) = -\frac{\pi^2 \hbar c}{240a^4} = -1.30 \text{ mN/m}^2.
\]

This force is attractive, pulling the two plates of the capacitor towards each other. Though small, it has been measured.

Problems for Chapter 17

1. In class, we quantized the free electromagnetic field. In homework, you will quantize the free massive scalar field, with classical energy

\[
E = \frac{1}{2} \int d^3r \left\{ \phi^2(r,t) + \mu^2 \left[ \nabla \phi(r,t) \right]^2 + \mu^2 \phi^2(r,t) \right\}
\]

This problem differs from the electromagnetic field in that: (i) there is no such thing as gauge choice; (ii) the field \( \phi(r,t) \) is not a vector field; it doesn’t have components, and (iii) there is a new term \( \mu^2 \phi^2 \), unlike anything you’ve seen before.

a) Write such a classical field \( \phi(r,t) \) in terms of Fourier modes \( \hat{\phi}_k(t) \). What is the relationship between \( \hat{\phi}_k(t) \) and \( \hat{\phi}_{-k}(t) \)?

b) Substitute your expression for \( \phi(r,t) \) into the expression for \( E \). Work in finite volume \( V \) and do as many integrals and sums as possible.

c) Restrict the sum using only positive values of \( k \). Argue that you now have a sum of classical complex harmonic oscillators. What is the formula for \( \omega_k \), the frequency for each of these oscillators?

d) Reinterpret \( H \) as a Hamiltonian, and quantize the resulting theory. Find an expression for the Hamiltonian in terms of creation and annihilation operators.

2. How do we create the classical analog of a plane wave quantum mechanically? Naively, you simply use a large number of quanta.

a) Suppose the EM field is in the quantum state \( |n,k,\sigma\rangle \), where \( n \) is large. Find the expectation value of the electric \( \langle \mathbf{E}(r) \rangle \) and magnetic fields \( \langle \mathbf{B}(r) \rangle \) for this quantum state. For definiteness, choose \( k = k\hat{z} \) and \( \epsilon_{k\sigma} = \hat{x} \)

b) Now try a coherent state, given by \( |\psi_c\rangle = e^{i\sigma/\hbar} \sum_{n=0}^{\infty} \left( c^n / \sqrt{n!} \right) |n,k,\sigma\rangle \), where \( c \) is an arbitrary complex number. Once again, find the expectation value of the electric and magnetic field. Coherent states were described in section 5D.
3. Suppose we measure the instantaneous electric field using a probe of finite size, so that we actually measure \( E_f (r) = \int E(r+s) f(s) \, ds \), where \( f(s) = \pi^{-3/2} a^{-3} e^{-s^2/a^2} \), where \( a \) is the characteristic size of the probe. For the vacuum state, find the expectation value of \( \langle E_f (r) \rangle \) and \( \langle E^2_f (r) \rangle \). You should take the infinite volume limit, and make sure your answer is independent of \( V \).

4. Define, for the electric and magnetic field, the annihilation and creation parts as

\[
E_+ (r) = i \sum_{k, \sigma} \sqrt{\frac{\hbar \omega_k}{2 \varepsilon_0 V}} \mathbf{e}_{k\sigma} e^{i k \cdot r} a_{k\sigma}^+, \quad E_- (r) = -i \sum_{k, \sigma} \frac{\hbar \omega_k}{2 \varepsilon_0 V} \mathbf{e}_{k\sigma} e^{-i k \cdot r} a_{k\sigma}^-,
\]

\[
B_+ (r) = i \sum_{k, \sigma} \frac{\hbar}{2 \varepsilon_0 V \omega_k} \mathbf{k} \times \mathbf{e}_{k\sigma} e^{i k \cdot r} a_{k\sigma}^+, \quad B_- (r) = -i \sum_{k, \sigma} \frac{\hbar}{2 \varepsilon_0 V \omega_k} \mathbf{k} \times \mathbf{e}_{k\sigma} e^{-i k \cdot r} a_{k\sigma}^-.
\]

It should be obvious that \( E(r) = E_+ (r) + E_- (r) \) and \( B(r) = B_+ (r) + B_- (r) \).

(a) Define the normal-ordered energy density as

\[
\tilde{u}(r) \equiv \frac{1}{2} \varepsilon_0 \left[ \epsilon_r^2 (r) + 2 \epsilon_r (r) \cdot E_- (r) + E_-^2 (r) \right] + c^2 \left[ B^2_+ (r) + 2 B_+ (r) \cdot B_- (r) + B_-^2 (r) \right].
\]

Prove that this normal-ordered energy density differs from the usual definition by a constant, i.e., that the difference contains no operators (the constant will be infinite).

(b) Prove that the expectation value of this operator for the vacuum is zero.

(c) Consider the quantum state \( |\psi\rangle = \frac{\sqrt{2}}{2} |0\rangle + \frac{\sqrt{2}}{2} |2, q, \tau\rangle \); i.e., a quantum superposition of the vacuum and a two photon state with wave number \( q \) and polarization \( \tau \).

To keep things simple, let the polarization \( \mathbf{e}_{q\tau} \) be real. Work out the expectation value \( \langle \tilde{u}(r) \rangle \) for this quantum state.

(d) Sketch \( \langle \tilde{u}(r) \rangle \) as a function of \( q \cdot r \). Note that it is sometimes negative (less energy than the vacuum!). Find its integral over space, and check that it does, however, have total energy positive.
We have quantized the electromagnetic field, and we have discussed atoms as well in terms of quantum mechanics. It is time to put our knowledge together so that we can gain an understanding of how photons interact with matter. Our tool will be primarily time-dependent perturbation theory, in which we divide the Hamiltonian \( H \) into two pieces, \( H = H_0 + W \), where \( H_0 \) will be assumed to be solved, and \( W \) will be small.

The rate at which an interaction occurs is then given by Fermi’s Golden rule, (15.35), and the transition matrix (15.34), given by

\[
T_{FI} = \lim_{\epsilon \to 0} \left[ \sum_m \frac{W_{Fm} W_{ml}}{(E_l - E_m + i\epsilon)} + \sum_n \frac{W_{Fn} W_{mn} W_{nl}}{(E_l - E_m + i\epsilon)(E_l - E_n + i\epsilon)} + \cdots \right],
\]

where \( W_{nm} = \langle n \mid W \mid m \rangle \). Our first step will be to break the Hamiltonian up into a main part and a perturbation, and find the eigenstates of the main part.

A. The Hamiltonian

The Hamiltonian we wish to consider is one involving both atoms and (quantized) electromagnetic fields. The pure electromagnetic field energy \( H_{em} \) will be given by (17.5), which, after quantization, becomes (17.15). This Hamiltonian was the focus of the previous chapter. In addition, there will be the interaction of all the electrons in all of the atoms, etc. With the help of (9.20), we see that the full Hamiltonian will be

\[
H = \sum_{j=1}^{N} \left\{ \frac{1}{2m} \left[ p_j + eA(R_j) \right]^2 + \frac{e}{m} B(R_j) \cdot S_j \right\} + V(R_1, \ldots, R_N) + H_{em},
\]

where \( V \) contains all of the interactions of the electrons with each other, or with the nuclei or background fields, etc., and we have approximated \( g = 2 \) for the electron. Indeed, in general we might want to also include lesser effects in \( V \), such as the spin-orbit coupling within an atom, but what we want to exclude is any interaction with external electromagnetic fields, which are explicitly shown. Keep in mind that \( A \) and \( B \) are no longer mere functions (as they were previously) but now are full operators. Recall that we are working in Coulomb gauge, and therefore the electrostatic potential is determined by the instantaneous charge distribution.

We now wish to divide \( H \) up into “large” and “small” pieces, which we do as follows:

\[
H = H_{\text{atom}} + H_{\text{em}} + W^{(1)} + W^{(2)},
\]

where
\[ H_{\text{atom}} = \sum_{j=1}^{N} \frac{1}{2m} \mathbf{p}_j^2 + V(\mathbf{r}_1, \ldots, \mathbf{r}_N), \]

and the perturbations are\(^1\)

\[ W^{(1)} = \frac{e}{m} \sum_{j=1}^{N} \left[ A \left( \mathbf{r}_j \right) \cdot \mathbf{p}_j + B \left( \mathbf{r}_j \right) \cdot \mathbf{S}_j \right] \quad \text{and} \quad W^{(2)} = \frac{e^2}{2m} \sum_{j=1}^{N} \mathbf{A}^2 \left( \mathbf{r}_j \right). \quad (18.2) \]

This distinction between the two perturbative terms is a natural one, because \( W^{(1)} \) is first order in the charge \( e \), while \( W^{(2)} \) is second order, and one way we will keep track of our perturbation theory is by counting factors of \( e \). Hence if we perform a computation to second order, we will allow up to two factors of \( W^{(1)} \), but only one factor of \( W^{(2)} \). The explicit form of (18.2) will be needed later in terms of creation and annihilation operators, and using (17.13) and (17.14b), these can be written

\[ W^{(1)} = \frac{e}{m} \sum_{j=1}^{N} \sum_{k\sigma} \sqrt{\frac{\hbar}{2e_0 V \omega_k}} \left\{ e^{i \mathbf{k} \cdot \mathbf{r}} \mathbf{e}_{k\sigma} \cdot \left[ \mathbf{p}_j + i \left( \mathbf{S}_j \times \mathbf{k} \right) \right] a_{k\sigma} + e^{-i \mathbf{k} \cdot \mathbf{r}} \mathbf{e}_{k\sigma}^* \cdot \left[ \mathbf{p}_j - i \left( \mathbf{S}_j \times \mathbf{k} \right) \right] a_{k\sigma}^* \right\}, \]

\[ W^{(2)} = \frac{e^2 \hbar}{4m \epsilon_0 V} \sum_{j=1}^{N} \sum_{k\sigma} \sum_{k'\sigma'} \left( e^{i \mathbf{k} \cdot \mathbf{r}} \mathbf{e}_{k\sigma} a_{k\sigma} + e^{-i \mathbf{k} \cdot \mathbf{r}} \mathbf{e}_{k\sigma}^* a_{k\sigma}^* \right) \frac{\mathbf{e}_{k'\sigma'} a_{k'\sigma'} + e^{-i \mathbf{k} \cdot \mathbf{r}} \mathbf{e}_{k'\sigma'}^* a_{k'\sigma'}}{\sqrt{\omega_k \omega_{k'}}}. \]

\[ (18.3) \]

We now let our first two terms be our unperturbed Hamiltonian,

\[ H_0 = H_{\text{atom}} + H_{\text{em}}. \]

Fortunately, these two terms are completely decoupled, and therefore to find eigenstates of \( H_0 \), we need only find eigenstates of these two parts separately. We will now assume that we have somehow managed to find the exact eigenstates of \( H_{\text{atom}} \), which we call \( |\phi_n\rangle \), where \( n \) describes all the quantum numbers associated with the atom itself, such as \( l, j, m_j \), etc. This state will have energy \( \epsilon_n \). In addition, there will be eigenstates of \( H_{\text{em}} \), which we computed in the previous chapter. The overall eigenstates will then be given by

\[ |\phi_n; n_1, k_1, \sigma_1; n_2, k_2, \sigma_2; \ldots; n_M, k_M, \sigma_M \rangle. \]

and will have energy

\[ E = \epsilon_n + \hbar \left( n_1 \omega_1 + n_2 \omega_2 + \cdots + n_M \omega_M \right). \]

We are now ready to start putting in the effect of our perturbations.

\(^1\) Note that \( \mathbf{p}_j \cdot A \left( \mathbf{r}_j \right) = A \left( \mathbf{r}_j \right) \cdot \mathbf{p}_j \), because we are working in Coulomb gauge where \( \nabla \cdot A \left( \mathbf{r} \right) = 0 \).
B. Absorption and Emission of Photons by Atoms

Consider first the effect of our interaction terms to leading order in $e$. To this order, we need only consider $W^{(1)}$, given in (18.3). We immediately note that this perturbation can only create or destroy one photon. It follows that the final state must be identical with the initial state, save for a single photon. The only matrix elements we will consider, therefore, will be of the form

$$\langle \phi_f; n_i \pm 1, k_1, \sigma_1; \ldots | W^{(1)} | \phi_i; n_i, k_i, \sigma_i; \ldots \rangle.$$ 

The energy difference $E_F - E_i$ will therefore be

$$E_F - E_i = \varepsilon_F + \hbar \left[ (n_i \pm 1) \omega + n_2 \omega_2 + \cdots + n_M \omega_M \right] - \varepsilon_i - \hbar \left[ n_i \omega + n_2 \omega_2 + \cdots + n_M \omega_M \right]$$

$$= \varepsilon_F - \varepsilon_i \pm \hbar \omega = \hbar (\omega_{fi} \pm \omega),$$

where we have renamed $\omega_k = \omega$ since this is the frequency that will interest us most. Since this must vanish by Fermi’s Golden rule, we conclude that $\varepsilon_F - \varepsilon_i = \mp \hbar \omega$, so we either emit a photon and decrease the atom’s energy, or we absorb a photon and increase the atom’s energy.

Consider first the case where we absorb a photon, so that $\varepsilon_F > \varepsilon_i$. Then only the annihilation part of $W^{(1)}$ will contribute. Furthermore, in the sum, only the single term that matches the wave number $k_1$ and polarization $\sigma_1$ will contribute. We will rename these as $k$ and $\sigma$, so we have

$$W^{(1)}_{fi} = \frac{e}{m} \sum_{j=1}^{\infty} \sqrt{\frac{\hbar}{2\varepsilon_0 V \omega}} \langle \phi_f; n-1, k, \sigma; \ldots | e^{ik \cdot \mathbf{R}} \mathbf{e}_{k_{\sigma}} \left[ \mathbf{P}_j + i (\mathbf{S}_j \times k) \right] | a_{k_{\sigma}} \phi_i; n, k, \sigma; \ldots \rangle.$$ 

The photon part of this expression is easy to find; it is just

$$\langle n-1, k, \sigma; \ldots | a_{k_{\sigma}} | n, k, \sigma; \ldots \rangle = \sqrt{n} \langle n-1, k, \sigma; \ldots | n-1, k, \sigma; \ldots \rangle = \sqrt{n},$$

so the whole matrix element is just

$$W^{(1)}_{fi} = \frac{e}{m} \sum_{j=1}^{\infty} \sqrt{\frac{\hbar n}{2\varepsilon_0 V \omega}} \langle \phi_F | e^{ik \cdot \mathbf{R}} \mathbf{e} \left[ \mathbf{P}_j + i (\mathbf{S}_j \times k) \right] | \phi_i \rangle.$$ 

Comparing this, for example, with (15.25), we see that our computation now closely follows what we found before. To leading order, we can approximate $e^{ik \cdot \mathbf{R}} = 1$, and ignore the spin term to obtain the electric dipole approximation

$$W^{(1)}_{fi} \approx \frac{e}{m} \sqrt{\frac{\hbar n}{2\varepsilon_0 V \omega}} \langle \phi_F | \mathbf{P} | \phi_i \rangle.$$ 

In a manner identical to before, we can again derive equation (15.27), relating $\langle \phi_F | \mathbf{P} | \phi_i \rangle$ and $\langle \phi_F | \mathbf{R} | \phi_i \rangle$, and rewrite this as
\[ W^{(i)}_{FI} \approx ie \sqrt{\frac{n\hbar \omega}{2\epsilon_0 V}} \mathbf{e} \cdot \langle \phi_f | R | \phi_i \rangle = ie \sqrt{\frac{n\hbar \omega}{2\epsilon_0 V}} \mathbf{e} \cdot \mathbf{r}_{fi}. \]

where we used the fact that we must have \( \omega = \omega_{fi} \). Substituting this into (18.1) then yields

\[ \Gamma (I \rightarrow F) = \frac{\pi e^2 n\omega}{\epsilon_0 V} |\mathbf{e} \cdot \mathbf{r}_{fi}|^2 \delta (\omega_f - \omega) = 4\pi^2 \alpha \hbar^{-1} \left( \frac{cn\hbar \omega}{V} \right) |\mathbf{e} \cdot \mathbf{r}_{fi}|^2 \delta (\omega_f - \omega), \]

where we have replaced factors of \( e \) using the fine structure constant \( \alpha = e^2 / 4\pi e_0 hc \). A careful comparison with (15.29) will convince you that they are identical formulas. The factor of \( (cn\hbar \omega/V) \) is simply the energy of the photons \( n\hbar \omega \), divided by the volume (yielding energy density) and multiplied by the speed of light, and hence is just the intensity \( I \).

Let’s now consider the reverse case, where the final energy is less than the initial. Then we must increase the number of photons by one, which means we need the other half of our perturbation, so

\[ W^{(i)}_{FI} = \frac{e}{m} \sum_{j=1}^{N} \sqrt{\frac{\hbar}{2\epsilon_0 V \omega}} \langle \phi_f ; n+1, \mathbf{k}, \sigma, \ldots | e^{i \mathbf{R} \cdot \mathbf{P}_j - i \mathbf{e} \cdot (\mathbf{k} \times \mathbf{e}^* \cdot \mathbf{S}_j) \rangle a^\dagger_{\mathbf{k} \sigma} | \phi_i ; n, \mathbf{k}, \sigma, \ldots \rangle. \]

This time the photon part of this matrix element is \( \langle n+1, \mathbf{k}, \sigma, \ldots | a^\dagger_{\mathbf{k} \sigma} | n, \mathbf{k}, \sigma, \ldots \rangle = \sqrt{n+1} \), which yields

\[ W^{(i)}_{FI} = \frac{e}{m} \sum_{j=1}^{N} \sqrt{\frac{\hbar(n+1)}{2\epsilon_0 V \omega}} \langle \phi_f | e^{-i \mathbf{R} \cdot \mathbf{P}_j - i \mathbf{e} \cdot (\mathbf{k} \times \mathbf{e}^* \cdot \mathbf{S}_j) \rangle | \phi_i \rangle. \]

In the dipole approximation, this can be simplified in a manner very similar to before. Skipping steps, the final answer is

\[ \Gamma (I \rightarrow F) = \left( \frac{4\pi^2 \alpha \hbar}{\hbar} \right) \left[ \frac{c(n+1)\hbar \omega}{V} \right] |\mathbf{e} \cdot \mathbf{r}_{fi}|^2 \delta (\omega - \omega_f). \]

This time, the result is not what we got before. If you keep the term proportional to \( n \), we will get exactly the same result as before. The interesting thing is that we have a new term which does not require the presence of photons in the initial state at all. This new process is called spontaneous emission. The rate is given, in the electric dipole approximation, by

\[ \Gamma (I \rightarrow F) = \left( \frac{4\pi^2 \alpha c \omega}{V} \right) |\mathbf{e} \cdot \mathbf{r}_{fi}|^2 \delta (\omega - \omega_{fi}). \]

As it stands this formula is a bit difficult to interpret. It has a delta function, which makes it look infinite, but it also has the reciprocal of the volume, which just means that in the limit of infinite volume, the probability of it going to a particular wave number \( \mathbf{k} \) is vanishingly small. The way to avoid these double difficulties is to sum over
all possible outgoing wave numbers, and then take the limit of infinite volume, which gives us

\[ \Gamma(I \to F) = \frac{4\pi^2 \alpha_c \omega}{V} |\mathbf{e}^* \cdot \mathbf{r}_{F_f}|^2 \sum_k \delta(\omega - \omega_{IF}) = 4\pi^2 \alpha_c \omega \int \frac{d^3 \mathbf{k}}{(2\pi)^3} |\mathbf{e}^* \cdot \mathbf{r}_{F_f}|^2 \delta(\omega - \omega_{IF}) \]

\[ = \frac{\alpha_c \omega}{2\pi} \int d\Omega_k |\mathbf{e}^* \cdot \mathbf{r}_{F_f}|^2 \int k^2 dk \delta(kc - \omega_{IF}) = \frac{\alpha_c \omega_k^3}{2\pi} \int d\Omega_k |\mathbf{e}^* \cdot \mathbf{r}_{F_f}|^2, \]

\[ \frac{d\Gamma(I \to F)}{d\Omega_k} = \frac{\alpha_c \omega_k^3}{2\pi c^3} |\mathbf{e}^* \cdot \mathbf{r}_{F_f}|^2. \]

This rate is the angle-dependant polarized rate. If we do not measure the polarization of the outgoing wave, then the effective rate is the sum of this expression over the two polarizations. Without working through the details too much, the result is

\[ \frac{d\Gamma(I \to F)}{d\Omega_k} = \frac{\alpha_c \omega_k^3}{2\pi c^3} \left( |\mathbf{r}_{F_f}|^2 - |\mathbf{k} \cdot \mathbf{r}_{F_f}|^2 \right). \]

where \( \mathbf{\hat{k}} \) denotes a unit vector in the direction of \( \mathbf{k} \). For example, if \( \mathbf{r}_{F_f} \) is a real vector, the last factor is simply \( \sin^2 \theta \), where \( \theta \) is the angle between \( \mathbf{k} \) and \( \mathbf{r}_{F_f} \). If we don’t measure the direction of the final photon, we can integrate over solid angle to get a final answer

\[ \Gamma(I \to F) = \frac{4\alpha_c \omega_k^3 |\mathbf{r}_{F_f}|^2}{3c^2}. \] (18.5)

Of course, we can go beyond the dipole approximation when needed. The atomic matrix elements can be expanded in powers of \( \mathbf{k} \cdot \mathbf{R}_j \) to any order desired. The first three terms can be written in the form

\[ \frac{e}{m} \langle \phi_f | e^{-i\mathbf{k} \cdot \mathbf{R}_j} \left[ \mathbf{e}^* \cdot \mathbf{P}_j - i(\mathbf{k} \times \mathbf{e}^*) \cdot \mathbf{S}_j \right] |\phi_i \rangle = \langle \phi_f | \left[ i\epsilon \mathbf{e} \cdot \mathbf{R}_j + \frac{i}{\epsilon} \epsilon \mathbf{e} \cdot \mathbf{R}_j \left( \mathbf{e}^* \cdot \mathbf{R}_j \right) \right] (\mathbf{k} \cdot \mathbf{R}_j) \]

\[ - \frac{i\epsilon}{m} (\mathbf{k} \times \mathbf{e}^*) \cdot \left( \mathbf{S}_j + \frac{1}{2} \mathbf{L}_j \right) + \cdots \rangle |\phi_i \rangle. \]

The three terms correspond to the electric dipole, electric quadrupole, and magnetic dipole terms respectively. Equations analogous to (18.5) can then be derived for each of these other two cases.
C. The Self-Energy of the Electron

We now see that to leading order in perturbation theory, \( W^{(1)} \) causes atoms to absorb or emit a single photon. What is the effect of \( W^{(2)} \)? There are a variety of effects, but notice in particular that it will shift the energy of an arbitrary atomic state \( |\phi\rangle \). To first order in time-independent perturbation theory, the shift in energy of a single electron will be \( \epsilon' = \langle \phi | W^{(2)} | \phi \rangle = e^2 \langle \phi | A^2 (R) | \phi \rangle / 2m \). We therefore compute

\[
A(R)|\phi\rangle = \sum_{k,\sigma} \sqrt{\frac{\hbar}{2\varepsilon_0 \omega_k V}} (e^{i\mathbf{k}\cdot\mathbf{R}} d_{k,\sigma}^* \mathbf{e}_{k,\sigma} + e^{-i\mathbf{k}\cdot\mathbf{R}} d_{k,\sigma} \mathbf{e}_{k,\sigma}^*) |\phi\rangle = \sum_{k,\sigma} \sqrt{\frac{\hbar}{2\varepsilon_0 \omega_k V}} e^{-i\mathbf{k}\cdot\mathbf{R}} \mathbf{e}_{k,\sigma}^* |\phi;1,k,\sigma\rangle.
\]

We therefore have

\[
\epsilon' = \frac{e^2}{2m} \langle \phi | A^2 (R) | \phi \rangle = \frac{\hbar e^2}{4m\varepsilon_0 V} \sum_{k,\sigma,k,\sigma'} \frac{\mathbf{e}_{k,\sigma} \cdot \mathbf{e}_{k,\sigma'}}{\omega_k \omega_{k'}} \langle \phi;1,k,\sigma | e^{i\mathbf{k}\cdot\mathbf{R}} e^{-i\mathbf{k'}\cdot\mathbf{R}} | \phi;1,k',\sigma' \rangle
\]

\[
= \frac{\hbar e^2}{4m\varepsilon_0 V} \sum_{k,\sigma} \frac{1}{\omega_k} = \frac{\hbar e^2}{2m\varepsilon_0} \frac{d^3 k}{(2\pi)^3} \frac{4\pi}{\omega_k} = \frac{\hbar e^2}{2m\varepsilon_0} \frac{4\pi}{(2\pi)^3} \frac{k^2 dk}{ck} = \frac{\alpha \hbar^2}{\pi m} \int_0^\infty k^2 dk = \infty.
\]

Hence the electron will be shifted up in energy by an infinite amount! Note, however, that this shift is completely independent of what state the electron is in. It is therefore an intrinsic property of the electron, and is irrelevant.

It is sometimes suggested that these infinities that seem to be appearing have something inherent to do with quantum mechanics. But the same problems plague classical mechanics. An electron at the origin is surrounded by an electric field \( \mathbf{E}(r) = k_e \mathbf{\hat{r}} / r^2 \), which has an energy density \( \frac{1}{2} \varepsilon_0 \mathbf{E}^2 (r) = k_e e^2 / 8\pi r^4 = \alpha \hbar c / 8\pi r^4 \). If we integrate the resulting energy density over all space, we find there is an infinite energy associated with the electric field around an electron. Hence this problem of infinities exists classically as well.

D. Photon Scattering

We turn our attention now to the subject of photon scattering. We will consider situations where both the initial and final state includes a photon. We will assume only a single photon, so the initial and final states are

\[
|I\rangle = |\phi_i; k_i, \sigma_i\rangle \quad \text{and} \quad |F\rangle = |\phi_F; k_F, \sigma_F\rangle,
\]

where there is an implied “1” describing the number of photons in each case. The initial and final state energy are \( E_i = \varepsilon_i + \hbar \omega_i \) and \( E_F = \varepsilon_F + \hbar \omega_F \). We are interested only in the case where the final photon differs from the initial one. We would like to calculate the probability \( P(I \rightarrow F) \) to second order in perturbation theory. The amplitude is

\[
T_{IF} = \langle F | W^{(2)} | I \rangle + \lim_{\delta \to 0} \sum_m \frac{\langle F | W^{(1)} | m \rangle \langle m | W^{(1)} | I \rangle}{E_i - E_m + i\delta}, \quad (18.6)
\]
where we have changed our small variable to $\delta$ to avoid confusion with the atomic energies $\varepsilon_m$. We have left out $W^{(1)}$ in the first term because it can only create or annihilate one photon, and therefore $\langle F | W^{(1)} | I \rangle = 0$. We have left out $W^{(2)}$ in the second term because this has more factors of the coupling $\epsilon$, and therefore is higher order.

In the second term, what sort of intermediate states are possible? Since we have assumed the final photon state is distinct from the initial, and each of the two perturbations can only annihilate or create one photon, we conclude that one factor of $W^{(1)}$ must be eliminating the initial photon and the other creating the final photon. Depending on which is which, the intermediate state will either have no photons in it, or it will have two photons, one each in the initial and final state. So the only possibilities for the intermediate state are $\ket{m} = \ket{\phi_m}$ or $\ket{m} = \ket{\phi_m; k_f, \sigma_f; k_f, \sigma_f}$. These states have energies of $E_m = \varepsilon_m$ or $E_m = \varepsilon_m + \hbar \omega_{F}$. Thus the sum in (18.6) will become a double sum, one term for each of these cases, and we have

$$T_{FI} = \langle \phi_f; k_f, \sigma_f | W^{(2)} | \phi_i; k_f, \sigma_f \rangle + \lim_{\delta \to 0} \sum_m \left\{ \frac{\langle \phi_f; k_f, \sigma_f | W^{(1)} | \phi_m \rangle \langle \phi_m | W^{(1)} | \phi_i; k_f, \sigma_f \rangle}{E_f - \varepsilon_m + \hbar \omega_{F} + i\delta} \right. $$

$$\left. + \frac{\langle \phi_f; k_f, \sigma_f | W^{(1)} | \phi_m; k_f, \sigma_f; k_f, \sigma_f \rangle \langle \phi_m; k_f, \sigma_f; k_f, \sigma_f | W^{(1)} | \phi_i; k_f, \sigma_f \rangle}{E_f - \varepsilon_m - \hbar \omega_{F} + i\delta} \right\} \quad (18.7)$$

We now turn our attention to evaluating and using (18.7) in certain special cases, but before we go on, it is helpful to come up with a more compact notation to keep track of what is going on. When necessary, we will refer to the three terms in (18.7) as $T_{FI} (1)$, $T_{FI} (2)$, and $T_{FI} (3)$ respectively.

E. A Diagrammatic Approach

When the expressions are getting as complicated as (18.7) it is a good idea to try to come up with an easier way to keep track of what is going on. Fortunately, there is a neat diagrammatic notation that helps us understand what is happening. In this notation, an electron (or atom) will be denoted by a solid line with an arrow on it, and a photon by a wiggly line. Time will increase from left to right, and interactions will be denoted by a dot. For example, $W^{(1)}$ can create or annihilate one photon, and there are correspondingly two corresponding diagrams, as illustrated in Fig. 18-1. In contrast, $W^{(2)}$ can either create two photons, annihilate two photons, or create and annihilate one of each. Hence there are three corresponding diagrams, as illustrated in Fig. 18-2.
An equation like (18.7) could be written in the simple diagrammatic notation as:

We see that in the first diagram, the photons are emitted and reabsorbed at the same place and time ($W(2)$), while in the other two diagrams, there are two factors of $W(1)$. By looking at the diagrams half way through we can see that the intermediate state contains only an atom in the middle diagram, but there is an atom and two photons in the intermediate state in the final diagram.

**F. Thomson Scattering**

As our first computation of this process, consider a free electron, for which the “atomic” Hamiltonian is simply given by $H_{\text{atom}} = \frac{P^2}{2m}$. The eigenstates of this Hamiltonian will be proportional to $e^{ia\mathbf{r}}$, and since we are working in finite volume, we can normalize them simply as $\phi_p(\mathbf{r}) = e^{ia\mathbf{r}}/\sqrt{V}$. Hence our initial state will be labeled $|I\rangle = |\mathbf{q}_I; \mathbf{k}_I, \sigma_I\rangle$. For definiteness, and to keep our computations simple, we will assume the initial electron is at rest, so $\mathbf{q}_I = 0$.

Which of the three terms in (18.7) will be most important? The first one, $T_{FI}(1)$, has a matrix element of $W(2) = e^2 A^2 / 2m$. The second and third term each have factors that looks like $eA \cdot P/m$ and $eB \cdot S/m$. Because we chose the initial state to have momentum zero, it is easy to see that $P$ vanishes in the second factor of each of the second and third terms. Hence we must look at $eB \cdot S/m$. The eigenvalues of $S$ are of order $\hbar$, while $B = \nabla \times A$ which will turn into a factor of $\mathbf{k} \times \mathbf{A}$, where $\mathbf{k}$ is the wave number of one of the photons. The first factor has a similar term. Noting that the energy denominators contain terms like $\hbar \omega$, we therefore estimate the order of magnitude of the second and third terms in (18.7) as

$$T_{FI}(2) \sim T_{FI}(3) \sim \frac{e^2 \hbar^2 \Lambda^2 k^2}{\hbar \omega m^2}$$

The relative size of these compared to the first term is therefore of order

$$\frac{T_{FI}(2)}{T_{FI}(1)} \sim \frac{T_{FI}(3)}{T_{FI}(1)} \sim \frac{\hbar k^2}{\hbar \omega m} = \frac{\hbar \omega}{mc^2}$$

Thus the second and third terms will be dominated by the first if the photons have an energy small compared to the rest energy of the electron. Indeed, if this is not the case,

---

1 This simplifies the computation, but does not affect the answer.
then we need to consider relativistic corrections for the electron, and our entire formalism is wrong. We therefore ignore $T_{FI} \left(2\right)$, and $T_{FI} \left(3\right)$ and have

$$\begin{align*}
T_{FI} &= T_{FI} \left(1\right) = \left\langle q_F; k_F, \sigma_F \mid W^{(2)} \mid q_i; k_i, \sigma_I \right\rangle = \frac{e^2}{2m} \left\langle q_F; k_F, \sigma_F \mid A^2 \left(R\right) \mid q_i; k_i, \sigma_I \right\rangle \\
&= \frac{e^2}{2m} \sum_{k_0} \sum_{k_0} \frac{h}{2\omega_f \sqrt{\omega_i \omega_f}} \left[ \left\langle q_F; k_F, \sigma_F \mid \left( e^{-ik_0 R} a_{k_0 \sigma} e_{k_0 \sigma} + e^{-ik_0 R} a_{k_0 \sigma}^* e_{k_0 \sigma}^* \right) \left\mid q_i; k_i, \sigma_I \right\rangle \right. \\
&\quad \times \left. \left( e^{ik_0 R} a_{k_0 \sigma}^* e_{k_0 \sigma} + e^{ik_0 R} a_{k_0 \sigma} e_{k_0 \sigma}^* \right) \left\mid q_i; k_i, \sigma_I \right\rangle \right].
\end{align*}$$

This expression is not nearly as fearsome as it appears. The only non-vanishing terms in this double sum occur when we annihilate the correct photon in the initial state, and create the correct photon in the final state. There are exactly two terms in the double sum which survive, and they are effectively identical.

$$\begin{align*}
T_{FI} &= \left( \mathbf{e}_F^* \cdot \mathbf{e}_I \right) \frac{e^2 \hbar}{2m \varepsilon_0 V \sqrt{\omega_f \omega_f}} \left\langle q_F \mid e^{-ik_0 R} a_{k_0 \sigma}^* e_{k_0 \sigma} + e^{-ik_0 R} a_{k_0 \sigma}^* e_{k_0 \sigma}^* \left\mid q_i \right\rangle \right. \\
&\quad \times \left. \left( e^{ik_0 R} a_{k_0 \sigma}^* e_{k_0 \sigma} + e^{ik_0 R} a_{k_0 \sigma} e_{k_0 \sigma}^* \right) \left\mid q_i \right\rangle \right. \\
&= \left( \mathbf{e}_F^* \cdot \mathbf{e}_I \right) \frac{e^2 \hbar}{2m \varepsilon_0 V \sqrt{\omega_f \omega_f}} \left\langle q_F \right\rangle e^{ik_0 R} a_{k_0 \sigma}^* e_{k_0 \sigma} \left\langle q_i \right\rangle.
\end{align*}$$

This last matrix element is simple to work out:

$$\begin{align*}
\left\langle q_F \mid e^{ik_0 R} a_{k_0 \sigma}^* e_{k_0 \sigma} \left\mid q_i \right\rangle &= \frac{1}{V} \int e^{-i q_{k_0} \cdot q} e^{i k_0 \cdot r} e^{-i q_{k_0} \cdot r} e^{i q_{k_0} \cdot r} d^3r = \delta_{q_{k_0} + q \sigma + k_0, q_i}.
\end{align*}$$

This matrix element does nothing more than ensure conservation of momentum. Substituting it into (18.8) and then into (18.1), and noting that squaring a Kronecker delta function has no effect, we find

$$\begin{align*}
T_{FI} &= \left( \mathbf{e}_F^* \cdot \mathbf{e}_I \right) \frac{e^2 \hbar}{2m \varepsilon_0 V \sqrt{\omega_f \omega_f}} \delta_{q_{k_0} + q \sigma + k_0, q_i}, \\
\Gamma \left( I \rightarrow F \right) &= \frac{2\pi}{\hbar} \left| \mathbf{e}_F^* \cdot \mathbf{e}_I \right|^2 \frac{e^4 \hbar^2}{4m^2 \varepsilon_0^2 V^2 \omega_f \omega_f} \delta_{q_{k_0} + q \sigma + k_0, q_i} \delta \left( E_F - E_I \right).
\end{align*}$$

Equation (18.9b) is difficult to understand as it stands, because we are still in the infinite volume limit, and because the probabilities of scattering into a particular state are small. To make sense of it, sum over all possible final quantum states; that is, sum over the wave numbers for both the electron and the photon, and summing over polarizations of the photon. Let’s also put in the explicit form for the energy difference, where the energy is just the energy of the electron plus the photon. We find

$$\begin{align*}
\Gamma \left( I \rightarrow F \right) &= \sum_{k_F, \sigma_F} \sum_{q_i} \left| \mathbf{e}_F^* \cdot \mathbf{e}_I \right|^2 \frac{\pi e^4 \hbar}{2m^2 \varepsilon_0^2 V^2 \omega_f \omega_f} \delta_{q_{k_F} + q_i + k_0, q_i} \delta \left( E_F - E_I \right) \\
&= \sum_{k_F, \sigma_F} \left| \mathbf{e}_F^* \cdot \mathbf{e}_I \right|^2 \frac{\pi e^4 \hbar}{2m^2 \varepsilon_0^2 V^2 \omega_f \omega_f} \delta \left( \hbar \omega_f + \frac{\hbar q_F^2}{2m} - \hbar \omega_I \right),
\end{align*}$$

© 2011, Eric D. Carlson 321
\[ \Gamma(I \rightarrow F) = \frac{1}{V} \sum_{\sigma_F} \int \frac{d^3k_F}{(2\pi)^3} |\epsilon_F^* \cdot \epsilon_i|^2 \frac{\pi e^4 \hbar}{2m^2 \epsilon_0^2 \omega \omega_F} \delta \left( \hbar \omega_F + \frac{\hbar q_F^2}{2m} - \hbar \omega_i \right). \quad (18.10) \]

Now, if the initial wave number is \( k_i \), conservation of momentum will tend to make the final photon and electron wave numbers comparable. We have, therefore,

\[ \frac{\hbar^2 q_F^2}{2m} \sim \frac{\hbar^2 k_i^2}{m} \sim \frac{\hbar^2 \omega_i^2}{mc^2} = \hbar \omega_i \left( \frac{\hbar \omega_i}{mc^2} \right) \ll \hbar \omega_i. \]

Hence in (18.10) we can ignore the electron energy compared to the photon energies, so \( \omega_i = \omega_F = \omega \), and we have

\[ \Gamma(I \rightarrow F) = \frac{1}{V} \frac{e^4 \hbar}{16\pi^2 m^2 \epsilon_0^2 \omega^2} \sum_{\sigma_F} \int k_F^2 dk_F \int d\Omega_k |\epsilon_F^* \cdot \epsilon_i|^2 \delta (\hbar ck_F - \hbar c k_i) \]

\[ = \frac{1}{V} \frac{e^4 k_F^2}{16\pi^2 m^2 \epsilon_0^2 \omega^2 c} \int d\Omega_k \sum_{\sigma_F} |\epsilon_F^* \cdot \epsilon_i|^2 = \frac{1}{V} \frac{\alpha^2 \hbar^2}{m^2 c^2} \int d\Omega_k \sum_{\sigma_F} |\epsilon_F^* \cdot \epsilon_i|^2, \]

where in the last step we wrote our expression in terms of the fine structure constant \( \alpha = e^2 / 4\pi \epsilon_0 c \).

One disturbing fact is that the rate is still inversely proportional to the total volume. This is simply because we are trying to collide a single photon in the universe with a single electron. Under such circumstances, the correct description is not in terms of rates, but in terms of cross section, where the rate is the cross section times the density of targets (\( 1/V \) in this case) times the relative velocity (\( c \) in this case). Hence the cross section is

\[ \sigma = \frac{V}{c} \Gamma(I \rightarrow F) = \frac{\alpha^2 \hbar^2}{m^2 c^2} \int d\Omega_k \sum_{\sigma_F} |\epsilon_F^* \cdot \epsilon_i|^2 = \frac{\alpha^2 \hbar^2}{m^2 c^2} \int d\Omega_k \sin^2 \theta = \frac{8\pi \alpha^2 \hbar^2}{3m^2 c^2}. \]

If the polarized or unpolarized differential cross-section is desired instead, we can simply not perform the final sums and integral. The angle \( \theta \) is the angle between the polarization of the incoming wave and the direction of the outgoing wave. This cross-section can be calculated classically, and is called the Thomson cross-section, but it is comforting to find the quantum result agrees with it. With a bit more work, it can be shown that this cross-section is still valid if the initial electron is moving, provided such motion is non-relativistic.

\[ \text{G. Scattering Away From a Resonance} \]

Let us now move to the other extreme. Suppose that instead of a free electron, you have bound electrons in the ground state of the atom, and that the energy of the photon is insufficient to free the electron. We’ll also assume we are far from resonance, so that the photon energy is not right, nor close to right, to excite the atom to an intermediate state, so that \( \epsilon_i + \hbar \omega_i \neq \epsilon_m \) for any intermediate state. We will furthermore
assume that the final state of the atom is identical to the initial state, so \( \omega_f = \omega_i = \omega \) and \( \phi_f = \phi_i \). We need to calculate the matrix elements in eq. (18.7), so we need

\[
T_{FI} = \langle \phi_f; k_f, \sigma_f | W^{(2)} | \phi_i; k_i, \sigma_i \rangle + \lim_{\delta \to 0^+} \sum_{m} \left\{ \frac{\langle \phi_i; k_f, \sigma_f | W^{(1)} | \phi_m \rangle \langle \phi_m | W^{(1)} | \phi_i; k_i, \sigma_i \rangle}{\varepsilon_i - \varepsilon_m + \hbar \omega + i\delta} \right. \\
+ \left. \frac{\langle \phi_f; k_f, \sigma_f | W^{(1)} | \phi_m; k_i, \sigma_i; k_f, \sigma_f \rangle \langle \phi_m; k_i, \sigma_i; k_f, \sigma_f | W^{(1)} | \phi_f; k_i, \sigma_i \rangle}{\varepsilon_i - \varepsilon_m - \hbar \omega + i\delta} \right\}. 
\tag{18.11}
\]

We start with the first term, which in a manner similar to before, yields

\[
T_{FI}(1) = \left( \mathbf{e}_F \cdot \mathbf{e}_i \right) \frac{\varepsilon_i}{2m\varepsilon_0 V} \sum_{j=1}^{N} \langle \phi_j | e^{i k_f \cdot \mathbf{R}_j - i k_i \cdot \mathbf{R}_i} | \phi_i \rangle. 
\]

Now, as we have argued before, the wavelength of light that causes transitions tends to be much larger than the size of an atom, and we are by assumption working at energies of the same order or lower, so we can approximate \( e^{i k_f \cdot \mathbf{R}_j - i k_i \cdot \mathbf{R}_i} = 1 \), and we have

\[
T_{FI}(1) = \left( \mathbf{e}_F \cdot \mathbf{e}_i \right) \frac{\varepsilon_i}{2m\varepsilon_0 V} \sum_{j=1}^{N} \langle \phi_j | \phi_i \rangle = \left( \mathbf{e}_F \cdot \mathbf{e}_i \right) \frac{\varepsilon_i N}{2m\varepsilon_0 V}. 
\tag{18.12}
\]

Now let’s look at the second term in (18.11), which, when expanded out a bit, is

\[
T_{FI}(2) = \frac{\varepsilon^2}{m^2} \lim_{\delta \to 0^+} \sum_{m} \sum_{j,i=1}^{N} \langle \phi_j; k_f, \sigma_f | \left( \mathbf{A}_j \cdot \mathbf{P}_j + \mathbf{B}_j \cdot \mathbf{S}_j \right) | \phi_m \rangle \langle \phi_m | \left( \mathbf{A}_i \cdot \mathbf{P}_i + \mathbf{B}_i \cdot \mathbf{S}_i \right) | \phi_i; k_i, \sigma_i \rangle \frac{\varepsilon_i - \varepsilon_m + \hbar \omega + i\delta}{\varepsilon_i - \varepsilon_m - \hbar \omega + i\delta}. 
\]

As we have found before, the \( \mathbf{A}_j \cdot \mathbf{P}_j \) terms tend to dominate the \( \mathbf{B}_j \cdot \mathbf{S}_j \) terms. Furthermore, in the sum over all the photon operators, only the term that annihilates the correct photon and then creates the correct photon contributes. So we find

\[
T_{FI}(2) = \frac{\varepsilon^2}{m^2} \frac{\hbar}{2\varepsilon_0 V} \lim_{\delta \to 0^+} \sum_{m} \sum_{j,i=1}^{N} \langle \phi_j | e^{-i k_f \cdot \mathbf{R}_j} e^{i k_i \cdot \mathbf{P}_i} | \phi_m \rangle \langle \phi_m | e^{i k_i \cdot \mathbf{R}_i} e^{-i k_f \cdot \mathbf{P}_j} | \phi_i \rangle. 
\]

As usual, we then approximate \( e^{i k_f \cdot \mathbf{R}_j} = 1 = e^{-i k_f \cdot \mathbf{R}_j} \), and rewrite the sums as the total momentum. We also replace \( \varepsilon_i - \varepsilon_m = -\hbar \omega_{nl} \) to yield

\[
T_{FI}(2) = -\frac{\varepsilon^2}{2\varepsilon_0 V m^2} \lim_{\delta \to 0^+} \sum_{m} \langle \phi_j | \mathbf{e}_F \cdot \mathbf{P} | \phi_m \rangle \langle \phi_m \mathbf{e}_i \cdot \mathbf{P} | \phi_i \rangle \frac{1}{\omega (\omega_{nl} - \omega - i\delta)}. 
\tag{18.13}
\]

Since we are assuming we are not near resonance, we can take the limit \( \delta \to 0 \). We rewrite the denominator as

\[
\frac{1}{\omega (\omega_{ml} - \omega)} = \frac{1}{\omega \omega_{ml}} + \frac{1}{\omega_{ml} (\omega_{ml} - \omega)}. 
\]

Substituting this into (18.13), we have
\[ T_{F_1} (2) = -\frac{e^2}{2\varepsilon_0 V m^2} \sum_{m} \left\langle \phi_I \left| e_F^\ast \cdot P \right| \phi_m \right\rangle \left\langle \phi_m \left| e_J \cdot P \right| \phi_I \right\rangle \left( \frac{1}{\omega} + \frac{1}{\omega_{ml} - \omega} \right). \]

We now use the same trick we found in equation (15.27), \( \left\langle \phi_m \left| P \right| \phi_I \right\rangle = i m \omega_{ml} \left\langle \phi_m \left| R \right| \phi_I \right\rangle \), but for the first term we use this only once, while for the second we use it twice, to yield

\[ T_{F_1} (2) = \frac{e^2}{2\varepsilon_0 V} \sum_{m} \left\langle \phi_I \left| e_F^\ast \cdot R \right| \phi_m \right\rangle \left( i \frac{\left\langle \phi_m \left| e_J \cdot P \right| \phi_I \right\rangle}{m \omega} - \frac{\omega_{ml} \left\langle \phi_m \left| e_J \cdot R \right| \phi_I \right\rangle}{\omega_{ml} - \omega} \right). \]

In the first term, we now note that we have a complete sum over intermediate states, and no other factors depending on the intermediate state. On the final term we rewrite \( \left\langle \phi_m \left| R \right| \phi_I \right\rangle = r_{ml} \) so we can rewrite this term in a nicer form to yield

\[ T_{F_1} (2) = \frac{i e^2}{2 m \varepsilon_0 V \omega} \left\langle \phi_I \left| \left( e_F^\ast \cdot R \right) \left( e_J \cdot P \right) \right| \phi_I \right\rangle - \frac{e^2}{2 \varepsilon_0 V} \sum_{m} \frac{\omega_{ml} \left( e_F^\ast \cdot r_{ml} \right) \left( e_J \cdot r_{ml} \right)}{\omega_{ml} - \omega}. \quad (18.14) \]

We have two terms done; we still have one to go. In a manner similar to before, we keep only those terms that create the final state photon, and then annihilate the initial state photon. We again ignore the exponential factors, and combine the momenta to obtain the total momentum. We eventually arrive at an equation analogous to (18.13):

\[ T_{F_1} (3) = -\frac{e^2}{2 \varepsilon_0 V m^2} \lim_{\delta \to 0} \sum_{m} \left\langle \phi_I \left| e_J \cdot P \right| \phi_m \right\rangle \left\langle \phi_m \left| e_F^\ast \cdot P \right| \phi_I \right\rangle \frac{\omega_{ml}}{\omega \left( \omega_{ml} + \omega - i \delta \right)}. \]

Notice the denominator is slightly different, since in this case we have an extra photon in the intermediate state. We again expand out the denominator and then use the same trick we always do for turning momentum matrix elements into position matrix elements, so

\[ T_{F_1} (3) = \frac{e^2}{2 \varepsilon_0 V} \sum_{m} \left( -\frac{i}{m \omega} \left\langle \phi_I \left| e_J \cdot P \right| \phi_m \right\rangle + \frac{\omega_{ml}}{\omega_{ml} + \omega} \left\langle \phi_I \left| e_F^\ast \cdot R \right| \phi_m \right\rangle \right) \left\langle \phi_m \left| e_F^\ast \cdot R \right| \phi_I \right\rangle \]

\[ = -\frac{i e^2}{2 m \varepsilon_0 V \omega} \left\langle \phi_I \left| \left( e_J \cdot P \right) \left( e_F^\ast \cdot R \right) \right| \phi_I \right\rangle + \frac{e^2}{2 \varepsilon_0 V} \sum_{m} \frac{\omega_{ml} \left( e_J \cdot r_{ml} \right) \left( e_F^\ast \cdot r_{ml} \right)}{\omega_{ml} + \omega}. \quad (18.15) \]

Now that we have the three pieces, all we need to do is put them together. To simplify matters slightly, I’ll assume we are working with real polarizations or real dipole moments \( r_{ml} \). Combining (18.12), (18.14), and (18.15), we have

\[ T_{F_1} = T_{F_1} (1) + T_{F_1} (2) + T_{F_1} (3) \]

\[ = \frac{e^2 \hbar N}{2 m \varepsilon_0 V \omega} \left( e_F^\ast \cdot e_J \right) + \frac{i e^2}{2 m \varepsilon_0 V \omega} \left\langle \phi_I \left| \left( e_F^\ast \cdot R \right) \left( e_J \cdot P \right) - \left( e_J \cdot P \right) \left( e_F^\ast \cdot R \right) \right| \phi_I \right\rangle \]

\[ + \frac{e^2}{2 \varepsilon_0 V} \sum_{m} \left( e_F^\ast \cdot r_{ml} \right) \left( e_J \cdot r_{ml} \right) \left( \frac{\omega_{ml}}{\omega_{ml} + \omega} - \frac{\omega_{ml}}{\omega_{ml} - \omega} \right). \quad (18.16) \]

Now, in the second term, we note that we have a commutator. Keeping in mind that we have several electrons, this commutator can be worked out to yield
\[
\left[ \varepsilon_F^* \cdot R, \varepsilon_i \cdot P \right] = \sum_{j=1}^{N} \sum_{i=1}^{N} \left[ \varepsilon_F^* \cdot R_j, \varepsilon_i \cdot P_i \right] = \sum_{j=1}^{N} \sum_{i=1}^{N} i \hbar \delta_{\mu j} \left( \varepsilon_F^* \cdot \varepsilon_i \right) = N \hbar \left( \varepsilon_F^* \cdot \varepsilon_i \right).
\]

Substituting this into (18.16), we find the first two terms cancel exactly, so we have

\[
T_{FI} = -\frac{e^2}{\epsilon_0 V} \sum_m \frac{\omega \omega_{ml}}{\omega_{ml}^2 - \omega^2} \left( \varepsilon_F^* \cdot \mathbf{r}_{ml} \right) \left( \varepsilon_i \cdot \mathbf{r}_{ml} \right). \tag{18.17}
\]

We can then substitute this into (18.1) to obtain a rate per unit volume, which in turn can be converted into a cross-section.

Let us compare the size of the amplitude (18.17) we found for a bound electron to that for a free electron, (18.9a). Ignoring matching factors, and making only order of magnitude estimates, we find

\[
\frac{T_{FI} (\text{bound})}{T_{FI} (\text{free})} \sim \frac{m \omega^2 |\mathbf{r}_{ml}|^2}{\hbar (\omega_{ml}^2 - \omega^2)}.
\]

For a typical atom, we would expect \( \omega_{ml} \sim mc^2 \alpha^2 / \hbar \) and \( |\mathbf{r}_{ml}| \sim (\hbar / mc) \), so we have

\[
\frac{T_{FI} (\text{bound})}{T_{FI} (\text{free})} \sim \frac{\omega^2}{\omega_{ml}^2 - \omega^2}.
\]

Suppose that we have tightly bound atom, so that \( \omega / \omega_{ml} \ll 1 \). Then this ratio is of order \( (\omega / \omega_{ml})^2 \), and this gets squared to \( (\omega / \omega_{ml})^4 \) when you calculate the rate (which ultimately becomes a cross-section). Hence a tightly bound atom has a much lower cross-section than a free electron for photon scattering.

This has considerable significance in the early universe. The early universe was at very high temperatures, hot enough that electrons were not generally bound into atoms, but rather free. This meant that the early universe was opaque, and photons effectively were in near perfect equilibrium. However, when the temperature of the universe dropped to about 3000 K, at an age of about 380,000 years, the electrons became bound to the free protons to produce neutral hydrogen atoms. Since the typical photon had an energy of about \( 3k_B T \approx 0.8 \text{ eV} \), and the first excited state of hydrogen requires about 10.2 eV of energy, the cross-section dropped suddenly by several orders of magnitude, and the universe became transparent. These photons have since been red-shifted in the expanding universe, and now appear as a nearly uniform 2.73 K background. This background gives us a snapshot of what the universe looked like at \( t = 380,000 \text{ y} \).

Our derivation of (18.17) assumed that we were working away from resonance. However, as \( \omega \rightarrow \omega_{ml} \), it is clear from (18.17) that something dramatic may occur, since there will be a sudden increase in the amplitude. Under such resonance conditions, we need to rethink how we are computing our amplitude. This complication is the subject of the next section.

It should be understood that, as usual, we have included only the dipole approximation in our scattering amplitude. This is generally a good approximation, since we are summing over all intermediate states, and those which have non-zero dipole
moment will tend to dominate. But if we approach resonance, so \( \omega \approx \omega_{ml} \) for some state that does not have a dipole moment, we will have to include “smaller” effects like the electric quadrupole or magnetic dipole. These small matrix elements will be greatly enhanced by the nearly vanishing denominators.

H. Scattering Near a Resonance

It is evident from equation (18.17) that our amplitudes may become large when \( \omega \approx \omega_{ml} \). The large amplitude can be traced back to the contribution from \( T_{Fl} (2) \), corresponding to the diagram sketched above. It is easy to see why this term dominates: the intermediate state has almost exactly the right energy to match the incoming energy. Indeed, we will focus exclusively on this diagram, and only include one term in the sum, the one which has a small energy denominator. The amplitude for this is given by (18.14), repeated here, but with the standard substitution

\[
\langle \phi_m | p | \phi_f \rangle = i m \omega_{ml} r_{ml}.
\]

\[
T_{Fl}^{(0)} = \frac{e^2 \omega_{ml}^2}{2 e_0 V} \lim_{\delta \rightarrow 0} \frac{(\epsilon_F^* \cdot r_{ml}^*) (\epsilon_{\ell} \cdot r_{ml})}{(\omega - \omega_{ml} + i\delta)}.
\] (18.18)

My notation has changed slightly; I have dropped the (2), because when we are close to resonance, this will effectively be the only term we need to worry about, and the (0) is because in a moment we will be adding higher and higher order terms to try to figure out what is going on. This diagram has zero loops in it.

If we were to continue with this expression, we would ultimately obtain a cross section that diverges at \( \omega = \omega_{ml} \). What is the solution to this problem? The answer turns out, somewhat surprisingly, to add more diagrams. Consider now the one-loop diagram sketched at right. In this diagram, the ground state atom merges with a photon to produce the resonant excited state, then it splits into an intermediate photon of wave number \( k \) and the ground state again, which then remerges to produce the excited state before splitting back into a photon and the ground state.

Fortunately, nearly all the work has already been done, and we simply reuse it. The two new vertices (in the middle) are virtually identical to the ones we had before, and we now have two intermediate states that are purely the resonant state. The main new feature is an energy denominator, which looks like \( \epsilon_{\ell} + \hbar \omega - \epsilon_{\ell} - \hbar \omega_k + i\delta \), so we have

\[
T_{Fl}^{(1)} (k, \sigma) = \left( \frac{e^2 \omega_{ml}^2}{2 e_0 V} \right)^2 \lim_{\delta \rightarrow 0} \frac{(\epsilon_F^* \cdot r_{ml}^*) (\epsilon_{\ell} \cdot r_{ml}^*) (\epsilon_{\ell} \cdot r_{ml})}{(\omega - \omega_{ml} + i\delta)^2 (\hbar \omega - \hbar \omega_k + i\delta)},
\]
\[ T^{(1)}_{F1}(k, \sigma) = T^{(0)}_{F1} \left( \frac{e^2 \omega_{ml}^2}{2\epsilon_0 V \omega} \right) \frac{(\epsilon_{k\sigma} \cdot r_{ml}) (\epsilon_{k\sigma} \cdot r_{ml}^*)}{(\omega - \omega_{ml} + i\delta)(\hbar \omega - \hbar \omega_k + i\delta)}. \]

This is the amplitude from this diagram for a particular intermediate momentum and polarization. Of course, we don’t measure this intermediate state, so we must actually sum over all such intermediate states. We then turn the sum into an integral in the infinite volume limit in the usual way, and have

\[ T^{(1)}_{F1} = T^{(0)}_{F1} \frac{1}{\omega - \omega_{ml} + i\delta} \frac{e^2 \omega_{ml}^2}{2\epsilon_0 \omega \hbar} \sum_\sigma \int \frac{d^3k}{(2\pi)^3} \frac{|\epsilon_{k\sigma} \cdot r_{ml}|^2}{\omega - \omega_k + i\delta}. \]  

(18.19)

Note that there is an implied \( \lim_{\delta \to 0^+} \) in \( T^{(0)}_{F1} \), which applies to every factor in (18.19).

Now, we will be interested in the integral appearing in (18.19). We will split it into a real and imaginary part, which we write as

\[ \Delta(\omega) - \frac{i}{2} \Gamma(\omega) \equiv \frac{e^2 \omega_{ml}^2}{2\epsilon_0 \omega \hbar} \sum_\sigma \int \frac{d^3k}{(2\pi)^3} \frac{|\epsilon_{k\sigma} \cdot r_{ml}|^2}{\omega - \omega_k + i\delta}, \]

(18.20)

where \( \Delta(\omega) \) and \( \Gamma(\omega) \) are real functions of \( \omega \). Switching to spherical coordinates, and remembering that \( k = \omega_k / c \), and substituting the fine structure constant \( \alpha = e^2 / 4\pi\epsilon_0 \hbar c \), we have

\[ \Delta(\omega) - \frac{i}{2} \Gamma(\omega) = \sum_\sigma \int d\Omega_k \int^\infty_0 \omega_k^2 d\omega_k \frac{\alpha \omega_{ml}^2 |\epsilon_{k\sigma} \cdot r_{ml}|^2}{4\pi^2 c^2 \omega (\omega - \omega_k + i\delta)}. \]  

(18.21)

At the moment, I am not particularly interested in the real part. Indeed, if you continue the computation, you will discover that it is divergent, but this divergence is only present because we approximate the phase factors as \( e^{ik \cdot r} = 1 \) in all our matrix elements. The infinity is coming from the high frequency modes, when this approximation breaks down, and one would ultimately find some finite contribution to \( \Delta(\omega) \). In the limit \( \delta \to 0^+ \), the integral is zero almost everywhere, except for a small region right around \( \omega_k = \omega \).

We can therefore concentrate on this small region to find the imaginary part. Let us take the limit \( \delta \to 0^+ \), but let’s distort the path of integration in this region so that we avoid the divergence. The \( i\delta \) tells us how to do this: We want to keep the combination \( \omega - \omega_k + i\delta \) positive imaginary, which we do by letting \( \omega_k \) have a small negative imaginary part, as illustrated in Fig. 18-3. We can now let \( \delta = 0 \) in the denominator, but for the little hop below the real axis, we let \( \omega_k = \omega - \delta e^{i\theta} \).

Thus as \( \theta \) progresses from 0 to \( \pi \), the little half-loop will be followed. We will treat \( \delta \) as so small that effectively it equals zero, except in the critical denominator. Focusing on the

**Figure 18-3**: The path for integration for Eq. (18.21). The only part that contributes to the imaginary part is the little “hop” below that real axis near \( \omega_k = \omega \).
imaginary part, we substitute \( \omega_k = \omega - \delta e^{i\theta} \) into (18.21), and find

\[
\Gamma(\omega) = -2 \text{Im} \left\{ \sum_{\sigma} d\Omega_k \int_{\omega = 0}^{\pi} (\omega - \delta e^{i\theta})^2 d(\omega - \delta e^{i\theta}) \frac{\alpha \omega_{ml}^2 |\mathbf{e}_{k\sigma} \cdot \mathbf{r}_{ml}|^2}{4\pi^2 c^2 \omega (\omega - \omega + \delta e^{i\theta})} \right\}
\]

\[
= 2 \text{Im} \left\{ \sum_{\sigma} d\Omega_k \frac{\alpha \omega_{ml}^2 |\mathbf{e}_{k\sigma} \cdot \mathbf{r}_{ml}|^2}{4\pi^2 c^2 \omega} \int_{\omega = 0}^{\pi} i \delta e^{i\theta} d\theta \right\} = \frac{\alpha \omega_{ml}^2 \omega}{2\pi c^2} \sum_{\sigma} d\Omega_k |\mathbf{e}_{k\sigma} \cdot \mathbf{r}_{ml}|^2
\]

\[
= \frac{\alpha \omega_{ml}^2 \omega |\mathbf{r}_{ml}|^2}{2\pi c^2} \int \sin^2 \theta_k d\Omega_k = \frac{4 \alpha \omega_{ml}^2 \omega |\mathbf{r}_{ml}|^2}{3c^2}.
\]

(18.22)

Compare the result (18.22) with eq. (18.5). It is virtually identical. It is something like a frequency dependent decay rate for the intermediate state \(|\phi_m\rangle\). Indeed, if we are working near resonance, then they are identical.

Substituting (18.20) into (18.19) will then yield

\[
T_F^{(1)} = T_F^{(0)} \frac{\Delta(\omega) - \frac{i}{2} \Gamma(\omega)}{\omega - \omega_{ml} + i\delta}.
\]

(18.23)

We don’t know what \( \Delta(\omega) \) exactly is, but we won’t need it in the subsequent discussion, but we do know the explicit form of \( \Gamma(\omega) \). However, all we need to know for now is that they are both finite.

Now, as we look at (18.23), it seems that our situation is worse than ever. Recall that our original amplitude \( T_F^{(0)} \) given by (18.18) diverges at resonance. The new term (18.23) has yet another denominator, and diverges even worse. Let’s try adding two loops and see what we get, as sketched at right. This time we can write down the answer instantly: The effect of adding another loop is simply to put yet another factor similar to the one in (18.23). The amplitude, we immediately see, is given by

\[
T_F^{(2)} = T_F^{(0)} \left[ \frac{\Delta(\omega) - \frac{i}{2} \Gamma(\omega)}{\omega - \omega_{ml} + i\delta} \right]^2.
\]

Now that we’ve established the pattern, we can continue indefinitely. We are most interested in the sum of all these diagrams, which looks like

\[
T_F = T_F^{(0)} + T_F^{(1)} + T_F^{(2)} + \ldots = T_F^{(0)} \left\{ 1 + \frac{\Delta(\omega) - \frac{i}{2} \Gamma(\omega)}{\omega - \omega_{ml} + i\delta} + \left[ \frac{\Delta(\omega) - \frac{i}{2} \Gamma(\omega)}{\omega - \omega_{ml} + i\delta} \right]^2 + \ldots \right\}.
\]

This is nothing more than a geometric series. Summing it and substituting our explicit form (18.18) for \( T_F^{(0)} \), we have

328 © 2011, Eric D. Carlson
\[
T_{FI} = T_{FI}^{(0)} \sum_{n=0}^{\infty} \left[ \frac{\Delta(\omega) - \frac{1}{2} i \Gamma(\omega)}{\omega - \omega_{ml} + i \delta} \right] \]
\[
= e^2 \omega_{ml}^2 \lim_{\delta \to 0^+} \frac{(\mathbf{e}_F \cdot \mathbf{r}_{ml})(\mathbf{e}_i \cdot \mathbf{r}_{ml})}{\omega - \omega_{ml} + i \delta} \frac{\omega - \omega_{ml} + i \delta}{\omega - \omega_{ml} + i \delta - \left[ \Delta(\omega) - \frac{1}{2} i \Gamma(\omega) \right]} \]
\[
= e^2 \omega_{ml}^2 \lim_{\delta \to 0^+} \frac{(\mathbf{e}_F \cdot \mathbf{r}_{ml})(\mathbf{e}_i \cdot \mathbf{r}_{ml})}{\omega - \omega_{ml} + \Delta(\omega) + \frac{1}{2} i \Gamma(\omega) + i \delta} \cdot (18.24) \]

Now, as if by miracle, we note that we can take the limit $\delta \to 0^+$ with impunity. The decay rate $\Gamma(\omega)$ makes the denominator finite. We also now understand the meaning of $\Delta(\omega)$: it is a shift of the intermediate energy for the state $|\phi_m\rangle$.

We would like to go ahead and calculate the scattering cross-section from (18.24). Substituting into (18.1), we have

\[
\Gamma(I \to F) = \frac{2\pi}{\hbar} \left( \frac{e^2 \omega_{ml}^2}{2e_0 V^2} \right)^2 \sum_{k, \sigma} \frac{|\mathbf{e}_F \cdot \mathbf{r}_{ml}|^2 |\mathbf{e}_i \cdot \mathbf{r}_{ml}|^2}{\omega - \omega_{ml} - \Delta(\omega) + \frac{1}{4} \Gamma^2(\omega)} \delta(E_F - E_i) \]

Let's assume we're working near resonance. We'll neglect the small shift in the energy $\Delta(\omega)$. We also want to sum over final state polarizations and outgoing final state directions. Then we have

\[
\Gamma(I \to F) = \frac{2\pi}{\hbar} e^2 \omega_{ml}^2 \sum_{k, \sigma} \frac{|\mathbf{e}_F \cdot \mathbf{r}_{ml}|^2 |\mathbf{e}_i \cdot \mathbf{r}_{ml}|^2}{(2\pi)^3 (\omega - \omega_{ml})^2 + \frac{1}{4} \Gamma^2(\omega)} \delta(E_F - E_i) \]
\[
= \frac{\pi e^4 \omega_{ml}^4 |\mathbf{e}_i \cdot \mathbf{r}_{ml}|^2}{2\hbar e_0 V^2 \omega^2} \sum_{\sigma} \int \frac{d^3k}{(2\pi)^3} \frac{|\mathbf{e}_F \cdot \mathbf{r}_{ml}|^2}{(\omega - \omega_{ml})^2 + \frac{1}{4} \Gamma^2(\omega)} \delta(h\omega_F - h\omega) \]
\[
= \frac{e^4 \omega_{ml}^4 |\mathbf{e}_i \cdot \mathbf{r}_{ml}|^2}{16\pi^2 h^2 e_0^2 V^2 \omega^2} \sum_{\sigma} \int d\Omega_k \frac{|\mathbf{e}_F \cdot \mathbf{r}_{ml}|^2}{(\omega - \omega_{ml})^2 + \frac{1}{4} \Gamma^2(\omega)} \delta(h\omega_F - h\omega) \]
\[
= \frac{\alpha^2 \omega_{ml}^4 |\mathbf{e}_i \cdot \mathbf{r}_{ml}|^2}{Vc \left[ (\omega - \omega_{ml})^2 + \frac{1}{4} \Gamma^2(\omega) \right]} \sum_{\sigma} \int d\Omega_k |\mathbf{e}_F \cdot \mathbf{r}_{ml}|^2 = \frac{8\pi\alpha^2 \omega_{ml}^4 |\mathbf{e}_i \cdot \mathbf{r}_{ml}|^2 |\mathbf{r}_{ml}|^2}{3Vc \left[ (\omega - \omega_{ml})^2 + \frac{1}{4} \Gamma^2(\omega) \right]}.
\]

Because decay rates tend to be rather small, the denominator will have a dramatic increase right at resonance. Note that the rate is still dependant on $V$, which is hardly surprising, since we are trying to collide one photon with one atom. We need to convert this to a cross section, which we do in the usual way by using the formula $\Gamma = n\sigma |\Delta\mathbf{v}|$, where $n = 1/V$ is the density of target atoms, and $|\Delta\mathbf{v}| = c$ is the relative velocity. We find

\[
\sigma(\omega) = \frac{8\pi\alpha^2 |\mathbf{r}_{ml}|^2 |\mathbf{e}_i \cdot \mathbf{r}_{ml}|^2}{3c^2 \left[ (\omega - \omega_{ml})^2 + \frac{1}{4} \Gamma^2(\omega) \right]}.
\]
If we treat $\Gamma(\omega)$ as a constant, this shape is called a *Lorentzian line shape*.

Now, in all our computations, we have been assuming there is only one possible intermediate state, and sometimes this is the case. More typically, however, there will be several degenerate intermediate states, related to each other by rotational symmetry. Normally we can choose one of them to be the state with the dipole moment in the direction of the polarization of the incoming light, so that $|\mu| = \mu_m$, and the others will be perpendicular to the dipole moment, and we have

$$\sigma(\omega) = \frac{8\pi\alpha^2 \omega_m^4 |r_m|^4}{3c^2 \left[ (\omega - \omega_m)^2 + \frac{1}{4} \Gamma^2(\omega) \right]}.$$  \hspace{1cm} (18.25)

If we perform the scattering right on resonance, at $\omega = \omega_m$, and use our explicit form (18.22) for the rate, we have

$$\sigma(\omega_m) = \frac{8\pi\alpha^2 \omega_m^4 |r_m|^4}{3c^2 \left[ \frac{1}{4} \alpha \omega_m^3 |r_m|^2 / c^2 \right]^2} = \frac{3\pi c^2}{2\omega_m^2}.$$  \hspace{1cm} (18.26)

Notice that all the factors of the coupling, matrix elements, and so on, have disappeared. We have simplified the discussion by assuming that the state $|\phi_m\rangle$ can only decay to the state $|\phi\rangle$. When this assumption is false, there will be an additional factor representing essentially the fraction of the time that $|\phi_m\rangle$ decays to $|\phi\rangle$. Nonetheless, the cross-section (18.26) is *enormous*; it is of the order of the wavelength squared, which will be tens of thousands of times greater than scattering off of a free electron. This is the power of resonant scattering.

One way to think of this process is that the atom is actually absorbing the photon, staying in an excited state for a brief time, and then re-emitting the photon. Over what range of energies will this absorption occur? We can see from (18.25) that the cross section will be of order its maximum value if $|\omega - \omega_m| < \frac{1}{2} \Gamma$. This means that the energy is “wrong” by an approximate quantity $\Delta E \sim \frac{1}{2} \hbar \Gamma$. The atom will “hang onto” the energy of the photon for a typical time $\Delta t \approx \Gamma^{-1}$. Multiplying these two quantities yields the time-energy uncertainty relationship, $\Delta E \Delta t \sim \frac{1}{2} \hbar$. 


Problems for Chapter 18

1. An electron is trapped in a 3D harmonic oscillator potential, $H = \mathbf{P}^2/2m + \frac{1}{2}m\omega^2\mathbf{R}^2$. It is in the quantum state $|n_x, n_y, n_z\rangle = |2,1,0\rangle$

   (a) Calculate every non-vanishing matrix element of the form $\langle n_x', n_y', n_z'|\mathbf{R}|2,1,0\rangle$ where the final state is lower in energy than the initial state.

   (b) Calculate the decay rate $\Gamma(210 \rightarrow n_x', n_y', n_z')$ for this decay in the dipole approximation for every possible final state.

2. A hydrogen atom is initially in a 3d state, specifically, $|n, l, m\rangle = |3,2,+2\rangle$.

   (a) Find all non-zero matrix elements of the form $\langle n', l', m'|\mathbf{R}|3,2,+2\rangle$, where $n' < n$.

   Which state(s) will it decay into?

   (b) Calculate the decay rate in s$^{-1}$.

3. An electron is trapped in a 3D harmonic oscillator potential, $H = \mathbf{P}^2/2m + \frac{1}{2}m\omega^2\mathbf{R}^2$. It is in the quantum state $|n_x, n_y, n_z\rangle = |0,0,2\rangle$. It is going to decay directly into the ground state $|0,0,0\rangle$.

   (a) Convince yourself that it cannot go there via the electric dipole transition. It can, however, go there via the electric quadrupole transition.

   (b) Calculate every non-vanishing matrix element of the form $\langle 0,0,0|\mathbf{R}_x\mathbf{R}_y|0,0,2\rangle$.

   (b) Calculate the polarized differential decay rate $d\Gamma_{\text{pol}}(002 \rightarrow 000)/d\Omega$ for this decay. This will require, among many other things, converting a sum to an integral in the infinite volume limit.

   (c) Sum it over polarizations and integrate it over angles to determine the total decay rate $\Gamma(002 \rightarrow 000)$.

4. The present day density of electrons is about 0.21/m$^3$; however, these electrons are not all free.

   (a) Calculate the cross-section for free photon-electron scattering in m$^2$. If all the electrons were free, what would be the rate at which they scatter, today? How many scatterings would have occurred to an average photon in the age of the universe, 13.7 Gyr?

   (b) The last time the electrons were free was when the universe was 1092 times smaller in all three directions, and it was only 380,000 years old. The number of electrons was about the same then (though the number density was much higher, since the universe was smaller). Redo part (a) at this time.
5. An electron is trapped in a 3D harmonic oscillator potential, \( H = \frac{P^2}{2m} + \frac{1}{2} m \omega^2 R^2 \). It is in the ground state \( |n_x, n_y, n_z\rangle = |0, 0, 0\rangle \). Photons are fired at the electron with frequency \( \omega \ll \omega_0 \), going in the z-direction, and polarized in the x-direction, \( \varepsilon = \hat{x} \). Calculate the differential cross section \( d\sigma/d\Omega \) (summed over final polarizations, not integrated), and the total cross section \( \sigma \).

6. A Hydrogen atom in the 1s-state is hit by photons polarized in the z-direction very close to \( \frac{1}{4} \omega_{\text{Kyd}} \); that is, quite close to the 2p-resonance, where \( \omega_{\text{Kyd}} = \alpha^2 mc^2 / 2h \) is the binding frequency of the ground state of Hydrogen.
   (a) Find all non-zero components of the matrix elements \( \langle 210 | R | 100 \rangle \). Find the decay rate \( \Gamma(\omega) \) for this state. \( \Gamma(\omega) \) should be proportional to \( \omega \); it should be \( \omega \) times a pure number.
   (b) Calculate a formula for the cross-section as a function of \( \omega \). Assume that only one diagram (the one calculated in class) is relevant, the one that has a resonance at the 2p state. Treat the energy shift \( \Delta(\omega) \) as zero.
   (c) Sketch the cross-section in units of \( a_0^2 \) as a function of \( \omega \) in the range 0.74 – 0.76 \( \omega_{\text{Kyd}} \). I recommend making the vertical scale logarithmic.
Appendix A. Some Mathematical Tools of Quantum Mechanics

A. Calculus with One Variable

Quantum Mechanics takes extensive advantage of calculus, and you should familiarize yourself with techniques of calculus. The two basic functions of calculus are the derivative and the integral. The derivative of a function of one variable \( f(x) \) can be written as \( df(x)/dx \) or \( f'(x) \). Finding derivatives of most functions is not difficult. The following table gives derivatives of the most common simple functions

\[
\begin{align*}
\frac{d}{dx}c &= 0, \\
\frac{d}{dx}x^n &= nx^{n-1}, \\
\frac{d}{dx}e^x &= e^x, \\
\frac{d}{dx}\ln x &= \frac{1}{x}, \\
\frac{d}{dx}\sin x &= \cos x, \\
\frac{d}{dx}\cos x &= -\sin x, \\
\frac{d}{dx}\tan x &= \sec^2 x, \\
\frac{d}{dx}\tan^{-1} x &= \frac{1}{1+x^2}.
\end{align*}
\]

Notice that many other functions can be rewritten in terms of these functions; for example, \( \sqrt{x} = x^{1/2} \) and \( 1/x = x^{-1} \). In addition, there are rules for taking the derivatives of sums, differences, products, quotients, and composition of functions:

\[
\begin{align*}
\frac{d}{dx}[f(x) \pm g(x)] &= f'(x) \pm g'(x), \\
\frac{d}{dx}[f(x)g(x)] &= f'(x)g(x) + f(x)g'(x), \\
\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] &= \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}, \\
\frac{d}{dx}[f(g(x))] &= f'(g(x))g'(x).
\end{align*}
\]

As a special case, note that when a function has a constant added to it, the constant has no effect; when a function has a constant multiplied by it, the constant simply multiplies the derivative as well.

The other thing you do in the calculus of one variable is to integrate. There are two types of integrals, definite and indefinite. A definite integral has limits of integration and is defined as the area under a specific curve between two limits. It looks like this:

\[
\int_a^b f(x) \, dx.
\]
The other type of integral is an indefinite integral. It uses the same symbol, except that no limits are specified, and it is defined as the anti-derivative; that is, the function whose derivative is \( f(x) \). In other words, if we let \( F(x) \) be the indefinite integral of \( f(x) \), then

\[
F(x) = \int f(x) \, dx \quad \Rightarrow \quad F'(x) = f(x).
\]

As we already mentioned, if you add a constant to a function \( F \), its derivative is unchanged, so that the indefinite integral is ambiguous up to an additive constant. Hence, in general, the indefinite integrals should always end with \( +C \), an unknown constant. The fundamental theorem of calculus tells us that these two types of integrals are related. Specifically, if the indefinite integral of \( f(x) \) is \( F(x) \), then

\[
\int_a^b f(x) \, dx = \left[ F(x) \right]_a^b = F(b) - F(a). \tag{A.1}
\]

Because of the subtraction, the constant of integration cancels out, and therefore is unnecessary and irrelevant when calculating an indefinite integral. Note that if you ever have to exchange the two limits of an integral, the resulting integral changes sign, as is clear from (A.1). Also note that in any definite integral, the variable being integrated (\( x \) in this case) disappears after the substitution of the limits, and hence can be replaced by any other unused variable with impunity.

Unlike derivatives, there are no simple rules for doing integrals. Generally you use a few steps to try to convert your integral into smaller, more manageable pieces, and then either look up the integral in an integral table, or use some tool like Maple to do the integral. Two of the simple rules that allow you to convert complicated integrals into simpler ones are

\[
\int c f(x) \, dx = c \int f(x) \, dx,
\]

\[
\int [f(x) \pm g(x)] \, dx = \int f(x) \, dx \pm \int g(x) \, dx.
\]

In addition, it is possible to change variables. Suppose we have an integral and we wish to change variables \( x = g(y) \). Then we can substitute this integral in to find

\[
\int f(x) \, dx = \int f(g(y)) d[g(y)] = \int f(g(y)) g'(y) \, dy.
\]

Note that if you are dealing with an indefinite integral, you will have to substitute back in to convert to the original variable \( x \) using \( y = g^{-1}(x) \). If you are doing a definite integral, this may not be necessary, but you will have to redo the limits of integration appropriately.

Once the equation has been massaged into a relatively well-controlled form, you will often be left with some integrals that you need to complete. I recommend looking them up in an integral table or using a program like Maple. For example, suppose you have the surprisingly difficult integral \( \int_{-\infty}^{\infty} \exp(-Ax^2) \, dx \). This integral is defined only if \( A \) is positive, or at least has a positive real part; otherwise, the integral is undefined. To
determine the integral, you must tell Maple that this constant is positive. To find this integral, open Maple, and then enter the commands

\[
\texttt{assume(A>0); integrate(exp(-A*x^2),x=-infinity..infinity);}
\]

and it will come back with the answer

\[
\int_{-\infty}^{\infty} \exp(-Ax^2) \, dx = \sqrt{\frac{\pi}{A}}.
\]

A very useful formula from the calculus of one variable is the Taylor expansion, which says that a function in the neighborhood of any point \(a\) can be expanded in terms of the value of the function and its derivatives at the point \(a\). The Taylor expansion is

\[
f(x) = f(a) + (x-a)f'(a) + \frac{1}{2!}(x-a)^2 f''(a) + \frac{1}{3!}(x-a)^3 f'''(a) + \cdots,
\]

where \(n!\) is defined as \(1 \cdot 2 \cdot 3 \cdots n\). This formula will strictly speaking be true only if the function \(f\) is sufficiently smooth between \(a\) and \(x\), and if the sum converges. If the sum does not converge, a finite sum often will still serve as an excellent approximation for the function, though it will not be perfect.

The Taylor expansion comes up in lots of situations, but it is most commonly used for a power of \(1 + x\), for \(e^x\), and for \(\cos x\) and \(\sin x\). For these functions, the Taylor expansion gives

\[
(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \cdots,
\]

\[
e^x = \exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots,
\]

\[
\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots,
\]

\[
\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots.
\]

The notation \(\exp(x)\) is an alternate notation for \(e^x\), useful because we will often make the argument of this function quite complicated, and it is easier to read, for example, \(\exp\left(-\frac{1}{2}Ax^2\right)\) then it is to read \(e^{-Ax^2}\). (A.2a) is convergent for any complex number \(|x| < 1\); the other three converge for all complex \(x\). Another nice feature about the equations (A.2) is that if we replace \(x\) with any square matrix, the formulas still make sense, so we can now take exponentials of matrices, for example. Euler’s theorem, which we will use extensively, can be derived from (A.2b), (A.2c), and (A.2d):

\[
e^{ix} = \exp(ix) = \cos x + i \sin x.
\]
B. Calculus with Several Variables

In math we often work in one dimension; in physics, at least when we are doing realistic calculations, there are at least three. A scalar function in three dimensions would be a single ordinary function like \( f(\mathbf{r}) = f(x, y, z) \). A vector function in three dimensions would actually be a triplet of functions \( \mathbf{A}(\mathbf{r}) \), with

\[
\mathbf{A}(\mathbf{r}) = \mathbf{\hat{x}} A_x(\mathbf{r}) + \mathbf{\hat{y}} A_y(\mathbf{r}) + \mathbf{\hat{z}} A_z(\mathbf{r}),
\]

where \( \mathbf{\hat{x}} \), \( \mathbf{\hat{y}} \), and \( \mathbf{\hat{z}} \) are unit vectors in the directions of increasing \( x \), \( y \), and \( z \) respectively. We can then take the derivatives of scalar functions or the components of vector functions using partial derivative, which we write as \( \partial/\partial x \), \( \partial/\partial y \), and \( \partial/\partial z \).\(^{1}\) It is helpful to define a sort of vector derivative, which we denote as

\[
\nabla \equiv \mathbf{\hat{x}} \frac{\partial}{\partial x} + \mathbf{\hat{y}} \frac{\partial}{\partial y} + \mathbf{\hat{z}} \frac{\partial}{\partial z}.
\]

This vector derivative, in three dimensions, can be used to produce a vector function from a scalar function using the gradient, defined by

\[
\nabla f \equiv \mathbf{\hat{x}} \frac{\partial f}{\partial x} + \mathbf{\hat{y}} \frac{\partial f}{\partial y} + \mathbf{\hat{z}} \frac{\partial f}{\partial z}.
\]

From a vector function \( \mathbf{A}(\mathbf{r}) \), we can produce a scalar function by using the divergence, defined as

\[
\nabla \cdot \mathbf{A} \equiv \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}.
\]

We also define the curl, a vector quantity, as

\[
\nabla \times \mathbf{A} \equiv \mathbf{\hat{x}} \left( \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right) + \mathbf{\hat{y}} \left( \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \right) + \mathbf{\hat{z}} \left( \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right).
\]

Finally, it is useful to define the Laplacian, a second derivative of a scalar function, given by

\[
\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}.
\]

The gradient, divergence, and Laplacian generalize in an obvious way to higher dimensions; the curl, in contrast, does not so simply generalize.

Just as you can take derivatives for functions of multiple variables, you can also do integrals. We will focus almost exclusively on volume integrals, which are a type of

\(^{1}\) Note that in this notation, expressions like \( f'(\mathbf{r}) \) make no sense. We must either write out the partial derivative explicitly, or develop a new, more sophisticated shorthand.
definite integral where we add up the value of some function \( f(\mathbf{r}) \) over a volume \( V \). A volume integral would be written something like

\[ \iiint_V f(\mathbf{r}) \, d\mathbf{r} = \int \int \int_V f(\mathbf{r}) \, d^3\mathbf{r}. \]

Such an integral is really three one dimensional integrals, one over each of the three Cartesian coordinates \( x, y, \) and \( z \). Writing out more explicitly, you can think of the integral as of the form

\[ \iiint_V f(\mathbf{r}) \, d^3\mathbf{r} = \int dx \int dy \int dz \, f(x, y, z). \]  

(A.9)

In other words, we have three nested integrals, one inside the other (they can generally be performed in any order). As set up here, one would first perform the inner \( z \) integral, then the intermediate \( y \) integral, and finally the outermost \( x \) integral. The limits can be quite tricky; for example, the innermost integral \( (z) \) may have limits that depend on the outer two variable \( (x \) and \( y \)), while the intermediate integral may have limits that depend on the outermost variable \( (x) \), but not the inner variable \( (z) \), and the outermost integral \( (x) \) may not depend on the other two variables. Furthermore, when performing the integral, the variables \( x \) and \( y \) should be treated as constants inside the inner-most integral, while \( z \) is a variable; when you perform the intermediate integral, \( y \) is a variable, \( x \) is a constant, and \( z \) should have disappeared, and in the outermost integral, only the variable \( x \) should appear.

For example, suppose that I were given a region \( V \) defined by the four conditions

\[ \{(x, y, z) : x > 0, y > 0, z > 0, x + y + z < 1\} \]

Now, suppose I was asked to find the volume of this region, which is

\[ \int \int \int_V d^3\mathbf{r} = \int dx \int dy \int dz \]

What limits should I use? For the innermost integral, the simultaneous conditions \( z > 0 \) and \( x + y + z < 1 \) set the interval of integration as \( (0,1-x-y) \). For the intermediate integration, we don’t know what \( z \) is (other than positive), but clearly \( y > 0 \) and \( x + y < 1 \), and for the outermost integration, we know only that \( 0 < x < 1 \). So the correct limits are

\[ \int \int \int_V 1 \, d^3\mathbf{r} = \int_0^1 dx \int_0^{1-x} dy \int_0^{1-x-y} dz = \int_0^1 dx \int_0^{1-x} dy \int_0^{1-x-y} (1-x-y) \, dz = \int_0^1 dx \left( y - xy - \frac{1}{2} y^2 \right) \bigg|_{y=0}^{y=1-x} = \int_0^1 (1-x)^2 \, dx = \frac{1}{2} \int_0^1 (1-x)^2 \, dx = -\frac{1}{6} (1-x)^3 \bigg|_{x=0}^{x=1} = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{8}. \]

As a notational matter, it is worth noting that when given a multidimensional integral, it is common to use only a single integral symbol, rather than the triple integral. If the integral is over all space, it is common to leave the limit symbol \( V \) out of it entirely.

Is there any multi-dimensional analog of the fundamental theorem of calculus, (A.5), that allows you to find the integral of a derivative? There are several, in fact,
depending on what type of integral you are dealing with. The one we will need most often is Gauss’s Law, which relates an integral over a volume \( V \) to a surface integral over the boundary of \( V \) when you are integrating a divergence of a vector function.

\[
\iiint_V \nabla \cdot \mathbf{A}(\mathbf{r}) \, dV = 
\iint_S \hat{n} \cdot \mathbf{A}(\mathbf{r}) \, dS,
\]  
(A.10)

where \( S \) is the boundary of the volume \( V \) and \( \hat{n} \) is a unit normal sticking out of the integration region. The funky integral symbol on the right side of (A.10) merely indicates that the integral is being taken over a closed surface. The reason (A.10) still has integrals left over is because we started with a triple integral, and the derivative ‘cancels’ only one of the integrals, leaving you still with the two-dimensional surface integral. Equation (A.10) generalizes easily to other dimensions as well. In particular, in one dimension, the left side has only a single integral, and the right side is not an integral, but merely surface terms coming from the two limits of integration. In other words, the one-dimensional version of (A.10) is (A.1).

Stokes’ Theorem will also prove useful. It relates the integral of the curl of a vector field over a surface to an integral around the boundary of that same surface:

\[
\iint_S \hat{n} \cdot [\nabla \times \mathbf{A}(\mathbf{r})] \, dS = 
\oint_C \mathbf{A}(\mathbf{r}) \cdot d\mathbf{l}.
\]  
(A.11)

Once again, the number of integrals is reduced by one.

Finally, there is the divergence theorem, which assists you in performing a path integral of the divergence of a function from one point to another along a path:

\[
\int_A^B [\nabla f(\mathbf{r})] \cdot d\mathbf{l} = f(B) - f(A).
\]  
(A.12)

C. Coordinates in Two and Three Dimensions

In dimensions higher than one, it is common to use alternate coordinates besides the conventional Cartesian coordinates. When you change coordinates from a set of coordinates \( (x_1, x_2, \ldots, x_N) \) to a new set of coordinates \( (y_1, y_2, \ldots, y_N) \), any integration over the \( N \)-dimensional volume must be transformed as well. The rule is

\[
\int f(\mathbf{x}) \, d^N \mathbf{x} = 
\int f(\mathbf{y}) \left| \det \left( \frac{\partial x_i}{\partial y_j} \right) \right| \, d^N \mathbf{y}.
\]  
(A.13)

Furthermore, all of the derivative operators must be rewritten in the new basis as well.

Let’s start with two dimensions. In addition to standard Cartesian coordinates \( (x, y) \), the most common set of coordinates are polar coordinates \( (\rho, \phi) \). The relations between these two coordinate systems are
\[
\begin{aligned}
x &= \rho \cos \phi \\
y &= \rho \sin \phi \\
\end{aligned}
\] and
\[
\begin{aligned}
\rho &= \sqrt{x^2 + y^2} \\
\phi &= \tan^{-1}(y/x) \\
\end{aligned}
\] (A.14)

For any pair \((x, y)\) there are multiple values of \((\rho, \phi)\) that satisfy these equations. To make the relationship (A.14) unambiguous, we therefore restrict the coordinates \((\rho, \phi)\) by

\[
0 \leq \rho , \quad 0 \leq \phi < 2\pi .
\] (A.15)

The inverse tangent must be considered carefully, since there will be two values of \(\phi\) in the range \(0 \leq \phi < 2\pi\) that satisfy \(\phi = \tan^{-1}(y/x)\). The ambiguity can be removed with the help of \(y = \rho \sin \phi\): we choose \(0 < \phi < \pi\) if \(y\) is positive and \(\pi < \phi < 2\pi\) if \(y\) is negative. If \(y\) vanishes, then we pick \(\phi = 0\) if \(x\) is positive and \(\phi = \pi\) if \(x\) is negative. If \(x\) and \(y\) both vanish, then \(\phi\) is ambiguous; it is a bad spot in the coordinate system.

Scalar functions in one coordinate system are unchanged in the new coordinate system, but a vector quantity will now take the form

\[
\mathbf{A} = \hat{\rho} A_\rho + \hat{\phi} A_\phi.
\]

These components can be related to the ordinary Cartesian components by

\[
\begin{pmatrix}
A_x \\
A_y
\end{pmatrix} = \begin{pmatrix}
\cos \phi & -\sin \phi \\
\sin \phi & \cos \phi
\end{pmatrix} \begin{pmatrix}
A_\rho \\
A_\phi
\end{pmatrix},
\quad
\begin{pmatrix}
A_\rho \\
A_\phi
\end{pmatrix} = \begin{pmatrix}
\cos \phi & \sin \phi \\
-\sin \phi & \cos \phi
\end{pmatrix} \begin{pmatrix}
A_x \\
A_y
\end{pmatrix}.
\] (A.16)

We present, without proof, formulas for the gradient, divergence, Laplacian, and integral in these new coordinates:

\[
\nabla f = \frac{\partial f}{\partial \rho} \hat{\rho} + \frac{1}{\rho} \frac{\partial f}{\partial \phi} \hat{\phi},
\] (A.17a)

\[
\nabla \cdot \mathbf{A} = \frac{1}{\rho} \frac{\partial}{\partial \rho} (\rho A_\rho) + \frac{1}{\rho} \frac{\partial A_\phi}{\partial \phi},
\] (A.17b)

\[
\nabla^2 f = \frac{1}{\rho} \frac{\partial}{\partial \rho} \left( \rho \frac{\partial f}{\partial \rho} \right) + \frac{1}{\rho^2} \frac{\partial^2 f}{\partial \phi^2},
\] (A.17c)

\[
\int \int f(\mathbf{r}) d^2 \mathbf{r} = \int \rho d\rho \int d\phi f(\rho, \phi).
\] (A.17d)

If one integrates over all space, the limits of integration in (A.17d) are implied by (A.15).

In three dimensions, there are two coordinate systems that are commonly used in addition to Cartesian. The first are cylindrical, where \((x, y, z)\) are replaced with \((\rho, \phi, z)\), defined by the relations

\[
\begin{pmatrix}
x = \rho \cos \phi \\
y = \rho \sin \phi \\
z = z
\end{pmatrix}
\] and
\[
\begin{pmatrix}
\rho = \sqrt{x^2 + y^2} \\
\phi = \tan^{-1}(y/x) \\
z = z
\end{pmatrix},
\] (A.18)
where the last expression in each triplet means simply that the coordinate $z$ is unchanged. These coordinates are obviously very similar to polar coordinates in two dimensions. The restrictions (A.15) apply in this case as well, and the inverse tangent must be interpreted carefully in a manner similar to the comments after (A.15). Vector functions in the two coordinates are related in a manner almost identical to (A.9):

\[
\begin{pmatrix}
A_x \\
A_y \\
A_z
\end{pmatrix} = \begin{pmatrix}
\cos \phi & -\sin \phi & 0 \\
\sin \phi & \cos \phi & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
A_x \\
A_y \\
A_z
\end{pmatrix},
\begin{pmatrix}
A_x \\
A_y \\
A_z
\end{pmatrix} = \begin{pmatrix}
\cos \phi & \sin \phi & 0 \\
-\sin \phi & \cos \phi & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
A_x \\
A_y \\
A_z
\end{pmatrix}.
\]

The various differential and integral relations look nearly identical to (A.17), except we include the curl:

\[
\nabla f = \frac{\partial f}{\partial \rho} \hat{\rho} + \frac{1}{\rho} \frac{\partial f}{\partial \phi} \hat{\phi} + \frac{\partial f}{\partial z} \hat{z},
\]

(A.19a)

\[
\nabla \cdot \mathbf{A} = \frac{1}{\rho} \frac{\partial}{\partial \rho} (\rho A_\rho) + \frac{1}{\rho} \frac{\partial A_\phi}{\partial \phi} + \frac{\partial A_z}{\partial z},
\]

(A.19b)

\[
\nabla \times \mathbf{A} = \left[ \frac{1}{\rho} \frac{\partial A_z}{\partial \phi} - \frac{\partial A_\phi}{\partial z} \right] \hat{\rho} + \left[ \frac{1}{\rho} \frac{\partial A_\rho}{\partial \phi} - \frac{\partial A_z}{\partial \rho} \right] \hat{\phi} + \frac{1}{\rho} \frac{\partial}{\partial \rho} \left[ \rho A_\phi - \partial A_z/\partial z \right] \hat{z},
\]

(A.19c)

\[
\nabla^2 f = \frac{1}{\rho} \frac{\partial}{\partial \rho} \left( \rho \frac{\partial f}{\partial \rho} \right) + \frac{1}{\rho^2} \frac{\partial^2 f}{\partial \phi^2} + \frac{\partial^2 f}{\partial z^2},
\]

(A.19d)

\[
\int \int \int f(\mathbf{r}) d^3\mathbf{r} = \int \rho d\rho \int d\phi \int dz f(\rho, \phi, z).
\]

(A.19e)

Spherical coordinates $(r, \theta, \phi)$ are related to Cartesian coordinates by the relations

\[
\begin{pmatrix}
x = r \sin \theta \cos \phi \\
y = r \sin \theta \sin \phi \\
z = r \cos \theta
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
r = \sqrt{x^2 + y^2 + z^2} \\
\theta = \tan^{-1}(\sqrt{x^2 + y^2 / z}) \\
\phi = \tan^{-1}(y / x)
\end{pmatrix},
\]

(A.20)

with coordinate restrictions

\[0 \leq r, \quad 0 \leq \theta \leq \pi, \quad \text{and} \quad 0 \leq \phi < 2\pi.\]

The ambiguity in the variable $\phi$ is resolved in a manner identical to that in polar coordinates in two dimensions or cylindrical in three. There is no corresponding ambiguity for $\theta$. They are related to cylindrical coordinates by

\[
\begin{pmatrix}
\rho = r \sin \theta \\
\phi = \phi \\
z = r \cos \theta
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
r = \sqrt{\rho^2 + z^2} \\
\theta = \tan^{-1}(\rho / z) \\
\phi = \phi
\end{pmatrix}.
\]

(A.21)

Vector functions in spherical coordinates are related to Cartesian by
\begin{align*}
\begin{pmatrix}
A_x \\
A_y \\
A_z
\end{pmatrix} &=
\begin{pmatrix}
\sin \theta \cos \phi & \cos \theta \cos \phi & -\sin \phi \\
\sin \theta \sin \phi & \cos \theta \sin \phi & \cos \phi \\
\cos \theta & -\sin \theta & 0
\end{pmatrix}
\begin{pmatrix}
A_r \\
A_\theta \\
A_\phi
\end{pmatrix}, \\
\begin{pmatrix}
A_r \\
A_\theta \\
A_\phi
\end{pmatrix} &=
\begin{pmatrix}
\sin \theta \cos \phi & \sin \theta \sin \phi & \cos \theta \\
\cos \theta \cos \phi & \cos \theta \sin \phi & -\sin \phi \\
-\sin \phi & \cos \phi & 0
\end{pmatrix}
\begin{pmatrix}
A_x \\
A_y \\
A_z
\end{pmatrix}.
\end{align*}

and they are related to cylindrical coordinates by
\begin{align*}
\begin{pmatrix}
A_\rho \\
A_\phi \\
A_z
\end{pmatrix} &=
\begin{pmatrix}
\sin \theta & \cos \theta & 0 \\
0 & 0 & 1 \\
\cos \theta & -\sin \theta & 0
\end{pmatrix}
\begin{pmatrix}
A_r \\
A_\theta \\
A_\phi
\end{pmatrix}, \\
\begin{pmatrix}
A_r \\
A_\theta \\
A_\phi
\end{pmatrix} &=
\begin{pmatrix}
\sin \theta & 0 & \cos \theta \\
0 & \cos \theta & -\sin \theta \\
0 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
A_\rho \\
A_\phi \\
A_z
\end{pmatrix}.
\end{align*}

The various differential and integral identities in this coordinate system look like
\begin{align*}
\nabla f &= \frac{\partial f}{\partial r} \hat{r} + \frac{1}{r} \frac{\partial f}{\partial \theta} \hat{\theta} + \frac{1}{r \sin \theta} \frac{\partial f}{\partial \phi} \hat{\phi}, \quad (A.22a) \\
\nabla \cdot \mathbf{A} &= \frac{1}{r} \frac{\partial}{\partial r} \left( r^2 A_r \right) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} \left( A_\theta \sin \theta \right) + \frac{1}{r \sin \theta} \frac{\partial A_\phi}{\partial \phi}, \quad (A.22b) \\
\nabla \times \mathbf{A} &= \frac{1}{r \sin \theta} \left[ \frac{\partial}{\partial r} \left( A_\phi \sin \theta \right) - \frac{\partial A_\theta}{\partial \phi} \right] \hat{r} + \frac{1}{r} \left( \frac{1}{\sin \theta} \frac{\partial A_\phi}{\partial \phi} - \frac{\partial}{\partial r} (r A_\phi) \right) \hat{\theta} + \frac{1}{r} \left( \frac{\partial}{\partial r} (r A_\phi) - \frac{\partial A_\phi}{\partial \phi} \right) \hat{\phi}, \quad (A.22c) \\
\nabla^2 f &= \frac{1}{r} \frac{\partial^2 f}{\partial r^2} (rf) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial f}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2}, \quad (A.22d) \\
\iiint f(r) d^3r &= \iiint f(r) d^3r \int \sin \theta d\theta \int d\phi f(r, \theta, \phi). \quad (A.22e)
\end{align*}

The Laplacian can also be rewritten as
\begin{align*}
\nabla^2 f &= \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial f}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial f}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2}. \quad (A.23)
\end{align*}
D. Special Functions

Certain types of functions come up commonly enough in physics that you should become familiar with them. One has already been encountered, the factorial, defined by

\[ 0! \equiv 1, \quad n! \equiv 1 \cdot 2 \cdot 3 \cdots n. \]

A closely related function is the Gamma function, usually defined as

\[ \Gamma(n) \equiv \int_0^\infty x^{n-1} e^{-x} \, dx. \quad \text{(A.24)} \]

It is not hard to show by integration by parts that this satisfies the recursion relationship

\[ \Gamma(n + 1) = n \Gamma(n). \quad \text{(A.25)} \]

This relationship together with the easily derived fact that \( \Gamma(1) = 1 \) allows you to prove that if \( n \) is a positive integer, then \( \Gamma(n + 1) = n! \). The reason the Gamma function is more generally useful is that (A.24) can be defined also for non-integers, indeed for any positive number or even any complex number with a positive real part. Indeed, with the help of (A.25), you can extend the definition to all complex numbers (except for zero and negative integers, for which \( \Gamma(n) \) diverges). In particular, half-integers come up a lot, and it is helpful to know that

\[ \Gamma\left( \frac{1}{2} \right) = \sqrt{\pi}. \quad \text{(A.26)} \]

Together with (A.25), you can use this to get \( \Gamma(n) \) for any half-integer \( n \). Equation (A.26) can be proven with the help of a Gaussian integral, to which we turn next.

A simple Gaussian is a function of the form

\[ e^{-Ax^2}. \]

For \( A \) real and positive, this function peaks at the origin, and quickly dies out as \( x \) goes to positive or negative infinity. If \( A \) is complex but has a positive real part, it still has a maximum magnitude at the origin, but it will not only diminish but also rotate in the complex plane as you move away from zero. We very often will encounter Gaussian integrals of the form

\[ I = \int_{-\infty}^\infty e^{-Ax^2} \, dx. \]

A clever trick allows us to evaluate this integral. First square it, writing the right side as the integral times itself, and then change the variable of integration of one of the two resulting integral from \( x \) to \( y \):

\[ I^2 = \left[ \int_{-\infty}^\infty e^{-Ax^2} \, dx \right]^2 = \int_{-\infty}^\infty e^{-Ax^2} \, dx \int_{-\infty}^\infty e^{-Ay^2} \, dy = \int_{-\infty}^\infty \int_{-\infty}^\infty e^{-Ax^2 - Ay^2} \, dx \, dy. \]
Now change this two-dimensional area integral into polar coordinates with the help of (A.17d)

\[ I^2 = \int_0^\infty \int_0^{2\pi} e^{-Ap^2} \rho d\rho d\phi = 2\pi \frac{1}{2A} \int_0^{\infty} e^{-Ap^2} \rho \left( Ap^2 \right) = -\frac{\pi}{A} \left. e^{-Ap^2} \right|_0^\infty = \frac{\pi}{A}. \]

Taking the square root, we have \( I = \sqrt{\pi/A} \), so in conclusion,

\[ \int_{-\infty}^{\infty} e^{-Ax^2} dx = \sqrt{\frac{\pi}{A}} \quad \text{for} \quad \text{Re}(A) \geq 0. \quad (A.27) \]

Since \( A \) might be complex, we should clarify that the square root is chosen such that the real part is positive. The formula also works for \( A \) pure imaginary, provided \( A \) is not zero. If you let \( A = 1 \) and substitute \( x = \sqrt{y} \), in (A.27), it is not hard to derive (A.26).

Starting from (A.27), it isn’t hard to derive a more general integral, which I call a shifted Gaussian, given by

\[ \int_{-\infty}^{\infty} e^{-Ax^2+Bx} dx = \sqrt{\frac{\pi}{A}} \exp\left(\frac{B^2}{4A}\right) \quad \text{for} \quad \text{Re}(A) \geq 0. \quad (A.28) \]

If there are extra powers of \( x \) in front, it is possible to see that these can be found by taking derivatives of (A.28) with respect to \( B \), since this brings down factors of \( x \). Hence we have

\[ \int_{-\infty}^{\infty} x^n e^{-Ax^2+Bx} dx = \sqrt{\frac{\pi}{A}} \frac{\partial^n}{\partial B^n} \left[ \exp\left(\frac{B^2}{4A}\right) \right] \quad \text{for} \quad \text{Re}(A) \geq 0. \quad (A.29) \]

This formula takes a particular simple form if \( B = 0 \). Though it can be derived directly from (A.29), it is easier to change variables and use the definition of the Gamma function to show that

\[ \int_{-\infty}^{\infty} x^n e^{-Ax^2} dx = \Gamma\left(\frac{n+1}{2}\right) \frac{A^{\frac{n+1}{2}}}{\sqrt{\pi}} \quad \text{for} \quad \text{even}. \quad (A.30) \]

For \( n \) odd, the integral vanishes because the positive and negative parts cancel out.

Some additional functions that come up occasionally are hyperbolic functions. Three of them are defined as

\[ \cosh(x) \equiv \frac{1}{2}\left(e^x + e^{-x}\right), \quad \sinh(x) \equiv \frac{1}{2}\left(e^x - e^{-x}\right), \quad \text{and} \quad \tanh(x) \equiv \frac{\sinh(x)}{\cosh(x)}. \]

Like \( e^{\pm x} \), both \( \cosh \) and \( \sinh \) satisfy the differential equation \( f''(x) = f(x) \). However, they are even and odd functions respectively. Indeed, using Euler’s Theorem (A.3), it is easy to see that these are closely related to the trigonometric functions through the complex numbers; specifically

\[ \cosh(ix) = \cos(x), \quad \sinh(ix) = i \sin(x), \quad \text{and} \quad \tanh(ix) = i \tan(x). \]
E. The Dirac Delta function

One other function requires special attention, though strictly speaking, it isn’t a function at all. The Dirac delta function is written $\delta(x)$, and is defined to be the function such that for any function $f(x)$, we have

$$\int_{-\infty}^{\infty} f(x) \delta(x) \, dx = f(0). \quad (A.31)$$

You should think of the Dirac delta function as a function that vanishes everywhere except the origin, but it is infinitely large at the origin, and has a total area of 1. As such, it is not, in the mathematical sense, a function at all (sometimes it is called a distribution), but we can think of it as the limit of a peak with very narrow width and very large height. For example, consider the limit

$$\lim_{A \to \infty} \frac{1}{A} e^{-\frac{x^2}{A}}. \quad (A.32)$$

For large but finite $A$, this represents a function whose height is very large, but which quickly diminishes outside of a small region surrounding the origin. If, for example, the function $f(x)$ is continuous near the origin, its value will change little in the region where $\delta(x)$ is large, and we can treat it as a constant $f(x) \approx f(0)$ in this region. We can therefore demonstrate that (A.31) implies (A.32) because

$$\int_{-\infty}^{\infty} f(x) \delta(x) \, dx = \lim_{A \to \infty} \int_{-\infty}^{\infty} f(x) \frac{1}{\pi} e^{-\frac{x^2}{A}} \, dx = f(0) \lim_{A \to \infty} \frac{1}{\pi} \int_{-\infty}^{\infty} e^{-\frac{x^2}{A}} \, dx = f(0). \quad (A.33)$$

This is not to imply that (A.32) is the only way to define the Dirac delta function. Furthermore, the arguments given assume that the function $f(x)$ is continuous at the origin; if this is not the case, then the integral (A.31) will not work, and indeed, the integral is poorly defined.

Several variants of (A.31) will prove useful. First of all, because the only place that contributes to the integral is the origin, (A.31) will remain true if we change the upper and lower limits of integration, provided 0 is still within the range of integration (if it isn’t, the integral is just 0). We can shift the argument of the delta function, which simply shifts the point where we evaluate the function $f$:

$$\int_{-\infty}^{\infty} f(x) \delta(x-a) \, dx = f(a).$$

We can even imagine having functions inside the Dirac delta function. Suppose we are given an integral of the form

$$\int_{-\infty}^{\infty} f(x) \delta(g(x)) \, dx. \quad (A.34)$$
For the moment, let us assume that $g(x)$ is a monotonic function, which either increases from $-\infty$ to $+\infty$ as $x$ increases, or decreases from $+\infty$ to $-\infty$. The key to performing this integration is to define a new variable $y = g(x)$, and rewrite the integral in terms of $y$. We imagine finding the inverse function $x = g^{-1}(y)$. The result will be

$$\int_{-\infty}^{\infty} f(x) \delta(g(x)) \, dx = \int_{-\infty}^{\infty} f(g^{-1}(y)) \delta(y) \, dx = \int_{g^{-1}(\infty)}^{g^{-1}(0)} f(g^{-1}(y)) \delta(y) \, dy \, dx \quad \text{(A.35)}$$

Now, suppose for the moment that $g$ is an increasing function, and therefore the limits are exactly as before, $-\infty$ to $+\infty$. Then this integral is straightforward, and we have

$$\int_{-\infty}^{\infty} f(x) \delta(g(x)) \, dx = \left. f(g^{-1}(y)) \frac{dx}{dy} \right|_{y = 0} = \frac{f(g^{-1}(0))}{g'(g^{-1}(0))}, \quad \text{(A.36)}$$

if $g$ is increasing, where $g'$ is the derivative of $g$. Note that in this case, $g'$ is positive everywhere. On the other hand, if $g$ is a decreasing function, the limits on the final integral in (A.35) will be switched, which produces a minus sign. If you work through the same steps as we did in (A.36), you find

$$\int_{-\infty}^{\infty} f(x) \delta(g(x)) \, dx = -\frac{f(g^{-1}(0))}{g'(g^{-1}(0))}, \quad \text{(A.37)}$$

if $g$ is decreasing. Note that in this case, $g'$ is negative everywhere. We can therefore combine equations (A.36) and (A.37) into a single case, which we write as

$$\int_{-\infty}^{\infty} f(x) \delta(g(x)) \, dx = \frac{f(x)}{|g'(x)|}, \quad \text{(A.38)}$$

where $x_i = g^{-1}(0)$; i.e., the unique point where $g(x_i) = 0$.

What do we do if the function $g$ does not go from $-\infty$ to $+\infty$, or vice versa, but instead has only a finite range, or perhaps even turns around and is sometimes increasing and sometimes decreasing? The answer can be understood by thinking more carefully about the integral (A.34). Recall that the Dirac delta function vanishes except when its argument is zero. Therefore, if $g(x)$ doesn’t vanish anywhere, the integral (A.23) vanishes. If $g(x)$ vanishes at one point, but doesn’t go all the way to $-\infty$ or $+\infty$, it doesn’t matter; just as in (A.31), the upper and lower limits don’t matter, so long as the range includes 0. Equation (A.38) is still valid. If $g(x)$ vanishes more than once, then each time it vanishes there will be a contribution to the integral (A.38), and that contribution will be of the form (A.38), so as a consequence the result will be a sum.

$$\int_{-\infty}^{\infty} f(x) \delta(g(x)) \, dx = \sum_{x_i: g(x_i) = 0} \frac{f(x_i)}{|g'(x_i)|}. \quad \text{(A.39)}$$
Indeed, if the upper and lower limits of (A.39) are changed, the right hand side will remain unchanged, except that the sum will be restricted to those roots of $g$ which lie within the range of integration. Equation (A.39) can be rewritten in non-integral form as

$$\delta(g(x)) = \sum_{x = x_i} \frac{1}{g'(x_i)} \delta(x-x_i).$$  (A.40)

As an example, suppose $g(x) = x^2 - a^2$, then $g'(x) = 2x$, and the two roots of $g(x) = 0$ are $x = \pm a$, and we find

$$\int_{-\infty}^{\infty} f(x) \delta(x^2 - a^2) dx = \frac{1}{2a} [f(a) + f(-a)].$$

The Dirac delta function can be generalized to more than one dimension by simply multiplying one-dimensional Dirac delta functions. In three dimensions,

$$\delta^3(r) \equiv \delta(x) \delta(y) \delta(z).$$

It easily follows that

$$\iiint f(r) \delta^3(r - a) d^3r = f(a).$$

Analogs of (A.39) and (A.40) can be derived for when the Dirac delta function has a vector argument, $\delta^3(g(r))$. We give the results without proof.

$$\iiint f(r) \delta^3(g(r)) d^3r = \sum_{r_i : \nabla g(r_i) = 0} f(r_i) \left| \det \left[ \frac{\partial g(r)}{\partial r} \right]_{r_i} \right|^{-1}.$$  (A.41a)

$$\delta^3(g(r)) = \sum_{r_i : \nabla g(r_i) = 0} \delta^3(r - r_i) \left| \det \left[ \frac{\partial g(r)}{\partial r} \right]_{r_i} \right|^{-1}.$$  (A.41b)

F. Fourier Transforms

We start with the following “identity”:

$$\int_{-\infty}^{\infty} e^{ikx} dx = 2\pi \delta(k).$$  (A.42)

This identity cannot really be proven, simply because it isn’t quite true, but it can be justified by considering the following integral, which can be evaluated with the help of (A.28) and (A.32):

$$\int_{-\infty}^{\infty} e^{ikx} dx = \lim_{\epsilon \to 0} \int_{-\infty}^{\infty} e^{ikx - \epsilon^2} dx = \sqrt{\frac{\pi}{\epsilon}} \exp \left( \frac{(ik)^2}{4\epsilon} \right) = 2\pi \lim_{\epsilon \to 0} \sqrt{\frac{1}{4\epsilon \pi}} \exp \left( -\frac{1}{4\epsilon} k^2 \right) = 2\pi \delta(k).$$
Strictly speaking, the interchange of the limit with the integration is not valid, but this formula is so widely used in physics that we will treat it as true.

Let \( f(x) \) be any reasonably smooth function. We will normally also demand that \( f \) vanish not too slowly, say, faster than \( 1/x \) at large \( x \). Then define the Fourier transform \( \tilde{f}(k) \) as

\[
\tilde{f}(k) \equiv \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}} f(x) e^{-ikx}.
\]

Consider now the Fourier transform of the Fourier transform, which we will evaluate with the help of (A.42)

\[
\tilde{\tilde{f}}(x) = \int_{-\infty}^{\infty} \frac{dk}{\sqrt{2\pi}} \tilde{f}(k) e^{ikx} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} f(x) e^{ikx} dx \right] e^{ikx} dk = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} e^{ik(x+y)} dy \right] f(y) dy
\]

\[
= \int_{-\infty}^{\infty} \delta(x+y) f(y) dy = f(-x).
\]

Up to sign, the Fourier transform of a Fourier transform is the function back again. We therefore summarize this information as two relations, here presented together.

\[
\tilde{f}(k) = \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}} f(x) e^{-ikx}, \quad f(x) = \int_{-\infty}^{\infty} \frac{dk}{\sqrt{2\pi}} \tilde{f}(k) e^{ikx}.
\]

One other relation that is not hard to prove is that the magnitude squared integral of a function and its Fourier transform are the same. The proof is given below.

\[
\int_{-\infty}^{\infty} \left| \tilde{f}(k) \right|^2 dk = \int_{-\infty}^{\infty} \tilde{f}(k)^* \tilde{f}(k) dk = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}} f(x) e^{-ikx} dx \right] \left[ \int_{-\infty}^{\infty} \frac{dy}{\sqrt{2\pi}} f(y) e^{-iky} dy \right] dk
\]

\[
= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) dx \int_{-\infty}^{\infty} f^*(y) dy \int_{-\infty}^{\infty} e^{-ikx} e^{iky} dk = \int_{-\infty}^{\infty} f(x) dx \int_{-\infty}^{\infty} f^*(y) dy \delta(y-x)
\]

\[
= \int_{-\infty}^{\infty} f^*(x) f(x) dx = \int_{-\infty}^{\infty} |f(x)|^2 dx.
\]

All of these formulas can be generalized to three (or more) dimensions easily, so we give the resulting equations without proof.

\[
\int_{-\infty}^{\infty} e^{ik\cdot r} d^3 \mathbf{r} = (2\pi)^3 \delta^3(\mathbf{k}),
\]

\[
\tilde{f}(\mathbf{k}) = \int_{-\infty}^{\infty} \frac{d^3 \mathbf{r}}{(2\pi)^{3/2}} f(\mathbf{r}) e^{-i\mathbf{k} \cdot \mathbf{r}},
\]

\[
f(\mathbf{r}) = \int_{-\infty}^{\infty} \frac{d^3 \mathbf{k}}{(2\pi)^{3/2}} \tilde{f}(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{r}},
\]

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| \tilde{f}(\mathbf{k}) \right|^2 d^3 \mathbf{k} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(\mathbf{r})|^2 d^3 \mathbf{r}.
\]
### Appendix B. Units and Constants

#### Fundamental SI Units

<table>
<thead>
<tr>
<th>Measures</th>
<th>Name</th>
<th>Abbr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>meter</td>
<td>m</td>
</tr>
<tr>
<td>Time</td>
<td>second</td>
<td>s</td>
</tr>
<tr>
<td>Mass</td>
<td>kilogram</td>
<td>kg</td>
</tr>
<tr>
<td>Temperature</td>
<td>Kelvin</td>
<td>K</td>
</tr>
<tr>
<td>Charge</td>
<td>Coulomb</td>
<td>C</td>
</tr>
</tbody>
</table>

#### Metric Abbreviations

<table>
<thead>
<tr>
<th>Name</th>
<th>Abbr.</th>
<th>Multiple</th>
</tr>
</thead>
<tbody>
<tr>
<td>tera</td>
<td>T</td>
<td>$10^{12}$</td>
</tr>
<tr>
<td>giga</td>
<td>G</td>
<td>$10^9$</td>
</tr>
<tr>
<td>mega</td>
<td>M</td>
<td>$10^6$</td>
</tr>
<tr>
<td>kilo</td>
<td>k</td>
<td>$10^3$</td>
</tr>
<tr>
<td>centi</td>
<td>c</td>
<td>$10^{-2}$</td>
</tr>
<tr>
<td>milli</td>
<td>m</td>
<td>$10^{-3}$</td>
</tr>
<tr>
<td>micro</td>
<td>µ</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td>nano</td>
<td>n</td>
<td>$10^{-9}$</td>
</tr>
<tr>
<td>pico</td>
<td>p</td>
<td>$10^{-12}$</td>
</tr>
<tr>
<td>femto</td>
<td>f</td>
<td>$10^{-15}$</td>
</tr>
</tbody>
</table>

#### Derived SI Units

<table>
<thead>
<tr>
<th>Measures</th>
<th>Name</th>
<th>Abbr.</th>
<th>Equiv.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Force</td>
<td>Newton</td>
<td>N</td>
<td>kg·m/s²</td>
</tr>
<tr>
<td>Energy</td>
<td>Joule</td>
<td>J</td>
<td>N·m</td>
</tr>
<tr>
<td>Power</td>
<td>Watt</td>
<td>W</td>
<td>J/s</td>
</tr>
<tr>
<td>Frequency</td>
<td>Hertz</td>
<td>Hz</td>
<td>s⁻¹</td>
</tr>
<tr>
<td>Pressure</td>
<td>Pascal</td>
<td>Pa</td>
<td>N/m²</td>
</tr>
<tr>
<td>Current</td>
<td>Ampere</td>
<td>A</td>
<td>C/s</td>
</tr>
<tr>
<td>Elec. Potential</td>
<td>Volt</td>
<td>V</td>
<td>J/C</td>
</tr>
<tr>
<td>Mag. Field</td>
<td>Tesla</td>
<td>T</td>
<td>N·s/m/C</td>
</tr>
</tbody>
</table>

#### Non SI Units

<table>
<thead>
<tr>
<th>Name</th>
<th>Abbr.</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day</td>
<td>d</td>
<td>86,400 s</td>
</tr>
<tr>
<td>Year</td>
<td>y</td>
<td>$3.1558 \times 10^7$ s</td>
</tr>
<tr>
<td>Electron volt</td>
<td>eV</td>
<td>$1.6022 \times 10^{-19}$ J</td>
</tr>
<tr>
<td>Unified mass unit</td>
<td>u</td>
<td>$1.6605 \times 10^{-27}$ kg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$= 931.494$ MeV/c²</td>
</tr>
</tbody>
</table>

#### Physical Constants

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planck’s Constant</td>
<td>$h$</td>
<td>$6.6261 \times 10^{-34}$ J·s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$= 4.1357 \times 10^{-15}$ eV·s</td>
</tr>
<tr>
<td>Planck’s Reduced Constant</td>
<td>$\hbar$</td>
<td>$1.0546 \times 10^{-34}$ J·s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$= 6.5821 \times 10^{-16}$ eV·s</td>
</tr>
<tr>
<td>Speed of light</td>
<td>$c$</td>
<td>$2.9979 \times 10^8$ m/s</td>
</tr>
<tr>
<td>Fundamental Charge</td>
<td>$e$</td>
<td>$1.6022 \times 10^{-19}$ C</td>
</tr>
<tr>
<td>Coulomb’s Constant</td>
<td>$k_e$</td>
<td>$8.9876 \times 10^9$ N·m²/C²</td>
</tr>
<tr>
<td>Permittivity of free space</td>
<td>$\varepsilon_0$</td>
<td>$8.8542 \times 10^{-12}$ C²/N·m²</td>
</tr>
<tr>
<td>Permeability of free space</td>
<td>$\mu_0$</td>
<td>$4\pi \times 10^{-7}$ N·s²/C²</td>
</tr>
<tr>
<td>Avogadro’s number</td>
<td>$N_A$</td>
<td>$6.0221 \times 10^{23}$ /mol</td>
</tr>
<tr>
<td>Electron mass</td>
<td>$m_e$</td>
<td>$9.1094 \times 10^{-31}$ kg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$= 511.00$ keV/c²</td>
</tr>
<tr>
<td>Proton mass</td>
<td>$m_p$</td>
<td>$1.6726 \times 10^{-27}$ kg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$= 938.27$ MeV/c²</td>
</tr>
<tr>
<td>Neutron mass</td>
<td>$m_n$</td>
<td>$1.6749 \times 10^{-27}$ kg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$= 939.56$ MeV/c²</td>
</tr>
<tr>
<td>Boltzmann’s Constant</td>
<td>$k_B$</td>
<td>$1.3807 \times 10^{-23}$ J/K</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$= 8.6173 \times 10^{-5}$ eV/K</td>
</tr>
</tbody>
</table>

#### Useful Combinations

<table>
<thead>
<tr>
<th>Combination</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$hc$</td>
<td>$1.9864 \times 10^{-25}$ J·m</td>
</tr>
<tr>
<td></td>
<td>$= 1239.8$ eV·nm</td>
</tr>
<tr>
<td>$\hbar c$</td>
<td>$3.1615 \times 10^{-26}$ J·m</td>
</tr>
<tr>
<td></td>
<td>$= 197.33$ eV·nm</td>
</tr>
<tr>
<td>$a_0 \equiv \frac{h^2}{k_e m_e e^2}$</td>
<td>0.052918 nm</td>
</tr>
<tr>
<td>$\alpha \equiv \frac{k_e e^2}{\hbar c}$</td>
<td>$0.0072974 = \frac{1}{137.04}$</td>
</tr>
<tr>
<td>MeV/c</td>
<td>$5.344 \times 10^{-22}$ kg·m/s</td>
</tr>
<tr>
<td>MeV/c²</td>
<td>$1.78266 \times 10^{-30}$ kg</td>
</tr>
</tbody>
</table>