Principles of Quantum Mechanics

University of Cambridge Part II Mathematical Tripos

Yue Wu

Yusuf Hamied Department of Chemistry
Lensfield Road,
Cambridge, CB2 1EW

yw628@cam.ac.uk

Acknowledgements

Nothing in these lecture notes is original. They are largely based on the notes by Prof. David Skinner, who lectured this course in 2021, and his notes can be found here. Moreover, they are nowhere near accurate representations of what was actually lectured, and in particular, all errors are almost surely mine.

Contents

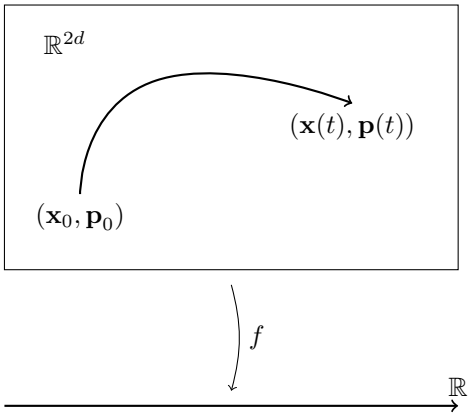


Figure 1: A particle's trajectory in phase space. Observables are represented by functions $f : \mathbb{R}^{2d} \to \mathbb{R}$, evaluated along a given particle's trajectory.

1 Introduction

In classical mechanics, a particle's motion is governed by Newton's laws. These are second order ODEs, so to determine the fate of our particle we must specify two initial conditions. We could take these to be the particle's initial position $\mathbf{x}(t_0)$ and velocity $\mathbf{v}(t_0)$, or its initial position $\mathbf{x}(t_0)$ and momentum $\mathbf{p}(t_0)$. Once these are specified the motion is determined (provided of course we understand how to describe the forces that are acting). This means that we can solve Newton's second law to find the values $(\mathbf{x}(t), \mathbf{p}(t))$ for $t > t_0$. So as time passes, our classical particle traces out a trajectory in the space M of possible positions and momenta, as sketched in figure 1. The space M is known as the *phase space* and in our case, for motion in three dimensions, M is just $\mathbb{R}^6$. In general M comes with a rich geometry known as a *Poisson structure*; you'll study this structure in detail if you are taking the Part II courses on Classical Dynamics or Integrable Systems, and we'll touch on it later in this course, too.

Classical observables are represented by functions

$$\begin{aligned} f : M &\longrightarrow \mathbb{R} \\ (\mathbf{x}, \mathbf{p}) &\longmapsto f(\mathbf{x}, \mathbf{p})\,. \end{aligned} \tag{1.1}$$

For example, we may be interested in the kinetic energy $T = \mathbf{p}^2/2m$, potential energy $V(\mathbf{x})$, angular momentum $\mathbf{L} = \mathbf{x} \times \mathbf{p}$, or a host of other possible quantities. A priori, these functions are defined everywhere over M, but if we want to know about the energy or angular momentum of our specific particle then we should evaluate them not at some random point $(\mathbf{x}, \mathbf{p}) \in M$, but along the particle's phase space trajectory. For example, if at time t the particle has position $\mathbf{x}(t)$ and momentum $\mathbf{p}(t)$, then its angular momentum is $\mathbf{x}(t) \times \mathbf{p}(t)$. Thus the values of the particle's energy, angular momentum *etc.* may depend on time, though of course our definition of these quantities does not. In this way, everything we could possibly want to know about a single, point like particle is encoded in its phase space trajectory.

But that is not our World. To the best of our current experimental knowledge, our world is quantum, not classical. Initially, these experiments were based on careful studies of atomic spectroscopy and blackbody radiation, but nowadays I'd prefer to say that the best evidence of quantum mechanics is simply that we use it constantly in our daily lives. Each time you listen to music on your stereo, post a photo on Instagram or make a call on your phone you're relying on technology that's only become possible due to our understanding of the quantum structure of matter. Whenever you plug something into the mains, you're using electricity that's in part generated by nuclear reactions in which quantum mechanics is essential, while much of modern medicine relies on new drugs designed with the benefit of the improved understanding of chemistry that quantum mechanics provides.

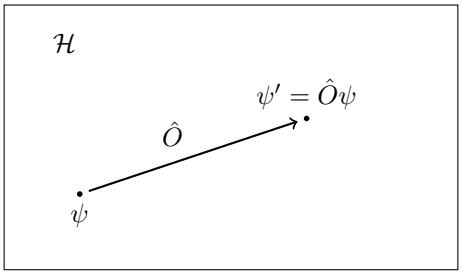


Figure 2: In Quantum Mechanics, complete knowledge of a particle's state is determined by a vector in Hilbert space. Observables are represented by Hermitian linear operators $\hat{O} : \mathcal{H} \to \mathcal{H}$.

In such a quantum world, instead of a phase space trajectory, everything we could want to know about a particle is encoded in a vector ψ in Hilbert space $\mathcal{H}$. As you met in IB Quantum Mechanics, this *state vector* evolves in time according to Schrödinger's equation. In the quantum world, observables are represented by certain operators $\hat{O}$. The operators you saw in IB Quantum Mechanics had the same sort of form as observables in classical mechanics, such as the kinetic operator $\hat{T} = \hat{\mathbf{p}}^2/2m$ or angular momentum operator $\hat{\mathbf{L}} = \hat{\mathbf{x}} \times \hat{\mathbf{p}}$. However, rather than being functions, these operators are (roughly) linear maps

$$\hat{O} : \mathcal{H} \longrightarrow \mathcal{H}\,. \tag{1.2}$$

Again, in the first instance these operators are defined throughout $\mathcal{H}$, but if we're interested in knowing about the energy or angular momentum of our particular quantum particle, then we should find out what happens when they act on the specific $\psi \in \mathcal{H}$ that describes the state of our particle at time t.

In the following chapters we'll study what Hilbert space is and what its operators do in a more general framework than you saw last year, building your insight into the mathematical structure of quantum mechanics. Much of this is just linear algebra, but the Hilbert spaces we'll care about in Quantum Mechanics are often infinite-dimensional, so we also make contact with Functional Analysis. Furthermore, although last year you 'guessed' the form of quantum operators by analogy with their classical counterparts, we'll see that at a deeper level many of them can be understood to have their origins in symmetries of space and time; the operators just reflect the way these symmetry transforms act on Hilbert space, rather that on (non-relativistic) space-time. In this way, Quantum Mechanics makes contact with Representation Theory.

So, mathematically, much of Quantum Mechanics boils down to a mix of Functional Analysis and Representation Theory. It's even true that it provides a particularly interesting example of these subjects. But this is not the reason we study it. We study Quantum Mechanics in an effort to understand and appreciate our world, not some abstract mathematical one. You're all intimately familiar with vector spaces, and you're also very good at solving Sturm–Liouville type eigenfunction/eigenvalue problems. But the real skill is in understanding how this formalism relates to the world we see around us.

It's not obvious. Newton's laws are (at least generically) non-linear differential equations and we can't superpose solutions. General Relativity teaches us that spacetime is not flat. So it's not at all clear that our particle should in fact be described by a point in a vector space, any more than it was obvious to Aristotle that bodies actually stay in uniform motion unless acted on by a force, or clear to the Ancients that the arrival of solar eclipses, changes of the weather, or any other natural phenomenon are actually governed by calculable laws, rather than the whims of various gods. For this reason, instead of emphasizing how weird and different Quantum Mechanics is, I'd prefer to make you appreciate how it actually underpins the physics you're already familiar with. No matter how good you are at solving eigenvalue problems, if you don't see how these relate to your everyday physical intuition, knowledge you've built up since first opening your eyes and learning to crawl, then you haven't really understood the subject.

2 Hilbert Space

The realm of Quantum Mechanics is Hilbert space[1], so we'll begin by exploring the properties of these. This chapter will necessarily be almost entirely mathematical; the physics comes later.

2.1 Definition of Hilbert Spaces

Hilbert space is a vector space $\mathcal{H}$ over $\mathbb{C}$ equipped with a complete inner product. Let's take a moment to understand what this means; much of it will be familiar from IB Linear Algebra or IB Methods.

Saying that $\mathcal{H}$ is a vector space means that it is a set on which we have an operation of addition $+ : \mathcal{H} \times \mathcal{H} \to \mathcal{H}$, obeying

- *commutativity.* $\psi + \phi = \phi + \psi$;
- *associativity.* $\psi + (\phi + \chi) = (\psi + \phi) + \chi$;
- *identity.* $\exists! \mathbf{0} \in \mathcal{H}$ s.t. $\psi + \mathbf{0} = \psi$

for all $\psi, \phi, \chi \in \mathcal{H}$. Furthermore, we can multiply our vectors by numbers in $\mathbb{C}$ called scalars. This multiplication $\cdot : \mathbb{C} \times \mathcal{H} \to \mathcal{H}$ is

- *distributive over* $\mathcal{H}$. $c \cdot (\psi + \phi) = c \cdot \psi + c \cdot \phi$;
- *distributive in* $\mathbb{C}$. $(a + b) \cdot \psi = a \cdot \psi + b \cdot \psi$

for all $a, b, c \in \mathbb{C}$. In addition, $\mathcal{H}$ comes equipped with an *inner product*. This is a map $(\ ,\ ) : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ that obeys

- *conjugate symmetry.* $(\psi, \phi) = (\phi, \psi)^*$;
- *linearity.* $(\phi, a\psi) = a(\phi, \psi)$;
- *additivity.* $(\phi, \psi + \chi) = (\phi, \psi) + (\phi, \chi)$;
- *positive definiteness.* $(\psi, \psi) \geq 0$, with equality iff $\psi = \mathbf{0}$.

Note that the first two of these imply $(a\phi, \psi) = a^*(\phi, \psi)$ so that $(\ ,\ )$ is *antilinear* in its first argument[2]. Note also that $(\psi, \psi) = (\psi, \psi)^*$ so it is necessarily real. Whenever we have inner product, we can define the *norm* of a state to be

$$\|\psi\| = \sqrt{(\psi, \psi)}\,. \tag{2.1}$$

These properties ensure that the Cauchy–Schwarz inequality

$$|(\phi, \psi)|^2 \leq \|\phi\|^2 \, \|\psi\|^2 \tag{2.2}$$

holds.

As always, a set of vectors $\{\phi_1, \phi_2, \ldots, \phi_n\}$ is *linearly independent* if the only solution to

$$c_1\phi_1 + c_2\phi_2 + \cdots + c_n\phi_n = \mathbf{0} \tag{2.3}$$

[1] In this course, we'll focus on Dirac's formulation of QM, which is based on Hilbert space. This is by far the most commonly used approach. However, there are some other approaches to QM (notably deformation quantization and the theory of C^*-algebra) in which Hilbert spaces do not play a prominent role. We won't discuss any of them.

[2] In the math literature, the inner product is often taken to be linear in the first entry and antilinear in the second. We will follow the QM convention, which is opposite.

for $c_i \in \mathbb{C}$ is $c_1 = \cdots = c_n = 0$. The *dimension* of the vector space is the largest possible number of linearly independent vectors we can find. If there is no such largest number, then we say the vector space has infinite dimension. A set of vectors $\{\phi_1, \phi_2, \ldots, \phi_n\}$ is orthonormal with respect to the inner product $(\ ,\ )$ if

$$(\phi_a, \phi_b) = \begin{cases} 0 & \text{if } a \neq b \\ 1 & \text{if } a = b\,. \end{cases} \tag{2.4}$$

A set $\{\phi_1, \phi_2, \ldots, \phi_n\}$ forms a *basis* of an n-dimensional Hilbert space if every $\psi \in \mathcal{H}$ can be uniquely expressed as a sum

$$\psi = \sum_{i=1}^{n} c_i \phi_i\,, \tag{2.5}$$

with some coefficients $c_i \in \mathbb{C}$. The basis is usually taken to be orthonormal. In this case, we can easily determine these coefficients by taking the inner product with ϕ_i, since

$$(\phi_i, \psi) = \left(\phi_i, \sum_{j=1}^{n} c_j \phi_j\right) = \sum_j c_j (\phi_i, \phi_j) = c_i \tag{2.6}$$

by linearity and orthonormality.

Quantum mechanics makes use of both finite and infinite dimensional Hilbert spaces, as we'll see. In the infinite dimensional case, we have to decide what we mean by an 'infinite linear combination' of (e.g. basis) vectors. Not every such infinite sum makes sense, because infinite sums such as $\sum_{i=1}^{\infty} c_i \phi_i$ might not converge. Technically, we consider the partial sums $S_N = \sum_{i=1}^{N} c_i \phi_i$ of just the first N terms, and say that the Cauchy sequence $\{S_1, S_2, \ldots\}$ *converges in the norm* if there is some vector Ψ to which it converges in the sense that

$$\lim_{N \to \infty} \|S_N - \Psi\| = 0\,. \tag{2.7}$$

To say that $\mathcal{H}$ is complete (or complete in the norm) means that every Cauchy sequence $\{S_1, S_2, \ldots\}$ indeed converges in $\mathcal{H}$. This captures the heuristic idea that there are no points 'missing' from $\mathcal{H}$. Importantly, it also allows us to analyse: being able to differentiate vectors in $\mathcal{H}$ requires that we can take limits, and to do this we need to know whether the limits exist.

The interplay between the vector space structure and this requirement of completeness can be very subtle in infinite dimensions — you'll see much more of this if you take the Part II course on Linear Analysis or (next term) Functional Analysis. In this course, we'll largely ignore such subtleties, not because they're not interesting, but because they're a distraction from all the interesting physics we need to learn.

2.1.1 Examples

Let's now look at a few examples of Hilbert spaces, pointing out where they're relevant to physics.

The simplest case is when $\mathcal{H}$ is finite dimensional. In this case, as a vector space we have $\mathcal{H} \cong \mathbb{C}^n$ for some n. You might think we still have lots of choice in picking an inner product, but it turns out that finite dimensional Hilbert space is always isomorphic to one with inner product

$$(\mathbf{v}, \mathbf{u}) = \sum_{i=1}^{n} v_i^* u_i \tag{2.8}$$

and corresponding norm

$$\|u\| = \sqrt{\sum_{i=1}^{n} |u_i|^2}\,. \tag{2.9}$$

Here $\{u_i\}$ are just complex numbers — the components of the vector $\mathbf{u}$ in the canonical basis.

In physics, finite dimensional Hilbert spaces often arise as idealizations or "toy models". We may wish to illustrate some quantum phenomenon by first considering an especially simple case that has only finitely many (perhaps just two) different things it can do. Alternatively, we may be concerned with just a finite dimensional subspace of a larger physical system. For example, we may be interested in states at a particular energy level of a degenerate Hamiltonian, which can easily be finite dimensional. Also, we'll see later that the angular behavior of systems with a fixed (and finite) total angular momentum is also captured by a finite dimensional Hilbert space.

A simple infinite dimensional generalisation of this is the space of infinite sequences of complex number $\mathbf{u} = (u_1, u_2, \dots)$ such that

$$\sum_{i=1}^{\infty} |u_i|^2 < \infty \,. \tag{2.10}$$

This space is known as ℓ^2 and, heuristically, you can think of it as "$\mathbb{C}^n$ with $n = \infty$". The inner product between two such sequences $\mathbf{u}$ and $\mathbf{v}$ is defined as

$$(\mathbf{v}, \mathbf{u}) = \sum_{i=1}^{\infty} v_i^* u_i \tag{2.11}$$

as an obvious generalisation of the finite dimensional case. The Cauchy–Schwarz inequality gives $|(\mathbf{v}, \mathbf{u})| \leq \|\mathbf{v}\| \, \|\mathbf{u}\| < \infty$, so this inner product converges provided the norms do. Again, the notion of completeness of ℓ^2 with respect to this norm enables us to meaningfully take limits and, ultimately, differentiate vectors $\mathbf{u} \in \ell^2$.

In physics, ℓ^2 arises in many places. For example, the space of energy eigenstates in an infinite square well, or in the harmonic oscillator potential, may be thought of ℓ^2. Similarly, the space of *bound states* ($E < 0$) of the hydrogen atom is ℓ^2, though the scattering states ($E \geq 0$) are not.

We'll often also meet infinite dimensional vector spaces that are spaces of functions, such as the wavefunctions you dealt with throughout IB Quantum Mechanics. Given two functions $\psi, \phi : \mathbb{R} \to \mathbb{C}$, we can define linear combinations $a\psi + b\phi$ for any $a, b \in \mathbb{C}$ in the obvious way

$$(a\psi + b\phi)(x) = a\psi(x) + b\phi(x) \,, \tag{2.12}$$

where the multiplication and addition on the right-hand side are just those in $\mathbb{C}$, so spaces of functions are naturally infinite dimensional vector spaces (as you saw in IB Methods). To turn this space of functions into a Hilbert space, we first give it the norm

$$\|\psi\| = \sqrt{\int_{\mathbb{R}} \mathrm{d}x \; |\psi(x)|^2} \tag{2.13}$$

and require that $\|\psi\| < \infty$, i.e. the integral converges. In physics, functions for which $\|\psi\| < \infty$ holds are called *normalisable*. Note that just asking (2.13) to converge is not a very strong restriction, and so our normed function space contains a very wide class of functions, including all piecewise continuous functions that decay sufficiently rapidly as $|x| \to \infty$, and even some functions that are singular for some discrete values of x, provided these singularities are not strong enough to cause the integral to diverge[3].

[3]The requirement that the space be complete in the norm (2.13) is rather subtle. If $\|\psi - \phi\| = 0$, then we must identify ψ and ϕ as the same object in our space. This does not necessarily mean that they're identical as functions, because e.g. they could take different at some discrete points $x_i \subset \mathbb{R}$, as the non-zero value of $\psi - \phi$ at these discrete points would not contribute to the norm (2.13). In particular, any function that is non-zero only at a set of points that has a *Lebesgue measure* zero should be identified with the zero function. The resulting space is known as $L^2(\mathbb{R}, \mathrm{d}x)$ or sometimes just L^2 for short. (The L stands for Lebesgue, and is an example of a more general type of normed function space.) $L^2(\mathbb{R}, \mathrm{d}x)$ consists of equivalence classes of Cauchy sequences of functions that are convergent in the norm. In this course we'll mostly gloss over such technicalities, and they're certainly non-examinable. For a deeper discussion of Hilbert space, see the Part II Linear Analysis and Functional Analysis courses.

Finally, we take the inner product between two functions ψ and ϕ whose norms are finite to be

$$(\phi, \psi) = \int_{\mathbb{R}} \mathrm{d}x\, \phi(x)^* \psi(x)\,, \tag{2.14}$$

and again this converges by Cauchy–Schwarz. In physics, we often call $\psi(x)$ the *wavefunction* of the particle, and sometimes call (ϕ, ψ) the *overlap integral* between two wavefunctions.

2.1.2 Dual Spaces

As with any vector space, the *dual* $\mathcal{H}^*$ of a Hilbert space $\mathcal{H}$ is the space of linear maps $\mathcal{H} \to \mathbb{C}$. That is, an element $\varphi \in \mathcal{H}^*$ defines a map $\varphi : \psi \mapsto \varphi(\psi)$ for every $\psi \in \mathcal{H}$, such that

$$\varphi : a\psi_1 + b\psi_2 \longmapsto a\varphi(\psi_1) + b\varphi(\psi_2) \tag{2.15}$$

for all $\psi_1, \psi_2 \in \mathcal{H}$ and $a, b \in \mathbb{C}$.

One way to construct such a map is to use the inner product. Given some state $\phi \in \mathcal{H}$, we can define an element $(\phi, \ \ ) \in \mathcal{H}^*$ which acts on $\psi \in \mathcal{H}$ by

$$(\phi, \ \ ) : \psi \longmapsto (\phi, \psi)\,, \tag{2.16}$$

i.e. we take the inner product of $\psi \in \mathcal{H}$ with our chosen element ϕ. The linearity properties of the inner product transfer to ensure that $(\phi, \ \ )$ is indeed a linear map — note that since the inner product is antilinear in its first entry, it's important that our chosen element ϕ to sit in this first entry. In IB Linear Algebra you proved that, in finite dimensions, every element of $\mathcal{H}$ arises this way. That is, any linear map $\varphi : \mathcal{H} \to \mathbb{C}$ can be written as $(\phi, \ \ )$ for some $\phi \in \mathcal{H}$. This means in particular that the inner product $(\ ,\ )$ provides a vector space isomorphism

$$\begin{aligned} &\mathcal{H} \cong \mathcal{H}^*\,, \\ &\phi \leftrightarrow (\phi, \ \ ) = \phi^\dagger\,. \end{aligned} \tag{2.17}$$

The notation † is known as the *Hermitian conjugate* that transform a vector $\phi \in \mathcal{H}$ into its dual $\phi^\dagger \in \mathcal{H}$. Comfortingly, the same result also holds in infinite dimensions[4], but it's non-trivial to prove and is known as the *Riesz Representation Theorem*. The isomorphism $\mathcal{H}^* \cong \mathcal{H}$ is what is special about Hilbert spaces among various other infinite dimensional vector spaces, and makes them especially easy to handle.

2.1.3 Dirac Notation and Continuum States

From now on, in this course we'll use a notation for Hilbert spaces that was introduced by Dirac and is standard throughout the theoretical physics literature. Dirac denotes an element of $\mathcal{H}$ as $|\psi\rangle$, where the symbol "$|\ \ \rangle$" is known as a *ket*. An element of the dual space is written $\langle\phi|$ and the symbol "$\langle\ \ |$" is called a *bra*. The relation between the ket $|\phi\rangle \in \mathcal{H}$ and the bra $\langle\phi| \in \mathcal{H}^*$ is what we would previously have written as ϕ versus $(\phi, \ )$. We can again use Hermitian conjugate to transform a ket into a bra: $|\phi\rangle^\dagger \equiv \langle\phi|$, or vice versa, $\langle\phi|^\dagger \equiv |\phi\rangle$. The inner product between two states $|\psi\rangle\,, |\phi\rangle \in \mathcal{H}$ is then written $\langle\phi|\psi\rangle$ forming a *bra-ket* or bracket. Note that everything above implicitly uses the

[4] We should really be more careful here, though we won't be concerned with the following subtleties in this course. In infinite dimensions we distinguish the *algebraic dual space* — the space of all linear functionals $\varphi : H \to \mathbb{C}$ from the *continuous dual*, where the number $\varphi(\psi)$ is required to vary continuously as ψ varies in $\mathcal{H}$. The Riesz Representation Theorem applies to the continuous dual. For example, the algebraic dual also contains distributions such as the Dirac delta, δ, acting as $\delta : \psi \mapsto \psi(\mathbf{0})$"$= \int_{\mathbb{R}} \mathrm{d}x\, \delta(x)\psi(x)$", with $\psi \in L^2(\mathbb{R}, \mathrm{d}x)$. Certainly $\delta(x)$ is not itself square-integrable, so is not in the Hilbert space. However, since $L^2(\mathbb{R}, \mathrm{d}x\,)$ contains discontinuous functions, $\delta[\psi]$ does not have to vary smoothly with ψ. Functional analysis is almost always interested in the continuous dual, and thus this is often called just the *dual space*.

isomorphism $\mathcal{H}^* \cong \mathcal{H}$ provided by the inner product, building it into the notation. Recall also that in all physics courses, $\langle\phi|\psi\rangle$ is antilinear in $\langle\phi|$.

Given an orthonormal basis $\{|e_i\rangle\}$ of $\mathcal{H}$, at least in the cases $\mathcal{H} \cong \mathbb{C}^n$ or $\mathcal{H} \cong \ell^2$, in Dirac notation we can expand a general ket $|\psi\rangle \in \mathcal{H}$ as

$$|\psi\rangle = \sum_i \psi_i \, |e_i\rangle \tag{2.18}$$

in terms of this basis. Then the inner product of $\chi, \psi \in \mathcal{H}$ can be expanded as

$$\langle\chi|\psi\rangle = \sum_{i,j} \chi_j^* \psi_i \, \langle e_j|e_i\rangle = \sum_i \chi_i^* \psi_i \tag{2.19}$$

as usual.

It's very useful to be able to extend this idea also to function spaces. In this case, we introduce a *continuum basis* with elements $|a\rangle$ labelled by a continuous variable a, normalised so that

$$\langle a'|a\rangle = \delta(a' - a) \tag{2.20}$$

using the Dirac δ-function. To expand a $|\psi\rangle$ in a continuum basis, we need to replace the summation in (2.18) and write

$$|\psi\rangle = \int \mathrm{d}a \, \psi(a) \, |a\rangle \ , \tag{2.21}$$

where $\psi(a)$ is the component of $|a\rangle$. The point of the normalisation (2.20) is that

$$\langle\chi|\psi\rangle = \int \mathrm{d}b \int \mathrm{d}a \, \chi(b)^* \psi(a) \, \langle b|a\rangle = \int \mathrm{d}b \int \mathrm{d}a \, \chi(b)^* \psi(a) \delta(b-a) = \int \mathrm{d}a \, \chi(a)^* \psi(a) \, , \tag{2.22}$$

which is just the inner product (and also norm) we defined for $L^2(\mathbb{R}, \mathrm{d}a)$ before. Indeed, a key example of a continuum basis is the *position basis* $\{|x\rangle\}$, where $x \in \mathbb{R}$. Expanding a general state $|\psi\rangle$ as an integral gives

$$|\psi\rangle = \int_{\mathbb{R}} \mathrm{d}x' \, \psi(x') \, |x'\rangle \ . \tag{2.23}$$

We see that the complex coefficients are

$$\langle x|\psi\rangle = \int_{\mathbb{R}} \mathrm{d}x' \, \psi(x') \, \langle x|x'\rangle = \psi(x) \, . \tag{2.24}$$

In other words, the position space wavefunctions we're familiar with are nothing but the coefficients of a state $|\psi\rangle \in \mathcal{H}$ in a particular position continuum basis.

As always, we could equally choose to expand this same vector in different bases. For example, our state $|\psi\rangle = \int_{\mathbb{R}} \mathrm{d}x \, \psi(x) \, |x\rangle$ from above can equally be expanded in the momentum basis as $|\psi\rangle = \int_{\tilde{\mathbb{R}}} \mathrm{d}p \, \tilde{\psi}(p) \, |p\rangle$, where the new coefficients $\tilde{\psi}(p) = \langle p|\psi\rangle$ are the *momentum space wavefunction.* Later, we'll show that $\langle x|p\rangle = \mathrm{e}^{\mathrm{i}xp/\hbar}/\sqrt{2\pi\hbar}$, so these two sets of coefficients are related by

$$\langle x|\psi\rangle = \int_{\tilde{\mathbb{R}}} \mathrm{d}p \, \tilde{\psi}(p) \, \langle x|p\rangle = \frac{1}{\sqrt{2\pi\hbar}} \int_{\tilde{\mathbb{R}}} \mathrm{d}p \, \mathrm{e}^{\mathrm{i}xp/\hbar} \tilde{\psi}(p) \, , \tag{2.25}$$

$$\langle p|\psi\rangle = \int_{\mathbb{R}} \mathrm{d}x \, \psi(x) \, \langle p|x\rangle = \frac{1}{\sqrt{2\pi\hbar}} \int_{\mathbb{R}} \mathrm{d}x \, \mathrm{e}^{-\mathrm{i}xp/\hbar} \psi(x) \, . \tag{2.26}$$

Equations (2.25) and (2.26) are just the statements that the position and momentum space wavefunctions are each other's Fourier transforms, again familiar from IB Quantum Mechanics.

The real point I wish to make is that the fundamental object is the abstract vector $|\psi\rangle \in \mathcal{H}$. All the physical information about a quantum system is encoded in its state vector $|\psi\rangle$; the wavefunctions

$\psi(x)$ or $\tilde{\psi}(p)$ are merely the expansion coefficients in some basis. Like any other choice of basis, this expansion may be useful for some purposes and unhelpful for others.

Finally, a technical point. Although using continuum bases such as $\{|x\rangle\}$ and $\{|p\rangle\}$ is convenient, because it emphasizes the similarities between the finite and infinite-dimensional cases, it blurs the analysis. In particular, if $\langle x'|x\rangle = \delta(x'-x)$ then the norm

$$\||x\rangle\|^2 = \delta(x-x) = \delta(0)\,. \tag{2.27}$$

So whatever these objects $|x\rangle$ are, they certainly do not lie in our Hilbert space. It is possible to make good mathematical sense of these by appropriately enlarging our spaces to include spaces of distributions, but for the most part (and certainly in this course) physicists are content to say that continuum states such as $|x\rangle$ are allowable as basis elements, but call them non-normalizable states: actual physical particles are never represented by a non-normalizable state.

2.2 Operators

A *linear operator* A is a map $A : \mathcal{H} \to \mathcal{H}$ that is compatible with the vector space structure in the sense that

$$A(c_1\,|\psi_1\rangle + c_2\,|\psi_2\rangle) = c_1 A\,|\psi_1\rangle + c_2 A\,|\psi_2\rangle \tag{2.28}$$

in Dirac notation[5]. All the operators we meet in Quantum Mechanics will be linear, so henceforth we'll just call them "operators". Operators form an algebra: given two such linear operators A, B, we define their sum $\alpha A + \beta B$ as

$$(\alpha A + \beta B) : |\psi\rangle \mapsto \alpha A\,|\psi\rangle + \beta B\,|\psi\rangle \tag{2.29}$$

for all $\alpha, \beta \in \mathbb{C}$ and all $|\psi\rangle \in \mathcal{H}$, and take their product AB to be the composition

$$AB : |\psi\rangle \mapsto A \circ B\,|\psi\rangle \equiv A(B\,|\psi\rangle) \tag{2.30}$$

for $|\psi\rangle \in \mathcal{H}$. One can check that both the sum and product of two linear operators is again a linear operator[6].

The operator algebra is associative, so $A(BC) = (AB)C$, but not commutative and in general $AB \neq BA$ so the order in which the operators act is important. The difference between these two actions is known as the *commutator*

$$[A, B] = AB - BA\,. \tag{2.32}$$

The commutator obeys the following properties:

- *anti-symmetry.* $[A, B] = -[B, A]$;

[5]Here we are being a little sloppy again. In the infinite dimensional case, it often happens that operators are not defined on the whole of $\mathcal{H}$, but just on some domain $\mathrm{Dom}\,A \subseteq \mathcal{H}$ which depends on the operator itself. For example, the momentum operator P acts on position space wavefunctions $\psi(x) \in L^2(\mathbb{R}, \mathrm{d}x)$ by $-\mathrm{i}\hbar\partial/\partial x$. If $\psi(x)$ is discontinuous, then it could be that $\int \mathrm{d}x\,|\psi'|^2$ diverges even though $\int \mathrm{d}x\,|\psi|^2$ itself is finite. Understanding the correct domain of various operators is an important part of functional analysis. We'll largely ignore such subtleties in this course.

[6]A linear operator is *bounded* if, for all $|\psi\rangle \in \mathcal{H}$,

$$\|A\,|\psi\rangle\| \leq M\,\||\psi\rangle\| \tag{2.31}$$

for some $M > 0$. The space of such bounded linear operators on $\mathcal{H}$ is denoted as $B(\mathcal{H})$. Bounded linear operators thus map normalisable states to normalisable states, and so act on the whole Hilbert space $\mathcal{H}$. This usually makes them the nicest ones to deal with. One of the reasons we'll be largely ignoring the analysis aspects of Hilbert space in this course is that the operators we commonly deal with in quantum mechanics, such as the position and momentum operators X and P, are unbounded. (For example, even when its wavefunction is correctly normalised, a particle can be located at arbitrarily large $\mathbf{x} \in \mathbb{R}^3$, or have arbitrarily large momentum.) It's of course possible to handle such unbounded operators rigorously, but doing so involves an extra layer of technicality that would take us too far afield here. For further discussion, see that Part II course on Analysis of Functions.

- *linearity.* $[\alpha_1 A_1 + \alpha_2 A_2, B] = \alpha_1 [A_1, B] + \alpha_2 [A_2, B]\ \forall \alpha_1, \alpha_2 \in \mathbb{C}$;
- *Leibniz identity.* $[A, BC] = [A, B]C + B[A, C]$;
- *Jacobi identity.* $[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0$.

These are easy to verify from the definitions. Commutators of operators arise frequently in quantum mechanics.

A state $|\psi\rangle \in \mathcal{H}$ is said to be an *eigenstate* of an operator A if

$$A\,|\psi\rangle = \lambda\,|\psi\rangle\ , \tag{2.33}$$

where the number $\lambda \in \mathbb{C}$ is known as its *eigenvalue*. Thus, if $|\psi\rangle$ is an eigenstate of A, acting on $|\psi\rangle$ with A returns exactly the same vector, except that it is rescaled by a number that may depend both on the state and the particular operator. The set of all eigenvalues of an operator A is sometimes called the *spectrum* of A, while the number of linearly independent eigenstates having the same eigenvalue is called the degeneracy of this eigenvalue.

One place where Dirac notation is particularly convenient is that it allows us to label states by their eigenvalues. We let $|q\rangle$ denote an eigenstate of some operator Q with eigenvalue q, so that $Q\,|q\rangle = q\,|q\rangle$. For example, a state that is an eigenstate of the position operator $\mathbf{X}$ with position $\mathbf{x} \in \mathbb{R}^3$ (representing a particle that is definitely located at $\mathbf{x}$) is written $|\mathbf{x}\rangle$, while the state $|\mathbf{p}\rangle$ is an eigenstate of the momentum operator $\mathbf{P}$ with eigenvalue $\mathbf{p}$. There's a potential confusion here: $|\mathbf{x}\rangle$ does not refer to the function $\mathbf{x}$, but rather a state in $\mathcal{H}$ whose only support is at $\mathbf{x}$. We saw above that the wavefunction corresponding to $|\mathbf{x}\rangle$ was really a δ function.

In this course, I'll mostly use the convention that operators are written using capital letters, with their eigenvalues being labelled by the same letter in lowercase. However, I won't stick to this religiously; a notable and deeply ingrained exception is to use E for the eigenvalues of the Hamiltonian operator H.

The extra structure of the inner product on $\mathcal{H}$ allows us to also define the *Hermitian conjugate* or *adjoint* $A^\dagger$ of an operator A by

$$\langle\phi|\,A^\dagger = (A\,|\phi\rangle)^\dagger\ , \tag{2.34}$$

or, equivalently,

$$\big\langle\phi\big|A^\dagger\big|\psi\big\rangle = (A\,|\phi\rangle)^\dagger\,|\psi\rangle = \langle\psi|A|\phi\rangle^* \tag{2.35}$$

for all $|\phi\rangle\,, |\psi\rangle \in \mathcal{H}$. One can easily check that[7]

$$(A+B)^\dagger = A^\dagger + B^\dagger\,,\ (AB)^\dagger = (BA)^\dagger\,,\ (\alpha A)^\dagger = \alpha^* A^\dagger \text{ and } (A^\dagger)^\dagger = A\,, \tag{2.36}$$

from which it follows that $[A, B]^\dagger = [B^\dagger, A^\dagger]$. Note also that the adjoint of the eigenvalue equation $A\,|a\rangle = a\,|a\rangle$ is

$$\langle a|\,A^\dagger = \langle a|\,a^*\,, \tag{2.37}$$

where $A^\dagger$ is taken to act to the left on the dual vector $\langle a|$.

An operator Q is called *Hermitian*[8] if $Q^\dagger = Q$, so that $\langle\phi|Q|\psi\rangle = \langle\psi|Q|\phi\rangle^*$ for all $|\phi\rangle\,, |\psi\rangle \in \mathcal{H}$. Hermitian operators are very special and have a number of important properties. Firstly, suppose $|q\rangle$

[7] At least for the finite dimensional case. We'll assume without being careful that it holds in the infinite dimensional case too.

[8] The terminology *Hermitian* is used in the physics literature, coming from the fact that if $\mathcal{H}$ is finite dimensional, we can represent linear operators obeying $Q^\dagger = Q$ by Hermitian matrices. In functional analysis, we'd have to be a little more careful about exactly which states we allow our operator to act on. For example, differentiating a non-smooth but square-integrable function will typically sharpen its singularities and may render it non-normalisable. Thus the momentum operator $-\mathrm{i}\hbar\mathrm{d}/\mathrm{d}x$ may take a state in $\mathcal{H} \cong L^2(\mathbb{R}, \mathrm{d}x)$ out of $L^2(\mathbb{R}, \mathrm{d}x)$, and to keep analytic control we should first limit the states on which we allow the momentum operator to act. Being more careful, we'd say that operators obeying $\langle\psi|\,Q\,|\phi\rangle = \langle\phi|\,Q\,|\psi\rangle^*$ for all $|\phi\rangle\,, |\psi\rangle \in \mathrm{Dom}\,Q \subseteq \mathrm{Dom}\,Q^\dagger \subseteq \mathcal{H}$ and $\mathrm{Dom}\,Q$ dense in $\mathcal{H}$ are *symmetric*, while symmetric operators for which $\mathrm{Dom}\,Q = \mathrm{Dom}\,Q^\dagger$ are called *self-adjoint*. The distinction between operators that are self-adjoint and those that are merely symmetric is a little technical, but has important consequences for their spectrum and the existence of eigenstates. See e.g. this Wikipedia page for a basic discussion with examples. We'll largely ignore such subtleties in our course.

is an eigenstate of a Hermitian operator Q with eigenvalue q, then

$$q \langle q|q\rangle = \langle q|Q|q\rangle = \langle q| Q |q\rangle^* = q^* \langle q|q\rangle \ , \tag{2.38}$$

so $q = q^*$. Hence, the eigenvalues of a Hermitian operator are real. Second, suppose $|q_1\rangle$ and $|q_2\rangle$ are both eigenstates of Q with distinct eigenvalues q_1 and q_2. Then

$$(q_1 - q_2) \langle q_1|q_2\rangle = \langle q_1| Q^\dagger |q_2\rangle - \langle q_1| Q |q_2\rangle = 0 \, , \tag{2.39}$$

where the first equality uses the fact that q_1 is real and the second uses the fact that Q is Hermitian. Since $q_1 \neq q_2$ by assumption, we must have $\langle q_1|q_2\rangle = 0$, so the eigenstates of Hermitian operators with distinct eigenvalues are orthogonal, and after rescaling, we can choose them to be orthonormal.

In the finite dimensional case, you proved in IB Linear Algebra that the set of eigenvectors of a given Hermitian operator form a basis of $\mathcal{H}$; that is, they form a complete, orthonormal set. This property allows us to express Hermitian operators in a form that is often useful. If $\{|n\rangle\}$ is an orthonormal basis of eigenstates of a Hermitian operator Q, with eigenvalues $\{\lambda_n\}$, then we can write

$$Q = \sum_n \lambda_n |n\rangle \langle n| \ , \tag{2.40}$$

where we think of this as acting on $|\psi\rangle \in \mathcal{H}$ by

$$Q |\psi\rangle = \sum_n \lambda_n |n\rangle \langle n|\psi\rangle \ . \tag{2.41}$$

Note that indeed $Q|n\rangle \in \mathcal{H}$, since the $\langle n|\psi\rangle$ are just numbers in $\mathbb{C}$, whereas each term in the sum also involves $|n\rangle \in \mathcal{H}$. In particular, expressing $|\psi\rangle$ in this basis as $|\psi\rangle = \sum_m c_m |m\rangle$ gives

$$Q|\psi\rangle = \sum_{n,m} \lambda_n c_m |n\rangle \langle n|m\rangle = \sum_n \lambda_n c_n |n\rangle \ . \tag{2.42}$$

One reason this representation of Q is useful is that it allows us to define functions of operations. We set

$$f(Q) = \sum_n f(\lambda_n) |n\rangle \langle n| \ , \tag{2.43}$$

provided $f(\lambda_n)$ is defined for all eigenvalues λ_n of the original operator. For example, the inverse of an operator Q is the operator

$$Q^{-1} = \sum_n \frac{1}{\lambda_n} |n\rangle \langle n| \tag{2.44}$$

provided $\lambda_n \neq 0$ for all n, while

$$\ln Q = \sum_n \ln(\lambda_n) |n\rangle \langle n| \tag{2.45}$$

when the eigenvalues of Q are all positive.

A particularly important example of this is that we can represent the identity operator Id on $\mathcal{H}$ as

$$\mathrm{Id} = \sum_n |n\rangle \langle n| \ \text{ or } \ \mathrm{Id} = \int \mathrm{d}q \ |q\rangle \langle q| \ , \tag{2.46}$$

where $\{|n\rangle\}$ or $\{|q\rangle\}$ are any bases of $\mathcal{H}$ respectively for discrete or continuous labels. This is often convenient for moving between different bases. For example, if $\{|p\rangle\}$ denotes a complete set of eigenstates of the momentum operator P, with $P|p\rangle = p|p\rangle$, then to represent this operator acting on a state $|\psi\rangle$ in the position basis we write

$$\begin{aligned}\langle x|P|\psi\rangle &= \int \mathrm{d}p \ \langle x|P|p\rangle \langle p|\psi\rangle = \int \mathrm{d}p \ \langle x|p\rangle \, p\tilde{\psi}(p) \\ &= \frac{1}{\sqrt{2\pi\hbar}} \int \mathrm{d}p \, \mathrm{e}^{\mathrm{i}xp/\hbar} p\tilde{\psi}(p) = -\mathrm{i}\hbar \frac{\partial}{\partial x} \left[\frac{1}{\sqrt{2\pi\hbar}} \int \mathrm{d}p \, \mathrm{e}^{\mathrm{i}xp/\hbar} \tilde{\psi}(p) \right] \ . \end{aligned} \tag{2.47}$$

In the first equality here we've squeezed the identity operator $\mathrm{Id} = \int \mathrm{d}p\, |p\rangle \langle p|$ in between the momentum operator P and the general state $|\psi\rangle$, then used the fact that $|p\rangle$ is an eigenstate of P, and finally used our previous expression for $\langle x|p\rangle$ and standard properties of Fourier transforms. We recognise the integral in the final expression as the position space wavefunction of $|\psi\rangle$, so altogether we have

$$\langle x|P|\psi\rangle = -\mathrm{i}\hbar \frac{\partial}{\partial x} \langle x|\psi\rangle \ , \tag{2.48}$$

or $\hat{P}\psi(x) = -\mathrm{i}\hbar\partial\psi/\partial x$ in IB notation. This and similar manipulations will be used repeatedly throughout the course, so it's important to make sure you're comfortable with them.

If $\mathcal{H}$ is finite dimensional, we can represent any linear operator on $\mathcal{H}$ by a matrix. We pick an orthonormal basis $|n\rangle$ and define

$$A_{nm} = \langle n|A|m\rangle \tag{2.49}$$

as usual. In particular, inserting the identity operator in the form $\mathrm{Id} = \sum_n |n\rangle \langle n|$ shows that the matrix elements of the product operator AB as

$$\begin{aligned}(AB)_{km} = \langle k|AB|m\rangle &= \left\langle k \middle| A \left(\sum_n |n\rangle \langle n| \right) B \middle| m \right\rangle \\ &= \sum_n \langle k|A|n\rangle \langle n|B|m\rangle = \sum_n A_{kn} B_{nm} \ ,\end{aligned} \tag{2.50}$$

which just corresponds to usual matrix multiplication. Hermitian operators are represented by Hermitian matrices, i.e. ones whose elements obey $(\mathsf{A}^\dagger)_{mn} := A^*_{nm} = A_{mn}$, hence the name.

2.3 Composite Systems

We often have to deal with systems with more than a single degree of freedom. This could be just a single particle moving in three dimensions rather than one, or a composite system such as a hydrogen atom consisting of both an electron and a proton, or a diamond which consists of a very large number of carbon atoms. In the case of a macroscopic object, our system could be made up of a huge number of constituent parts, each able to behave differently, at least in principle.

It should be intuitively clear that the more complicated our system is, meaning the more independent degrees of freedom it possesses, the larger a Hilbert space we'll need if we wish to encode all its possible states. We'll now understand how quantum mechanics handles systems with more than one degree of freedom by taking the tensor product of the Hilbert spaces of its constituent parts.

2.3.1 Tensor Product of Hilbert Spaces

Hilbert spaces are vector spaces over $\mathbb{C}$, so we can try to define their tensor product in the same way as we would for any pair of vector spaces. Recall from IB Linear Algebra that if $\mathcal{H}_1$ and $\mathcal{H}_2$ are two finite Hilbert spaces, with $\{|e_a\rangle\}_{a=1}^m$ a basis of $\mathcal{H}_1$ and $\{|f_\alpha\rangle\}_{\alpha=1}^n$ a basis of $\mathcal{H}_2$, then the *tensor product space* $\mathcal{H}_1 \otimes \mathcal{H}_2$ is again a vector space over $\mathbb{C}$ spanned by all pairs of elements $|e_a\rangle \otimes |f_\alpha\rangle$ chosen from the two bases[9]. It follows that

$$\dim(\mathcal{H}_1 \otimes \mathcal{H}_2) = \dim(\mathcal{H}_1)\dim(\mathcal{H}_2) \tag{2.52}$$

[9] If you're a purist (and the PQM Tripos examiners are not), then there's something slightly unsatisfactory about this definition of the tensor product vector space, which is that it apparently depends on a choice of basis on both $\mathcal{H}_1$ and $\mathcal{H}_2$. We can do better, just using the abstract vector space structure, as follows. For any pair of vector spaces V and W, we first define the *free vector space* $F(V \times W)$ to be the set of all possible pairs of elements chosen from V and W. This is a vector space with the sum of a pair (v, w) and a pair (v', w') denoted simply by $(v, w) + (v', w')$ (Do not confuse the pair $(v, w) \in F(V \times W)$ with an inner product! Here v and w live in different spaces.) By itself, $F(V \times W)$ does not care that V and W are already vector spaces. For example, if e_i is a basis for V and $v = c_1 e_1 + c_2 e_2$ for some coefficients c_1, c_2, the elements (v, w) and $c_1(e_1, w) + c_2(e_2, w)$ are nonetheless distinct in $F(V \times W)$. We'd like

provided these are finite. This should be familiar from IB Linear Algebra.

It's important to stress that, as with any vector space, a general element of $\mathcal{H}_1 \otimes \mathcal{H}_2$ is a linear combination of its basis elements $|e_a\rangle \otimes |f_\alpha\rangle$. In particular, although general elements of $\mathcal{H}_1$ and $\mathcal{H}_2$ can be written

$$|\psi_1\rangle = \sum_{a=1}^{m} c_a |e_a\rangle \quad \text{and} \quad |\psi_2\rangle = \sum_{\alpha=1}^{n} d_\alpha |f_\alpha\rangle \ , \tag{2.53}$$

it is not true that a general element of $\mathcal{H}_1 \otimes \mathcal{H}_2$ necessarily takes the form

$$|\psi_1\rangle \otimes |\psi_2\rangle = \sum_{a,\alpha} c_a d_\alpha |e_a\rangle \otimes |f_\alpha\rangle \ . \tag{2.54}$$

Rather, a general element of $\mathcal{H}_1 \otimes \mathcal{H}_2$ is written as

$$|\Psi\rangle = \sum_{a,\alpha} r_{a,\alpha} |e_a\rangle \otimes |f_\alpha\rangle \ . \tag{2.55}$$

In particular, elements of the form $|\psi_1\rangle \otimes |\psi_2\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ are only specified by $\dim \mathcal{H}_1 + \dim \mathcal{H}_2$ coefficients, vastly fewer than is required to specify a generic element. Elements of the form $|\psi\rangle \otimes |\phi\rangle$ are sometimes called simple, while in physics generic elements of the form (2.55) are said to be entangled. In section 10 we'll explore some of the vast resources this entanglement opens up to quantum mechanics; you'll see much more of it next term if you take the course on Quantum Information and Computation.

In order to make $\mathcal{H}_1 \otimes \mathcal{H}_2$ a Hilbert space, rather than just a vector space, we must give it an inner product. We do this in the obvious way. Let $\langle \ | \ \rangle_1$ and $\langle \ | \ \rangle_2$ denote inner products on $\mathcal{H}_1$ and $\mathcal{H}_2$. We first define the inner product between each pair of basis elements of $\mathcal{H}_1 \otimes \mathcal{H}_2$ by

$$(\langle e_a| \otimes \langle f_\alpha|)(|e_b\rangle \otimes |f_\beta\rangle) = \langle e_a|e_b\rangle_1 \langle f_\alpha|f_\beta\rangle_2 \ , \tag{2.56}$$

and then extend it to arbitrary states $|\Psi\rangle$ by linearity. Note in particular that if $\{|e_a\rangle\}$ and $\{f_\alpha\}$ are orthonormal bases of their respective Hilbert spaces, then $\{|e_a\rangle \otimes |f_\alpha\rangle\}$ will be an orthonormal basis of $\mathcal{H}_1 \otimes \mathcal{H}_2$.

The most common occurrence of this is simply a single particle moving in more than one dimension. For example, suppose a quantum particle moves in $\mathbb{R}^2$, described by Cartesian coordinates (x, y). If $\{|x\rangle\}_{x\in\mathbb{R}}$ is a complete set of position eigenstates for the x-direction, and $\{|y\rangle\}_{y\in\mathbb{R}}$ is likewise a complete set for the y-direction, then the state of the particle can be expanded as

$$|\psi\rangle = \int_{\mathbb{R}\times\mathbb{R}} \mathrm{d}x\, \mathrm{d}y\, \psi(x, y)\, |x\rangle \otimes |y\rangle \ , \tag{2.57}$$

where $\{|x\rangle \otimes |y\rangle\}_{x,y\in\mathbb{R}}$ form a continuum basis of the tensor product space. Note that in general, our wavefunction is not a simple product $\psi(x, y) \neq \psi_1(x)\psi_2(y)$ of wavefunctions in the two directions separately, though it may be possible to write it as the sum of such products, as you met repeatedly when studying the separation of variables in IB Methods. In this continuum case, the inner product between any two states is

$$\begin{aligned}
\langle\chi|\psi\rangle &= \int_{\mathbb{R}^2\times\mathbb{R}^2} \mathrm{d}x'\, \mathrm{d}y'\, \mathrm{d}x\, \mathrm{d}y\, \chi(x', y')^* \psi(x, y)\, \langle x'|x\rangle\, \langle y'|y\rangle \\
&= \int_{\mathbb{R}^2\times\mathbb{R}^2} \mathrm{d}x'\, \mathrm{d}y'\, \mathrm{d}x\, \mathrm{d}y\, \chi(x', y')^* \psi(x, y) \delta(x - x')\delta(y - y') \\
&= \int_{\mathbb{R}^2} \mathrm{d}x\, \mathrm{d}y\, \chi(x, y)^* \psi(x, y) \, .
\end{aligned} \tag{2.58}$$

to teach the free vector space about the structure inherited from V and W individually. We thus define the tensor product $V \otimes W$ to be the *quotient* $F(V \times W)/\sim$ under the equivalence relations

$$(v, w) + (v', w) \sim (v + v', w)\, , \ (v, w) + (v, w') \sim (v, w + w') \text{ and } c(v, w) \sim (cv, w) \sim (v, cw) \tag{2.51}$$

for all $v \in V$, $w \in W$ and $c \in \mathbb{C}$. We denote the equivalence class of the pair (v, w) by $v \otimes w$.

This is exactly the structure of $L^2(\mathbb{R}^2, \mathrm{d}^2\mathbf{x})$, and we identify[10] $L^2(\mathbb{R}^2, \mathrm{d}^2\mathbf{x}) \cong L^2(\mathbb{R}, \mathrm{d}x) \otimes L^2(\mathbb{R}, \mathrm{d}y)$.

More commonly in physics, we consider quantum systems that live in $\mathbb{R}^3$, described by a Hilbert space $\mathcal{H} \cong L^2(\mathbb{R}^3, \mathrm{d}^3\mathbf{x})$ obtained from the tensor product of three copies of the Hilbert space for a one-dimensional system. We can expand a general state in this Hilbert space as

$$|\psi\rangle = \int_{\mathbb{R}^3} \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,\psi(x,y,z)\,|x\rangle \otimes |y\rangle \otimes |z\rangle = \int_{\mathbb{R}^3} \mathrm{d}^3\mathbf{x}\,\psi(\mathbf{x})\,|\mathbf{x}\rangle\ , \tag{2.61}$$

where the last expression is just a convenient shorthand. Going further, to describe a composite system such as a hydrogen atom that consists of two particles (an electron and a proton), we need to take the tensor product of the individual Hilbert spaces of each particle. Thus (neglecting the spin of the electron and proton) the atom is described by a state

$$|\Psi\rangle = \int_{\mathbb{R}^6} \mathrm{d}^3\mathbf{x}_e\,\mathrm{d}^3\mathbf{x}_p\,\Psi(\mathbf{x}_e, \mathbf{x}_p)\,|\mathbf{x}_e\rangle \otimes |\mathbf{x}_p\rangle\ , \tag{2.62}$$

where $\Psi(\mathbf{x}_e, \mathbf{x}_p) = (\langle\mathbf{x}_e| \otimes \langle\mathbf{x}_p|)\,|\Psi\rangle$ is the amplitude for the atom's electron to be found at $\mathbf{x}_e$ while its proton is at $\mathbf{x}_p$.

Once we know the Hilbert spaces appropriate to describe the fundamental constituents of our system, we can build up the Hilbert space for the combined system by taking tensor products. We should then ask, *"What is the correct Hilbert space to use to describe the fundamental particles in our system?"*. Ultimately, this question can only be determined by carrying out an experiment. For example, experiments performed by Stern and Gerlach (see section 5.3.2) showed that a single electron is in fact described by a Hilbert space $\mathcal{H}_e = L^2(\mathbb{R}^3, \mathrm{d}^3\mathbf{x}) \otimes \mathbb{C}^2$, formed from the tensor product of the electron's position space wavefunction with the two-dimensional Hilbert space $\mathbb{C}^2$ describing the electron's internal degree of freedom. Thus, letting $\{|\uparrow\rangle, |\downarrow\rangle\}$ be an orthonormal basis of $\mathbb{C}^2$, a generic state of the electron (whether or not it's part of a hydrogen atom) is

$$|\Psi\rangle = |\phi\rangle \otimes |\uparrow\rangle + |\chi\rangle \otimes |\downarrow\rangle\ , \tag{2.63}$$

given in the position representation by

$$\langle\mathbf{x}_e|\Psi\rangle = \phi(\mathbf{x}_e)\,|\uparrow\rangle + \chi(\mathbf{x}_e)\,|\downarrow\rangle\ , \tag{2.64}$$

involving a pair of wavefunction $\phi, \chi \in L^2(\mathbb{R}^3, \mathrm{d}^3\mathbf{x})$. We'll investigate this in detail in section 5.3.

Let's now understand how operators act on our composite system. Given linear operators $A : \mathcal{H}_1 \to \mathcal{H}_1$ and $B : \mathcal{H}_2 \to \mathcal{H}_2$, we define the linear operator $A \otimes B : \mathcal{H}_1 \otimes \mathcal{H}_2 \to \mathcal{H}_1 \otimes \mathcal{H}_2$ by

$$A \otimes B : |e_a\rangle \otimes |f_\alpha\rangle \longmapsto (A\,|e_a\rangle) \otimes (B\,|e_\alpha\rangle)\,, \tag{2.65}$$

[10] There is actually one more step in the infinite dimensional case: we must take the *completion* of $L^2(\mathbb{R}, \mathrm{d}x) \otimes L^2(\mathbb{R}, \mathrm{d}y)$ in the norm associated to this inner product. There is no problem if our general state $\psi(x,y)$ can be written as a finite sum of products of square-integrable wavefunctions, say $\phi_a(x)$ and $\rho_b(y)$ on each of the two directions separately, for then

$$\begin{aligned}\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,|\psi(x,y)|^2 &= \int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y \sum_{a,a',b,b'} \phi_{a'}(x)^*\rho_{b'}(y)^*\phi_a(x)\phi_b(y) \\ &= \left[\sum_{a,a'} \int_{\mathbb{R}} \mathrm{d}x\,\phi_{a'}(x)^*\phi_a(x)\right]\left[\sum_{b,b'} \int_{\mathbb{R}} \mathrm{d}y\,\rho_{b'}(y)^*\rho_b(y)\right] < \infty\,,\end{aligned} \tag{2.59}$$

with convergence guaranteed by the properties of the inner product on each copy of $L^2(\mathbb{R}, \mathrm{d}x)$. However, there are functions in $L^2(\mathbb{R}^2, \mathrm{d}^2\mathbf{x})$ that cannot be so expressed as a finite sum. A simple example is the function $f : \mathbb{R}^2 \to \mathbb{C}$ given by

$$f : (x,y) \mapsto \begin{cases} 1 & \text{when } \sqrt{x^2+y^2} \le a^2 \\ 0 & \text{otherwise.} \end{cases} \tag{2.60}$$

It is obvious that this function is square-integrable over $\mathbb{R}^2$, but it cannot be written as a finite sum of products of functions of x and y respectively. It thus lies in the completion of $L^2(\mathbb{R}, \mathrm{d}x) \otimes L^2(\mathbb{R}, \mathrm{d}y)$, but not in this space itself. The Hilbert space completion of $\mathcal{H}_1 \otimes \mathcal{H}_2$ is often denoted $\mathcal{H}_1 \hat{\otimes} \mathcal{H}_2$, so more correctly we have $L^2(\mathbb{R}^2, \mathrm{d}^2\mathbf{x}) \cong L^2(\mathbb{R}, \mathrm{d}x) \hat{\otimes} L^2(\mathbb{R}, \mathrm{d}y)$. We'll ignore such subtleties in this course, saving a proper treatment for the Functional Analysis course.

and we extend this definition to arbitrary states in $\mathcal{H}_1 \otimes \mathcal{H}_2$ by linearity. In particular, the operator A acting purely on $\mathcal{H}_1$ corresponds to the operator $A \otimes \mathrm{Id}_{\mathcal{H}_2}$ when acting on the tensor product, and similarly $\mathrm{Id}_{\mathcal{H}_1} \otimes B$ acts non-trivially just on the second factor in the tensor product. Note that since they act on separate Hilbert spaces,

$$[A \otimes \mathrm{Id}_{\mathcal{H}_2}, \mathrm{Id}_{\mathcal{H}_1} \otimes B] = 0 \tag{2.66}$$

for any operators A, B, even if these operators happen not to commute when acting on the same Hilbert space.

As an example, let's again consider the case of the hydrogen atom, where the Hamiltonian takes the form

$$H = \frac{\mathbf{P}_e^2}{2m_e} \otimes \mathrm{Id}_p + \mathrm{Id}_e \otimes \frac{\mathbf{P}_p^2}{2m_p} + V(\mathbf{X}_e, \mathbf{X}_p)\,. \tag{2.67}$$

The kinetic terms for the electron and proton each act only on one factor in the tensor product, while the Coulumbic interaction

$$V(\mathbf{X}_e, \mathbf{X}_p) = -\frac{e^2}{|\mathbf{X}_e - \mathbf{X}_p|} \tag{2.68}$$

depends on the location of each particle. In this case, since V depends only on the relative separation of the two, it's actually more convenient to view the tensor product differently, writing

$$\mathcal{H}_{\mathrm{Hyd}} = \mathcal{H}_e \otimes \mathcal{H}_p = \mathcal{H}_{\mathrm{com}} \otimes \mathcal{H}_{\mathrm{rel}} \tag{2.69}$$

to split it in terms of the states describing the behaviour of the centre of mass and those describing the relative states. Defining the centre of mass and relative operators

$$\mathbf{X}_{\mathrm{com}} = \frac{m_e \mathbf{X}_e + m_p \mathbf{X}_p}{m_e + m_p}\,,\ \mathbf{P}_{\mathrm{com}} = \mathbf{P}_e + \mathbf{P}_p\,, \tag{2.70}$$

$$\mathbf{X} = \mathbf{X}_e - \mathbf{X}_p\,,\ \mathbf{P} = \frac{m_p \mathbf{P}_e - m_e \mathbf{P}_p}{m_e + m_p}\,, \tag{2.71}$$

the Hamiltonian becomes

$$H = \frac{\mathbf{P}_{\mathrm{com}}^2}{2M} \otimes \mathrm{Id}_{\mathrm{rel}} + \mathrm{Id}_{\mathrm{com}} \otimes \left[\frac{\mathbf{P}^2}{2\mu} - \frac{e^2}{|\mathbf{X}|}\right]\,, \tag{2.72}$$

where $M = m_e + m_p$ is the total mass and $\mu = m_e m_p / M$ is the *reduced mass*. Such calculations should be familiar from studying planetary orbits in IA Dynamics and Relativity.

Henceforth we'll often omit the $\otimes$ symbol — both in states and in operators — when the meaning is clear.

2.4 Postulates of Quantum Mechanics

So far, we've just been doing linear algebra, talking about Hilbert spaces in a fairly abstract way. Let's now begin to connect this to some physics.

The first postulate of quantum mechanics says that our system is described (up to a redundancy discussed below) by some state $|\psi\rangle \in \mathcal{H}$, and that any complete set of orthogonal states $\{|\phi_1\rangle, |\phi_2\rangle, \dots\}$ is in one-to-one correspondence with all the possible outcomes of the measurement of some quantity. Further, if a system is prepared in some general state

$$|\psi\rangle = \sum_m c_m |\phi_m\rangle\,, \tag{2.73}$$

then the probability that the measurement will yield an outcome corresponding to the state ϕ_n is

$$\mathrm{Prob}(|\psi\rangle \to |\phi_n\rangle) = \frac{|\langle\phi_n|\psi\rangle|^2}{\langle\psi|\psi\rangle\,\langle\phi_n|\phi_n\rangle}\,, \tag{2.74}$$

or equivalently $\mathrm{Prob}(|\psi\rangle \to |\phi_n\rangle) = |c_n|^2 / \langle\phi_n|\phi_n\rangle\,\langle\psi|\psi\rangle$. The denominator in the above expression allows the states $|\psi\rangle$ and $\{\phi_n\}$ to have arbitrary norm. Notice that $\mathrm{Prob}(|\psi\rangle \to |\phi_n\rangle) \geq 0$ and also, since $\langle\psi|\psi\rangle = \sum_{m,n} c_m^* c_n \langle\phi_m|\phi_n\rangle = \sum_m |c_m|^2 \langle\phi_m|\phi_m\rangle$ by the orthogonality of $\{|\phi_m\rangle\}$, that

$$\sum_n \mathrm{Prob}(|\psi\rangle \to |\phi_n\rangle) = \sum_b \frac{|c_n|^2 \langle\phi_n|\phi_n\rangle}{\langle\psi|\psi\rangle} = \frac{\langle\psi|\psi\rangle}{\langle\psi|\psi\rangle} = 1\,. \tag{2.75}$$

The probabilities sum to one.

Let us make some remarks. Firstly, note that we've not specified exactly which Hilbert space $\mathcal{H}$ we should use; is $\mathcal{H}$ to be $\mathbb{C}^n$, or ℓ^2, or $L^2(\mathbb{R}^3, \mathrm{d}^3\mathbf{x})$, or something else? In fact, the appropriate choice depends on the system we're trying to describe. For example, a single point particle with no internal structure could be described using[11] $\mathcal{H} \cong L^2(\mathbb{R}^3, \mathrm{d}^3\mathbf{x})$ whereas to specify the state of a more complicated system with several degrees of freedom we'll need a larger Hilbert space, encoding for example the location of its centre of mass, but also details of the system's orientation.

Second, the fact that quantum mechanics only predicts the probabilities of obtaining a given experimental outcome — even in principle — is one of its most puzzling features. Originally, this probabilistic interpretation of wavefunctions was due to Max Born, and (2.74) is sometimes known as the *Born rule*. He found that, solving the Schrödinger equation for scattering a particle of a generic obstacle in $\mathbb{R}^3$, the particle's wavefunction would typically become very spread out so as to have appreciable magnitude over a wide region. This is in contrast to our experience of particles bouncing off targets, each in a certain specific way. For example, recall that when performing his gold foil experiment to probe the structure of the atom, Rutherford usually observed α-particles to plough straight through, but occasionally saw them rebound back revealing the presence of the dense nucleus. Each α-particle was observed at a specific angle from the target, and did not itself 'spread out'. Reconciling this observation with the behaviour of the wavefunction is what led Born to suggest that quantum mechanics should be inherently probabilistic. His reasoning was perhaps not totally sound, and we'll explore the interpretative meaning of QM further and from a more modern perspective in section 10.

As a final remark on this postulate, to clean up the basic probability rule (2.74), it's usually convenient to assume our states are normalised, meaning that

$$\langle\psi|\psi\rangle = 1\,, \tag{2.76}$$

and to expand them in an orthonormal basis $\{|\phi_b\rangle\}$. In this case, the above probability is simply

$$\mathrm{Prob}(|\psi\rangle \to |\phi_b\rangle) = |\langle\phi_b|\psi\rangle|^2\,, \tag{2.77}$$

while the quantity $\langle\phi_b|\psi\rangle \in \mathbb{C}$ itself is often called the *probability amplitude*, or just *amplitude* for short. In what follows, we'll almost always use this simpler expression, but it's important to recall that it only holds in the case of a properly normalised state expanded in an orthonormal basis. Even when all states are properly normalised, we're still free to redefine $|\psi\rangle \to \mathrm{e}^{\mathrm{i}\alpha}|\psi\rangle$ for some constant phase α. This phase drops out of both the normalisation condition and the Born rule (2.74) for the probability. Taking into account both the normalisation condition and the phase freedom, in quantum mechanics, physical states do not correspond to elements $|\psi\rangle \in \mathcal{H}$ but rather *rays through the origin*. That is, provided $|\psi\rangle \neq \mathbf{0}$, the entire family of states

$$|\psi_\lambda\rangle = \lambda\,|\psi\rangle\ ,\ \lambda \in \mathbb{C} \setminus \{0\} \tag{2.78}$$

[11] Note that we only require $\mathcal{H}$ to be isomorphic to the space $L^2(\mathbb{R}^3, \mathrm{d}^3\mathbf{x})$ of normalisable position-space wavefunctions. We could equally well describe our structureless particle using a momentum space wavefunction, and indeed $L^2(\tilde{\mathbb{R}}^3, \mathrm{d}^3\mathbf{p}) \cong L^2(\mathbb{R}^3, \mathrm{d}^3\mathbf{x})$, with the isomorphism provided by the Fourier transform.

define the physical system — by tuning $|\lambda| = 1/\langle\psi|\psi\rangle$ we can always find a member of this family that's correctly normalised, and the constant phase λ cannot affect the probability of any outcome of a physical system. (This is the redundancy referred to above.) Note in particular that the zero vector $\mathbf{0}$ never represents a physical state. Geometrically, the equivalence $|\psi\rangle \sim \lambda|\psi\rangle$ for all $|\psi\rangle \in \mathcal{H} \setminus \{\mathbf{0}\}$ and all $\lambda \in \mathbb{C}\setminus\{0\}$ means that physical states actually correspond to elements of the *projective Hilbert space* $\mathbb{P}\mathcal{H}$. As with any projective space, it's often most convenient to work simply with normalised vectors $|\psi\rangle \in \mathcal{H}$ themselves, and recall that the overall phase is irrelevant.

The second postulate of quantum mechanics states that observable quantities are represented by Hermitian linear operators. In particular, upon measurement of a quantity corresponding to a Hermitian operator Q, a state is certain to return the definite value q if and only if it is an eigenstate of Q with eigenvalue q. Let $\{|n\rangle\}$ be a complete orthonormal set of eigenstates of some Hermitian operator Q, with $Q|n\rangle = q_n|n\rangle$. Then we can expand an arbitrary state $|\psi\rangle$ in this basis as $|\psi\rangle = \sum_n c_n|n\rangle$. The *expectation value* of Q in this state is

$$\langle Q\rangle_\psi = \langle\psi|Q|\psi\rangle = \sum_{m,n} c_m^* c_n \langle m|Q|n\rangle = \sum_n q_n |c_n|^2 \,, \tag{2.79}$$

using the orthonormality of the basis. This is just the sum of values of Q possessed by the states $|n\rangle$, weighted by the probability that $|\psi\rangle$ agrees with $|n\rangle$.

Since operators representing observables are Hermitian, for any such operator we have

$$\langle\psi|Q^2|\psi\rangle = \langle\psi|Q^\dagger Q|\psi\rangle = \|Q|\psi\rangle\|^2 \geq 0 \,, \tag{2.80}$$

and hence

$$0 \leq \left\langle\psi\middle|(Q - \langle Q\rangle_\psi)^2\middle|\psi\right\rangle = \langle Q^2\rangle_\psi - \langle Q\rangle_\psi^2 \,. \tag{2.81}$$

This shows that $\langle Q^2\rangle_\psi \geq \langle Q\rangle_\psi^2$, with equality iff $|\psi\rangle$ is an eigenstate of Q. We define the *rms deviation* $\Delta_\psi Q$ of Q from its mean $\langle Q\rangle_\psi$ in a state $|\psi\rangle$ by

$$\Delta_\psi Q = \sqrt{\left\langle\psi\middle|(Q - \langle Q\rangle_\psi)^2\middle|\psi\right\rangle} \,. \tag{2.82}$$

This is just the usual definition familiar from probability. As always, it gives us a measure of how 'spread' a state is around the eigenstate of Q. This implies that we can be sure of the value we'll obtain when measuring an observable quantity only if we somehow know that our state is in an eigenstate of the corresponding operator before carrying out the measurement.

Let me emphasize that these two postulates do not say anything about how the physical process of actually carrying out a measurement is described in the formalism of QM. (Nor do they even tell us what constitutes 'makes a measurement'.) In particular, we do not say that measuring the observable corresponding to some Hermitian operator Q has anything to do with the mathematical operation of acting on our state $|\psi\rangle$ with Q. According to the Copenhagen interpretation, if we measure the observable corresponding Q and find the result q, then immediately after this measurement, our system must be the state $|q\rangle$, because we've just learned that it does indeed have this value. The wavefunction is assumed to have *collapsed* from whatever it was before we measured it to $|q\rangle$. The Copenhagen interpretation is the most widely accepted version of quantum mechanics, but it's not uncontroversial. We'll try to understand measurement from a deeper perspective in section 10.4.2.

The final postulate of quantum mechanics is that the state $|\psi\rangle$ of our system evolves in time according to the Schrödinger equation

$$\mathrm{i}\hbar\frac{\partial}{\partial t}|\psi\rangle = H|\psi\rangle \,, \tag{2.83}$$

where H is some distinguished operator known as the Hamiltonian.

2.4.1 The Generalised Uncertainty Principle

Suppose we have two Hermitian operators A and B and $|\psi\rangle \in \mathcal{H}$. Let's define $|\psi_A\rangle = A|\psi\rangle - \langle A\rangle_\psi |\psi\rangle$ so that $(\Delta_\psi A)^2 = \||\psi_A\rangle\|^2$, and similarly $|\psi_B\rangle = B|\psi\rangle - \langle B\rangle_\psi |\psi\rangle$. Then the Cauchy–Schwarz inequality says

$$(\Delta_\psi A)^2(\Delta_\psi B)^2 = \||\psi_A\rangle\|^2 \, \||\psi_B\rangle\|^2 \geq |\langle\psi_A|\psi_B\rangle|^2 \ . \tag{2.84}$$

Expanding the right-hand side, we have

$$\langle\psi_A|\psi_B\rangle = \left\langle\psi\middle|(A - \langle A\rangle_\psi)(B - \langle B\rangle_\psi)\middle|\psi\right\rangle = \left\langle\psi\middle|AB - \langle A\rangle_\psi \langle B\rangle_\psi\middle|\psi\right\rangle \ . \tag{2.85}$$

Since we're considering Hermitian operators A, B, we have $\langle\psi|AB|\psi\rangle = \langle\psi|BA|\psi\rangle^*$, so

$$\operatorname{Im}\langle\psi_A|\psi_B\rangle = \frac{1}{2\mathrm{i}}\langle\psi|[A,B]|\psi\rangle \ . \tag{2.86}$$

Combining this with the Cauchy–Schwarz inequality gives the generalised uncertainty relation

$$(\Delta_\psi A)^2(\Delta_\psi B)^2 \geq |\langle\psi_A|\psi_B\rangle|^2 \geq \frac{1}{4}\left|\langle[A,B]\rangle_\psi\right|^2 \ . \tag{2.87}$$

In particular, if $[A,B] \neq 0$ we cannot in general find states that simultaneously have definite values for both the quantities represented by A or B. As a particular case, recall from IB Quantum Mechanics that the position and momentum operators obey the commutation relation $[X,P] = \mathrm{i}\hbar$. This is the *Heisenberg principle of uncertainty*

$$\Delta_\psi X \Delta_\psi P \geq \frac{\hbar}{2} \, , \tag{2.88}$$

which states that no quantum state can have a definite value for both position and momentum[12].

[12] The inequality is in fact saturated by states whose position space wavefunctions are Gaussians, so despite the fact that our derivation above neglected several positive semi-definite terms, we cannot place a higher bound on the minimum uncertainty.

3 The Harmonic Oscillator

We can now use Dirac's formalism to study a simple system — the one-dimensional harmonic oscillator — with which you should already be familiar. Our aim here is not to learn new things about harmonic oscillators; indeed, we'll mostly just recover results you've known about since you first heard of simple harmonic motion. Rather, our aim is to get accustomed to this more abstract approach to quantum mechanics, seeing how it's a very powerful way to think about the subject.

The steps we follow in our treatment of the harmonic oscillator will form a good prototype for the way we'll approach many other problems in quantum mechanics. We'll see these same steps repeated in various contexts throughout the course.

3.1 Raising and Lowering Operators

The Hamiltonian of a harmonic oscillator of mass m and classical frequency ω is

$$H = \frac{P^2}{2m} + \frac{1}{2}m\omega^2 X^2 \,, \tag{3.1}$$

where X and P are the position and momentum operators, respectively. To analyse this, we begin by defining the dimensionless combination

$$A \coloneqq \frac{1}{\sqrt{2m\hbar\omega}}(m\omega X + \mathrm{i}P) \,. \tag{3.2}$$

A is not a Hermitian operator, but since X and P are both Hermitian, we see that the adjoint of A is

$$A^\dagger = \frac{1}{\sqrt{2m\hbar\omega}}(m\omega X - \mathrm{i}P) \,. \tag{3.3}$$

Roughly, the motivation for introducing these operators is that they allow us to 'factorise' the Hamiltonian. We have

$$\begin{aligned} A^\dagger A &= \frac{1}{2m\hbar\omega}(m\omega X - \mathrm{i}P)(m\omega X + \mathrm{i}P) \\ &= \frac{1}{2m\hbar\omega}(P^2 + m^2\omega^2 X^2 + im\omega[X,P]) \\ &= \frac{H}{\hbar\omega} - \frac{1}{2} \,, \end{aligned} \tag{3.4}$$

so we can write our Hamiltonian as

$$H = \hbar\omega\left(A^\dagger A + \frac{1}{2}\right) = \hbar\omega\left(N + \frac{1}{2}\right) \,, \tag{3.5}$$

where

$$N \coloneqq A^\dagger A \,. \tag{3.6}$$

A and $A^\dagger$ are often called *lowering* and *raising* operators, respectively, whilst N is called the *number* operator. The reason for these names will become apparent soon. Notice that computing the spectrum of H is obviously equivalent to computing the spectrum of N.

Whenever we're presented with some new operators, the first thing to do is to work out their commutation relations. In this case, the fundamental commutation $[X,P] = \mathrm{i}\hbar$ shows that

$$\begin{aligned} [A, A^\dagger] &= \frac{1}{2m\hbar\omega}(m^2\omega^2[X,X] - im\omega[X,P] + im\omega[P,X] + [P,P]) \\ &= -\frac{im\omega}{2m\hbar\omega}([X,P] - [P,X]) \\ &= 1 \,, \end{aligned} \tag{3.7}$$

whilst $[A, A] = [A^\dagger, A^\dagger] = 0$ trivially. It will also be useful to compute commutators involving N. We have

$$[N, A^\dagger] = [A^\dagger A, A^\dagger] = \big(A^\dagger[A, A^\dagger] + [A^\dagger, A^\dagger]A\big) = A^\dagger \,. \tag{3.8}$$

Similarly, $[N, A] = -A$ by taking the adjoint.

Let's see what these commutators teach us about the possible energy levels. Suppose that $|n\rangle$ is a correctly normalised eigenstate of N, so that $N|n\rangle = n|n\rangle$ and $\langle n|n\rangle = 1$. We have

$$NA^\dagger|n\rangle = (A^\dagger N + [N, A^\dagger])|n\rangle = A^\dagger(N+1)|n\rangle = (n+1)A^\dagger|n\rangle \,, \tag{3.9}$$

so $A^\dagger|n\rangle$ is an eigenstate of N with eigenvalue $n+1$. Similarly,

$$NA|n\rangle = (AN + [N, A])|n\rangle = (n-1)A|n\rangle \,, \tag{3.10}$$

so $A|n\rangle$ is an eigenstate of N with eigenvalue $n-1$. Acting with $A^\dagger$ thus raises the eigenvalue of N by one unit, while acting with A lowers it by one, giving the operators their names. The above algebra shows that if we can find just one energy eigenstate $|n\rangle$, then we can construct a whole family of them by repeatedly applying either A or $A^\dagger$. The energies of these new states will be $(n+k+\frac{1}{2})\hbar\omega$ for some $k \in \mathbb{Z}$. However, we can't yet conclude that the energy levels are quantised, because we don't yet know that there isn't a continuum of possible starting-points $|n\rangle$ (or indeed any).

This brings us to the second key step: we must investigate the norm of our states. Since N is Hermitian all its eigenvalues must certainly be real, but in fact they're also non-negative, because

$$n = n\langle n|n\rangle = \langle n|N|n\rangle = \big\langle n\big|A^\dagger A\big|n\big\rangle = \|A|n\rangle\|^2 \geq 0 \,, \tag{3.11}$$

with $n = 0$ iff $A|n\rangle = \mathbf{0}$, by the properties of the norm. If $A|n\rangle \neq 0$ then it is an eigenstate of N with eigenvalue $n-1$, but we've just shown that there are no states in $\mathcal{H}$ with negative eigenvalues for N, so the lowering process must terminate, That will be the case iff n is a non-negative integer.

Putting all these together, we have a ground state $|0\rangle$, we have a ground state $|0\rangle$ of energy $\frac{1}{2}\hbar\omega$ and an infinite tower of excited states $|n\rangle$ of energy

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega \,,\ n \in \mathbb{N}_0 \,. \tag{3.12}$$

These energy levels should be familiar from IB Quantum Mechanics.

Acting on an energy eigenstate with the raising operator does not necessarily give us an excited state that is correctly normalised. In $d = 1$, energy eigenstates of the harmonic oscillator are non-degenerate[13], so we must have $A^\dagger|n\rangle = c_n|n+1\rangle$ for some constant c_n (which may depend on n). Taking the norm of both sides shows that

$$|c_n|^2 = \big\|A^\dagger|n\rangle\big\|^2 = \big\langle n\big|AA^\dagger\big|n\big\rangle = \langle n|N+1|n\rangle = n+1 \,. \tag{3.13}$$

Therefore, the correctly normalised state is

$$|n+1\rangle = \frac{1}{\sqrt{n+1}}A^\dagger|n\rangle = \frac{1}{\sqrt{(n+1)!}}(A^\dagger)^{n+1}|0\rangle \,. \tag{3.14}$$

Likewise, you can check that $|n-1\rangle = \frac{1}{n}A|n\rangle$ for $n \geq 1$, whilst again $A|0\rangle = 0$.

We've seen that everything about the energy levels of the harmonic oscillator follows from (i) the algebra (i.e. the commutation relations) of raising and lowering operators, together with (ii) considering the norm. Again, we'll follow these same two steps in analysing the eigenvalues and eigenstates of various operators throughout this course. This also parallels what you did in IB Quantum Mechanics: there, by looking for series solutions of Schrödinger's equation, you could find

[13] A proof is given in IB Quantum Mechanics.

an energy eigenstate for all $E \in \mathbb{R}$. However, these eigenstates were only normalizable if the series you found terminated. It was requiring this termination (i.e. normalisability) that lead to quantization of the energies.

I hope I've persuaded you that the algebraic approach above is somewhat cleaner and more efficient than looking for series solutions to Schrödinger's equation. Nothing has been lost, and we can easily use Dirac's formalism to recover the explicit wavefunctions of our eigenstates. For example, in the position representation, the defining equation $A\,|0\rangle = \mathbf{0}$ of the ground state become

$$0 = \sqrt{2m\hbar\omega}\,\langle x|A|0\rangle = \langle x|m\omega X + \mathrm{i}P|0\rangle = \left(m\omega x + \hbar\frac{\mathrm{d}}{\mathrm{d}x}\right)\psi_0(x)\,, \tag{3.15}$$

where $\psi_0(x) = \langle x|0\rangle$ is the position space wavefunction of our ground state. This is a first order ODE, whose solution is

$$\psi_0(x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4}\exp(-\frac{m\omega}{2\hbar}x^2)\,, \tag{3.16}$$

where the overall constant is fixed by normalisation $\langle 0|0\rangle = \int \mathrm{d}x\,|\psi_0(x)|^2 = 1$ up to phase. This is the same Gaussian wavefunction for the ground state of the harmonic oscillator familiar form IB Quantum Mechanics. If needed, the wavefunctions of the excited states may be found by acting with $A^\dagger = (m\omega X - \mathrm{i}P)/\sqrt{2m\hbar\omega}$, which in the position representation is the differential operator $\frac{1}{\sqrt{2}}(-\alpha\frac{\mathrm{d}}{\mathrm{d}x} + x/\alpha)$, where $\alpha = \sqrt{\hbar/m\omega}$. You can check that these operators generate the usual Hermite polynomials.

3.2 Dynamics of Oscillators

At this stage, we've learned everything about the set $\{|n\rangle\}$ of energy eigenstates of the quantum harmonic oscillator and their corresponding energies $E_n = (n + \frac{1}{2})\hbar\omega$. However, we still haven't said anything about the physics of how quantum oscillators actually behave. Classically, we know that a harmonic oscillator would undergo periodic motion with a period $T = 2\pi/\omega$. Furthermore, the energy of the classical oscillator is independent of the period, but is proportional to the square of the amplitude of oscillation. To what extent is the same true of our quantum oscillator?

To say anything about the motion of a quantum system, we need to examine the time-dependent Schrödinger equation. To get started, first suppose our system is prepared at time $t = 0$ to be in some energy eigenstate $|n\rangle$. Then by the time-dependent Schrödinger equation, at a later time t it will have evolved to

$$|n, t\rangle = \mathrm{e}^{-\mathrm{i}(n+1/2)\omega t}\,|n\rangle\ . \tag{3.17}$$

Consequently, no eigenstate has a time dependence which oscillates at the classical frequency ω. Even worse, no matter which energy eigenstate our oscillator is in, the expected position of the oscillator at any given time t is

$$\langle x, t|X|n, t\rangle = \mathrm{e}^{+\mathrm{i}(n+1/2)\omega t}\mathrm{e}^{-\mathrm{i}(n+1/2)\omega t}\,\langle n|X|n\rangle = \langle n|X|n\rangle\ , \tag{3.18}$$

where we've used the linearity/antilinearity properties of the inner product. The right-hand side is independent of time, so none of these states move — our oscillator does not appear to be oscillating.

To find some interesting dynamics, we must consider not a single energy eigenstate, but rather a *superposition*. This is a much more realistic assumption: there's no practical way we could prepare a macroscopic system to be in just one energy eigenstate. Let's now suppose our oscillator is prepared at $t = 0$ to be in some generic state $|\psi, 0\rangle = \sum_n c_n\,|n\rangle$. The c_n should be chosen so that $\langle\psi|\psi\rangle = 1$, but are otherwise arbitrary. Then at time t this state will have evolved to

$$|\psi, t\rangle = \sum_{n=0}^{\infty} c_n \mathrm{e}^{-\mathrm{i}E_n t/\hbar}\,|n\rangle \tag{3.19}$$

by the time-dependent Schrödinger equation.

Now let's examine where we expect to find such a generic state. We have

$$\begin{aligned}\langle\psi,t|X|\psi,t\rangle &= \left(\sum_m c_m^* \mathrm{e}^{\mathrm{i}E_m t/\hbar}\langle m|\right) X \left(\sum_n c_n \mathrm{e}^{-\mathrm{i}E_n t/\hbar}|n\rangle\right)\\ &= \sum_{m,n} c_m^* c_n \mathrm{e}^{\mathrm{i}(m-n)\omega t}\langle m|X|n\rangle\ . \end{aligned} \tag{3.20}$$

To evaluate the inner product $\langle m|X|n\rangle$, we write[14]

$$X = \sqrt{\frac{\hbar}{2m\omega}}\left(A + A^\dagger\right) \tag{3.22}$$

in terms of the raising and lowering operators. Recalling that $A^\dagger|n\rangle = \sqrt{n+1}\,|n+1\rangle$ and $A|n\rangle = \sqrt{n}\,|n-1\rangle$, we have

$$\langle m|X|n\rangle = \sqrt{\frac{\hbar}{2m\omega}}(\sqrt{n}\,\langle m|n-1\rangle + \sqrt{n+1}\,\langle m|n+1\rangle)\,, \tag{3.23}$$

showing that $\langle m|X|n\rangle$ is non-zero only when $m = n \pm 1$. The double sum in (3.20) thus reduces to

$$\begin{aligned}\langle\psi,t|X|\psi,t\rangle &= \sqrt{\frac{\hbar}{2m\omega}}\left(\sum_{n=1}^{\infty}\sqrt{n}c_{n-1}^* c_n \mathrm{e}^{-\mathrm{i}\omega t} + \sqrt{n}c_n^* c_{n-1}\mathrm{e}^{+\mathrm{i}\omega t}\right)\\ &= \sum_n x_n \cos(\omega t + \phi_n)\,, \end{aligned} \tag{3.24}$$

where the real numbers x_n and ϕ_n are defined by

$$\sqrt{\frac{2n\hbar}{m\omega}}c_n^* c_{n-1} = x_n \mathrm{e}^{\mathrm{i}\phi_n}\ . \tag{3.25}$$

Equation (3.24) shows that $\langle X\rangle$ oscillates sinusoidally at exactly the classical frequency ω whenever our oscillator is prepared in any generic[15] superposition of energy eigenstates. Furthermore, as for the classical oscillator, the frequency of oscillation is independent of the energy. In the calculation above, this occurs because the separation between every adjacent pair of energy levels is always $\hbar\omega$.

For a macroscopic oscillator, the only non-negligible amplitudes will be those where $n \approx n_{\mathrm{cl}}$ for some $n_{\mathrm{cl}} \gg 1$. Consequently, a measurement of the energy is certain to yield some value close to $E_{n_{\mathrm{cl}}} = (n_{\mathrm{cl}} + \frac{1}{2})\hbar\omega \approx n_{\mathrm{cl}}\hbar\omega$. For each eigenstate, we have

$$\left\langle n\middle|X^2\middle|n\right\rangle = \frac{\hbar}{2m\omega}\left\langle n\middle|AA^\dagger + A^\dagger A\middle|n\right\rangle = \frac{E_n}{m\omega^2}\,, \tag{3.26}$$

so the mean value $\left\langle X^2\right\rangle = E_{n_{\mathrm{cl}}}/m\omega^2$. Classically, the time average of x^2 is proportional to the average potential energy, which for a harmonic oscillator is just a half of the total energy. Thus, classically we have $\left\langle x^2\right\rangle = E/m\omega^2$ in agreement with the quantum result. The correspondence principle requires that the quantum and classical results agree for large value of E. They agree even for low energies is a coincidence due to special symmetries of the harmonic oscillator.

[14] To do this calculation using techniques of IB Quantum Mechanics, you would have to work in the position representation and said

$$\langle m|X|n\rangle = \int_{-\infty}^{+\infty}\mathrm{d}x\,\left(H_m(x)\mathrm{e}^{-x^2/2\alpha}\right)^* x H_n(x)\mathrm{e}^{-x^2/2\alpha}\,, \tag{3.21}$$

where H_n are the Hermite polynomials of degree n and $\alpha = \hbar/m\omega$. This is correct but the integral looks rather unpleasant to evaluate. Fortunately, our operator formalism means we never even have to try!

[15] Here, "generic" means we must have non-zero amplitudes $c_n = \langle n|\psi\rangle$ for at least one pair of adjacent energy levels.

3.3 Anharmonic Oscillations

Suppose that, instead of the pure harmonic oscillator, we have a potential that has a minimum at $x = 0$ around which it grows approximately quadratically while $|x| \ll a$, but then asymptotes to a constant value at large $|x|$. Particles of low energy will have position space wavefunctions that are supported near the minimum of $V(x)$, so we'd expect the low-lying energy levels to be roughly equal to those of a corresponding harmonic oscillator. Particles of higher energy would start to see the fact that the potential is not purely harmonic. In fact, for any such asymptotically constant potential with a single extremum, the separation between energy levels gets smaller and smaller as we approach the asymptotic value $\lim_{x\to\infty} V(x)$ of the potential (beyond which we have a continuum of non-bound states).

Let's prepare a state to be in two adjacent energy levels, say

$$|\psi\rangle = c_n\,|n\rangle + c_{n+1}\,|n+1\rangle\ , \tag{3.27}$$

where $|n\rangle$ is the nth energy eigenstate of our anharmonic potential, whatever it may be. Since the potential is symmetric about $x = 0$, its eigenstates have definite parity so $\langle n|X|n\rangle = 0$ for all $|n\rangle$. Therefore, at time t,

$$\begin{aligned}\langle\psi,t|X|\psi,t\rangle &= c_n^* c_{n+1} \mathrm{e}^{\mathrm{i}(E_n - E_{n+1})t/\hbar}\,\langle n|X|n+1\rangle + c_{n+1}^* c_n \mathrm{e}^{\mathrm{i}(E_{n+1}-E_n)t/\hbar}\,\langle n+1|X|n\rangle \\ &= 4\,\mathrm{Re}[c_n^* c_{n+1}\,\langle n|X|n+1\rangle]\cos\left(\frac{E_{n+1}-E_n}{\hbar}t\right)\ .\end{aligned} \tag{3.28}$$

This is again a sinusoidally oscillating function of time, but now the period $T = 2\pi\hbar/(E_{n+1} - E_n)$ depends on n since we no longer expect the energy levels to be equally spaced. In fact, since the energy levels get closer together as n increases, if we give our particle a larger amplitude of oscillation — and hence more energy — it will take longer to execute a complete an oscillation. This is just what we'd expect classically.

For example, consider the *Pöschl–Teller potential*

$$V(x) = -V_0\,\mathrm{sech}^2(\kappa X) \tag{3.29}$$

for some constant $V_0 > 0$ and length scale $a = 1/\kappa$. This has bound state energy levels

$$E_n = -\frac{\hbar^2\kappa^2}{2m}(\nu - n)^2\ ,\ n \in \mathbb{Z}\ ,\ 0 \le n < \nu\ , \tag{3.30}$$

where ν is the positive root of $\nu(\nu+1) = 2mV_0/\hbar^2\kappa^2$. The separation between adjacent bound state energy levels is thus

$$E_{n+1} - E_n = (2(\nu - n) - 1)\hbar^2\kappa^2/2m\,, \tag{3.31}$$

which decreases as n increases towards ν. These formulae break down when $n \ge \nu$, where we enter the continuum of non-normalisable states.

If $V_0 \gg \hbar^2\kappa^2/2m$ then the potential is very deep and contains many bound states. in this regime, we have $\nu \approx \sqrt{2mV_0/\hbar^2\kappa^2} \gg 1$ and so from (3.31) we find that a superposition of low-lying states will oscillate with a frequency

$$\omega \approx \frac{\hbar\kappa^2}{m}\nu \approx \kappa\sqrt{\frac{2V_0}{m}}\,. \tag{3.32}$$

For $x \ll 1/\kappa$, we may approximate $-V_0\,\mathrm{sech}^2(\kappa x) = -V_0 + V_0\kappa^2 x^2$, so this frequency is just what we'd expect for the corresponding harmonic oscillator. On the other hand, a superposition of states clustered around $n = \nu/2$ will oscillate at around half this frequency as neighbouring energy levels in this region are only separated by about half as much.

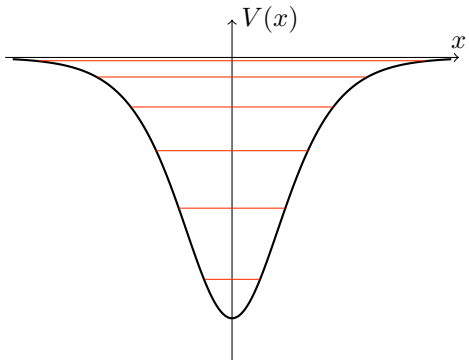


Figure 3: The Pöschl–Teller potential $V(x) = -V_0 \operatorname{sech}^2(\kappa x)$ and its energy levels.

If we include a wider range of states in our initial superposition, then we'll instead find a sequence of terms

$$\begin{aligned}\langle \psi, t|X|\psi, t\rangle = \\ &\ldots + c_{n-1}^* c_n \mathrm{e}^{\mathrm{i}(E_{n-1}-E_n)t/\hbar} \langle n-1|X|n\rangle + c_{n+1}^* c_n \mathrm{e}^{\mathrm{i}(E_{n+1}-E_n)t/\hbar} \langle n+1|X|n\rangle \\ &\quad + c_{n+3}^* c_n \mathrm{e}^{\mathrm{i}(E_{n+3}-E_n)t/\hbar} \langle n+3|X|n\rangle + \ldots \\ &+ \ldots + c_n^* c_{n+1} \mathrm{e}^{\mathrm{i}(E_n - E_{n+1}t/\hbar)} \langle n|X|n+1\rangle + \ldots . \end{aligned} \tag{3.33}$$

The potential is symmetric, so we again have $\langle n+2k|X|n\rangle = 0$ by parity, but this is no longer a harmonic potential, so we do not generally expect $\langle x+2k+1|X|n\rangle = 0$. Let's define a frequency $\Omega = (E_{n_{\mathrm{cl}}+1} - E_{n_{\mathrm{cl}}})$, where $n_{\mathrm{cl}} \gg 1$. Provided the c_n's are clustered around this large central value $n = n_{\mathrm{cl}}$ sufficiently tightly that the difference between adjacent energy levels is roughly constant over the range of n for which the c_n are appreciable, then, to reasonable accuracy, all the terms that contribute to (3.33) oscillate with frequencies that are integer multiples of $\Omega_{n_{\mathrm{cl}}}$. Thus the motion will be periodic, but anharmonic, just as we expect classically.

If we release the anharmonic oscillator from some large extension x_0, then initially the wavefunction will be a superposition of many energy levels, with coefficients that ensure $\langle x|\psi\rangle = \sum_n c_n \langle x|n\rangle$ is sharply peaked around x_0. At time t, this state will evolve to

$$|\psi, t\rangle = \mathrm{e}^{-\mathrm{i}E_{n_{\mathrm{cl}}}t/\hbar} \sum_n c_n \mathrm{e}^{\mathrm{i}(E_{b_{\mathrm{cl}}}-E_n)t/\hbar} |n\rangle \ . \tag{3.34}$$

Since the separation between energy levels varies, the frequencies appearing in this sum are not all integer multiples of any Ω_N. Consequently, after a time of order $2\pi/\Omega_N$, most terms in the sum will have not quite returned to their original values, so the wavefunction at $t = 2\pi/\Omega$ will be less sharply peaked around X. With each subsequent passing of time $2\pi/\Omega_N$, the wavefunction will become more and more diffusely spread. Classically, if we release an oscillator with a rather uncertain value for its energy, then for a pure harmonic oscillator we always know where to find the oscillator at later times, since it's period is independent of the energy. However, for an anharmonic oscillator, the period depends on the energy, so after a long time our oscillator is equally likely to be located anywhere.

4 Transformations and Symmetries

While classical physics is firmly rooted in space-time, quantum mechanics takes place in the more abstract Hilbert space. Actions such as translations or rotations around a given origin have a straightforward space-time interpretation, but how can these affect a quantum particle living in Hilbert space?

In this chapter, we'll return to the general formalism of quantum mechanics. We'll understand how transformations such as spatial translations or rotations affect states in Hilbert space, linking together the world of quantum mechanics with our familiar experience in "normal" space $\mathbb{R}^3$. In so doing, we'll obtain a deeper appreciation of the origin of many of the most common commutation relations, including most of the ones you're familiar with from IB. We'll also understand why the momentum operator is $P = -\mathrm{i}\hbar\partial/\partial x$, a fact which probably seemed rather mysterious in IB Quantum Mechanics.

4.1 Transformations of States and Operations

If we apply a transformation such as a translation or rotation to our physical system, we'd expect states to be different before and after the transformation. For example, if a particle is prepared in a state $|\psi\rangle$ that is strongly peaked around the origin $\mathbf{0} \in \mathbb{R}^3$ and we translate our system through a vector $\mathbf{a}$, after this transformation we should expect to find our particle in a new state $|\psi'\rangle$ whose wavefunction is strongly peaked around $\mathbf{x} = \mathbf{a}$. This suggests that the effect of any given spatial transformation should be represented on Hilbert space by a linear operator

$$\begin{aligned} U : \mathcal{H} &\longrightarrow \mathcal{H} \\ |\psi\rangle &\longmapsto U|\psi\rangle = |\psi'\rangle \ , \end{aligned} \tag{4.1}$$

where the operator depends on the specific type of transformation.

Whatever state we start with, after applying the transformation, we still expect to find it somewhere in space. Thus, properly normalised states should remain so after applying U, or in other words

$$1 = \langle\psi|\psi\rangle = \langle\psi'|\psi'\rangle = \langle\psi|U^\dagger U|\psi\rangle \tag{4.2}$$

for all $|\psi\rangle \in \mathcal{H}$. We'll now show that the requirement that Y preserves the norm of our state fixes it to be a *unitary* operator[16]. The required trick is sometimes called the *polarisation identity.* We first write $|\psi\rangle$ as $|\psi\rangle = |\phi\rangle + \lambda|\chi\rangle$ for some $|\phi\rangle, |\chi\rangle \in \mathcal{H}$ and some $\lambda \in \mathbb{C}$. Equation (4.2) becomes

$$\begin{aligned} &\langle\phi|\phi\rangle + \lambda^*\langle\chi|\phi\rangle + \lambda\langle\phi|\chi\rangle + |\lambda|^2\langle\chi|\chi\rangle \\ &\qquad = \langle\phi|U^\dagger U|\phi\rangle + \lambda^*\langle\chi|U^\dagger U|\phi\rangle + \lambda\langle\phi|U^\dagger U|\chi\rangle + |\lambda|^2\langle\chi|U^\dagger U|\chi\rangle \ . \end{aligned} \tag{4.3}$$

By assumption, $\langle\phi|U^\dagger U|\phi\rangle = \langle\phi|\phi\rangle$ and $\langle\chi|U^\dagger U|\chi\rangle = \langle\chi|\chi\rangle$, so this simplifies to

$$\lambda\left(\langle\phi|\chi\rangle - \langle\phi|U^\dagger U|\chi\rangle\right) = \lambda^*\left(\langle\chi|U^\dagger U|\phi\rangle - \langle\chi|\phi\rangle\right) \ . \tag{4.4}$$

In order for this to hold for every $|\psi\rangle \in \mathcal{H}$, it must continue to hold as we vary λ. In particular, as the phase of λ varies, the only way for (4.4) to be satisfied is if

$$\langle\phi|\chi\rangle = \langle\phi|U^\dagger U|\chi\rangle \ . \tag{4.5}$$

[16] Because physical states are represented by rays in $\mathcal{H}$, rather than actual vectors, there is one other possibility, transformations may be represented by operators that are both *anti-linear*, in the sense that $U(c_1|\psi_1\rangle + c_2|\psi_2\rangle) = c_1^* U|\psi_1\rangle + c_2^* U|\psi_2\rangle$, and *anti-unitary*, meaning that $\langle\chi|U^\dagger U|\phi\rangle = \langle\phi|\chi\rangle = \langle\chi|\phi\rangle^*$ for all $|\phi\rangle, \phi\chi \in \mathcal{H}$. Such anti-unitary operators turn out to be related to reversing the direction of time; we will not discuss them further in this course. Wigner showed that all transformations are represented by either linear, unitary operators or anti-linear, anti-unitary ones

Finally, for this hold for arbitrary states $|\phi\rangle$ and $|\psi\rangle$ we must have $U^\dagger U = \mathrm{Id}$. Multiplying through on the right by U^{-1} shows that

$$U^\dagger = U^{-1}\,, \tag{4.6}$$

which says that the adjoint of U is equal to its inverse. Operators with these properties are said to be *unitary*[17].

There's one further condition on our operator U. Transformations of space such as translations and rotations form a *group*. For example, the composition of two rotations is again a rotation, the trivial rotation is the identity, and the inverse of a rotation is a rotation of the same amount around the same axis, but in the opposite direction. Let's suppose that our spatial transformations form a group G. To reflect this group structure, our operators should[18] provide a *homomorphism* from group G to the group of unitary operators in the sense that, for all $g_1, g_2 \in G$,

$$U(g_2) \circ U(g_1) = U(g_2 \cdot g_1) \quad \text{and} \quad U(1_G) = 1_{\mathcal{H}}\,, \tag{4.7}$$

where $\cdot$ denotes the group multiplication in G and $\circ$ denotes the composition of linear operators acting on $\mathcal{H}$. We'll typically suppress these composition symbols from now on. Note that if U_1 and U_2 are such unitary operators, then $(U_2U_1)^{-1} = U_1^{-1}U_2^{-1} = U_1^\dagger U_2^\dagger = (U_2U_1)^\dagger$ and so the composite operator U_2U_1 is also unitary. Note also that the identity operator $1_{\mathcal{H}}$ is trivially unitary.

A particularly important class of transformations are those that depend smoothly on some parameter θ. For example, we can smoothly vary both the axis about which and the angle through which we rotate, or the magnitude and direction of a translation vector. If $\theta = 0$ is trivial transformation, represented on $\mathcal{H}$ by the identity operator, then for infinitesimal transformation of $\delta\theta$, we have

$$U(\delta\theta) = 1 - \mathrm{i}\delta\theta T + O(\delta\theta^2)\,, \tag{4.8}$$

where T is some operator that is independent of θ. The factor $-\mathrm{i}$ in this equation is just a convention for later convenience. T is called the *generator* of the transformation. For such infinitesimal transformations, the condition of unitarity of U (4.6) becomes

$$1 + \mathrm{i}\delta\theta T^\dagger + O(\delta\theta^2) = 1 + \mathrm{i}\delta\theta T + O(\delta\theta^2)\,, \tag{4.9}$$

which, to first order in $\delta\theta$, gives

$$T = T^\dagger\,. \tag{4.10}$$

Thus the generator T is Hermitian and hence a good candidate for an observable quantity. A finite transformation can be generated by repeatedly performing an infinitesimal one. Specifically, if we set $\delta\theta = \theta/N$ and transform N times with $U(\delta\theta)$, then in the limit $N \to \infty$ we have

$$U(\theta) = \lim_{N\to\infty}\left(1 - \mathrm{i}\frac{\theta}{N}T\right)^N = \mathrm{e}^{-\mathrm{i}\theta T}\,, \tag{4.11}$$

where the exponential of an operator may be defined by (2.43), or equivalently by its power series expansion. This form of the unitary operator $U(\theta)$ is especially useful when states are expressed in terms of the basis of eigenstates of the Hermitian generator T.

We obtain an important equation by using (4.8) to evaluate $|\psi'\rangle = U(\delta\theta)\,|\psi\rangle$. Subtracting $|\psi\rangle$ from both sides and dividing through by $\delta\theta$, in the limit $\delta\theta \to 0$ we obtain

$$\mathrm{i}\frac{\partial\,|\psi\rangle}{\partial\theta} = T\,|\psi\rangle\ . \tag{4.12}$$

There is no assumption here that the wavefunction $\psi(\mathbf{x}) = \langle\mathbf{x}|\psi\rangle$ should be differentiable as a function of space. We merely say that as our transformation varies smoothly away from the identity, so too does the state $|\psi\rangle$ vary smoothly inside $\mathcal{H}$. The rate at which $|\psi\rangle$ varies is governed by the generator.

[17] Note that unitary operators are certainly bounded, and in fact have unit norm in the operator topology.

[18] This statement is not quite accurate — there's an important refinement that we'll return to later in the course.

Having described how transformations act on states in $\mathcal{H}$, we should now understand how they act on operators. Suppose A is some operator whose matrix elements we are interested in. If our transformation maps $|\psi\rangle \to |\psi'\rangle$, then the expectation value of A will be mapped as

$$\langle\psi|A|\psi\rangle \longmapsto \langle\psi'|A|\psi'\rangle = \big\langle\psi\big|U^\dagger(\theta)AU(\theta)\big|\psi\big\rangle \ . \tag{4.13}$$

Consequently, we can find the expectation value A will take after the transformation by working with the original states, but instead we can also transform the operator as

$$A \longmapsto A' = U^\dagger(\theta)AU(\theta) = U^{-1}(\theta)AU(\theta)\,, \tag{4.14}$$

using the fact that U is unitary. This is known as a *similarity transform*. Note that

$$[A', B'] = [U^{-1}AU, U^{-1}BU] = U^{-1}[A, B]U\,, \tag{4.15}$$

so the commutator of two similarity transforms is the similarity transform of the commutator. Also, similarity transforms do not change the spectrum of any operator: if $|\psi\rangle$ is an eigenstate of A with eigenvalue a, then $U^{-1}|\psi\rangle$ is an eigenstate of A' with the same eigenvalue.

Finally, for an infinitesimal transformation, we have

$$U^\dagger(\delta\theta)AU(\delta\theta) = A + \mathrm{i}\delta\theta[T, A] + O(\delta\theta^2)\,, \tag{4.16}$$

so $\delta A = \mathrm{i}\delta\theta[T, A]$. Thus, while the rate of change of states themselves is given by the action of the generator, the rate of change of operators under a transformation is determined by the commutator of the generator with the operator. We'll see that most of the commutation relations we meet in quantum mechanics (and certainly all the familiar ones) arise this way. Furthermore, because Hermitian operators represent observable quantities with which we are familiar, in practice it's often easier to understand how a transformation should act on an operator, rather than on the more abstract notion of a state in Hilbert space.

The above discussion has been rather abstract. Let's ground it by looking at a few of the most important examples of transformations and their corresponding generators.

4.2 Translations

When we move an object around, we expect to find it in a new place. Specifically, suppose $\langle\psi|\mathbf{X}|\psi\rangle = \mathbf{x}_0$ for some normalised state $|\psi\rangle$. Since $\mathbf{x}_0$ just labels a spatial point, it must behave under translations and rotations like any vector. Thus, after the translation we expect our object to be described by a new state $|\psi'\rangle$ with $\langle\psi'|\mathbf{X}|\psi'\rangle = \mathbf{x}_0 + \mathbf{a}$. The general theory of the previous section asserts that this translation is represented on Hilbert space by a unitary operator[19] $U(\mathbf{a})$ so that $|\psi'\rangle = U(\mathbf{a})\,|\psi\rangle$. Therefore we can write the expected location of the translated state as

$$\langle\psi'|\mathbf{X}|\psi'\rangle = \langle\psi|\mathbf{X}|\psi\rangle + \mathbf{a} = \langle\psi|(\mathbf{X} + 1_{\mathcal{H}}\mathbf{a})|\psi\rangle \tag{4.17}$$

where $1_{\mathcal{H}}$ represents the identity operator on $\mathcal{H}$. We'll often drop the identity operator where it's unambiguous. Since this is true for any state $|\psi\rangle$, the transformed operator may be expressed as

$$U^{-1}(\mathbf{a})\mathbf{X}U(\mathbf{a}) = \mathbf{X} + \mathbf{a}1_{\mathcal{H}} \tag{4.18}$$

by (4.14). Let's be clear what this equation means. In this course, we're interested in quantum mechanics on $\mathbb{R}^3$, so the position operator $\mathbf{X} : \mathcal{H} \to \mathcal{H}$ is really a collection of three operators corresponding to the three coordinates of $\mathbb{R}^3$. Taking the standard Cartesian basis of $\mathbb{R}^3$, in more detail (4.18) says

$$\begin{pmatrix} U^{-1}(\mathbf{a})XU(\mathbf{a}) \\ U^{-1}(\mathbf{a})YU(\mathbf{a}) \\ U^{-1}(\mathbf{a})ZU(\mathbf{a}) \end{pmatrix} = \begin{pmatrix} X + a_x 1_{\mathcal{H}} \\ Y + a_y 1_{\mathcal{H}} \\ Z + a_z 1_{\mathcal{H}} \end{pmatrix} \ . \tag{4.19}$$

[19] Strictly, we should label this as $U_{\text{trans}}(\mathbf{a})$ to emphasise that it is the unitary operator generating translations, as opposed to anything else we might associate to $\mathbf{a}$. We typically drop such additional labels. Exactly which unitary operator we're talking about should hopefully be clear from the context.

In other words, conjugating the X (component) operator by $U(\mathbf{a})$ gives a new operator $U^{-1}(\mathbf{a})XU(\mathbf{a}) : \mathcal{H} \to \mathcal{H}$ that we identify with $X+a_x$, and similarly for Y and Z. If $U(\mathbf{a})$ acted on the position operator in any other way, it could not represent a translation, so equation (4.18) may be taken as the defining property of the translation operator, distinguishing it from other unitary transformations.

For an infinitesimal translation $\delta\mathbf{a}$, equation (4.8) states that the translation operator can be written as[20]

$$U(\delta\mathbf{a}) = 1 - \frac{\mathrm{i}}{\hbar}\delta\mathbf{a}\cdot\mathbf{P} + O(\delta\mathbf{a}^2)\,, \tag{4.20}$$

where $\mathbf{P}$ is Hermitian. I've named this operator $\mathbf{P}$, so I'm not going to be able to fool any of you that it will turn out to be anything other than the momentum operator. But that's not how I want you to think of it yet. By definition, $\mathbf{P}/\hbar$ is the generator of infinitesimal translations.

The translation generator $\mathbf{P}/\hbar$ must have units 1/(length) so that it makes sense to add the second term of (4.20) to the first. Because we've included a factor of $\hbar$, our $\mathbf{P}$ must have dimensions of (mass $\times$ velocity). In fact, this is just a convention. We could choose to measure masses in units of 1/(length $\times$ velocity), using $\hbar$ as a conversion factor for our units. In relativity, the speed of light c provides a further natural conversion factor between lengths and times, so in units of $\hbar/c$, masses are equivalent to inverse lengths. In natural units, we work with length and time scales adapted to Nature rather than our civilisation, so we set $\hbar = 1$ and $c = 1$. In these units, the translation generator is precisely $\mathbf{P}$. Incidentally, you may be worried about setting $\hbar = 1$ when it is such a "small" number; $\hbar \approx 1.0547 \times 10^{-34}$ J s rad^{-1}. But this is quite wrong. Every atom in the Universe knows that $\hbar$ is not at all small or insignificant. It only appears small when measured in units such as Joules and seconds that are relevant for steam engines and pocketwatches. The real, deep question is not why $\hbar$ appears small, but why humans are so big.

Plugging (4.20) into the defining equation (4.18) for the translation operator we find

$$\frac{\mathrm{i}}{\hbar}[\delta\mathbf{a}\cdot\mathbf{P}, \mathbf{X}] = \delta\mathbf{a}\,, \tag{4.21}$$

and since this holds for any infinitesimal translation,

$$[X_i, P_j] = \mathrm{i}\hbar\delta_{ij}\,. \tag{4.22}$$

These, as I'm sure you recognise, are the fundamental commutation relations of quantum mechanics. I'm prepared to bet that the first time you met them (and perhaps even up until now) you thought they were a weird and mysterious feature of quantum mechanics. Here they're revealed as little more than common sense: asking where you are and then translating is not the same as first translating and then asking where you are. What is weird and quantum about this equation is not that $\mathbf{X}$ and $\mathbf{P}$ don't commute, but the fact that they are operators acting on a Hilbert space.

By repeatedly performing the same infinitesimal translation, we can write

$$U(\mathbf{a}) = \exp\left(-\frac{\mathrm{i}}{\hbar}\mathbf{a}\cdot\mathbf{P}\right) \tag{4.23}$$

for a translation through the finite vector $\mathbf{a}$. Since $U(\mathbf{a})$ provides a homomorphism of this group of translations to the group of linear operators on $\mathcal{H}$, we have

$$U(\mathbf{b})U(\mathbf{a}) = U(\mathbf{b}+\mathbf{a}) = U(\mathbf{a}+\mathbf{b}) = U(\mathbf{a})U(\mathbf{b})\,, \tag{4.24}$$

where in the second equality we used the fact that $\mathbf{a}+\mathbf{b} = \mathbf{b}+\mathbf{a}$, stating that the order in which we perform two translations doesn't matter[21]. Using (4.20) to expand the far left and far right of this equation to lowest non-trivial order in $a_i b_j$ shows that the generators $\mathbf{P}$ obey

$$[P_i, P_j] = 0 \text{ for all } i, j\,, \tag{4.25}$$

[20] Here, $\delta\mathbf{a}\cdot\mathbf{P} = \sum_i \delta a_i P_i$ involves the standard dot product in $\mathbb{R}^3$. Note that the result is still an operator on $\mathcal{H}$, a linear combination of the operators $\mathbf{P}$.

[21] Here this is of course the statement that the group of spatial translations is $(\mathbb{R}^3, +)$, where the group operation "+" is commutative. Groups for which the group operation is commutative are often called *Abelian*.

and so they commute with each other. Again, the fundamental meaning of this equation is simply that translations form an Abelian group — the order of translations doesn't matter. It only says anything about our ability to simultaneously know all three components of a particle's momentum once we identify $\mathbf{P}$ as corresponding to momentum, which we haven't yet done.

Now let's think about what the translation operator does to states. Let $|\mathbf{x}\rangle$ represent an eigenstate of the position operator with eigenvalue $\mathbf{x}$, so that $\mathbf{X}\,|\mathbf{x}\rangle = \mathbf{x}\,|\mathbf{x}\rangle$, with normalisation condition

$$\langle \mathbf{x}'|\mathbf{x}\rangle = \delta(\mathbf{x}' - \mathbf{x}) \tag{4.26}$$

as usual for continuum states[22]. This state represents a particle which is definitely located at $\mathbf{x} \in \mathbb{R}^3$. Applying the translation operator,

$$\mathbf{X}(U(\mathbf{a})\,|\mathbf{x}\rangle) = ([\mathbf{X}, U(\mathbf{a})] + U(\mathbf{a})\mathbf{X})\,|\mathbf{x}\rangle\ . \tag{4.27}$$

We can evaluate the commutator by multiplying (4.18) through by $U(\mathbf{a})$ on the left, finding $[\mathbf{X}, U(\mathbf{a})] = U(\mathbf{a})\mathbf{a}$. Since $\mathbf{a}$ is just a constant vector (not an operator on $\mathcal{H}$) it commutes with the translation operator, so

$$\mathbf{X}(U(\mathbf{a})\,|\mathbf{x}\rangle) = (\mathbf{x} + \mathbf{a})(U(\mathbf{a})\,|\mathbf{x}\rangle)\,. \tag{4.28}$$

As anticipated, our new state $U(\mathbf{a})\,|\mathbf{x}\rangle$ is certainly located at $\mathbf{x} + \mathbf{a}$. Consequently, $U(\mathbf{a})\,|\mathbf{x}\rangle$ must be proportional to the state $|\mathbf{x} + \mathbf{a}\rangle$, the normalised state that is definitely located at $\mathbf{x} + \mathbf{a}$. Setting $U(\mathbf{a})\,|\mathbf{x}\rangle = c\,|\mathbf{x} + \mathbf{a}\rangle$ for some $c \in \mathbb{C}$ and taking the inner product with another position eigenstate $|\mathbf{x}'\rangle$, we have

$$\begin{aligned} c\delta(\mathbf{x}' - \mathbf{x} - \mathbf{a}) &= c\,\langle \mathbf{x}'|\mathbf{x} + \mathbf{a}\rangle = \langle \mathbf{x}'|U(\mathbf{a})|\mathbf{x}\rangle = \left(U^{-1}(\mathbf{a})\,|\mathbf{x}'\rangle\right)^\dagger |\mathbf{x}\rangle \\ &= \left(\frac{1}{c}\,|\mathbf{x}' - \mathbf{a}\rangle\right)^\dagger |\mathbf{x}\rangle = \frac{1}{c^*}\delta(\mathbf{x}' - \mathbf{a} - \mathbf{x})\,, \end{aligned} \tag{4.29}$$

where we used the fact that $U(\mathbf{a})$ is unitary. Comparing both sides shows that $|c|^2 = 1$, so c is a pure phase and without loss of generality we can set $c = 1$.

Now suppose we consider translating an arbitrary state $|\psi\rangle$. The position space wavefunction of this state is simply its coefficient

$$\psi(\mathbf{x}) = \langle \mathbf{x}|\psi\rangle \tag{4.30}$$

in the position basis. Thus the wavefunction $\psi_{\text{trans}}(\mathbf{x})$ of the translated state $|\psi_{\text{trans}}\rangle = U(\mathbf{a})\,|\psi\rangle$ is

$$\begin{aligned} \psi_{\text{trans}}(\mathbf{x}) = \langle \mathbf{x}|\psi_{\text{trans}}\rangle &= \langle \mathbf{x}|U(\mathbf{a})|\psi\rangle = (U^{-1}(\mathbf{a})\,|\mathbf{x}\rangle)\,|\psi\rangle \\ &= \langle \mathbf{x} - \mathbf{a}|\psi\rangle = \psi(\mathbf{x} - \mathbf{a})\,. \end{aligned} \tag{4.31}$$

This says that the wavefunction of the translated state takes the same value at $\mathbf{x}$ as the original wavefunction took at $\mathbf{x} - \mathbf{a}$. This is completely natural as we've translated our state through $\mathbf{a}$.

In particular, for an infinitesimal translation $\delta\mathbf{a}$, on the one hand we can Taylor expand the translated wavefunction to find[23]

$$\psi_{\text{trans}}(\mathbf{x}) - \psi(\mathbf{x}) = -\delta\mathbf{a} \cdot \boldsymbol{\nabla}\psi(\mathbf{x})\,, \tag{4.32}$$

while on the other hand we expand the operator $U(\delta\mathbf{a})$ to find

$$\psi_{\text{trans}}(\mathbf{x}) - \psi(\mathbf{x}) = \left\langle \mathbf{x}\middle|1 - \frac{\mathrm{i}}{\hbar}\delta\mathbf{a} \cdot \mathbf{P}\middle|\psi\right\rangle - \langle \mathbf{x}|\psi\rangle = -\frac{\mathrm{i}}{\hbar}\delta\mathbf{a} \cdot \langle \mathbf{x}|\mathbf{P}|\psi\rangle \tag{4.33}$$

[22] Technically, $|\mathbf{x}\rangle$ is a non-normalisable state, which really means it doesn't live in the Hilbert space $L^2(\mathbb{R}, \mathrm{d}^3\mathbf{x})$. This is related to the fact that the position operator $\mathbf{X}$ is unbounded — there is no constant M such that $\|\mathbf{X}\,|\psi\rangle\| \leq M\,\||\psi\rangle\|$ for all $|\psi\rangle \in \mathcal{H}$. Naively, this means that "the eigenvalue of $\mathbf{X}$ can be arbitrarily large", which is physically reasonable since our particle may be located very far away. However, in general, unbounded operators do not have any eigenstate in $\mathcal{H}$, so do not really have eigenvalues. The distinction between bounded and unbounded operators is tremendously important in functional analysis, but will not play a role in this course.

[23] At least if the wavefunction is sufficiently smooth

using (4.20) to the lowest order in $\delta\mathbf{a}$. Consequently, for any state $|\psi\rangle$ the momentum operator $\mathbf{P}$ acts in the position representation as

$$\langle\mathbf{x}|\mathbf{P}|\psi\rangle = -\mathrm{i}\hbar\boldsymbol{\nabla}\psi(\mathbf{x})\,. \tag{4.34}$$

This is just what you would have said in IB Quantum Mechanics last year, here derived from the point of view of the effect a translation of $\mathbb{R}^3$ has on states in Hilbert space.

As a special case, let's apply this argument to the state $|\mathbf{p}\rangle$ that obeys[24] $\mathbf{P}\,|\mathbf{p}\rangle = \mathbf{p}\,|\mathbf{p}\rangle$, representing a particle whose momentum is certainly $\mathbf{p}$. In this case the above argument becomes

$$\begin{aligned}\psi_{\mathbf{p}}(\mathbf{x}-\mathbf{a}) &= \langle\mathbf{x}-\mathbf{a}|\mathbf{p}\rangle = \langle\mathbf{x}|U(\mathbf{a})|\mathbf{p}\rangle = \left\langle\mathbf{x}\middle|\mathrm{e}^{-\mathrm{i}\mathbf{a}\cdot\mathbf{P}/\hbar}\middle|\mathbf{p}\right\rangle \\ &= \mathrm{e}^{-\mathrm{i}\mathbf{a}\cdot\mathbf{p}/\hbar}\,\langle\mathbf{x}|\mathbf{p}\rangle = \mathrm{e}^{-\mathrm{i}\mathbf{a}\cdot\mathbf{p}/\hbar}\psi_{\mathbf{p}}(\mathbf{x})\,.\end{aligned} \tag{4.35}$$

Comparing the first and last expression, we deduce that the position space wavefunction of a momentum eigenstate must take the form

$$\psi_{\mathbf{p}}(\mathbf{x}) = C\mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{x}/\hbar} \tag{4.36}$$

for some $\mathbf{x}$-independent factor $C \in \mathbb{C}$. It's convenient to choose C to be $(2\pi\hbar)^{-3/2}$ since in this case we have

$$\begin{aligned}\langle\mathbf{p}'|\mathbf{p}\rangle &= \int \mathrm{d}^3\mathbf{y}\,\mathrm{d}^3\mathbf{x}\;\langle\mathbf{p}'|\mathbf{y}\rangle\,\langle\mathbf{y}|\mathbf{x}\rangle\,\langle\mathbf{x}|\mathbf{p}\rangle = \int \mathrm{d}^3\mathbf{x}\,\mathrm{d}^3\mathbf{y}\;\frac{1}{(2\pi\hbar)^3}\mathrm{e}^{-\mathrm{i}\mathbf{p}'\cdot\mathbf{y}/\hbar}\delta(\mathbf{y}-\mathbf{x})\mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{x}/\hbar} \\ &= \int \mathrm{d}^3\mathbf{x}\;\frac{1}{(2\pi\hbar)^3}\mathrm{e}^{-\mathrm{i}(\mathbf{p}'-\mathbf{p})\cdot\mathbf{x}/\hbar} = \delta(\mathbf{p}'-\mathbf{p})\,,\end{aligned} \tag{4.37}$$

so that the momentum eigenstates are normalised in the same way as the position eigenstates (and in the same way as for any continuum states). Again, the form

$$\psi_{\mathbf{p}}(\mathbf{x}) = \frac{1}{(2\pi\hbar)^{3/2}}\mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{x}/\hbar} \tag{4.38}$$

of the wavefunction for momentum eigenstates is familiar, but here we've derived it from first principles rather than using the position space representation (4.34) of the momentum operator. (We assumed this result earlier when showing that the position and momentum space wavefunctions are each other's Fourier transformations.) Note also that $|\psi_{\mathbf{p}}(\mathbf{x})|^2 = 1/(2\pi\hbar)^3$, so $\int_{\mathbb{R}^3}\mathrm{d}^3\mathbf{x}\;|\psi_{\mathbf{p}}(\mathbf{x})|^2$ diverges, Like position eigenvalues, momentum eigenstates are also non-normalisable.

4.3 Rotations

We now consider the effect rotating $\mathbb{R}^3$ has on a quantum state $|\psi\rangle \in \mathcal{H}$. For a vector $\mathbf{v} \in \mathbb{R}^3$, a rotation anticlockwise around the axis $\hat{\boldsymbol{\alpha}}$ by an amount $|\boldsymbol{\alpha}|$ is a linear transformation

$$\begin{aligned}\mathsf{R}(\boldsymbol{\alpha}) : \mathbb{R}^3 &\longrightarrow \mathbb{R}^3 \\ \mathbf{v} &\longmapsto \mathbf{v}' = \mathsf{R}(\boldsymbol{\alpha})\mathbf{v}\end{aligned} \tag{4.39}$$

that obeys

$$\mathbf{v}'\cdot\mathbf{v}' = \mathbf{v}\cdot\mathbf{v} \quad\text{and}\quad \det(\mathsf{R}(\boldsymbol{\alpha})) = +1\,. \tag{4.40}$$

The first of these conditions says that rotations preserve lengths, and implies that $\mathsf{R}(\boldsymbol{\alpha})$ must be an *orthogonal* transformation. The second ensures that $\mathsf{R}(\boldsymbol{\alpha})$ preserves the orientation. For infinitesimal rotations, figure 4 shows that

$$\mathbf{v}' = \mathbf{v} + \delta\boldsymbol{\alpha}\times\mathbf{v} + O(\delta\boldsymbol{\alpha}^2)\,. \tag{4.41}$$

[24] As with the position operator, the momentum operator $\mathbf{P}$ is unbounded, so the "momentum eigenstates" $|\mathbf{p}\rangle$ are necessarily non-normalisable, as we'll soon see explicitly.

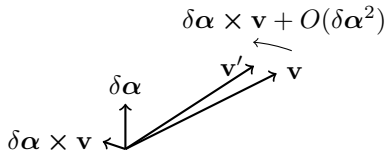


Figure 4: An infinitesimal rotation transforms $\mathbf{v} \mapsto \mathbf{v} + \delta\boldsymbol{\alpha} \times \mathbf{v}$.

Note that this obeys (4.40). Given two orthogonal transformations $\mathsf{R}(\boldsymbol{\alpha})$ and $\mathsf{R}(\boldsymbol{\beta})$, the composite $\mathsf{R}(\boldsymbol{\beta})\mathsf{R}(\boldsymbol{\alpha})$ is again an orthogonal transformation with unit determinant. Spatial rotations form a group, known as SO(3), with the identity being the trivial rotation and the inverse of $\mathsf{R}(\boldsymbol{\alpha})$ being a rotation of the same amount around the same axis $\boldsymbol{\alpha}$, but now clockwise. However, in general $\mathsf{R}(\boldsymbol{\beta})\mathsf{R}(\boldsymbol{\alpha}) = \mathsf{R}(\boldsymbol{\alpha})\mathsf{R}(\boldsymbol{\beta})$, so the order in which we apply rotations is important and the rotation is *non-Abelian*.

As for any group of transformations, in quantum mechanics the group of rotations is represented on $\mathcal{H}$ by unitary operators. We denote the unitary operator corresponding to $\mathsf{R}(\boldsymbol{\alpha})$ as $U(\boldsymbol{\alpha})$. $U(\boldsymbol{\alpha})$ acts on the position operator as

$$U^{-1}(\boldsymbol{\alpha})\mathbf{X}U(\boldsymbol{\alpha}) = \mathsf{R}(\boldsymbol{\alpha})\mathbf{X}\,, \tag{4.42}$$

where the left-hand side involves composition of operators in $\mathcal{H}$, while the right-hand side is the usual action of a rotation on the three components of $\mathbf{X}$. For example, if $\boldsymbol{\alpha} = \alpha\hat{\mathbf{z}}$ so that the rotation is around the z-axis in $\mathbb{R}^3$, then in more detail (4.42) says

$$\begin{pmatrix} U^{-1}(\boldsymbol{\alpha})XU(\boldsymbol{\alpha}) \\ U^{-1}(\boldsymbol{\alpha})YU(\boldsymbol{\alpha}) \\ U^{-1}(\boldsymbol{\alpha})ZU(\boldsymbol{\alpha}) \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}. \tag{4.43}$$

Thus, conjugating the operator X by this rotation operator changes it to a new operator $U^{-1}(\boldsymbol{\alpha})XU(\boldsymbol{\alpha})$ on Hilbert space, which we identity with the usual linear combination $X\cos\alpha - Y\sin\alpha$. Again, $U(\boldsymbol{\alpha})$ must act on $\mathbf{X}$ this way if it is indeed to correspond to a rotation.

For an infinitesimal transformation, following (4.8) we can write

$$U(\delta\boldsymbol{\alpha}) = 1 - \frac{\mathrm{i}}{\hbar}\delta\boldsymbol{\alpha}\cdot\mathbf{J} + O(\delta\boldsymbol{\alpha}^2) \tag{4.44}$$

for some Hermitian generators $\mathbf{J}/\hbar$. Since angles are dimensionless, $\mathbf{J}$, like $\hbar$, must have dimensions of (length $\times$ momentum). Later we will see that $\mathbf{J}$ corresponds to the angular momentum operator, though by definition we have that $\mathbf{J}/\hbar$ is the generator of rotations. Using this and (4.41) in (4.42) shows that we have

$$\frac{\mathrm{i}}{\hbar}[\delta\boldsymbol{\alpha}\cdot\mathbf{J}, \mathbf{X}] = \delta\boldsymbol{\alpha}\times\mathbf{X}\,, \tag{4.45}$$

and since this is true for any axis of rotation, we have the commutation relations

$$[J_i, X_j] = \mathrm{i}\hbar\sum_k \epsilon_{ijk}X_k\,. \tag{4.46}$$

These relations are nothing more than the statement that the operator $\mathbf{X}$ transforms as a vector under rotations. They are the infinitesimal version of (4.40) — they must hold if $\mathbf{J}$ is indeed to generate rotations.

Just as for translations, the mutual commutation relations among the components of $\mathbf{J}$ themselves follow from the fact that the $U(\boldsymbol{\alpha})$'s provide a homomorphism from the group[25] SO(3) of rotations to the space of unitary operators on $\mathcal{H}$. However, the non-Abelian nature of the rotation group means we

[25] Again, later we'll see that this statement needs to be slightly refined.

should expect these commutators to be non-trivial in general. Performing two infinitesimal rotations through angles $\delta\boldsymbol{\alpha}$ and $\delta\boldsymbol{\beta}$ for any vector $\mathbf{v}$ gives

$$\begin{aligned}\mathsf{R}(\delta\boldsymbol{\beta})\mathsf{R}(\delta\boldsymbol{\alpha})\mathbf{v} &= \mathsf{R}(\delta\boldsymbol{\beta})(\mathbf{v}+\delta\boldsymbol{\alpha}\times\mathbf{v}) = (\mathbf{v}+\delta\boldsymbol{\alpha}\times\mathbf{v}) + \delta\boldsymbol{\beta}\times(\mathbf{v}+\delta\boldsymbol{\alpha}\times\mathbf{v}) \\ &= \mathbf{v}+\delta\boldsymbol{\alpha}\times\mathbf{v}+\delta\boldsymbol{\beta}\times\mathbf{v}+\delta\boldsymbol{\beta}\times(\delta\boldsymbol{\alpha}\times\mathbf{v}) \end{aligned} \tag{4.47}$$

to the lowest non-trivial order in $\delta\boldsymbol{\alpha}$ and $\delta\boldsymbol{\beta}$. Consequently, the difference between these rotations and the two rotations performed in the opposite order is

$$\begin{aligned}[\mathsf{R}(\delta\boldsymbol{\beta})\mathsf{R}(\delta\boldsymbol{\alpha}) - \mathsf{R}(\delta\boldsymbol{\alpha})\mathsf{R}(\delta\boldsymbol{\beta})] &= \delta\boldsymbol{\beta}\times(\delta\boldsymbol{\alpha}\times\mathbf{v}) - \delta\boldsymbol{\alpha}\times(\delta\boldsymbol{\beta}\times\mathbf{v}) \\ &= (\delta\boldsymbol{\beta}\times\delta\boldsymbol{\alpha})\times\mathbf{v} \\ &= \mathsf{R}(\delta\boldsymbol{\beta}\times\delta\boldsymbol{\alpha})\mathbf{v} - \mathbf{v} \end{aligned} \tag{4.48}$$

using the standard properties of the vector triple product. The right-hand side again involves a rotation, through $\delta\boldsymbol{\beta}\times\delta\boldsymbol{\alpha}$. Applying the homomorphism U, we obtain

$$[U(\delta\boldsymbol{\beta}), U(\delta\boldsymbol{\alpha})] = U(\delta\boldsymbol{\beta}\times\delta\boldsymbol{\alpha}) - 1_{\mathcal{H}} \tag{4.49}$$

for our operators in Hilbert space. Using (4.44), this is

$$-\frac{1}{\hbar^2}[\delta\boldsymbol{\beta}\cdot\mathbf{J}, \delta\boldsymbol{\alpha}\cdot\mathbf{J}] = -\frac{\mathrm{i}}{\hbar}(\delta\boldsymbol{\beta}\times\delta\boldsymbol{\alpha})\cdot\mathbf{J}\,, \tag{4.50}$$

and since this must hold for arbitrary successive infinitesimal rotations, we have

$$[J_i, J_j] = \mathrm{i}\hbar\sum_k \epsilon_{ijk}J_k \tag{4.51}$$

as the commutation relations among the rotation generators.

Again, there's nothing "weird" or "quantum" about (4.51), beyond the fact that they involve operators on Hilbert space. In particular, their form just reflects the fact that the order of rotations around different axes matters, and that the difference between the two ordering is itself a rotation around the axis perpendicular the original two. Compared to the relations $[P_i, P_j] = 0$, the non-triviality of the commutation relations (4.51) arises purely because SO(3) is a non-Abelian group, unlike the group of translations. These non-trivial commutation relations do not prevent us from exponentiating the infinitesimal rotation operator (4.44) to write $U(\boldsymbol{\alpha}) = \mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{J}/\hbar}$ for a finite rotation around a fixed axis $\boldsymbol{\alpha}$, because the exponentiation always involve the same component $\hat{\boldsymbol{\alpha}}\cdot\mathbf{J}$ of the rotation generator, which certainly commutes with itself.

The combined group of translations and rotations of Euclidean space $\mathbb{R}^3$ is known as the *Euclidean group*, denoted E(3) or ISO(3). For translations through $\mathbf{a}_1$ and $\mathbf{a}_2$, and rotations R_1 and R_2, E_3 has the group composition law

$$(\mathbf{a}_2, \mathsf{R}_2)\cdot(\mathbf{a}_1, \mathsf{R}_1) = (\mathbf{a}_2 + \mathsf{R}_2\mathbf{a}_1, \mathsf{R}_2\mathsf{R}_1)\,, \tag{4.52}$$

where we note that the second rotation also acts on the first translation. Clearly, both $\mathbb{R}^3$ and SO(3) are subgroups of E(3), but the fact that the rotations act non-trivially on previous translations means that E(3) is the *semi-direct product* $\mathrm{E}(3) \cong \mathbb{R}^3 \rtimes \mathrm{SO}(3)$. This group composition law shows that rotations and translations do not commute: translating any vector $\mathbf{v}$ through $\mathbf{a}$ then rotating around $\hat{\boldsymbol{\alpha}}$ gives us $\mathsf{R}(\boldsymbol{\alpha})(\mathbf{v}+\mathbf{a})$, whereas first rotating then translating gives $\mathbf{a}+\mathsf{R}(\boldsymbol{\alpha})\mathbf{v}$ instead. In particular, for infinitesimal translations and rotations we have

$$\mathsf{R}(\delta\boldsymbol{\alpha})(\mathbf{v}+\delta\mathbf{a}) - (\delta\mathbf{a} + \mathsf{R}(\delta\boldsymbol{\alpha})\mathbf{v}) = \delta\boldsymbol{\alpha}\times\delta\mathbf{a} \tag{4.53}$$

independent of $\mathbf{v}$. Applying the homomorphism U, in Hilbert space we obtain commutation relations

$$[U_R(\delta\boldsymbol{\alpha}), U_T(\delta\mathbf{a})] = U_T(\delta\boldsymbol{\alpha}\times\delta\mathbf{a}) - 1\,, \tag{4.54}$$

where, for the sake of clarity, we have labelled the translations operator by U_T and rotation operator by U_R. Equivalently, we have

$$[J_i, P_j] = \mathrm{i}\hbar \sum_k \epsilon_{ijk} P_k \tag{4.55}$$

in terms of the rotation and translation generators.

More generally, and 3-component linear operator $\mathbf{V} : \mathcal{H} \to \mathcal{H}$ is said to *transform as a vector under rotations* if it obeys $U^{-1}(\boldsymbol{\alpha})\mathbf{V}U(\boldsymbol{\alpha}) = \mathsf{R}(\boldsymbol{\alpha})\mathbf{V}$ for any $\boldsymbol{\alpha}$. Depending on its behaviour under parity transformations, which will be discussed in section 4.6, such a vector $\mathbf{V}$ could either be a vector or a pseudovector operator. Just as above, for infinitesimal rotations this implies the commutation relations

$$[J_i, V_j] = \mathrm{i}\hbar \sum_k \epsilon_{ijk} V_k \tag{4.56}$$

for the components of $\mathbf{V}$ and $\mathbf{J}$. We see that $\mathbf{X}, \mathbf{P}$ and $\mathbf{J}$ itself[26] each transform as vectors under rotations. This of course is just what we'd expect for position, momentum and angular momentum in classical mechanics.

On the other hand, if an operator S obeys

$$U^{-1}(\boldsymbol{\alpha})SU(\boldsymbol{\alpha}) = S \tag{4.57}$$

for any $\boldsymbol{\alpha}$, so that it is unchanged by the rotation operator, we say S *transforms as a scalar under rotations*. Again, depending on its behaviours under parity transformations it could either be a scalar or a pseudoscalar operator. The corresponding infinitesimal version is

$$[\mathbf{J}, S] = 0\,. \tag{4.58}$$

Just as we can form a scalar by taking the Euclidean inner product $\mathbf{v}\cdot\mathbf{w}$ of the two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, we can form scalar operators from Euclidean inner product of two vector operators. Indeed, if $U^{-1}(\boldsymbol{\alpha})\mathbf{V}U(\boldsymbol{\alpha}) = \mathsf{R}(\boldsymbol{\alpha})\mathbf{V}$ and similarly for $\mathbf{W}$, then

$$\begin{aligned} U^{-1}(\boldsymbol{\alpha})(\mathbf{V}\cdot\mathbf{W})U(\boldsymbol{\alpha}) &= (U^{-1}(\boldsymbol{\alpha})\mathbf{V}U(\boldsymbol{\alpha}))\cdot(U^{-1}(\boldsymbol{\alpha})\mathbf{W}U(\boldsymbol{\alpha})) \\ &= (\mathsf{R}(\boldsymbol{\alpha})\mathbf{V})\cdot(\mathsf{R}(\boldsymbol{\alpha})\mathbf{W}) \\ &= \mathbf{V}\cdot\mathbf{W} \end{aligned} \tag{4.59}$$

by the standard rotational invariance of the dot product. As always, we can write this in terms of commutators by considering infinitesimal rotations:

$$\begin{aligned} \left[J_i, \sum_j V_j W_j\right] &= \sum_j [J_i, V_j] W_j + \sum_j V_j [J_i, W_j] \\ &= \mathrm{i}\hbar \sum_{j,k} \epsilon_{ijk} V_k W_j + \mathrm{i}\hbar \sum_{j,k} V_j \epsilon_{ijk} W_k \\ &= \mathrm{i}\hbar \sum_{j,k} \epsilon_{ijk}(-V_j W_k + V_j W_k) = 0\,, \end{aligned} \tag{4.60}$$

where in going to the final line we relabelled the dummy indices $j \leftrightarrow k$ in the first term and used the antisymmetry of ϵ_{ijk}. Note that our calculations always preserved the order of $\mathbf{V}$ and $\mathbf{W}$, so our result holds irrespective of whether $\mathbf{V}$ and $\mathbf{W}$ commute or not.

An important special case of (4.58) is to take $S = \mathbf{J}^2 = \mathbf{J}\cdot\mathbf{J}$. The fact that the rotation generators obey the commutation relations

$$[J_i, J_j] = \mathrm{i}\hbar \sum_k \epsilon_{ijk} J_k \quad \text{and} \quad [J_i, \mathbf{J}^2] = 0 \tag{4.61}$$

[26] This is always true of $\mathbf{X}$ and $\mathbf{P}$, but turns out to be a three-dimensional coincidence for $\mathbf{J}$: the rotation group $\mathrm{SO}(d)$ of $\mathbb{R}^d$ has dimension $d(d-1)/2$ and the generators are generically represented by antisymmetric matrices $\mathcal{J}_{ij}$. When $d = 3$, a 3×3 antisymmetric matrix has 3 independent components, so we can equivalently package the $\mathcal{J}_{ij}$ as vectors through $J_i = \epsilon_{ijk}\mathcal{J}_{jk}$.

means that we cannot find a complete set of simultaneous eigenstates of all of the J_i's, but we can find a complete set of simultaneous eigenstates of $\mathbf{J}^2$ and any one component of $\mathbf{J}$. Looking ahead to our interpretation of $\mathbf{J}$ as the angular momentum operator, Born's 2nd postulate of quantum mechanics tells us that we cannot say a particle has a definite angular momentum vector, but we can know the magnitude of the angular momentum and the amount aligned along any given axis.

4.3.1 Translations Around a Circle

In IB Quantum Mechanics, we defined the orbital angular momentum operator $\mathbf{L} = \mathbf{X} \times \mathbf{P}$ which also has the commutation relations (4.56) with itself and with X and P. We'll understand the relation between $\mathbf{J}$ and $\mathbf{L}$ later, but it's important to understand that in general $\mathbf{J} \neq \mathbf{L}$. We can understand $\mathbf{L}$ from the present perspective as follows.

When a system is displaced through the vector $\mathbf{a}$, its state is transformed by the unitary operator $U(\mathbf{a}) = \mathrm{e}^{-\mathrm{i}\mathbf{a}\cdot\mathbf{P}/\hbar}$. We now imagine successively performing n translations successively through the set of vectors $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$. Each translation is represented by a unitary operator $U(\mathbf{a}_i)$, so the final state will be

$$U(\mathbf{a}_n) \ldots U(\mathbf{a}_1) \left|\psi\right\rangle = U(\mathbf{b}) \left|\psi\right\rangle \ , \tag{4.62}$$

where $\mathbf{b} = \mathbf{a}_1 + \ldots \mathbf{a}_n$ is the total displacement vector. Since the net result depends only on the total translation $\mathbf{b}$, the change in $|\psi\rangle$ is independent of the particular path that the system takes. In particular, if the path is closed so that $\mathbf{a}_1 + \cdots + \mathbf{a}_n = \mathbf{0}$, then $|\psi\rangle$ is unchanged.

Now consider the effect of translating the system around an arc of a circle centred on the origin. We can approximate the circle by an N-sided regular polygon, with the approximation improving as $N \to \infty$, so we move around a circle by applying a succession of small translations, each in a slightly different direction. Specifically, if the system initially lies at some location $\mathbf{x}$, making an angle α with some axis $\mathbf{n}$, then when move it in the plane normal to $\mathbf{n}$ to lie at an angle $\alpha + \delta\alpha$, we translate it through $\delta\mathbf{a} = \delta\alpha\mathbf{n} \times \mathbf{x}$. Thus the associated unitary translation operator obeys

$$U^{-1}(\delta\mathbf{a})\mathbf{X}U(\delta\mathbf{a}) = \mathbf{X} + \delta\alpha(\mathbf{n} \times \mathbf{X}) \, , \tag{4.63}$$

and so can be written as

$$U(\delta\mathbf{a}) = 1 - \frac{\mathrm{i}}{\hbar}\delta\alpha(\mathbf{n} \times \mathbf{X}) \cdot \mathbf{P} + O(\delta\alpha^2) = 1 - \frac{\mathrm{i}}{\hbar}\delta\alpha\mathbf{n} \cdot \mathbf{L} + O(\delta\alpha^2) \, , \tag{4.64}$$

where

$$\mathbf{L} = \mathbf{X} \times \mathbf{P} \tag{4.65}$$

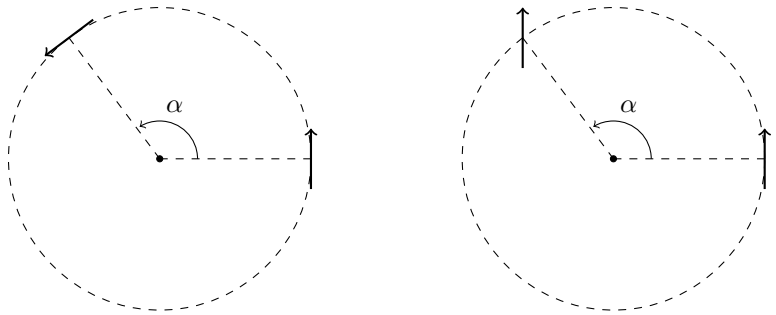
is a Hermitian operator. We now see that $\mathbf{L}$ is the generator of circular transformations. (4.64) contains only one operator $\mathbf{n} \cdot \mathbf{L}$, so it inevitably commutes with itself and we can exponentiate to find $U(\mathbf{a}_{\mathrm{circ}}) = \mathrm{e}^{-\mathrm{i}\alpha\mathbf{n}\cdot\mathbf{L}/\hbar}$ for finite translations around our circle.

4.3.2 Spin

The commutation relations $[L_i, X_j]$, $[L_i, P_j]$ and $[L_i, L_j]$ of the composite operator $\mathbf{L}$ all follow from the more primitive commutation relations $[X_i, P_j]$, $[X_i, X_j]$ and $[P_i, P_j]$. If we're only concerned with these operators, then $U_{\mathrm{circ}} = \mathrm{e}^{-\mathrm{i}\alpha\mathbf{n}\cdot\mathbf{L}/\hbar}$ is indistinguishable from the rotation operator $U(\boldsymbol{\alpha}) = \mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{J}/\hbar}$. However, if our system has any internal structure, circular translations using the centre of mass position operator $\mathbf{X}$ and centre of mass momentum $\mathbf{P}$ are not the same as rotations: the difference is best explained by a picture, which you can find in figure 5.

We define the *spin* operator $\mathbf{S}$ to be the difference

$$\mathbf{S} = \mathbf{J} - \mathbf{L} \tag{4.66}$$



(a) rotation (b) translation around the circle

Figure 5: The rotation operator $U(\boldsymbol{\alpha})$ swings a system around the origin and also rotates its orientation in $\mathbb{R}^3$, while a circular translation merely moves the system around a circular path, without affecting its orientation. The difference is a rotation of the body around its own centre of mass, reorientating it without changing its location.

so that $\mathbf{J} = \mathbf{L} + \mathbf{S}$. From figure 5 we see that difference between rotating an object around some fixed origin and a translating it's centre of mass along a circular path is a rotation around the object's centre of mass. generates a rotation of the body around its own centre of mass. We thus expect (and will confirm below) that $\mathbf{S}$ generates rotations of a body around its own centre of mass, reorienting it in space. This is why $\mathbf{S}$ is called the spin operator.

In the case of a macroscopic body, made up of many constituent particles, it's reasonable to suppose that the spin operator just account for the difference between translating the body as a whole and translating each individual particle around their own arcs, with slightly different radii according to where in the body the particle is located. That is,

$$\mathbf{S} \stackrel{?}{=} \left(\sum_a \mathbf{X}_a \times \mathbf{P}_a \right) - \mathbf{X} \times \mathbf{P} = \sum_a \mathbf{x}_a \times \mathbf{p}_a \,, \tag{4.67}$$

where $\mathbf{X}_a$ and $\mathbf{P}_a$ are position and translation operators for each individual particle and $\mathbf{x}_a$ and $\mathbf{p}_a$ are their positions and momenta relative to the centre of mass[27]. However, for an object consisting of many particles such a description is clearly going to be very cumbersome. More fundamentally, we do not know what the "fundamental" constituents of our object really are. (For example, if I sit on a merry-go-round, is the "right" quantum description of my motion given in terms of my cells, or my atoms, or protons and neutrons, or quarks and gluons, or bits of string, or ...?) It's thus crucial that we can consider objects as a whole in quantum mechanics, just as we can classically. For rotations, understanding $\mathbf{S}$ will allow us to do this. In addition, as we'll see in section 5.3.1, one of the surprises of quantum mechanics is that even fundamental particles such as electrons and photons may have an "intrinsic" spin that (as far as we know) is not related to any composite structure.

Fortunately, commutation relations involving the spin operator $\mathbf{S}$ are easy to obtain from the ones we already have for $\mathbf{J}$ and $\mathbf{L}$. Firstly, the fundamental relations

$$[J_i, X_j] = \mathrm{i}\hbar \sum_k \epsilon_{ijk} X_k \quad \text{and} \quad [J_i, P_j] = \mathrm{i}\hbar \sum_k \epsilon_{ijk} P_k \tag{4.68}$$

show that

$$[J_i, L_j] = \mathrm{i}\hbar \sum_k \epsilon_{ijk} L_k \,, \tag{4.69}$$

so that the angular momentum operator $\mathbf{L}$ also transforms as a vector under rotations. Recalling

[27] We'll understand how to properly describe the quantum mechanics of a system with many degrees of freedom (e.g. many particles) in section 7.

from IB Quantum Mechanics that $[L_i, L_j] = \mathrm{i}\hbar \sum_k \epsilon_{ijk} L_k$, we have

$$\begin{aligned}[S_i, S_j] &= [J_i - L_i, J_j - L_j] = [J_i, J_j] - [J_i, L_j] - [L_i, J_j] + [L_i, L_j] \\ &= \mathrm{i}\hbar \sum_k \epsilon_{ijk}(J_k - L_k) = \mathrm{i}\hbar \sum_k \epsilon_{ijk} S_k \,. \end{aligned} \tag{4.70}$$

It then immediately follows that $[\mathbf{S}, \mathbf{S}^2] = 0$. This algebra — the same as that obeyed by the components of $\mathbf{J}$ and $\mathbf{L}$ — confirms that the spin operator $\mathbf{S}$ generate some form of rotation. On the other hand, since $[L_i, X_j] = \mathrm{i}\hbar \sum_k \epsilon_{ijk} X_k$ and $[L_i, P_j] = \mathrm{i}\hbar \sum_k \epsilon_{ijk} P_k$, we have

$$[S_i, X_j] = [J_i, X_j] - [L_i, X_j] = 0 \,, \tag{4.71}$$

$$[S_i, P_j] = [J_i, P_j] - [L_i, P_j] = 0 \,. \tag{4.72}$$

These commutation relations confirm that the spin operator has nothing to do with an object's location in or motion through space, but is purely to do with rotating its intrinsic orientation. Finally, since $\mathbf{S}$ commutes with both $\mathbf{X}$ and $\mathbf{P}$, it also commutes with $\mathbf{L}$:

$$[S_i, L_j] = 0 \quad \forall i, j \,. \tag{4.73}$$

This allows us to factorize the operator $U(\boldsymbol{\alpha})$ describing finite rotations as

$$U(\boldsymbol{\alpha}) = \mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{J}/\hbar} = \mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot(\mathbf{L}+\mathbf{S})/\hbar} = \mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{L}/\hbar}\mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{S}/\hbar} = \mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{S}/\hbar}\mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{L}/\hbar} \,. \tag{4.74}$$

As in figure 5, these equations confirm that we can think of a quantum rotation as consisting of a translation of a body's centre of mass along an arc centred on the origin together with a simultaneous rotation of the body around its own centre of mass by the same amount. The order in which we perform these two operations makes no difference.

4.4 Time Translations

It's not only transformations of space that are represented on $\mathcal{H}$ by unitary operators. Consider translations in time. These again form an Abelian group, since sending t_0 to $t_0 + t$ and then to $t_0 + t + t'$ gives the same end result as does $t_0 \mapsto t_0 + t' \mapsto t_0 + t' + t$. If we want the total probability of our particle being found somewhere in $\mathbb{R}^3$ at all times, then time translation must also be represented by a unitary operator $U(t) : \mathcal{H} \to \mathcal{H}$. Since we can translate through an arbitrarily small time, the time translation operator takes the form

$$U(t) = \exp\left(-\frac{\mathrm{i}}{\hbar} H t\right) \,, \tag{4.75}$$

where $H/\hbar$ must be Hermitian and is, by definition, the generator of time translations. This generator must have dimensions 1/(time), so the conventional factor of $\hbar$ means that H itself has dimensions of energy. Of course, H will turn out to be the Hamiltonian, but for now I'd like you to think of it more abstractly just as the generator of time translations.

The fact that $U(t)$ represents the effect of a time translation on $\mathcal{H}$ means that if a particle is in some state $|\psi(0)\rangle$ at time $t_0 = 0$, translating forward to time t we will find it in the state

$$|\psi(t)\rangle = U(t)\,|\psi(0)\rangle = \mathrm{e}^{-\mathrm{i}Ht/\hbar}\,|\psi(0)\rangle \,. \tag{4.76}$$

In particular, the difference between $|\psi(t)\rangle$ and the state we find a short time later is

$$|\psi(t+\delta t)\rangle - |\psi(t)\rangle = -\frac{\mathrm{i}}{\hbar}\delta t H\,|\psi(t)\rangle + U(\delta t^2) \,, \tag{4.77}$$

or, taking the limit $\delta t \to 0$,

$$\mathrm{i}\hbar \frac{\partial}{\partial t}\,|\psi(t)\rangle = H\,|\psi(t)\rangle \,. \tag{4.78}$$

This, famously, is the time dependent Schrödinger equation. It's just the infinitesimal version of the statement that all states in $\mathcal{H}$ evolve in time according to the action of a unitary operator $U(t)$.

4.4.1 The Heisenberg Picture

In IB Quantum Mechanics, we were used to the idea that states evolve in time according to the time dependent Schrödinger equation. However, just as we did for spatial transformations, we can instead work in a picture where the states are time independent and instead the operators evolve in time. To see how this works, suppose O_S is some operator that contains no explicit time dependence. Then the amplitude for the state $O_S\,|\psi(t)\rangle$ to agree at time t with the state $|\chi(t)\rangle$ is

$$\langle\chi(t)|O_S|\psi(t)\rangle = \big\langle\chi(0)\big|U^{-1}(t)O_S U(t)\big|\psi(0)\big\rangle\ . \tag{4.79}$$

The right hand side here only makes explicit reference to the initial values of the states $|\psi\rangle$ and $|\chi\rangle$. Since it holds true for any pair of states, just as above we can obtain the same results by always working with the initial states and instead evolving the operators as

$$O_H(t) = U^{-1}(t)O_S U(t)\,. \tag{4.80}$$

The version of quantum mechanics where we use the time dependent Schrödinger equation to evolve the states in time, leaving operators unaltered, is known as the Schrödinger picture, whereas the version where the states are fixed at their initial values and instead the operators evolve in time is called the Heisenberg picture.

In fact, all this has a precise analogue in classical mechanics. Classically, there are also two ways of thinking about time evolution. On the one hand, we can think of a particle moving in some way through phase space M. If we know it's location $(\mathbf{x}(t), \mathbf{p}(t)) \in M$ for every time t we can compute any quantity we wish, represented by some function $f : M \to \mathbb{R}$, by evaluating f at the location of our particle, obtaining the value $f(\mathbf{x}(t), \mathbf{p}(t))$. (This is the perspective we took in the Introduction.) However, Newton's Laws are deterministic, so, given a force, the entire trajectory is determined by the initial conditions $(\mathbf{x}_0, \mathbf{p}_0)$. This suggests a perspective in which the "state" of our particle is simply a choice of initial conditions. These initial conditions do not themselves evolve, rather, it is the quantities we measure that vary in time. Thus, instead of thinking of a physical quantity f as a map from phase space, we treat it just as a map from time, so $f : [t_0, \infty) \to \mathbb{R}$. You'll examine these classical pictures further if you're taking the Classical Dynamics course, in the context of Hamiltonian mechanics. (It's really for this reformulation of classical mechanics, not the particular H, that Hamilton is famous.)

Differentiating (4.79) with respect to t by product rule shows that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}O_H(t) &= \frac{\mathrm{d}}{\mathrm{d}t}(U^{-1}(t)O_S U(t)) = \frac{\mathrm{i}}{\hbar}(U^{-1}(t)HO_S U(t) - U^{-1}O_S HU(t)) \\ &= \frac{\mathrm{i}}{\hbar}[H, O_H(t)]\,,\end{aligned} \tag{4.81}$$

where in the last step we used the fact that $[U(t), H] = 0$, since $U(t)$ depends only on H. This is the *Heisenberg equation of motion*. It's completely equivalent to the Schrödinger equation, and is also just a (very important) special case of our general argument of how operators transform under the action of unitary groups of transformations. More generally, if the original operator O_S had some explicit time dependence of its own — independent of any particular particle's motion — then we would obtain a further term in this equation, modifying it to

$$\frac{\mathrm{d}}{\mathrm{d}t}O_H(t) = U^{-1}(t)\frac{\partial O_S}{\partial t}U(t) + \frac{\mathrm{i}}{\hbar}[H, O_H(t)]\,. \tag{4.82}$$

For example, we may wish to understand the behaviour of a charged particle in the presence of an applied electric field. If the electric field is itself changing in time, then the potential in which the particle moves will be time-dependent even in the Schrödinger picture.

4.5 Dynamics

So far, we've simply defined some unitary operators $U(\mathbf{a})$, $U(\boldsymbol{\alpha})$ and $U(t)$ that are responsible for translating, rotating or evolving our system through space and time. The commutation relations of the generators $\mathbf{P}$ and $\mathbf{J}$, and their commutation relations with $\mathbf{X}$ were determined purely by the properties of the corresponding group of transformations of $\mathbb{R}^3$. However, while we've seen that commutators such as $[H, \mathbf{X}]$, $[H, \mathbf{P}]$ and $[H, \mathbf{J}]$ are important in the Heisenberg picture, telling us how these operators change in time, we haven't yet given any way to actually calculate what such commutators should be. Similarly, although we've said that in the Schrödinger picture, states evolve in time according to $|\psi(t)\rangle = U(t)\,|\psi(0)\rangle$, we haven't given any way to work out what the action of the generator H on a state should actually be.

To do so, we must specify the dynamics: we must provide a relation $H = H(\mathbf{X}, \mathbf{P})$ giving H in terms of the other operators whose commutators we already understand. The simplest (non-trivial) example of such relation is[28]

$$H = \frac{1}{2m}\mathbf{P}^2 \tag{4.83}$$

where m is a constant with dimensions of mass. Equation (4.83) relates how a state evolves in time (via H) to how its location is translated through space (via $\mathbf{P}$); in other words, it's telling us something about how particles move! In particular, with this form of Hamiltonian the difference between the expected location of a state $|\psi\rangle$ at an initial time and a short time δt later is

$$\begin{aligned}\left\langle\psi\middle|U^{-1}(\delta t)\mathbf{X}U(\delta t)\middle|\psi\right\rangle - \langle\psi|\mathbf{X}|\psi\rangle &= \frac{\mathrm{i}}{\hbar}\delta t\,\langle\psi|[H, \mathbf{X}]|\psi\rangle + O(\delta t^2) \\ &= \frac{\delta t}{m}\,\langle\psi|\mathbf{P}|\psi\rangle + O(\delta t^2)\,. \end{aligned} \tag{4.84}$$

In the limit $\delta t \to 0$, we see that the expected velocity of the particle is $\langle\psi|\mathbf{P}|\psi\rangle\,/m$. Equivalently, since this is true for all $|\psi\rangle \in \mathcal{H}$, we can write

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{X}(t) = \frac{\mathbf{P}(t)}{m} \tag{4.85}$$

in the Heisenberg picture. Furthermore, since $[H, \mathbf{P}] = 0$ for this Hamiltonian, $P(t) = P(0)$ and all states travel in uniform motion.

In the real world, we observe that particles do not always travel with constant velocities: they may slow down, speed up, or change direction as they encounter various obstacles. These obstacles are typically located at various different points in space. To allow for this, we generalise our dynamical relation (4.83) to

$$H = \frac{1}{2m}\mathbf{P}^2 + V(\mathbf{X})\,. \tag{4.86}$$

The first term on the right is the kinetic term and is the contribution to the Hamiltonian due to the state is traveling through space. The second, potential term is the contribution due to the state's location. This familiar form is still a very special case of the general statement $H = H(\mathbf{X}, \mathbf{P})$ and later in the course we'll meet examples of Hamiltonians that don't fit this form.

Repeating the calculation we still find that $\mathrm{d}\mathbf{X}(t)/\mathrm{d}t = \mathbf{P}(t)/m$, but now $[H, \mathbf{P}] \neq 0$ so the rate of motion through space will not always be the same. Rather, using (4.81), the Heisenberg picture

[28] Note that this is the first time we've pinned ourselves down to a non-relativistic theory. Up to this point, everything we've said holds good in relativistic quantum mechanics and, suitably interpreted, also in quantum field theory. In the lectures we studied translations and rotations, and you can extend this to also consider Galilean boosts and can show that the non-relativistic Hamiltonian is compatible with the Galilean algebra. If one instead wishes to study a relativistic version of quantum theory, one starts by finding unitary operators representing the action of the Poincaré group $\mathbb{R}^{3,1} \rtimes \mathrm{SO}(3,1)$ of special relativity. The relation (4.83) between H and $\mathbf{P}$ does not respect the commutation relations appropriate for Poincaré symmetry, but of course the dynamical relation $H^2 = c^2\mathbf{P}^2 + m^2c^4$ does. The difficulty with relativistic quantum mechanics lies not with symmetries, but with interactions. The proper treatment of these requires Quantum Field Theory, a far deeper subject.

operators obey

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{P}(t) = \frac{\mathrm{i}}{\hbar}[H, \mathbf{P}(t)] = -\boldsymbol{\nabla} V(t)\,, \tag{4.87}$$

where $\boldsymbol{\nabla} V(t) = U^{-1}(t)(\partial V/\partial \mathbf{X})U(t) = -\boldsymbol{\nabla} V(\mathbf{X}(t))$. Thus the motion slows down or speeds up according to the gradient of $V(\mathbf{X})$.

Most of our common intuition about momentum comes from equation (4.87). We "feel" that a tennis ball traveling with a certain speed has less momentum than a cannonball traveling with the same speed, and less than the same tennis ball traveling faster, because we've known from an early age that it will cause us less damage to stand in the way of the first tennis ball than either of the other two. This is really a statement about what our unfortunate bodies will have to do in order to bring these projectiles to rest (or how much effort we will have to exert to launch them in the first place). In other words, our intuitive notion of momentum is built on a feeling for the energy our body will gain as we slow the projectile down during the impact. This energy then excites the atoms that ultimately make up our nerves, muscles and bones, becoming dissipated through our bodies. Ultimately, it is the dynamical relation between the Hamiltonian and other operators that justifies us identifying the operator $\mathbf{P}/\hbar$ — by definition, the generator of translations — with our pre-existing notion of momentum (in units of $\hbar$). Only after we specify our dynamics do commutation relations such as $[X_i, P_j] = \mathrm{i}\hbar\delta_{ij}$ tell us, via Born's 2nd postulate of Quantum Mechanics, that a particle cannot simultaneously be in a state of well-defined position and of well-defined momentum.

We'll often find it useful to write the kinetic term in a slightly different form. Keeping careful track of the operator ordering, you'll show in the problem set that

$$\mathbf{L}^2 = (\mathbf{X}\times\mathbf{P})\cdot(\mathbf{X}\times\mathbf{P}) = \mathbf{X}^2(\mathbf{P}^2 - P_r^2)\,, \tag{4.88}$$

where for $\hat{\mathbf{X}} = \mathbf{X}/\,\|\mathbf{X}\|$,

$$P_r = \frac{1}{2}\left(\hat{\mathbf{X}}\cdot\mathbf{P} + \mathbf{P}\cdot\hat{\mathbf{X}}\right) \tag{4.89}$$

is the radial angular momentum operator. Consequently, we can write the Hamiltonian (4.86) in terms of the generator of circular translations and the radial momentum operator P_r as

$$H = \frac{P_r^2}{2m} + \frac{\mathbf{L}^2}{2m\mathbf{X}^2} + V(\mathbf{X})\,, \tag{4.90}$$

where we note that $[\mathbf{L}, \mathbf{X}^2] = 0$, so the order of the second term is irrelevant provided we keep $\mathbf{X}^2$ together as a composite operator. (In particular, in the position representation, $\mathbf{X}^2$ acts as multiplication by $\mathbf{x}^2 = r^2$, which is certainly unaffected by translation in a circle.) Most of our intuitive feeling about angular momentum — like our happy childhood hours spent on merry-go-rounds — is encapsulated in this relation between the Hamiltonian and $\mathbf{L}^2$. Again, it's the form of the Hamiltonian which is the basis of our identification of the generator $\mathbf{L}/\hbar$ of circular transformations with angular momentum (in units of $\hbar$).

4.5.1 Symmetries and Conservation Laws

One of the immediate uses of the Heisenberg picture is to deduce a relation between symmetries and conserved quantities[29]. If it happens that an operator Q commutes with the Hamiltonian, then it also commutes with $\mathrm{e}^{-\mathrm{i}Ht/\hbar}$, so the Heisenberg picture operator

$$Q(t) = U^{-1}(t)QU(t) = U^{-1}U(t)Q = Q\,, \tag{4.91}$$

coinciding with the Schrödinger operator at all t. Infinitesimally, provided Q has no explicit time dependence, this is

$$\frac{\mathrm{d}}{\mathrm{d}t}Q(t) = \frac{\mathrm{i}}{\hbar}[H, Q(t)] = \frac{\mathrm{i}}{\hbar}U^{-1}(t)[H, Q]U(t) = 0\,. \tag{4.92}$$

[29] In Classical Dynamics course, you will see a corresponding relation in Hamilton's approach to classical physics

Operators that are time independent even in the Heisenberg picture are said to be conserved. We've shown that conserved operators are just that commute with the Hamiltonian.

Suppose we prepare a particle to be in an eigenstate of some conversed operator at time $t = 0$, so $Q\,|\psi(0)\rangle = q\,|\psi(0)\rangle$. Then at any later time we have

$$Q\,|\psi(t)\rangle = QU(t)\,|\psi(0)\rangle = U(t)Q\,|\psi(0)\rangle = qU(t)\,|\psi(0)\rangle = q\,|\psi(t)\rangle \tag{4.93}$$

so our particle remains in an eigenstate of Q at all subsequent times (provided the state evolves according to the time-dependent Schrödinger equation). For this reason, it's usually sensible to expand our states in a basis of eigenstates of a maximal set of conserved operators, rather than a maximal commuting set of any old operators. We label the states in this basis by their corresponding eigenvalues, because the same labelling will remain valid at subsequent times. For example, provided H has no explicit time dependence[30] $[H, H] = 0$ trivially, so the Hamiltonian itself is always conserved, and it is often useful to work in a basis of energy eigenstates.

By far the most important source of conserved quantities is symmetries of the Hamiltonian. These are transformations that leave the Hamiltonian invariant. We've seen that a generic unitary operator $U(\theta) = \mathrm{e}^{\mathrm{i}\theta T}$ representing some transformation of space on H acts on operators as in equation (4.14). In particular, its action on the Hamiltonian is

$$H \longmapsto U^{-1}(\theta)HU(\theta)\,. \tag{4.94}$$

If the Hamiltonian is invariant under this transformation — i.e. if the transformation is a symmetry — then

$$U^{-1}(\theta)HU(\theta) = H\,, \tag{4.95}$$

or equivalently

$$[T, H] = 0 \tag{4.96}$$

for an infinitesimal symmetry transformation, where T is the Hermitian generator. This is exactly the same condition (4.92) we had for the generator T to be conserved, so symmetries of the Hamiltonian correspond to conserved quantities[31]. For example, if the Hamiltonian is translationally invariant then $[\mathbf{P}, H] = 0$, so for such Hamiltonians, momentum will be conserved. Note that it's not necessary for the Hamiltonian to be independent of each particle's position $\mathbf{x}_a$ if H is to be translationally invariant, as long as any potential term $V(\mathbf{x}_a, \mathbf{x}_b)$ depends only on the relative positions. Similarly, if the Hamiltonian is rotationally invariant then $[\mathbf{J}, H] = 0$ so angular momentum will be conserved. In this case, J_z, J^2 and H form a set of mutually commuting operators, so we can expand a general state in a basis $\{|n, \ell, m, \ldots\rangle\}$ where the labels (n, ℓ, m) refer to the eigenvalues of H, J^2 and J_z. We've left open the possibility that our states have further conserved properties, indexed by further labels. These will often contain information about the "internal" state of our system.

4.6 Parity

Not all transformations are continuous. In physics, the most prominent example of a discrete transformation is the *parity transformation* $\mathcal{P}$, acting on $\mathbb{R}^3$ as $\mathcal{P} : \mathbf{x} \mapsto -\mathbf{x}$. Since $\det(\mathcal{P}) = -1$, this is different from a rotation, so may have consequences that cannot be deduced by considering only rotations.

On Hilbert space, parity transformations will still be represented by a unitary operator, but this operator will no longer involve a parameter that may be taken infinitesimally small. Consequently

[30] Even if the Hamiltonian does contain explicit time dependence, for example because it describes the dynamics of a charged particle in a varying electric field, $[H(t), H(t)] = 0$ trivially. However, $[H(t'), H(t)]$ may not vanish as the form of the Hamiltonian itself changes.

[31] In the Classical Dynamics or Integrable Systems courses, you'll meet the corresponding statement in classical mechanics in called Nöther's theorem.

there's no generator associated with parity transformations. Calling the unitary parity operator Π, we have

$$\Pi^{-1}\mathbf{X}\Pi = \mathcal{P}\mathbf{X} = -\mathbf{X}$$

as its defining property. Multiplying through on the right by Π, we can also write this as

$$\{\Pi, \mathbf{X}\} \coloneqq \Pi\mathbf{X} + \mathbf{X}\Pi = 0\,, \tag{4.97}$$

where the bracket $\{A, B\}$ is the *anticommutator* of A and B. Translating through $\mathbf{a}$ then applying the parity operator $\mathcal{P} : \mathbb{R}^3 \to \mathbb{R}^3$ is the same as first applying $\mathcal{P}$, then translating through $-\mathbf{a}$, so

$$\Pi^{-1}U(\mathbf{a})\Pi = U(-\mathbf{a})\,, \tag{4.98}$$

so that the momentum operator also anticommutes with the parity operator: $\Pi^{-1}\mathbf{P}\Pi = -\mathbf{P}$. More generally, if $\mathbf{V}$ is any vector that transforms as a vector under rotations, we say $\mathbf{V}$ is a *vector operator* if also

$$\Pi^{-1}\mathbf{V}\Pi = -\mathbf{V}\,, \tag{4.99}$$

so that $\{\Pi, \mathbf{V}\} = 0$. If instead

$$\Pi^{-1}\mathbf{V}\Pi = +\mathbf{V}\,, \tag{4.100}$$

so equivalently $[\Pi, \mathbf{V}] = 0$, then $\mathbf{V}$ is a *pseudovector operator* (provided it transforms appropriately under rotations). The most prominent example of a pseudovector operator is the rotation generator $\mathbf{J}$: since the parity transformation $\mathcal{P}$ acts on $\mathbb{R}^3$ as $-\operatorname{Id}_{3\times 3}$, rotations obey $\mathsf{R}(\boldsymbol{\alpha})\mathcal{P} = \mathcal{P}\mathsf{R}(\boldsymbol{\alpha})$, or $\mathcal{P}^{-1}\mathsf{R}(\boldsymbol{\alpha})\mathcal{P} = \mathsf{R}(\boldsymbol{\alpha})$. This gives

$$\Pi^{-1}U(\boldsymbol{\alpha})\Pi = U(\boldsymbol{\alpha}) \tag{4.101}$$

for the parity and rotation operators in the Hilbert space, or $\Pi^{-1}\mathbf{J}\Pi = \mathbf{J}$ for the rotation generator. Likewise, the orbital angular momentum operator obeys

$$\Pi^{-1}\mathbf{L}\Pi = \Pi^{-1}(\mathbf{X}\times\mathbf{P})\Pi^{-1} = (\Pi^{-1}\mathbf{X}\Pi)\times(\Pi^{-1}\mathbf{P}\Pi) = (-\mathbf{X})\times(-\mathbf{P}) = \mathbf{L}\,, \tag{4.102}$$

so it is a pseudovector as in classical mechanics. It follows that the spin operator $\mathbf{S} = \mathbf{J} - \mathbf{L}$ is also a pseudovector.

Similarly, operators S that are invariant under rotations are scalar operators if in addition $\Pi S\Pi = S$ so that they are unchanged under parity, and are pseudoscalar operators if instead $\Pi S\Pi = S$. (Notice that the signs for scalars and pseudoscalars are the opposite way round to those for vectors and pseudovectors.) As usual, the parity of a system will be conserved if $[H, \Pi] = 0$. As a simple example, if our system is governed by the dynamical relation $H = \mathbf{P}^2/2m + V(\mathbf{X})$ where the potential is an even function, then

$$\begin{aligned}
\Pi^{-1}H\Pi &= \frac{1}{2m}(\Pi^{-1}\mathbf{P}\Pi)\cdot(\Pi^{-1}\mathbf{P}\Pi^{-1}) + \Pi^{-1}V(\mathbf{X})\Pi && (4.103)\\
&= \frac{\mathbf{P}^2}{2m} + V(\Pi^{-1}\mathbf{X}\Pi) && (4.104)\\
&= \frac{\mathbf{P}^2}{2m} + V(\mathbf{X}) = H\,, && (4.105)
\end{aligned}$$

so parity is conserved for such Hamiltonians.

Parity transformations form the group $G = \mathbb{Z}_2$, since carrying out a non-trivial parity transformation twice just brings us back to where we were. Thus $P_2 = 1_{\mathbb{R}^3}$ and applying our homomorphism shows that

$$\Pi^2 = 1_{\mathcal{H}} \tag{4.106}$$

for the parity operator on Hilbert space, too. Now suppose that $|\psi\rangle$ is an eigenstate of Π with eigenvalue η. Then[32]

$$|\psi\rangle = \Pi^2\,|\psi\rangle = \eta\Pi\,|\psi\rangle = \eta^2\,|\psi\rangle\ , \tag{4.107}$$

[32]This argument is adequate for our course, and leads to a correct conclusion, but it hides a number of subtleties. See e.g. Weinberg The Quantum Theory of Fields, vol. 1, section 4.7 for a fuller discussion.

so the spectrum of Π is just $\{+1, -1\}$.

If $|\mathbf{v}\rangle$ is an eigenstate of a vector operator $\mathbf{V}$, with $\mathbf{V}|\mathbf{v}\rangle = \mathbf{v}|\mathbf{v}\rangle$, then from 4.99, the parity reversed state $|\mathbf{v}'\rangle = \Pi|\mathbf{v}\rangle$ obeys

$$\mathbf{V}|\mathbf{v}'\rangle = \mathbf{V}\Pi|\mathbf{v}\rangle = -\Pi\mathbf{V}|\mathbf{v}\rangle = -\mathbf{v}\Pi|\mathbf{v}\rangle = -\mathbf{v}|\mathbf{v}'\rangle \tag{4.108}$$

so is also an eigenstate of $\mathbf{V}$, but with opposite sign eigenvalue. In particular, $\Pi|\mathbf{x}\rangle$ is an eigenstate of the position operator with eigenvalue $-\mathbf{x}$, so

$$\Pi|\mathbf{x}\rangle = c|-\mathbf{x}\rangle \tag{4.109}$$

for some constant $c \in \mathbb{C}$. This is not an eigenvalue equation, because the right hand side involves $|-\mathbf{x}\rangle$ rather than $|\mathbf{x}\rangle$. However, we can immediately identify c with η because applying Π a second time gives[33]

$$|\mathbf{x}\rangle = \Pi^2|\mathbf{x}\rangle = c^2|\mathbf{x}\rangle\ , \tag{4.110}$$

so again $c^2 = 1$ and $c = \pm 1$. More generally, given any state $|\psi\rangle$, the wavefunction of the parity transformed state $\Pi|\psi\rangle$ is

$$\langle\mathbf{x}|\Pi|\psi\rangle = \eta^{-1}\langle-\mathbf{x}|\psi\rangle = \eta\psi(-\mathbf{x})\,, \tag{4.111}$$

so, up to a sign, the wavefunction of the new state takes the same value at $\mathbf{x}$ as the original one did at $-\mathbf{x}$.

For example, suppose that for some particle, $|\mathbf{x}\rangle$ obey $\Pi|\mathbf{x}\rangle = |-\mathbf{x}\rangle$, with $\eta = +1$. Recalling from IB Quantum Mechanics that the spherical harmonics $Y_\ell^m(\mathbf{x})$ obey $Y_\ell^m(-\mathbf{x}) = (-1)^\ell Y_\ell^m(\mathbf{x})$, we see that if this particle is in the state $|n,\ell,m\rangle$ whose wavefunction $\langle\mathbf{x}|n,\ell,m\rangle = R_n(|\mathbf{x}|)Y_\ell^m(\mathbf{x})$ (where the radial part $R_n(r)$ of the wavefunction determines the energy level n), applying the parity operator gives

$$\langle\mathbf{x}|\Pi|n,\ell,m\rangle = \langle-\mathbf{x}|n,\ell,m\rangle = (-1)^\ell\langle\mathbf{x}|n,\ell,m\rangle \tag{4.112}$$

and consequently $\Pi|n,\ell,m\rangle = (-1)^\ell|n,\ell,m\rangle$.

Parity transformations act in a more complicated way on macroscopic objects, or systems with some internal structure. This is clear classically: the parity transformation of a book initially located at $\mathbf{x}$ is not simply a book located at $-\mathbf{x}$, but rather a mirror image (simultaneously left-right, up-down and front-back) of the book. As with the distinction between $\mathbf{J}$ and $\mathbf{L}$, for a macroscopic body we might hope to account for this by defining a separate parity operator for each constituent particle. However, it turns out that subatomic particles such as the photon or pion have their own "intrinsic" parities. Thus, whether a given Π-eigenstate has $\eta = +1$ or $\eta = -1$ depends not only on details of the spatial wavefunction, but also on the species of particle the state describes.

We'll explore intrinsic parity further in section 7.3.1, but as an example of its use, consider transitions between different energy levels of an atom. These are usually mediated by electromagnetic interactions: during such a transition, the atom's electrons either emit or absorb a photon whose energy $\hbar\omega$ is equal to the difference in energy E between the two levels involved in the transition. As we'll study in more detail later, when the corresponding wavelength is much larger than the typical size of the atom, the transition rate $\Gamma(i \to f)$ is given by

$$\Gamma(i \to f) = \frac{4(\Delta E)^3}{c^3\hbar^4}\left|\langle f|\mathbf{D}|i\rangle\right|^2\ , \tag{4.113}$$

where $|i\rangle$ and $|f\rangle$ are the initial and final states of the atom, and $\mathbf{D} = \sum e_a\mathbf{X}_a$ is the operator corresponding to the atom's electric dipole moment. (The sum runs over all the electrons in the atom.) Reversing the parity of space affects all the electrons equally, so the parity operator obeys $\Pi^{-1}\mathbf{X}_a\Pi = -\mathbf{X}_a$ for each $\mathbf{X}_a$. Thus $\Pi\mathbf{D}\Pi = -\mathbf{D}$.

[33] Note that it is important here that $\Pi|\mathbf{x}\rangle = c|-\mathbf{x}\rangle$ for all values of $\mathbf{x}$.

Now suppose that the initial and final states of the atom are eigenstates of Π, with eigenvalues η_i and η_f, respectively. Then[34]

$$\eta_i \eta_f \langle f|\mathbf{D}|i\rangle = \left\langle f\middle|\Pi^{-1}\mathbf{D}\Pi\middle|i\right\rangle = -\langle f|\mathbf{D}|i\rangle \tag{4.114}$$

so the amplitude for the transition will vanish unless the initial and final atomic states have opposite parity, $\eta_i\eta_f = 1$. In the most common case, just a single electron is involved in the transition and in this case equation (4.112) shows that $\eta_i\eta_f = (-1)^{l_i+l_f}$, where ℓ_i, ℓ_f are the total orbital angular momentum quantum numbers of the initial and final states of this electron. Thus, in any radiative atomic transition involving just a single electron, the orbital angular momentum quantum number ℓ must change by an odd number. If our initial and final states kept track of the photon as well as the atom, we'd be able to say overall parity is conserved in such transitions if we assign an intrinsic parity -1 to the photon that is emitted or absorbed in the transition. (The intrinsic parity of the photon is related to the fact that the electromagnetic vector potential $\mathbf{A}$, whose quantum version provides the description of the photon, indeed transforms as a vector, not a pseudovector.) In fact, $[H, \Pi] = 0$ holds in general for particles travelling in the presence of any type of electromagnetic interaction.

Having $[H, \Pi]$ be zero means that if at time $t = 0$ you set up a system to be a mirror image of another system, then their subsequent evolution will be identical in the sense that if you observe them at a later time, they will still be mirror images of one another. Hence, when $[H, \Pi] = 0$ it is impossible to tell whether a system is being observed directly or through a mirror. One of the major surprises of twentieth-century physics was an experiment by Wu *et al.* in 1957, which showed that in fact the Hamiltonian of our Universe has $[H, \Pi] \neq 0$ — you can see things in a mirror that are impossible in our World!

[34]Note that since the parity operator Π is unitary, its quantum numbers η behave multiplicatively, whilst those of a Hermitian generator of a continuous symmetry (such as P) behave additively.

5 Angular Momentum

In the previous chapter we obtained the fundamental commutation relations among the position, momentum and angular momentum operators, together with an understanding of how a dynamical relation $H = H(\mathbf{X}, \mathbf{P})$ allow us to understand how such quantities evolve in time. However, with the exception of the parity operator, we have not yet said anything about the spectrum — the set of possible eigenvalues — of these operators, nor have we been specific about the precise Hilbert space on which they act. Answering such questions is an important part of the subject of representation theory in mathematics; the study of how group actions can be realised on a given vector space. In physics, understanding this will provide essential information about the nature of our quantum system.

To a large extent, both questions are determined by the operator algebra itself. For example, taking the trace of the commutation relations $[X_i, P_j] = \mathrm{i}\hbar\delta_{ij} I_{\mathcal{H}}$ in any finite dimensional Hilbert space $\mathcal{H}$ gives

$$\dim(\mathcal{H})\delta_{ij} = -\frac{\mathrm{i}}{\hbar}\,\mathrm{tr}_{\mathcal{H}}(X_i P_j - P_j X_i) = 0 \tag{5.1}$$

for all i, j by the cyclic property of the trace. So there is no realisation of the position and translation operators on any non-trivial finite dimensional Hilbert space (and if $\dim(\mathcal{H}) = 0$ then $\mathbf{X}$ and $\mathbf{P}$ necessarily act trivially). The above argument fails in an infinite dimensional Hilbert space, where neither $\mathrm{tr}_{\mathcal{H}}(I_{\mathcal{H}})$ nor $\mathrm{tr}_{\mathcal{H}}(X_i P_j)$ is defined. Thus, if we wish to discuss the position and momentum of a quantum system, we are necessarily in the world of function space.[35]

On the other hand, taking the trace of the commutation relations

$$[J_i, J_j] = \mathrm{i}\hbar \sum_k \epsilon_{ijk} J_k \tag{5.2}$$

for the rotation generators just gives

$$\mathrm{i}\hbar\epsilon_{ijk}\,\mathrm{tr}_{\mathcal{H}}(J_k) = \mathrm{tr}_{\mathcal{H}}(J_i J_j - J_j J_i) = 0\,, \tag{5.3}$$

so it is possible to represent each component of $\mathbf{J}$ on a finite dimensional Hilbert space in terms of a traceless matrix[36]. This is often a useful thing to do, particularly when discussing the internal structure of a system.

In this chapter, we'll consider finite dimensional Hilbert spaces $\mathcal{H}_j$ whose elements transform among themselves under rotations. In mathematics, the $\mathcal{H}_j$ are known as *irreducible representations* of the rotation group, whilst in physics they're called multiplets of definite total angular momentum. This chapter will also substantiate the link between the rotation generators $\mathbf{J}$ and the physics of angular momentum, by examining a simple Hamiltonian for rotating diatomic molecule. We'll also flesh out the details of exactly how the orientation of a system is encoded in the amplitudes for it to be found in different eigenstates of appropriate angular momentum operators.

5.1 Angular Momentum Eigenstates

Let's begin by seeing how the algebra $[J_i, J_j] = \mathrm{i}\hbar\sum_k \epsilon_{ijk} J_k$ determines the spectrum of the angular momentum operators. Since no two components of $\mathbf{J}$ commute, we cannot find a complete set of simultaneous eigenstates of two components of $\mathbf{J}$. However, since $[\mathbf{J}, \mathbf{J}^2] = 0$ we can find a complete set of simultaneous eigenstates of any one component of $\mathbf{J}$ and $\mathbf{J}^2$. Without loss of generality, we orient

[35] Even here we should really be more careful. As mentioned in a previous footnote, $\mathbf{X}$ and $\mathbf{P}$ are not defined as linear operators on the whole of a Hilbert space such as $L^2(\mathbb{R}^3, \mathrm{d}^3x)$, because a state that is initially square-integrable may not remain so after the application of the unbounded operator $\mathbf{X}$ or $\mathbf{P}$.

[36] This traceless condition in $\mathcal{H}$ is closely related to but distinct from the fact that the generators of SO(3) can also be represented by antisymmetric (hence traceless) matrices acting on $\mathbb{R}^3$.

our coordinate system so that the chosen component of $\mathbf{J}$ is J_z. We let $|\beta, m\rangle$ denote a simultaneous eigenstate of J_z and $\mathbf{J}^2$, where

$$\mathbf{J}^2 |\beta, m\rangle = \beta\hbar^2 |\beta, m\rangle \qquad J_z |\beta, m\rangle = m\hbar |\beta, m\rangle \tag{5.4}$$

and the factors of $\hbar$ are for later convenience. We also choose all the states $|\beta, m\rangle$ to be properly normalised. Since they are eigenstates of Hermitian operators, they must then be orthonormal

$$\langle\beta', m'|\beta, m\rangle = \delta_{\beta\beta'}\delta_{mm'} \,. \tag{5.5}$$

Finally, since we are looking for the simplest possible way to realize our commutation relations, we'll assume that the states $\{|\beta, m\rangle\}$ are non-degenerate; that is, we want a Hilbert space where $|\beta, m\rangle$ is the unique eigenstate of $\mathbf{J}^2$ and J_z with the given eigenvalues. We'll comment more on this assumption in the next chapter.

We now define

$$J_\pm = J_x \pm \mathrm{i}J_y \,, \tag{5.6}$$

which obey $J_\pm^\dagger = J_\mp$ and so they are each other's adjoint. As they are built from linear combinations of the components of $\mathbf{J}$, clearly $J_\pm$ each commute with $\mathbf{J}^2$, while their commutation relations with J_z are

$$[J_z, J_\pm] = [J_z, J_x] \pm \mathrm{i}[J_z, J_y] = \mathrm{i}\hbar(J_y \mp iJ_x) = \pm\hbar J_\pm \,. \tag{5.7}$$

We learn that

$$\mathbf{J}^2(J_\pm |\beta, m\rangle) = J_\pm \mathbf{J}^2 |\beta, m\rangle = \hbar^2\beta J_\pm |\beta, m\rangle \tag{5.8}$$

and

$$J_z(J_\pm |\beta, m\rangle) = ([J_z, J_\pm] + J_\pm J_z) |\beta, m\rangle = (m \pm 1)\hbar(J_\pm |\beta, m\rangle) \,. \tag{5.9}$$

These show that the new states $J_\pm |\beta, m\rangle$, if not zero, are still eigenstates of both $\mathbf{J}^2$ and J_z, with the same eigenvalues $\beta\hbar^2$ for $\mathbf{J}^2$. However, their J_z eigenvalue is shifted up or down (respectively for $J_+ |\beta, m\rangle$ and $J_- |\beta, m\rangle$) by one unit of $\hbar$. We see that the role of $J_\pm$ in angular momentum is similar to the role of A and $A^\dagger$ for energy of the harmonic oscillator. Again, J_+ and J_- are often called the *raising* and *lowering operators*. Their role is to rotate our system, aligning more or less of its total angular momentum along the z-axis without changing the total angular momentum available.

Just as for the harmonic oscillator, examining the algebra of our raising and lowering operators has told us the separation between angular momentum eigenstates, given a starting point $|\beta, m\rangle$. To fix our initial states, we must examine the norm. By assumption, $|\beta, m\rangle$ itself is correctly normalized and we compute

$$\begin{aligned} \|J_+ |\beta, m\rangle\|^2 &= \langle\beta, m|J_- J_+|\beta, m\rangle = \langle\beta, m|(J_x - \mathrm{i}J_y)(J_x + \mathrm{i}J_y)|\beta, m\rangle \\ &= \big\langle\beta, m\big|\mathbf{J}^2 - J_z^2 - \hbar J_z\big|\beta, m\big\rangle = \hbar^2(\beta - m(m+1)) \,. \end{aligned} \tag{5.10}$$

But since this is a norm, whatever state $|\beta, m\rangle$ we started we must have

$$\beta - m(m+1) \geq 0 \,, \tag{5.11}$$

with equality iff $J_+ |\beta, m\rangle = 0$ is the trivial state. This shows that it cannot be possible to keep applying J_+, repeatedly raising J_z eigenvalue m while leaving β. There must be some maximum value of m — let's call it j — for which $J_+ |\beta, j\rangle = 0$. This can only be the case if β obeys

$$\beta = j(j+1) \tag{5.12}$$

and so is fixed in terms of the maximum allowed value of m.

Similarly for a generic m value we find

$$\|J_- |\beta, m\rangle\|^2 = \hbar^2(\beta - m(m-1)) \tag{5.13}$$

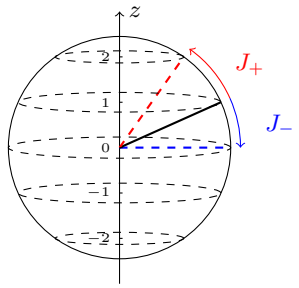


Figure 6: **J** eigenvectors for $j = 2$ states. Only the magnitude and the projection along the z axis is definite so each state is represented by a circle on a sphere. The raising and lowering operators makes the **J** more/less aligned with the z axis.

so it also cannot be possible to keep lowering the J_z eigenvalue whilst remaining in the Hilbert space. There must be some minimum value of m, say j', for which $J_- |\beta, j'\rangle = 0$ and this can only be the case if

$$\beta = j'(j' - 1)\,. \tag{5.14}$$

Applying $J_\pm$ does not change the $\mathbf{J}^2$ eigenvalue β, so these two values of β must agree. Comparing them, we obtain a quadratic equation for j' that we solve in terms of j, finding $j' = j + 1$ or $j' = -j$. By definition j' can't be greater than j so we must have $j' = -j$. The eigenvalue $\beta\hbar^2 = j(j+1)\hbar^2$ is determined by j, so we henceforth label our states as $|j, m\rangle$; this labelling is simple less cluttered than $|j(j+1), m\rangle$.

Finally, we note that since applying J_- repeatedly will take us from the highest state $|j, j\rangle$ through $|j, j-1\rangle, \ldots$ down to $|j, -j\rangle$, so it must be that $2j$ is a non-negative integer. In other words, the possible values for j are

$$j \in \left\{0, \frac{1}{2}, 1, \frac{3}{2}, \ldots\right\} = \frac{1}{2}\mathbb{N}_0\,. \tag{5.15}$$

Once the total angular momentum quantum number j is fixed, we have

$$m \in \{-j, -j+1, \ldots, j-1, j\}\,. \tag{5.16}$$

Thus there is in total of $2j+1$ states in any given multiplet, and $\mathcal{H}_j$ is the Hilbert space of dimension $2j+1$ spanned by the $\{|j, m\rangle\}$ for fixed j. We can move between states in $\mathcal{H}_j$ using raising and lowering operators $J_\pm$ which obey

$$J_+ |j, m\rangle = \hbar\sqrt{j(j+1) - m(m+1)}\,|j, m+1\rangle \tag{5.17}$$

$$J_- |j, m\rangle = \hbar\sqrt{j(j+1) - m(m-1)}\,|j, m-1\rangle\,. \tag{5.18}$$

These operators only change the J_z eigenvalue, not the $\mathbf{J}^2$ one. They just realign a given system, placing more (J_+) or less (J_-) of its angular momentum along the z-axis.

5.2 Rotations and Orientations

Note that, since J_x and J_y can be written in terms of the raising and lowering operators $J_\pm$, rotations around an arbitrary axis transform states in a given $\mathcal{H}_j$ into other states in the same $\mathcal{H}_j$. Since $J_x = (J_+ + J_-)/2$, we see that when $j > 0$, the state $|j, m\rangle$ is never an eigenstate of J_x. Hence, for $j > 0$, we can never be certain of the outcome of measurements of both J_x and J_z. This argument applies equally to J_y, so it's impossible to be certain of the outcome of measurements of more than one component of **J**. If $j = 0$, every component of **J** yields zero, but a null vector in $\mathbb{R}^3$ has no direction. This situation contrasts with the momentum vector **P** which can have well-defined direction since all its components commute with one another.

In the mathematical literature, states with $m = j$ are known as *highest weight states.* They play a key role in the representation theory of any group G, because once we know the highest weight state the rest of the multiplet can be constructed by applying lowering operators such as J_-. Physically, a highest weight state is one in which the body's angular momentum is most nearly aligned along the z-axis. In this state, the ratio of the squared angular momentum that lies in the xy-plane to that parallel to the z-axis is

$$\frac{\langle j,j|J_x^2 + J_y^2|j,j\rangle}{\langle j,j|J_z^2|j,j\rangle} = \frac{\langle j,j|\mathbf{J}^2 - J_z^2|j,j\rangle}{\hbar^2 j^2} = \frac{1}{j}\,. \tag{5.19}$$

As $J_x = (J_+ + J_-)/2$ we have $\langle j,j|J_x|j,j\rangle = 0$ and similarly $\langle j,j|J_y|j,j\rangle = 0$. Thus we have no information about the direction in the xy-plane any angular momentum points.

Macroscopic bodies, for which $j \gg 1$, have only a tiny fraction of their total angular momentum in the xy-plane when they're in state $|j,j\rangle$, so the uncertain direction of this component of angular momentum does not lead to significant uncertainty in the total direction of the angular momentum vector. By contrast, when $j = 1/2$, even in state$|1/2, 1/2\rangle$ there's twice as much angular momentum associated with the xy-plane as with the z-axis and the uncertain direction of this planar component makes it impossible to say anything more specific than that the angular momentum vector lies somewhere in the northern, rather than southern, hemisphere. Even when $j = 1$ and $m = 1$, the state $|1,1\rangle$ has as much angular momentum along the xy-plane as it has parallel to the z-axis, so the direction of the angular momentum vector is still very uncertain. This is no surprise: since $\dim(\mathcal{H}_j) = 2j + 1$, when $j = 1/2$ there are only two independent states the orientation of our body can take. With such a limited Hilbert space it's no wonder we only have a fuzzy notion of where the body's angular momentum lies.

5.2.1 Rotations of Diatomic Molecules

Our deduction of the possible eigenvalues and eigenstates of $\mathbf{J}^2$ and J_z came from considering rotations: the states $|j,m\rangle$ simply enable us to describe what happens when an object is rotated around some axis, as we'll understand in more detail and with many examples later in the chapter. In particular, it's important to understand that *a priori* these states have nothing to do with the energy levels of any given particle. As always, there's no way to tell how rotating an object may or may not change its energy until we specify a form for the Hamiltonian. In this section, we'll choose a simple form of dynamical relation $H = H(\mathbf{J})$ that will enable us to understand an important part of the dynamics of a diatomic molecule.

For some purposes, a diatomic molecules such as CO molecule can be considered to consist of two point masses, the nuclei of the oxygen and carbon atoms, joined by a "light rod" provided by the electrons. Following the analogous formula in classical mechanics, we model the dynamics of this molecule by the Hamiltonian

$$H = \frac{1}{2}\left(\frac{J_x^2}{I_x} + \frac{J_y^2}{I_y} + \frac{J_z^2}{I_z}\right)\,, \tag{5.20}$$

where I_x is the moment of inertia along the x-axis, and similarly for I_y and I_z. Though motivated by classical mechanics, the best justification for this form of Hamiltonian is ultimately that it agrees with detailed experiments.

In our axisymmetric case, we choose coordinates whose origin is at the centre of mass of the molecule (somewhere between the C and O atoms). Then $I_x = I_y = I$, say, whilst I_z is different. In fact, since the centre of mass of the C and O atoms lies along the z-axis, the molecule's moment of inertia around this axis is negligible, $I_z \ll I$. We can thus rewrite our Hamiltonian as

$$H = \frac{1}{2}\left[\frac{\mathbf{J}^2}{I} + J_z^2\left(\frac{1}{I_z} - \frac{1}{I}\right)\right]\,. \tag{5.21}$$

The virtue of expressing H in terms of $\mathbf{J}^2$ and J_z is that our knowledge of the spectrum of these operators immediately allows us to write down the spectrum of this Hamiltonian: the state $|j,m\rangle$ is

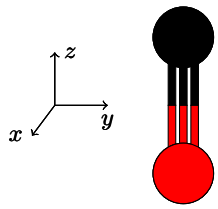


Figure 7: A simple model of carbon monoxide. By symmetry, $I_x = I_y$ whilst I_z is much smaller.

an energy eigenstate with

$$E_{jm} = \frac{\hbar^2}{2}\left[\frac{j(j+1)}{I} + m^2\left(\frac{1}{I_z} - \frac{1}{I}\right)\right], \tag{5.22}$$

with $|m| \leq j$. Since $I_z \ll I$ for our diatomic molecule, the coefficient of m^2 is very much greater than that of $j(j+1)$, so states with $m \neq 0$ will only occur very far above the ground state. Consequently, the low lying states have energies of the form

$$E_j = \frac{\hbar^2}{2I} j(j+1) \tag{5.23}$$

for some j. As we saw in the previous section, only discrete values of j are allowed, so the energy levels are quantized. One can excite a CO molecule, causing it to rotate faster, only by supplying a definite amount of energy. Similarly, for the molecule to relax down to a state of lower angular momentum, it must emit a quantized lump of energy.

Carbon monoxide is a significantly dipolar molecule. The carbon atom has a smaller share of the binding electrons than does the oxygen, with the result that it is positively charged while the oxygen atom gains negative charge. In Maxwell's theory of electrodynamics, a rotating electric dipole is expected to emit electromagnetic radiation. Because we're in the quantum regime, this radiation emerges as photons which, as we'll see later in the chapter, can add or carry away only one unit $\hbar$ of angular momentum. Thus the energies of the photons that can be emitted or absorbed by a rotating dipolar molecule are

$$E_\gamma = \pm(E_j - E_{j-1}) = \pm\frac{j\hbar^2}{I}. \tag{5.24}$$

Using the relation $E = \hbar\omega$, the angular frequencies in the rotation spectrum of the molecule are

$$\omega_j = \frac{j\hbar}{I}. \tag{5.25}$$

In the case of ^{12}CO, $\hbar/2\pi I$ evaluates to a frequency $\nu \sim 115.271$ GHz and spectral lines occur at multiples of this frequency. In the classical limit of large j, the molecule's total angular momentum $|\mathbf{J}| \approx j\hbar$. This is related to the angular frequency Ω at which the molecule rotates by $|\mathbf{J}| = I\Omega$. Comparing to (5.25) we see that in the classical limit, $\omega_{j\gg 1} = \Omega$, so the frequency of the emitted radiation is just the frequency at which the molecule rotates.

Measurements of the radiation at 115.271 GHz provide one of the two most important probes of interstellar gas[37]. In denser, cooler regions, hydrogen atoms combine to form H_2 molecules which are centrosymmetric and do not have an electric dipole moment when they rotate. Consequently, these molecules, which together with similarly uncommunicative Helium make up the vast majority of cold interstellar gas, lack readily observable spectral lines. Astronomers are thus obliged to study the interstellar medium through the rotation spectrum of the few parts in a million of CO it contains.

[37]The other key probe is the hyperfine line of atomic hydrogen that will be discussed in section 8.

5.3 Spin

We saw that the orbital angular momentum operator $\mathbf{L}$ and spin operator $\mathbf{S}$ each obey an identical algebra to that of the rotation generators

$$[L_i, L_j] = \mathrm{i}\hbar \sum_k \epsilon_{ijk} L_k \qquad [L_i, \mathbf{L}^2] = 0 \tag{5.26}$$

$$[S_i, S_j] = \mathrm{i}\hbar \sum_k \epsilon_{ijk} S_k \qquad [S_i, \mathbf{S}^2] = 0 \tag{5.27}$$

as well as $[S_i, L_j] = 0$. Since the algebra of the $\mathbf{J}$'s was all we used to deduce their spectra, it follows immediately that $(\mathbf{L}^2, L_z)$ and $(\mathbf{S}^2, S_z)$ have the same possible spectra as we found for $(\mathbf{J}^2, J_z)$. It is conventional to label eigenstates of $(\mathbf{L}^2, L_z)$ as $|\ell, m\rangle$ and those of $(\mathbf{S}^2, S_z)$ as $|s, \sigma\rangle$, where ℓ and s correspond to the eigenvalues of $\mathbf{L}^2$ and $\mathbf{S}^2$, respectively, and m and σ label the eigenvalues of L_z and S_z. (Unfortunately, it's traditional to label the eigenvalue of both J_z and L_z by the same letter m. Which is meant is usually clear from the context.)

5.3.1 Large Rotations

Our claim that the possible values for (j, m) are

$$j \in \left\{0, \frac{1}{2}, 1, \frac{3}{2}, \dots\right\} \qquad \text{and then} \qquad m \in \{-j, \dots, j-1, j\} \tag{5.28}$$

was based on examining the algebra $[J_i, J_j] = \mathrm{i}\hbar \sum_k \epsilon_{ijk} J_k$ and likewise for (ℓ, m_ℓ) or (s, σ). In turn, this algebra originally came from considering the behaviour of objects under very small rotations. We now check whether these spectra are also compatible with large rotations.

Suppose we rotate the state $|j, m\rangle$ through an amount α around the z-axis. Then

$$|j, m\rangle \longrightarrow U(\alpha\hat{\mathbf{z}})\,|j, m\rangle = \mathrm{e}^{\mathrm{i}\alpha J_z/\hbar}\,|j, m\rangle = \mathrm{e}^{-\mathrm{i}m\alpha}\,|j, m\rangle \tag{5.29}$$

since it is an eigenstate of J_z. Rotations through 2π around any axis return us to our starting point, so are equivalent to no rotation. Thus, for $U(\boldsymbol{\alpha})$ to be a homomorphism from SO(3) to the group of unitary operators on $\mathcal{H}$, we must have $U(2\pi\hat{\boldsymbol{\alpha}}) = 1_{\mathcal{H}}$. In particular, when $\hat{\boldsymbol{\alpha}} = \hat{\mathbf{k}}$, from above we must have $\mathrm{e}^{-2\pi\mathrm{i}m} = 1$. This is true if m is an integer, but not if it is an odd-half-integer. Since m is an integer or (odd) half-integer iff j is, we conclude that odd half-integer values of j are in fact not compatible with the behaviour of objects under large rotations: they are ruled out by our basic requirement that $U(\boldsymbol{\alpha})$ represents the action of SO(3) on $\mathcal{H}$.

The fact that global properties of rotations rule out some of the eigenvalues allowed by the rotation algebra is an example of a phenomenon familiar from IB Methods. The behaviour of function in the neighbourhood of a point may be governed by some differential equation. The associated differential operator (if it's linear) typically has a large spectrum, which is cut down by boundary conditions or periodicity conditions. For example, on any open set $U \subset \mathbb{R}$ the linear operator $-\mathrm{i}\mathrm{d}/\mathrm{d}x$ has eigenfunctions $\mathrm{e}^{\mathrm{i}kx}$ for any $k \in \mathbb{C}$. However, if we know that globally $\psi : S^1 \to \mathbb{C}$ so is periodic, then k must be quantized in units of the circle's radius. The only difference in our case is that the non-trivial algebra of the rotation generators already restricted the possible eigenvalues of $\mathbf{J}^2$ and J_z to be quantized in units of $\hbar$. Global properties of the rotation group still remove some of the eigenvalues that were allowed locally. More succinctly, states where $j \in \mathbb{N}_0 + \frac{1}{2}$ do form representations of the rotation algebra $\mathfrak{so}(3)$, but not of the rotation group SO(3).

In fact, we'll see that our arguments above have been rather to hasty.

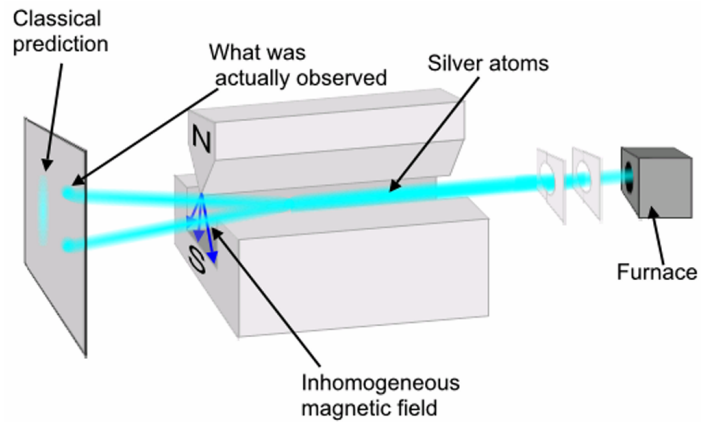


Figure 8: A schematic picture of the Stern–Gerlach experiment, adapted from the Astronomy Cast site.

5.3.2 The Stern–Gerlach Experiment

Quantization of angular momentum in units of $\hbar$ was originally proposed by Bohr and Sommerfeld as a means by which the stability of atomic orbits could be understood; we've now seen how this quantization arises automatically in the full mathematical framework of quantum mechanics. However, back in 1922 it still seemed very mysterious, so Stern and Gerlach designed and conducted experiments to check whether angular momentum is really quantized in Nature.

In the Stern–Gerlach experiments, a beam of uncharged atoms of the same type is passed through a region of slowly varying magnetic field. If the atoms have mass M, the Hamiltonian for this process is[38]

$$H = \frac{\mathbf{P}^2}{2M} - \boldsymbol{\mu} \cdot \mathbf{B}\,, \tag{5.30}$$

where $\mathbf{B}$ is the applied magnetic field, and $\boldsymbol{\mu}$ is the *magnetic dipole moment* of the atom. Ultimately, the reason the atom can be treated as a magnetic dipole is because of detailed properties of the distribution of its electrons. It would take us too far afield to explain this precisely, but the dipole moment arises because the atom has an orientation: a perfectly spherically symmetric atom cannot have any dipole moment as there is no preferred direction for the "north" or "south" poles. Thus the dipole moment is only non-zero for atoms that transform non-trivially under $\mathbf{S}$. If the atom is in the spin s representation, we can write $\boldsymbol{\mu} = (\mu/\hbar s)\mathbf{S}$.

Choosing our coordinates so that the z-axis points along the direction of the magnetic field, the Hamiltonian becomes

$$H = \frac{\mathbf{P}^2}{2M} - \frac{\mu}{\hbar s} B S_z\,, \tag{5.31}$$

where $B = |\mathbf{B}|$. The equations of motion obtained using the Heisenberg picture tell us that the expectation values of position and momentum in any state obey

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \mathbf{X}\rangle = \frac{\langle \mathbf{P}\rangle}{M} \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}t}\langle \mathbf{P}\rangle = \frac{\mu}{\hbar s}\langle (\boldsymbol{\nabla} B) S_z\rangle\,. \tag{5.32}$$

The second of these equations is analogous to the classical $\mathbf{F} = -\boldsymbol{\nabla} V$.

We see that the force experienced by any given atom depends on its value of S_z. For a spin s particle, these are $\hbar\{-s, -s+1, \ldots, s\}$ and in particular can be of either sign. If an atom is in the state $|s, s\rangle$ with all its spin aligned along the direction of $\mathbf{B}$, then it will experience a force pushing

[38] The potential energy resulting from coupling a magnetic dipole to an applied magnetic field is similar to the (perhaps more familiar) energy $\mathbf{p} \cdot \mathbf{E}$ of an electric dipole $\mathbf{p}$ in an applied electric field $\mathbf{E}$. Note that the atoms carry no net electric charge, so there's no Lorentz force.

it in the direction of increasing magnetic field. On the other hand, those atoms in the state $|s, -s\rangle$ will be pushed in the direction of decreasing B, while atoms with $\sigma = 0$ (when s is an integer) are unaffected. In total, when an initial beam of atoms in which the spins are randomly aligned passes through the region of magnetic field, it will be split into $2s + 1$ different trajectories. Thus, if the atoms are perfectly spherical, they will pass through unaffected, whereas if they have $s = 1$ the beam will be split into three, corresponding to the three possible eigenvalues $\{-\hbar, 0, \hbar\}$, if they have $s = 2$ the beam will split into five, and so on. Finally, if the atoms have very large spin (and so a very definite orientation in space) the beam will be split into so many paths that we can no longer distinguish the individual paths, seeing instead a broad smear. This is the same result we'd expect to find if angular momentum is not quantised, where the amount by which an atom is deflected would depend smoothly on its orientation with respect to $\mathbf{B}$.

Stern and Gerlach's original experiment in fact used silver atoms. Their magnetic was controlled by a dial. Initially, as $B = 0$ the atoms all passed straight through. As they turned up the magnetic field, the beam split into **two** separate paths. As $\mathbf{B}$ was further increased, the separation between these two beams increased, but no further splitting was observed. Thus, not only is angular momentum quantized as Bohr and Sommerfeld had predicted, but silver atoms have spin $s = \frac{1}{2}$!

5.3.3 Spinors and Projective Representation (Non-examinable)

The Stern–Gerlach experiment shows that, despite our arguments, half integer values of s actually arise in Nature. In fact, the chemical properties of the elements and the structure of the periodic table, together with properties of materials such as metals, conductors and insulators depends crucially on particles having half-integer values of spin. These values are in conflict with our current mathematical formalism, so we must have made a mistake, imposing too strong a condition that ruled out the possibility of half-integer spins.

The error lay in our claim that we needed to represent the action a transformation group G on Hilbert space $\mathcal{H}$, rather than just on projective Hilbert space $\mathbb{P}\mathcal{H}$. (Recall that states which differ only by an overall constant — which does not need to have modulus 1 provided we use the general form of the Born rule — yield the same results in all experiments. Thus physical systems are represented by states in projective Hilbert space.) Projectively, it's enough to require

$$U(g_2) \circ U(g_1) = \mathrm{e}^{\mathrm{i}\phi(g_2,g_1)} U(g_2 \cdot g_1) \tag{5.33}$$

rather than the homomorphism condition (4.7), where $\phi(g_1, g_2)$ is a real phase. This phase does not affect which ray in $\mathcal{H}$ the state lies in, so leaves the physics unchanged. The operator algebra is associative, so we must have

$$U(g_3) \circ (U(g_2) \circ U(g_1)) = (U(g_3) \circ U(g_2)) \circ U(g_1)\,, \tag{5.34}$$

which implies the phase obey

$$\phi(g_2, g_1) + \phi(g_3, g_2 \cdot g_1) = \phi(g_3, g_2) + \phi(g_3 \cdot g_2, g_1)\,. \tag{5.35}$$

Phase factors $\mathrm{e}^{\mathrm{i}\phi(g_2,g_1)}$ obeying this condition are known as *cocycles* on the group G. One possible solution of (5.35) is to take

$$\phi(g_2, g_1) = \beta(g_2 \cdot g_1) - \beta(g_2) - \beta(g_1) \tag{5.36}$$

for arbitrary smooth function $\beta : G \to \mathbb{R}$. This solution is "trivial", because if $\phi(g_2, g_1)$ takes this form, then we can define a new unitary transformation operator $U'(g) = \mathrm{e}^{\mathrm{i}\beta(g)} U(g)$ which obeys our original condition (4.7). By agreeing to work with the new operator, the phases never arise. The interesting question is whether there are other, non-trivial solutions to (5.35).

A theorem in group cohomology[39] states that it is possible for non-trivial projective representations

[39] Unfortunately, we won't prove this here. If you're interested, you can find a proof (given in the context of quantum mechanics) in either Weinberg's *The Quantum Theory of Fields*, vol. 1 (chapter 2, appendix B), or else in Hall's *Quantum Theory for Mathematicians*.

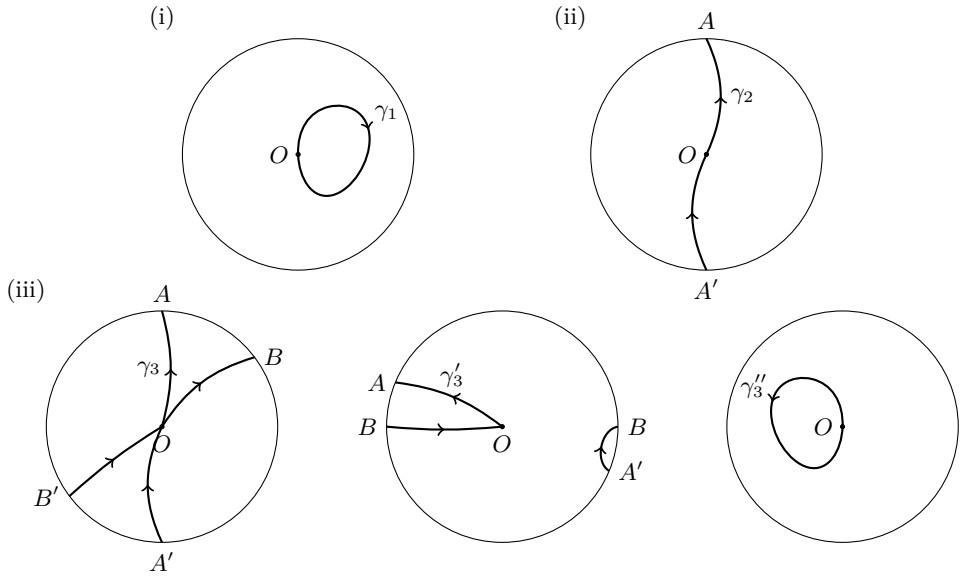


Figure 9: Paths in SO(3) (i) contractible (ii) a loop with angles increasing by 2π that is non-contractible (iii) a loop with angles increased by 4π but it is again contractible.

to arise for groups that are not simply connected[40]. In fact, this is the case for SO(3). The topology of the rotation group can be seen by viewing each rotation as parameterised by a vector $\boldsymbol{\alpha}$. If we allow the axis of rotation $\hat{\boldsymbol{\alpha}}$ to point in any direction, then rotations with $|\boldsymbol{\alpha}| \in (0, \pi)$ are uniquely specified. However, when rotating through π we get the same rotation whether rotating around $\hat{\boldsymbol{\alpha}}$ or $-\hat{\boldsymbol{\alpha}}$. Thus, we can picture the space of all rotations as a solid, three-dimensional ball $|\boldsymbol{\alpha}| \leq \pi$, but with antipodal points on the surface $|\boldsymbol{\alpha}| = \pi$ identified the same.

The description show that SO(3) contains smooth, closed paths, beginning and ending at the identity rotation (the origin of the 3-ball) that cannot be continuously shrunk to a point. For example, consider the loops in figure 9 which all start and end at the identity rotation, i.e. the centre of the sphere. Figure (i) shows a loop which is contractible; it can obviously be shrunk to a point. On the other hand, figure (ii) shows a loop for which the angle of rotation all starts at zero, smoothly increasing to π at the point A. At this point it reaches the boundary and reappears at the antipodal point A', before continuing to increase to 2π back at the identity. It should be intuitively clear that this loop cannot be shrunk to a point whilst keeping both its ends fixed at the identity.

Next, consider the loop in figure (iii), along which the angle of rotation increases from 0 to 4π, reaching the boundary of the ball twice, once at A, reappearing at A' and then again at B reappearing at B'. By moving the two antipodal pairs, the section of the path between A' and B can be pulled across the boundary, smoothly deforming the loop back to the situation as in figure (i). Thus, after two complete rotations the situation becomes simple once more. In general, the angle along any closed path must increase by an integer multiple of 2π. The resulting loop will be contractible if this integer is even, but non-contractible if it is odd.

The topological properties of paths in G become important when we consider the behaviour of $U(g)$ as the group element g varies around a loop L, beginning and ending at some fixed g_0. It is quite possible for $U(g)$ to change smoothly with g along L in such a way that the operators at each end of the loop do not coincide: they may differ by a phase

$$U(g_0) \xrightarrow{L} \alpha_L U(g_0)\,, \tag{5.37}$$

where the phase α_L depends on the loop. This is consistent with the projective homomorphism (5.33).

[40]There's one other possibility: that the Lie algebra $\mathfrak{g}$ contains central elements (vectors $e \in \mathfrak{g}$ s.t. $[e, \mathfrak{g}] = 0$) that cannot be removed by a redefinition of the generators. This case is not relevant for us.

To investigate the possibilities for α_L, let's specialize to the case in which the operators act on a Hilbert space of finite dimension N, so that each $U(g)$ can be regarded as an $N \times N$ unitary matrix. We can use up some of the freedom in (5.33) by requiring that each such matrix has unit determinant. This does not fix things completely, but the residual freedom is discrete:

$$\det U(g) = +1 \ \forall g \in G \implies \mathrm{e}^{\mathrm{i}N\phi(g_1,g_2)} = 1\,. \tag{5.38}$$

If G is simply connected, meaning that any closed loop is contractible, then the determinant condition implies $\alpha_L = 1$ for all loops L. This follows straightforwardly using continuity: α_L must be an Nth root of unity, but if it varies continuously as we vary our loop L and if all such loops are contractible to the single point g_0, then we must have the same matrix at the beginning and ending of each loop, so $\alpha_L = 1$.

If G is not simply connected, the same argument still shows that $\alpha_L = 1$ for any contractible L, but there may also be non-contractible loops with $\alpha_L \neq 1$. In this case, we can at least deduce that $\alpha_L = \alpha_{L'}$ whenever the loops L and L' are in the same homotopy class, meaning that they can be smoothly deformed into one another. Furthermore, if loops L and L' are traversed successively, then $U(g_0) \xleftarrow{L} \alpha_L U(g_0) \xleftarrow{L'} \alpha_L \alpha_{L'} U(g_0)$, so there are self-consistency constraints.

We can now give a unique definition of $U(g)$ for all g in a simply connected Lie group G. Recalling that $U(e) = \mathrm{Id}_{\mathcal{H}}$, we define $U(g_0)$ by choosing any path from the identity e to g_0 and demanding that $U(g)$ changes smoothly along this path. The values along the path are unique by the determinant and continuity conditions, but the end result $U(g_0)$ is also unique, because by traversing them in opposite directions, any two paths $e \to g_0$ can be combined to form a closed loop at g_0. This loop is contractible since our group is simply connected. With this definition, continuity also ensures that all cocycles are equal to one, so we have a genuine representation (not just a projective representation) of G.

Carrying out the same construction when G is not simply connected, we'll encounter paths from e to g_0 which cannot be smoothly deformed into one another. Thus, starting from $U(e) = \mathrm{Id}_{\mathcal{H}}$, in general we obtain different values for $U(g_0)$ depending on the path we take. In this way we're forced to consider multi-valued functions on G (just as when defining a continuous square root in the complex plane). The ambiguity, or multi-valuedness, in $U(g_0)$ can be resolved only by keeping track of the path we used to reach g_0, just as for the complex square root we must keep track of how many times we've encircled the origin to be sure which branch we're on. Such a multi-valued definition inevitably means that non-trivial cocyles appear.

As we saw above, the rotation group $G = \mathrm{SO}(3)$ is not simply connected, but there are just two topological classes of loops, depending on whether the net angle of rotation is an even or odd multiple of 2π. Any loop L in the first class is contractible and so has $\alpha_L = 1$. Any loop L in the second class is non-contractible, but if we traverse L twice it becomes contractible again. Thus $\alpha_{L'}^2 = 1$ and $L = \pm 1$. The finite-dimensional spaces $\mathcal{H}_j$ on which the rotation operators act are nothing but the multiplets of total angular momentum $j(j+1)\hbar^2$, with a basis $|j,m\rangle$ where

$$m \in \{-j, -j+1, \ldots, j-1, j\}\,. \tag{5.39}$$

Thus $\dim \mathcal{H}_j = 2j+1$. From the determinant condition $(\alpha_L)^{2j+1} = 1$ for any loop, but our topological considerations $\alpha_L = \pm 1$. If j is an integer then $2j+1$ is odd and it follows that $\alpha_L = 1$ for all loops, contractible or not. However, if j is an odd half-integer, then $2j+1$ is even and $\alpha_L = -1$ is possible for non-contractible loops. This is exactly the behaviour we found above: under the rotation operator $U(\alpha\hat{\mathbf{z}}) = \mathrm{e}^{-\mathrm{i}\alpha J_z/\hbar}$, a generic state in the j multiplet transforms as

$$|\psi\rangle = \sum_{m=-j}^{j} c_m\,|j,m\rangle \longmapsto U(\boldsymbol{\alpha})\,|\psi\rangle = \sum_{m=-j}^{j} c_m \mathrm{e}^{-\mathrm{i}\alpha m}\,|j,m\rangle\ . \tag{5.40}$$

Any state with j (and hence all m) odd-half-integral changes sign under a rotation through an odd multiple of 2π. The discussion of this section shows that the origin of this unexpected sign is the non-trivial topology of the rotation group — in our new terminology, we have a projective representation

of SO(3). These projective representations play an important role in physics, and are often called *spinors*.

We have, finally, obtained the correct mathematical framework in which to describe rotations and angular momentum in quantum mechanics. The history of the subject developed rather differently from the exposition we've given here. By the 1920's, careful studies of atomic spectroscopy had revealed that many spectral lines were in fact doubled, composed of two lines, very close in frequency. In 1924 (even before Schrödinger published his famous equation) Pauli proposed that this doubling indicated that, in addition to the quantum numbers n, ℓ, m labelling their energy levels, electrons possessed a further quantum number that took just two values. A year later, Uhlenbeck and Goudsmit suggested that this could be associated with some form of internal angular momentum. Their idea was initially treated with suspicion. If one wished to suppose the electron was a small, rotating sphere, then for it to have the needed angular momentum $\hbar/2$ and yet keep its radius small enough so that its finite size would not have been detected by experiment[41], the surface of the sphere would need to be travelling faster than light.

Nevertheless, the Stern–Gerlach experiment showed that particles could indeed have spin, contributing to the Hamiltonian in the presence of a magnetic field in just the same way as would a classical spinning magnetic dipole. Heisenberg, Jordan and C. G. Darwin then showed that the internal spin of the electron exactly accounted for the fine splitting of spectral lines that had puzzled Pauli, as we'll see in section 8.1.1.

We now understand that spin is an intrinsic property of fundamental particles: the Hilbert space of a fundamental particle is not simply $H_{\text{spat}} = L^2(\mathbb{R}^3, \mathrm{d}^3x)$ describing its spatial wavefunction, but rather a tensor product $H_{\text{spat}} \otimes H_s$. However we may excite, crash into, or generally interfere with the motion of an electron or W boson, for as long as they remain electrons and W bosons, their spin will always be $s = \frac{1}{2}$ and $s = 1$, respectively.

5.3.4 Spin Matrices

Since $\sigma \in \{-s, -s+1, \ldots, s-1, s\}$, states of definite total spin s can be described by a finite dimensional Hilbert space $H_s \cong \mathbb{C}^{2n+1}$. As always, once we pick a basis on $\mathcal{H}$ we can describe the action of linear operators such as $\mathbf{S}$ explicitly in terms of matrices. Let's now carry this out the first few spin representations, working in the basis $\{|s, \sigma\rangle\}$ of eigenstates of S_z.

The simplest case is $s = 0$, for which the only possible value of σ is also zero. This state obeys $\mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{S}/\hbar}\,|0,0\rangle = |0,0\rangle$ for any $|\alpha\rangle$. Hence a spin-zero object, like a perfect sphere, is completely unchanged under any rotation. In view of this, spin-zero particles are known as *scalar particles*. The discovery of the Higgs boson, announced on 4th July 2012 at CERN in Geneva, was the first time a fundamental scalar particle had been observed in Nature.

The next case is $s = \frac{1}{2}$. Electrons, neutrinos and quarks are fundamental particles of spin $\frac{1}{2}$, whilst protons, neutrons and the silver atoms used by Stern and Gerlach are examples of composite spin-half particles. When $s = \frac{1}{2}$, σ can only take one of the two values $\pm\frac{1}{2}$. For shorthand, we let $|\uparrow\rangle$ denote the state $|s, \sigma\rangle = \left|\frac{1}{2}, \frac{1}{2}\right\rangle$ and $|\downarrow\rangle$ denote $\left|\frac{1}{2}, -\frac{1}{2}\right\rangle$. A generic state of a spin-half system can thus be expanded as

$$|\psi\rangle = a\,|\uparrow\rangle + b\,|\downarrow\rangle \tag{5.41}$$

in this basis, where $a, b \in \mathbb{C}$ and $|a|^2 + |b|^2 = 1$ so that $|\psi\rangle$ is correctly normalised.

[41] To date, no experiment has ever detected a finite size, or any other non-trivial internal spatial structure in the electron.

In this basis, we can write the spin operators themselves as the 2×2 matrices

$$S_x = \begin{pmatrix} \langle\uparrow|S_x|\uparrow\rangle & \langle\uparrow|S_x|\downarrow\rangle \\ \langle\downarrow|S_x|\uparrow\rangle & \langle\downarrow|S_x|\downarrow\rangle \end{pmatrix} \tag{5.42}$$

$$S_y = \begin{pmatrix} \langle\uparrow|S_y|\uparrow\rangle & \langle\uparrow|S_y|\downarrow\rangle \\ \langle\downarrow|S_y|\uparrow\rangle & \langle\downarrow|S_y|\downarrow\rangle \end{pmatrix} \tag{5.43}$$

$$S_z = \begin{pmatrix} \langle\uparrow|S_z|\uparrow\rangle & \langle\uparrow|S_z|\downarrow\rangle \\ \langle\downarrow|S_z|\uparrow\rangle & \langle\downarrow|S_z|\downarrow\rangle \end{pmatrix} . \tag{5.44}$$

Since $|\uparrow\rangle$ and $|\downarrow\rangle$ are eigenstates of S_z, evaluating S_z in this basis is immediate. To also evaluate S_x and S_y, we note that $S_x = (S_+ + S_-)/2$ and $S_y = (S_+ - S_-)/2\mathrm{i}$ where $S_\pm$ are the spin raising and lowering operators defined just as for $J_\pm$. Using (5.17) and (5.18) with $j = s = \frac{1}{2}$ gives $S_+ |\downarrow\rangle = \hbar |\uparrow\rangle$ and $S_- |\uparrow\rangle = \hbar |\downarrow\rangle$. In this way, we obtain

$$S_x = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \quad S_y = \frac{\hbar}{2}\begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} , \quad S_z = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{5.45}$$

The coefficients for $\hbar/2$ here are known as the *Pauli matrices* and usually written as $(\sigma_x, \sigma_y, \sigma_z)$. Thus, for $s = \frac{1}{2}$, we can write $\mathbf{S} = \frac{\hbar}{2}\boldsymbol{\sigma}$.

Proceeding to spin-one, we find three possible values $\sigma \in \{-1, 0, 1\}$. Thus the spin-one Hilbert space is three (complex) dimensional, and when $s = 1$ we can represent each component of $\mathbf{S}$ by a 3×3 matrix. In this case, the spin raising and lowering operators $S_\pm$ act as

$$\begin{aligned} S_+ |1, -1\rangle &= \sqrt{2}\hbar |1, 0\rangle \qquad & S_+ |1, 0\rangle &= \sqrt{2}\hbar |1, +1\rangle \\ S_- |1, +1\rangle &= \sqrt{2}\hbar |1, 0\rangle \qquad & S_- |1, 0\rangle &= \sqrt{2}\hbar |1, -1\rangle . \end{aligned} \tag{5.46}$$

Using this result, one can check that

$$S_x = \frac{\hbar}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} , \quad S_y = \frac{\hbar}{\sqrt{2}}\begin{pmatrix} 0 & -\mathrm{i} & 0 \\ \mathrm{i} & 0 & -\mathrm{i} \\ 0 & \mathrm{i} & 0 \end{pmatrix} , \quad S_z = \hbar\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} . \tag{5.47}$$

Sadly, these matrices don't have any special name. Examples of fundamental particles with spin-one are the somewhat unimaginatively named W and Z bosons[42]. These are responsible for the weak interactions that, among other things, allows two hydrogen nuclei to fuse into Deuterium, powering nuclear fusion in the core of stars.

In just the same way, we can represent each component of the spin operator by a $(2s+1) \times (2s+1)$ matrix for any finite s. Since our basis is adapted to S_z, the matrix representation of S_z will be

$$S_z = \hbar \,\mathrm{diag}(s, s-1, \ldots, -s+1, -s) . \tag{5.48}$$

Matrices for S_x and S_y may be constructed using the spin raising and lowering operators as above. Since $S_\pm$ only change the z-component of the spin by 1 unit, these matrices are very sparse. One finds that they have non-zero entries only along the subleading diagonals, given by

$$\begin{aligned} (S_x)_{\sigma'\sigma} &= \frac{\hbar}{2}[\alpha(\sigma)\delta_{\sigma'-1,\sigma} + \alpha(\sigma-1)\delta_{\sigma'+1,\sigma}] \\ (S_y)_{\sigma'\sigma} &= \frac{\hbar}{2\mathrm{i}}[-\alpha(\sigma)\delta_{\sigma'-1,\sigma} + \alpha(\sigma-1)\delta_{\sigma'+1,\sigma}] , \end{aligned} \tag{5.49}$$

where $\alpha(\sigma) = \sqrt{(s-\sigma)(s+\sigma+1)}$. Note that, whatever the value of s, we always have three matrices (S_x, S_y, S_z) describing reorientations around the three independent axes in $\mathbb{R}^3$. Note also that each

[42] Although the photon carries one unit of $\hbar$ in intrinsic angular momentum, it has only two possible states corresponding to left- or right- circular polarized light. Because the photon is massless, one needs to consider representations of the Lorentz group, rather than the spatial rotation group, in order to describe it accurately. We won't do this in this course.

of these matrices is indeed traceless, as required by $\mathrm{i}\hbar\epsilon_{ijk}\,\mathrm{tr}_{\mathcal{H}}(S_k) = \mathrm{tr}_{\mathcal{H}}([S_i, S_j]) = 0$ as we said at the beginning of the chapter.

In elementary particle physics (and also in Tripos questions), one rarely encounters spins higher than 1. Nonetheless, it's interesting to consider the limits $s \gg 1$ of very high spins to see how our classical intuition emerges. For example, an electric motor that is roughly 1 cm in diameter and weighs about 10 g might spin at ~ 100 revolutions per second. Its intrinsic angular momentum is then 10^3 kg m^2 s$^{-1} \approx 10^3 1\hbar$. The classical world thus involves huge values of s.

Let's now show that, when $s \gg 1$, there is very little uncertainty in the direction of a system's spin. Suppose our system is in the state $|s, s\rangle$ so that its spin in maximally aligned with the z-axis. Let's compute $\langle s, s|\mathbf{n}\cdot\mathbf{S}|s, s\rangle$ where $\mathbf{n} = (\sin\theta, 0, \cos\theta)$. That is, we want to know how much spin we expect to measure along a direction in the xz-plane, inclined at angle θ to the z-axis. Classically this would just be $\hbar s\cos\theta$, the projection of $\hbar s$ onto this axis. Quantum mechanically, we have

$$\langle s, s|\mathbf{n}\cdot\mathbf{S}|s, s\rangle = \sin\theta\,\langle s, s|S_x|s, s\rangle + \cos\theta\,\langle s, s|S_z|s, s\rangle = \hbar s\cos\theta \tag{5.50}$$

using the fact that $S_x = (S_+ + S_-)/2$. Thus, on average the quantum result agrees with the classical intuition. This holds for any value of s, but now let's ask what the uncertainty in this result is. We compute

$$\begin{aligned}\left\langle s, s\middle|(\mathbf{n}\cdot\mathbf{S})^2\middle|s, s\right\rangle &= \sin^2\theta\,\left\langle s, s\middle|S_x^2\middle|s, s\right\rangle + \sin\theta\cos\theta\,\langle s, s|S_xS_z + S_zS_x|s, s\rangle + \cos^2\theta\,\left\langle s, s\middle|S_z^2\middle|s, s\right\rangle \\ &= \frac{1}{4}\sin^2\theta\,\langle S_+S_-|s, s|S_+S_-\rangle\,s, s + \hbar^2 s^2\cos^2\theta \\ &= \hbar^2\left(\frac{s}{2}\sin^2\theta + s^2\cos^2\theta\right)\,,\end{aligned} \tag{5.51}$$

with all other terms vanishing. Consequently

$$\sqrt{\langle(\mathbf{n}\cdot\mathbf{S})^2\rangle - \langle\mathbf{n}\cdot\mathbf{S}\rangle^2} = \sqrt{\frac{s}{2}}\hbar\,|\sin\theta| \tag{5.52}$$

and so, in the classical limit of large s, the relative uncertainty is small compared to $\langle\mathbf{n}\cdot\mathbf{S}\rangle$ (ratio $\sim\sqrt{\frac{1}{s}}$).

5.3.5 Paramagnetic Resonance and MRI Scanners

In the presence of an external magnetic field, a classical magnetic dipole $\boldsymbol{\mu}$ experiences a torque

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = \boldsymbol{\mu}\times\mathbf{B}\,. \tag{5.53}$$

The dipole will thus turn until it aligns itself along the direction of the field, minimizing its energy. This is familiar from a compass. However, suppose the dipole is already spinning around its centre of mass, such that $\boldsymbol{\mu} = \gamma\mathbf{L}$ where the constant γ is known as the gyromagnetic ratio. Then instead one finds that the torque causes the dipole to precess around $\mathbf{B}$.

As in our discussion of the Stern–Gerlach experiment, particles such as a proton or electron do have magnetic dipole moments $\boldsymbol{\mu} = (2\mu/\hbar)\mathbf{S}$ proportional to their spin. We'll see that in quantum mechanics, this spin does indeed precess around the direction of an applied magnetic field. This is the basis of MRI scanners, which have become an enormously important diagnostic tool for both chemistry and medicine.

We can understand the basic principles of an MRI machine by using our spin matrices. Unlike the beam of silver atoms in the Stern–Gerlach experiment, here the protons are not free to move, because they're held in place by the electromagnetic binding forces of a complex molecule, and this molecule is also held in a (roughly) fixed place in our cells. Thus we take the Hamiltonian to be

$$H = -\boldsymbol{\mu}\cdot\mathbf{B} = -\frac{2\mu B}{\hbar}S_z\,, \tag{5.54}$$

with no kinetic term. Again, we've chosen the direction of the magnetic field to define the $\hat{\mathbf{z}}$ axis. In particular, for a spin-$\frac{1}{2}$ particle such as a proton, the eigenvalues of this Hamiltonian are $E_\pm = \mp\mu B$, where $B = |\mathbf{B}|$.

Initially, we cannot expect the protons in our body to have their spins already aligned along $\mathbf{B}$. Suppose instead that some proton has its spin aligned along some axis $\mathbf{n} = (0, \sin\theta, \cos\theta)$ inclined at angle θ to $\mathbf{B}$. That is, we consider a proton in the state $|\theta_\uparrow\rangle$ defined by

$$\mathbf{n}\cdot\mathbf{S}\,|\theta_\uparrow\rangle = \frac{\hbar}{2}\,|\theta_\uparrow\rangle\ . \tag{5.55}$$

We can always choose to expand this state in terms of our basis $\{|\uparrow\rangle, |\downarrow\rangle\}$, representing states of definite spin along $\hat{\mathbf{z}}$, as

$$|\theta_\uparrow\rangle = a\,|\uparrow\rangle + b\,|\downarrow\rangle\ , \tag{5.56}$$

where $|a|^2 + |b|^2 = 1$. In terms of our Pauli matrices, the eigenvalue equation (5.55) becomes

$$\frac{\hbar}{2}\sin\theta\begin{pmatrix}0 & -\mathrm{i}\\ \mathrm{i} & 0\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix} + \frac{\hbar}{2}\cos\theta\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix}\cos\theta & -\mathrm{i}\sin\theta\\ \mathrm{i}\sin\theta & -\cos\theta\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix}a\\ b\end{pmatrix}\ . \tag{5.57}$$

Solving this eigenvalue problem and the normalisation condition yields $a = \cos\frac{\theta}{2}$ and $b = \mathrm{i}\sin\frac{\theta}{2}$ (up to a possible phase), so a proton whose spin is aligned along the $\mathbf{n}$ axis is in the state

$$|\theta_\uparrow\rangle = \cos\frac{\theta}{2}\,|\uparrow\rangle + \mathrm{i}\sin\frac{\theta}{2}\,|\downarrow\rangle\ . \tag{5.58}$$

Note that $|\theta_\uparrow\rangle$ has the expected behaviour at $\theta = 0$ and $\theta = \pi$, and it yields $-|\uparrow\rangle$ as θ is continuously increased to 2π. (It's straightforward to generalised this example to the case of a state with spin aligned along an arbitrary axis $\mathbf{n}$.)

Applying the time evolution operator $U(t) = \mathrm{e}^{-\mathrm{i}Ht/\hbar}$, we find that at time t the proton's state has evolved to

$$\begin{aligned}|\psi(t)\rangle = U(t)\,|\theta_\uparrow\rangle &= \cos\frac{\theta}{2}\mathrm{e}^{-\mathrm{i}Ht/\hbar}\,|\uparrow\rangle + \mathrm{i}\sin\frac{\theta}{2}\mathrm{e}^{-\mathrm{i}Ht/\hbar}\,|\downarrow\rangle\\ &= \cos\frac{\theta}{2}\mathrm{e}^{-\mathrm{i}\omega t/2}\,|\uparrow\rangle + \mathrm{i}\sin\frac{\theta}{2}\mathrm{e}^{+\mathrm{i}\omega t/2}\,|\downarrow\rangle\ ,\end{aligned} \tag{5.59}$$

where $\omega = 2\mu B/\hbar$ is the *Larmor frequency*. One can check that this state is an eigenstate of the spin operator aligned along the axis $\mathbf{n}(t) = (\sin\theta\sin\omega t, \sin\theta\cos\omega t, \cos\theta)$, so that at time t, the proton's spin is definitely aligned along $\mathbf{n}(t)$. Consequently, as time passes, the spin of any proton will precess around the direction of $\mathbf{B}$ with frequency that is independent of the angle of inclination i.e. independent of the proton's initial orientation (provided it was not pointing exactly along $\hat{\mathbf{B}}$ in the first place). This is exactly the same behaviour as we found classically.

When a material that contains chemically bound hydrogen atoms is immersed in a strong magnetic field, over time the protons will emit radiation (not accounted for by our above Hamiltonian) so as to sit in the ground state. Thus, eventually the precession will cease and most of the protons' spins will be aligned along the direction of $\mathbf{B}$.

Now suppose that, in addition to the static external $\mathbf{B}$ field, we apply a small additional magnetic field $\mathbf{b}\cos(\omega t)$ that varies at the same Larmor frequency $\omega = 2\mu B/\hbar$. The Hamiltonian felt by the protons will then be

$$H = -\boldsymbol{\mu}\cdot(\mathbf{B} + \mathbf{b}\cos\omega t) = -\frac{2\mu}{\hbar}\begin{pmatrix}B & b\cos\omega t\\ b\cos\omega t & -B\end{pmatrix} \tag{5.60}$$

if the new field is in the $\hat{\mathbf{x}}$-direction. The time dependent Schrödinger equation of this system is thus

$$\begin{pmatrix}\dot{a}\\ \dot{b}\end{pmatrix} = \frac{2\mu\mathrm{i}}{\hbar^2}\begin{pmatrix}B & b\cos\omega t\\ b\cos\omega t & -B\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix}\ , \tag{5.61}$$

varies with frequency $\omega = 2\mu B/\hbar$. This radiation has just the right energy to excite these protons into the state where their spin is aligned against $\mathbf{B}$. Consequently, such radiation is readily absorbed by the sample, whereas radiation at nearby frequencies is not. As we have seen, interference between the two states causes the spin to precess around the direction of $\mathbf{B}$, and this precessing magnetic moment couples resonantly to the applied radiation field.

MRI scanners are typically tuned to determine the concentration of a single type of atom, usually hydrogen. However, in a complex molecule, not every hydrogen nucleus (proton) feels the same magnetic field, because of additional contributions from the electrons that bind the atoms together in the molecule. For example, in methanol (CH_3OH) the magnetic field experienced by the proton that is attached to the oxygen atom differs from those experienced by the protons attached to the carbon atom. Now, the frequency of precession is proportional to the strength of the magnetic field at the location of the proton, so for any fixed strength of applied magnetic field, methanol has different resonant frequencies. Clues to the chemical structure of a substance can thus be obtained by determining the frequencies at which magnetic resonance occurs in a given imposed field. If we choose $\mathbf{B}$ to have a spatial gradient, then only a thin slice of our sample material will have ω tuned to its resonant frequency, so we excite transitions to higher energy levels only in this thin slice. Varying this field in an orthogonal direction as the nuclear spins decay back down to the ground state allows us to recover three dimensional images.

5.4 Orbital Angular Momentum

The topological considerations that allowed us to admit half-integer values of s do not apply to ℓ. To understand this, recall that $\mathbf{L}$ could be interpreted as the generator of circular translations — transformations that translate a state around a circle in space, without adjusting its orientation. Unlike the space $S^3/\mathbb{Z}_2$ of rotations, in $\mathbb{R}^3$ the space of such circular paths is contractible. Consequently, translation around a circular path always leaves our state unchanged. In particular, we must have

$$\mathrm{e}^{-2\pi \mathrm{i}\hat{\mathbf{z}}\cdot\mathbf{L}/\hbar}\,|\ell, m\rangle = \mathrm{e}^{-2\pi \mathrm{i} m}\,|\ell, m\rangle = |\ell, m\rangle \tag{5.62}$$

so that $m \in \mathbb{Z}$ and hence $\ell \in \mathbb{N}_0$.

5.4.1 Spherical Harmonics

The commutation relations of the $\mathbf{L}$'s are exactly the same as those of the $\mathbf{S}$'s, so for any finite ℓ we could choose to represent the orbital angular momentum operators by the same $(2\ell+1)\times(2\ell+1)$ matrices as we obtained above. However, in practical applications we're usually interested in states whose orbital angular momentum quantum number ℓ may change, perhaps as a result of the particle being excited from one energy level to another. Thus it's more convenient to use a formalism that allows us to treat all values of ℓ simultaneously. Furthermore, we often want to know about $\mathbf{L}$ at the same time as knowing about linear momentum $\mathbf{P}$ or position $\mathbf{X}$, and we have seen that the $[X_i, P_j]$ commutation relations do not have any finite dimensional representation and we can only represent them as operators acting on (wave)functions. Let's now see how to reconstruct eigenfunctions of orbital angular momentum — the Legendre polynomials and spherical harmonics you met in IB — from the operator formalism.

In the position representation, the orbital angular momentum operator become

$$\mathbf{L} = \mathbf{X}\times\mathbf{P} = -\mathrm{i}\hbar\mathbf{x}\times\boldsymbol{\nabla}\,, \tag{5.63}$$

so that in particular,

$$L_z = -\mathrm{i}\hbar\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right)\,. \tag{5.64}$$

In terms of spherical polar coordinates we have $(x, y, z) = (r\sin\theta\cos\phi, r\sin\theta\sin\phi, r\cos\theta)$, so

$$\frac{\partial}{\partial\phi} = \frac{\partial x}{\partial\phi}\frac{\partial}{\partial x} + \frac{\partial y}{\partial\phi}\frac{\partial}{\partial y} + \frac{\partial z}{\partial\phi}\frac{\partial}{\partial z} = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y} \tag{5.65}$$

so that $L_z = -\mathrm{i}\hbar\frac{\partial}{\partial\phi}$. Thus, using these coordinates, the eigenvalue equation for L_z becomes

$$\langle\mathbf{x}|L_z|\ell, m\rangle = -\mathrm{i}\hbar\frac{\partial}{\partial\phi}\langle\mathbf{x}|\ell, m\rangle = m\hbar\langle\mathbf{x}|\ell, m\rangle \tag{5.66}$$

or

$$-\mathrm{i}\frac{\partial}{\partial\phi}\psi_{\ell,m}(\mathbf{x}) = m\psi_{\ell,m}(\mathbf{x})\,, \tag{5.67}$$

where $\psi_{\ell,m}(\mathbf{x}) = \langle\mathbf{x}|\ell, m\rangle$ is the position space wavefunction. This is solved by

$$\psi_{\ell,m}(\mathbf{x}) = K(r,\theta)\mathrm{e}^{\mathrm{i}m\phi} \tag{5.68}$$

for some function $K(r,\theta)$. Since $m \in \mathbb{Z}$, $\psi_{\ell,m}$ is a single-valued function of the azimuthal angle ϕ. This is often given as a further reason why only integer values of m (and hence ℓ) should be allowed.

A straightforward, though somewhat tedious calculation shows that the raising and lowering operators

$$L_\pm = L_x \pm iL_y = \pm\hbar\mathrm{e}^{\pm\mathrm{i}\phi}\left(\frac{\partial}{\partial\theta} \pm \mathrm{i}\cot\theta\frac{\partial}{\partial\phi}\right) \tag{5.69}$$

in the position representation. The condition $L_+\psi_{\ell,\ell} = 0$ then fixes

$$\psi_{\ell,\ell}(\mathbf{x}) = R(r)\sin^\ell\theta\mathrm{e}^{\mathrm{i}\ell\phi}\,. \tag{5.70}$$

Applying the lowering operators, one finds that all the other $\psi_{\ell,m}(\mathbf{x})$'s are of the form

$$\psi_{\ell,m}(\mathbf{x}) = R(r)Y_\ell^m(\theta,\phi)\,, \tag{5.71}$$

where the *spherical harmonics*

$$Y_\ell^m(\theta,\phi) = (-1)^m\sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}}P_\ell^m(\cos\theta)\mathrm{e}^{\mathrm{i}m\phi} \tag{5.72}$$

is given in terms of the *associated Legendre polynomials*

$$P_\ell^m(x) = \frac{1}{2^\ell\ell!}(1-x^2)^{m/2}\frac{\mathrm{d}^{\ell+m}}{\mathrm{d}x^{\ell+m}}(x^2-1)^\ell\,. \tag{5.73}$$

In particular, the spherical harmonics with $m = 0$ are proportional to the ordinary Legendre polynomials by

$$Y_\ell^0(\theta) = \sqrt{\frac{2\ell+1}{4\pi}}P_\ell(\cos\theta)\,, \tag{5.74}$$

which are odd or even polynomials in $\cos\theta$, according to whether ℓ is odd or even. In particular, the P_ℓ^m's are only single-valued as functions on S^2 when ℓ is an integer, providing another reason why the half integer values allowed by general considerations of the algebra must in fact be discarded for $\mathbf{L}$.

We won't be concerned with the detailed form of these spherical harmonics, though you may wish to note the orthogonality condition

$$\int_{S^2}\mathrm{d}\theta\,\mathrm{d}\phi\,\sin\theta Y_{\ell'}^{m'*}(\theta,\phi)Y_\ell^m(\theta,\phi) = \delta_{\ell\ell'}\delta_{mm'} \tag{5.75}$$

and the fact that

$$r^2\nabla^2 Y_\ell^m(\theta,\phi) = -\ell(\ell+1)Y_\ell^m(\theta,\phi)\,, \tag{5.76}$$

where ∇^2 is the Laplacian. Furthermore, since $\mathbf{L}$ is invariant under parity, $\Pi^{-1}\mathbf{L}\Pi = +\mathbf{L}$, it follows that so too are the raising and lowering operators $L_\pm$. Therefore, all states in a given ℓ multiplet have the same parity. To determine what that parity is, note that in spherical polar coordinates the transformation $\mathbf{x} \mapsto -\mathbf{x}$ becomes

$$(r, \theta, \phi) \mapsto (r, \pi - \theta, \phi + \pi) \tag{5.77}$$

so in particular $\cos\theta \mapsto -\cos\theta$. Since $P_\ell(-\cos\theta) = (-1)^\ell P_\ell(\cos\theta)$, we see that the parity of Y_ℓ^0 is odd or even, according to whether ℓ is odd or even, and hence

$$Y_\ell^m(\theta - \pi, \phi + \pi) = (-1)^\ell Y_\ell^m(\theta, \phi) \tag{5.78}$$

for all spherical harmonics with a given value of ℓ.

It's occasionally useful to have an alternative form of the spherical harmonics. Consider the polynomial in $\mathbb{R}^3$

$$\psi(\mathbf{x}) = \sum_{i_1, i_2, \ldots, i_l = 1}^{3} \psi_{i_1 i_2 \ldots i_\ell} x^{i_1} x^{i_2} \ldots x^{i_\ell} \tag{5.79}$$

that is homogeneous of degree ℓ. The coefficients $\psi_{i_1 i_2 \ldots i_\ell} \in \mathbb{C}$ are necessarily totally symmetric in their indices, and the polynomial is harmonic, i.e. $\nabla^2\psi(x) = 0$ if $\psi_{i_1 i_2 \ldots i_\ell}$ are traceless on any pair of indices.

Finally, notice that so far we've said nothing about the radial profile of the wavefunction, $R(r)$. Since this function is certainly spherically symmetric, the rotation generators cannot tell us anything about it and to determine $R(r)$ we'd need further information, such as the Hamiltonian.

6 Addition of Angular Momentum

Back in section 2.3, we understood how to describe composite systems using the tensor product of the Hilbert spaces of the individual subsystems. However, in many circumstances, the basis of the tensor product formed by taking all possible pairs of basis elements from the individual subspaces is not the most convenient. A key case is when the subspaces transform under the action of a group G. We'd like to understand the effect of a G-transformation on the combined space, and the "obvious" tensor product basis usually obscures this.

In this chapter, we'll see how this works for the case of the rotation group $G = \mathrm{SO}(3)$. Specifically, suppose two subsystems are each in states of definite total angular momentum — mathematically they transform in irreducible representations of SO(3) with total angular momentum quantum numbers j_1 and j_2, respectively. We'd like to express the tensor product of the subsystems in terms of a sum of Hilbert spaces,

$$\mathcal{H}_{j_1} \otimes \mathcal{H}_{j_2} = \bigoplus_j \mathcal{H}_j \,, \tag{6.1}$$

where each $\mathcal{H}_j$ describes a state of the whole system with definite total angular momentum quantum number j.[43] Physically, to understand this decomposition is to understand how to add the angular momentum of the two subsystems so as to find the possible values of the combined angular momentum of the whole system. For example, in a hydrogen atom, both the proton and electron carry angular momentum $\hbar/2$ by virtue of their spins, and further angular momentum may be present depending on the electrons orbit around the proton. If we wish to treat the atom as a whole, then we'll be concerned with the angular momentum of the combined system rather than that of the electron and proton individually. An even more familiar example is your bike: the total angular momentum comes from a combination of the back and front wheels which can be independent (at least in principle, though I don't recommend trying this while you're riding along!).

Our second task in this chapter is to fill in a gap left in our knowledge: While we know how to use the raising and lowering operators $J_\pm$ to realign a given amount of angular momentum along or away from some axis, we've not yet learnt how to perform a mathematical operation that changes the total angular momentum of our state. We'll understand that, as well as states, operators themselves can carry angular momentum. Applying such an operator to a state yields a new state whose total angular momentum can differ from that of the original state. We'll illustrate this using various physical examples, including radiative transitions induced by an electric dipole moment, and the special "dynamical" symmetries present in the Coulomb and harmonic oscillator potentials.

6.1 Combining the Angular Momenta of Two States

Suppose we have two subsystems ("gyros") enclosed in a box, where the first system has total angular momentum quantum number j_1, and the second has total angular momentum quantum number j_2. A basis of states of the first system is

$$\{|j_1, m_1\rangle\} \quad \text{where} \quad m_1 \in \{-j_1, -j_1+1, \ldots, j_1\} \,, \tag{6.2}$$

while the second system may be described in terms of the basis

$$\{|j_2, m_2\rangle\} \quad \text{where} \quad m_2 \in \{-j_2, -j_2+1, \ldots, j_2\} \,. \tag{6.3}$$

[43] In quantum mechanics we're concerned with Hilbert space, but in fact the inner product plays very little role in this story. The whole of this section carries over more generally to the case of representations of a general Lie group G, where we decompose the tensor product $R_{j_1} \otimes R_{j_2}$ of two irreducible representations (vector spaces) R_{j_1} and R_{j_2} in terms of a sum $\bigoplus_j R_j$ of further irreducible representations. You can learn more by taking the Part II Representation Theory course next term (and more still in the Part III courses).

The state of the combined system can be expressed as

$$|\psi\rangle = \sum_{m_1=-j_1}^{j_1} \sum_{m_2=-j_2}^{j_2} c_{m_1 m_2} |j_1, m_1\rangle \otimes |j_2, m_2\rangle \tag{6.4}$$

with some coefficients $c_{m_1 m_2}$. We can choose m_1 and m_2 independently, so there are a total of $(2j_1+1)(2j_2+1)$ states here — this is the dimension of the tensor product $\mathcal{H}_{j_1} \otimes \mathcal{H}_{j_2}$. We want to understand how the states (6.4) behave under rotations; that is, we'd like to understand which linear combinations in (6.4) correspond to a definite amount of angular momentum for the system as a whole.

Let's first consider the corresponding classical situation, where we may wish to know the combined angular momentum of two gyroscopes. Although the total angular momentum of the individual components is fixed, that of the whole system is variable because it depends on the relative orientation of the two gyros: if they are aligned with each other, the whole system might be expected to have total angular momentum labelled by $j_1 + j_2$, while if the two subsystems are aligned exactly against each other, then you may expect the system as a whole to have total angular momentum $|j_1 - j_2|$. It's important to realise that were not saying anything at all about how the individual subsystems may or may not be coupled to one another dynamically that is, we're not assuming any form of interaction between them in the Hamiltonian. Rather, we're just considering the different relative alignments of their angular momenta that are possible in principle.

Quantum mechanically, the angular momentum operator for the combined system is

$$\mathbf{J} = (\mathbf{J}_1 \otimes 1_{\mathcal{H}_{j_2}}) + (1_{\mathcal{H}_{j_1}} \otimes \mathbf{J}_2)\,, \tag{6.5}$$

where $\mathbf{J}_1$ and $\mathbf{J}_2$ are the angular momentum operators for the two subsystems. It follows that

$$\mathbf{J}^2 = (\mathbf{J}_1^2 \otimes 1_{\mathcal{H}_{j_2}}) + (1_{\mathcal{H}_{j_1}} \otimes \mathbf{J}_2^2) + 2(\mathbf{J}_1 \otimes \mathbf{J}_2)\,, \tag{6.6}$$

where in the final term we take the scalar product of the two spin operators over their spatial indices. We will often abuse notation by writing

$$\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2 \quad \text{and} \quad \mathbf{J}^2 = \mathbf{J}_1^2 + \mathbf{J}_2^2 + 2\mathbf{J}_1 \cdot \mathbf{J}_2\,, \tag{6.7}$$

with the tensor product and identity operators being understood.

We can rewrite the total angular momentum squared operator (6.6) in a way that allows us to understand its action on states $|j_1, m_1\rangle |j_2, m_2\rangle$ (we've again dropped the symbol $\otimes$), which form our basis of (6.4). Using the fact

$$J_x = \frac{J_+ + J_-}{2} \quad \text{and} \quad J_y = \frac{J_+ - J_-}{2\mathrm{i}}\,, \tag{6.8}$$

we have

$$\begin{aligned} 2\mathbf{J}_1 \cdot \mathbf{J}_2 &= 2(J_{1x}J_{2x} + J_{1y}J_{2y} + J_{1z}J_{2z}) \\ &= 2\left(\frac{J_{1+}+J_{1-}}{2}\frac{J_{2+}+J_{2-}}{2} + \frac{J_{1+}-J_{1-}}{2\mathrm{i}}\frac{J_{2+}-J_{2-}}{2\mathrm{i}} + J_{1z}J_{2z}\right) \\ &= J_{1+}J_{2-} + J_{1-}J_{2+} + 2J_{1z}J_{2z}\,. \end{aligned} \tag{6.9}$$

Using this eliminate $\mathbf{J}_1 \cdot \mathbf{J}_2$ from (6.6) allows us to write the total angular momentum operator as

$$\mathbf{J}^2 = \mathbf{J}_1^2 + \mathbf{J}_2^2 + J_{1+}J_{2-} + J_{1-}J_{2+} + 2J_{1z}J_{2z}\,, \tag{6.10}$$

where we now understand how each term acts on any state of the form $|j_1, m_1\rangle |j_2, m_2\rangle$.

Let's consider the action of this operator on states of the whole system. We'll start by examining the state $|j_1, j_1\rangle |j_2, j_2\rangle$ in which both gyros are maximally aligned with the $\hat{\mathbf{z}}$-axis. Since $J_z = J_{1z} + J_{2z}$, we have

$$J_z |j_1, j_1\rangle |j_2, j_2\rangle = (j_1 + j_2)\hbar |j_1, j_1\rangle |j_2, j_2\rangle \tag{6.11}$$

so indeed this state is an eigenstate of J_z with the expected eigenvalue. Also, using (6.9) we have

$$\begin{aligned}\mathbf{J}^2 \left|j_1, j_1\right\rangle \left|j_2, j_2\right\rangle &= (\mathbf{J}_1^2 + \mathbf{J}_2^2 + J_{1+}J_{2-} + J_{1-}J_{2+} + 2J_{1z}J_{2z}) \left|j_1, j_1\right\rangle \left|j_2, j_2\right\rangle \\ &= (j_1(j_1+1) + j_2(j_2+1) + 2j_1j_2)\hbar^2 \left|j_1, j_1\right\rangle \left|j_2, j_2\right\rangle \\ &= (j_1+j_2)(j_1+j_2+1)\hbar^2 \left|j_1, j_1\right\rangle \left|j_2, j_2\right\rangle \ . \end{aligned} \tag{6.12}$$

Thus, by setting $j = j_1 + j_2$, we may write

$$\left|j, j\right\rangle = \left|j_1, j_1\right\rangle \left|j_2, j_2\right\rangle \ , \tag{6.13}$$

since $\left|j_1, j_1\right\rangle \left|j_2, j_2\right\rangle$ satisfies both the defining equations for a state of total angular momentum labelled by $j_1 + j_2$, with it all aligned along $\hat{\mathbf{z}}$. (This is a highest weight state for $j = j_1 + j_2$.)

Now that we've found one mutual eigenstate of the combined $\mathbf{J}^2$ and J_z, we may easily construct others by applying the lowering operator $J_- = J_{1-} + J_{2-}$. Acting on the left using (5.18) we have

$$J_- \left|j, j\right\rangle = \hbar\sqrt{j(j+1) - j(j-1)} \left|j, j-1\right\rangle = \hbar\sqrt{2j} \left|j, j-1\right\rangle \ . \tag{6.14}$$

On the other hand, applying $J_- = J_{1-} + J_{2-}$ to the right-hand side gives

$$\begin{aligned}(J_{1-} + J_{2-}) \left|j_1, j_1\right\rangle \left|j_2, j_2\right\rangle &= \hbar \Big[\sqrt{j_1(j_1+1) - j_1(j_1-1)} \left|j_1, j_1-1\right\rangle \left|j_2, j_2\right\rangle \\ &\qquad + \sqrt{j_2(j_2+1) - j_2(j_2-1)} \left|j_1, j_1\right\rangle \left|j_2, j_2-1\right\rangle\Big] \\ &= \hbar\sqrt{2j_1} \left|j_1, j_1-1\right\rangle \left|j_2, j_2\right\rangle + \hbar\sqrt{2j_2} \left|j_1, j_1\right\rangle \left|j_2, j_2-1\right\rangle \ . \end{aligned} \tag{6.15}$$

Comparing the two sides we learn that

$$\left|j, j-1\right\rangle = \sqrt{\frac{j_1}{j}} \left|j_1, j_1-1\right\rangle \left|j_2, j_2\right\rangle + \sqrt{\frac{j_2}{j}} \left|j_1, j_1\right\rangle \left|j_2, j_2-1\right\rangle \ . \tag{6.16}$$

Note that the left-hand side is an eigenstate of J_z with eigenvalue $(j-1)\hbar$, and indeed the right-hand side is an eigenstate of $J_{1z} + J_{2z}$ with eigenvalue $(j_1 + j_2 - 1)\hbar$. A further application of J_- to the left-hand side and $J_{1-} + J_{2-}$ to the right-hand side would produce an expression for $\left|j, j-2\right\rangle$ and so on.

Since $[\mathbf{J}^2, J_-] = 0$, all the states we produced by acting on $\left|j, j\right\rangle$ with J_- have total angular momentum quantum number $j = j_1 + j_2$. This corresponds to the individual angular momenta of the subsystems always being aligned with each other, with the different $\left|j, m\right\rangle$ telling us about how closely their mutual direction of alignment coincides with the z-axis.

It's perfectly possible for the individual subsystems to not line up with each other. In such a configuration the net total angular momentum of the combined system will be less than the maximum value j_1+j_2. Let's seek an expression for the state $\left|j-1, j-1\right\rangle$ in which the total angular momentum of the whole system is just less than the maximum possible, but where this angular momentum still points along the z-axis. It's trivial to verify that any simple state $\left|j_1, m_1\right\rangle \left|j_2, m_2\right\rangle$ is an eigenstate of $J_z = J_{1z} + J_{2z}$, with eigenvalue $(m_1 + m_2)$. We require $m_1 + m_2 = j - 1 = j_1 + j_2 - 1$, so either $(m_1, m_2) = (j_1 - 1, j_2)$ or $(m_1, m_2) = (j_1, j_2 - 1)$. Also, the state $\left|j-1, j-1\right\rangle$ must be orthogonal to the state $\left|j, j-1\right\rangle$ we found since they are each eigenstates of the Hermitian operator $\mathbf{J}^2$ with distinct eigenvalues. Therefore we must have

$$\left|j-1, j-1\right\rangle = \sqrt{\frac{j_2}{j}} \left|j_1, j_1-1\right\rangle \left|j_2, j_2\right\rangle - \sqrt{\frac{j_1}{j}} \left|j_1, j_1\right\rangle \left|j_2, j_2-1\right\rangle \ . \tag{6.17}$$

Again, this is a highest weight state, now with $j = j_1 + j_2 - 1$. Given this state, we may now proceed to construct all the states $\left|j_1, m\right\rangle$ with $-j + 1 \leq m \leq j - 1$ by applying the lowering operator J_-.

In order to construct $\left|j-2, j-2\right\rangle$, we note that it must be in the span of $\{\left|j_1, j_1-2\right\rangle \left|j_2, j_2\right\rangle, \left|j_1, j_1-1\right\rangle \left|j_2, j_2-1\right\rangle, \left|j_1, j_1\right\rangle \left|j_2, j_2-2\right\rangle\}$, and that $\left|j, j-2\right\rangle$ and $\left|j-1, j-2\right\rangle$ also lie in the span

of these states. Having constructed both $|j, j-2\rangle$ and $|j-1, j-2\rangle$, the state $|j-2, j-2\rangle$ are determined uniquely by orthogonality.

From our classical considerations, it's reasonable to conjecture that the smallest possible total angular momentum quantum number of the whole system is $|j_1 - j_2|$, coming when the two gyros are aligned against one another. Let's check this by working out the total number of states the conjecture leads to. WLOG, we can assume $j_1 \geq j_2$. If the combined system has j values running from $j_1 + j_2$ to $j_1 - j_2$, then we count a total of

$$\begin{aligned}\sum_{j=j_1-j_2}^{j_1+j_2} (2j+1) &= 2\left(\sum_{j=j_1-j_2}^{j_1+j_2} j\right) + (2j_2+1) \\ &= (2j_1+1)(2j_2+1)\,, \end{aligned} \tag{6.18}$$

in agreement with the dimension of the tensor product space we found earlier using the "obvious" tensor product basis in (6.4). We've thus accounted for all the possible states of the system.

The numbers

$$C_{j,m}(j_1, m_1; j_2, m_2) = \langle j, m| \left(|j_1, m_1\rangle \otimes |j_2, m_2\rangle\right) \tag{6.19}$$

are known as the *Clebsch–Gordan coefficients*. Physically, they represent the amplitude that, when the total system is in state $|j, m\rangle$, we will find the subsystems to be in states $|j_1, m_1\rangle$ and $|j_2, m_2\rangle$ when we peer inside the box. For example, we have $C_{3,2}(2, 1; 1, 1) = \sqrt{2/3}$, so if we know our box contains a gyro of spin 2 and a gyro of spin 1, if the box is in state $|3, 2\rangle$ then there's a 2/3 chance that the second gyro has its spin maximally aligned with the z-axis with the first significantly inclined, but only a 1/3 chance that it is the first gyro that is maximally aligned with the z-axis.

Let's take a look at a few examples to familiarise ourselves with the above general formalism.

6.1.1 $j \otimes 0 = j$

First, the trivial case: if one of the subsystems (say the second) has $j_2 = 0$, then it always lies in $|j_2, m_2\rangle = |0, 0\rangle$. Hence the tensor product is trivial

$$\mathcal{H} \cong \mathcal{H}_{j_1} \otimes \mathcal{H}_0 = \mathcal{H}_{j_1} \otimes \mathbb{C} \cong \mathcal{H}_{j_1} \tag{6.20}$$

because we have no choice about the state of the second subsystem. The states

$$\{|j, m\rangle = |j_1, m_1\rangle\, |0, 0\rangle\} \tag{6.21}$$

therefore form a basis of the whole system, and are immediately a single irreducible representation with $j = j_1$.

6.1.2 $\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$

The first non-trivial case is when $j_1 = j_2 = 1/2$. Physically this is relevant, e.g. to the case of the ground state of the hydrogen atom where to understand the spin of the whole atom we must combine the spins of both the proton and the electron. From above, we have

$$|1, 1\rangle_{\mathrm{H}} = |\uparrow\rangle_{\mathrm{e}}\, |\uparrow\rangle_{\mathrm{p}}\,, \tag{6.22}$$

where the subscripts refer to hydrogen, the electron and the proton, respectively. In this state, the spins of the electron and proton are aligned with each other, and they each have maximal possible spin along the z-axis. Applying $J_{-\mathrm{H}} = J_{-\mathrm{e}} + J_{-\mathrm{p}}$ we obtain

$$|1, 0\rangle_{\mathrm{H}} = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle_{\mathrm{e}}\, |\downarrow\rangle_{\mathrm{p}} + |\downarrow\rangle_{\mathrm{e}}\, |\uparrow\rangle_{\mathrm{p}}\right)\,, \tag{6.23}$$

while applying $J_{-\mathrm{H}}$ a second time gives

$$|1,-1\rangle_{\mathrm{H}} = |\downarrow\rangle_{\mathrm{e}} |\downarrow\rangle_{\mathrm{p}} \ . \tag{6.24}$$

Since both the electron and proton are in the $|\downarrow\rangle$ state here, any further applications of either lowering operator will annihilate it, in agreement with the action of the total lowering operator $J_{-\mathrm{H}}$ on the state $|1,-1\rangle_{\mathrm{H}}$ of the whole atom.

Let me point out a perhaps surprising feature of this multiplet. In state (6.23), we are certain to find that the z-component of the spins of the electron and proton are "antiparallel". This may seem surprising given that the atom is still in a spin-1 state, corresponding to the spins of the subsystems being aligned. The resolution of this paradox is that while the z-components are indeed antiparallel, the components in the xy-plane are aligned, although their precise direction is unknown to us. Similarly, in both the $|1,1\rangle$ and $|1,-1\rangle$ states, while the z-components of the electron and proton spins were aligned, their components in the xy-plane were in fact antiparallel. The poor alignment in the xy-plane accounts for the fact that $\sqrt{\langle \mathbf{J}^2 \rangle} = \sqrt{2}\hbar$ for the atom, which is less than the sum of $\sqrt{\langle \mathbf{J}^2 \rangle} = \sqrt{3/4}\hbar$ for the electron and proton individually.

The remaining state of the atom is $|0,0\rangle$, in which the atom has no angular momentum. This should be orthogonal to (6.23), so we find

$$|0,0\rangle_{\mathrm{H}} = \frac{1}{\sqrt{2}} \Big(|\uparrow\rangle_{\mathrm{e}} |\downarrow\rangle_{\mathrm{p}} - |\downarrow\rangle_{\mathrm{e}} |\uparrow\rangle_{\mathrm{p}} \Big) \ . \tag{6.25}$$

The change in sign ensures that the electron and proton spins are antiparallel in the xy-plane as well as in the z direction. In fact, one can check that this state may also be written as

$$|0,0\rangle_{\mathrm{H}} = \frac{1}{\sqrt{2}} \Big(|\uparrow_x\rangle_{\mathrm{e}} |\downarrow_x\rangle_{\mathrm{p}} - |\downarrow_x\rangle_{\mathrm{e}} |\uparrow_x\rangle_{\mathrm{p}} \Big) \ , \tag{6.26}$$

where $|\uparrow_x\rangle$ and $|\downarrow_x\rangle$ are eigenstates for J_x. Similarly,

$$|1,0\rangle_{\mathrm{H}} = \frac{1}{\sqrt{2}} \Big(|\uparrow_x\rangle_{\mathrm{e}} |\uparrow_x\rangle_{\mathrm{p}} - |\downarrow_x\rangle_{\mathrm{e}} |\downarrow_x\rangle_{\mathrm{p}} \Big) \tag{6.27}$$

as one can check straightforwardly.

Altogether, combining two states each with angular momentum 1/2, we've found a triplet of states with total angular momentum $j = 1$:

$$|1,1\rangle = |\uparrow\rangle |\uparrow\rangle \ , \quad |1,0\rangle = \frac{1}{\sqrt{2}} \left(|\uparrow\rangle |\downarrow\rangle + |\downarrow\rangle |\uparrow\rangle \right) \ , \quad |1,-1\rangle = |\downarrow\rangle |\downarrow\rangle \ , \tag{6.28}$$

and a singlet with total angular momentum $j = 0$:

$$|0,0\rangle = \frac{1}{\sqrt{2}} \left(|\uparrow\rangle |\downarrow\rangle - |\downarrow\rangle |\uparrow\rangle \right) \ . \tag{6.29}$$

Nota that each state in the triplet is symmetric under exchange of the angular momenta of the two subsystem, while the singlet state is antisymmetric under this exchange.

6.1.3 $\mathbf{1 \otimes \frac{1}{2} = \frac{3}{2} \otimes \frac{1}{2}}$

If our subsystems have $j_1 = 1$ and $j_2 = \frac{1}{2}$, then the total system can have $j = \frac{3}{2}$ or $j = \frac{1}{2}$. Let's start as always with the highest state

$$\left| \frac{3}{2}, \frac{3}{2} \right\rangle = |1,1\rangle |\uparrow\rangle \ . \tag{6.30}$$

Applying the lowering operator, we find successively the states

$$\left|\frac{3}{2},\frac{1}{2}\right\rangle = \sqrt{\frac{2}{3}}\,|1,0\rangle\,|\uparrow\rangle + \sqrt{\frac{1}{3}}\,|1,1\rangle\,|\downarrow\rangle \tag{6.31}$$

$$\left|\frac{3}{2},-\frac{1}{2}\right\rangle = \sqrt{\frac{1}{3}}\,|1,-1\rangle\,|\uparrow\rangle + \sqrt{\frac{2}{3}}\,|1,0\rangle\,|\downarrow\rangle \tag{6.32}$$

$$\left|\frac{3}{2},-\frac{3}{2}\right\rangle = |1,-1\rangle\,|\downarrow\rangle\ , \tag{6.33}$$

which completes the $j = \frac{3}{2}$ quartet. The remaining $j = \frac{1}{2}$ doublet is

$$\left|\frac{1}{2},\frac{1}{2}\right\rangle = \sqrt{\frac{1}{3}}\,|1,0\rangle\,|\uparrow\rangle - \sqrt{\frac{2}{3}}\,|1,1\rangle\,|\downarrow\rangle \tag{6.34}$$

$$\left|\frac{1}{2},-\frac{1}{2}\right\rangle = \sqrt{\frac{2}{3}}\,|1,-1\rangle\,|\uparrow\rangle - \sqrt{\frac{1}{3}}\,|1,0\rangle\,|\downarrow\rangle\ . \tag{6.35}$$

6.1.4 The Classical Limit

We should expect to recover a classical picture when we combine two systems each with a large amount of angular momentum. Let's see how this occurs. Classically, if we add two angular momentum vectors $\mathbf{j}_1$ and $\mathbf{j}_2$ then the resultant vector has magnitude

$$\mathbf{j}^2 = (\mathbf{j}_1 + \mathbf{j}_2)^2 = \mathbf{j}_1^2 + \mathbf{j}_2^2 + 2\mathbf{j}_1 \cdot \mathbf{j}_2\,. \tag{6.36}$$

If we know nothing about the relative orientations of $\mathbf{j}_1$ and $\mathbf{j}_2$, then all points on a sphere of radius $|\mathbf{j}_2|$ centred on the end of $\mathbf{j}_1$ are equally likely. Consequently, the probability $\mathrm{d}P$ that $\mathbf{j}_1$ and $\mathbf{j}_2$ have alignment in the range $\theta + \mathrm{d}\theta$ is proportional to the area of a band with polar angle between θ and $\theta + \mathrm{d}\theta$ on the unit sphere. The magnitudes of $\mathbf{j}_1$ and $\mathbf{j}_2$ are fixed, so $\mathrm{d}\mathbf{j}^2 = 2\,|\mathbf{j}_1|\,|\mathbf{j}_2|\,\mathrm{d}\cos\theta$. Hence the probability of this alignment is

$$\mathrm{d}P = \frac{2\pi\sin\theta\mathrm{d}\theta}{4\pi} = \frac{|\mathbf{j}|\,\mathrm{d}\,|\mathbf{j}|}{2\,|\mathbf{j}_1|\,|\mathbf{j}_2|}\,, \tag{6.37}$$

and hence

$$\frac{\mathrm{d}P}{\mathrm{d}\,|\mathbf{j}|} = \frac{|\mathbf{j}|}{2\,|\mathbf{j}_1|\,|\mathbf{j}_2|}\,. \tag{6.38}$$

In the quantum case, suppose $j_1, j_2 \gg 1$. Then the fraction of states in the combined system with some amount j of angular momentum is

$$\frac{2j+1}{(2j_1+1)(2j_2+1)} \approx \frac{j}{2j_1j_2} \tag{6.39}$$

provided $|j_1 - j_2| \le j \le j_1 + j_2$. Thus, if we know nothing about the original state of the subsystems, the probability that the combined system has total angular momentum j agrees with the classical probability computed above.

6.2 Angular Momentum of Operators

Suppose that $\mathbf{V}$ is a vector operator, so that in particular $U^{-1}(\boldsymbol{\alpha})\mathbf{V}U(\boldsymbol{\alpha}) = \mathsf{R}(\boldsymbol{\alpha})\mathbf{V}$ under rotations, or

$$[J_i, V_j] = \mathrm{i}\hbar\sum_k \epsilon_{ijk}V_k \tag{6.40}$$

infinitesimally. We define the spherical components of $\mathbf{V}$ by

$$V^{+1} = -\frac{1}{\sqrt{2}}(V_x + iV_y)\,, \quad V^{-1} = \frac{1}{\sqrt{2}}(V_x - iV_y)\,, \quad V^0 = V_z\,. \tag{6.41}$$

Then the commutation relations are equivalent to

$$[J_i, V^m] = m\hbar V^m\,, \tag{6.42}$$

$$[J_\pm, V^m] = \hbar\sqrt{2 - m(m \pm 1)}V^{m\pm 1}\,. \tag{6.43}$$

6.2.1 The Wigner–Eckart Theorem and Transition Dipole Moment

Wigner–Eckart Theorem states that given a tensor operator $\mathbf{T}^{(k)}$ and two states of angular momenta j and j', there exists a constant $\left\langle j \middle\| \mathbf{T}^{(k)} \middle\| j' \right\rangle$ known as the *reduced matrix element* such that for all m, m' and q, we have

$$\left\langle j', m' \middle| T_q^{(k)} \middle| jm \right\rangle = C_{j',m'}(k, q; j, m) \left\langle j \middle\| \mathbf{T}^{(k)} \middle\| j' \right\rangle\,, \tag{6.44}$$

where $T_q^{(k)}$ is the qth component of the spherical tensor operator $\mathbf{T}^{(k)}$ of rank k.

We are not proving it, but we are doing a motivating example.

Let's say we want to calculate transition dipole moments for an electron transition from a 4d to a 2p orbital of a hydrogen atom, i.e. the matrix elements of the form $\langle 2\text{p}, m_1|r_i|4\text{d}, m_2\rangle$, where r_i is either the x, y, or z component of the position operator, and m_1, m_2 are the magnetic quantum numbers that distinguish different orbitals within the 2p or 4d subshell. If we do this directly, it involves calculating 45 different integrals: there are 3 possibilities for m_1: $m_1 \in \{-1, 0, 1\}$, 5 possibilities for m_2: $m_2 \in \{-2, -1, 0, 1, 2\}$, and 3 possibilities for i.

The Wigner–Eckart theorem allows one to obtain the same information after evaluating just one of those 45 integrals (any of them can be used, as long as it is nonzero). Then the other 44 integrals can be inferred from that first one — without the need to write down any wavefunctions or evaluate any integrals — with the help of Clebsch–Gordan coefficients, which can be easily looked up in a table or computed by hand or computer.

The Wigner–Eckart theorem works because all 45 of these different calculations are related to each other by rotations. If an electron is in one of the 2p orbitals, rotating the system will generally move it into a different 2p orbital (usually it will wind up in a quantum superposition of all three basis states, $m = +1, 0, -1$). Similarly, if an electron is in one of the 4d orbitals, rotating the system will move it into a different 4d orbital. Finally, an analogous statement is true for the position operator: when the system is rotated, the three different components of the position operator are effectively interchanged or mixed.

If we start by knowing just one of the 45 values (say, we know $\langle 2\text{p}, m_1|r_i|4\text{d}, m_2\rangle = K$) and then we rotate the system, we can infer that K is also the matrix element between the rotated version of $\langle 2\text{p}, m_1|$, the rotated version of r_i, and the rotated version of $|4\text{d}, m_2\rangle$. This gives an algebraic relation involving K and some or all of the 44 unknown matrix elements. Different rotations of the system lead to different algebraic relations, and it turns out that there is enough information to figure out all of the matrix elements in this way.

6.2.2 The Runge–Lenz Vector in Hydrogen

For a spinless particle of mass M in a generic central potential, the Hamiltonian takes the form

$$H = \frac{\mathbf{P}^2}{2M} + V(|\mathbf{X}|)\,. \tag{6.45}$$

By definition, any such Hamiltonian is rotationally invariant, so it follows that

$$[\mathbf{L}, H] = 0 \qquad \text{and} \qquad [\mathbf{L}^2, H] = 0\,. \tag{6.46}$$

Since we also have $[L_i, \mathbf{L}^2] = 0$ we can find a complete set of simultaneous eigenstates of H, $\mathbf{L}^2$ and L_z, so we can label our states as $|n, \ell, m\rangle$. The Hamiltonian does not involve L_z, except through $\mathbf{L}^2$, so the energy level $E_{n,\ell,m}$ of such a state must certainly be independent of m. However, the Hamiltonian does involve $\mathbf{L}^2$ through the kinetic operator, so we do not generically expect the energy levels to be independent of ℓ. Physically, if we give our particle more or less angular momentum, it will whizz round faster or slower. Typically, we'd expect this to change the energy.

Nonetheless, it occasionally happens that a Hamiltonian is invariant under transformations beyond those inherited from the transformations of the space $\mathbb{R}^3$ in which the quantum system lives. If the algebra of the corresponding generators closes, meaning that the commutator of any pair of generators is equal to another generator (perhaps including the Hamiltonian), then we say we have a *dynamical symmetry*. In such cases, it may be possible to change the total angular momentum of a state without changing its energy, so that states of different orbital angular momentum quantum number ℓ can be degenerate.

Hamiltonians with dynamical symmetry are very rare: for single particle quantum systems moving on $\mathbb{R}^3$ with a central potential, the only cases are the harmonic oscillator and motion in a Coulomb potential. However, these two cases are so important that it's worth considering them from this point of view.

The first case we'll study is for the Coulomb potential, relevant e.g. to the hydrogen atom. Classically, conservation of angular momentum in any central potential implies that the motion is conned to a plane, but the Kepler problem has a further conserved quantity known as the *Runge–Lenz vector* $\mathbf{r}$, given by

$$\mathbf{r} = -e^2 \frac{\mathbf{x}}{|\mathbf{x}|} + \frac{1}{M}\mathbf{p} \times \mathbf{L}\,. \tag{6.47}$$

Conservation of this vector implies that classical bound states have closed orbits; the motion is not just in a plane, but follows a fixed ellipse. (The observation of a gradual shift in Mercury's orbit — the precession of its perihelion — was one of the early confirmations of General Relativity's modifications to Newton's Laws of Gravity.)

Now let's consider the quantum mechanical case. We construct the Runge–Lenz operator

$$\begin{aligned} \mathbf{R} &= -\frac{e^2}{|\mathbf{X}|}\mathbf{X} + \frac{1}{2M}(\mathbf{P} \times \mathbf{L} - \mathbf{L} \times \mathbf{P}) \\ &= -\frac{e^2}{|\mathbf{X}|}\mathbf{X} + \frac{1}{M}\mathbf{P} \times \mathbf{L} - \frac{\mathrm{i}\hbar}{M}\mathbf{P}\,. \end{aligned} \tag{6.48}$$

Classically, the two terms in brackets on the first line of (6.48) are equal; we've written it in this way to ensure that $\mathbf{R}^\dagger = \mathbf{R}$.

Since $\mathbf{R}$ is a vector operator built from $\mathbf{X}$ and $\mathbf{P}$, it follows that its commutation relations with $\mathbf{L}$ are

$$[L_i, R_j] = \mathrm{i}\hbar \sum_k \epsilon_{ijk} R_k \tag{6.49}$$

so that it transforms as a vector under rotations. Therefore, the Wigner–Eckart theorem says that

$$\left\langle n', \ell', m' \middle| R_1^k \middle| n, \ell, m \right\rangle = C_{\ell',m'}(1, k; \ell, m) \left\langle n' \| R \| n \right\rangle\,. \tag{6.50}$$

Furthermore, a somewhat tedious calculation shows that $[H, \mathbf{R}] = 0$ so the Runge–Lenz vector is conserved. Thus the reduced matrix element $\langle n' \| R \| n \rangle$ vanishes unless $n' = n$.

The orbital angular momentum $\mathbf{L} = \mathbf{X} \times \mathbf{P}$ is orthogonal to each of the three terms in the expression of the Runge–Lenz vector (6.48), so $\mathbf{R} \cdot \mathbf{L} = \mathbf{L} \cdot \mathbf{R} = 0$. We also have

$$\begin{aligned} \mathbf{L}^2 &= (\mathbf{X} \times \mathbf{P}) \cdot \mathbf{L} = \mathbf{X} \cdot (\mathbf{P} \times \mathbf{L}) \\ &= (\mathbf{P} \times \mathbf{L}) \cdot \mathbf{X} + \sum_{i,j,k=1}^{3} \epsilon_{ijk} [X_i, P_j L_k] \\ &= (\mathbf{P} \times \mathbf{L}) \cdot \mathbf{X} + 2\mathrm{i}\hbar \mathbf{P} \cdot \mathbf{X} \end{aligned} \tag{6.51}$$

as well as

$$\mathbf{P} \cdot (\mathbf{P} \times \mathbf{L}) = 0\,, \quad (\mathbf{P} \times \mathbf{L}) \cdot \mathbf{P} = 2\mathrm{i}\hbar \mathbf{P}^2\,, \quad (\mathbf{P} \times \mathbf{L})^2 = \mathbf{P}^2 \mathbf{L}^2\,, \tag{6.52}$$

where we've been careful to keep track of operator orders. Using these results, we find that the length-squared of the Runge–Lenz operator is given by

$$\mathbf{R}^2 = Z^2 e^4 + \frac{2H}{\mu}(L^2 + \hbar^2)\,. \tag{6.53}$$

The point of this expression is that it relates $\mathbf{R}$ to H; since we know the spectrum of $\mathbf{L}^2$, we can find the eigenvalues of H if we can find the eigenvalues of $\mathbf{R}$. A rather longwinded calculation shows that the components of the Runge–Lenz operator obey the commutation relations

$$[R_i, R_j] = -\frac{2\mathrm{i}\hbar}{\mu} H \sum_k \epsilon_{ijk} L_k\,, \tag{6.54}$$

where we note that H commutes with $\mathbf{L}$.

We now introduce the operators

$$\mathbf{A}_\pm = \frac{1}{2}\left(\mathbf{L} \pm \sqrt{\frac{\mu}{-2H}}\mathbf{R}\right) \tag{6.55}$$

built from the orbital angular momentum operator, the Runge–Lenz operator and the Hamiltonian, where we recall that functions (such as the square root) of an operator are defined as in (2.43). The point of introducing these strange-looking combinations is that commutation relations (6.49) and (6.54) show that the $\mathbf{A}_\pm$ obey the algebra

$$[A_{\pm i}, A_{\pm j}] = \mathrm{i}\hbar \sum_k \epsilon_{ijk} A_{\pm k} \quad \text{and} \quad [\mathbf{A}_\pm, \mathbf{A}_\mp] = 0\,, \tag{6.56}$$

which we recognise[44] as two independent copies of the angular momentum $\mathfrak{so}(3)$. In particular, just as for the total angular momentum, the eigenvalues of $\mathbf{A}_\pm^2$ take the form $a_\pm(a_\pm + 1)\hbar^2$ where $a_\pm \in \{0, 1/2, 1, 3/2\}$, with the eigenvalues of any component of $\mathbf{A}_\pm$ running from $-\hbar a_\pm$ to $+\hbar a_\pm$ in integer steps of $\hbar$.

Since $[\mathbf{A}_\pm, H] = 0$ we can find simultaneous eigenstates of H, $\mathbf{A}_\pm^2$ and, say, A_{+z} and A_{-z}. However, taking the square of (6.55) and using the fact $\mathbf{L} \cdot \mathbf{R} = \mathbf{R} \cdot \mathbf{L} = 0$,

$$\mathbf{A}_+^2 = \mathbf{A}_-^2 = \frac{1}{4}\left(L^2 + \frac{\mu}{2H}\mathbf{R}^2\right) = -\frac{Z^2 e^4 \mu}{8H} - \frac{\hbar^2}{4}\,. \tag{6.57}$$

In particular, we cannot choose these eigenvalues of $\mathbf{A}_+^2$ and $\mathbf{A}_-^2$ independently, but must have $a_+ = a_-$. Calling the common eigenvalue a, an simultaneous eigenstate of H and $\mathbf{A}_\pm^2$ obeys

$$-\frac{Z^2 e^4 \mu}{8E} = \left(a(a+1) + \frac{1}{4}\right)\hbar^2 = \frac{\hbar^2}{4}(2a+1)^2\,. \tag{6.58}$$

[44] For bound states of the Coulomb potential, the spectrum of H is negative-definite, so the operators $\mathbf{A}_\pm$ are indeed Hermitian when acting on such states. At energies above the ionisation threshold, the square root in (6.55) means that $\mathbf{A}_\pm$ can no longer be treated as Hermitian, and this algebra is then better described as $\mathfrak{sl}(2)$.

For a any non-negative half integer, $2a+1 \in \{1,2,3,\dots\}$, so the energy levels can be written as

$$E_n = -\frac{Z^2 e^4 \mu}{2\hbar^2 n^2}\,, \quad n \in \mathbb{N}\,. \tag{6.59}$$

These are the energy levels of hydrogenic atoms that you obtained in IB QM. The eigenvalues of A_{+z} and A_{-z} can be chosen independently, and each run from $-a\hbar$ to $a\hbar$ in steps of $\hbar$. Thus the degeneracy of the state with energy E_n is $(2a+1)^2 = n^2$. Again, this is bigger than the degeneracy of a generic central potential, and the degenerate states transform in representations of the enhanced $\mathfrak{so}(3) \times \mathfrak{so}(3)$ symmetry algebra. The diagonal part of this, generated by $\mathbf{L} = \mathbf{A}_+ + \mathbf{A}_-$ corresponds to spatial rotations.

Incidentally, it's useful to note that the energy levels of a hydrogenic atom can be written as

$$E_n = -\frac{Z^2\alpha^2}{2n^2} mc^2\,, \tag{6.60}$$

where the dimensionless number

$$\alpha = \frac{e^2}{4\pi\epsilon_0 \hbar c} \tag{6.61}$$

is known as the *fine structure constant*. Experimentally, $\alpha \approx 1/137$. The significance of this is that, at least for small atomic number Z, $\left|E_n/mc^2\right| \ll 1$ providing an *a posteriori* justification of our use of non-relativistic quantum mechanics to study the hydrogen atom.

6.2.3 Enhanced Symmetry of the 3D Isotropic Harmonic Oscillator

As a second example, consider an isotropic 3D harmonic oscillator, with Hamiltonian

$$H = \frac{\mathbf{P}^2}{2M} + \frac{1}{2} M\omega^2 \mathbf{X}^2\,. \tag{6.62}$$

The fact that H is quadratic in $\mathbf{X}$ as well as in $\mathbf{P}$ means it has an enhanced symmetry beyond the obvious SO(3) rotational invariance. We can see this from the form

$$H = \hbar\omega \left(\sum_i A_i^\dagger A_i + \frac{3}{2} \right) \tag{6.63}$$

in terms of the raising and lowering operators. Written this way, H is clearly invariant under the transformations

$$A_i \longmapsto A_i' = \sum_{j=1}^{3} u_{ij} A_j \tag{6.64}$$

$$A_i^\dagger \longmapsto (A_i')^\dagger = \sum_{k=1}^{3} A_k^\dagger (u^\dagger)_{ki} \tag{6.65}$$

for any 3×3 unitary matrix u. These matrices must be unitary if the new $\mathbf{A}'^\dagger$ is indeed to be the Hermitian conjugate of the new $\mathbf{A}'$. It is important to realise that, although they are unitary, the matrices u_{ij} do not act on the Hilbert space of the harmonic oscillator, but just on the spatial indices of the vector operators $\mathbf{A}$ and $\mathbf{A}^\dagger$. They thus mix $\mathbf{X}$ and $\mathbf{P}$ together[45]. Note also that the new raising and lowering operators obey the same commutation relations as the original ones.

A 3×3 unitary matrix has 9 independent real components, whereas a rotation in three dimensions is completely specified by only 3 independent Euler angles, so the U(3) symmetry of this Hamiltonian

[45] For those taking Integrable Systems: classically these are symmetries of phase space that do not come from symmetries of space itself, and reflect the fact that the harmonic oscillator is (super-)integrable.

is much bigger than the SO(3) rotational invariance of a generic central potential. The fact that the Hamiltonian is invariant under the U(3) transformations (6.64) implies that there should be a total of nine conserved quantities corresponding to the nine generators of this symmetry. It's easy to check that these generators form the tensor operator $\mathbf{T}$ with components $T_{ij} = A_i^\dagger A_j$. We have $[H, \mathbf{T}] = 0$ and hence each component T_{ij} is conserved. To understand these operators, we decompose $\mathbf{T}$ into its irreducible parts, obtaining

$$T_{ij} = \mathbf{A}^\dagger \cdot \mathbf{A} \frac{\delta_{ij}}{3} + \frac{A_i^\dagger A_j - A_j^\dagger A_i}{2} + \left[\frac{A_i^\dagger A_j + A_j^\dagger A_i}{2} - \mathbf{A}^\dagger \cdot \mathbf{A} \frac{\delta_{ij}}{3} \right] . \tag{6.66}$$

The trace $\mathbf{A}^\dagger \cdot \mathbf{A}$ is (up to a constant) just the Hamiltonian itself, which is certainly conserved. Recalling the definition of the raising and lowering operators, the anti-symmetric part can be combined into a vector with components

$$\begin{aligned} \sum_{j,k} \epsilon_{ijk} A_j^\dagger A_k &= \frac{1}{2M\hbar\omega} \sum_{j,k} \epsilon_{ijk} (M\omega X_j - \mathrm{i} P_j)(M\omega X_k + \mathrm{i} P_k) \\ &= \frac{\mathrm{i}}{2\hbar} \sum_{j,k} \epsilon_{ijk} (X_j P_k - P_j X_k) = \frac{\mathrm{i}}{\hbar} \mathbf{L}_i \,, \end{aligned} \tag{6.67}$$

where in the last equality we used the fact $\epsilon_{ijk}(X_j P_k - P_j X_k) = 2\epsilon_{ijk} X_k P_k$, with the ϵ symbol ensuring that the components of $\mathbf{X}$ and $\mathbf{P}$ commute. Thus

$$\mathbf{L} = -\mathrm{i}\hbar(\mathbf{A}^\dagger \times \mathbf{A}) \,, \tag{6.68}$$

so that the vector part of $\mathbf{T}$ is nothing but the usual orbital angular momentum operator for the 3d oscillator. It is no surprise that this is conserved.

The remaining five conserved quantities — the symmetric, traceless part $(T_{ij} + T_{ji})/2 - \sum_i T_{ii}/3$ of $\mathbf{T}$ — generate transformations that mix $\mathbf{X}$ and $\mathbf{P}$, appropriately rescaled. They're less familiar, as they're special to the simple harmonic oscillator. The five components of this traceless symmetric part correspond to the spin-2 part of the operator $\mathbf{T}$. We can see this degeneracy by examining the excited states of the oscillator. Just as in 1d, these may be obtained by acting on $|\mathbf{0}\rangle$ with any component of the raising operators $\mathbf{A}^\dagger$. Since the ground state has energy $3\hbar\omega/2$ and each application of a raising operator increases the energy by one unit, each of the states

$$A_{i_N}^\dagger \ldots A_{i_2}^\dagger A_{i_1}^\dagger \, |\mathbf{0}\rangle \tag{6.69}$$

has the same energy $(N + 3/2)\hbar\omega$, where we can choose any $\{i_1, i_2, \ldots, i_N\} \in \{1, 2, 3\}$. We often call (correctly normalised) such state $|n_x, n_y, n_z\rangle$, where n_x is the total number of times we've applied the x-component of $\mathbf{A}^\dagger$, and similarly for n_y and n_z. Of course, $N = n_x + n_y + n_z$.

Since the components of the raising operators all commute with one another, we can consider the indices $i_1, \ldots, i_N$ in (6.69) to be symmetrized. Consequently, the degeneracy of the Nth energy level is the same as the number of independent components of a rank N tensor in 3 dimensions, which is $(N + 1)(N + 2)/2$ (Think "stars & bars", or the number of ways to partition N into three non-negative integers (n_x, n_y, n_z).) This is the larger degeneracy that we promised. It's easy to derive this degeneracy in number of ways; our derivation here makes clear that the degenerate energy states transform in representations of the larger $\mathfrak{su}(3)$ algebra.

6.3 The Radial Hamiltonian for the Coulomb Potential

For some purposes, it's useful to study hydrogen from a point of view that emphasises the fact that the potential is central, so energy eigenstates may also be chosen to have definite angular momentum. In the first problem set you showed that the kinetic part of the Hamiltonian could be written as

$$T = \frac{1}{2\mu} \left(P_r^2 + \frac{\mathbf{L}^2}{R^2} \right) , \tag{6.70}$$

where $R^2 = \mathbf{X}\cdot\mathbf{X}$ and $P_r = \frac{1}{2}(\hat{\mathbf{X}}\cdot\mathbf{P}+\mathbf{P}\cdot\hat{\mathbf{X}})$ is the radial momentum. Thus, when acting on an orbital angular momentum eigenstate $|\ell, m\rangle$, the Hamiltonian of a hydrogenic atom with atomic number Z reduces to

$$H_\ell = \frac{P_r^2}{2\mu} + \frac{\ell(\ell+1)\hbar^2}{2\mu R^2} - \frac{Ze^2}{4\pi\epsilon_0 R}\,. \tag{6.71}$$

As in IA Dynamics and Relativity, this is just what we'd find for a particle moving in one dimension under the influence of the *effective potential*

$$V_{\text{eff}}(R) = \frac{\ell(\ell+1)\hbar^2}{2\mu R^2} - \frac{Ze^2}{4\pi\epsilon_0 R}\,. \tag{6.72}$$

H_ℓ governs the oscillation of the particle around the minima of this potential. Note that since H_ℓ depends only on the radius R and radial momentum P_r, in position space its eigenstates will determine the radial dependence $R_E(r) = \langle r|E(\ell)\rangle$ of the full wavefunction. A basis of eigenstates $|E, \ell, m\rangle$ of the full Hamiltonian can be chosen to be (tensor) products of the eigenstates $|E(\ell)\rangle$ of H_ℓ and the eigenstates $|\ell, m\rangle$ of $\mathbf{L}^2$ and L_z, whose position space wavefunction $Y_\ell^m(\theta,\phi) = \langle\hat{\mathbf{r}}|\ell, m\rangle$ is a spherical harmonic.

All three terms in H_ℓ must have the same dimensions, and in particular the ratio of the two terms in V_{eff} must be dimensionless. Thus

$$\frac{\hbar^2}{\mu R^2} \times \frac{4\pi\epsilon_0 R}{e^2} = \frac{4\pi\epsilon_0\hbar^2}{\mu e^2} \times \frac{1}{R} \tag{6.73}$$

is dimensionless. It follows that the *Bohr radius*

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{\mu e^2} \tag{6.74}$$

is a characteristic length scale of the hydrogen atom. We have

$$H_\ell = \frac{\hbar^2}{2\mu}\left(\frac{P_r^2}{\hbar^2} + \frac{\ell(\ell+1)}{R^2} - \frac{2Z}{a_0 R}\right) \tag{6.75}$$

in terms of a_0.

Let's define a new operator A_ℓ by

$$A_\ell = \frac{a_0}{\sqrt{2}}\left(\frac{\mathrm{i}}{\hbar}P_r - \frac{\ell+1}{R} + \frac{Z}{(\ell+1)a_0}\right)\,. \tag{6.76}$$

Then one finds

$$\begin{aligned} A_\ell^\dagger A_\ell &= \frac{a_0^2}{2}\left(-\frac{\mathrm{i}}{\hbar}P_r - \frac{\ell+1}{R} + \frac{Z}{(\ell+1)a_0}\right)\left(\frac{\mathrm{i}}{\hbar}P_r - \frac{\ell+1}{R} + \frac{Z}{(\ell+1)a_0}\right) \\ &= \frac{a_0^2}{2}\left(\frac{P_r^2}{\hbar^2} + \left(\frac{Z}{(\ell+1)a_0} - \frac{\ell+1}{R}\right)^2 + \mathrm{i}\frac{\ell+1}{\hbar}\left[P_r, \frac{1}{R}\right]\right)\,. \end{aligned} \tag{6.77}$$

By the commutator identity $[A, B^{-1}] = -B^{-1}[A, B]B^{-1}$ and $[P_r, R] = -\mathrm{i}\hbar$, this gives

$$\begin{aligned} A_\ell^\dagger A_\ell &= \frac{a_0^2}{2}\left(\frac{P_r^2}{\hbar^2} + \frac{Z^2}{(\ell+1)^2 a_0^2} + \frac{(\ell+1)^2}{R^2} - \frac{2Z}{a_0 R} - \frac{\mathrm{i}}{\hbar}\frac{\ell+1}{R}[P_r, R]\frac{1}{R}\right) \\ &= \frac{a_0^2}{2}\left(\frac{P_r^2}{\hbar^2} + \frac{Z^2}{(\ell+1)^2 a_0^2} + \frac{\ell(\ell+1)}{R^2} - \frac{2Z}{a_0 R}\right) \\ &= \frac{a_0^2\mu}{\hbar^2}H_\ell + \frac{Z^2}{2(\ell+1)^2}\,. \end{aligned} \tag{6.78}$$

Had we instead evaluated $A_\ell A_\ell^\dagger$ the only difference would have been the sign in front of the commutator, so we would instead find

$$A_\ell A_\ell^\dagger = \frac{a_0^2\mu}{\hbar^2} H_{\ell+1} + \frac{Z^2}{2(\ell+1)^2} \tag{6.79}$$

which now involves the Hamiltonian appropriate for states whose orbital angular momentum is given by $\ell + 1$. Rearranging (6.78) allows us to write the radial Hamiltonian as

$$H_\ell = \frac{\hbar^2}{\mu a_0^2}\left(A_\ell^\dagger A_\ell - \frac{Z^2}{2(\ell+1)^2}\right) \tag{6.80}$$

while taking the difference of (6.78) and (6.79) gives the commutation relations

$$\left[A_\ell, A_\ell^\dagger\right] = \frac{a_0^2\mu}{\hbar^2}(H_{\ell+1} - H_\ell)\,. \tag{6.81}$$

The final result we need is the commutation relation

$$[A_\ell, H_\ell] = \frac{\hbar^2}{\mu a_0^2}\left[A_\ell, A_\ell^\dagger A_\ell\right] = \frac{\hbar^2}{\mu a_0^2}\left[A_\ell, A_\ell^\dagger\right] A_\ell = (H_{\ell+1} - H_\ell)A_\ell\,. \tag{6.82}$$

Cancelling the $H_\ell A_\ell$ term on both sides, this simplifies to

$$A_\ell H_\ell = H_{\ell+1} A_\ell\,, \tag{6.83}$$

and we also have $A_\ell^\dagger H_{\ell+1} = H_\ell A_\ell^\dagger$ by taking the Hermitian conjugate.

Now let's understand these operators. Let $|E,\ell\rangle$ be an eigenstate of H_ℓ and $\mathbf{L}^2$ with $H_\ell|E,\ell\rangle = E|E,\ell\rangle$ and $\mathbf{L}^2|E,\ell\rangle = \ell(\ell+1)\hbar^2|E,\ell\rangle$. Then

$$H_{\ell+1}(A_\ell|E,\ell\rangle) = A_\ell H_\ell|E,\ell\rangle = E(A_\ell|E,\ell\rangle) \tag{6.84}$$

so that $A_\ell|E,\ell\rangle$ is an eigenstate of $H_{\ell+1}$ with the same energy E. Thus acting on $|E,\ell\rangle$, A_ℓ creates a new state with the same energy, but where the effective potential corresponds to the angular momentum quantum number $\ell+1$. Acting on the new state $A_\ell|E,\ell\rangle$ with $A_{\ell+1}$ again creates a new state with the same energy, but even more orbital angular momentum labelled by $\ell+2$, and so on. We already know that our energy levels are not infinitely degenerate, so this process must eventually terminate. That is, just as in the harmonic oscillator, there must be a maximum value $\ell_{\max}$ such that

$$A_{\ell_{\max}}|E,\ell_{\max}\rangle = 0 \tag{6.85}$$

and again, by analogy with the classical case where a circular orbit has the highest possible angular momentum for fixed energy, we can view this state as the quantum analogue of a circular orbit. Taking the norm and using equation (6.78) we have

$$0 = \|A_{\ell_{\max}}|E,\ell_{\max}\rangle\| = \left\langle A_{\ell_{\max}}^\dagger A_{\ell_{\max}}\right\rangle E,\ell_{\max} = \frac{a_0^2\mu}{\hbar^2}E + \frac{Z^2}{2(\ell_{\max}+1)^2}\,, \tag{6.86}$$

or equivalently

$$E = -\frac{Z^2\hbar^2}{2\mu a_0^2}\frac{1}{n^2}\,, \tag{6.87}$$

where $n = \ell_{\max}+1$ labels the energy level. This agrees with the spectrum of the Hamiltonian computed earlier using the Runge–Lenz vector, or in IB QM. The virtue of the Runge–Lenz derivation is that it makes transparent the role if the additional, dynamical symmetry of hydrogen, explaining why the energy levels fall into multiplets of $\mathfrak{so}(3)\times\mathfrak{so}(3)$. The virtue of the present derivation is that it makes transparent how the degeneracy arises because, aside from simply rotating our system, we can trade radial kinetic energy and Coulomb potential for different orbital kinetic energy whilst keeping the total energy constant.

6.4 Angular Momentum of the Isotropic Oscillator

Same as for the hydrogen atom, we can write the Hamiltonian of the 3d isotropic harmonic oscillator acting on an angular momentum eigenstate $|\ell, m\rangle$ as

$$H_\ell = \frac{P_r^2}{2\mu} + \frac{\ell(\ell+1)\hbar^2}{2\mu R^2} + \frac{1}{2}\mu\omega^2 R^2 \,. \tag{6.88}$$

We can find the spectrum of H_ℓ using raising and lowering operators in a similar fashion to our treatment of the full H. We introduce

$$A_\ell = \frac{1}{\sqrt{2\mu\hbar\omega}}\left(\mathrm{i}P_r - \frac{(\ell+1)\hbar}{R} + \mu\omega R\right) \tag{6.89}$$

and a short computation using $[R, P_r]$ shows that

$$H_\ell = \hbar\omega\left(A_\ell^\dagger A_\ell + \left(\ell + \frac{3}{2}\right)\right) \,. \tag{6.90}$$

We have the commutation relations

$$[A_\ell, A_\ell^\dagger] = 1 + \frac{(\ell+1)\hbar}{\mu\omega R^2} = \frac{H_{\ell+1} - H_\ell}{\hbar\omega} + 1 \tag{6.91}$$

and

$$[A_\ell, H_\ell] = \hbar\omega[A_\ell, A_\ell^\dagger A_\ell] = \hbar\omega[A_\ell, A_\ell^\dagger]A_\ell = (H_{\ell+1} - H_\ell + \hbar\omega)A_\ell \,, \tag{6.92}$$

or equivalently

$$H_{\ell+1}A_\ell - A_\ell H_\ell = -\hbar\omega A_\ell \,. \tag{6.93}$$

This is very close to showing that A_ℓ is a lowering operator, but the presence of $H_{\ell+1}$ means that the algebra does not close unless we consider all values of ℓ. To understand the implications of this, let $|E(\ell)\rangle$ be an eigenstate of H_ℓ, with $H_\ell\,|E(\ell)\rangle = E\,|E(\ell)\rangle$ where we emphasise that the energies depend on the total orbital angular momentum quantum number ℓ. Then

$$H_{\ell+1}(A_\ell\,|E(\ell)\rangle) = (E - \hbar\omega)A_\ell\,|E(\ell)\rangle \,, \tag{6.94}$$

which implies that $A_\ell\,|E(\ell)\rangle$ is an eigenstates of $H_{\ell+1}$ with eigenvalue $E - \hbar\omega$. The Hamiltonian $H_{\ell+1}$ is the one appropriate for the radial part of wavefunctions with total angular momentum labelled by $\ell + 1$, so applying A_ℓ to $|E(\ell)\rangle$ has created a state with lower energy, but with a radial wavefunction corresponding to a state of greater orbital angular momentum.

By a now familiar argument, in any H_ℓ eigenstate we have

$$\frac{E}{\hbar\omega} - \ell - \frac{3}{2} = \left\langle E(\ell)\middle|\frac{H_\ell}{\hbar\omega} - \ell - \frac{3}{2}\middle|E(\ell)\right\rangle = \left\langle E(\ell)\middle|A_\ell^\dagger A_\ell\middle|E(\ell)\right\rangle = \|A_\ell\,|E(\ell)\rangle\|^2 \geq 0 \,. \tag{6.95}$$

Thus, for a given energy E the maximum angular momentum a state can have is

$$\ell_{\max} = \frac{E}{\hbar\omega} - \frac{3}{2} \,. \tag{6.96}$$

As $\ell_{\max}$ is a non-negative integer (since it is a possible value for ℓ at given E) we see that, the ground state energy is $3\hbar\omega/2$, which occurs when $\ell_{\max} = 0$, meaning that this ground state is necessarily spherically symmetric.

For any given energy, the state $|E(\ell_{\max})\rangle$ which saturates the bound (6.95) obeys $A_{\ell_{\max}}\,|E(\ell_{\max})\rangle = 0$ and has the maximum possible total angular momentum out of all the states with this energy. It is thus the quantum equivalent of a circular orbit. In the position representation we have

$$0 = \langle r|A_{\ell_{\max}}|E(\ell_{\max})\rangle = \frac{1}{\sqrt{2\mu\hbar\omega}}\left(\frac{\partial}{\partial r} + \frac{1}{r} - \frac{\ell_{\max}+1}{r} + \frac{m\omega r}{\hbar}\right)\langle r|E(\ell_{\max})\rangle \,, \tag{6.97}$$

where we use $|r\rangle$ to denote a state that is definitely located at radius r. This equation is solved by

$$\langle r|E(\ell_{\max})\rangle = Cr^{\ell_{\max}}\mathrm{e}^{-r^2/4r_0^2}\,, \tag{6.98}$$

where $r_0 = \sqrt{\hbar/2\mu\omega}$ and C is a constant. Note that the exponential here is just the product of the exponentials we'd expect for Cartesian oscillators. This wavefunction varies with r, so quantum mechanically even a 'circular' orbit has some radial kinetic energy. In the limit of large $\ell_{\max}$ in which classical mechanics applies, the radial kinetic energy is negligible compared to the tangential kinetic energy. The full wavefunction is $\langle\mathbf{r}|\,(|E(\ell_{\max})\rangle \otimes |\ell_{\max}, m\rangle)$ and so involves the spherical harmonics $Y^m_{\ell_{\max}}(\theta,\phi)$. Bearing in mind that $\mathrm{d}^3\mathbf{x} = r^2\sin\theta\,\mathrm{d}\phi\,\mathrm{d}\theta\,\mathrm{d}r$, the radial probability $P(r)$ for this state is $P(r) \sim r^{2\ell_{\max}+2}\mathrm{e}^{-r^2/2r_0^2}$. For $r/r_0 \ll \sqrt{2\ell_{\max}+2}$ this rises as $r^{2\ell_{\max}+2}$ and it falls as the Gaussian takes over when $r/r_0 > \sqrt{2\ell_{\max}+2}$. The root-mean-square deviation in the radial location is $\sim r_0$, which is a small fraction of the expected radial location $\langle R\rangle$ when $\ell_{\max}$ is large, as expected classically.

We can obtain the radial wavefunctions of more eccentric orbits using $A^\dagger_\ell$. Taking the adjoint of (6.93) we obtain

$$H_\ell A^\dagger_\ell - A^\dagger_\ell H_{\ell+1} = \hbar\omega A^\dagger_\ell\,. \tag{6.99}$$

Acting this on $|E(\ell+1)\rangle$ gets

$$H_\ell(A^\dagger_\ell\,|E(\ell+1)\rangle) = (E+\hbar\omega)(A^\dagger_\ell\,|E(\ell+1)\rangle) \tag{6.100}$$

so that $A^\dagger_\ell\,|E(\ell+1)\rangle$ is a state of increased energy, whose radial wavefunction is appropriate for angular momentum ℓ. Thus $A^\dagger_\ell\,|E(\ell+1)\rangle = c\,|E(\ell)+\hbar\omega\rangle$ for some constant c. The radial wavefunctions of the eccentric orbits follow by writing this equation in the position representation.[46]

[46] If we introduce a radial quantum number n_r and label our state by $|n_r,\ell,m\rangle = |n_r,\ell\rangle \otimes |\ell,m\rangle$, then the effect of $A^\dagger_\ell$ is to shift from $|n_r,\ell+1\rangle$ to $|n_r+1,\ell\rangle$. The energy of the oscillator is $E = \hbar\omega(N+\frac{3}{2})$, where the total quantum number $N = 2n_r+\ell$. Therefore by acting $A^\dagger$, we increase the energy by $\hbar\omega$ as stated above.

In fact, knowing this allows us to determine the degeneracy of the 3d isotropic oscillator in a different way. For the Nth excited state, the allowed ℓ values are $\ell = N-2n_r = 0,2,\ldots,N$ if N is even and $1,3,\ldots,N$ if N is odd. Each ℓ value is $2\ell+1$ degenerate. Hence the total degeneracy is

$$g_N = \sum_{\substack{\ell=0\\ \ell\equiv N \pmod 2}}^{N} (2\ell+1) = \frac{(N+1)(N+2)}{2}\,, \tag{6.101}$$

which matches with the degeneracy we calculated previously.

7 Identical Particles

One of the most striking facts that helps us understand the Universe is that, apart from details of their motion, all electrons are identical. This fact, though convenient, has no explanation within Quantum Mechanics. In Quantum Field Theory, one learns that particles are simply quantised modes of excitation of a field. We thus replace the vast number of electrons our Universe contains with a single quantum field that has hugely many excitations. From this perspective, the fact that all elementary particles of the same species are identical is more or less a tautology, no more surprising than the fact that one factor of x in the monomial x^n is just the same as another.

7.1 Bosons and Fermions

Consider a system consisting of N particles, individually described by some state in a Hilbert space $\mathcal{H}_a$ where $a = 1, \ldots, N$, and suppose $|\alpha_a\rangle \in \mathcal{H}_a$ gives a complete specification of the state of the ath particle, so α_a collectively labels all the quantum numbers describing the particle's energy, orbital angular momentum, spin *etc.* In the case that the particles are distinguishable — meaning they are each from different species, such as an electron, a proton or a neutron — a complete set of states of the full system is then labelled by linear combinations of states

$$|\alpha_1; \alpha_2; \ldots; \alpha_N\rangle = |\alpha_1\rangle\,|\alpha_2\rangle \ldots |\alpha_N\rangle \in \bigotimes_{a=1}^{N} \mathcal{H}_a\,. \tag{7.1}$$

However, suppose particles 1 and 2 are indistinguishable; for example, they might both be electrons. The state (7.1) thus describes a situation where there is one electron with quantum numbers given by α_1 and another whose same properties are labelled by α_2. Since both electrons are identical, it can make no difference which is which. Consequently, there can be no physical difference between this state and the state in which these two particles are exchanged. This does not imply that the two states are identical, but does imply that they can only differ by a phase:

$$|\alpha_2; \alpha_1; \ldots\rangle = \mathrm{e}^{\mathrm{i}\phi}\,|\alpha_1; \alpha_2\rangle\ , \tag{7.2}$$

where ϕ is independent[47] of the particular values of the labels α_a, but may depend on the species of particle we exchange.

If we exchange the pair of particles once again we find

$$|\alpha_1; \alpha_2; \ldots\rangle = \mathrm{e}^{\mathrm{i}\phi}\,|\alpha_2; \alpha_1; \ldots\rangle = \mathrm{e}^{2\mathrm{i}\phi}\,|\alpha_1; \alpha_2; \ldots\rangle \tag{7.3}$$

using the fact that the phase is independent of the particular values of the quantum numbers α_a. Thus $\mathrm{e}^{2\mathrm{i}\phi} = 1$, so

$$\mathrm{e}^{\mathrm{i}\phi} = \pm 1\,. \tag{7.4}$$

The choice of sign depends only on the species of the particle. Identical particles which are symmetric under exchange are called *bosons*, while those that are antisymmetric are said to be *fermions*.

If all N of the particles in our system are identical, then the argument immediately extends to say that the state must be either completely symmetric or completely antisymmetric under exchange of any pair, which case being determined by whether the particles are bosons or fermions. That is, for bosons the state of the system is actually described by a vector

$$|\alpha_1; \alpha_2; \ldots; \alpha_N\rangle \in \mathrm{Sym}^N\,\mathcal{H} \subset \bigotimes^{N} \mathcal{H}\,, \tag{7.5}$$

[47] There's a (non-examinable) exception to this in the case of two spatial dimensions, where the phase may depend on the homotopy class of the path the particles take whilst they are being exchanged. In $2+1$ dimensions, these paths can become braided (tangled) and the multi-particle state transforms in a representation of the braid group, first studied in the maths literature by Emil Artin. This leads to the possibility of anyonic statistics, which play an important role in some condensed matter systems. Fortunately, we're interested in quantum mechanics on $\mathbb{R}^3$ and in higher dimensions it turns out the paths can always be disentangled, so anyons do not arise.

whereas for fermions the state lies in the antisymmetric part

$$|\alpha_1;\alpha_2;\ldots;\alpha_N\rangle \in \bigwedge^N \mathcal{H} \subset \bigotimes^N \mathcal{H}\,. \tag{7.6}$$

As a simple example, a pair of identical fermions with quantum numbers α_1 and α_2 live in the antisymmetric state

$$|\psi\rangle_{\text{fermi}} = \frac{1}{\sqrt{2}}\left(|\alpha_1\rangle\,|\alpha_2\rangle - |\alpha_2\rangle\,|\alpha_1\rangle\right)\,, \tag{7.7}$$

whereas the system would have to be described by the symmetric combination

$$|\psi\rangle_{\text{bose}} = \frac{1}{\sqrt{2}}\left(|\alpha_1\rangle\,|\alpha_2\rangle + |\alpha_2\rangle\,|\alpha_1\rangle\right)\,. \tag{7.8}$$

A further fact that also comes from quantum field theory is that particles with integer spin are always bosons, whilst particles with (odd) half integer spin are always fermions. Among the elementary particles, examples of bosons thus include the W and Z bosons and the photon, while examples of fermions include electrons, neutrinos and quarks. You'll learn about how this connection between spin and statistics arises if you take the Part III QFT course.

The above considerations apply to composite particles as well as to elementary ones. When we exchange a pair of identical composite particles, we exchange all of their constituents. Thus, if the composites are made up from an odd number of fermions and any number of bosons, upon exchange we'll acquire an odd number of minus signs, so the whole state must be antisymmetric. For example, a proton consists of three (valence) quarks and many gluons, so is a fermion. So too is a neutron. On the other hand, composites consisting of an even number of fermions and any number of bosons must be symmetric under exchange with identical composite. For example, a hydrogen atom consists of a proton and an electron, each of which are fermions, so the composite hydrogen atom is itself a boson.

The behaviour is consistent with addition of angular momentum. When we add the angular momenta of an odd number of particles each with odd half-integer spin, the total angular momentum is also an odd half-integer, whereas when we combine the angular momenta of an even number of particles each with odd half-integer spin, the total angular momentum is an integer. Even without quantum field theory, it would be impossible for all integer spin particles to be fermions, because a composite consisting of an even number of such integer spin particles would also have integer spin, but would be a boson.

7.1.1 Pauli's Exclusion Principle

The fact that the state of identical fermions must be totally antisymmetric has an immediate, striking consequence: no two fermions can be put in exactly the same state simultaneously. This fact is known as *Pauli's exclusion principle.* As an example, since we must always have $\psi_{\sigma,\sigma'}(\mathbf{x},\mathbf{x}') = -\psi_{\sigma',\sigma}(\mathbf{x}',\mathbf{x})$, if both fermions happen to have the same S_z eigenvalue, we find $\psi_{\sigma,\sigma}(\mathbf{x},\mathbf{x}') = -\psi_{\sigma,\sigma}(\mathbf{x}',\mathbf{x})$. Hence the spatial wavefunction of the system necessarily vanishes at $\mathbf{x} = \mathbf{x}'$, so there is zero probability that the two particles are located at the same place.

More generally, the state $|\Psi\rangle$ of N identical fermions is represented in terms of the determinant

$$\langle\alpha_1,\alpha_2,\ldots,\alpha_N|\Psi\rangle = \frac{1}{\sqrt{N!}}\begin{vmatrix} \psi_1(\alpha_1) & \psi_1(\alpha_2) & \cdots & \psi_1(\alpha_N) \\ \psi_2(\alpha_1) & \psi_2(\alpha_2) & \cdots & \psi_2(\alpha_N) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_N(\alpha_1) & \psi_N(\alpha_2) & \cdots & \psi_N(\alpha_N) \end{vmatrix} \tag{7.9}$$

that in this context are known as the *Slater determinants.* Again, $|\alpha_1,\alpha_2,\ldots,\alpha_N\rangle$ represents a state in which some fermion has quantum number α_1, another has α_2 and so on, where the labels α_i denote

a complete set of quantum numbers for single particle states, including, say, their momentum, spin-z value *etc.*, while $\psi_1(\alpha_1) = \langle\alpha_1|\psi_1\rangle$ is the amplitude for the "first" fermion to have quantum numbers α_1. If $\alpha_i = \alpha_j$ for any $i \neq j$, then the two columns of this determinant coincide. The amplitude $\langle\ldots,\alpha_i,\ldots,\alpha_i,\ldots|\Psi\rangle$ thus vanishes identically, so there is zero probability for any two fermions being found in an identical state.

It's worth stressing that we've learned only that the state of a pair of identical fermions must be antisymmetric under exchange of every one of their quantum numbers. The Hilbert space of a single such spin-s fermion is itself a tensor product $\mathcal{H}_{\text{fermion}} = \mathcal{H}_{\text{spat}} \otimes \mathbb{C}^{2s+1}$, including a factor that describes the fermion's spatial wavefunction (encoding details of its likely positions, orbital angular momentum *etc.*) and a separate factor $\mathcal{H}_{\text{spin}} \cong \mathbb{C}^{2s+1}$ spanned by the possible spin states $\{|s\rangle, |s-1\rangle, \ldots, |-s\rangle\}$ of this spin-s particle. We must exchange both the spin and spatial parts of the state in order to find a physically equivalent state.

In the example of our electron pair, $\psi_{\sigma,\sigma'}(\mathbf{x},\mathbf{x}')$ is certainly equivalent to $\psi_{\sigma',\sigma}(\mathbf{x}',\mathbf{x})$, since both corresponds to the electron at $\mathbf{x}$ having spin σ while the electron found at $\mathbf{x}'$ having spin σ'. However, antisymmetry of the total state does not imply any relation between states $\psi_{\sigma,\sigma'}(\mathbf{x},\mathbf{x}')$ and $\psi_{\sigma',\sigma}(\mathbf{x},\mathbf{x}')$: these are distinct physical possibilities because the electron at $\mathbf{x}$ has different spins in the two cases. We saw that the spin wavefunctions of the two electrons could be combined into a triplet of spin-1 states

$$|1,1\rangle = |\uparrow\rangle|\uparrow\rangle\ , \quad |1,0\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle|\downarrow\rangle + |\downarrow\rangle|\uparrow\rangle\right)\ , \quad |1,-1\rangle = |\downarrow\rangle|\downarrow\rangle \tag{7.10}$$

which are symmetric, and a single antisymmetric state

$$|0,0\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle|\downarrow\rangle - |\downarrow\rangle|\uparrow\rangle\right) \tag{7.11}$$

of spin-0. Because the overall state of the electron pair must be antisymmetric, the spatial wavefunction must be antisymmetric if the electron pair is in any of the spin-1 states, while the spatial wavefunction must be symmetric if the electron pair is in the spin-0 state. Thus, for a given spatial wavefunction $\psi(\mathbf{x},\mathbf{x}')$ the pair has four possible states

$$\frac{1}{\sqrt{2}}\left(\psi(\mathbf{x},\mathbf{x}') - \psi(\mathbf{x}',\mathbf{x})\right) \otimes \begin{cases} |1,1\rangle \\ |1,0\rangle \\ |1,-1\rangle \end{cases} \quad \text{and} \quad \frac{1}{\sqrt{2}}\left(\psi(\mathbf{x},\mathbf{x}') + \psi(\mathbf{x}',\mathbf{x})\right) \otimes |0,0\rangle\ , \tag{7.12}$$

which span a basis of the two-electron Hilbert space.

In general, as always, an electron pair can be in a superposition of the four states in (7.12), with different spatial wavefunctions in each term of the superposition. Similarly, an N-fermion system may be in state $|\Psi\rangle$ that is a superposition of different states $|\alpha_1,\alpha_2,\ldots,\alpha_N\rangle$, with different distinct values of the α_i's, where each of the amplitudes $\langle\alpha_1,\alpha_2,\ldots,\alpha_n|\Psi\rangle$ are given in terms of the single-particle states by Slater determinants.

7.2 The Periodic Table

One of the outstanding successes of the early years of quantum mechanics was an understanding of the structure of the Periodic Table of the Elements. The exact dynamics of a heavy atom, with large atomic number Z, is very complicated because each electron not only feels an attractive Coulomb potential $-Ze^2/r$ from the Z protons in the nucleus, but also a repulsive electromagnetic interaction from all the other electrons. However, it turns out that a reasonable approximation known as the *Hartree approximation* is to imagine that each electron moves in a (roughly) central potential $V(r)$ created by the net effect of the protons and the other electrons, while neglecting the fact that the response of any given electron to this potential will itself modify the potential. Near the nucleus,

$V(r) \approx -Ze^2/r$ as the electron feels the full strength of the attraction of the protons, while at large distances $V(r) \approx -e^2/r$ since the protons' charge is shielded by the $(Z-1)$ other electrons.

As for any central potential, $[\mathbf{L}^2, H] = 0$ and $[\mathbf{L}, H] = 0$ so we can label the single-electron states by their values of ℓ and m, together with another integer n labelling the energy levels. However, because $V(r)$ is not precisely $1/r$, unlike the pure Coulomb potential, states with different ℓ do not have identical energy. Because $V(r) \sim 1/r$ near the nucleus, the single electron wavefunctions behave as r^ℓ for small r, just as in hydrogen. This means that states with high values of ℓ are less likely to be found at small r where the potential is deepest, and so typically have slightly larger energy than those of lower ℓ, compared to the pure Coulomb case. Numerical calculations show that the states of roughly equal energy are

1s,			
2s,	2p,		
3s,	3p,		
4s,	3d,	4p,	
5s,	4d,	5p,	
6s,	4f,	5d,	6p,
7s,	5f,	7p,	...

with energy increasing as one proceeds down the list. Here, the number corresponds to the label n, while $\mathrm{s}, \mathrm{p}, \mathrm{d}, \mathrm{f}, \ldots$ stand for sharp, principal, diffuse, faint and simply corresponds to $\ell = 0, 1, 2, 3, \ldots$. This historic notation is standard in atomic spectroscopy, which is still of great importance in the pharmaceutical industry.

Since electrons have spin-$\frac{1}{2}$ they are fermions. Pauli's exclusion principle prevents the electrons in a heavy atom from all burying down into the lowest energy $1s$ state, and forces them to gradually fill up the higher energy levels. There are $2(2\ell+1)$ distinct states with orbital angular momentum ℓ, with the extra factor of 2 coming from the two possible spin states of the electron. Thus, the number of possible states in each line of the above table is 2, $2+6=8$, $2+6=8$, $2+10+6=18$, $2+10+6=18$ and $2+14+10+6=32$, *etc.*

The first two elements, hydrogen and helium, have electrons just in the ground state 1s. The 8 elements from lithium to neon have electrons in the $n=1$ and $n=2$ states, while the next 8 elements from sodium to argon have electrons in states with $n=1,2,3$. The first excited states of hydrogen and helium are obtained by promoting (one of) their 1s electrons up to the 2s state, and lie 10.2 eV and 19.8 eV above the corresponding ground states. These energy differences correspond to the frequency of UV radiation, so this transition cannot be stimulated by optical frequency radiation. By contrast, the first excited state of lithium is obtained by promoting a 2s electron to a 2p state, and lies a mere 1.85 eV above its ground state. This corresponds to the frequency of rather deep red. Elements that lie beyond helium play an important role in astronomical measurements, even though they are present only in trace amounts compared to hydrogen and helium, because their absorption spectra contain lines at easily observed optical frequencies.

The chemical properties of an element are largely determined by the number of electrons in its highest energy level, since these are least tightly bound. Atoms with no electrons outside filled energy levels are particularly stable chemically. These elements are called *noble gases* and include helium ($Z=2$), neon ($Z=2+8=10$), argon ($Z=2+8+8=18$), krypton ($Z=2+8+8+18=36$), xenon ($Z=2+8+8+18+18=54$) and radon ($Z=2+8+8+18+18+32=86$).

Elements with either a few more or few less electrons than required to fill a shell have their chemical properties determined by this number, known as the *valence* and counted positive for extra electrons and negative for fewer. If there is just one electron in the highest energy level, then this electron is easily stripped away, so the element is chemically reactive. These are the *alkali metals* and include lithium ($Z=2+1=3$), sodium ($Z=2+8+1=11$), potassium ($Z=2+8+8+1=19$) *etc.* If a large number of such atoms combine to form a solid crystal, then it is energetically favourable for the valence electrons to be shared throughout the solid rather than clinging to their original atom. These electrons thus form a sort of fluid that is free to flow within the crystal when stimulated by a small

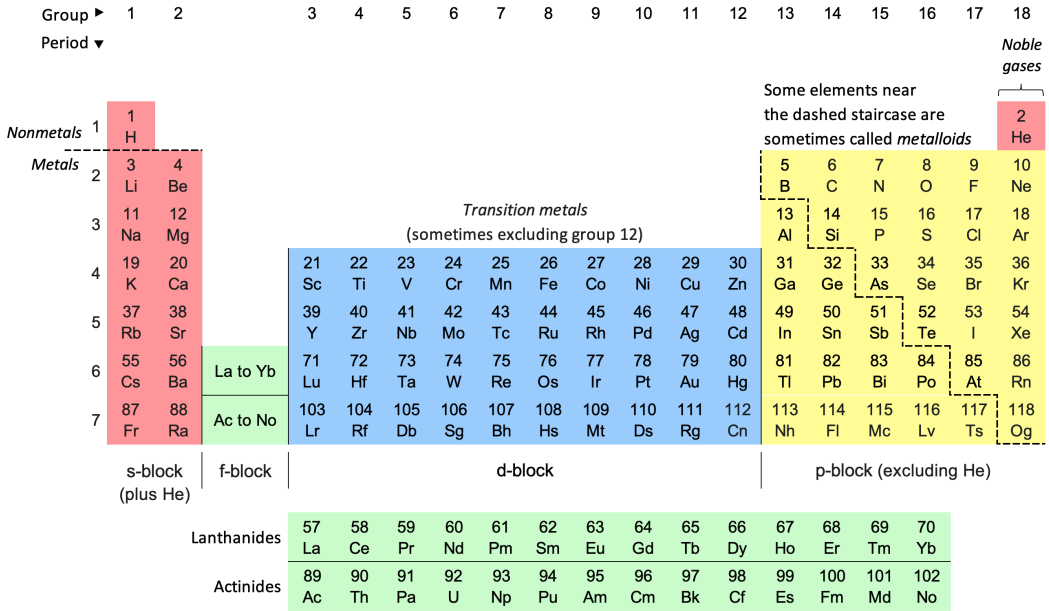


Figure 10: The Periodic Table of the Elements. For formatting reasons, the lanthanides and actinides are traditionally shown below the rest of the table. Figure adapted from Wikipedia.

external force. This gives the crystal high electrical and high thermal conductivity, making it a metal. (You'll study this in much more detail if you take the Applications of Quantum Mechanics course next term.) Atoms with two electrons more than the noble gases are also chemically reactive, though not as reactive as the alkalis. They are known as the *alkaline earth metals* and include beryllium ($Z = 2 + 2 = 4$), magnesium ($Z = 2 + 8 + 2 = 12$), calcium ($Z = 2 + 8 + 8 + 2 = 20$) *etc.*

On the other hand, atoms with one electron missing from their highest energy level tend to strongly attract other electrons and so are chemically reactive nonmetals, often reacting violently when brought into contact with metals. These elements are called *halogens* and include fluorine ($Z = 2+8-1 = 9$), chlorine ($Z = 2+8+8-1 = 17$), bromine ($Z = 2+8+8+18-1 = 35$), iodine ($Z = 2+8+8+18+18-1 = 53$) *etc.* Elements with two electrons missing from their highest energy level include oxygen ($Z = 2 + 8 - 2 = 8$), sulfur ($Z = 2 + 8 + 6 = 16$), *etc.* These elements are again chemically reactive, though less so than the halogens.

The inclusion of the 4f and 5f states in the sixth and seventh energy levels, respectively, are responsible for the long sequence of *rare earths* in the middle of the periodic table. Numerical calculations show that wavefunctions of the $2(2\times3+1) = 14$ different 4f states have small probability to lie outside the wavefunctions of the two 6s states, despite having slightly higher energy. Consequently, these elements are all rather similar chemically. The same is true of the 14 different 5f states compared to the wavefunctions of the two 7s states. The 14 elements in which the 4f states are being filled run from lanthanum ($Z = 2 + 8 + 8 + 18 + 18 + 2 + 1 = 57$) to ytterbium ($Z = 70$) and are known as *lanthanides* (or more officially, the stupid name *lanthanoids*), while the next chemically similar sequence are the *actinides* (or *actinoids*), running from actinium ($Z = 2+8+8+18+18+32+2+1 = 89$) to nobelium ($Z = 102$). Beyond this point the nuclei themselves become so unstable that the element tends to undergo radioactive decay before it has a chance to participate in any chemical reaction.

7.3 Exchange and Parity in the Centre of Momentum Frame

Let's now consider the effects of exchanging two identical particles on their spatial wavefunction. Letting $\mathbf{X}_1, \mathbf{P}_1$ and $\mathbf{X}_2, \mathbf{P}_2$ denote position and momentum operators of the two particles, exchanging

$1 \leftrightarrow 2$ implies that

$$\mathbf{X}_{\text{com}} = \frac{\mathbf{X}_1 + \mathbf{X}_2}{2} \qquad \longmapsto \qquad \frac{\mathbf{X}_2 + \mathbf{X}_1}{2} = \mathbf{X}_{\text{com}} \tag{7.13}$$

$$\mathbf{P}_{\text{com}} = \mathbf{P}_1 + \mathbf{P}_2 \qquad \longmapsto \qquad \mathbf{P}_2 + \mathbf{P}_1 = \mathbf{P}_{\text{com}}\,, \tag{7.14}$$

whilst

$$\mathbf{X}_{\text{rel}} = \mathbf{X}_1 - \mathbf{X}_2 \qquad \longmapsto \qquad \mathbf{X}_2 - \mathbf{X}_1 = -\mathbf{X}_{\text{rel}} \tag{7.15}$$

$$\mathbf{P}_{\text{rel}} = \frac{\mathbf{P}_1 - \mathbf{P}_2}{2} \qquad \longmapsto \qquad \frac{\mathbf{P}_2 - \mathbf{P}_1}{2} = -\mathbf{P}_{\text{rel}}\,. \tag{7.16}$$

Thus, exchange acts trivially on the centre-of-momentum coordinates, but acts on the relative coordinates just like a parity transformation. Since $Y_\ell^m(-\mathbf{x}) = (-1)^\ell Y_\ell^m(\mathbf{x})$ under parity, we see that if two identical particles have relative orbital angular momentum ℓ, the spatial part of the wavefunction will be either symmetric or antisymmetric under exchange according to whether ℓ is odd or even.

The behaviour of the entire state under exchange is determined by whether the particles in question are bosons or fermions. In the case of identical bosons, the spins must be combined to form a symmetric overall spin state when ℓ is even, or an antisymmetric overall spin state when ℓ is odd so as to ensure the overall state is always symmetric under exchange. For fermions, the opposite holds so as to ensure overall antisymmetry.

If the neutrons are produced in a state where their relative orbital angular momentum is ℓ in the centre of momentum frame, their spatial wavefunction acquires a factor of $(-1)^\ell$ under exchange of the two neutrons.

7.3.1 Identical Particles and Inelastic Collisions

The requirement that the state describing N identical bosons/fermions needs to be totally symmetric/antisymmetric under exchange of any pair of identical particles has important consequences in collision processes where such particles are created or destroyed.

For example, consider an exotic type of 'atom' consisting of a spin-0 pion (denoted by π^-) bound electromagnetically to a deuterium nucleus (denoted by D^+ — itself consisting of a proton and a neutron, bound together by the strong nuclear force). The bound state energy levels of this atom due to the electromagnetic Coulomb attraction between the π^- and D^+ have the same form as in Hydrogen and, in particular, the ground state is $|1,0,0\rangle$, an s-wave having $\ell = 0$. Because the pion is much heavier than the electron, the Bohr radius of the $D^+\pi^-$ 'atom' is much smaller than that of Hydrogen and the π^- wavefunction closely hugs the D^+ nucleus. The D^+ is known to have spin-1 whilst the π^- is spin zero, so the 1s ground state of the $D^+\pi^-$ 'atom' has total angular momentum $j = 1$.

Now, as well as their electromagnetic interactions, the π^- and D^+ also interact via the short-range strong nuclear force. This causes the 'atom' to be unstable, with the π^- rapidly being absorbed by the D^+, causing the system to disintegrate into a pair of neutrons (each denoted N):

$$\pi^- + D^+ \longrightarrow N + N\,. \tag{7.17}$$

(In particle physics terminology, processes such as this, where different types of particles appear in the initial and final states, are known as *inelastic*. An *elastic* process is one in which the initial and final states contain the same particles.)

Neutrons are fermions with spin-$\frac{1}{2}$, so the final state must be antisymmetric under their exchange. One possibility is for their spins to combine into one of the triplets

$$|\uparrow\rangle\,|\uparrow\rangle\,, \quad \frac{1}{\sqrt{2}}\left(|\uparrow\rangle\,|\downarrow\rangle + |\downarrow\rangle\,|\uparrow\rangle\right), \quad |\downarrow\rangle\,|\downarrow\rangle \tag{7.18}$$

of symmetric spin-1 states, combined with a spatial wavefunction that is antisymmetric under exchange. From above, this will be the case if the relative orbital angular momentum of the two neutrons is odd. The other possibility is for the spins to form the antisymmetric spin-0 state

$$\frac{1}{\sqrt{2}}\left(|\uparrow\rangle\,|\downarrow\rangle - |\downarrow\rangle\,|\uparrow\rangle\right) \tag{7.19}$$

combined with a relative spatial wavefunction of even ℓ.

Total angular momentum is conserved, so the spin and orbital angular momentum of the final state must combine to give $j = 1$ as for the initial ground state of the $\pi^- D^+$. A combined system with angular momenta ℓ and s have total angular momentum $j \in \{\ell + s, \ell + s - 1, \ldots, |\ell - s|\}$ depending on the relative alignment of the two subsystems. Thus, in the case that the neutrons spins combine to the antisymmetric spin-0 state, $j = \ell = 1$. However, since this case requires ℓ even for fermionic statistics, we see it is ruled out. The remaining case is that the neutrons combine to give the net spin-1, so that

$$j = 1 \in \{\ell + 1, \ell, \ell - 1\}\,. \tag{7.20}$$

Since fermionic statistics here require that ℓ is odd, the only possibility to get $j = 1$ is if $\ell = 1$ and the spin and orbital angular momenta of the neutron pair are neither perfectly aligned nor perfectly anti-aligned. Thus the final state has $j = \ell = s = 1$.

These considerations also help us to determine the intrinsic parity of the pion. Recall that the parity operator Π acts on a state $|\mathbf{x}\rangle$ as

$$\Pi\,|\mathbf{x}\rangle = \eta\,|\mathbf{x}\rangle\ , \tag{7.21}$$

where the value of $\eta \in \{+1, -1\}$ is known as the *intrinsic parity* that, like spin, depends on the type of particle that the state $|\mathbf{x}\rangle$ represents. More generally, if $|\mathbf{x}_1, \kappa_1; \mathbf{x}_2, \kappa_2; \ldots; \mathbf{x}_N, \kappa_N\rangle$ describe an N-particle state in which a particle of type κ_a is definitely located at $\mathbf{x}_a$, then the parity operator acts as

$$\Pi\,|\mathbf{x}_1, \kappa_1; \mathbf{x}_2, \kappa_2; \ldots; \mathbf{x}_N, \kappa_N\rangle = \eta_1\eta_2\ldots\eta_N\,|\mathbf{x}_1, \kappa_1; \mathbf{x}_2, \kappa_2; \ldots; \mathbf{x}_N, \kappa_N\rangle\ , \tag{7.22}$$

where $\eta_k = \pm 1$ is the intrinsic parity of the kth particle that only depends on κ_k. This holds whether or not the particles are distinguishable.

Like the spin of fundamental particles, these intrinsic parities are independent of any details of the spatial wavefunction. Intrinsic parities thus have no effect in elastic processes, where the same particles are present in both the initial and final states. The intrinsic parity of a particle can often be determined by examining inelastic processes in which the particle participates.

In the example $\pi^- D^+ \to NN$ above, equating the parities gives

$$\eta_\pi \eta_D = (-1)^\ell \eta_N^2 = -1 \tag{7.23}$$

since the initial atomic state $|1, 0, 0\rangle$ has $\ell = 0$ and hence no parity other than the intrinsic parities. Provided parity is conserved by the strong nuclear interactions causing this decay, we conclude that the π^- and D^+ must have opposite intrinsic parity. Now, the D^+ is predominantly an s-wave bound state of a proton and neutron, so $\eta_D = (-1)^\ell \eta_P \eta_N = \eta_P \eta_N$, and furthermore the proton and neutron can always be chosen to have the same intrinsic parity since they are related by an 'isospin' symmetry[48]. Thus $\eta_D = +1$ and hence the pion must have intrinsic parity $\eta_\pi = -1$.

As we mentioned in section 4.6, one of the great surprises of particle physics came in the 1950's. Cosmic rays were found to contain various types of particles, including two of similar mass called the τ^+ and the θ^+. Both particles decay quickly, the predominant channels being

$$\theta^+ \longrightarrow \pi^+ + \pi^0 \qquad \text{and} \qquad \tau^+ \longrightarrow \pi^+ + \pi^+ + \pi^-\,. \tag{7.24}$$

[48] The standard model has an (approximate) 'internal' SU(2) symmetry known as *isospin* that rotates protons into neutrons; they are different states of the same *nucleon*, somewhat like the two different spin states $|\uparrow\rangle, |\downarrow\rangle$ of the same electron

The angular distribution of the pions in the final states could be observed in a cloud chamber or bubble chamber, and studying these patterns showed that the pions were produced in an $\ell = 0$ state in both cases. Since $\eta_\pi = -1$ (irrespective of the electric charge of the pion, for the same isospin reason as above), the θ^+ should have intrinsic parity $\eta_\theta = \eta_\pi^2 = +1$, while the τ^+ should have $\eta_\tau = -1$. However, as measurements improved the masses and lifetimes of the two particles became indistinguishable. The puzzle was resolved in 1956 when Lee and Yang proposed that the two particles were in fact one and the same, but that parity was not conserved in their decay process. Their proposal was largely ignored until Lee persuaded his colleague, Madame Chien-Shiung Wu, to test it using the decay of a certain isotope of cobalt. Wu's experiments showed that parity is indeed violated in the weak interactions. The fact that $\mathbf{x} \to -\mathbf{x}$ is not a symmetry of Nature can be accommodated in QFT, though the deep reason for it remains mysterious[49].

[49]There are various natural ways for parity violation to originate in string theory, but needless to say none of them have been verified experimentally.

8 Perturbation Theory I: Time Independent Case

We've now come about as far as we can (in this course) relying purely on symmetry principles. The dynamics of systems of genuine physical interest is rarely simple enough to hope that we can solve it exactly, just using the general constraints of symmetry. In these circumstances we need to make some sort of approximation, treating our system as being 'close to' some other system whose dynamics is sufficiently simple to be controllable. That is, we treat the difference between our actual system, whose experimental properties we care about, and our model system, whose description is simple enough that we can handle it, as a *perturbation*. The whole art of a theoretical physicist lies in striking this balance; if one's model system is too simplistic, it might not provide a reasonable guide to the behaviour of the real case[50], while if the model is overly complicated it may itself prove impossible to understand.

There's nothing inherently quantum mechanical about the need to approximate a complicated system by a simpler one, but, fortunately, perturbation theory in quantum mechanics often turns out to be considerably easier than in classical dynamics, largely because of the vector space nature of $\mathcal{H}$. In this chapter and the next, we'll study various techniques to handle quantum mechanical perturbation theory, beginning with the simplest cases.

8.1 An Analytic Expansion

Let H be the Hamiltonian of the experimental system we wish to understand, and H_0 be the Hamiltonian of our model system whose eigenstates and eigenvalues we already know. We hope that $\Delta H = H - H_0$ is in some sense 'small', so it may be treated as a perturbation. More specifically, we look for a parameter — let's call it λ — that our true Hamiltonian depends on such that at $\lambda = 0$, $H = H_0$. For $\lambda \in [0, 1]$, define

$$H_\lambda = H_0 + \lambda \Delta H \,. \tag{8.1}$$

We can think of H_λ as the Hamiltonian of an apparatus that is equipped with a dial that allows us to vary λ. At $\lambda = 0$ the system is our model case, and at $\lambda = 1$ it's the case of genuine interest.

We now seek the eigenstates $|E_\lambda\rangle$ of H_λ. This may look as though we've made the problem even harder — we now need to find the eigenstates not just of our model and experimental systems, but of a 1-parameter family of interpolating systems. Our key assumption is that since H_λ depends analytically on λ, so too do its eigenstates. In essence, this amounts to the assumption that small changes in the system will lead to only small changes in the outcome. Every mountain climber knows that this assumption can be dreadfully false and we'll see that it can easily fail in QM too, but for now let's see where it takes us.

If indeed $|E_\lambda\rangle$ depends analytically in λ, then we can expand it as

$$|E_\lambda\rangle = |\alpha\rangle + \lambda\,|\beta\rangle + \lambda^2\,|\gamma\rangle + \dots \tag{8.2}$$

and similarly expand the eigenvalues

$$E(\lambda) = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \dots \,. \tag{8.3}$$

Plugging these expansions into the defining equation $H_\lambda\,|E_\lambda\rangle = E(\lambda)\,|E_\lambda\rangle$ we obtain

$$\begin{aligned}(H_0 + \lambda\Delta H)(|\alpha\rangle + \lambda\,|\beta\rangle + \lambda^2\,|\gamma\rangle + \dots) \\ = \left(E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \dots\right)(|\alpha\rangle + \lambda\,|\beta\rangle + \lambda^2\,|\gamma\rangle + \dots)\,.\end{aligned} \tag{8.4}$$

[50] A classic joke. Milk production at a dairy farm was low, so the farmer wrote to the local university, asking for help from academia. A multidisciplinary team of professors was assembled, headed by a theoretical physicist, and two weeks of intensive on-site investigation took place. The scholars then returned to the university, notebooks crammed with data, where the task of writing the report was left to the team leader. Shortly thereafter the physicist returned to the farm, saying to the farmer, "I have the solution, but it works only in the case of spherical cows in a vacuum."

Since we require this to hold as λ varies, it must hold for each power of λ separately. Thus we find an infinite system of equations

$$H_0 |\alpha\rangle = E^{(0)} |\alpha\rangle \ , \tag{8.5}$$

$$H_0 |\beta\rangle + \Delta H |\alpha\rangle = E^{(0)} |\beta\rangle + E^{(1)} |\alpha\rangle \ , \tag{8.6}$$

$$H_0 |\gamma\rangle + \Delta H |\beta\rangle = E^{(0)} |\gamma\rangle + E^{(1)} |\beta\rangle + E^{(2)} |\alpha\rangle \ , \tag{8.7}$$

$$\vdots$$

The first of these equations simply states that $|\alpha\rangle$ is an eigenstate of the model Hamiltonian H_0 with eigenvalue $E^{(0)}$. This is not surprising; under our analytic assumptions the terms of $O(\lambda^0)$ are all that would survive when the dial is set to $\lambda = 0$, which is indeed the model system. Henceforth, we relabel $|\alpha\rangle \to |n\rangle$ and $E^{(0)} \to E_n$ to reflect this understanding. (The notation n is intended to imply the nth energy eigenstate of the model Hamiltonian H_0, with eigenvalue E_n; we will distinguish the higher-order corrections to these states with superscripts.)

To determine the first-order correction $E_n^{(1)}$ to the nth energy level of the unperturbed Hamiltonian, we contract the second equation (8.6) with $|\alpha\rangle = |n\rangle$ to find

$$\langle n|H_0|\beta\rangle + \langle n|\Delta H|n\rangle = E_n \langle n|\beta\rangle + E_n^{(1)} \ . \tag{8.8}$$

Since the model Hamiltonian is Hermitian,

$$\langle n|H_0|\beta\rangle = \langle \beta|H_0|n\rangle^* = E_n \langle n|\beta\rangle \ , \tag{8.9}$$

so in the case that the unperturbed state is $|n\rangle$, the equation becomes

$$E_n^{(1)} = \langle n|\Delta H|n\rangle \ . \tag{8.10}$$

In other words, to first order in λ, the change in the energy of our system as we move away from the model system is given by the expectation value $\langle \Delta H \rangle_n$ of the change in the Hamiltonian when the system is in its original state $|n\rangle$.

To find the perturbed state to first order in λ, we must understand $|\beta\rangle$. We can expand $|\beta\rangle$ in the complete set $|n\rangle$ of eigenstates of the original system as

$$|\beta\rangle = \sum_n b_n |n\rangle \ . \tag{8.11}$$

Using this expression in the second of equations (8.6) and contracting with $\langle m|$, where $\langle m| \neq \langle n|$, the initial state we're perturbing, gives

$$(E_n - E_m) b_m = \langle m|\Delta H|n\rangle \tag{8.12}$$

and so, provided $E_m \neq E_n$,

$$b_m = \frac{\langle m|\Delta H|n\rangle}{E_n - E_m} \ . \tag{8.13}$$

This will hold provided the energy levels of our model system are non-degenerate; we'll examine how to handle the more general case including degeneracy in section 8.2. In the non-degenerate case, equation (8.13) determines all the expansion coefficients b_m in $|\beta\rangle$ except b_n. Fortunately, one can argue that $b_n = 0$ from the requirement that $|E_\lambda\rangle$ remains correctly normalised. This is left as an exercise, but the essential idea is that if we move a point on a unit sphere, then to first order $\mathbf{r} \cdot \delta\mathbf{r} = 0$ so that the variation is only nonzero in directions orthogonal to the original vector. With $b_n = 0$ we have

$$|\beta\rangle = \sum_{m \neq n} \frac{\langle m|\Delta H|n\rangle}{E_n - E_m} |m\rangle \tag{8.14}$$

as the first-order perturbation of the state when the unperturbed state is $|n\rangle$.

We can also examine the second-order perturbation, $E_n^{(2)}$, of the nth energy level. To do so, contract the third of equations (8.7) with $\langle n|$. Using the facts that $\langle n|H_0|\gamma\rangle = E_n \langle n|\gamma\rangle$ and $\langle n|\beta\rangle = 0$ we have

$$\begin{aligned} E_n^{(2)} = \langle n|\Delta H|\beta\rangle &= \sum_{m\neq n} \frac{\langle n|\Delta H|m\rangle \langle m|\Delta H|n\rangle}{E_n - E_m} \\ &= \sum_{m\neq n} \frac{|\langle n|\Delta H|m\rangle|^2}{E_n - E_m} \end{aligned} \tag{8.15}$$

using our expression for $|\beta\rangle$. We could go on to higher order in λ, next finding $|\gamma\rangle$ in terms of the original states $\{|n\rangle\}$ and then finding the third-order energy shift $E_n^{(3)}$ *etc.*, but in practice the summations become increasingly messy and we hope that the first few terms already provide a good guide to the behaviour of the system near $\lambda = 1$. (In high-energy quantum field theory, modern experiments typically cost many millions of dollars so it's especially important to have extremely accurate theoretical predictions to compare to. Consequently, there's a whole industry of people whose life's work is to compute higher and higher order terms in perturbation series such as these.) Fortunately, in the Tripos you'll never be asked to do anything beyond 2nd order.

Combining our results shows that, to second order in λ, the energy levels of the perturbed system are given by

$$E_n(\lambda) = E_n + \lambda \langle n|\Delta H|n\rangle + \lambda^2 \sum_{m\neq n} \frac{|\langle n|\Delta H|m\rangle|^2}{E_n - E_m} + O(\lambda^3)\,, \tag{8.16}$$

where E_n are the energies of our model system. Recall that this expression — much beloved of Tripos examiners — is derived under the assumptions (i) that the new energies $E(\lambda)$ and new states $|E_\lambda\rangle$ are analytic at $\lambda = 0$ and (ii) that the model system is non-degenerate so $E_m = E_n$ iff $|m\rangle = |n\rangle$.

Let's now take a look at the use of this formula in a number of examples.

8.1.1 Fine Structure of Hydrogen

Our treatment of the hydrogen atom in the last chapter assumed the electron was moving non relativistically in the Coulomb field of the proton. Since $|E/\mu c^2| = \alpha^2/2n^2 \ll 1$, non-relativistic quantum mechanics should indeed be a good approximation. Nonetheless, better agreement with experiment is obtained by describing the electron using the relativistic Dirac equation[51]. The energy levels obtained by assuming non-relativistic motion in a pure Coulomb potential are often called the *gross structure* of the hydrogen atom, whilst the small relativistic corrections to these energies implied by the Dirac equation are known as fine structure. We understood that gross structure energies are independent of ℓ because of an enhanced symmetry of the pure Coulomb potential, generated by the Runge–Lenz operator. We thus expect that relativistic corrections will lift the degeneracy among states of different ℓ.

The Dirac equation itself is beyond the scope of this course, but some of its consequences are easy to understand in perturbation theory. Firstly, expanding the relativistic dispersion relation around the non-relativistic limit gives

$$E = \sqrt{\mathbf{p}^2c^2 + \mu^2c^4} \approx \mu c^2 + \frac{\mathbf{p}^2}{2\mu} - \frac{\mathbf{p}^4}{8\mu^3c^2} + \dots \tag{8.17}$$

This shows that the relativistic correction to the kinetic energy is

$$\Delta H = -\frac{\mathbf{p}^4}{8\mu^3c^2} \tag{8.18}$$

[51] A better approximation still — in precise agreement with the most accurate measurement ever performed in any branch of science — comes from quantum field theory.

at first order. From (8.10), the first-order shift in the energy of state $|n,\ell,m\rangle$ of hydrogen due to this modified kinetic term is thus

$$E^{(1)}_{n,\ell} = \langle n,\ell,m|\Delta H|n,\ell,m\rangle \ . \tag{8.19}$$

(In making this claim, we should note that our perturbation (8.18) is rotationally invariant, so in particular $[L_z, \Delta H] = 0$ and $[\mathbf{L}^2, \Delta H] = 0$. Therefore $\langle n,\ell',m'|\Delta H|n,\ell,m\rangle = 0$ unless $\ell' = \ell$ and $m' = m$. Thus our perturbation does not mix degenerate states of the gross structure, so non-degenerate perturbation theory is sufficient.)

To compute the first order energy correction, first we note that

$$\Delta H = -\frac{(H_0 - V(r))^2}{2\mu c^2} \ , \tag{8.20}$$

where H_0 is the gross structure Hamiltonian and $V(r)$ is the usual Coulombic potential. Then

$$E^{(1)}_{n,\ell} = -\frac{1}{2\mu c^2}[(E_n)^2 - 2E_n \langle V(r)\rangle + \langle V(r)^2\rangle] \ , \tag{8.21}$$

where $E_n = -\frac{1}{2}\mu c^2\alpha^2/n^2$ is the energy of the unperturbed state and $\langle V(r)\rangle = \langle n,\ell,m|V(r)|n,\ell,m\rangle$. The virial theorem for the $1/r$ potential tells us that $2\langle T\rangle = -\langle V\rangle$, so $\langle V\rangle = 2E_n$. Next, to compute $\langle V^2\rangle$, observe that this $1/r^2$ term just modifies the effective potential due to the electron's orbital angular momentum. Consider the Hamiltonian for the radial part of the Hydrogen atom, but now with an effective potential

$$V_{\text{eff}}(r) = \frac{\hbar^2}{2\mu}\frac{\ell(\ell+1)}{r^2} + \frac{\alpha}{r} - \frac{e^2}{4\pi\epsilon_0 r} = \frac{\hbar^2}{2\mu}\frac{\ell'(\ell'+1)}{r^2} - \frac{e^2}{4\pi\epsilon_0 r} \ . \tag{8.22}$$

Here we've introduced ℓ' to absorb the extra α/r^2 term into the angular momentum contribution. Of course ℓ' does not correspond to any actual angular momentum. It's just a trick to help us compute $\langle V^2\rangle$. With this effective potential, repeating the calculations of section 6.3 would lead to just the same energy levels as before, but now in terms of (non-integer) ℓ' instead of ℓ:

$$E(\ell') = -\frac{1}{2}\mu c^2\alpha^2\frac{1}{(\ell'+1)^2} \ . \tag{8.23}$$

This is the exact result for a hydrogen atom with an additional $1/r^2$ term in its potential. However, in our perturbative context, it's only appropriate to keep the answer accurate to first order — there may be other effects we haven't yet accounted for (such as expanding (8.17) further) that contribute at higher order. Expanding ℓ' around ℓ we find

$$E(\ell') = E_n + \frac{1}{2}\mu c^2\frac{\alpha^4}{n^3(\ell+\frac{1}{2})} + \dots \tag{8.24}$$

where $n = \ell + 1$. Combining this with the other terms gives

$$E^{(1)}_{n,\ell} = -\frac{1}{2}\mu c^2\left(\frac{n}{\ell+\frac{1}{2}} - \frac{3}{4}\right)\frac{\alpha^4}{n^4} \tag{8.25}$$

as the first-order change in the energy of state $|n,\ell,m\rangle$ due to the relativistic correction to kinetic energy. Notice that this effect is suppressed by an additional power of $\alpha^2 \sim v^2/c^2$ compared to the gross structure, and that the degeneracy among states of different ℓ is lifted, as we expected.

There's a second consequence of the Dirac's equation which comes in at the same order as the above kinetic terms. Due to its spin, the electron has a magnetic dipole moment

$$\mathbf{m} = -\frac{e}{2\mu}\mathbf{S} \ . \tag{8.26}$$

When placed in a magnetic field, the dipole has energy $U = -\mathbf{m}\cdot\mathbf{B}$. Relativistically, the electron in a Hydrogen atom does experience a magnetic field because it moves through the electric field

$\mathbf{E}$ produced by the proton. If the electron has velocity $\mathbf{v} = \mathbf{p}/\mu\gamma$, then classically the Lorentz transformation of $\mathbf{E}$ produces a magnetic field

$$\mathbf{B} = \frac{\gamma}{c^2}\mathbf{v} \times \mathbf{E} = \frac{1}{\mu c^2}\mathbf{p} \times \left(\frac{e\hat{\mathbf{r}}}{4\pi\epsilon_0 r^2}\right) = -\frac{e}{4\pi\epsilon_0 \mu c^2}\frac{1}{r^3}\mathbf{L}\,. \tag{8.27}$$

Consequently, in QM the Hamiltonian of hydrogen receives a further fine-structure correction

$$H_{\mathrm{SO}} = -\mathbf{m}\cdot\mathbf{B} = \frac{\alpha\hbar^3}{2\mu^2 c}\frac{1}{r^3}\mathbf{S}\cdot\mathbf{L} \tag{8.28}$$

known as the *spin-orbit coupling*. Since the coefficient of the operator $\mathbf{S}\cdot\mathbf{L}$ is positive, spin-orbit coupling lowers the total energy when the spin and orbital angular momentum are antiparallel. In particular, $\mathbf{S}\cdot\mathbf{L}$ vanishes in any state of hydrogen in which $\ell = 0$, since then all components of $\mathbf{L}$ act trivially. As an important special case, spin-orbit coupling does not affect the ground state. However, excited states with $\ell \neq 0$ do generically feel the effects of spin-orbit coupling.

We first compute the effect of acting on our electron states with $\mathbf{L}\cdot\mathbf{S}$. To do this, clearly we must include a specification of the electron's spin. This could be done using the basis

$$\{|n,\ell,m\rangle \otimes |\uparrow\rangle\,, |n,\ell,m\rangle \otimes |\downarrow\rangle\}\,, \tag{8.29}$$

but it turns out to be more convenient to instead combine the electron's orbital and spin angular momenta to $\mathbf{J} = \mathbf{L} + \mathbf{S}$ and label states by[52] $|n,j,m_j;\ell\rangle$ where $j = \ell + \frac{1}{2}$ or $j = \ell - \frac{1}{2}$ are the two possible values for the combined angular momentum, and $|m_j| \leq j$. To make use of this, we write

$$\mathbf{S}\cdot\mathbf{L} = \frac{1}{2}(\mathbf{J}^2 - \mathbf{L}^2 - \mathbf{S}^2)\,. \tag{8.30}$$

Thus

$$\mathbf{S}\cdot\mathbf{L}\,|n,j,m_j;\ell\rangle = \frac{\hbar^2}{2}\left(j(j+1) - \ell(\ell+1) - \frac{3}{4}\right)|n,j,m_j,\ell\rangle \tag{8.31}$$

$$= \frac{\hbar^2}{2}\begin{cases}\ell\,|n,j,m_j;\ell\rangle & \text{when } j = \ell + \frac{1}{2}\\ -(\ell+1)\,|n,j,m_j;\ell\rangle & \text{when } j = \ell - \frac{1}{2}\,.\end{cases} \tag{8.32}$$

Consequently, the first order change in the energy of the $|n,j,m_j;\ell\rangle$ state of hydrogen due to spin-orbit coupling is

$$E^{(1)}_{n,j;\ell} = -\frac{1}{4\mu^2c^2}\frac{e^2\hbar^2}{4\pi\epsilon_0}\begin{Bmatrix}\ell\\ -(\ell+1)\end{Bmatrix}\left\langle\frac{1}{r^3}\right\rangle_{n,j;\ell}\,, \tag{8.33}$$

where the factor of ℓ or $-(\ell+1)$ chosen depending on whether $j = \ell \pm \frac{1}{2}$.

To compute the expectation value $\left\langle 1/r^3\right\rangle_{n,j;\ell}$ we could just perform the integral

$$\int_{\mathbb{R}^3} \mathrm{d}^3\mathbf{x}\;|\psi_{n,l,m}(\mathbf{x})|^2\,\frac{1}{r^3}\,. \tag{8.34}$$

However, there is a trick that allows us to short-circuit the evaluation of this integral. Recall that for states of definite ℓ, the Hamiltonian for the unperturbed atom can be written

$$H_\ell = -\frac{P_r^2}{2\mu} + \frac{\hbar^2\ell(\ell+1)}{2\mu r^2} - \frac{e^2}{4\pi\epsilon_0 r}\,, \tag{8.35}$$

where $P_r = (\hat{\mathbf{X}}\cdot\mathbf{P} + \mathbf{P}\cdot\hat{\mathbf{X}})/2$ is the radial momentum given by

$$P_r = -\mathrm{i}\hbar\left(\frac{\partial}{\partial r} + \frac{1}{r}\right) \tag{8.36}$$

[52] Since $[\mathbf{J},\mathbf{L}^2] = 0$, it's possible to find a basis of simultaneous eigenstates of $\mathbf{J}^2$, $\mathbf{J}_z$ and $\mathbf{L}^2$. In this basis, we do not necessarily know the z-components of the electron's orbital and spin angular momenta separately.

in the position representation. Now, the expectation value $\langle [P_r, H_\ell] \rangle = 0$ when evaluated in any normalisable eigenstate of H_ℓ. In our case, the commutator evaluates to

$$[P_r, H_\ell] = -\mathrm{i}\hbar \left(-\frac{\hbar^2 \ell(\ell+1)}{\mu r^3} + \frac{e^2}{4\pi\epsilon_0 r^2} \right) \tag{8.37}$$

and therefore, provided $\ell \neq 0$,

$$\left\langle \frac{1}{r^3} \right\rangle_{n,j;\ell} = \frac{1}{a_0} \frac{1}{\ell(\ell+1)} \left\langle \frac{1}{r^2} \right\rangle_{n,j;\ell} = \frac{1}{a_0^3} \frac{1}{\ell(\ell+\frac{1}{2})(\ell+1)} \frac{1}{n^3} , \tag{8.38}$$

where $a_0 = \hbar/\alpha\mu c$ is the Bohr radius and the second equality follows from our previous result for $\langle 1/r^2 \rangle$.

Putting all these together, the first-order energy shifts due to spin-orbit coupling are

$$E^{(1)}_{n,j;\ell} = +\frac{1}{2}\alpha^4 \mu c^2 \frac{1}{2n^3(\ell+\frac{1}{2})(\ell+1)} \tag{8.39}$$

when $j = \ell + \frac{1}{2}$, whereas when $j = \ell - \frac{1}{2}$ we have

$$E^{(1)}_{n,j;\ell} = -\frac{1}{2}\alpha^4 \mu c^2 \frac{1}{2n^3 \ell(\ell+\frac{1}{2})} . \tag{8.40}$$

(Also recall that $E^{(1)}_{n,j;\ell} = 0$ if $\ell = 0$.) As anticipated, states whose spin and orbital angular momenta are anti-aligned ($j = \ell - \frac{1}{2}$) have lower energy and so are more tightly bound than the pure Coulomb interaction, whereas if the spin and orbital angular momentum are aligned with each other ($j = \ell + \frac{1}{2}$) the state is less tightly bound. Notice also that the deviation from a pure Coulombic interaction partially lifts the degeneracy between states of different ℓ.

We can also see that the spin-orbit contribution to the energy is suppressed by a factor of α^2 compared to the gross structure energy $-\frac{1}{2}\mu\alpha^2 c^2/n^2$. This is the same suppression as we found for the $\mathbf{P}^4$ correction to the kinetic term, so it does not really make sense to consider these two effects on the energy separately. Combining (8.25) with (8.39) and (8.40), we find the net energy levels

$$E_{n,j;\ell} = -\frac{1}{2}\mu\alpha^2 c^2 \left[\frac{1}{n^2} - \frac{\alpha^2}{n^3} \left(\frac{3}{4n} - \frac{1}{j+\frac{1}{2}} \right) + \dots \right] \tag{8.41}$$

incorporating fine structure correct to order α^4. This formula holds regardless of whether $j = \ell + \frac{1}{2}$ or $j = \ell - \frac{1}{2}$. Even better, although the spin-orbit term contributes only when $\ell \neq 0$, it turns out that there is another term[53] that contributes only when ℓ does equal zero! Even better, this remaining term has the effect that (8.41) is the correct fine structure energy levels even for states with $\ell = 0$ where $j = \frac{1}{2}$ just from the spin.

Proceeding down the periodic table, if we attach a single electron to an atomic nucleus containing Z protons, the bare Coulomb interaction increases by a factor of Z. The formula (8.41) remains valid provided we replace $\alpha \sim e^2$ by $Z\alpha$. Hydrogen has $Z = 1$, so in the $n = 1$ ground state the fine structure contribution is suppressed by a factor of $\alpha^2 \sim 1/(137)^2$ compared to the gross structure, justifying our treatment of the fine structure as a small perturbation. However, for heavier elements with higher Z, the suppression is only by $(Z\alpha)^2$ and the small value of α can be negated by a sufficiently large Z. In particular, from (8.41) we see that the energy difference between the states with principal quantum number n but $j = \ell \pm \frac{1}{2}$ is

$$E_{n,\ell+\frac{1}{2};\ell} - E_{n,\ell-\frac{1}{2};\ell} = \frac{1}{2}\mu c^2 \frac{1}{n^3} \frac{Z^4\alpha^4}{\ell(\ell+1)} . \tag{8.42}$$

[53] This remaining relativistic effect is known as the Darwin term and gives a contribution $\frac{\alpha\pi\hbar^3}{2\mu^2 c}\delta(\mathbf{x})$. The origin of this term is rather subtle, and properly lies within QFT rather than QM. Because of the δ-function, it only affects states which have non-zero probability to be found at $r = 0$. The only such states are those with $\ell = 0$, since all others are prevented from reaching the origin by the effective potential $\ell(\ell+1)/2mr^2$.

This gives the splitting between the two 2p states of Hydrogen to be 4.53×10^{-5} eV, in excellent agreement with the measured value of 4.54×10^{-5} eV. These splittings reach around 10% of the gross energy as one reaches the middle of the periodic table, making perturbative treatments less reliable.

8.1.2 Hyperfine Structure of Hydrogen

A proton is a spin-$\frac{1}{2}$ particle, so like the electron it has a magnetic dipole moment $\mathbf{m}_p$ proportional to its spin. A magnetic dipole $\mathbf{m}_p$ placed at the origin generates its own magnetic field

$$\mathbf{B} = \frac{2\mu_0}{3}\mathbf{m}_p\delta(\mathbf{x}) + \frac{\mu_0}{4\pi r^3}(3(\mathbf{m}_p \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{m}_p)\,, \tag{8.43}$$

which is roughly the field you saw traced out by iron lings when first playing with magnets. The dipole moment $\mathbf{m}_e$ of the electron sits inside this field, leading to a new contribution

$$H_{\text{hfs}} = -\mathbf{m}_e \cdot \mathbf{B} \tag{8.44}$$

depending on both dipoles. We'll just investigate the effect of this perturbation on the $\ell = 0$ state of hydrogen, where it turns out that only the first term $\sim \delta(\mathbf{x})$ in $\mathbf{B}$ has non-vanishing contribution. For these $\ell = 0$ states, one finds

$$H_{\text{hfs}} = \frac{4}{3}\frac{m}{M}\alpha^4\mu c^2\frac{1}{n^3\hbar^2}\mathbf{S}\cdot\mathbf{I}\,, \tag{8.45}$$

where $\mathbf{I}$ is the spin operator for the proton, labelled using a different letter to avoid confusion with the electron spin operator $\mathbf{S}$.

The splittings in energy levels caused by this perturbation are suppressed not just by a factor of α^4, but by an additional factor of the electron-to-proton mass ratio $m/M \approx 1/1836$. The splittings caused by accounting for nuclear spin are thus much smaller than those of the fine structure considered above, and the perturbation (8.45) is known as the *hyperfine* contribution to the energy levels. We can compute the eigenvalues in much the same way as for the spin-orbit contribution to fine structure. Let $\mathbf{F} = \mathbf{I} + \mathbf{S}$ denote the total spin of the proton and electron. Then

$$(\mathbf{S}\cdot\mathbf{I}) = \frac{1}{2}(\mathbf{F}^2 - \mathbf{I}^2 - \mathbf{S}^2) = \frac{\hbar^2}{2}\left(f(f+1) - \frac{3}{2}\right)\,. \tag{8.46}$$

This is $-3\hbar^2/4$ when the proton and electron spins are antiparallel so $f = 0$, and $+\hbar^2/4$ when $f = 1$. Plugging these into (8.45) shows that the difference between the energies for the $n = 1$ ground state of hydrogen is

$$E_{n=1;f=+1} - E_{n=1;f=0} = \frac{4}{3}\frac{m}{M}\alpha^4\mu c^2 \approx 5.88 \times 10^{-6}\ \text{eV}\,, \tag{8.47}$$

corresponding to a frequency of 1.4 GHz and a wavelength of 21 cm. Thus transitions between the energy levels split by the hyperfine structure causes the hydrogen atom to have a microwave emission line.

The 21 cm line provides the most powerful way of tracing diffuse gas in interstellar and intergalactic space. This wavelength is much longer than the size of typical specks of dust, so 21 cm radiation can propagate with little absorption right through clouds of dust and gas that do absorb visible light. It was radiation at 1.4 GHz that first revealed the large scale structure of our galaxy. The line is intrinsically very narrow, so the temperature and radial (with the Earth as origin) velocity of the hydrogen that emitted the radiation can be accurately measured from the Doppler shift and broadening of the observed spectral line.

The hyperfine structure of hydrogen leading to this 21 cm emission line was first predicted theoretically in 1944 by the Dutch physicist H.C. van de Hulst as part of his doctoral thesis, carried out in Utrecht which was then under Nazi occupation. 21 cm radiation from our galaxy was first detected experimentally in 1951 by three groups working independently in the USA, Australia and the Netherlands. The Dutch group used a German radio antenna left over from the war.

8.1.3 The Ground State of Helium

After hydrogen, helium is the next most abundant element, making up about a quarter of the ordinary matter in the Universe. Much of this helium was created during the period of nucleosynthesis, a tiny fraction of a second after the Big Bang.

Let's use perturbation theory to find an estimate of the ground state energy of helium. Working in the centre of mass frame and treating the nucleus as stationary at the origin, the (gross structure) Hamiltonian is

$$H = \frac{\mathbf{P}_1^2}{2m} + \frac{\mathbf{P}_2^2}{2m} - \frac{2e^2}{4\pi\epsilon_0\,|\mathbf{X}_1|} - \frac{2e^2}{4\pi\epsilon_0\,|\mathbf{X}_2|} + \frac{e^2}{4\pi\epsilon_0}\frac{1}{|\mathbf{X}_1 - \mathbf{X}_2|}\,, \tag{8.48}$$

where $(\mathbf{X}_1, \mathbf{X}_2)$ are the position operators and $(\mathbf{P}_1, \mathbf{P}_2)$ are the momentum operators for the two electrons. The terms describing the electrons' kinetic energy and attraction to the nucleus are just the sum of those for hydrogen, expect with e^2 replaced by $2e^2$ since there are two protons in the nucleus of helium. We'll take these terms to be the model Hamiltonian H_0, treating the remaining electron-electron repulsion as a perturbation.

In the unperturbed Hamiltonian, single-electron states are just the usual hydrogenic states $|n, \ell, m\rangle$ with energy

$$-4 \times \frac{1}{2} m v^2 \alpha^2 \frac{1}{n^2}\,, \tag{8.49}$$

where the factor of 4 arise because each electron's attraction to the helium nucleus is twice as strong as it was for hydrogen. Since electrons are fermions, the total state of the neutral helium atom must be antisymmetric under their exchange. In particular, the ground state is

$$|\Psi_0\rangle = |1,0,0\rangle \otimes |1,0,0\rangle \otimes \frac{1}{\sqrt{2}}\left(|\uparrow\rangle\,|\downarrow\rangle - |\downarrow\rangle\,|\uparrow\rangle\right) \tag{8.50}$$

where the final factor is the antisymmetric spin state of the two spin-$\frac{1}{2}$ electrons. (Note that,at the level of the gross structure of helium, the Hamiltonian is independent of the spins of the two electrons, so we are free to seek simultaneous energy and spin eigenstates.) This state is an eigenstate of the model Hamiltonian H_0 with energy

$$E_1 = -\frac{4\alpha^2}{2}mc^2 - \frac{4\alpha^2}{2}mc^2 = -2\alpha^2 mc^2 \approx -108.8\ \text{eV}\,, \tag{8.51}$$

being the sum of the energies of the two electrons individually.

We now take account of the electron-electron repulsion. To use perturbation theory, we set

$$\Delta H = \frac{e^2}{4\pi\epsilon_0}\frac{1}{|\mathbf{X}_1 - \mathbf{X}_2|} = \frac{\alpha\hbar c}{|\mathbf{X}_1 - \mathbf{X}_2|}\,. \tag{8.52}$$

The first order shift in the ground state energy is

$$\begin{aligned}
\langle\Psi_0|\Delta H|\Psi_0\rangle &= \int \mathrm{d}^3\mathbf{x}_1'\,\mathrm{d}^3\mathbf{x}_2'\,\mathrm{d}^3\mathbf{x}_1\,\mathrm{d}^3\mathbf{x}_2\ \langle\Psi_0|\mathbf{x}_1'\mathbf{x}_2'\rangle\,\langle\mathbf{x}_1'\mathbf{x}_2'|\Delta H(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{x}_1\mathbf{x}_2\rangle\,\langle\mathbf{x}_1\mathbf{x}_2|\Psi_0\rangle \\
&= \int \mathrm{d}^3\mathbf{x}_1\,\mathrm{d}^3\mathbf{x}_2\ \Psi_0(\mathbf{x}_1,\mathbf{x}_2)^*\frac{\alpha\hbar c}{|\mathbf{x}_1 - \mathbf{x}_2|}\Psi_0(\mathbf{x}_1,\mathbf{x}_2) \\
&= \int \mathrm{d}^3\mathbf{x}_1\,\mathrm{d}^3\mathbf{x}_2\ \frac{\alpha\hbar c}{|\mathbf{x}_1 - \mathbf{x}_2|}\,|\psi_{100}(\mathbf{x}_1)|^2\,|\psi_{100}(\mathbf{x}_2)|^2\ .
\end{aligned} \tag{8.53}$$

Here $\psi_{100}(\mathbf{x})$ are the single-electron $n = 1$ wavefunctions

$$\psi_{100} = \frac{1}{\sqrt{\pi}}\left(\frac{2}{a_2}\right)^{\frac{3}{2}} \mathrm{e}^{-|\mathbf{x}|/a_2}\,, \tag{8.54}$$

where the length scale $a_2 = \frac{\hbar}{2\alpha mc}$ is half the Bohr radius in hydrogen. In going to the last line, we've used the fact that the interaction is independent of spin, and the spin states have unit norm.

Performing the integral is straightforward but somewhat tedious[54]. We have

$$\langle \Delta H \rangle \, \Psi_0 = \frac{5}{4}\alpha^2 mc^2 \,. \tag{8.58}$$

As we suspected, this perturbation is also of order α^2 just like the leading-order term. Hence we get the ground state energy of Helium corrected to first order as

$$E = -4\alpha^2 mc^2 \left(1 - \frac{5}{16}\lambda\right)_{\lambda=1} \approx -74.6 \text{ eV} \,. \tag{8.59}$$

This is in significantly better agreement with the experimental value of ≈ -79.0 eV than the zeroth order estimate. Including the second-order term leads to

$$E = -4\alpha^2 mc^2 \left(1 - \frac{5}{16}\lambda + \frac{25}{1024}\lambda^2\right)_{\lambda=1} \approx -77.5 \text{ eV} \,. \tag{8.60}$$

Given the crudity of our expansion this is remarkably close agreement with the experimental value.

In heavy atoms, the ground state energy itself is difficult to measure experimentally since it involves computing the energy required to strip off all the electrons from the nucleus. Fortunately, for the same reason it's also rarely the directly relevant quantity. Instead, chemists and spectroscopists are usually more interested in the first ionisation energy, defined to be the energy required to liberate a single electron from the atom. Hydrogen only has one electron, so its first ionization energy is just the 13.6 eV which we must supply to eject this electron from the ground state of the atom. In the case of helium, once one electron is stripped away the remaining electron sees the full force of the Coulomb potential from the $Z = 2$ nucleus, so will have binding energy -54.4 eV. The first ionisation energy is thus the difference $(-54.4 + 79.0)$ eV $= 24.6$ eV. This is significantly greater than the ionisation energy of hydrogen, and in fact helium has the greatest first ionisation energy of any of the elements, reflecting its status as the first noble gas.

8.1.4 The Quadratic Stark Effect

We'll now consider a different type of perturbation, caused not by relativistic corrections to the Hamiltonian of the isolated atom, but by allowing the atom to interact with some external system.

Suppose a hydrogen atom is placed in an external electric field $\mathbf{E} = -\boldsymbol{\nabla}\phi$. Most electric fields we can generate in the laboratory are small compared to the strength of the Coulomb field felt by the

[54] In case you are curious: Choose the z-axis of $\mathbf{x}_2$ to be aligned with whatever direction $\mathbf{r}_1$. Then $|\mathbf{x}_1 - \mathbf{x}_2| = \sqrt{|r_1^2 + r_2^2 - 2r_1 r_2 \cos\theta_2|}$, independent of ϕ_2. So

$$\langle \Delta H \rangle \, \Psi_0 = \frac{4\alpha\hbar c}{a_2^3} \int \mathrm{d}^3\mathbf{x}_1 \; |\psi_{100}(\mathbf{x}_1)|^2 \left[\int \mathrm{d}r_2 \, \mathrm{d}\theta_2 \, \frac{r_2^2 \sin\theta_2 \mathrm{e}^{-r_2/a_2}}{\sqrt{|r_1^2 + r_2^2 - 2r_1 r_2 \cos\theta_2|}} \right] . \tag{8.55}$$

The integral in square brackets can be done using

$$\int_0^\pi \mathrm{d}\theta_2 \, \frac{\sin\theta_2}{\sqrt{|r_1^2 + r_2^2 - 2r_1 r_2 \cos\theta_2|}} = \frac{1}{r_1 r_2} \int_0^\pi \mathrm{d}\theta_2 \, \frac{\mathrm{d}}{\mathrm{d}\theta_2} \sqrt{|r_1^2 + r_2^2 - 2r_1 r_2 \cos\theta_2|} = \begin{cases} 2/r_1 & \text{when } r_1 > r_2 \\ 2/r_2 & \text{when } r_1 < r_2 \,. \end{cases} \tag{8.56}$$

Thus the radial integral $\mathrm{d}r_2$ must be broken into two regions. Letting $\rho_1 = 2r_1/a_2$ be a rescaled radial coordinate, we have

$$\begin{aligned} \langle \Delta H \rangle \, \Psi_0 &= \frac{8\alpha\hbar c}{a_2^3} \int \mathrm{d}^3\mathbf{x}_1 \; |\psi_{100}(\mathbf{x}_1)|^2 \left[\int_0^{r_1} \mathrm{d}r_2 \, \frac{r_2^2}{r_1} \mathrm{e}^{-2r_2/a_2} + \int_{r_1}^\infty \mathrm{d}r_2 \, r_2 \mathrm{e}^{-2r_2/a_2} \right] \\ &= \frac{4\alpha\hbar c}{a_2} \int \mathrm{d}^3\mathbf{x}_1 \; |\psi_{100}(\mathbf{x}_1)|^2 \, \frac{2 - \mathrm{e}^{-\rho_1}(2 + \rho_1)}{\rho_1} \\ &= \frac{2\alpha\hbar c}{a_2} \int_0^\infty \mathrm{d}\rho_1 \, \rho_1 \mathrm{e}^{-\rho_1}(2 - \mathrm{e}^{-\rho_1}(2 + \rho_1)) = \frac{5\alpha\hbar c}{4a_2} \,. \end{aligned} \tag{8.57}$$

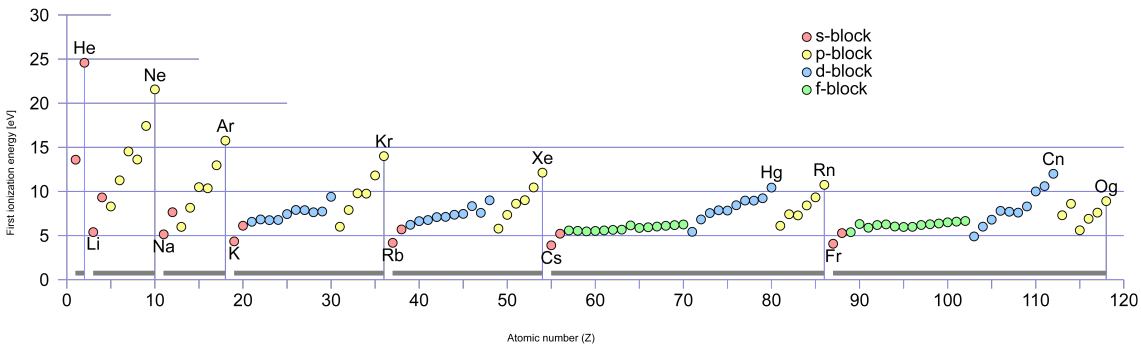


Figure 11: First ionisation energies of the ground states of the first few elements. Figure adapted from Wikipedia.

electron due to the nucleus[55], so perturbation theory should provide a good estimate of the energy shifts such an external field induces. By definition of the electrostatic potential ϕ, the field changes the energy of the atom by $E = e(\phi(\mathbf{x}_p) - \phi(\mathbf{x}_e))$ where $\mathbf{x}_p$ and $\mathbf{x}_e$ are the locations of the proton and electron. We assume the external field varies only slowly over scales of order the Bohr radius, so

$$\delta E \approx -e\mathbf{r} \cdot \boldsymbol{\nabla}\phi = -e\mathbf{r} \cdot \mathbf{E}\,, \tag{8.61}$$

where $\mathbf{r} = \mathbf{x}_e - \mathbf{x}_p$. Let's choose the direction of the electric field to define the z-axis, and set $\mathcal{E} = |\mathbf{E}|$ so as to avoid confusion with the energy. Thus $\mathbf{E} = \mathcal{E}\hat{\mathbf{z}}$ and the effect of the external electric field is to add a new term to the Hamiltonian H_0 of the unperturbed atom

$$H = H_0 - e\mathcal{E}X_3\,, \tag{8.62}$$

where X_3 is the position operator for the z-coordinate of the electron relative to the proton.

In this section, we'll just consider the effect of this electric field on the ground state $|n, \ell, m\rangle = |1, 0, 0\rangle$ of the hydrogen atom. This state is non-degenerate for the unperturbed Hamiltonian (and our perturbation is blind to spin), so again non-degenerate perturbation theory will suffice. From (8.10), the first order change in the ground state energy is

$$E_1^{(1)} = -e\mathcal{E}\,\langle 1, 0, 0|X_3|1, 0, 0\rangle\ . \tag{8.63}$$

However, applying the parity operator Π and noting that $\Pi^{-1}X_3\Pi = -X_3$ we have

$$\langle 1, 0, 0|X_3|1, 0, 0\rangle = -\left\langle 1, 0, 0\middle|\Pi^{-1}X_3\Pi\middle|1, 0, 0\right\rangle = -\langle 1, 0, 0|X_3|1, 0, 0\rangle\ , \tag{8.64}$$

where the last equality uses the fact that $\Pi\,|1, 0, 0\rangle = |1, 0, 0\rangle$ since the ground state wavefunction is spherically symmetric. Thus $\langle 1, 0, 0|X_3|1, 0, 0\rangle = 0$ and parity symmetry ensures that, to first order in the external electric field, the ground state energy is unaffected.

To see a non-trivial effect, we must go one order further. The second order shift in the ground state energy is given by (8.15) in general. In our case, the sum is to be taken over all states of hydrogen except the ground state, so

$$E_1^{(2)} = e^2\mathcal{E}^2 \sum_{n=2}^{\infty}\sum_{\ell<n}\sum_{m=-\ell}^{\ell} \frac{|\langle 1, 0, 0|X_3|n, \ell, m\rangle|^2}{E_1 - E_n}\,. \tag{8.65}$$

[55]With modern high intensity lasers, electric fields comparable to the nuclear Coulomb potential can be generated. The behaviour of a hydrogen atom in such a laser must be studied by other methods.

In the second problem set you proved that

$$\langle n', \ell', m' | \mathbf{X} | n, l, m \rangle = 0 \text{ unless } |\ell' - \ell| = 1 \,. \tag{8.66}$$

Applying this to our case, $\langle 1,0,0|X_3|n,\ell,m\rangle = 0$ unless $\ell = 1$, which greatly simplifies the sums. Furthermore, $[L_z, X_3] = 0$ so $X_3 \,|n,0,0\rangle$ is an eigenstate of L_z with eigenvalue zero and hence it is orthogonal to $|n,1,m\rangle$ whenever $m = 0$. In summary, to this order the electric field causes the ground state $|1,0,0\rangle$ to mix only with states $|n,1,0\rangle$. Thus the only non-vanishing terms are

$$E_1^{(2)} = e^2 \mathcal{E}^2 \sum_{n=2}^{\infty} \frac{|\langle 1,0,0|X_3|n,1,0\rangle|^2}{E_1 - E_n} \tag{8.67}$$

involving just a single sum. This is the second-order change in the ground state energy due to the presence of the external electric field.

This change in the ground state energy is known as the *quadratic Stark effect*, after its discoverer and since it comes in at order $\mathcal{E}^2$. There's a good physical reason why the change in energy comes in at this order: the hydrogen atom is electrically neutral and the unperturbed ground state $|1,0,0\rangle$ is spherically symmetric, so the atom in this state has no electric dipole moment. In response to the applied electric field, the electronic wavefunction changes by an amount that is proportional to the coefficients b_m, which are themselves proportional to $\mathcal{E}$. In other words, the applied electric field $\mathcal{E}$ *polarizes* the atom, generating a dipole moment $\mathbf{p} \propto \mathbf{E}$. The energy of any dipole $\mathbf{p}$ in an electric field is $U_{\text{dip}} = -\mathbf{p} \cdot \mathbf{E}$ and since the dipole moment itself is $\sim \mathcal{E}$, for the ground state of hydrogen $U_{\text{dip}} \sim \mathcal{E}^2$. Higher energy states $|n,l,m\rangle$ can have a permanent dipole moment due to the electron's elliptical orbits, but to study this we need degenerate perturbation theory.

8.2 Degenerate Perturbation Theory

Our derivation of the coefficients

$$b_m = \frac{\langle m|\Delta H|n\rangle}{E_n - E_m} \tag{8.68}$$

of the first-order shift in the state breaks down if there are states m that have the same energy as the state n that we're attempting to perturb. Indeed, naively applying (8.68) in this case appears to show that the small perturbation can cause an extremely dramatic shift in the state of the system.

There's nothing particularly surprising about this. Consider a marble lying in the bottom of a bowl at the minimum of the gravitational potential. If we tilt the bowl slightly, the marble will move a little, resettling in the bottom of the tilted bowl as it adjusts to the new minimum of the potential. However, if the marble initially lies at rest on a smooth table, so that it initially has the same energy no matter where on the table it lies, then tilting the table even very slightly will lead to a large change in the marbles location.

To handle this degenerate situation, we first observe that the only states that will acquire a significant amplitude as a result of the perturbation are those that are initially degenerate with the original state. In other words, to good approximation, the state to which the system changes will be a linear combination of those having the same zeroth-order energy as our initial state. In many situations, there are only a finite number of these. We then diagonalise the matrix $\langle r|\Delta H|s\rangle$ of the perturbing Hamiltonian in the subspace spanned by these degenerate states. Since this subspace was completely degenerate with respect to H_0, in this subspace the eigenstates of the perturbation ΔH are in fact eigenstates of the full Hamiltonian $H_0 + \Delta H$.

Let's see how this works in more detail. Suppose $V \subset \mathcal{H}$ is an N-dimensional subspace with $H_0 \,|\psi\rangle = E_V \,|\psi\rangle$ for all $|\psi\rangle \in V$. For $r = 1, \ldots, N$, let $\{|r\rangle\}$ be an orthonormal basis of V and define P_V to be the projection operator $P_V : \mathcal{H} \to V$. We can write this projection operator as

$$P_V = \sum_{r=1}^{N} |r\rangle \langle r| \tag{8.69}$$

in terms of our orthonormal basis. Similarly, we let $P_\perp = 1 - P_V$ denote the projection onto $V^\perp$, where

$$V^\perp = \{|\xi\rangle \in \mathcal{H} \mid \langle\psi|\xi\rangle = 0 \;\forall\, |\psi\rangle \in V\}\,. \tag{8.70}$$

Note that, as projection operators, $P_V^2 = P_V$ and $P_\perp^2 = P_\perp$ and that $P_V P_\perp = P_\perp P_V = 0$. Also, we have

$$[H_0, P_V] = 0 \quad \text{and} \quad [H_0, P_\perp] = 0 \tag{8.71}$$

since V was defined in terms of a degenerate subspace of H_0.

Now let's consider a perturbed Hamiltonian $H_\lambda = H_0 + \lambda\Delta H$. Any eigenstate $|\psi_\lambda\rangle$ of the full Hamiltonian obeys

$$\begin{aligned} 0 &= (H_0 - E(\lambda) + \lambda\Delta H)\,|\psi_\lambda\rangle \\ &= (H_0 - E(\lambda) + \lambda\Delta H)(P_V + P_\perp)\,|\psi_\lambda\rangle \\ &= (E_V - E(\lambda) + \lambda\Delta H)P_V\,|\psi_\lambda\rangle + (H_0 - E(\lambda) + \lambda\Delta H)P_\perp\,|\psi(\lambda)\rangle\;. \end{aligned} \tag{8.72}$$

Acting on the left with either P_V or $P_\perp$ and using (8.71), we obtain the two equations

$$(E_V - E(\lambda) + \lambda P_V\Delta H)P_V\,|\psi_\lambda\rangle + \lambda P_V\Delta H P_\perp\,|\psi_\lambda\rangle = 0 \tag{8.73}$$

$$\lambda P_\perp\Delta H P_V\,|\psi_\lambda\rangle + (H_0 - E(\lambda) + \lambda P_\perp\Delta H)P_\perp\,|\psi_\lambda\rangle = 0\,, \tag{8.74}$$

in which we have separated out the effects of the perturbation within V and its complement.

Suppose now that we're perturbing a state $|\psi_0\rangle$ that, in the absence of the perturbation, lies in V. We write

$$|\psi_\lambda\rangle = |\alpha\rangle + \lambda\,|\beta\rangle + \lambda^2\,|\gamma\rangle + \dots \tag{8.75}$$

$$E(\lambda) = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \dots \tag{8.76}$$

just as before, noting that $P_\perp\,|\psi_\lambda\rangle$ is necessarily at least first-order in λ. To zeroth order, equation (8.73) tells us that $E^{(0)} = E_V$ as expected. At first order, this equation becomes

$$(-E^{(1)} + P_V\Delta H)\,|\alpha\rangle = 0 \tag{8.77}$$

or equivalently

$$P_V\Delta H P_V\,|\alpha\rangle = E^{(1)}\,|\alpha\rangle\;. \tag{8.78}$$

This is a remarkable equation! It tells us that, if our assumption that the perturbation causes only a small change in the energies and states of the system is to hold and if our zeroth-order state $|\psi_0\rangle$ lies in some degenerate subspace, then $|\psi_0\rangle$ can't be just any eigenstate of H_0, but must also be an eigenstate of $P_V\Delta H P_V$ — the perturbation, restricted to the degenerate subspace V. In other words, in degenerate systems, states that are stable against small perturbations are those that are already eigenstates of the perturbation. If we started with some other $|\psi_0\rangle \in V$ that was not an eigenstate of ΔH, then as soon as the perturbation is turned on, the state would rapidly change into an eigenstate[56]. This large response to a small change in H would invalidate our use of perturbation theory; it's the quantum analogue of the fact that the location of a marble on a perfectly at table is extremely sensitive to any tilt.

Let's now suppose that we've chosen our basis $\{|r\rangle\}$ of V in such a way that $P_V\Delta H P_V\,|r\rangle = E_r^{(1)}\,|r\rangle$, so that $|r\rangle$ is indeed an eigenstate of the perturbation with some eigenvalue we'll call $E_r^{(1)}$. Then choosing our initial state $|\alpha\rangle$ to be $|r\rangle$, by (8.78) the first order shift in its energy is

$$E_r^{(1)} = \langle r|P_V\Delta H P_V|r\rangle = \langle r|\Delta H|r\rangle\;. \tag{8.79}$$

However, it is wrong to think that since $|r\rangle$ diagonalises both H and ΔH confined in V, with eigenvalues E_V and $E_r^{(1)}$ respectively, we have solved the perturbed Hamiltonian $H_0 + \lambda\Delta H^{(1)}$ exactly,

[56] To justify this, in particular to understand what happens "as a perturbation is turned on", we'll need to consider time-dependent perturbation theory. See section 9.

with eigenvalue $E_V + \lambda E_r^{(1)}$. This is because $|r\rangle$ only diagonalises ΔH in the subspace V, not in the full Hilbert space. If ΔH results in coupling between states in V and outside V, then $\Delta H\,|r\rangle \neq E_r^{(1)}\,|r\rangle$ in the full Hilbert space. This makes the second order degenerate perturbation theory a bit tricky to do because the first order wavefunction will not completely lie within the degenerate subspace V, but will also contain components outside V. We shall not discuss it further.

As we've seen, degeneracy in the energy spectrum often arises as a result of some symmetry. If this symmetry is dynamical, such as the enhanced $\mathrm{U}(d)$ symmetry of the d-dimensional harmonic oscillator, or the $\mathrm{SU}(2) \times \mathrm{SU}(2)$ symmetry the Runge–Lenz vector implies for the gross structure of hydrogen, it will almost inevitably be broken by a perturbation. Even symmetries such as rotations that are inherited from transformations of space may be broken in the presence of external fields. For example, although the ground state of an atom is typically spherically symmetric, the laboratory in which one conducts an experiment likely won't be and any small stray electric and magnetic fields in the laboratory will perturb the atom away from pure spherical symmetry. For this reason, perturbations typically lift degeneracy. As we've seen, even very small perturbations can have a dramatic effect in immediately singling out a preferred set of states in an otherwise degenerate system; we'll return to this point in section 10.4.2 when trying to understand 'collapse of the wavefunction' in quantum mechanics. If you take the Applications of Quantum Mechanics course, you'll see that the lifting of degeneracy is also important in giving crystals their electronic properties: when a large number of identical metal atoms are brought together, ignoring the inter-atomic interactions, all the valence electrons in the different atoms will be degenerate. The small effect of the potential from nearby atoms breaks this degeneracy and makes these valence electrons prefer to delocalise throughout the crystal.

8.2.1 The Linear Stark Effect

As an application of degenerate perturbation theory, let's again consider the Stark effect but now for the 2s state of our atom. This state is degenerate with the three[57] 2p states, so here our degenerate subspace

$$V = \mathrm{span}\{|2,0,0\rangle\,, |2,1,1\rangle\,, |2,1,0\rangle\,, |2,1,-1\rangle\}\,. \tag{8.80}$$

As before, the electric field induces a perturbation $\Delta H = e\mathcal{E}X_3$ and we need to diagonalise this perturbation within V. Parity implies

$$\langle 2,0,0|X_3|2,0,0\rangle = 0 \quad \text{and} \quad \langle 2,1,m|X_3|2,1,m'\rangle = 0\ \forall m,m' \in \{-1,0,1\}\,, \tag{8.81}$$

while since $[L_z, X_3] = 0$ we also have $\langle 2,0,0|X_3|2,1,\pm 1\rangle = 0$. Thus the matrix elements of ΔH in this basis are

$$\Delta|_V = e\mathcal{E}\begin{pmatrix} 0 & 0 & a & 0 \\ 0 & 0 & 0 & 0 \\ a^* & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}\,, \tag{8.82}$$

where

$$a = \langle 2,0,0|X_3|2,1,0\rangle = -3a_0\,, \tag{8.83}$$

with a_0 the Bohr radius. The eigenvalues of Δ in this subspace are thus $\{3e\mathcal{E}a_0, 0, 0, -3e\mathcal{E}a_0\}$ with corresponding eigenstates

$$\frac{1}{\sqrt{2}}(|2,0,0\rangle - |2,1,0\rangle)\,,\ |2,1,1\rangle\ ,\ |2,1,-1\rangle\ ,\ \frac{1}{\sqrt{2}}(|2,0,0\rangle + |2,1,0\rangle)\,. \tag{8.84}$$

In the presence of the electric field, the $n = 2$ state of lowest energy is $(|2,0,0\rangle + |2,1,0\rangle)/\sqrt{2}$ and we conclude that even for a tiny electric field, the $n = 2$ states will jump to this preferred state.

[57] Since the applied electric field does not coupled to spin, we ignore the effect of spin in this discussion at this order, the Stark effect will not lift the degeneracy of the gross structure of hydrogen with respect to the electrons spin.

From our discussion of the quadratic Stark effect above, we know that a change in energy $\propto \mathcal{E}$ requires the dipole moment $\mathbf{D}$ of an atom to be independent of $\mathcal{E}$. Since this $n = 2$ state has a first-order energy shift of $3e\mathcal{E}a_0$, we conclude that the $n = 2$ states of hydrogen do indeed have a permanent electric dipole moment of magnitude $3ea_0$. Classically, for an electron in an elliptical orbit, this result is expected because the electron would spend more time at the apocentre (the point of its orbit furthest from the atoms centre of mass) than at the pericentre (the point nearest to the centre of mass). If electrons orbit was a perfect Keplerian ellipse, the atom would have a permanent electric dipole moment aligned parallel to the orbit's major axis. However, any small deviation from the $1/r$ potential will cause the major axis of the ellipse to precess, and the time averaged dipole moment would vanish. For hydrogen, the potential does differ from $1/r$, though only very slightly due to the effects of fine structure. Hence even a weak external field can prevent precession and give rise to a steady electric dipole.

Quantum mechanically, in the presence of the electric field the lowest energy $n = 2$ state $(|2,0,0\rangle + |2,1,0\rangle)/\sqrt{2}$ is no longer an eigenstate of $\mathbf{L}^2$, because the field applies torque to the atom, changing its total orbital angular momentum. However, this lowest energy state is still an eigenstate of $\mathbf{L}_z$ with eigenvalue zero. Recalling that we took the z-direction to be the direction of the applied field, we see that the angular momentum is perpendicular to $\mathbf{E}$, as expected from the classical picture of an ellipse with major axis aligned along $\mathbf{E}$. Note also that the electric field does not lift the degeneracy between the $|2,1,1\rangle$ and $|2,1,-1\rangle$ states. These two states have their orbital angular momentum maximally aligned or anti-aligned with $\mathbf{E}$; classically, they correspond to orbits confined to a plane perpendicular to $\mathbf{E}$.

8.3 Does Perturbation Theory Converge?

We began our study of perturbation theory by assuming that the states and energy eigenvalues of the full Hamiltonian depended analytically on a dimensionless parameter λ controlling the perturbation. Even when the individual coefficients of powers of λ are finite, this is often not the case because the infinite perturbative series itself may fail to converge, or may converge only for some range of λ. The issue is that we really have an expansion in

$$\left|\lambda \frac{\langle m|\Delta H|n\rangle}{E_m - E_n}\right| \tag{8.85}$$

where $m \neq n$, and the coefficients of λ may grow too rapidly. Heuristically, the condition for convergence is thus that the typical energy splitting $\langle m|\Delta H|n\rangle$ induced by the perturbation should be much smaller than the initial energy difference $E_m - E_n$. However, a detailed criterion is often hard to come by since the higher terms in the perturbation expansion involve complicated sums (or integrals) over many different intermediate states.

To illustrate this in a simple context, let's consider in turn the following three perturbations of a 1d harmonic oscillator potential

$$H = \frac{P^2}{2m} + \frac{1}{2}m\omega^2 X^2 + \begin{cases} -\lambda m\omega^2 x_0 X \\ +\frac{1}{2}\lambda m\omega^2 X^2 \\ +\lambda\epsilon X^4\,, \end{cases} \tag{8.86}$$

where x_0 and ϵ are constants. Of course, the first two can be solved exactly — we'd never really use perturbation theory to study them.

In the first case, we have

$$H = \frac{P^2}{2m} + \frac{1}{2}m\omega^2(X - \lambda x_0)^2 - \frac{\lambda^2}{2}m\omega^2 x_0^2\,, \tag{8.87}$$

from which we easily see that the exact energies are

$$E_n(\lambda) = \left(n + \frac{1}{2}\right)\hbar\omega - \frac{\lambda^2}{2}m\omega^2 x_0^2 \tag{8.88}$$

with corresponding position space wavefunction $\langle x|n_\lambda\rangle = \langle x - \lambda x_0|n\rangle$ just a translation of the usual harmonic oscillator wavefunction $\langle x|n\rangle$. If we instead tackled this problem using perturbation theory, we'd find

$$\begin{aligned} E_n(\lambda) &= E_n - \lambda m\omega^2 x_0 \langle n|X|n\rangle + \lambda^2 m^2\omega^4 x_0^2 \sum_{k\neq n} \frac{|\langle k|X|n\rangle|^2}{(n-k)\hbar\omega} + O(\lambda^3) \\ &= \left(n + \frac{1}{2}\right)\hbar\omega - \frac{\lambda^2}{2} m\omega^2 x_0^2 + O(\lambda^3)\,. \end{aligned} \tag{8.89}$$

To obtain this result, we note that the first order term vanishes (e.g. by parity), while since X is a linear combination of creation and annihilation operators $A^\dagger$ and A, only the $k = n+1$ and $k = n-1$ terms can contribute to the sum in the second-order term. Going further, we'd find that there are no higher corrections in λ — the $O(\lambda^3)$ terms are in fact zero — though this is not easy to see directly. Thus, in this case, the perturbative result converges to the exact result, and the radius of convergence is infinite. This reflects the fact that the perturbation $-\lambda m\omega^2 x_0 X$ didn't really change the character of the original Hamiltonian. No matter how large λ is, for large enough x the perturbation remains negligible.

Turning to the second case, it's again immediate that the exact energy levels are

$$E_n(\lambda) = \left(n + \frac{1}{2}\right)\hbar\omega(\lambda)\,, \tag{8.90}$$

where $\omega(\lambda) = \omega\sqrt{1+\lambda}$ is the modified frequency. This has a branch cut starting at $\lambda = -1$, so the energy is only analytic in the disc $|\lambda| < 1$. Again using perturbation theory, we find

$$\begin{aligned} E_n(\lambda) &= E_n + \frac{\lambda}{2} m\omega^2 \langle n|X^2|n\rangle + \frac{\lambda^2}{4} m^2\omega^4 \sum_{k\neq n} \frac{|\langle k|X^2|n\rangle|^2}{(n-k)\hbar\omega} + O(\lambda^3) \\ &= \left(n + \frac{1}{2}\right)\hbar\omega\left(1 + \frac{\lambda}{2} - \frac{\lambda^2}{8} + O(\lambda^3)\right)\,, \end{aligned} \tag{8.91}$$

agreeing to this order with the Taylor expansion of the exact answer. Continuing further, we'd find that this Taylor series does indeed converge provided $|\lambda| < 1$, and that it then converges to the exact answer. The physical reason why the perturbation series diverges when $|\lambda| \geq 1$ is simply that if $\lambda = -1$, the 'perturbation' has completely cancelled the original harmonic oscillator potential, so we are no longer studying a system that can be treated as a harmonic oscillator in the first instance. Once $\lambda < -1$ the harmonic oscillator potential is turned upside down, and we do not expect our system to possess any stable bound states.

Finally, consider the case

$$H = H_0 + \lambda\epsilon X^4\,. \tag{8.92}$$

I do not know whether this model has been solved exactly, but it can be treated perturbatively. After a fair amount of non-trivial calculation[58] one obtains the series

$$E_0(\lambda) = \frac{1}{2}\hbar\omega + \sum_{n=1}^{\infty} (\lambda\epsilon)^n a_n \tag{8.93}$$

for the ground state energy including the quartic interaction, where the coefficients behave as

$$a_n = \frac{(-1)^{n+1}\sqrt{6}}{\pi^{3/2}} 3^n \Gamma\left(n + \frac{1}{2}\right)\left(1 - \frac{95}{72}\frac{1}{n} + O(n^{-2})\right)\,. \tag{8.94}$$

On account of the Γ-function, these grow faster than factorially with n, so the series (8.93) has radius of convergence $\lambda = 0$! Once again, this is easy to see from the form of the perturbed Hamiltonian:

[58] You can find the details in Bender, C. and Wu, T.T., *Anharmonic Oscillator II: A Study of Perturbation Theory in Large Order*, Phys. Rev. D7, 1620-1636 (1973)

even though we may only care about $\lambda > 0$, our assumption that the perturbation expansion is analytic in λ at $\lambda = 0$ means that, if it converges, it will do so for a disc $\lambda \in D \subset \mathbb{C}$. For any $\lambda \in \mathbb{R}_{<0}$, the Hamiltonian of the quartic oscillator is unbounded below, so there cannot be any stable bound states that are analytic in λ at $\lambda = 0$.

Let me comment that even when perturbative series do not converge, they may still provide very useful information as an asymptotic series. You'll learn far more about these if you take the Part II course on Asymptotic Methods, but briefly, we say a series $S_N(\lambda) = \sum_{n=0}^{N} a_n \lambda^n$ is *asymptotic* to an exact function $S(\lambda)$ as $\lambda \to 0^+$ (written $S_N(\lambda) \sim S(\lambda)$ as $\lambda \to 0^+$) if

$$\lim_{\lambda \to 0^+} \frac{1}{\lambda^N} \left| S(\lambda) - \sum_{n=0}^{N} a_n \lambda^n \right| = 0 \,. \tag{8.95}$$

In other words, if we just include a fixed number N of terms in our series, then for small enough $\lambda \geq 0$ these first N terms differ from the exact answer by less that $\epsilon \lambda^N$ for any $\epsilon > 0$ (so the difference is $o(\lambda^N)$). However, if we instead try to fix λ and improve our accuracy by including more and more terms in the series, then an asymptotic series will eventually diverge. Most of the perturbative series one meets in the quantum world (including most Feynman diagram expansions of Quantum Field Theory) are only asymptotic series. Just as in our toy examples above, the radius of convergence of such series is often associated with interesting physics.

9 Perturbation Theory II: Time Dependent Case

In this chapter we'll study perturbation theory in the case that the perturbation varies in time. Unlike the time independent case, where we mostly wanted to know the bound state energy levels of the system including the perturbation, in the time dependent case our interest is usually in the rate at which a quantum system changes its state in response to the presence of the perturbation.

9.1 The Interaction Picture

Consider the Hamiltonian

$$H(t) = H_0 + \Delta_{\mathrm{S}}(t)\,, \tag{9.1}$$

where H_0 is a time-independent Hamiltonian whose spectrum and eigenstates we understand, and $\Delta_{\mathrm{S}}(t)$ is a perturbation that we now allow to depend on time. (The footnote on $\Delta_{\mathrm{S}}(t)$ is to remind us that this operator has genuine, explicit time dependence even in the Schrödinger picture.) The idea is that H_0 models a system in quiescent state, and we want to understand how eigenstates of H_0 respond as they start to notice the perturbation. In the absence of the time-dependent perturbation, the time evolution operator would be $U_0(t) = \mathrm{e}^{\mathrm{i}H_0 t/\hbar}$ as appropriate for the model Hamiltonian H_0. We define the *interaction picture* state $|\psi_{\mathrm{I}}(t)\rangle$ as

$$|\psi_{\mathrm{I}}(t)\rangle = U_0^{-1}(t)\,|\psi_{\mathrm{S}}(t)\rangle\ , \tag{9.2}$$

where $|\psi_{\mathrm{S}}(t)\rangle$ is the Schrödinger picture state evolving in accordance with the time dependent Schrödinger equation for the full Hamiltonian (9.1). Thus, in the absence of $\Delta_{\mathrm{S}}(t)$, we'd simply have $|\psi_{\mathrm{I}}(t)\rangle = U_0^{-1}(t)U_0(t)\,|\psi(0)\rangle = |\psi_0\rangle$ so the state would in fact be independent of time. The presence of the perturbation means that our state do not evolve purely according to $U_0(t)$, so $|\psi_{\mathrm{I}}(t)\rangle$ has non-trivial time dependence.

To calculate the evolution of $|\psi_{\mathrm{I}}(t)\rangle$, we differentiate

$$\begin{aligned}
\mathrm{i}\hbar\frac{\partial}{\partial t}\,|\psi(t)\rangle &= -H_0\mathrm{e}^{\mathrm{i}H_0t/\hbar}\,|\psi_{\mathrm{S}}(t)\rangle + \mathrm{e}^{\mathrm{i}H_0t/\hbar}\mathrm{i}\hbar\frac{\partial}{\partial t}\,|\psi_{\mathrm{S}}(t)\rangle \\
&= -H_0\mathrm{e}^{\mathrm{i}H_0t/\hbar}\,|\psi_{\mathrm{S}}(t)\rangle + \mathrm{e}^{\mathrm{i}H_0t/\hbar}(H_0+\Delta_{\mathrm{S}}(t))\,|\psi_{\mathrm{S}}(t)\rangle \\
&= U_0^{-1}(t)\Delta_{\mathrm{S}}(t)U_0(t)\,|\psi_{\mathrm{I}}(t)\rangle\ ,
\end{aligned} \tag{9.3}$$

where we have used the full time dependent Schrödinger equation to evaluate $\partial\,|\psi_{\mathrm{S}}\rangle\,/\partial t$, and in the last line we've written our state back in terms of the interaction picture. Similarly, the expectation value of an operator A_{S} in the Schrödinger picture becomes

$$\langle\psi_{\mathrm{S}}(t)|A_{\mathrm{S}}|\psi_{\mathrm{S}}(t)\rangle = \big\langle\psi_{\mathrm{I}}(t)\big|U_0^{-1}(t)A_{\mathrm{S}}U_0(t)\big|\psi_{\mathrm{I}}(t)\big\rangle \tag{9.4}$$

in terms of the interaction picture states, so we define the interaction picture operator $A_{\mathrm{I}}(t)$ by

$$A_{\mathrm{I}}(t) = U_0^{-1}(t)A_{\mathrm{S}}U_0(t) \tag{9.5}$$

again using just the evolution operators of the unperturbed, rather than full, Hamiltonian. Infinitesimally, this is

$$\frac{\mathrm{d}}{\mathrm{d}t}A_{\mathrm{I}}(t) = \frac{\mathrm{i}}{\hbar}[H_0, A_{\mathrm{I}}(t)] + U_0^{-1}(t)\frac{\partial A_{\mathrm{I}}(t)}{\partial t}U_0(t)\,, \tag{9.6}$$

where the final term comes from the explicit time dependence present even in the Schrödinger picture operator. In this sense, the interaction picture is 'halfway between' the Schrödinger picture, in which states evolve and operators are time independent, and the Heisenberg picture, in which states are always taken at their initial value $|\psi(0)\rangle$, but operators evolve according to the full Hamiltonian. In the interaction picture, states evolve by just the time-dependent perturbation as

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\,|\psi_{\mathrm{I}}(t)\rangle = \Delta_{\mathrm{I}}(t)\,|\psi_{\mathrm{I}}(t)\rangle\ , \tag{9.7}$$

where $\Delta_{\mathrm{I}}(t) = U_0^{-1}(t)\Delta(t)U_0(t)$ is the perturbation in the interaction picture, whilst operators also evolve, but only via the model Hamiltonian.

It will turn our to be useful to introduce an interaction picture time evolution operator $U_{\mathrm{I}}(t)$ defined so that

$$|\psi_{\mathrm{I}}(t)\rangle = U_{\mathrm{I}}(t)\,|\psi_{\mathrm{I}}(0)\rangle \tag{9.8}$$

for any initial state $|\psi_{\mathrm{I}}(0)\rangle \in \mathcal{H}$. From (9.3) we find that $U_{\mathrm{I}}(t)$ satisfies the differential equation

$$\mathrm{i}\hbar\frac{\partial}{\partial t}U_{\mathrm{I}}(t) = \Delta_{\mathrm{I}}(t)U_{\mathrm{I}}(t)\,. \tag{9.9}$$

Note that since the operator $\Delta_{\mathrm{I}}(t)$ itself depends on time, we cannot immediately integrate this to write $U_{\mathrm{I}}(t)$ in terms of the exponential of $\Delta_{\mathrm{I}}(t)$: in particular it is generally not correct to claim

$$U_{\mathrm{I}}(t) \stackrel{?}{=} \exp\left(-\frac{\mathrm{i}}{\hbar}\int_0^t \mathrm{d}t'\,\Delta_{\mathrm{I}}(t')\right) \tag{9.10}$$

because since the operator $\Delta_{\mathrm{I}}(t)$ itself now depends on time, in general

$$[\Delta_{\mathrm{I}}(t), \Delta_{\mathrm{I}}(t')] \neq 0\,. \tag{9.11}$$

Thus, when expanding the exponential as a power series, we'd have to commute the operators at different times through one another.

So far our considerations have been exact, but to make progress we must now approximate. To do so, it will be convenient to rewrite (9.9) as an integral equation (keep in mind the initial condition $U_{\mathrm{I}}(0) = 1$)

$$U_{\mathrm{I}}(t) = 1 - \frac{\mathrm{i}}{\hbar}\int_0^t \mathrm{d}t'\,\Delta_{\mathrm{I}}(t')U_{\mathrm{I}}(t') \tag{9.12}$$

in which the operator $U_{\mathrm{I}}(t)$ we wish to understand also appears on the right-hand side, inside the integral. Now we replace the $U_{\mathrm{I}}(t')$ in the integral by its value according to its own equation, obtaining

$$U_{\mathrm{I}}(t) = 1 - \frac{\mathrm{i}}{\hbar}\int_0^t \mathrm{d}t_1\,\Delta_{\mathrm{I}}(t_1) + \left(-\frac{\mathrm{i}}{\hbar}\right)^2\int_0^t \mathrm{d}t_1\,\Delta_{\mathrm{I}}(t_1)\left[\int_0^{t_1}\mathrm{d}t_2\,\Delta_{\mathrm{I}}(t_2)U_{\mathrm{I}}(t_2)\right]. \tag{9.13}$$

The term containing $U_{\mathrm{I}}(t)$ on the right-hand side now has two explicit powers of the perturbation $\Delta_{\mathrm{I}}(t)$, so may be hoped to be of lower order in our perturbative treatment. Iterating this procedure gives

$$U_{\mathrm{I}}(t) = \sum_{n=0}^{\infty}\left(-\frac{\mathrm{i}}{\hbar}\right)^n\int_0^{t_1}\mathrm{d}t_1\,\ldots\int_0^{t_{n-1}}\mathrm{d}t_n\,\Delta_{\mathrm{I}}(t_1)\Delta_{\mathrm{I}}(t_2)\ldots\Delta_{\mathrm{I}}(t_n) \tag{9.14}$$

in general. Note that the operators $\Delta(t_i)$ at different times do not necessarily commute — they appear in this expression in a time-ordered sequence, with the integrals taken over the range

$$0 \leq t_n \leq t_{n-1} \leq \cdots \leq t_1 \leq t\,. \tag{9.15}$$

Instead of the naive expression (9.10), we often write the series as

$$U_{\mathrm{I}}(t) = \mathrm{T}\exp\left(-\frac{\mathrm{i}}{\hbar}\int_0^{\infty}\mathrm{d}t'\,\Delta_{\mathrm{I}}(t')\right) \tag{9.16}$$

for short, where the *time-ordered exponential*[59] $\mathrm{T}\exp$ defined by the series (9.13). It's a good exercise to check that this time-ordered exponential is equal to the usual exponential in the case that $[\Delta(t_1), \Delta(t_2)] = 0$ for all t_1 and t_2.

[59] Time ordered exponentials were introduced by Freeman Dyson and play a prominent role in QFT. They are also familiar in differential geometry in the context of computing the holonomy of a connection around a closed path. (This is not a coincidence, but it would take us too far afield to explain here.)

The interaction picture perturbation is $\Delta_{\mathrm{I}}(t) = U_0^{-1}(t)\Delta_{\mathrm{S}}(t)U_0(t)$, so we can also write (9.13) as

$$U_{\mathrm{I}}(t) = \sum_{n=0}^{\infty}\left(-\frac{\mathrm{i}}{\hbar}\right)^n \int_0^t \mathrm{d}t_1 \ldots \int_0^{t_{n-1}} \mathrm{d}t_n \\ U_0^{-1}(t_1)\Delta_{\mathrm{S}}(t_1)U_0(t_1-t_2)\Delta_{\mathrm{S}}(t_2)\ldots U_0(t_{n-1}-t_n)\Delta_{\mathrm{S}}(t_n)U_0(t_n) \tag{9.17}$$

using the fact that $U_0^{-1}(t_i)U_0(t_{i-1}) = U_0(t_{i-1}-t_i)$. In this form we see that the perturbation expansion of $U_{\mathrm{I}}(t)$ treats the effect of the perturbation as a sequence of impulses: we evolve according to H_0 for some time $t_n \geq 0$ then feel the effect $\Delta_{\mathrm{S}}(t_n)$ of the perturbation just at this time, before evolving according to H_0 for a further time $t_n - t_{n-1} \geq 0$ and feeling the next impulse from the perturbation, and so on. Finally, we integrate over all the possible intermediate times at which the effects of the perturbation were felt, and sum over how many times it acted. This method, closely related to Green's functions for ODEs and PDEs, is also the basis for Feynman diagrams in QFT, where H_0 is usually taken to be a free Hamiltonian.

Now let's put these ideas to work. Suppose that at some initial time t_0 we're in the state $|\psi(t_0)\rangle = \sum_n a_n(t_0)\,|n\rangle$, described in terms of some superposition of the basis $\{|n\rangle\}$ of eigenstates of the unperturbed Hamiltonian. Then at a later time t the interaction picture state will be $|\psi_{\mathrm{I}}(t)\rangle = U_{\mathrm{I}}(t-t_0)\,|\psi(t_0)\rangle$ and can again be described by a superposition $\sum_n a_n(t)\,|n\rangle$. Contracting with a state $\langle k|$ gives

$$a_k(t) = \langle k|U_{\mathrm{I}}(t-t_0)|\psi_{\mathrm{I}}(t_0)\rangle \tag{9.18}$$

in the interaction picture. Thus, using (9.13) we have

$$\begin{aligned} a_k(t) &\approx a_k(t_0) - \frac{\mathrm{i}}{\hbar}\sum_n a_n(t_0)\int_{t_0}^t \mathrm{d}t'\ \langle k|\Delta_{\mathrm{I}}(t')|n\rangle \\ &= a_k(t_0) - \frac{\mathrm{i}}{\hbar}\sum_n a_n(t_0)\int_{t_0}^t \mathrm{d}t'\ \left\langle k\middle|U_0^{-1}(t')\Delta_{\mathrm{S}}(t')U_0(t')\middle|n\right\rangle \\ &= a_k(t_0) - \frac{\mathrm{i}}{\hbar}\sum_n a_n(t_0)\int_{t_0}^t \mathrm{d}t'\ \mathrm{e}^{\mathrm{i}(E_k-E_n)t'/\hbar}\,\langle k|\Delta_{\mathrm{S}}(t')|n\rangle \end{aligned} \tag{9.19}$$

to the first non-trivial order in Δ, where we used the fact that $H_0\,|n\rangle = E_n\,|n\rangle$. In particular, if at $t = t_0$ we begin in an eigenstate $|m\rangle$ of H_0, so that

$$a_n(t_0) = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{otherwise,} \end{cases} \tag{9.20}$$

then the amplitude for our perturbation caused the system to make a transition into a different H_0 eigenstate $|k\rangle \neq |m\rangle$ is

$$a_k(t) = -\frac{\mathrm{i}}{\hbar}\int_{t_0}^t \mathrm{d}t'\ \mathrm{e}^{\mathrm{i}(E_k-E_m)t'/\hbar}\,\langle k|\Delta_{\mathrm{S}}(t')|m\rangle \tag{9.21}$$

to the first order of perturbation. Henceforth, we will drop the subscript on the Schrödinger picture perturbation — note that this is the form in which the perturbation enters the original Hamiltonian.

To go further, we must specify how the perturbation $\Delta(t)$ actually depends on time.

9.2 Prodding the Harmonic Oscillator

As a example, let's consider a $d = 1$ harmonic oscillator that receives a gentle 'prod' described by the time-dependent force $F_{\mathrm{prod}}(t) = F_0\mathrm{e}^{-t^2/\tau^2}$, in addition to the force $F = -kX$ from the spring constant. We can describe the kick using the potential $V(X) = -F_0 X\mathrm{e}^{-t^2/\tau^2}$ which we will treat as a perturbation. Now suppose that, long before the kick, the oscillator was in its ground state $|0\rangle$.

Then to first order in the perturbation, the amplitude for the oscillator to make a transition into the kth excited state as $t \to \infty$ is

$$\lim_{t\to\infty} a_k(t) = -\frac{\mathrm{i}}{\hbar}\int_{-\infty}^{\infty} \mathrm{d}t\ -F_0 \langle k|X|0\rangle\, \mathrm{e}^{\mathrm{i}k\omega t - \frac{t^2}{\tau^2}}\,. \tag{9.22}$$

Recalling that $X = \sqrt{\frac{\hbar}{2m\omega}}(A + A^\dagger)$, we see that $a_k(t) = 0$ for $k \neq 1$, while

$$\lim_{t\to\infty} a_1(t) = F_0\sqrt{\frac{\pi\hbar}{2m\omega}}\tau \mathrm{e}^{-\omega^2\tau^2/4}\,, \tag{9.23}$$

where the Gaussian integral is performed by completing the square in the exponent. Thus the probability of the kick causing a transition is

$$\mathrm{Prob}(|0\rangle \to |1\rangle) = F_0^2 \frac{\pi\hbar}{2m}\frac{\tau^2}{\omega}\mathrm{e}^{-\omega^2\tau^2/2} \tag{9.24}$$

to first order in perturbation theory. This grows like τ^2 for small τ, then falls as $\mathrm{e}^{-\omega^2 r^2/2}$ for $\tau \gg \omega^{-1}$. For an oscillator of fixed (classical) frequency ω, the maximum probability occurs when $\tau \sim \omega^{-1}$. Transitions to higher excited states are possible at higher order in perturbation theory, but if we really kick the oscillator, then perturbation theory will cease to be appropriate.

9.3 A Constant Perturbation Turned On at $t = 0$

As a further example, consider the case that

$$\Delta(t) = \begin{cases} 0 & \text{for } t < 0 \\ \Delta & \text{for } t > 0\,, \end{cases} \tag{9.25}$$

where $\Delta = \Delta(X, P, \dots)$ is a time-independent operator. Let's start our system so that it's in the mth eigenstate $|m\rangle$ of H_0 at $t = 0$. Then from (9.21) with $t_0 = 0$ we have

$$a_k(t) = -\frac{\mathrm{i}}{\hbar}\int_0^t \mathrm{d}t'\ \langle k|\Delta|m\rangle\, \mathrm{e}^{\mathrm{i}\omega_{km}t'} = \frac{\Delta_{km}}{E_k - E_m}(1 - \mathrm{e}^{\mathrm{i}\omega_{km}t})\,, \tag{9.26}$$

where $\Delta_{km} = \langle k|\Delta|m\rangle$ and $\omega_{km} = (E_k - E_m)/\hbar$. Thus the probability that the perturbation causes the system to be in the kth state at time t is

$$|a_k(t)|^2 = \frac{4\,|\Delta_{km}|^2}{\hbar^2\omega_{km}^2}\sin^2(\omega_{km}t/2)\,. \tag{9.27}$$

We see that the transition probability depends on the energy difference $\hbar\omega_{km}$ as well as the matrix element Δ_{km}.

Consider the graph of $\sin^2(\Omega t/2)/\Omega^2$ as a function of Ω for fixed t, as shown in figure 12 (Note that the plot has an extra t in the denominator.). The height of the central peak is $\sim t^2$ while its width is $\sim 1/t$, where we recall that t is the time over which the perturbation acted. For large t, $|a_k(t)|^2$ is only appreciable for states k whose energies obey

$$t \sim \frac{2\pi}{|\omega_{km}|} = \frac{2\pi\hbar}{|E_k - E_m|}\,. \tag{9.28}$$

In other words, if Δt is the time for which the perturbation has been switched on and ΔE the energy difference between states which have an appreciable probability of a transition, then we require

$$\Delta t \Delta E \sim \hbar\,. \tag{9.29}$$

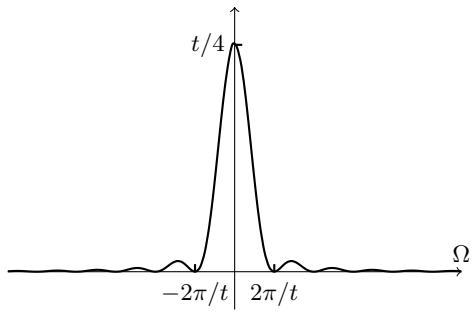


Figure 12: A plot of $\sin^2(\Omega t/2)/\Omega^2 t$ as a function of Ω for fixed t. It $\to \pi\delta(\Omega)/2$ as $t \to \infty$.

For small Δt, the central peak in figure 12 is broad and we can have appreciable probability of a transition even between states with large ΔE. However, if the constant perturbation has been switched on for a long time, we will only obtain transitions to states $|k\rangle$ that are (very nearly) degenerate with the original states $|m\rangle$. In fact, as $t \to \infty$, for $\Omega \neq 0$ we have

$$\lim_{t\to\infty} \frac{\sin^2(\Omega t/2)}{\Omega^2 t} \leq \lim_{t\to\infty} \frac{1}{\Omega^2 t} = 0\,, \tag{9.30}$$

whilst

$$\lim_{\Omega\to 0} \frac{\sin^2(\Omega t/2)}{\Omega^2 t} = \frac{t}{4} \tag{9.31}$$

which diverges if we then send $t \to \infty$. Furthermore, the integral of this function is

$$\int_{-\infty}^{\infty} \mathrm{d}\Omega\, \frac{\sin^2(\Omega t/2)}{\Omega^2 t} = \frac{\pi}{2}\,, \tag{9.32}$$

independent of t. This shows that we can replace

$$\lim_{t\to\infty} \frac{\sin^2(\Omega t/2)}{\Omega^2 t} = \frac{\pi}{2}\delta(\Omega)\,, \tag{9.33}$$

so that as $t \to \infty$, the transition probability behaves as

$$|a_k(t)|^2 \sim \frac{2\pi}{\hbar^2}\, |\langle k|\Delta|m\rangle|^2\, \delta(\omega_{km})t\,. \tag{9.34}$$

We define the *transition rate* $\Gamma(m \to k)$ by

$$\Gamma(m \to k) = \lim_{t\to\infty} \frac{\partial}{\partial t}\, |a_k(t)|^2\ . \tag{9.35}$$

In the case of a constant perturbation switched on at $t = 0$, we thus have

$$\Gamma(m \to k) = \frac{2\pi}{\hbar}\, |\langle k|\Delta|m\rangle|^2\, \delta(E_k - E_m)\,, \tag{9.36}$$

where we have written the argument of the δ-function in terms of energies, pulling out a factor of $\hbar$. As claimed, for this constant perturbation we only have a non-zero transition rate to states which are degenerate with our initial state.

9.4 Fermi's Golden Rule

Just a small extension of the above gives us one of the most important cases of time dependent perturbation theory. Suppose our perturbation takes the form

$$\Delta(t) = \begin{cases} 0 & \text{for } t < 0 \\ \Delta \mathrm{e}^{-\mathrm{i}\omega t} + \Delta^\dagger \mathrm{e}^{+\mathrm{i}\omega t} & \text{for } t > 0\,, \end{cases} \tag{9.37}$$

where again $\Delta = \Delta(X, P, \dots)$ is some time-independent operator. Thus, as before our Schrödinger picture perturbation $\Delta(t)$ is turned on at $t = 0$, but thereafter it now oscillates with fixed frequency ω. Since $\Delta(t)$ appears as a term in the Hamiltonian, it must be Hermitian, so we've included both an $\mathrm{e}^{+\mathrm{i}\omega t}$ and $\mathrm{e}^{-\mathrm{i}\omega t}$ term and without loss of generality we may take $\omega > 0$. We'll see below that these two terms are responsible for different physics. We call such perturbations *monochromatic* in anticipation of the case that they describe an applied electromagnetic field, corresponding to light of frequency ω.

We'll again consider the case where our system is initially in the mth eigenstate of the unperturbed H_0, so that $a_n(0) = \delta_{nm}$. Performing the time integral in (9.21) in this monochromatic case is again trivial and yields

$$a_k(t) = \frac{\langle k|\Delta|m\rangle}{\hbar(\omega_{km} - \omega)} \left[\mathrm{e}^{\mathrm{i}(\omega_{km}-\omega)t} - 1\right] + \frac{\langle k|\Delta^\dagger|m\rangle}{\hbar(\omega_{km} + \omega)} \left[\mathrm{e}^{\mathrm{i}(\omega_{km}+\omega)t} - 1\right] . \tag{9.38}$$

This is very similar to our result (9.26) for the constant perturbation, with the time-dependence of the perturbation just causing ω_{km} to be replaced by $\omega_{km} \pm \omega$. Consequently, we will obtain significant transitions amplitudes only to states $|m\rangle$ for which $E_k - E_m \approx \hbar\omega$, or else for which $E_k - E_m \approx -\hbar\omega$ to high accuracy. The first situation, where $E_k > E_m \approx \hbar\omega$ corresponds to our system *absorbing* energy from the perturbation, whilst the second is *stimulated emission* where the system is prompted to release energy by the perturbation. In this way, the time dependent monochromatic perturbation can be regarded as an enormous source and sink of energy that can be exchanged with the system.

In the case of absorption, the first term dominates at late times. Thus the probability that the perturbation excites the system from $|m\rangle \to |k\rangle$ is

$$|a_k(t)|^2 = \frac{4\,|\langle k|\Delta|m\rangle|^2}{\hbar(\omega_{km} - \omega)^2} \sin^2\left(\frac{\omega_{km} - \omega}{2} t\right) \tag{9.39}$$

with $E_k > E_m$, correct to order $|\Delta|^2$. Replacing the late-time behaviour of $|a_k(t)|^2$ by a δ-function as before shows that the transition rate is

$$\Gamma(m \to k) = \lim_{t\to\infty} \frac{\partial}{\partial t} |a_k(t)|^2 = \frac{2\pi}{\hbar} |\langle k|\Delta|m\rangle|^2 \, \delta(E_k - E_m - \hbar\omega) \tag{9.40}$$

in the case that the system absorbs energy from the perturbation. In the opposite case that the system loses energy back into the perturbation, we have $E_k < E_m$ and the corresponding transition rate is

$$\Gamma(m \to k) = \frac{2\pi}{\hbar} |\langle m|\Delta|k\rangle|^2 \, \delta(E_k - E_m + \hbar\omega) \tag{9.41}$$

with an opposite sign in the δ-function. These results were first obtained by Dirac. They proved to be so useful in describing atomic transitions that Fermi dubbed them 'golden rules'. The name stuck and now (9.40) and (9.41) are known as Fermi's golden rule.

On the one hand, we may be interested not in transitions to some specific final state $|k\rangle$, but to any one of a range of continuum eigenstates of H_0. A typical example here would be transitions between an initial bound state of an atom to any of the continuum of positive-energy (non-bound) states. In these circumstances, we let $n(E_k)\mathrm{d}E_k$ denote the number of states with energy in the interval $(E_k, E_k + \delta E_k)$. The function $n(E_k)$ — which needs to be calculated in each case — is known as the *density of states*; note that it must have dimensions of $1/(\text{energy})$ in order for $n(E_k)\mathrm{d}E_k$ to be a number of states. The total transition probability from $|m\rangle$ is then $\int \mathrm{d}E_k \; |a_k(t)|^2 \, n(E_k)$, generalizing the sum $\sum_k |a_k(t)|^2$ in the discrete case. In particular, in the case of absorption the total transition rate from our initial eigenstate $|m\rangle$ is given by

$$\int \mathrm{d}E_k \, \Gamma(m \to k) n(E_k) = \frac{2\pi}{\hbar} \int \mathrm{d}E_k \; |\langle k|\Delta|m\rangle|^2 \, \delta(E_k - E_m - \hbar\omega) n(E_k) \tag{9.42}$$

$$= \frac{2\pi}{\hbar} |\langle k|\Delta|m\rangle|^2 \; n(E_k)|_{E_k = E_m + \hbar\omega} \tag{9.43}$$

to lowest non-trivial order in perturbation theory. We see that we always make transitions through an energy of exactly $\hbar\omega$, with the density of states telling us how many states of energy $E_m + \hbar\omega$ there actually are.

On the other hand, Fermi's golden rule is also useful when our perturbation itself contains a continuum of different frequencies, perhaps thought of as a Laplace or Fourier transform of some a perturbation with generic time dependence. In this case, transitions between discrete bound states are generally possible, because there will always be some perturbation at exactly the right frequency $\omega = \omega_{km} = (E_k - E_m)/\hbar$, even though the right-hand side takes discrete values as k varies over different bound states as the final state.

Let's now consider an example of each of these two cases.

9.4.1 The Photoelectric Effect

First, we'll use Fermi's golden rule to calculate the rate at which an applied monochromatic electromagnetic field can ionize an atom, say hydrogen, stripping off the electron from a bound state into the continuum of ionized states. We'll assume that the electromagnetic field itself may be treated classically — this assumption is valid e.g. in a laser, or at the focus of the antenna of a radio telescope where the field is large. Of course, if we wish perturbation theory to produce reliable results, we'll still need this applied electromagnetic field to be small compared to the Coulomb fields binding the electron(s) to the nucleus of the atom.

In the vacuum, the electromagnetic field is divergence free and is entirely generated by Faraday's Law $\boldsymbol{\nabla} \times \mathbf{E} = -\partial \mathbf{B}/\partial t$. The whole electromagnetic field can be described by a vector potential $\mathbf{A}$ via

$$\mathbf{b} = \boldsymbol{\nabla} \times \mathbf{A} \quad \text{and} \quad \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}\,. \tag{9.44}$$

In the monochromatic case, Faraday's equation is solved by

$$\mathbf{A}(\mathbf{x}, t) = \boldsymbol{\epsilon}\cos(\mathbf{k}' \cdot \mathbf{x} - \omega t)\,, \tag{9.45}$$

where $\boldsymbol{\epsilon}$ is a constant vector describing the polarization of the wave and $\mathbf{k}'$ is the wavevector. (we reserve $\mathbf{k}$ for later use.) From the vacuum Maxwell equation $\boldsymbol{\nabla} \cdot \mathbf{E} = 0$, we learn that

$$\mathbf{k}' \cdot \boldsymbol{\epsilon} = 0\,, \tag{9.46}$$

saying that the wave is transverse. Faraday's law then reduces to the statement that $\omega = c\left|\mathbf{k}'\right|$ so that the wave travels at the speed of light.

The coupling of an elementary particle of charge q to such a vector potential is given by the minimal coupling rule, replacing

$$\mathbf{P} \longmapsto \mathbf{P} - q\mathbf{A}(\mathbf{X}, t) \tag{9.47}$$

in the Hamiltonian[60], so

$$H = \frac{(\mathbf{P} - q\mathbf{A})^2}{2\mu} + V(\mathbf{X}) = H_0 - \frac{q}{2\mu}(\mathbf{P}\cdot\mathbf{A} - \mathbf{A}\cdot\mathbf{P}) + \frac{q^2\mathbf{A}^2}{2\mu}\,, \tag{9.48}$$

where H_0 contains the original potential term $V(\mathbf{X})$ as well as the kinetic term. In our case of atomic transitions, an electron has charge $q = e$ so to first order in e the perturbing Hamiltonian is

$$\Delta(t) = \frac{e}{2\mu}\left(\mathbf{P}\cdot\boldsymbol{\epsilon}\cos(\mathbf{k}'\cdot\mathbf{X} - \omega t) + \cos(\mathbf{k}'\cdot\mathbf{X} - \omega t)\boldsymbol{\epsilon}\cdot\mathbf{P}\right)\,. \tag{9.49}$$

[60] This is called minimal coupling because more complicated modifications are possible if the charged particle is not elementary, or in other more complicated circumstances. You'll learn more about this in the Part II Classical Dynamics and Applications of Quantum Mechanics.

The component $\boldsymbol{\epsilon}\cdot\mathbf{P}$ of the momentum operator in the direction of the polarization of the electromagnetic wave commutes with $\cos(\mathbf{k}'\cdot\mathbf{X}-\omega t)$ by the transversality condition (9.46) (recall that $\boldsymbol{\epsilon}$ itself is a constant vector). Thus

$$\begin{aligned}\Delta(t) &= \frac{e}{\mu}\cos(\mathbf{k}'\cdot\mathbf{X}-\omega t)\boldsymbol{\epsilon}\cdot\mathbf{P}\\ &= \frac{e}{2\mu}(\mathrm{e}^{\mathrm{i}(\mathbf{k}'\cdot\mathbf{X}-\omega t)}+\mathrm{e}^{-\mathrm{i}(\mathbf{k}'\cdot\mathbf{X}-\omega t)})\boldsymbol{\epsilon}\cdot\mathbf{P}\,.\end{aligned} \tag{9.50}$$

Having found the form of our perturbation, we now need calculate its matrix element between our initial and final states. We'll consider the case where the hydrogen atom is initially in its ground state, so that $m=|1,0,0\rangle$ with position space wavefunction

$$\psi_0(\mathbf{x}) = \langle\mathbf{x}|1,0,0\rangle = \frac{1}{\sqrt{\pi a_0^3}}\mathrm{e}^{-r/a_0}\,, \tag{9.51}$$

where a_0 is the Bohr radius. For the final state, we are interested in the case where the electron has been liberated from the proton. If the energy of the ejected electron is sufficiently great, we can ignore its Coulomb attraction back to the ion and treat the final electron as a free particle of momentum $\hbar\mathbf{k}$, so we take our final state $|\mathbf{k}\rangle$ to be the plane-wave

$$\psi_{\mathbf{k}}(\mathbf{x}) = \langle\mathbf{x}|\mathbf{k}\rangle = \frac{1}{(2\pi\hbar)^{3/2}}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\,. \tag{9.52}$$

This transition requires that the electron absorbs energy from the radiation, so the matrix element we seek comes from the first term of the perturbation Hamiltonian (9.50). We have

$$\begin{aligned}\langle\mathbf{k}|\Delta|100\rangle &= \frac{e}{2\mu}\left\langle\mathbf{k}\middle|\mathrm{e}^{\mathrm{i}\mathbf{k}'\cdot\mathbf{X}}\boldsymbol{\epsilon}\cdot\mathbf{P}\middle|1,0,0\right\rangle\\ &= -\frac{\mathrm{i}\hbar e}{2\mu}\frac{1}{(2\pi\hbar)^{3/2}(\pi a_0^3)^{1/2}}\int_{\mathbb{R}^3}\mathrm{d}^3\mathbf{x}\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\mathrm{e}^{\mathrm{i}\mathbf{k}'\cdot\mathbf{x}}\boldsymbol{\epsilon}\cdot\boldsymbol{\nabla}(\mathrm{e}^{-r/a_0})\,.\end{aligned} \tag{9.53}$$

Integration by parts and using the fact that the bound state wavefunction decays exponentially as $|\mathbf{x}|\to\infty$, this is

$$\langle\mathbf{k}|\Delta|1,0,0\rangle = \frac{e\hbar}{2\mu}\frac{\boldsymbol{\epsilon}\cdot\mathbf{k}}{(2\pi\hbar)^{3/2}(\pi a_0^3)^{1/2}}\int_{\mathbb{R}^3}\mathrm{d}^3\mathbf{x}\,\mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot\mathbf{x}}\mathrm{e}^{-r/a_0}\,, \tag{9.54}$$

where $\mathbf{q}=\mathbf{k}-\mathbf{k}'$ is the momentum transfer and we recall that the polarisation vector is transverse, $\boldsymbol{\epsilon}'\cdot\mathbf{k}'=0$. We're left with just the Fourier transform of the ground state wavefunction. Choosing the z-axis to lie along the direction of $\mathbf{q}$ and working in polar coordinates, this is

$$\int\mathrm{d}^3\mathbf{x}\,\mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot\mathbf{x}}\mathrm{e}^{-r/a_0} = \frac{4\pi}{q}\int_0^\infty\mathrm{d}r\,r\mathrm{e}^{-r/a_0}\sin(qr) = \frac{8\pi a_0^3}{(1+a_0^2q^2)^2}\,, \tag{9.55}$$

where $q=|\mathbf{q}|$.

We now wish to use this matrix element in Fermi's golden rule for the transition rate. Since there are a range of possible final momenta, we define the differential rate $\Gamma_{|1,0,0\rangle\to\mathbf{k}}$ to be the rate at which the radiation ionizes an atom so that the momentum of the ejected electron lies in the range $\hbar(\mathbf{k},\mathbf{k}+\delta k)$, so

$$\begin{aligned}\mathrm{d}\Gamma_{|1,0,0\rangle\to|\mathbf{k}\rangle} &= \frac{2\pi}{\hbar}\left|\langle\mathbf{k}|\Delta|1,0,0\rangle\right|^2\delta(E_{\mathbf{k}}-E_{100}-\hbar\omega)\hbar^3\mathrm{d}^3\mathbf{k}\\ &= 2\pi\hbar^2\left|\langle\mathbf{k}|\Delta|1,0,0\rangle\right|^2\delta(E_{\mathbf{k}}-E_{100}-\hbar\omega)\,\mathrm{d}\Omega\,k^2\mathrm{d}k\,,\end{aligned} \tag{9.56}$$

where $\mathrm{d}\Omega=\sin\theta\mathrm{d}\theta\mathrm{d}\phi$. To relate this differential rate to our previous expression for the rate of transition into states of particular energy, recall that we're treating the final electron as a free particle, so its energy is $E_{\mathbf{k}}=\hbar^2\mathbf{k}^2/2\mu$. Thus we have

$$k^2\mathrm{d}k = k^2\left(\frac{\mathrm{d}E_k}{\mathrm{d}k}\right)^{-1}\mathrm{d}E_k = \left(\frac{\mu k}{\hbar^2}\right)\mathrm{d}E_k\,, \tag{9.57}$$

so for a free particle, the density of states $n(E_k) = \mu k/\hbar^2 = \mu\sqrt{2\mu E_k}/\hbar^3$. Putting all the pieces together and integrating over the final state energy, the transition rate for electrons to be emitted into the direction $\mathrm{d}\Omega$ of solid angle is

$$\frac{\mathrm{d}\Gamma_{|1,0,0\rangle\to|\mathbf{k}\rangle}}{\mathrm{d}\Omega} = \frac{4e^2(ka_0)^3}{\pi\mu\hbar}\frac{\left|\boldsymbol{\epsilon}\cdot\hat{\mathbf{k}}\right|^2}{(1+a_0^2q^2)^4}\,, \tag{9.58}$$

where $\hbar k = \sqrt{2\pi E_k}$ is frozen to $\sqrt{2\pi(\hbar\omega + E_{100})} \approx \sqrt{2\mu\hbar\omega}$.

9.4.2 Absorption and Stimulated Emission

Now let's consider the case where instead of being illuminated by monochromatic light such as from a laser, our atom is immersed in a bath of thermal radiation. This radiation bath supplies us with a continuous range of different frequencies, so transitions between different (discrete) bound states of the atom are now possible.

The problem simplifies if we assume that the typical electromagnetic wavelength $1/k'$ in the thermal bath is much larger than the Bohr radius of the atom. This is a good approximation provided the atomic transition occurs between states that are separated in energy by much less than $\alpha\mu c^2$, as will be the case for transitions between highly excited energy levels, perhaps by the valence electrons in a heavy atom. Such wavelengths correspond to radiation with frequencies that are less than those of soft X-rays. In this approximation, $\mathbf{k}'\cdot\mathbf{X} \ll 1$ for all locations in the atom or molecule at which there is significant probability of finding the electron. To lowest order, we can thus approximate the electric field as being constant over the scale of the atom. The Hamiltonian for the atom is then

$$H = H_0 + e\sum_r \mathbf{E}(t)\cdot\mathbf{X}_r\,, \tag{9.59}$$

where H_0 is the Hamiltonian of the unperturbed atom, $\mathbf{E}(t)$ is the fluctuating electric field due to the radiation and $\mathbf{X}_r$ is the position operator for the rth electron in the atom. This is known as the *dipole approximation* since the operator $e\sum_r \mathbf{X}_r$ represents the atom's dipole moment. We interpret the transition as occurring due to the interaction between the applied electric field and this dipole.

Whilst roughly constant over the atom, the electric field $\mathbf{E}(t)$ fluctuates rapidly in time as the atom is jostled by radiation of different frequencies in the thermal bath. In particular, on average $\overline{\mathbf{E}(t)} = 0$ since the electric field is equally likely to point in any direction at any given moment. We take $\mathbf{E}(t)$ to be correlated as

$$\overline{E_i(t_1)E_j(t_2)} = \delta_{ij}\int_{\mathbb{R}} \mathrm{d}\omega\,\mathcal{P}(\omega)\mathrm{e}^{-\mathrm{i}\omega(t_1-t_2)}\,. \tag{9.60}$$

The presence of δ_{ij} means that the fluctuation in different directions are uncorrelated, whilst fluctuations that are aligned are correlated equally no matter their direction. This just says that there is no preferred direction for the electric field and holds e.g. for a thermal bath of radiation.

We can understand the meaning of the function $\mathcal{P}(\omega)$ as follows. First, note that if the electric fields are real, then we must have

$$\mathcal{P}(-\omega) = \mathcal{P}(\omega) \quad \text{and} \quad \mathcal{P}(\omega)^* = \mathcal{P}(\omega) \tag{9.61}$$

so that $\mathcal{P}$ is a real, even function. Recall from IB electromagnetism that the energy density of an electromagnetic field is

$$\frac{\epsilon_0}{2}(\mathbf{E}^2(\mathbf{x},t) + c^2\mathbf{B}^2(\mathbf{x},t)) \tag{9.62}$$

and in purely radiative (source-free) field $\mathbf{E}^2 = c^2\mathbf{B}^2$. In our case, (9.60) shows that the average energy density we expect to find in the radiation at time t is

$$\epsilon_0\overline{E_i(t)E_i(t)} = 3\epsilon_0\int_{-\infty}^{\infty}\mathrm{d}\omega\,\mathcal{P}(\omega) = 6\epsilon_0\int_0^{\infty}\mathrm{d}\omega\,\mathcal{P}(\omega) \tag{9.63}$$

so $\rho(\omega) = 6\epsilon_0 \mathcal{P}(\omega)$ represents the average energy density in radiation at energy time due to frequencies in the range $(\omega, \omega + \mathrm{d}\omega)$. (If the radiation is truly thermal at temperature T, then $\rho(\omega)$ would be given by the Planck's blackbody formula

$$\rho(\omega) = \frac{\hbar\omega^3}{\pi^3 c^2} \frac{1}{\mathrm{e}^{-\hbar\omega/k_B T} - 1} \tag{9.64}$$

coming from the Bose–Einstein distribution. You'll learn about this in the Statistical Physics course.)

Now let's put these results to work. For a transition $|m\rangle \to |k\rangle$, we have

$$a_k(t) = -\frac{ie}{\hbar} \int_0^t \mathrm{d}t' \sum_r \langle k|\mathbf{E}(t') \cdot \mathbf{X}_r|m\rangle \, \mathrm{e}^{\mathrm{i}\omega_{km} t'} \,. \tag{9.65}$$

We write the probability that the atom is found in state $|k\rangle$ at time t as a double integral

$$|a_k(t)|^2 = \frac{e^2}{\hbar^2} \int_0^t \int_0^t \mathrm{d}t_1 \, \mathrm{d}t_2 \sum_{r,r'} \langle k|\mathbf{E}(t_1) \cdot \mathbf{X}_{r'}|m\rangle \, \langle m|\mathbf{E}(t_2) \cdot \mathbf{X}_r|k\rangle \, \mathrm{e}^{\mathrm{i}\omega_{km}(t_1 - t_2)} \,. \tag{9.66}$$

The correlation (9.60) in our thermal bath shows that, on average,

$$\begin{aligned}
\overline{|a_k(t)|^2} &= \frac{e^2}{\hbar^2} \int_0^t \int_0^t \mathrm{d}t_1 \, \mathrm{d}t_2 \, \overline{E_i(t_1) E_j(t_2)} \sum_{r,r'} \langle k|(X_{r'})_i|m\rangle \, \langle m|(X_r)_j|k\rangle \, \mathrm{e}^{\mathrm{i}\omega_{km}(t_1 - t_2)} \\
&= \frac{e^2}{\hbar^2} \left| \sum_r \langle k|\mathbf{X}_r|m\rangle \right|^2 \times \int_{-\infty}^{\infty} \mathrm{d}\omega \, \mathcal{P}(\omega) \int_0^t \int_0^t \mathrm{d}t_1 \, \mathrm{d}t_2 \, \mathrm{e}^{\mathrm{i}(\omega_{km} - \omega)(t_1 - t_2)} \\
&= \frac{e^2}{\hbar^2} \left| \sum_r \langle k|\mathbf{X}_r|m\rangle \right|^2 \times \int_{-\infty}^{\infty} \mathrm{d}\omega \, \mathcal{P}(\omega) \left| \int_0^t \mathrm{d}t_1 \, \mathrm{e}^{\mathrm{i}(\omega_{km} - \omega)t_1} \right|^2 \\
&= \frac{4e^2}{\hbar^2} \left| \sum_r \langle k|\mathbf{X}_r|m\rangle \right|^2 \times \int_{-\infty}^{\infty} \mathrm{d}\omega \, \mathcal{P}(\omega) \frac{\sin^2\left(\frac{(\omega_{km} - \omega)t}{2}\right)}{(\omega_{km} - \omega)^2} \,,
\end{aligned} \tag{9.67}$$

where the mod-square includes the Euclidean inner product of the two $\langle k|\mathbf{X}_r|m\rangle$ matrix elements. This form is again familiar, except that instead of a density of final states, we have $\mathcal{P}(\omega)$, representing the average density $\rho(\omega)$ of the frequency ω component of the radiation. Making our familiar placement, and performing the integral over ω using the resulting δ-function in the $t \to \infty$ limit, we have the transition rate

$$\Gamma(|m\rangle \to |k\rangle) = \frac{\pi e^2}{3\epsilon_0 \hbar^2} \left| \sum_r \langle k|\mathbf{X}_r|m\rangle \right|^2 \rho(\omega_{km}) \,. \tag{9.68}$$

in the case that the states $|m\rangle$ and $|k\rangle$ are discrete but we have a continuous range of perturbing frequencies.

This result is very intuitive: the rate of transition depends on the probability that the dipole moment transition can link states $|m\rangle$ and $|k\rangle$, together with the energy density in the radiation bath at the right frequency to match this transition. For hydrogenic states of the atom, we recall that the matrix element vanishes unless the selection rules

$$|\ell - \ell'| = 1 \quad \text{and} \quad |m - m'| \leq 1 \tag{9.69}$$

are obeyed. These selection rules were derived under the assumption that the transition is just due to the dipole approximation where the perturbation involves matrix elements of the dipole moment operator $e\mathbf{X}$. The effects of both higher orders in perturbation theory, and inclusion of interactions beyond those of an electric field with wavelength $\lambda \gg a_0$ mean that transitions that are 'forbidden' according to the above selection rules may in fact occur. They will typically do so, however, at a much smaller rate.

Now, whether or not the radiation bath is thermal, the reality conditions (9.61) show that the rate $\Gamma(|m\rangle \to |k\rangle)$ in (9.68) is an even function of ω_{km}. Consequently, for fixed $|E_k - E_m|$, this rate is the same whether $E_k > E_m$ or $E_k < E_m$. For $E_k > E_m$, the atom absorbs energy from the radiation, exciting to a higher energy level, whilst if $E_k < E_m$ the atom emits energy back into the radiation as it decays to a lower energy state. The important result we've obtained is that the rate for stimulated emission is the same as the rate of absorption.

9.4.3 Spontaneous Emission

Our treatment above calculated the rate at which an atom would decay (or excite) due to being immersed in a bath of radiation. However, we currently cannot understand decays of an isolated atom: if the atom is prepared in any eigenstate of H_0, quantum mechanics says that in the absence of a perturbation it will just stay there. Remarkably, Einstein was able to relate the rate of spontaneous decay of an atom to the rate of stimulated decay that we calculated above.

Suppose our atom is immersed in a bath of radiations where the energy density of photons with frequency in the range $(\omega + \mathrm{d}\omega)$ is $\rho(\omega)\mathrm{d}\omega$. Einstein defined $A_{m\to k}(\omega_{mk})$ as the rate at which an atom would spontaneously decay from a state of energy E_m to a state of lower energy E_k, where $\omega_{mk} = (E_m - E_k)/\hbar$ is the frequency of the emitted photon. We let $\rho(\omega_{km})B_{k\to m}(\omega_{km})$ denote the rate at which energy is absorbed from the radiation bath, exciting the atom from $|k\rangle$ to $|m\rangle$. Similarly, let $\rho(\omega_{km})B_{m\to k}(\omega_{km})$ denote the rate of decay of the excited atom due to stimulation by the presence of the radiation bath. We calculated $B_{m\to k}$ and $B_{k\to m}$ in the dipole approximation above, finding they were the same, but let's pretend that (like Einstein) we don't know this yet.

In thermodynamic equilibrium these rates must balance, so if there are n_k atoms in state $|k\rangle$ and n_m in state $|m\rangle$, we must have

$$n_m(A_{m\to k}(\omega_{mk}) + \rho(\omega_{km})B_{m\to k}(\omega_{km})) = n_k\rho(\omega_{km})B_{k\to m}(\omega_{km})\,. \tag{9.70}$$

We will borrow two results you will derive in Part II Statistical Physics. First, when a gas of atoms is in equilibrium at temperature T, the relative numbers of atoms in states with energies E_m and E_k is given by the *Boltzmann distribution*

$$\frac{n_k}{n_m} = \frac{\mathrm{e}^{-E_k/k_BT}}{\mathrm{e}^{-E_m/k_BT}} = \mathrm{e}^{\hbar\omega_{mk}/k_BT}\,, \tag{9.71}$$

where $k_B \approx 1.38\times10^{-23}\ \mathrm{J\,K^{-1}}$ is the *Boltzmann's constant*. Furthermore, if radiation is in equilibrium at the same temperature then the density of states of photons at frequency ω is

$$\rho(\omega) = \frac{\hbar\omega^3}{\pi^2c^3}\frac{1}{\mathrm{e}^{\hbar\omega_{mk}/k_BT} - 1}\,, \tag{9.72}$$

which is essentially the Bose–Einstein distribution. Using these in (9.70) gives

$$A_{m\to k}(\omega) = \frac{\hbar\omega^3}{\pi^2c^3}\frac{1}{\mathrm{e}^{\hbar\omega/k_BT} - 1}\left(\mathrm{e}^{\hbar\omega/k_BT}B_{k\to m}(\omega) - B_{m\to k}(\omega)\right)\,, \tag{9.73}$$

where $\omega = \omega_{m\to k}$. However, $A_{m\to k}$ is supposed to be the rate of spontaneous emission of radiation from the atom, so it cannot depend on any properties of the radiation bath. In particular, it must be independent of the temperature T. This is possible iff

$$B_{k\to m}(\omega) = B_{m\to k}(\omega)\,, \tag{9.74}$$

so that the rates of stimulated emission and absorption agree. In this case, we have

$$A_{m\to k}(\omega) = \frac{\hbar\omega^3}{\pi^2c^3}B_{m\to k}(\omega) \tag{9.75}$$

so that knowing the rate $B_{k\to m}\,|\omega\rangle$ at which a classical electromagnetic wave of frequency $\omega = (E_m - E_k)/\hbar$ is absorbed by an atom, we can also calculate the rate $A_{m\to k}(\omega)$ at which the atom spontaneously decays, even in the absence of radiation.

Now we have found from first-principle calculations that

$$\Gamma(|m\rangle \to |k\rangle) = \frac{\pi e^2}{3\epsilon_0\hbar^2}\left|\sum_r \langle k|\mathbf{X}_r|m\rangle\right|^2 \rho(\omega_{km}) \tag{9.76}$$

in the dipole approximation, to first-order in perturbation theory. Consequently, for our atom Einstein's B coefficient is

$$E_{k\to m}(\omega_{km}) = B_{m\to k}(\omega_{km}) = \frac{\pi e^2}{3\epsilon_0\hbar^2}\left|\sum_r \langle k|\mathbf{X}_r|m\rangle\right|^2 \tag{9.77}$$

for stimulated emission or absorption. From Einstein's statistical argument we thus learn that

$$A_{m\to k}(\omega) = \frac{e^2\omega^3}{3\epsilon_0\pi\hbar c^3}\left|\sum_r \langle k|\mathbf{X}_r|m\rangle\right|^2 \tag{9.78}$$

is the rate at which an excited atom must spontaneously decay, even in the absence of any applied radiation.

Ingenious though it is, there's something puzzling about this calculation of the rate of spontaneous decay. Where does our quantum mechanical calculation stating that isolated excited atoms do not decay go wrong? The answer is that while we treated the energy levels of the atom fully quantum mechanically, the electromagnetic field itself was treated classically. Spontaneous emission is only possible once one also quantizes the electromagnetic field, since it is fluctuations in the zero-point value of this field that allow the atom to emit a photon and decay; heuristically, the fluctuating value of the quantum electromagnetic field — even in the vacuum — 'tickles' the excited atom prompting the decay. Treating the EM field classically is appropriate if the radiation comes from a high intensity laser, where the energy density of the field is so high that the terms $B(\omega)\rho(\omega)$ dominate in (9.70). By contrast, emission of light from the atoms in a humble candle is an inherently quantum phenomenon, occurring purely by spontaneous emission once the flame is lit — candlelight shines even in ambient darkness.

Einstein gave his statistical argument in 1916, when Bohr's model of the atom (the first thing you met in IB QM) was still the deepest understanding physicists had of quantum theory. This was at the same time as he was revolutionising our understanding of gravity, space and time, which obviously wasn't enough to keep him fully occupied. The quantum treatment of stimulated emission given in the previous section is due to Dirac in 1926, the same year as Schrödinger first published his famous equation. A first principles calculation of the spontaneous emission rate $A_{m\to k}$, requiring the quantization of the electromagnetic field, was also given by Dirac just one year later. Things move fast. Dirac's 1927 results agreed with those we've found via Einsteins argument. The necessity also to quantize the electromagnetic field heralded the arrival of Quantum Field Theory.

Let's use these results to calculate the typical lifetime of an excited state of an isolated atom. When the radiation density $\rho(\omega)$ is very small, the number of atoms n_m in the excited state obeys

$$\frac{\partial n_m}{\partial t} = -A_{m\to k} n_m \tag{9.79}$$

so the atom decays exponentially with a characteristic timescale $\sim A_{m\to k}^{-1}$. For hydrogen or an alkali atom, typically the outermost electron changes its energy level in the transition, so we can drop the sum in (9.78) to find

$$A_{m\to k}(\omega) = \frac{e^2\omega^3}{3\pi\epsilon_0\hbar c^3}\,|\langle k|\mathbf{X}|m\rangle|^2 \tag{9.80}$$

where $\mathbf{X}$ is the position of the outermost electron. Unless $\langle k|X|m\rangle$ vanishes due to some selection rule, we'd expect this matrix element to be of the typical size $a_0 = 4\pi\epsilon_0\hbar^2/\mu e^2$ characteristic of the atom, so $|\langle k|X|m\rangle|^2 \sim a_0^2$. Hence the characteristic lifetime of the excited atomic state is roughly

$$\tau = A_{m\to k}^{-1} \sim \frac{3\pi\epsilon_0\hbar c^3}{e^2\omega^3 a_0^2} = \frac{3\mu c^3}{4\hbar\omega^3 a_0} \tag{9.81}$$

where μ is the mass of the electron. Optical light has a frequency of order $\omega \sim 10^{15}$ Hz so the timescale τ is around 10^7 times longer than the timescale E_{200} associated with the energy E_{200} of the first excited state of hydrogen. Thus around 10^7 oscillations of the hydrogen atom occur before it radiates a photon, decaying down into the ground state.

10 Interpreting Quantum Mechanics

In this final chapter, we will explore the process of *measurement* in quantum mechanics. According to the Copenhagen interpretation, when we perform a measurement the state of the particle collapses onto an eigenstate of the corresponding operator, with the probability of different results being given by the Born rule (2.74). This entails a departure from the unitary (and hence deterministic) time evolution of our system as described by Schrödinger's equation. However, the Copenhagen interpretation does not tell us exactly what physical process should count as a 'measurement': does the observer need to be alive? to be human? to have taken PQM? Without such a prescription, how can we know when it is appropriate to evolve our state as $|\psi(t)\rangle = U(t)\,|\psi(0)\rangle$ and when instead it should collapse?

A further problem with measurement is that the states we measure usually correspond to what we consider to be 'classically sensible' quantities. This seems to imply that measurements involve a preferred choice of basis for the system's Hilbert space. To give a famous example, we are all familiar with living cats and dead cats, but no-one has ever seen a cat that is simultaneously alive + dead. But why should quantum mechanics distinguish the basis

$$\{|\text{alive}\rangle\,, |\text{dead}\rangle\} \quad \text{over} \quad \left\{\frac{|\text{alive}\rangle + |\text{dead}\rangle}{\sqrt{2}}, \frac{|\text{alive}\rangle - |\text{dead}\rangle}{\sqrt{2}}\right\}? \tag{10.1}$$

In this chapter, we'll examine these issues from the perspective of *decoherence*. The formalism presented here is completely standard, and indeed decoherence is an important, well-established property of any quantum system. However, the jury is still out on whether this finally resolves the infamous problems with measurement in quantum mechanics.

10.1 The Density Operator

To get started, we must first realise that during a measurement, we cannot treat our quantum system as being isolated. Any form of measurement requires that we bring the system under study into contact with some form of measuring apparatus. Up to this point, we've assumed that we know the precise quantum state our system is in. While this may be possible for a small, isolated quantum system, we cannot possibly hope to know the exact quantum state of a macroscopic measuring device, which may easily contain $\sim 10^{23}$ atoms. Thus, to talk about measurements, we first need a way to describe systems even when we're not sure which state they're in. In fact, even in purely classical systems, there's always some uncertainty in our knowledge of the system: we never actually know the momentum of a single particle with infinite precision even in classical mechanics, because all our measurements are subject to some experimental error.

Let's now see how to incorporate such imprecision in our knowledge into quantum mechanics. Suppose we know only that our system is in one of the states $\{|\psi_\alpha\rangle\}$, and that the probability it is in state $|\psi_\alpha\rangle$ is p_α. It's important to be clear that we're not saying

$$|\Psi\rangle = \sum_\alpha \sqrt{p_\alpha}\,|\psi_\alpha\rangle \tag{10.2}$$

because $|\Psi\rangle$ itself is a well-defined quantum state. Rather, we're admitting that we don't know the true state of the system, which could be any of the states $\{|\psi_\alpha\rangle\}$. Indeed, these states do not need to form a complete set, and do not even need to be orthogonal, although we will take them each to be correctly normalized $\langle\psi_\alpha|\psi_\alpha\rangle = 1$ for each α.

In this case, the average result we obtain when measuring the value of some observable Q is

$$\overline{Q} = \sum_\alpha p_\alpha\,\langle\psi_\alpha|Q|\psi_\alpha\rangle\ . \tag{10.3}$$

This expression combines the quantum expectation value of Q in the state $|\psi_\alpha\rangle$ (which may not be an eigenstate of Q), together with our lack of knowledge of the system's state, represented by the p_α's. For future use, it'll be convenient to describe this using a *density operator* $\rho : \mathcal{H} \to \mathcal{H}$, defined by

$$\rho = \sum_\alpha p_\alpha \left|\psi_\alpha\right\rangle \left\langle\psi_\alpha\right| , \tag{10.4}$$

where the p_α are the probabilities introduced above. Then we can write the average (10.3) as

$$\overline{Q} = \mathrm{tr}_{\mathcal{H}}(\rho Q) . \tag{10.5}$$

To see this, suppose $\{|q_n\rangle\}$ is a complete set of eigenstates of Q, with eigenvalues $\{q_n\}$. Using $\{|q_n\rangle\}$ as a basis for $\mathcal{H}$ we have

$$\begin{aligned} \mathrm{tr}_{\mathcal{H}}(\rho Q) &= \sum_n \langle q_n|\rho Q|q_n\rangle = \sum_{n,\alpha} p_\alpha \langle q_n|\psi_\alpha\rangle \langle\psi_\alpha|Q|q_n\rangle \\ &= \sum_{n,\alpha} p_\alpha q_n \left|\langle q_n|\psi_\alpha\rangle\right|^2 = \sum_\alpha p_\alpha \langle\psi_\alpha|Q|\psi_\alpha\rangle . \end{aligned} \tag{10.6}$$

The density operator has the following three properties: First, it is a Hermitian operator

$$\rho^\dagger = \rho, \tag{10.7}$$

reflecting the fact that the probabilities p_α must be real. Second,

$$\mathrm{tr}_{\mathcal{H}}(\rho) = 1 \tag{10.8}$$

since the probabilities sum to one, and third,

$$\langle\psi|\rho|\psi\rangle \geq 0 \tag{10.9}$$

for all $|\psi\rangle \in \mathcal{H}$ since probabilities are non-negative. We often write this property as $\rho \geq 0$ for short. In fact, these three properties can be taken to be the defining properties of a density operator, in the sense that any operator obeying these three properties is the density operator for some system. To see this, suppose the eigenvectors of ρ are $|\phi_r\rangle$, with $\rho\,|\phi_r\rangle = \rho_r\,|\phi_r\rangle$. Then since $\rho = \rho^\dagger$ we have $\rho_r \in R$. The remaining two properties tell us that $\sum_r \rho_r = 1$ and $\rho_r \geq 0$. Any set of real numbers ρ_r obeying these conditions can be taken to be a probability distribution for some system. Note that since the $|\phi_r\rangle$ are eigenvectors of the Hermitian operator ρ, they're necessarily orthogonal, in contrast to the arbitrary set of states we used in (10.3).

If we have perfect knowledge of our system, meaning $\rho = |\psi\rangle\langle\psi|$ so that with probability 1 the system is in state $|\psi\rangle$, then we refer to it as *pure*. Correspondingly, if our knowledge of the state is incomplete, so that $\rho = \sum_\alpha p_\alpha |\psi_\alpha\rangle\langle\psi_\alpha|$ with more than one $p_\alpha > 0$, we say that the system is in an *impure* or *mixed* state. This terminology is somewhat misleading, because it is really just our knowledge that is incomplete — the system itself is presumably in some perfectly well-defined quantum state, it's just that we don't know which one.

The density operator and the operators for the Hamiltonian and other observables encapsulate a complete, self-contained theory of quantum dynamics. If we have incomplete knowledge of the system's quantum state, then use of this formalism is mandatory. If we do happen to know that our system is initially in the precise quantum state $|\psi\rangle$, we can still use this apparatus by setting $\rho = |\psi\rangle\langle\psi|$, rather than using time dependent Schrödinger's equation, though in this case use of the density operator is optional.

10.1.1 The Bloch Sphere

As a simple example, consider a single spin-$\frac{1}{2}$ particle with $\{|\uparrow\rangle, |\downarrow\rangle\}$ forming a basis of $\mathcal{H} \cong \mathbb{C}^2$. If we know for sure that the system is in the state $|\uparrow\rangle$, then

$$\rho = |\uparrow\rangle\langle\uparrow| . \tag{10.10}$$

However, if we think there's only a $\frac{1}{2}$ chance that the system is actually in state $|\uparrow\rangle$, with a $\frac{1}{2}$ chance it might actually be in the state $|\downarrow\rangle$, then

$$\rho = \frac{1}{2}|\uparrow\rangle\langle\uparrow| + \frac{1}{2}|\downarrow\rangle\langle\downarrow| = \frac{1}{2}1_{\mathcal{H}}\,. \tag{10.11}$$

In this case, the average value of the spin along an axis is $\mathrm{tr}_{\mathcal{H}}(\rho\mathbf{S}) = 0$ and we'll see later that having ρ proportional to the identity matrix means we're maximally ignorant about the state of our system. As a further example, let $|\uparrow_x\rangle$ denote the eigenstate $S_x|\uparrow_x\rangle = \frac{\hbar}{2}|\uparrow_x\rangle$ of spin-up along the x-axis. Then if we think there's probability $\frac{1}{2}$ our system is in state $|\uparrow\rangle$ and probability $\frac{1}{2}$ it is in the (non-orthogonal) state $|\uparrow_x\rangle$, then

$$\begin{aligned}\rho &= \frac{1}{2}|\uparrow\rangle\langle\uparrow| + \frac{1}{2}|\uparrow_x\rangle\langle\uparrow_x| \\ &= \frac{1}{2}|\uparrow\rangle\langle\uparrow| + \frac{1}{4}(|\uparrow\rangle\langle\downarrow|)(|\uparrow\rangle\langle\downarrow|) \\ &= \frac{1}{4}1_{\mathcal{H}} + \frac{1}{2}|\uparrow\rangle\langle\uparrow| + \frac{1}{4}|\uparrow\rangle\langle\downarrow| + \frac{1}{4}|\downarrow\rangle\langle\downarrow|\,,\end{aligned} \tag{10.12}$$

where we've used the result $|\uparrow_x\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle)$ (can be verified by acting $S_x = (S_+ + S_-)/2$). With this density matrix, we find

$$\overline{S_x} = \overline{S_z} = \frac{\hbar}{2} \quad \text{while} \quad \overline{S_y} = 0\,, \tag{10.13}$$

reflecting the fact that we're more likely to have spin up than down along both the x and z-axes, but know nothing about the spin along the y-axis.

More generally, since any 2×2 Hermitian matrix can be written as a linear combination of the identity matrix and the Pauli sigma matrices, we can write

$$\rho = \frac{1}{2}(1_{\mathcal{H}} + \mathbf{b}\cdot\boldsymbol{\sigma}) \tag{10.14}$$

for some vector $\mathbf{b}$, where we've used the fact that $\mathrm{tr}_{\mathcal{H}}(\boldsymbol{\sigma}) = 0$ and the condition $\mathrm{tr}_{\mathcal{H}}\,\rho = 1$ to fix the overall factor. Since $1 = \mathrm{tr}_{\mathcal{H}}\,\rho$ is the sum of its eigenvalues, at least one must be positive. The other will also be non-negative as required for their interpretation as probabilities. This is true if

$$\det\rho = \frac{1}{4}(1 - \mathbf{b}\cdot\mathbf{b}) \geq 0\,. \tag{10.15}$$

Hence (10.14) is a well-defined density operator for our two-state system provided

$$|\mathbf{b}| \leq 1\,. \tag{10.16}$$

This condition is known as the *Bloch Ball*. Density matrices with $|\mathbf{b}| = 1$ on the *Bloch Sphere* correspond to pure states, where the system definitely has spin $+\hbar/2$ along the $\hat{\mathbf{b}}$-axis. On the other hand, states with $|\mathbf{b}| < 1$ must have both eigenvalues strictly positive, so there is no way to write such a density matrix as $|\uparrow_{\mathbf{n}}\rangle|\uparrow_{\mathbf{n}}\rangle$ for any direction $\mathbf{n}$. For both mixed and pure states, the direction of $\mathbf{b}$ is said to define the *polarization* of the state: for any $\mathbf{b} \neq \mathbf{0}$, measurements of the spin will be preferentially aligned along $\mathbf{b}$.

10.2 Entropy

For pure states, where $\rho = |\psi\rangle\langle\psi|$ for some $|\psi\rangle$, it is easy to see that

$$\rho^n = \rho \tag{10.17}$$

provided $|\psi\rangle$ is normalised. If an impure density operator has $p_\alpha = 0.99999999$ for some particular $|\psi_\alpha\rangle$, with the remaining probability spread in some way among other states, then although our

knowledge of the system's state is imperfect, the effects of the impurity are likely to be negligible. On the other hand, if the density operator has $p_\alpha \leq 10^{-20}$ for every α so that a tremendous number of states are equally probable, then our knowledge of the true quantum state of a system is very poor indeed.

We'd like to have a way to quantify how much knowledge, or information, about a state we have once the probability distribution $\{p_i\}$ has been specified. To achieve this, define the *von Neumann entropy* $S(\rho)$ associated to a density operator by

$$S(\rho) = -\operatorname{tr}_{\mathcal{H}}(\rho \ln \rho)\,. \tag{10.18}$$

If $\{|\phi_r\rangle\}$ are the orthonormal eigenvectors of ρ with eigenvalues $\{\rho_r\}$ then we can write

$$\ln \rho = \sum_r \ln(\rho_r)\,|\phi_r\rangle\,\langle\phi_r|\,. \tag{10.19}$$

Choosing any basis $\{|n\rangle\}$ for $\mathcal{H}$, we thus have

$$\begin{aligned} -\operatorname{tr}_{\mathcal{H}}(\rho\ln\rho) &= -\sum_n \left\langle n \middle| \left(\sum_r \rho_r\,|\phi_r\rangle\,\langle\phi_r|\right)\left(\sum_{r'} \ln(\rho_{r'})\,|\phi_{r'}\rangle\,\langle\phi_{r'}|\right) \middle| n \right\rangle \\ &= -\sum_{n,r} \rho_r \ln(\rho_r)\,|\langle\phi_r|n\rangle|^2 = -\sum_r \rho_r \ln(\rho_r) \end{aligned} \tag{10.20}$$

in terms of the eigenvalues of the density operator.

Since $0 \leq \rho_r \leq 1$, it's easy to see from the form (10.20) that $S(\rho) \geq 0$ with $S(\rho) = 0$ if and only if ρ describes a pure state, where only one of the ρ_r's is non-zero (and hence equal to 1) as we have complete certainty about which state our system is in. We also claim that the maximum value of $S(\rho)$ is attained iff

$$\rho = \rho_{\max} = \frac{1}{\dim(\mathcal{H})} 1_{\mathcal{H}} \tag{10.21}$$

for a finite dimensional Hilbert space. When $\rho = \rho_{\max}$ all states are equally likely — meaning we have no idea about which state our system is actually in. To see that this indeed maximises the entropy, use the method of Lagrange multipliers to impose the constraint $\operatorname{tr}_{\mathcal{H}}(\rho) = 1$ and vary

$$S(\rho) - \lambda(\operatorname{tr}_{\mathcal{H}}(\rho) - 1) \tag{10.22}$$

with respect to the probabilities and Lagrange multiplier λ. At an extremum,

$$\begin{cases} 0 = -\operatorname{tr}_{\mathcal{H}}[\delta\rho \ln\rho + \rho\rho^{-1}\delta\rho + \lambda\delta\rho] \\ 0 = \delta(\lambda)(\operatorname{tr}_{\mathcal{H}}(\rho) - 1)\,. \end{cases} \tag{10.23}$$

In the first line, we've used the fact that $\operatorname{tr}(\rho\rho^{-1}\delta\rho) = \operatorname{tr}(\rho\delta\rho\rho^{-1})$ inside the trace, so that the order of the variation in the logarithm doesn't matter. These equations must hold for arbitrary variations $\delta\rho$ and $\delta\lambda$, so the first tells us that

$$\rho = \mathrm{e}^{-\lambda-1} 1_{\mathcal{H}} \tag{10.24}$$

for some constant $\mathrm{e}^{-\lambda-1}$. Taking the trace, the second equation fixes the constant of proportionality so that

$$\rho = \rho_{\max} = \frac{1}{\dim(\mathcal{H})} 1_{\mathcal{H}} \tag{10.25}$$

as claimed. The corresponding maximum entropy is

$$\begin{aligned} S(\rho_{\max}) &= -\operatorname{tr}_{\mathcal{H}}(\rho_{\max}\ln\rho_{\max}) \\ &= -\frac{\operatorname{tr}_{\mathcal{H}}(1_{\mathcal{H}})}{\dim(\mathcal{H})}\ln(\dim(\mathcal{H})^{-1}) = \ln\dim(\mathcal{H})\,. \end{aligned} \tag{10.26}$$

Because it was defined as a trace, $S(\rho)$ doesn't depend on which basis we use to describe our Hilbert space.

10.2.1 The Gibbs Distribution

One of the main uses of the density operator and von Neumann entropy is in Quantum Statistical Mechanics. As an example, suppose we wish to extremise the entropy subject to both $tr_{\mathcal{H}}(\rho) = 1$ and $tr_{\mathcal{H}}(\rho H) = U$, saying that we know our system has a fixed average energy U. Then using two Lagrange multipliers λ and β, at an extremum we have

$$0 = \delta[S(\rho) - \lambda(\mathrm{tr}(\rho) - 1) - \beta(\mathrm{tr}(\rho H) - U)]\,, \tag{10.27}$$

which gives three conditions

$$\begin{cases} 0 = -\,\mathrm{tr}_{\mathcal{H}}[\delta\rho(\ln\rho + 1 + \beta H + \lambda)] \\ 0 = \delta\lambda(\mathrm{tr}_{\mathcal{H}}(\rho) - 1) \\ 0 = \delta\beta(\mathrm{tr}_{\mathcal{H}}(\rho H) - U)\,. \end{cases} \tag{10.28}$$

Since these must hold for arbitrary variations, the first equation gives

$$\rho = \mathrm{e}^{-\beta H}\mathrm{e}^{-\lambda-1} \tag{10.29}$$

at a maximum of $S(\rho)$ with fixed energy. The other two conditions just enforce our constraints: to ensure $\mathrm{tr}_{\mathcal{H}}(\rho) = 1$ we must set $\mathrm{e}^{\lambda+1} = Z(\beta)$ where the constant

$$Z(\beta) = \mathrm{tr}_{\mathcal{H}}(\mathrm{e}^{-\beta H}) \tag{10.30}$$

is known as the *partition function* of our system. Thus, in a state of maximum entropy for fixed average energy, the density operator takes the form

$$\rho = \frac{1}{Z(\beta)}\mathrm{e}^{-\beta H} = \frac{1}{Z(\beta)}\sum_n \mathrm{e}^{-\beta E_n}\,|E_n\rangle\,\langle E_n| \tag{10.31}$$

where in the final expression we have inserted a complete set of H eigenstates. This form of density operator is known as the *Gibbs distribution*. It plays a fundamental role in quantum statistical mechanics. β is usually denoted $1/k_BT$ where T is called the temperature and k_B the Boltzmann constant. For fixed average energy U of the system, the temperature is determined by the constraint $\mathrm{tr}_{\mathcal{H}}(\rho H) = U$. In other words, the temperature T is determined by the average energy of the system. You'll work much more with the density operator and entropy in Part II Statistical Physics.

10.3 Reduced Density Operators

If our system comprises two (or more) identifiable subsystems A and B, then $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ so the full Hilbert space is the tensor product of the Hilbert spaces of the subsystems. Recall that a state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is called *entangled* if it cannot be written as a single product $|\phi\rangle \otimes |\psi\rangle$ of states $|\phi\rangle \in \mathcal{H}_A$ and $|\psi\rangle \in \mathcal{H}_B$.

We'll suppose A describes the system we're really interested in, whilst B is the 'environment'. That is, B describes the quantum state of everything in the Universe except our immediate object of study A. Of course, we can't hope to know the precise quantum state of B.

We're going to be interested in the average value we obtain for measurements of a quantity Q that is an observable purely of A, represented on the full Hilbert space by $Q \otimes 1_B$, when the whole Universe is described by a density operator ρ_{AB}. We have

$$\overline{Q} = \mathrm{tr}_{\mathcal{H}_A\otimes\mathcal{H}_B}((Q \otimes 1_B)\rho_{AB}) = \mathrm{tr}_{\mathcal{H}_A}(Q_{\rho_A})\,, \tag{10.32}$$

where we've used the fact that the traces can be performed independently. The second equality introduces the *reduced density operator* ρ_A of subsystem A, defined by

$$\rho_A = \mathrm{tr}_{\mathcal{H}_B}(\rho_{AB})\,, \tag{10.33}$$

taking the traces only over subsystem B.

The reduced density operator enables us to obtain expectation values of subsystem A's observables without bothering about the states of B. For example, suppose an atom is situated in a low-intensity radiation field. Every so often, a photon comes along. This photon may scatter off the atom, or be absorbed by the atom into an excited state which subsequently decays re-emitting the photon, or may even cause the atom to be temporarily ionized. If we wish to keep track of the whole system, then as more and more photons interact with the atom, we'd need to use a larger and larger Hilbert space encompassing further and further tensor products of the Hilbert spaces of individual photons. This is inconvenient, to say the least. However, if we're only really interested in the state of the atom, it's enough to keep track of the atoms reduced density operator, which refers solely to the Hilbert space of the atom.

10.4 Decoherence

We now show a very important result. Suppose the Universe itself is in a pure quantum state, so that $\rho_{AB} = |\Psi\rangle\langle\Psi|$ for some state $|\Psi\rangle$ written as

$$|\Psi\rangle = \sum_{\alpha,\beta} c_{\alpha,\beta} |\alpha\rangle |\beta\rangle \tag{10.34}$$

in terms of the orthonormal bases $\{|\alpha\rangle\}$ for A and $\{|\beta\rangle\}$ for B. Then the reduced density matrix ρ_A is

$$\begin{aligned}\rho_A = \mathrm{tr}_{\mathcal{H}_B}(\rho_{AB}) &= \mathrm{tr}_{\mathcal{H}_B}\left(\sum_{\alpha,\alpha',\beta,\beta'} c_{\alpha,\beta} c^*_{\alpha',\beta'} |\alpha\rangle |\beta\rangle \langle\alpha'| \langle\beta'|\right) \\ &= \sum_{a,a'} C_{a,a'} |a\rangle \langle a'| \,, \end{aligned} \tag{10.35}$$

where now

$$C_{a,a'} = \sum_{\beta} c_{a,\beta} c^*_{a',\beta} \,. \tag{10.36}$$

If the original state $|\Psi\rangle$ was *simple*, so that the $c_{a,\beta}$ are non-zero only for a single value of β, for which $c_{a,\beta} = c_a$, then $C_{a,a'} = c_a c^*_{a'}$ with no sum. Then (10.35) is a pure density operator for the state $|\phi\rangle = \sum_a c_a |a\rangle$. However, if $|\Psi\rangle$ is a more general, entangled state then the sum in (10.36) means that ρ_A is the density operator for a mixed state.

For example, suppose our Universe consists of just two spin-$\frac{1}{2}$ particles, prepared in the pure but entangled state

$$|\psi\rangle = \frac{1}{\sqrt{2}} \left(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle\right) \,, \tag{10.37}$$

where $|\uparrow\downarrow\rangle = |\uparrow\rangle |\downarrow\rangle$ *etc.* The associated density operator is

$$\begin{aligned}\rho_{AB} &= |\psi\rangle\langle\psi| \\ &= \frac{1}{2}\left(|\uparrow\downarrow\rangle\langle\uparrow\downarrow| + |\downarrow\uparrow\rangle\langle\downarrow\uparrow| - |\uparrow\downarrow\rangle\langle\downarrow\uparrow| - |\downarrow\uparrow\rangle\langle\uparrow\downarrow|\right) \,. \end{aligned} \tag{10.38}$$

Tracing over the second spin gives the reduced density operator

$$\rho_A = \mathrm{tr}_{\mathcal{H}_B}(\rho_{AB}) = \frac{1}{2}\left(|\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|\right) = \frac{1}{2} 1_{\mathcal{H}_A} \,, \tag{10.39}$$

which is mixed, and the state of maximum entropy.

Although we won't prove here, the von Neumann entropy obeys

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B) \,, \tag{10.40}$$

where ρ_A and ρ_B are the reduced density matrices for the two subsystems. The equality is saturated if and only if the two subsystems are uncorrelated (unentangled) so that $\rho_{AB} = \rho_A \otimes \rho_B$. This property is known as *subadditivity* and it tells us that the entropy of a whole is no greater than the sum of entropies of its parts. It follows from subadditivity that if $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$ then

$$S(\rho_{ABC}) \leq S(\rho_A) + S(\rho_{BC}) \leq S(\rho_A) + S(\rho_B) + S(\rho_C)\,. \tag{10.41}$$

In fact, something stronger is true: we have

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC})\,, \tag{10.42}$$

which is known as *strong subadditivity* and was proved in 1973 by Elliott Lieb and Mary Beth Ruskai. To interpret it, we consider AB and BC are each subsystems of ABC, with $AB \cap BC = B$. Strong subadditivity states that the entropy of the whole is no greater than the entropies of the overlapping subsystems AB and BC, even when the entropy of the overlap B is removed.

10.4.1 Time Evolution of Density Operators and Reduced Density Operators

In the Schrödinger's picture, states evolve in time according to

$$|\psi(t)\rangle = U(t)\,|\psi_0\rangle\ , \tag{10.43}$$

where $U(t) = \mathrm{e}^{-\mathrm{i}Ht/\hbar}$ in the case that the Hamiltonian itself is time-independent. This implies that the density operator evolves, like any other operator, as

$$\rho(t) = U(t)\rho(0)U^{-1}(t)\,, \tag{10.44}$$

or

$$\mathrm{i}\hbar\frac{\mathrm{d}\rho}{\mathrm{d}t} = U(t)(H\rho(0) - \rho(0)H)U^{-1}(t) = [H, \rho(t)] \tag{10.45}$$

infinitesimally. This is the quantum analogue of Liouville's equation $\frac{\mathrm{d}\rho}{\mathrm{d}t} = \{H, \rho\}$ in Classical Dynamics, which governs the time evolution of a probability density on phase space. In particular, if the density operator can be written purely in terms of the Hamiltonian then it is time independent. The Gibbs ensemble we obtained above is a good example.

To obtain the time evolution of an arbitrary expectation value (of a quantity that has no explicit time dependence in the Schrödinger picture) we use (10.45) to find

$$\mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{tr}_{\mathcal{H}}(\rho Q) = \mathrm{tr}_{\mathcal{H}}([H, \rho]Q) = \mathrm{tr}_{\mathcal{H}}(\rho[Q, H])\,, \tag{10.46}$$

where the last equality uses the cyclicity of the trace. We know from before that the rate of change of the expectation value of Q in any pure quantum state is given by the expectation value of $[Q, H]/\mathrm{i}\hbar$. Equation (10.46) states that — even when our knowledge of the quantum state is imprecise — the expected rate of change of Q is the appropriately weighted average of the rates of change of Q for each of the possible states of the system.

Let's now consider how the reduced density operator evolves. Suppose that at $t = 0$ both A and B are in pure quantum states $|\phi\rangle$ and $|\chi\rangle$, respectively. Initially then,

$$\rho_{AB}(0) = |\Psi_0\rangle\,\langle\Psi_0| \tag{10.47}$$

where

$$|\Psi_0\rangle = |\phi\rangle\,|\chi\rangle\ , \tag{10.48}$$

so that the two systems are unentangled. The whole system will evolve unitarily in time via the operator $U_{AB}(t)$ built from the Hamiltonian of the full system. This means that the reduced density operator for system A evolves as

$$\rho_A(t) = \mathrm{tr}_{\mathcal{H}_B}(U_{AB}(t)\,|\Psi_0\rangle\,\langle\Psi_0|\,U_{AB}^{-1}(t)) = \sum_\beta \langle\beta|U_{AB}(t)|\Psi_0\rangle\,\big\langle\Psi_0\big|U_{AB}^{-1}(t)\big|\beta\big\rangle\ , \tag{10.49}$$

where we're using the orthonormal basis $\{|\beta\rangle\}$ of $\mathcal{H}_B$ to perform the trace.

We are motivated to define operators $M_\beta(t) : \mathcal{H}_A \to \mathcal{H}_A$ by

$$M_\beta(t) = \langle\beta|U_{AB}(t)|\chi\rangle = \mathrm{tr}_{\mathcal{H}_B}(U_{AB}(t)\,|\chi\rangle\,\langle\beta|)\,. \tag{10.50}$$

This is indeed a operator $\mathcal{H}_A \to \mathcal{H}_A$ since we've used the time-evolution operator $U_{AB}(t)$ for $\mathcal{H}_A \otimes \mathcal{H}_B$ and contracted only the $\mathcal{H}_B$. Since $U_{AB}(t)$ is unitary, the $M_\beta(t)$'s obey a completeness relation

$$\sum_\beta M_\beta^\dagger(t) M_\beta(t) = \sum_\beta \langle\chi|U_{AB}^{-1}(t)|\beta\rangle\,\langle\beta|U_{AB}(t)|\chi\rangle = 1_{\mathcal{H}_A} \tag{10.51}$$

provided that $|\chi\rangle$ was properly normalised. Putting all these together,

$$\rho_A(t) = \sum_\beta M_\beta(t)\rho_A(0)M_\beta^\dagger(t) \tag{10.52}$$

is the evolution of the reduced density matrix.

In the exceptional case that the full Hamiltonian does not couple A and B, so that $H = H_A \otimes 1_{\mathcal{H}_B} + 1_{\mathcal{H}_A} \otimes H_B$ and $U_{AB}(t) = U_A(t) \otimes U_B(t)$, we have

$$M_\beta(t) = \langle\beta|U_B(t)|\chi\rangle\, U_A(t)\,. \tag{10.53}$$

The completeness relation then shows that

$$\rho_A = U_A(t)\rho_A(0)U_A^{-1}(t)\,. \tag{10.54}$$

Thus, if A starts in a pure state and it does not interact with the environment then it will remain in a pure state. However, in every realistic case, subsystems are coupled to each other, however weak — there is some term H_{AB} in the Hamiltonian that is not diagonal with respect to the splitting $H_A \otimes H_B$. In the presence of an interaction term H_{AB}, the time evolution operator is not generically a product of the time evolution operators of the two subsystems, and states of A and B will typically become entangled, leading to $\rho_A(t)$ describing a mixed state at some later time t.

In general, interactions between an experimental system and the wider environment mean that the state of the whole Universe rapidly becomes entangled. Since we don't keep track of all the details of the environment, sooner or later we're obliged to describe our experimental system by its reduced density operator, which will be impure. The tendency for subsystems to evolve from pure quantum states to impure states through interactions with the environment is known as *quantum decoherence*. Trying to isolate a system from the environment so as to prevent it from becoming impure is one of the main challenges to be overcome in building a practical quantum computer.

10.4.2 Decoherence and Measurement

We're at last ready to explore what all this has to do with measurements in quantum mechanics.

Let's suppose our system A consists of a single qubit, either $|\uparrow\rangle$ or $|\downarrow\rangle$. To keep things simple, we'll imagine the environment (or measuring apparatus) has only three possible states, $|0\rangle$, $|1\rangle$ and $|2\rangle$. An ideal measurement will change the state of the measuring apparatus without affecting the system A under study. Let's suppose the measurement process is described via evolution by a unitary operator U, representing the usual evolution of the system and apparatus by a coupled Hamiltonian. We suppose our apparatus is designed in such a way that U is defined by[61]

$$\begin{aligned} U\,|\uparrow\rangle \otimes |0\rangle &= |\uparrow\rangle\,(\sqrt{1-p}\,|0\rangle + \sqrt{p}\,|1\rangle) \\ U\,|\downarrow\rangle \otimes |0\rangle &= |\downarrow\rangle\,(\sqrt{1-p}\,|0\rangle + \sqrt{p}\,|2\rangle) \end{aligned} \tag{10.55}$$

[61] By assigning appropriate values to U acting on $|1\rangle$ and $|2\rangle$, this U can indeed be completed to a unitary operator. This is left as an exercise.

In other words, the apparatus starts in the 'quiescent' state $|0\rangle$. When we bring it into contact with our system A, the apparatus changes its state with probability p, becoming $|1\rangle$ if A is in $|\uparrow\rangle$, or $|2\rangle$ if A is in $|\downarrow\rangle$. The apparatus is not perfectly efficient, so stays in the quiescent state with probability $1-p$. Now let's suppose the system A is initially described by some density matrix

$$\rho_A = \begin{pmatrix} \rho_{\uparrow\uparrow} & \rho_{\uparrow\downarrow} \\ \rho_{\downarrow\uparrow} & \rho_{\downarrow\downarrow} \end{pmatrix} . \tag{10.56}$$

For this evolution, we have

$$\begin{aligned} M_0 &= \langle 0|U|0\rangle = \sqrt{1-p}\, 1_{\mathcal{H}_A} \\ M_1 &= \langle 1|U|0\rangle = \sqrt{p}\, |\uparrow\rangle \langle\uparrow| \\ M_2 &= \langle 2|U|0\rangle = \sqrt{p}\, |\downarrow\rangle \langle\downarrow| \end{aligned} \tag{10.57}$$

and these indeed obey $\sum_\beta M_\beta^\dagger M_\beta = 1_{\mathcal{H}_A}$. Contact with our measuring apparatus thus causes ρ_A to evolve as

$$\rho_A = \begin{pmatrix} \rho_{\uparrow\uparrow} & \rho_{\uparrow\downarrow} \\ \rho_{\downarrow\uparrow} & \rho_{\downarrow\downarrow} \end{pmatrix} \longmapsto \begin{pmatrix} \rho_{\uparrow\uparrow} & (1-p)\rho_{\uparrow\downarrow} \\ (1-p)\rho_{\downarrow\uparrow} & \rho_{\downarrow\downarrow} \end{pmatrix} , \tag{10.58}$$

suppressing the off-diagonal components. These off-diagonal components encode possible superpositions between $|\uparrow\rangle$ and $|\downarrow\rangle$, so as our system becomes entangled with the measuring apparatus, we're less likely to find it in such a superposition.

Let's go further and look at successive evolution. The probability the apparatus changed away from the quiescent state during one measurement period was p, so if we suppose this measurement took a short time δt, then we can define a rate $\Gamma = p/\delta t$. After a total time $t = N\delta t$, the off diagonal terms will thus be suppressed by

$$(1-p)^N = \left(1 - \Gamma\frac{t}{N}\right)^N \approx \mathrm{e}^{-\Gamma t} \tag{10.59}$$

for large N. In particular, if we prepare A to be in the superposition

$$|\psi\rangle = a\,|\uparrow\rangle + b\,|\downarrow\rangle \ , \tag{10.60}$$

where $|a|^2 + |b|^2 = 1$, then eventually, the density matrix will become

$$\lim_{t\to\infty} \rho_A(t) = \lim_{t\to\infty} \begin{pmatrix} |a|^2 & ab^*\mathrm{e}^{-\Gamma t} \\ a^*b\mathrm{e}^{-\Gamma t} & |b|^2 \end{pmatrix} = \begin{pmatrix} |a|^2 & 0 \\ 0 & |b|^2 \end{pmatrix} . \tag{10.61}$$

This is sometimes called *phase damping*, because the late time density matrix only has real entries.

Now we come to the punchline. What exactly was it that made our measurement cause A to evolve into either $|\uparrow\rangle$ or $|\downarrow\rangle$, but not a superposition? Clearly, this must have had something to do with our choice of U in (10.55). To get an idea, let's imagine the two state system A actually corresponds to a dust particle which can sit either at x_0 or x_1. The measuring apparatus may be a photon which, with probability p, can scatter into a different direction, depending on where the dust is located. The fact that U is defined with respect to the preferred basis $|\uparrow\rangle = |x_0\rangle$ and $|\downarrow\rangle = |x_1\rangle$ then corresponds to the fact that the interactions are *local*: The interactions we can describe between the dust and photons will be built out of operators such as $\mathbf{X}_{\text{dust}}$, so decoherence will take place in the basis $\{|x_0\rangle, |x_1\rangle\}$ where the dust particle has a definite location, rather than the $(|x_0\rangle \pm |x_1\rangle)/\sqrt{2}$ basis.

Locality of interactions is one of the key features of all physical laws, and has deep rooted origins in quantum field theory. Combined with decoherence, many physicists believe[62] that this explains why we see cats either in the state alive or the state dead, but never $(|\text{alive}\rangle + |\text{dead}\rangle)/\sqrt{2}$.

[62]The matter is still not fully resolved. The leading proponents of the 'measurement=decoherence' paradigm are H.D. Zeh and W.Zurek, see for example Zurek, W., *Quantum Darwinism*, Nature Physics 5(3), 181 (2009) for a review. Prominent opponents include R. Kastner, see e.g. Kastner, R., Stud. Hist. Phil. Science B48 56 (2014).

10.5 Quantum Mechanics or Hidden Variables

Einstein was never happy with the probabilistic nature of quantum mechanics. He, Podolsky and Rosen devised a thought experiment that they hoped would show quantum mechanics was incomplete as a theory of Nature.

In the EPR thought experiment, an electron and a positron[63] are produced in a state with net spin 0, perhaps by the decay of some nucleus from a spin-0 excited state to a lower spin-0 state. The electron and positron travel in opposite directions, each carrying the same amount of momentum. At some distance from the decaying nucleus Alice detects the electron and measures the component of its spin in a direction $\mathbf{a}$ of her choice. Since electrons have spin-$\frac{1}{2}$, Alice inevitably discovers either $+\hbar/2$ or $-\hbar/2$. Meanwhile Bob, who is sitting at a similar distance on the opposite side of the nucleus, detects the positron and measures its spin in some direction $\mathbf{b}$ of his choice.

We're free to choose the z-axis to be aligned with Alice's direction $\mathbf{a}$. Since the electron positron system has combined spin zero, it must be in the state

$$|\mathrm{EPR}\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle\,|\downarrow\rangle - |\downarrow\rangle\,|\uparrow\rangle\right) \tag{10.62}$$

that entangles the separate spins of the electron and positron. We'll call this the EPR state. According to the Copenhagen interpretation, when Alice measures $+\hbar/2$ for the electron spin, the system collapses into the state

$$|\psi'\rangle = |\uparrow\rangle\,|\downarrow\rangle\ . \tag{10.63}$$

Thus, whilst before Alice's measurement the amplitude for the positron to have spin $+\hbar/2$ along $\mathbf{a}$ was $1/2$, after she has measured the electron spin, there is no chance that the positron also has spin up along the same axis.

The state of the positron corresponding to definitely having spin $+\hbar/2$ along the $\mathbf{b}$-axis is

$$|\uparrow_{\mathbf{b}}\rangle = \cos\left(\frac{\theta}{2}\right)\mathrm{e}^{-\mathrm{i}\phi/2}\,|\uparrow\rangle + \sin\left(\frac{\theta}{2}\right)\mathrm{e}^{\mathrm{i}\phi/2}\,|\downarrow\rangle \tag{10.64}$$

as we found in the second example sheet, where $\theta = \cos^{-1}(\mathbf{a}\cdot\mathbf{b})$ and ϕ is the azimuthal angle around $\hat{\mathbf{z}} = \hat{\mathbf{a}}$. Given that after Alice's measurement the positron is certainly in state $|\downarrow\rangle$, it follows that the probability Bob measures spin up along $\mathbf{b}$ is

$$|\langle\uparrow_{\mathbf{b}}|\downarrow\rangle| = \sin^2\left(\frac{\theta}{2}\right)\ . \tag{10.65}$$

In particular, there is only a small probability he will find spin-up along a direction closely aligned with Alice's choice $\mathbf{a}$.

We've supposed that Alice measures first, but if the electron and positron are far apart when the measurements are made, a light signal sent to Bob by Alice when she makes her measurement would not have arrived at Bob by the time he makes his measurement, and vice versa. In these circumstances, relativity tells us that the order in which the measurements appear to be made depend on the velocity of the observer who is judging the matter. Consequently, for consistency the predictions of quantum mechanics must be independent of who is supposed to make the first measurement and thus collapse the state. This condition is satisfied by the above discussion, since the final probability depends only on $\mathbf{a}\cdot\mathbf{b}$ and is thus symmetric between Alice and Bob.

What bothered EPR is that after Alice's measurement there is a direction $\mathbf{a}$ along which Bob can never find $+\hbar/2$ for the positron's spin, and this direction depends on what exactly Alice chooses to

[63] Here we'll describe a slightly sharper version of EPR's original thought experiment, due to Bohm. The positron is the antiparticle of the electron, predicted in the relativistic theory by Dirac's equation. It has the same mass and spin as an electron, but opposite sign electric charge.

measure. This fact seems to imply that the positron somehow 'knows' what Alice measured for the electron, and the collapse of the entangled wavefunction

$$\frac{1}{\sqrt{2}}\left(|\uparrow\rangle\,|\downarrow\rangle - |\downarrow\rangle\,|\uparrow\rangle\right) \quad\longrightarrow\quad |\uparrow\rangle\,|\downarrow\rangle \tag{10.66}$$

apparently confirms this suspicion. Since relativity forbids news of Alice's work on the electron from influencing the positron at the time of Bob's measurement, EPR argued that the required information must have travelled out in the form of a *hidden variable* which was correlated at the time of the nuclear decay with a matching hidden variable in the electron. These hidden variables would then explain the probabilistic nature of quantum mechanics — QM would contain no uncertainties once replaced by a 'better' theory taking into account these hidden variables.

10.5.1 Bell's Inequality

Remarkably, Bell was able to show that the predictions of any theory of hidden variables are in conflict with the predictions of quantum mechanics.

Suppose we assume that the result of measuring the electrons spin in the $\mathbf{a}$-direction is completely determined by the values taken by hidden variables in addition to $\mathbf{a}$. We suppose there are n such hidden variables, so that the result of measuring the electrons spin is a function

$$s_e : \mathbb{R}^3 \times \mathbb{R}^n \longrightarrow \left\{-\frac{\hbar}{2}, +\frac{\hbar}{2}\right\} \tag{10.67}$$

that returns either $+\hbar/2$ or $-\hbar/2$, depending only on the direction $\mathbf{a} \in \mathbb{R}^3$ along which we measure the spin and the values $\mathbf{v} \in \mathbb{R}^n$ of the hidden variables carried by the electron. In other words, if Alice knew the value of the hidden variable $\mathbf{v} \in \mathbb{R}^n$, we could predict with certainty the result of measuring the component of the electrons spin along any direction $\mathbf{a}$. Alice is only uncertain of the outcome because she does not know the values of the hidden variables. Similarly, the result of measuring the positron's spin along $\mathbf{b}$ is some function $s_p(\mathbf{b}, \mathbf{v})$. We have

$$s_e(\mathbf{a}, \mathbf{v}) + s_p(\mathbf{a}, \mathbf{v}) = 0 \tag{10.68}$$

by the conservation of momentum.

Let's suppose that $\mathbf{v}$ has a probability distribution $p(\mathbf{v})$, such that the probability $\mathrm{d}P$ that $\mathbf{v}$ lies in the infinitesimal volume $\mathrm{d}^n\mathbf{v}$ is

$$\mathrm{d}P = p(\mathbf{v})\mathrm{d}^n\mathbf{v}\,. \tag{10.69}$$

We are interested in the expectation value

$$\begin{aligned}\langle s_e(\mathbf{a}, \mathbf{v}) s_p(\mathbf{b}, \mathbf{v})\rangle &= \int \mathrm{d}^n\mathbf{v}\, p(\mathbf{v}) s_e(\mathbf{a}, \mathbf{v}) s_p(\mathbf{b}, \mathbf{v}) \\ &= -\int \mathrm{d}^n\mathbf{v}\, p(\mathbf{v}) s_e(\mathbf{a}, \mathbf{v}) s_e(\mathbf{b}, \mathbf{v})\,.\end{aligned} \tag{10.70}$$

Now suppose Bob sometimes measures the spin of the positron parallel to $\mathbf{b}'$ rather than $\mathbf{b}$. The fact that $s_p(\mathbf{b}, \mathbf{v})^2 = \frac{\hbar^2}{4}$ allows us to write

$$\begin{aligned}\langle s_e(\mathbf{a}, \mathbf{v}) s_p(\mathbf{b}, \mathbf{v})\rangle - \langle s_e(\mathbf{a}, \mathbf{v}) s_p(\mathbf{b}', \mathbf{v})\rangle &= -\int \mathrm{d}^n\mathbf{v}\, p(\mathbf{v}) s_e(\mathbf{a}, \mathbf{v}) \left[s_p(\mathbf{b}, \mathbf{v}) - s_p(\mathbf{b}', \mathbf{v})\right] \\ &= -\int \mathrm{d}^n\mathbf{v}\, p(\mathbf{v}) s_e(\mathbf{a}, \mathbf{v}) s_e(\mathbf{b}, \mathbf{v}) \left[1 - \frac{4}{\hbar^2} s_e(\mathbf{b}, \mathbf{v}) s_e(\mathbf{b}', \mathbf{v})\right]\,.\end{aligned} \tag{10.71}$$

The expression $[1-4s_e(\mathbf{b},\mathbf{v})s_e(\mathbf{b}',\mathbf{v})/\hbar^2]$ is non-negative, while the product $s_e(\mathbf{a},\mathbf{v})s_e(\mathbf{b},\mathbf{v})$ fluctuates between $\pm\hbar^2/4$. Hence we obtain the bound

$$\begin{aligned}\big|\langle s_e(\mathbf{a},\mathbf{v})s_p(\mathbf{b},\mathbf{v})\rangle - \big\langle s_e(\mathbf{a},\mathbf{v})s_p(\mathbf{b}',\mathbf{v})\big\rangle\big| &\leq \frac{\hbar^2}{4}\int \mathrm{d}^n\mathbf{v}\, p(\mathbf{v})\left[1-\frac{4}{\hbar^2}s_e(\mathbf{b},\mathbf{v})s_e(\mathbf{b}',\mathbf{v})\right] \\ &= \frac{\hbar^2}{4} - \big\langle s_e(\mathbf{b},\mathbf{v})s_p(\mathbf{b}',\mathbf{v})\big\rangle\,. \end{aligned} \tag{10.72}$$

This is *Bell's inequality.* It must hold for any three unit vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{b}'$ if the probabilistic nature of QM really comes from some underlying hidden variables.

We now show that quantum mechanics itself violates Bell's inequality. To do this, we must treat the spins as operators and compute their expectation values in some state, and we'll choose the EPR state (10.62). Because $|\mathrm{EPR}\rangle$ has total spin zero, it obeys

$$(\mathbf{S}_e\otimes 1_p + 1_e\otimes\mathbf{S}_p)\,|\mathrm{EPR}\rangle = 0 \tag{10.73}$$

so that we always find the spin of the electron and positron to be anti-aligned whenever we measure them along any one given axis, no matter in which direction this is. In particular, this allows us to write

$$\begin{aligned}(\mathbf{a}\cdot\mathbf{S}_e\otimes 1_p)(1_e\otimes\mathbf{b}\cdot\mathbf{S}_p) &= -(\mathbf{a}\cdot\mathbf{S}_e\otimes 1_p)(\mathbf{b}\cdot\mathbf{S}_e\otimes 1_p) \\ &= -(\mathbf{a}\cdot\mathbf{S}_e\mathbf{b}\cdot\mathbf{S}_e)\otimes 1_p \end{aligned} \tag{10.74}$$

when acting on $|\mathrm{EPR}\rangle$. For any single spin-$\frac{1}{2}$ particle, the spin operator obeys

$$\mathbf{a}\cdot\mathbf{S}_e\mathbf{b}\cdot\mathbf{S}_e = \frac{\hbar^2}{4}\mathbf{a}\cdot\mathbf{b} + \frac{\mathrm{i}\hbar}{2}(\mathbf{a}\times\mathbf{b})\cdot\mathbf{S}\,. \tag{10.75}$$

Therefore

$$\langle\mathbf{a}\cdot\mathbf{S}_e\mathbf{b}\cdot\mathbf{S}_p\rangle_{\mathrm{EPR}} = -\frac{\hbar^2}{4}\mathbf{a}\cdot\mathbf{b} - \frac{\mathrm{i}\hbar}{2}(\mathbf{a}\times\mathbf{b})\cdot\langle\mathbf{S}_e\rangle_{\mathrm{EPR}}\,. \tag{10.76}$$

Finally, we note that the expectation value of the electrons spin

$$\langle\mathbf{S}_e\otimes 1_p\rangle\,\mathrm{EPR} = 0 \tag{10.77}$$

along any axis. This is clear for the z-axis, but since $|\mathrm{EPR}\rangle$ has no preferred direction it must be true of the other directions also. Thus we find that the EPR state obeys

$$\langle\mathbf{a}\cdot\mathbf{S}_e\otimes\mathbf{b}\cdot\mathbf{S}_p\rangle_{\mathrm{EPR}} = -\frac{\hbar^2}{4}\mathbf{a}\cdot\mathbf{b} \tag{10.78}$$

for any two directions $\mathbf{a}$ and $\mathbf{b}$. Using this correlation in either side of Bell's inequality we find

$$\mathrm{LHS} = \frac{\hbar^2}{4}\big|\mathbf{a}\cdot(\mathbf{b}-\mathbf{b}')\big| \quad \text{whereas} \quad \mathrm{RHS} = \frac{\hbar^2}{4}(1-\mathbf{b}\cdot\mathbf{b}')\,. \tag{10.79}$$

In particular, suppose $\mathbf{a}$ and $\mathbf{b}$ are unit vectors, with $\mathbf{a}\cdot\mathbf{b}=0$ and $\mathbf{b}'=\mathbf{b}\cos\alpha+\mathbf{a}\sin\alpha$. Then we find

$$\mathrm{LHS} = \frac{\hbar^2}{4}|\sin\alpha| \quad \text{whereas} \quad \mathrm{RHS} = \frac{\hbar^2}{4}(1-\cos\alpha)\,, \tag{10.80}$$

and it is easy to see that Bell's inequality is violated for $0<\alpha<\pi/2$. The predictions of quantum mechanics are thus inconsistent with the existence of hidden variables.

10.5.2 The CHSH Inequality

There's a slightly simpler context in which we can see the essentials of the conflict between quantum probability and hidden variables theories, discovered by Clauser, Horne, Shimony and Holt.

Suppose Alice and Bob are each sent a two-state system as in the EPR experiment. Alice chooses to measure one of two possible observables, either A_1 or A_2. Similarly, Bob can choose to measure either B_1 or B_2. To keep things simple, let's assume that there are only two possible outcomes, $+1$ and -1, for the result when measuring any of the four quantities A_i or B_i. We'll require that

$$[A_i, B_j] = 0 \tag{10.81}$$

for $i = 1, 2$ so that the measurement Alice makes does not interfere with the one made by Bob. However, we do not require that either $[A_1, A_2] = 0$ or $[B_1, B_2] = 0$.

Now consider the observable

$$C = (A_1 + A_2)B_1 + (A_1 - A_2)B_2 \,. \tag{10.82}$$

In a hidden variable theory, the outcomes of measuring the A_i or B_j would be entirely determined by the value $\mathbf{v}$ of some hidden variables carried by the state, so we'd have functions

$$a_i, b_i : \mathbf{R}^n \longrightarrow \{+1, -1\} \,. \tag{10.83}$$

The average value of C in a hidden variable theory is

$$\langle C \rangle = \int \mathrm{d}^n\mathbf{v}\, p(\mathbf{v}) \left([a_1(\mathbf{v}) + a_2(\mathbf{v})]\mathbf{b}_1(\mathbf{v}) + [a_1(\mathbf{v}) - a_2(\mathbf{v})]\mathbf{b}_2(\mathbf{v}) \right) \,, \tag{10.84}$$

where again $p(\mathbf{v})$ is the probability density for the hidden variables. Since each $a_i(\mathbf{v})$ can take only the values ± 1, either $\mathbf{v}$ is such that the outcomes of A_1 and A_2 are different, in which case

$$a_1(\mathbf{v}) + a_2(\mathbf{v}) = 0 \quad \text{while} \quad a_1(\mathbf{v}) - a_2(\mathbf{v}) = \pm 2 \,, \tag{10.85}$$

or else the value of $\mathbf{v}$ ensures that the outcomes of A_1 and A_2 are the same, so that

$$a_1(\mathbf{v}) + a_2(\mathbf{v}) = \pm 2 \quad \text{while} \quad a_1(\mathbf{v}) - a_2(\mathbf{v}) = 0 \,. \tag{10.86}$$

Thus, whatever the value of $\mathbf{v}$, only one of the two terms in the integral (10.84) can be non-zero. Multiplying the non-zero term by $b_i(\mathbf{v})$ at most changes its sign, so we always have

$$[a_1(\mathbf{v}) + a_2(\mathbf{v})]\mathbf{b}_1(\mathbf{v}) + [a_1(\mathbf{v}) - a_2(\mathbf{v})]\mathbf{b}_2(\mathbf{v}) = \pm 2 \tag{10.87}$$

fluctuating as $\mathbf{v}$ moves around. Consequently, we can bound the average by

$$-2 \leq \langle C \rangle \leq 2 \,. \tag{10.88}$$

This is known as the *CHSH inequality* and it's obeyed in any hidden variables theory.

Now let's look at the same observable in quantum theory. Since the eigenvalues of A_i and B_j are just ± 1, we have $A_i^2 = 1$ and $B_j^2 = 1$. Consequently one finds

$$\begin{aligned} C^2 &= (A_1 + B_1)^2 B_1^2 + (A_1 - A_2)^2 B_2^2 + (A_1 + A_2)(A_1 - A_2)B_1B_2 + (A_1 - A_2)(A_1 + A_2)B_2B_1 \\ &= 4 - A_1A_2B_1B_2 + A_2A_1B_1B_2 - A_1A_2B_2B_1 + A_2A_1B_2B_1 \\ &= 4 - [A_1, A_2][B_1, B_2] \,, \end{aligned} \tag{10.89}$$

where the first equality uses our assumption $[A_i, B_j] = 0$. We have that

$$|\langle [A_1, A_2] \rangle| \leq |\langle A_1A_2 \rangle| + |\langle A_2, A_1 \rangle| \leq 2 \,, \tag{10.90}$$

with the final bound again coming since the eigenvalues of A_i are just ± 1. Thus we have $\langle C^2 \rangle \leq 8$ in quantum theory. Finally, since $\langle C \rangle^2 \leq \langle C^2 \rangle$ for any Hermitian operator, we obtain the *Cirel'son bound*

$$-2\sqrt{2} \leq \langle C \rangle \leq 2\sqrt{2} \tag{10.91}$$

in quantum theory. This shows that quantum theory permits a wider range of values for C than allowed by the CHSH bound (10.88) for hidden variable theories.

Again, it's straightforward to show that the EPR state saturate the Cirel'son inequality. Recall from that

$$\left\langle \hat{\boldsymbol{\alpha}} \cdot \boldsymbol{\sigma}_e \otimes \hat{\boldsymbol{\beta}} \cdot \boldsymbol{\sigma}_p \right\rangle_{\text{EPR}} = -\hat{\mathbf{a}} \cdot \hat{\mathbf{b}} = -\cos\theta \tag{10.92}$$

for any two unit vectors. (We're using the Pauli matrices rather than spins $\mathbf{S} = \hbar\boldsymbol{\sigma}/2$ to ensure the eigenvalues are ± 1 as in the CHSH and Cirel'son bounds.) To apply this to the Cirel'son case, let A_i be the Pauli matrices for the electron and B_j those for the positron and choose A_2, B_1, A_1, B_2 to all lie in (say) the xz-plane at angles 0, $\pi/4$, $\pi/2$ and $3\pi/4$ to the z-axis, respectively. Then

$$\langle A_1 B_1 \rangle = \langle A_1 B_2 \rangle = \langle A_2 B_1 \rangle = -\frac{1}{\sqrt{2}} \tag{10.93}$$

for the EPR state, while

$$\langle A_2 B_2 \rangle = +\frac{1}{\sqrt{2}}\,. \tag{10.94}$$

Consequently, we have

$$\langle C \rangle_{\text{EPR}} = -2\sqrt{2}\,, \tag{10.95}$$

saturating the Cirel'son bound.

Impressively, this inequality has actually been tested experimentally by Aspect *et al.*, following the suggestion of Clauser *et al.*[64] In the experiment, two photons are emitted from successive decays of excited states of calcium. The first comes from the decay of a parity-even state with $j = 0$ to a short-lived parity-odd state with $j = 1$, while the second photon comes from the decay of this short-lived state to a further parity-even state of $j = 0$ (of lower energy than the initial state). The photons are directed into a combination of polarizers and photomultipliers which read out ± 1 according to whether the photons are found to be linearly polarized along some directions $\mathbf{a}$ and $\mathbf{b}$. The experiment found

$$\left| \langle C \rangle_{\text{expt}} \right| = 2.697 \pm 0.0515\,. \tag{10.96}$$

This is slightly less than the idealised result $\langle C \rangle_{\text{EPR}} = 2\sqrt{2} \approx 2.828$, with most of the disagreement coming because the polarizers used in the experiment were not perfectly efficient. When the efficiency of the polarizers is taken into account, Aspect's result is in good agreement with what was predicted by quantum mechanics, and in clear violation of the CHSH bound for hidden variable theories.

In the final problem set, you'll explore an even more striking conflict between the predictions of Quantum Mechanics and classical hidden variables by considering entanglement between three qubits rather than two.

[64] See Freedman, S. & Clauser, J., *Experimental Test of Local Hidden Variable Theories*, Phys. Rev. Lett. 28, 938 (1972) and Aspect, A. & Roger, G. *Experimental Realization of the Einstein–Podolsky–Rosen–Bohm Gedankenexperiment: A New Violation of Bells Inequalities*, Phys. Rev. Lett. 49, 91 (1982). An earlier version of the experiment was performed by Kocher, C. and Commins, E. Phys. Rev. Lett. 18, 575 (1979).