WAVE MECHANICS

B. Zwiebach
September 13, 2013

Contents

1 The Schrödinger equation

In classical mechanics the motion of a particle is usually described using the time-dependent position $\vec{x}(t)$ as the dynamical variable. In wave mechanics the dynamical variable is a wavefunction. This wavefunction depends on position and on time and it is a complex number – it belongs to the complex numbers $\mathbb{C}$ (we denote the real numbers by $\mathbb{R}$). When all three dimensions of space are relevant we write the wavefunction as

$$\Psi(\vec{x},t) \in \mathbb{C}\,. \tag{1.1}$$

When only one spatial dimension is relevant we write it as $\Psi(x,t) \in \mathbb{C}$. The wavefunction satisfies the Schrödinger equation. For one-dimensional space we write

$$\boxed{i\hbar\frac{\partial\Psi}{\partial t}(x,t) \;=\; \Big(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x,t)\Big)\,\Psi(x,t)\,.} \tag{1.2}$$

This is the equation for a (non-relativistic) particle of mass m moving along the x axis while acted by the potential $V(x,t) \in \mathbb{R}$. It is clear from this equation that the wavefunction must be complex: if it were real, the right-hand side of (1.2) would be real while the left-hand side would be imaginary, due to the explicit factor of i.

Let us make two important remarks:

1. The Schrödinger equation is a *first order* differential equation in time. This means that if we prescribe the wavefunction $\Psi(x, t_0)$ for all of space at an arbitrary initial time t_0, the wavefunction is determined for all times.

2. The Schrödinger equation is a *linear* equation for Ψ: if Ψ_1 and Ψ_2 are solutions so is $a_1\Psi_1 + a_2\Psi_2$ with a_1 and a_2 arbitrary *complex numbers*.

Given a complex number $z = a + ib$, $a, b \in \mathbb{R}$, its complex conjugate is $z^* = a - ib$. Let $|z|$ denote the norm or length of the complex number z. The norm is a positive number (thus real!) and it is given by $|z| = \sqrt{a^2 + b^2}$. If the norm of a complex number is zero, the complex number is zero. You can quickly verify that

$$|z|^2 = zz^* \,. \tag{1.3}$$

For a wavefunction $\Psi(x, t)$ its complex conjugate $(\Psi(x, t))^*$ will be usually written as $\Psi^*(x, t)$.

We define the probability density $P(x, t)$, also denoted as $\rho(x, t)$, as the norm-squared of the wavefunction:

$$P(x,t) \;= \rho(x,t) \;\equiv\; \Psi^*(x,t)\Psi(x,t) \;=\; |\Psi(x,t)|^2 \,. \tag{1.4}$$

This probability density so defined is positive. The physical interpretation of the wavefunction arises because we declare that

$$\boxed{P(x,t)\,dx \;\text{ is the probability to find the particle in the interval } [x, x+dx] \text{ at time } t\,.} \tag{1.5}$$

This interpretation requires a *normalized* wavefunction, namely, the wavefunction used above must satisfy, for all times,

$$\int_{-\infty}^{\infty} dx\, |\Psi(x,t)|^2 \;=\; 1\,, \quad \forall\, t\,. \tag{1.6}$$

By integrating over space, the left-hand adds up the probabilities that the particle be found in all of the tiny intervals dx that comprise the real line. Since the particle must be found somewhere this sum must be equal to one.

Suppose you are handed a wavefunction that is normalized at time t_0:

$$\int_{-\infty}^{\infty} dx\, |\Psi(x,t_0)|^2 \;=\; 1\,, \quad \forall\, t\,. \tag{1.7}$$

As mentioned above, knowledge of the wavefunction at one time implies, via the Schrödinger equation, knowledge for all times. The Schrödinger equation must guarantee that the wavefunction remains normalized for all times. Proving this is a good exercise:

Exercise 1. Show that the Schrödinger equation implies that the norm of the wavefunction does not change in time:

$$\frac{d}{dt}\int_{-\infty}^{\infty} dx\,|\Psi(x,t)|^2 \;=\; 0\,. \tag{1.8}$$

You will have to use both the Schrödinger equation and its complex-conjugate version. Moreover you will have to use $\Psi(x,t) \to 0$ as $|x| \to \infty$, which is true, as no normalizable wavefunction can take a non-zero value as $|x| \to \infty$. While generally the derivative $\frac{\partial}{\partial x}\Psi$ also goes to zero as $|x| \to \infty$ you only need to assume that it remains bounded.

- - - - - - - - - - - - - - -

Associated to the probability density $\rho(x,t) = \Psi^*\Psi$ there is a **probability current** $J(x,t)$ that characterizes the flow of probability and is given by

$$J(x,t) \;=\; \frac{\hbar}{m}\mathrm{Im}\left(\Psi^*\frac{\partial\Psi}{\partial x}\right). \tag{1.9}$$

The analogy in electromagnetism is useful. There we have the current density vector $\vec{J}$ and the charge density ρ. The statement of charge conservation is the differential relation

$$\nabla\cdot\vec{J} + \frac{\partial\rho}{\partial t} \;=\; 0\,. \tag{1.10}$$

This equation applied to a fixed volume V implies that the rate of change of the enclosed charge $Q_V(t)$ is only due to the flux of $\vec{J}$ across the surface S that bounds the volume:

$$\frac{dQ_V}{dt}(t) \;=\; -\oint_S \vec{J}\cdot d\vec{a}\,. \tag{1.11}$$

Make sure you know how to get this equation from (1.10)! While the probability current in more than one spatial dimension is also a vector, in our present one-dimensional case, it has just one component. The conservation equation is the analog of (1.10):

$$\frac{\partial J}{\partial x} + \frac{\partial\rho}{\partial t} \;=\; 0\,. \tag{1.12}$$

You can check that this equation holds using the above formula for $J(x,t)$, the formula for $\rho(x,t)$, and the Schrödinger equation. The integral version is formulated by first defining the probability $P_{ab}(t)$ of finding the particle in the interval $x \in [a,b]$

$$P_{ab}(t) \;\equiv\; \int_a^b dx|\Psi(x,t)|^2 \;=\; \int_a^b dx\,\rho(x,t)\,. \tag{1.13}$$

You can then quickly show that

$$\frac{dP_{ab}}{dt}(t) \;=\; J(a,t) - J(b,t)\,. \tag{1.14}$$

Here $J(a,t)$ denotes the rate at which probability flows in (in units of one over time) at the left boundary of the interval, while $J(b,t)$ denotes the rate at which probability flows out at the right boundary of the interval.

It is sometimes easier to work with wavefunctions that are not normalized. The normalization can be perfomed if needed. We will thus refer to wavefunctions in general without assuming normalization, otherwise we will call them *normalized* wavefunction. In this spirit, two wavefunctions Ψ_1 and Ψ_2 solving the Schrödinger equation are *declared* to be physically equivalent if they differ by multiplication by a complex *number*. Using the symbol $\sim$ for equivalence, we write

$$\Psi_1 \sim \Psi_2 \quad \longleftrightarrow \quad \Psi_1(x,t) = \alpha\, \Psi_2(x,t)\,, \quad \alpha \in \mathbb{C}\,. \tag{1.15}$$

If the wavefunctions Ψ_1 and Ψ_2 are normalized they are equivalent if they differ by an overall constant phase:

$$\text{Normalized wavefunctions:} \quad \Psi_1 \sim \Psi_2 \quad \longleftrightarrow \quad \Psi_1(x,t) = e^{i\theta}\, \Psi_2(x,t)\,, \quad \theta \in \mathbb{R}\,. \tag{1.16}$$

2 Stationary Solutions

In a large class of problems the Schrödinger potential $V(x,t)$ has no time dependence and it is simply a function $V(x)$ of position. We focus on that case now. The Schrödinger equation (1.2) can be written more briefly as

$$i\hbar \frac{\partial \Psi}{\partial t}(x,t) \;=\; \hat{H}\, \Psi(x,t)\,, \tag{2.1}$$

where we have introduced the Hamiltonian operator $\hat{H}$:

$$\hat{H} \;\equiv\; -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x)\,. \tag{2.2}$$

$\hat{H}$ is an operator in the sense that it acts on functions of x and t to give functions of x and t: it acts on the space of complex functions, a space that contains wavefunctions. Note that $V(x)$ acts just by multiplication. Note that the operator $\hat{H}$ is time independent – it does not involve time at all.

A **stationary** state of energy $E \in \mathbb{R}$ is a state $\Psi(x,t)$ that takes the form

$$\Psi(x,t) \;=\; e^{-iEt/\hbar}\, \psi(x)\,, \tag{2.3}$$

where $\psi(x) \in \mathbb{C}$ is a function of x only that solves an equation that will be discussed below. All the time dependence of the stationary state is carried by the exponential prefactor. Such a state is called stationary because physical observables of the state are actually time independent. Consider, for example, the norm of the state. We see that the time dependence drops out

$$P(x,t) \;=\; \Psi^*(x,t)\, \Psi(x,t) \;=\; e^{+iEt/\hbar}\, \psi^*(x)\, e^{-iEt/\hbar}\, \psi(x) \;=\; \psi^*(x)\psi(x) \;=\; |\psi(x)|^2\,. \tag{2.4}$$

Had the energy E been a complex number $E = E_0 - i\Gamma$, with E_0 and Γ real, the time dependence would not drop out:

$$\begin{aligned} P(x,t) &= \Psi^*(x,t)\,\Psi(x,t) = e^{+iE^*t/\hbar}\,\psi^*(x)\,e^{-iEt/\hbar}\,\psi(x) \\ &= e^{i(E^*-E)t/\hbar}\psi^*(x)\psi(x) = e^{-2\Gamma t/\hbar}\,|\psi(x)|^2\,. \end{aligned} \tag{2.5}$$

This kind of state is not acceptable: the normalization cannot be preserved in time.

Let us consider the equation that $\psi(x)$ must satisfy. Plugging (2.3) into (2.1) we find

$$i\hbar\frac{\partial}{\partial t}e^{-iEt/\hbar}\,\psi(x) = \hat{H}e^{-iEt/\hbar}\,\psi(x)\,, \tag{2.6}$$

The time derivative on the left-hand side only acts on the exponential and the $\hat{H}$ operator on the right-hand side can be moved through the exponential (it commutes with it!) so we get

$$i\hbar\Big(\frac{\partial}{\partial t}e^{-iEt/\hbar}\Big)\,\psi(x) = e^{-iEt/\hbar}\,\hat{H}\psi(x)\,. \tag{2.7}$$

Taking the derivative and canceling the exponentials we get

$$i\hbar\Big(-i\frac{E}{\hbar}\Big)\,\psi(x) = \hat{H}\,\psi(x)\,, \tag{2.8}$$

which we write as

$$\boxed{\hat{H}\,\psi(x) = E\,\psi(x)\,.} \tag{2.9}$$

Recalling the expression for the Hamiltonian we have

$$\boxed{\Big(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V(x)\Big)\,\psi(x) = E\,\psi(x)\,.} \tag{2.10}$$

Note that the derivatives along x need not be denoted as partial derivatives since the functions they act on have no other argument except x. Using primes to denote derivatives with respect to the argument, the above equation is

$$\boxed{-\frac{\hbar^2}{2m}\psi''(x) + V(x)\psi(x) = E\,\psi(x)\,.} \tag{2.11}$$

This is the equation for ψ that makes $\Psi(x,t) = e^{-iEt/\hbar}\psi(x)$ a stationary state of energy E. Any of the three boxed equations above is referred to as the **time-independent** Schrödinger equation.

Since the time-independent Schrödinger equation is a second-order differential equation in x, a solution is completely determined by the value of ψ and ψ' at some point x_0 ($\neq \infty$). If $V(x)$

is nowhere infinite, $\psi = \psi' = 0$ at some point implies $\psi = 0$ everywhere. Alternatively, if we know the solution for *any* size x-interval, the full solution is fully determined. A full solution means finding all the values E for which acceptable solutions $\psi(x)$ exist and, of course, finding those solutions for each E.

A solution $\psi(x)$ associated with an energy E is called an **energy eigenstate** of energy E. The set of all allowed values of E is called the **spectrum** of the Hamiltonian $\hat{H}$. A **degeneracy** in the spectrum occurs when there is more than one solution $\psi(x)$ for a given value of the energy.

The solutions depend on the properties of the potential $V(x)$. We will consider potentials $V(x)$ that can fail to be continuous (but are piece-wise continuous, like the finite square well) and can fail to be bounded (like the potential for the harmonic oscillator). We allow delta function contributions in the potential but do not allow worse singularities, such as squares or derivatives of delta functions. We allow hard walls, as in the infinite square-well.

On the wavefunction we impose the following regularity condition:

$$\boxed{\psi(x) \text{ is continuous and bounded and its derivative } \psi'(x) \text{ is bounded.}} \tag{2.12}$$

We do not impose the requirement that $\psi(x)$ be normalizable. This would be too restrictive. There are energy eigenstates that are not normalizable. Momentum eigenstates of a free particle are also not normalizable. Solutions for which ψ is not normalizable do not have a direct physical interpretation, but are very useful: suitable superpositions of them give normalizable solutions that can represent a particle.

In the spectrum of a Hamiltonian, localized energy eigenstates are particularly important. This motivates the definition:

$$\boxed{\text{An energy eigenstate } \psi(x) \text{ is a } \textbf{bound} \text{ state if } \psi(x) \to 0 \text{ when } |x| \to \infty\,.} \tag{2.13}$$

Since a normalizable eigenstate must have a wavefunction that vanishes as $|x| \to \infty$, a bound state is just a normalizable eigenstate.

The eigenstates of $\hat{H}$ provide a useful set of functions. Let us denote the possible energies by E_n with $n = 1, 2, \ldots$, ordered as follows

$$E_1 \leq E_2 \leq E_3 \leq \ldots \tag{2.14}$$

and let the corresponding eigenstates be $\psi_n(x)$, with

$$\hat{H}\psi_n(x) \;=\; E_n\,\psi_n(x)\,, \quad n \geq 1\,. \tag{2.15}$$

For simplicity we discuss the case when the spectrum is denumerable so that, as above, we can label the states and energies with the integer label n. In general a potential $V(x)$ can result in a

spectrum that contains a discrete part and a continuous part. The discrete part is denumerable but the continuous part is not. The formulae we will write below require some modification when there spectrum contains a continuous part. The eigenstates of the continuum spectrum are not normalizable.

It is a known result about differential equations that for rather general potentials the $\hat{H}$ eigenstates $\psi_n(x)$ can be chosen to be orthonormal. What does it mean for two functions to be orthogonal? Orthogonal vectors have a vanishing dot product, where the dot product is a (clever) rule to obtain a single number from two vectors. For two functions f_1 and f_2 an *inner* product can be defined by integrating the product function $f_1 f_2$ over all x, thus giving us a number. Since our functions are complex valued, a small modification is needed: the inner product of f_1 and f_2 is taken to be $\int f_1^* f_2$. The functions f_1 and f_2 are orthogonal if this integral vanishes. An orthonormal set of functions is one in which each function is orthogonal to all others, while its inner product with itself gives one (this requires the complex conjugation in the definition, can you see that?). As a result, orthonormality means that

$$\boxed{\text{Orthonormality:} \quad \int_{-\infty}^{\infty} dx\ \psi_m^*(x)\,\psi_n(x) \ = \ \delta_{m,n}\,.} \tag{2.16}$$

Recall that the Kronecker delta $\delta_{m,n}$ is defined to be zero if $m \neq n$ and one otherwise.

The energy eigenstates are also *complete* in the sense that any reasonable (see (2.12) wavefunction $\psi(x)$ can be expanded as a superposition of energy eigenstates. Namely, there exist complex numbers b_n such that

$$\psi(x) \ = \ \sum_{n=1}^{\infty} b_n\,\psi_n(x)\,, \qquad b_n \in \mathbb{C}\,. \tag{2.17}$$

This is a very powerful statement: it means that if the energy eigenstates are known, the general solution of the Schrödinger equation is known. Indeed assume that the wavefuntion at time equal zero is the $\psi(x)$ above. Then we have

$$\boxed{\Psi(x, t=0) \ = \ \psi(x) \ = \ \sum_{n=1}^{\infty} b_n\,\psi_n(x)\,.} \tag{2.18}$$

If this wavefunction is normalized then we have

$$\int_{-\infty}^{\infty} dx\,\psi^*(x)\psi(x) \ = \ 1 \quad \rightarrow \quad \sum_{n=1}^{\infty} |b_n|^2 \ = \ 1\,. \tag{2.19}$$

We now claim that the wavefunction at all times can be written down immediately

$$\boxed{\Psi(x,t) \ = \ \sum_{n=1}^{\infty} b_n\, e^{-iE_n t/\hbar}\psi_n(x)\,.} \tag{2.20}$$

To prove that this is the solution we first note that we have produced a solution to the Schrödinger equation: this follows by linearity because each term in the above sum is a solution (a stationary state). Second, the solution reduces for $t = 0$ to the correct value $\Psi(x, t = 0)$ in (2.18). By the first remark below (1.2) this is all that is needed.

It should be emphasized that the superposition of stationary states is generally *not* a stationary state. The expansion coefficients b_n used above can be calculated explicitly if we know the energy eigenstates. Indeed using (2.16) and (2.17) a one-line computation (do it!) gives

$$b_n \;=\; \int_{-\infty}^{\infty} dx\, \psi_n^*(x)\psi(x)\,. \tag{2.21}$$

A curious identity can be derived by substituting this result back into (2.17):

$$\psi(x) \;=\; \sum_{n=1}^{\infty}\Big(\int_{-\infty}^{\infty} dx'\, \psi_n^*(x')\psi(x')\Big)\,\psi_n(x) = \;\int_{-\infty}^{\infty} dx'\Big(\sum_{n=1}^{\infty}\psi_n^*(x')\psi_n(x)\Big)\,\psi(x')\,, \tag{2.22}$$

where we interchanged the order of integration and summation (a safe operation in most cases!). The above equation is of the form

$$f(x) \;=\; \int_{-\infty}^{\infty} dx' K(x', x) f(x') \tag{2.23}$$

and is supposed to hold for any function $f(x)$. It is intuitively clear that that $K(x', x)$ must vanish for $x' \neq x$ for otherwise we could cook up a contradiction by choosing a peculiar function $f(x)$. Taking $f(x) = \delta(x - x_0)$ the equation gives

$$\delta(x - x_0) \;=\; \int dx' K(x', x)\delta(x' - x_0) \;=\; K(x_0, x)\,. \tag{2.24}$$

We therefore conclude that $K(x', x) = \delta(x - x')$ (recall that $\delta(x) = \delta(-x)$). Back in (2.22) we thus find

$$\boxed{\text{Completeness:}\quad \sum_{n=1}^{\infty}\psi_n^*(x')\psi_n(x) \;=\; \delta(x - x')\,.} \tag{2.25}$$

Let us compare the completeness relation above with the orthonormality relation (2.16). In the completeness relation we set equal the labels of the eigenfunctions and sum over them while keeping the two position arguments fixed. In the orthogonality relation we set equal the position arguments of the eigenfunctions and integrate (sum) over them while keeping the two labels fixed. On the right-hand sides we find "delta functions": a Kronecker delta setting equal the two labels in the orthonormality relation and a true delta function setting equal the two positions in the completeness relation. The two relations are obtained from each other by exchange of labels: position labels and energy labels. This is a neat duality!

It is fun to calculate the expectation value of the Hamiltonian in the solution $\Psi(x,t)$ in (2.20). For arbitrary time-independent operators $\hat{A}$ one defines the (generally) time-dependent expectation value on a *normalized* state Ψ by

$$\langle\, \hat{A}\,\rangle_\Psi(t) \;\equiv\; \int_{-\infty}^{\infty} dx\, \Psi^*(x,t)(\hat{A}\Psi(x,t))\,. \tag{2.26}$$

What happens when we take the operator to be $\hat{H}$? Using (2.20) twice, we get

$$\begin{aligned}
\langle\, \hat{H}\,\rangle_\Psi(t) \;&=\; \int_{-\infty}^{\infty} dx\, \Psi^*(x,t)(\hat{H}\Psi(x,t)) \\
&=\; \sum_{n,n'} \int_{-\infty}^{\infty} dx\, b_n^*\, e^{iE_n t/\hbar}\psi_n^*(x)\;\; b_{n'}\, e^{-iE'_n t/\hbar}\hat{H}\psi_{n'}(x) \\
&=\; \sum_{n,n'} b_n^* b_{n'} E_{n'}\, e^{i(E_n - E_{n'})t/\hbar} \int_{-\infty}^{\infty} dx\;\, \psi_n^*(x)\,\psi_{n'}(x) \\
&=\; \sum_{n,n'} b_n^* b_{n'} E_{n'}\, e^{i(E_n - E_{n'})t/\hbar}\delta_{n,n'}\,,
\end{aligned} \tag{2.27}$$

so that we get

$$\langle\, \hat{H}\,\rangle_\Psi(t) \;=\; \sum_{n=1}^{\infty} \;|b_n|^2 E_n\,. \tag{2.28}$$

The expectation value of the Hamiltonian is time-independent: this is the quantum version of energy conservation. This is the expected value of the energy: a weighted sum of the possible energies with weights the norm-squared of the expansion coefficients.

If the wavefunction $\Psi(x,t)$ is not normalized but is normalizable, then the wavefunction

$$\frac{\Psi(x,t)}{\sqrt{\int dx\, \Psi^*\Psi}} \tag{2.29}$$

is normalized. We can thus use this normalized wavefunction in the definition on $\langle\hat{A}\rangle$ to find the expectation value is given by

$$\langle\, \hat{A}\,\rangle_\Psi(t) \;\equiv\; \frac{\int_{-\infty}^{\infty} dx\, \Psi^*(x,t)(\hat{A}\Psi(x,t))}{\int dx\, \Psi^*(x,t)\Psi(x,t)}\;\;. \tag{2.30}$$

This formula can be used for any normalizable Ψ. If the Ψ is normalized the formula reduces to the earlier expression for $\langle\hat{A}\rangle$.

Another operator often used to explore the physics of states is the momentum operator $\hat{p}$. Acting on wavefuntions that depend on a coordinate x it takes the form of a differential operator:

$$\hat{p} \;\equiv\; \frac{\hbar}{i}\frac{\partial}{\partial x}\,. \tag{2.31}$$

3 Properties of energy eigenstates in one dimension

In order to simplify our notation we rewrite the time-independent Schrödinger equation (2.10) as follows

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}(E - V(x))\psi \;=\; 0\,. \tag{3.1}$$

We then define energy-like quantities $\mathcal{E}$ and $\mathcal{V}$ using a common rescaling factor:

$$\mathcal{E} \;\equiv\; \frac{2m}{\hbar^2}E\,, \quad \mathcal{V}(x) \;\equiv\; \frac{2m}{\hbar^2}V(x)\,. \tag{3.2}$$

With this the Schrödinger equation (3.1) becomes

$$\boxed{\psi'' + (\mathcal{E} - \mathcal{V}(x))\psi \;=\; 0\,.} \tag{3.3}$$

We are now ready to consider a basic result: in a one-dimensional potential there cannot be two or more bound states for any given energy.

Theorem 1. There is no degeneracy for bound states in one-dimensional potentials.

Proof. Suppose there is such degeneracy so that there are $\psi_1(x)$ and $\psi_2(x)$, different from each other and both corresponding to the same energy E, thus same value of $\mathcal{E}$. If so, we have that the following equations hold

$$\begin{aligned} \psi_1'' + (\mathcal{E} - \mathcal{V}(x))\psi_1 \;&=\; 0\,, \\ \psi_2'' + (\mathcal{E} - \mathcal{V}(x))\psi_2 \;&=\; 0\,. \end{aligned} \tag{3.4}$$

Multiplying the top equation by ψ_2 and the bottom one by ψ_1 and subtracting them we find

$$\psi_2\psi_1'' - \psi_1\psi_2'' \;=\; 0\,. \tag{3.5}$$

The left-hand side is actually a derivative

$$(\psi_2\psi_1' - \psi_1\psi_2')' \;=\; 0\,. \tag{3.6}$$

It follows from this that the expression inside the parenthesis must be a constant c,

$$\psi_2\psi_1' - \psi_1\psi_2' \;=\; c\,. \tag{3.7}$$

The constant can be evaluated by examining the left-hand side for $|x| \to \infty$. We then have that $\psi_1 \to 0$ and $\psi_2 \to 0$, since they are bound states, while the derivatives are bounded, as assumed in (2.12). It follows that the left-hand side vanishes as $|x| \to \infty$ and therefore $c = 0$. We thus have

$$\psi_2\psi_1' = \psi_1\psi_2' \quad \to \quad \frac{\psi_1'}{\psi_1} \;=\; \frac{\psi_2'}{\psi_2} \quad \to \quad \frac{d}{dx}(\ln\psi_1 - \ln\psi_2) = 0\,. \tag{3.8}$$

This implies that we have for some constant c'

$$\ln \psi_1 = \ln \psi_2 + \ln c' \quad \rightarrow \quad \psi_1(x) = c' \psi_2(x) \,. \tag{3.9}$$

We have thus shown that the wavefunctions ψ_1 and ψ_2 are equivalent. In contradiction with the initial assumption, they are the same energy eigenstate. This concludes the proof.
For our second theorem we show that the reality of $\mathcal{V}$ allows us to work with real wavefunctions.

Theorem 2. The energy eigenstates $\psi(x)$ can be chosen to be real.

Proof. Consider our main equation and a possibly complex wavefunction that correspond

$$\psi'' + (\mathcal{E} - \mathcal{V}(x))\psi \;=\; 0 \,, \tag{3.10}$$

Since $(\psi'')^* = (\psi^*)''$ the complex conjugation of the above equation gives

$$(\psi^*)'' + (\mathcal{E} - \mathcal{V}(x))\psi^* \;=\; 0 \,. \tag{3.11}$$

So ψ^* if different from ψ defines a degenerate solution. By superposition we can then get two real (degenerate) solutions

$$\psi_r \equiv \frac{1}{2}(\psi + \psi^*) \,, \quad \psi_{im} \;\equiv\; \frac{1}{2i}(\psi - \psi^*) \,. \tag{3.12}$$

These are, of course, the real and imaginary parts of ψ.

If we are dealing with bound states of one-dimensional potentials more can be said: any such solution is, up to a phase, *equal* to a real solution. Indeed, the absence of degenerate bound states means that the two real solutions ψ_r and ψ_{im} must be equal up to a constant that can only be real:

$$\psi_{im} = c\, \psi_r \,, \quad \text{with} \;\; c \in \mathbb{R} \tag{3.13}$$

It then follows that $\psi = \psi_r + i\psi_{im} = (1 + ic)\psi_r$. Writing $1 + ic = \sqrt{1 + c^2}\, e^{i\beta}$ with real beta, shows that ψ is, up to a phase β, equal to a real solution.

Theorem 3. If the potential is an even function of x: $V(-x) = V(x)$ the eigenstates can be chosen to be even or odd under $x \to -x$.

Proof. Again, we begin with our main equation

$$\psi''(x) + (\mathcal{E} - \mathcal{V}(x))\psi(x) \;=\; 0 \,. \tag{3.14}$$

Recall that primes denote here derivative with respect to the argument, so $\psi''(x)$ means the function "second-derivative-of-ψ" evaluated at x. Similarly $\psi''(-x)$ means the function "second-derivative-of-ψ" evaluated at $-x$. Thus we can change x for $-x$ with impunity in the above equation getting

$$\psi''(-x) + (\mathcal{E} - \mathcal{V}(x))\psi(-x) \;=\; 0 \,, \tag{3.15}$$

where we used that V, and thus $\mathcal{V}$, is even. We now want to make clear that the above equation implies that $\psi(-x)$ is another solution of the Schrödinger equation with the same energy. For this let us define

$$\varphi(x) \equiv \psi(-x) \quad \rightarrow \quad \frac{d}{dx}\varphi(x) = \psi'(-x)\cdot(-1)\,. \tag{3.16}$$

Taking a second derivative and using (3.15)

$$\frac{d^2}{dx^2}\,\varphi(x) = \psi''(-x) = -(\mathcal{E}-\mathcal{V}(x))\varphi(x)\,, \tag{3.17}$$

so that indeed $\varphi(x)=\psi(-x)$ provides a degenerate solution to the Schrödinger equation:

$$\frac{d^2}{dx^2}\,\varphi(x) + (\mathcal{E}-\mathcal{V}(x))\varphi(x) = 0\,. \tag{3.18}$$

Equipped with the degenerate solutions $\psi(x)$ and $\psi(-x)$ we can now form symmetric (s) and antisymmetric (a) combinations that are, respectively, even and odd under $x\rightarrow -x$:

$$\psi_s(x) \equiv \frac{1}{2}(\psi(x)+\psi(-x))\,, \quad \psi_a(x) \equiv \frac{1}{2}(\psi(x)-\psi(-x))\,. \tag{3.19}$$

These are the solutions claimed to exist in Theorem 3.

Again, if we focus on bound states of one-dimensional potentials the absence of degeneracy implies that $\psi(x)$ and $\psi(-x)$ must be the same solution. Because of Theorem 2 we can choose $\psi(x)$ to be real and thus we must have

$$\psi(-x) = c\,\psi(x)\,, \quad \text{with } c\in\mathbb{R}\,. \tag{3.20}$$

Letting $x\rightarrow -x$ in the above equation we get $\psi(x)=c\psi(-x)=c^2\psi(x)$ from which we learn that $c^2=1$. The only possibilities are $c=\pm 1$. So $\psi(x)$ is *automatically* even or odd under $x\rightarrow -x$. Any one-dimensional bound state solution with an even potential *must be* either even or odd under $x\rightarrow -x$.

4 The nature of the spectrum

Consider the time-independent Schrödinger equation written as

$$\psi'' = -\frac{2m}{\hbar^2}(E-V(x))\,\psi\,. \tag{4.1}$$

We always have that $\psi(x)$ is continuous, otherwise ψ'' has singularities worse than delta functions and we would require potentials $V(x)$ that are worse than delta functions – something we will not consider. Consider now three possibilities concerning the potential:

1. $V(x)$ is continuous. In this case the continuity of $\psi(x)$ and (4.1) imply ψ'' is also continuous. This requires ψ' continuous.

2. $V(x)$ has finite jumps. In this case ψ'' has finite jumps (it is the multiplication of ψ with no jumps times V with jumps). But then ψ' can have no jumps (it is continuous, with non-continuous derivative).

3. $V(x)$ contains delta functions. In this case ψ'' also contains delta functions (it is the multiplication of the continuous ψ times a delta function in V). Thus ψ' has finite jumps.

4. $V(x)$ contains a hard wall. A potential that is finite immediately to the left of $x = a$ and becomes infinite for $x > a$ is said to have a hard wall at $x = a$. In such a case, the wavefunction will vanish for $x \geq a$. The slope ψ' will be finite as $x \to a$ from the left, and will vanish for $x > a$. Thus ψ' is discontinuous at the wall.

In conclusion

Both ψ and ψ' are continuous unless the potential has delta functions or hard walls in which cases ψ' may have finite jumps.	(4.2)

The origin of the discrete and continuous spectrum can be seen from simple examples. We have three situations to discuss, as shown in Figure 1 as (a), (b), and (c). We will consider the number of parameters needed to write a solution and the number of constraints due to boundary conditions. Without loss of generality we can consider real solutions, and therefore the parameters will be real.

(a) Here the energy E is below the potential far to the left and far to the right, but not in the middle. On the left the solution must be a decaying exponential $\alpha_1 \exp(-\kappa|x|)$, where α_1 is a constant to be determined and κ is known if the energy E is known. So thus far we got one unknown constant α_1. In the middle region where $E > V$ the solution is oscillatory $\alpha_2 \cos kx + \alpha_3 \sin kx$, with two unknown constants α_2 and α_3, and k determined if E is known. Finally to the right we have a solution $\alpha_4 \exp(-\kappa x)$ since the wavefunction must vanish as $x \to \infty$. So we got four (real) unknown constants α_i, $i = 1, 2, 3, 4$. Since ψ and $c\psi$ are the same solution we can scale the solution and thus we only have three unknown constants to determine. There are, however, four constraints from boundary conditions: the continuity of ψ and ψ' at each of the two interfaces. With three coefficients and four conditions we cannot expect a solution to exist. If we view the energy E, however, as unknown, then we have four unknowns and four conditions. Indeed solutions exist for discrete values of the energy. We get a discrete spectrum.

(b) Here we have one unknown constant for the solution to the left of the interface (multiplying a decaying exponential) and two unknown constants for the oscillatory solution to the right of the interface, for a total of three unknowns, or just two unknowns once the overall scale

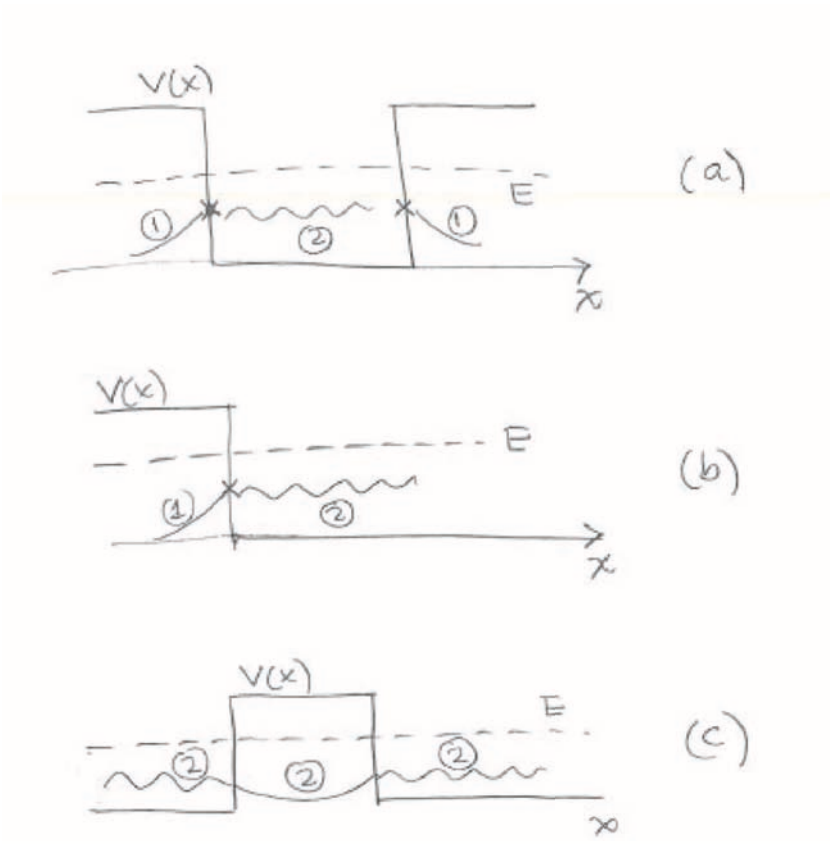


Figure 1: Discussing the number of constants needed to specify a solution. (a) Energy is smaller than the potential for $x \to \pm\infty$. (b) Energy is smaller than the potential for $x \to -\infty$ and larger than the potential for $x \to \infty$. (c) Energy is larger than the potential for $x \to \pm\infty$.

freedom is accounted in. We also have two boundary conditions at the interface. So we can expect a solution. Indeed there should be a solution for each value of the energy. The spectrum here is continuous and non-degenerate.

(c) Two constants are needed here in each of the three regions: they multiply sines and cosines to the left and to the right, and multiply the two exponentials in the middle. Thus six constants and due to scaling just five unknowns. We still have four boundary conditions so there should be solutions. In fact, there are two solutions for each energy. We can understand this as follows. Think of using just one coefficient to the far left, say the coefficient multiplying the sine function. With one less coefficient we have the same number of unknowns as constraints so we should get one solution (for any E). We get another solution if we use the cosine function to the far left. So we have two solutions for each energy. The spectrum is continuous and doubly degenerate.

Figure 2 illustrates the spectrum of the Hamiltonian for a rather generic type of potential. Here V_+ is the top asymptote of the potential, V_- is the bottom asymptote of the potential, and V_0 is the lowest value of the potential. In the figure we indicate the type of spectrum for energies in the various intervals defined: $E > V_+$, then $V_- < E < V_+$, then $V_0 < E < V_-$ and finally $E < V_0$.

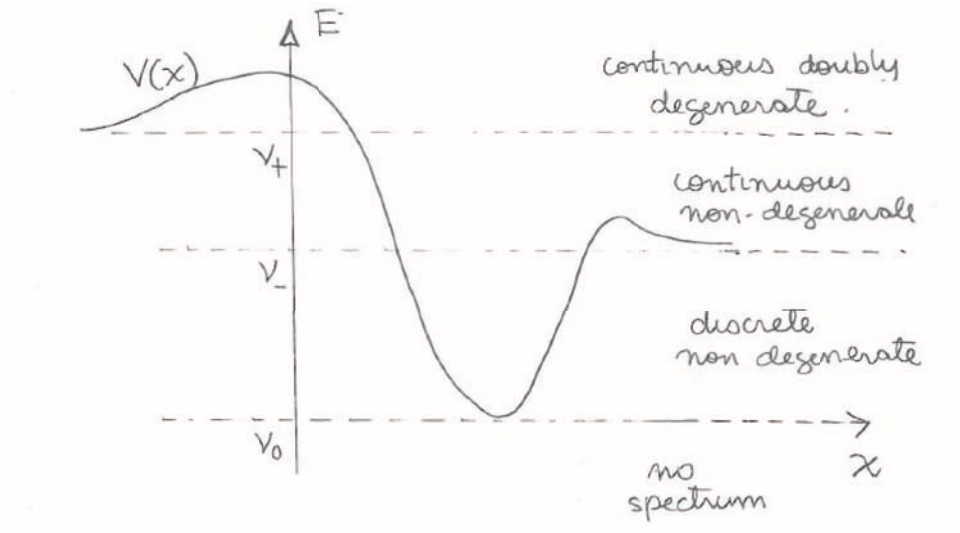


Figure 2: A generic potential and the type of spectrum for various energy ranges.

A **node** in a wavefunction is a point x_0 where $\psi(x_0) = 0$ (a zero of ψ) and $\psi'(x_0) \neq 0$. For a bound state the zeroes are nodes or the points at infinity (where typically $\psi' \to 0$).

Theorem 4 For the discrete bound-state spectrum of a one-dimensional potential let the allowed energies be $E_1 < E_2 < E_3 < \ldots$ with E_1 the ground state energy. Let the associated energy eigenstates be $\psi_1, \psi_2, \psi_3, \ldots$. The wavefunction ψ_1 has no nodes, ψ_2 has one node, and each consecutive wavefunction has one additional node. In conclusion ψ_n has $n-1$ nodes.

We will not prove this theorem here. In fact you will show in the homework that ψ_{k+1} has at least one node between two consecutive zeroes of ψ_k. This implies that ψ_{k+1} has at least one more node than ψ_k. This can be illustrated in Figure 3 that shows a bound state $\psi_4(x)$ with three nodes at x_1, x_2, and x_3 and zeroes at $x = -\infty$ and $x = \infty$. For ψ_5 there must be a node w_1 in $(-\infty, x_1]$, a node $w_2 \in (x_1, x_2)$ and so on until a last node $w_4 \in (x_3, \infty)$.

Example: Potential with five delta functions. We will discuss the bound states of the Schrödinger equation with potential

$$V(x) \;=\; -V_0 a \sum_{n=-2}^{2} \delta(x - na)\,. \tag{4.3}$$

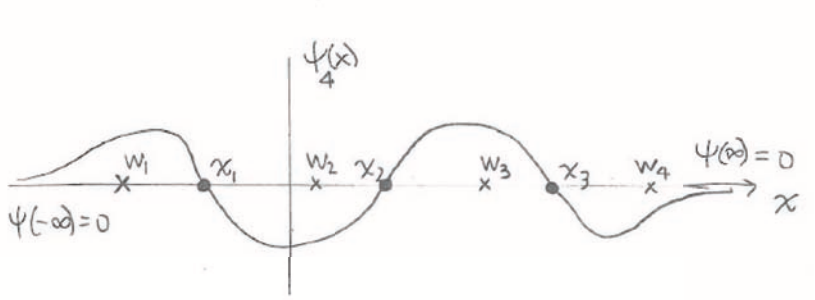


Figure 3: A wavefunction ψ_4 with three nodes (x_1, x_2, x_3) and zeroes at $x \pm \infty$. The next wavefunction ψ_5 must have four nodes, with positions indicated by w_1, w_2, w_3 and w_4.

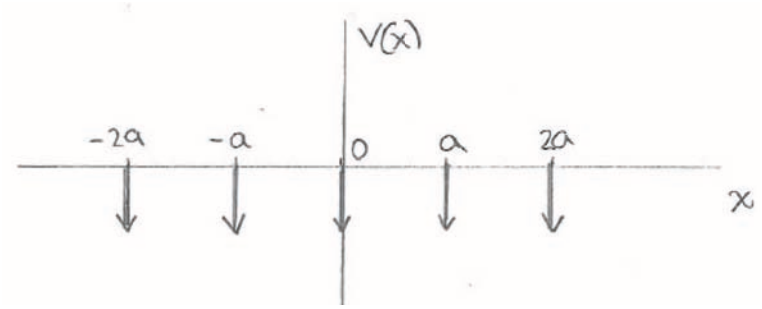


Figure 4: A potential $V(x)$ with five downwards pointing delta-functions.

This potential has delta functions at x equal to $-2a, -a, 0, a$, and $2a$, as shown in Figure 4.

We first examine the effect of the delta functions on the eigenstates. We will see that they produce discontinuities in ψ' at the position of the delta functions. We begin with the Schrödinger equation

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + V(x)\psi(x) \;=\; E\psi(x)\,, \tag{4.4}$$

and integrate this equation from $a-\epsilon$ to $a+\epsilon$, where ϵ is a small value that we will take down to zero. By doing this we will get one out of the five delta functions to fire. We find

$$-\frac{\hbar^2}{2m}\int_{a-\epsilon}^{a+\epsilon} dx\frac{d^2\psi}{dx^2} + \int_{a-\epsilon}^{a+\epsilon} dx V(x)\psi(x) \;=\; E\int_{a-\epsilon}^{a+\epsilon} dx\,\psi(x)\,. \tag{4.5}$$

The first term involves a total derivative, the second term just picks up the delta function at $x = a$, and the right hand side is evaluated by noting that since ψ is continuous its value at $x = a$ gives the leading contribution:

$$-\frac{\hbar^2}{2m}\frac{d\psi}{dx}\Big|_{\alpha-\epsilon}^{a+\epsilon} \;-\; V_0 a\int_{a-\epsilon}^{a+\epsilon} dx\delta(x-a)\psi(x) \;=\; E(2\epsilon)\psi(a) + \mathcal{O}(\epsilon^2)\,. \tag{4.6}$$

In the limit as $\epsilon \to 0$ we will denote $a+\epsilon$ as a^+ and $a-\epsilon$ as a^-. These labels are needed since ψ' has to be discontinuous at x. Indeed, we get

$$-\frac{\hbar^2}{2m}\left(\psi'(a^+)-\psi'(a^-)\right) \;-\; V_0\, a\psi(a) \;=\; 0\,. \tag{4.7}$$

This implies that the discontinuity $\Delta\psi'$ of ψ' is given by

$$\boxed{\Delta\psi'(a) \;\equiv\; \psi'(a^+)-\psi'(a^-) \;=\; \frac{2m}{\hbar^2}\,(-V_0 a)\,\psi(a)\,.} \tag{4.8}$$

The discontinuity of ψ' at the position of the delta function is proportional to the value of ψ at this point. The constant of proportionality is linear on the strength $V_0 a$ of the delta function. It follows that if the delta function of the potential is at a point where ψ vanishes then both ψ and ψ' are continuous and the delta function has no effect.

Let us now focus on bound states. These will be states with $E<0$. The Schrödinger equation away from the delta functions is just

$$\psi'' \;=\; -\frac{2mE}{\hbar^2}\,\psi \;=\; \kappa^2\,\psi\,, \quad \text{with} \;\; \kappa^2 \equiv -\frac{2mE}{\hbar^2} > 0\,,\; \kappa > 0\,. \tag{4.9}$$

The solutions are therefore the linear combinations

$$\psi(x) \;=\; ae^{-\kappa x} + be^{\kappa x}\,, \tag{4.10}$$

with a and b real constants to be determined (recall the wavefunction can be taken to be real). In Figure 5 we show these functions for $a>b>0$. Note the the curves intersect just once. It follows that the wavefunction will never have a zero if a and b have the same sign and it will have exactly one zero if a and b have opposite signs.

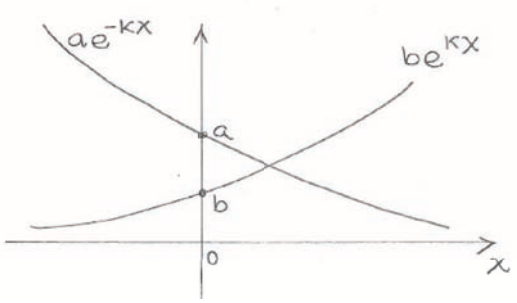


Figure 5: Plots of $ae^{-\kappa x}$ and $be^{\kappa x}$ with $a,b>0$. This can be used to show that any linear superposition of these two functions can at most have one zero.

Let us then make the following remarks:

1. *There cannot be zeroes of the wavefunction for $x \geq 2a$ (nor for $x \leq -2a$).* For $x \geq 2a$ the solution, if non vanishing, must be of the form $ce^{-\kappa x}$. This can only have a zero if $c = 0$. In this case the wavefunction would vanish identically for $x \geq 2a$. This does not look good. Since $\psi(2a) = 0$ then ψ' is not discontinuous and, by continuity, a bit to the left of $2a$ both ψ and ψ' vanish. This is enough to make the solution vanish over the next interval $x \in (a, 2a)$. Continuing in this way we find that the solution for ψ would have to be zero everywhere. This is not acceptable.

2. *There is at most one zero in between each pair of contiguous δ-functions.* This follows because the solution must take the form (4.10) and we argued that such function can at most have one zero.

3. Zeroes appear at $x = 0$ for all the antisymmetric bound states. In those cases, there cannot be another zero in the interval $[-a, a]$. Zeroes may appear at $x = \pm a$, but this is presumably not generic. There are at most five bound states because the maximum number of nodes is four; one in between each delta function. All these five bound states exist if the delta functions are strong enough. The ground state is even, has no nodes and presumably looks like the one drawn in Figure 6.

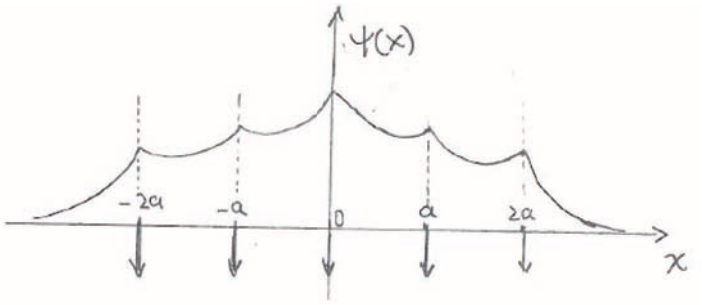


Figure 6: A sketch of the ground state wavefunction.

Exercise. Sketch the expected shapes of the four excited bound states of the potential.

5 Variational Principle

Consider a system with Hamiltonian $\hat{H}$ and focus on the time-independent Schrödinger equation:

$$\hat{H}\psi(\vec{x}) \;=\; E\psi(\vec{x})\,. \tag{5.11}$$

Let us assume that the system is such that it has a collection of energy eigenstates that are normalizable. This collection includes a ground state with ground state energy E_{gs}. Note the

use of $\vec{x}$: our discussion applies to quantum systems in any number of spatial dimensions. Our first goal is to learn something about the ground state energy without solving the Schrödinger equation nor trying to figure out the ground state wavefunction.

For this purpose, consider an *arbitrary* normalized wavefunction $\psi(\vec{x})$:

$$\int d\vec{x}\, \psi^*(\vec{x})\psi(\vec{x}) \;=\; 1\,. \tag{5.12}$$

By arbitrary we mean a wavefunction that need not satisfy the time-independent Schrödinger equation, a wavefunction that need not be an energy eigenstate. Then we claim the ground state energy E_{gs} of the Hamiltonian is smaller or equal than the expectation value of $\hat{H}$ in this arbitrary normalized ψ, namely,

$$\boxed{E_{gs} \;\leq\; \langle \hat{H} \rangle_\psi \;=\; \int d\vec{x}\, \psi^*(\vec{x})\, \hat{H}\psi(\vec{x})\,, \quad \text{Normalized } \psi}\,. \tag{5.13}$$

The wavefunction $\psi(\vec{x})$ here is sometimes called a *trial* wavefunction. When the right-hand side of the above inequality is evaluated we get an energy and learn that the ground state energy must be smaller or equal to the value we get. Thus any trial wavefunction provides an *upper bound* for the ground state energy. Better and better trial wavefunctions will produce lower and lower upper bounds. Note that if the trial wavefunction was set equal to the (unknown) ground-state wavefunction, the expectation value of $\hat{H}$ becomes exactly E_{gs} and the inequality is saturated.

Let us prove (5.13). For simplicity, we will consider here the case where the energy eigenstates $\psi_n(\vec{x})$ of $\hat{H}$ are denumerable and their corresponding energies E_n are ordered as

$$E_{gs} = E_1 \leq E_2 \leq E_3 \leq \ldots\,. \tag{5.14}$$

Of course $\hat{H}\psi_n = E_n\psi_n$. Since the energy eigenstates are complete, any trial wavefunction can be expanded in terms of them (see (2.18)):

$$\psi(\vec{x}) \;=\; \sum_{n=1}^{\infty} b_n\, \psi_n(\vec{x})\,. \tag{5.15}$$

Such a ψ is not an energy eigenstate in general. The normalization condition (5.12) gives us,

$$\sum_{n=1}^{\infty} |b_n|^2 = 1\,. \tag{5.16}$$

The evaluation of the right-hand side in (5.13) was done before in (2.28) so we have

$$\int d\vec{x}\, \psi^*(\vec{x})\, \hat{H}\psi(\vec{x}) \;=\; \sum_{n=1}^{\infty} |b_n|^2 E_n\,. \tag{5.17}$$

Since $E_n \geq E_1$ for all n, we can replace the E_n on the above right-hand side for E_1 getting a smaller or equal value:

$$\int d\vec{x}\, \psi^*(\vec{x})\, \hat{H}\psi(\vec{x}) \;\geq\; \sum_{n=1}^{\infty} |b_n|^2 E_1 \;=\; E_1 \sum_{n=1}^{\infty} |b_n|^2 \;=\; E_1 = E_{gs}\,, \tag{5.18}$$

where we used (5.16). This is in fact the claim in (5.13).

Is is sometimes more convenient not to worry about the normalization of the trial wavefunctions. Given a trial wavefunction ψ that is not normalized, the wavefunction

$$\frac{\psi(x)}{\sqrt{N}} \quad \text{with} \quad N = \int d\vec{x}\psi^*(\vec{x})\psi(\vec{x})\,, \tag{5.19}$$

is normalized and can be used in (5.13). We therefore find that

$$\boxed{E_{gs} \;\leq\; \frac{\int d\vec{x}\, \psi^*(\vec{x})\, \hat{H}\psi(\vec{x})}{\int d\vec{x}\, \psi^*(\vec{x})\psi(\vec{x})} \;\equiv\; \mathcal{F}[\psi]\,.} \tag{5.20}$$

This formula can be used for trial wavefunctions that are not normalized. We also introduced the definition of the functional $\mathcal{F}[\psi]$. A functional is a machine that given a function gives us a number. Our result states that the ground state energy arises as the minimum value that the functional can take.

One application of this variational principle is to find good upper bounds for the ground state energy of quantum systems that are not exactly solvable. For this purpose it is useful to construct trial wavefunctions $\psi(\vec{x}\,; \beta_1, \beta_2, \cdots \beta_m)$ that depend on a set of parameters β. One then computes the expectation value $\langle \hat{H} \rangle_\psi$ which, of course, is a function of the parameters. Any random values for the parameters will give an upper bound for the ground state energy, but by minimizing $\langle \hat{H} \rangle_\psi$ over the parameter space we get the lowest possible upper bound consistent with the chosen form for the trial wavefunction.

Example. (Griffiths). Consider a one-dimensional problem with the delta function potential

$$V(x) \;=\; -\alpha\, \delta(x)\,, \quad \alpha > 0\,. \tag{5.21}$$

In this problem the ground state energy is calculable exactly and one has

$$E_{gs} \;=\; -\frac{m\alpha^2}{2\hbar^2}\,. \tag{5.22}$$

So this problem is just for illustration. Consider an unnormalized gaussian trial wavefunction, with a real parameter β:

$$\psi(x) \;= e^{-\frac{1}{2}\beta^2 x^2}\,, \qquad \int_{-\infty}^{\infty} dx\, \psi^2 \;=\; \frac{\sqrt{\pi}}{\beta}\,. \tag{5.23}$$

The functional $\mathcal{F}$ in (5.20) is then[1]

$$\begin{aligned}
\frac{\int dx\, \psi^*(x)\, \hat{H}\psi(x)}{\int dx\, \psi^*(x)\psi(x)} &= \frac{\beta}{\sqrt{\pi}} \int dx\, e^{-\frac{1}{2}\beta^2 x^2} \Big(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} - \alpha\delta(x) \Big) e^{-\frac{1}{2}\beta^2 x^2} \\
&= \frac{\beta}{\sqrt{\pi}} \frac{\hbar^2}{2m} \int dx \left[\frac{d}{dx} e^{-\frac{1}{2}\beta^2 x^2} \right]^2 - \frac{\beta}{\sqrt{\pi}}\alpha \\
&= \frac{\beta}{\sqrt{\pi}} \frac{\hbar^2}{2m} \frac{\beta\sqrt{\pi}}{2} - \frac{\beta}{\sqrt{\pi}}\alpha \\
&= \frac{\beta^2\hbar^2}{4m} - \frac{\beta}{\sqrt{\pi}}\alpha\,.
\end{aligned} \tag{5.24}$$

The first term on the last right-hand side is the kinetic energy and the second term is the potential energy. For any value of β the final expression above provides an upper bound for the ground state energy, and the best upper bound is the lowest one. We thus have that the ground state energy satisfies

$$E_{gs} \leq \mathrm{Min}_\beta \Big(\frac{\beta^2\hbar^2}{4m} - \frac{\beta}{\sqrt{\pi}}\alpha \Big)\,. \tag{5.25}$$

The minimum is easily found

$$\beta = \frac{2m\alpha}{\hbar^2\sqrt{\pi}} \qquad \rightarrow \qquad E_{gs} \leq -\frac{m\alpha^2}{\pi\hbar^2} \;=\; \frac{2}{\pi}\Big(-\frac{m\alpha^2}{2\hbar^2} \Big)\,. \tag{5.26}$$

Comparing with (5.22) we see that the bound we found is in fact $\frac{2}{\pi}E_{gs} \simeq 0.64 E_{gs}$. The trial wavefuntion brought us to about 64% of the correct value.

In the exercises you will develop the following results:

1. With trial wavefunctions orthogonal to the ground state, the functional $\mathcal{F}$ gives upper bounds for the energy of the first excited state.

2. For any attractive one-dimensional potential (a nowhere positive potential that approaches zero at infinity) there is a bound state, namely, a state with energy less than zero.

3. We have shown that the functional $\mathcal{F}[\psi]$ has a minimum for ψ equal to the ground state wavefunction. Interestingly, this functional is stationary at each and every energy eigenstate. For eigenstates of energies higher than the ground state $\mathcal{F}$ has a saddle point.

[1] We use the integrals $\int du e^{-u^2} = \sqrt{\pi}$ and $\int du u^2 e^{-u^2} = \frac{1}{2}\sqrt{\pi}$.

6 Position and momentum

In quantum mechanics the position operator $\hat{x}$ and the momentum operator $\hat{p}$ do not commute. They satisfy the commutation relation

$$[\hat{x}\,,\hat{p}\,] \;=\; i\hbar\,. \tag{6.27}$$

When we deal with wavefuntions $\psi(x)$ the position operator acts on them in a simple way.[2] We define

$$\hat{x}\,\psi(x) \;\equiv\; x\psi(x)\,. \tag{6.28}$$

In words the position operator acting on an x-dependent wavefuntion simply multiplies the wavefunction by x. In quantum mechanics it is useful think of states as vectors and operators as matrices. A wavefuntion for a particle on the box $0 \leq x \leq a$, for example can be thought as vector with many components, each one giving the value of the function at a specific point. To make this concrete one discretizes the space into small intervals of size ϵ such that $N\epsilon = a$. In that case we can represent the information in $\psi(x)$ in a large column vector

$$\psi(x) \quad \longleftrightarrow \quad \begin{pmatrix} \psi(0) \\ \psi(\epsilon) \\ \psi(2\epsilon) \\ \vdots \\ \psi(N\epsilon) \end{pmatrix}\,. \tag{6.29}$$

The $N+1$ component column vector summarizes the values of the wavefunction at equally separated points. N is some kind of regulator: a precise description requires $N \to \infty$ or $\epsilon \to 0$. Associated with the description (6.29) the operator $\hat{x}$ can be viewed as the $(N+1)\times(N+1)$ diagonal matrix

$$\hat{x} \quad \longleftrightarrow \quad \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & \epsilon & 0 & \dots & 0 \\ 0 & 0 & 2\epsilon & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & N\epsilon \end{pmatrix}\,. \tag{6.30}$$

You can see that the action of the matrix $\hat{x}$ on the vector (6.29) gives the vector

$$\begin{pmatrix} 0\cdot\psi(0) \\ \epsilon\cdot\psi(\epsilon) \\ 2\epsilon\cdot\psi(2\epsilon) \\ \vdots \\ N\epsilon\cdot\psi(N\epsilon) \end{pmatrix}\,, \tag{6.31}$$

[2]The time dependence is irrelevant to the present discussion, which applies without changes to time-dependent wavefunctions $\Psi(x,t)$.

which is indeed the representation of $x\psi(x)$. Given our definition of the action of $\hat{x}$, expectation values in normalized states are naturally defined by

$$\langle \hat{x} \rangle \;\equiv\; \int dx'\, \psi^*(x')\, (x'\psi(x'))\,. \tag{6.32}$$

Are there eigenstates of the $\hat{x}$ operator? Yes, but their are not normalizable. An eigenstate of $\hat{x}$ must be a localized state, and the obvious candidate is a delta function. Defining

$$\psi_{x_0}(x) \;\equiv\; \delta(x - x_0)\,, \tag{6.33}$$

we verify that

$$\hat{x}\psi_{x_0}(x) \;=\; x\,\psi_{x_0}(x) \;=\; x\delta(x-x_0) \;=\; x_0\delta(x-x_0) \;=\; x_0\psi_{x_0}(x)\,, \tag{6.34}$$

confirming that $\psi_{x_0}(x)$ is an eigenstate of $\hat{x}$ with eigenvalue x_0. A delta function cannot be normalized, so the position eigenstates are not normalizable.

When we speak about the state of a particle and describe it with the wavefunction $\psi(x)$ we are using the "position" representation of the state. The momentum operator in the position representation is given by

$$\hat{p} \;\equiv\; \frac{\hbar}{i}\frac{\partial}{\partial x} \tag{6.35}$$

This is to say that acting on a wavefunction we have

$$\hat{p}\,\psi(x) \;=\; \frac{\hbar}{i}\frac{d\psi}{dx}\,. \tag{6.36}$$

Note that the commutation relation (6.27) is satisfied by the above definitions, as we can check acting on any wavefuntion:

$$\begin{aligned}
[\hat{x}\,,\hat{p}\,]\psi(x) \;&=\; (\hat{x}\hat{p} - \hat{p}\hat{x})\psi(x) \\
&= \hat{x}\hat{p}\,\psi(x) - \hat{p}\hat{x}\,\psi(x) \\
&= \hat{x}\,\frac{\hbar}{i}\frac{d\psi}{dx} - \hat{p}\,x\psi(x) \\
&= \frac{\hbar}{i}\,x\frac{d\psi}{dx} - \frac{\hbar}{i}\frac{d}{dx}(x\psi) \\
&= \frac{\hbar}{i}\,x\frac{d\psi}{dx} - \frac{\hbar}{i}\psi - \frac{\hbar}{i}x\frac{d\psi}{dx} \\
&= i\hbar\,\psi(x)\,.
\end{aligned} \tag{6.37}$$

Since the wavefunction is arbitrary, we have verified that our explicit representation of the operators $\hat{x}$ (by multiplication) and $\hat{p}$ (by differentiation) satisfies the commutation relation $[\hat{x},\hat{p}] = i\hbar$.

Eigenstates of the momentum operator exist and are not normalizable. Defining

$$\psi_p(x) \;\equiv\; \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}}\,, \tag{6.38}$$

we readily confirm that

$$\hat{p}\psi_p(x) \;=\; \frac{\hbar}{i}\frac{\partial}{\partial x}\,\frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}} \;=\; p\frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}} \;=\; p\,\psi_p(x)\,. \tag{6.39}$$

So $\psi_p(x)$ is a momentum eigenstate with momentum eigenvalue p. It is a plane wave.

The so-called momentum representation is mathematically described by Fourier transforms. The Fourier transform $\tilde{\psi}(p)$ of $\psi(x)$ is defined by

$$\tilde{\psi}(p) \;\equiv\; \int_{-\infty}^{\infty} dx\,\frac{e^{-ipx/\hbar}}{\sqrt{2\pi\hbar}}\,\psi(x)\,. \tag{6.40}$$

The function $\tilde{\psi}(p)$ encodes the same amount of information as $\psi(x)$. We call $\tilde{\psi}(p)$ the momentum space representation of the state. Clearly for each value of p, $\tilde{\psi}$ is a linear superposition of values of $\psi(x)$ for all x. We can view the Fourier transformation as a linear transformation, the action of a matrix that depends on p on the vector that represents $\psi(x)$. The inverse Fourier transform is written as

$$\psi(x) \;=\; \int_{-\infty}^{\infty} dp\,\frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}}\,\tilde{\psi}(p)\,. \tag{6.41}$$

We can view this formula as an expansion of $\psi(x)$ in a basis of momentum eigenstates, with $\tilde{\psi}(p)$ the expansion coefficients.

We have seen that $\psi(x)$ and $\tilde{\psi}(p)$ are just two different representations of the same state:

$$\psi(x) \;\longleftrightarrow\; \tilde{\psi}(p)\,. \tag{6.42}$$

The arrow above is implemented by Fourier Transformation. Calculate now the action of $\frac{1}{i}\frac{d}{dx}$ on (6.41)

$$\frac{\hbar}{i}\frac{d}{dx}\,\psi(x) \;=\; \frac{\hbar}{i}\frac{d}{dx}\int_{-\infty}^{\infty} dp\,\frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}}\,\tilde{\psi}(p) \;=\; \int_{-\infty}^{\infty} dp\,\frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}}\,p\,\tilde{\psi}(p)\,. \tag{6.43}$$

In the language of (6.42) we write this as

$$\frac{\hbar}{i}\frac{d}{dx}\,\psi(x) \;\longleftrightarrow\; p\,\tilde{\psi}(p)\,. \tag{6.44}$$

We see that the momentum operator, viewed as the action of $\frac{\hbar}{i}\frac{d}{dx}$ in coordinate space, is simply multiplication by p on momentum space wavefunctions $\tilde{\psi}$:

$$\hat{p}\,\tilde{\psi}(p) \;=\; p\,\tilde{\psi}(p)\,. \tag{6.45}$$

This is, of course, perfectly analogous to the way that $\hat{x}$ acts on position space wavefunctions.

Exercise. Verify that acting on momentum space wavefunctions the $\hat{x}$ operator is represented by

$$\hat{x} \equiv i\hbar \frac{d}{dp}, \quad \text{(momentum representation)} \tag{6.46}$$

You can do this in two ways. Working with Fourier transforms, or by verifying (as in (6.37)) that it is consistent with $[\hat{x}, \hat{p}] = i\hbar$ acting on momentum space wavefunctions.

SPIN ONE-HALF, BRAS, KETS, AND OPERATORS

B. Zwiebach

September 17, 2013

Contents

1 The Stern-Gerlach Experiment

In 1922, at the University of Frankfurt (Germany), Otto Stern and Walther Gerlach, did fundamental experiments in which beams of silver atoms were sent through inhomogeneous magnetic fields to observe their deflection. These experiments demonstrated that these atoms have quantized magnetic moments that can take two values. Although consistent with the idea that the electron had spin, this suggestion took a few more years to develop.

Pauli introduced a "two-valued" degree of freedom for electrons, without suggesting a physical interpretation. Kronig suggested in 1925 that it this degree of freedom originated from the self-rotation of the electron. This idea was severely criticized by Pauli, and Kronig did not publish it. In the same year Uhlenbeck and Goudsmit had a similar idea, and Ehrenfest encouraged them to publish it. They are presently credited with the discovery that the electron has an intrinsic spin with value "one-half". Much of the mathematics of spin one-half was developed by Pauli himself in 1927. It took in fact until 1927 before it was realized that the Stern-Gerlach experiment did measure the magnetic moment of the electron.

A current on a closed loop induces a magnetic dipole moment. The magnetic moment vector $\vec{\mu}$ is proportional to the current I on the loop and the area A of the loop:

$$\vec{\mu} \;=\; I\vec{A}\,. \tag{1.1}$$

The vector area, for a planar loop is a vector normal to the loop and of length equal to the value of the area. The direction of the normal is determined from the direction of the current and the right-hand rule. The product μB of the magnetic moment times the magnetic field has units of energy, thus the units of μ are

$$[\mu] \;=\; \frac{\text{erg}}{\text{gauss}} \quad \text{or} \quad \frac{\text{Joule}}{\text{Tesla}} \tag{1.2}$$

When we have a change distribution spinning we get a magnetic moment and, if the distribution has mass, an angular momentum. The magnetic moment and the angular momentum are proportional to each other, and the constant of proportionality is universal. To see this consider rotating radius R ring of charge with uniform charge distribution and total charge Q. Assume the ring is rotating about an axis perpendicular to the plane of the ring and going through its center. Let the tangential velocity at the ring be v. The current at the loop is equal to the linear charge density λ times the velocity:

$$I \;=\; \lambda\, v \;=\; \frac{Q}{2\pi R} v\,. \tag{1.3}$$

It follows that the magnitude μ of the dipole moment of the loop is

$$\mu \;=\; IA \;=\; \frac{Q}{2\pi R}\, v\, \pi R^2 \;=\; \frac{Q}{2}\, Rv\,. \tag{1.4}$$

Let the mass of the ring be M. The magnitude L of the angular momentum of the ring is then $L = R(Mv)$. As a result

$$\mu \;=\; \frac{Q}{2M}\, RMv \;=\; \frac{Q}{2M}\, L\,, \tag{1.5}$$

leading to the notable ratio

$$\boxed{\frac{\mu}{L} \;=\; \frac{Q}{2M}\,.} \tag{1.6}$$

Note that the ratio does not depend on the radius of the ring, nor on its velocity. By superposition, any rotating distribution with uniform mass and charge density will have a ratio μ/L as above, with Q the total charge and M the total mass. The above is also written as

$$\mu \;=\; \frac{Q}{2M}\, L\,. \tag{1.7}$$

an classical electron going in a circular orbit around a nucleus will have both orbital angular momentum and a magnetic moment, related as above, with Q the electron charge and M the electron mass. In quantum mechanics the electron is not actually going in circles around the proton, but the right quantum analog exhibits both orbital angular momentum and magnetic moment.

We can ask if the electron can have an intrinsic μ, as if it were, a tiny spinning ball. Well, it has an intrinsic μ but it cannot really be viewed as a rotating little ball of charge (this was part of Pauli's objection to the original idea of spin). Moreover, we currently view the electron as an elementary particle with zero size, so the idea that it rotates is just not sensible. The classical relation, however, points to the correct result. Even if it has no size, the electron has an intrinsic spin S –intrinsic angular momentum. One could guess that

$$\mu \;=\; \frac{e}{2m_e}\, S\;? \tag{1.8}$$

Since angular momentum and spin have the same units we write this as

$$\mu \;=\; \frac{e\hbar}{2m_e}\, \frac{S}{\hbar}\;? \tag{1.9}$$

This is not exactly right. For electrons the magnetic moment is twice as large as the above relation suggests. One uses a constant "g-factor" to describe this

$$\mu \;=\; g\,\frac{e\hbar}{2m_e}\,\frac{S}{\hbar}\,, \quad g=2 \;\text{ for an electron.} \tag{1.10}$$

This factor of two is in fact predicted by the Dirac equation for the electron, and has been verified experimentally. To describe the above more briefly, one introduces the canonical value μ_B of the dipole moment called the Bohr-magneton:

$$\mu_B \;\equiv\; \frac{e\hbar}{2m_e} \;=\; 9.27\times 10^{-24}\frac{\mathrm{J}}{\mathrm{Tesla}}\,. \tag{1.11}$$

With this formula we get

$$\mu \;=\; g\,\mu_B\,\frac{S}{\hbar}\,, \quad g=2 \;\text{ for an electron.} \tag{1.12}$$

Both the magnetic moment and the angular momentum point in the same direction if the charge is positive. For the electron we thus get

$$\boxed{\vec{\mu} \;=\; -\,g\,\mu_B\,\frac{\vec{S}}{\hbar}\,, \quad g=2\,.} \tag{1.13}$$

Another feature of magnetic dipoles is needed for our discussion. A dipole placed in a non-uniform magnetic field will experience a force. An illustration is given in Figure 1 below, where to the left we show a current ring whose associated dipole moment $\vec{\mu}$ points upward. The magnetic field lines diverge as we move up, so the magnetic field is stronger as we move down. This dipole will experience a force pointing down, as can be deduced by basic considerations. On a small piece of wire the force $d\vec{F}$ is proportional to $\vec{I}\times\vec{B}$. The vectors $d\vec{F}$ are sketched in the right part of the figure. Their horizontal components cancel out, but the result is a net force downwards.

In general the equation for the force on a dipole $\vec{\mu}$ in a magnetic field $\vec{B}$ is given by

$$\vec{F} \;=\; \nabla(\vec{\mu}\cdot\vec{B})\,. \tag{1.14}$$

Note that the force points in the direction for which $\vec{\mu}\cdot\vec{B}$ increases the fastest. Given that in our situation $\vec{\mu}$ and $\vec{B}$ are parallel, this direction is the direction in which the magnitude of $\vec{B}$ increases the fastest.

The Stern-Gerlach experiment uses atoms of silver. Silver atoms have 47 electrons. Forty-six of them fill completely the $n=1,2,3$, and 4 levels. The last electron is an $n=5$ electron with zero orbital angular momentum (a $5s$ state). The only possible angular momentum is the intrinsic angular momentum of the last electron. Thus a magnetic dipole moment is also that of the last electron (the nucleus has much smaller dipole moment and can be ignored). The silver is vaporized in an oven and with a help of a collimating slit a narrow beam of silver atoms is send down to a magnet configuration.

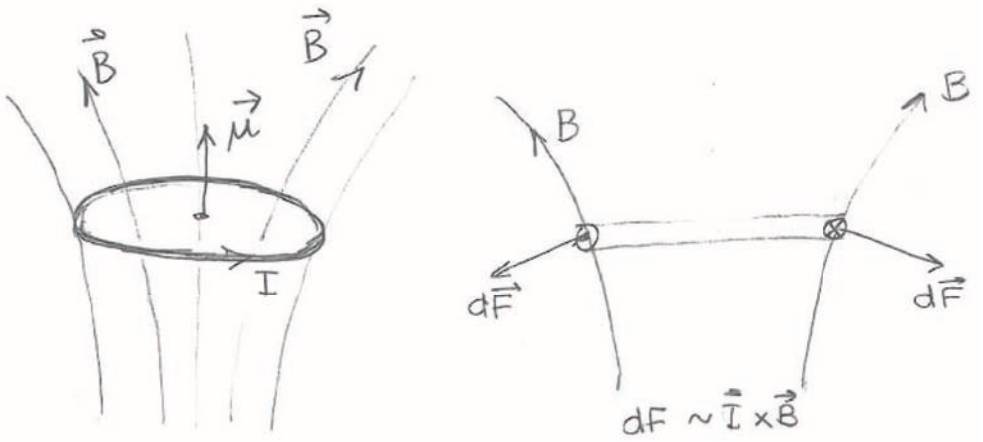


Figure 1: A magnetic dipole in a non-uniform magnetic field will experience a force. The force points in the direction for which $\vec{\mu} \cdot \vec{B}$ grows the fastest. In this case the force is downward.

In the situation described by Figure 2 the magnetic field points mostly in the positive z direction, and the gradient is also in the positive z-direction. As a result, the above equation gives

$$\vec{F} \simeq \nabla(\mu_z B_z) = \mu_z \nabla B_z \simeq \mu_z \frac{\partial B_z}{\partial z} \vec{e}_z \,, \tag{1.15}$$

and the atoms experience a force in the z-direction proportional to the z-component of their magnetic moment. Undeflected atoms would hit the detector screen at the point P. Atoms with positive μ_z should be deflected upwards and atoms with negative μ_z should be deflected downwards.

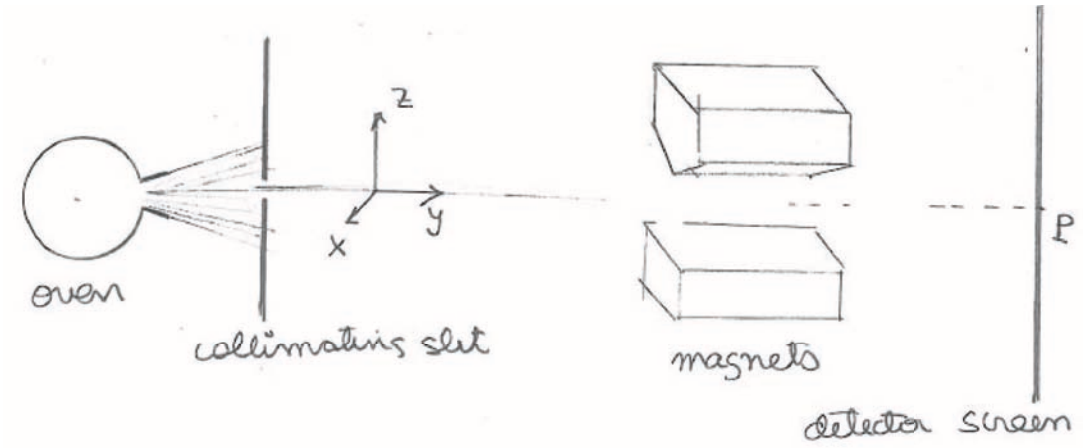


Figure 2: A sketch of the Stern-Gerlach apparatus. An oven and a collimating slit produces a narrow beam of silver atoms. The beam goes through a region with a strong magnetic field and a strong gradient, both in the z-direction. A screen, to the right, acts as a detector.

The oven source produces atoms with magnetic moments pointing in random directions and thus

the expectation was that the z-component of the magnetic moment would define a smooth probability distribution leading to a detection that would be roughly like the one indicated on the left side of Figure 3. Surprisingly, the observed result was two separate peaks as if all atoms had either a fixed positive μ_z or a fixed negative μ_z. This is shown on the right side of the figure. The fact that the peaks are spatially separated led to the original cumbersome name of "space quantization." The Stern Gerlach experiment demonstrates the quantization of the dipole moment, and by theoretical inference from (1.13), the quantization of the spin (or intrinsic) angular momentum.

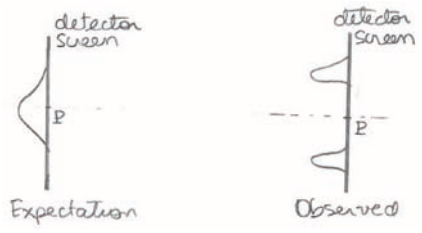


Figure 3: Left: the pattern on the detector screen that would be expected from classical physics. Right: the observed pattern, showing two separated peaks corresponding to up and down magnetic moments.

It follows from (1.13) that

$$\mu_z \;=\; -\,2\,\mu_B\,\frac{S_z}{\hbar}\,. \tag{1.16}$$

The deflections calculated using the details of the magnetic field configuration are consistent with

$$S_z \;=\; \pm\frac{\hbar}{2}\,, \quad \text{or} \quad \frac{S_z}{\hbar} \;=\; \pm\frac{1}{2}\,. \tag{1.17}$$

A particle with such possible values of $S_z/\hbar$ is called a spin one-half particle. The magnitude of the magnetic moments is one Bohr magneton.

With the magnetic field and its gradient along the z-direction, the Stern-Gerlach apparatus measures the component of the spin $\vec{S}$ in the z direction. To streamline our pictures we will denote such apparatus as a box with a $\hat{z}$ label, as in Figure 4. The box lets the input beam come in from the left and lets out two beams from the right side. If we placed a detector to the right, the top beam would be identified as having atoms with $S_z = \hbar/2$ and the bottom having atoms with $S_z = -\hbar/2$.[1]

Let us now consider thought experiments in which we put a few SG apparatus in series. In the first configuration, shown at the top of Figure 5, the first box is a $\hat{z}$ SG machine, where we block the $S_z = -\hbar/2$ output beam and let only the $S_z = \hbar/2$ beam go into the next machine. This machine acts as a filter. The second SG apparatus is also a $\hat{z}$ machine. Since all ingoing particles have $S_z = \hbar/2$ the second machine lets those out the top output and nothing comes out in the bottom output. The quantum mechanical lesson here is that $S_z = \hbar/2$ states have no component or amplitude along $S_z = -\hbar/2$. These are said to be orthogonal states.

[1] In the quantum mechanical view of the experiment, a single atom can be in both beams, with different amplitudes. Only the act of measurement, which corresponds to the act of placing the detector screen, forces the atom to decide in which beam it is.

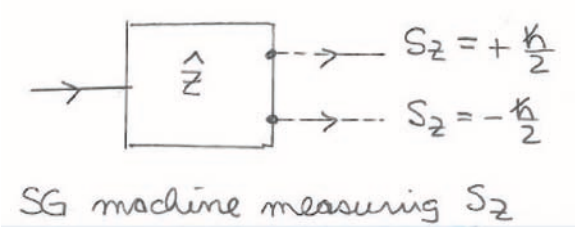


Figure 4: Left: A schematic representation of the SG apparatus, minus the screen.

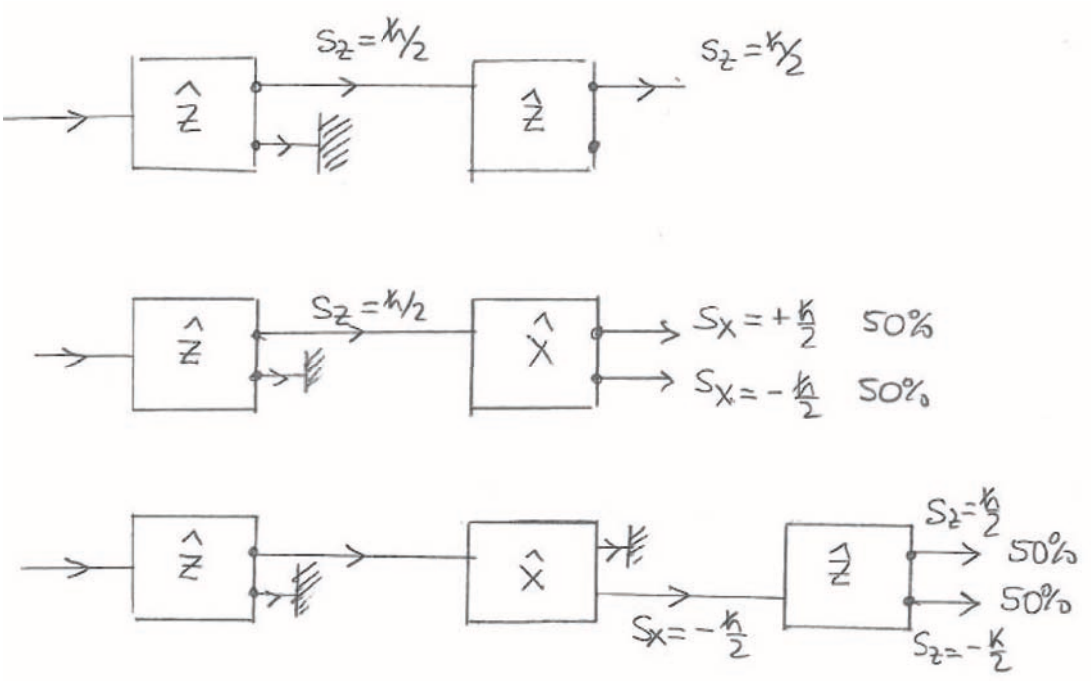


Figure 5: Left: Three configurations of SG boxes.

The second configuration in the figure shows the outgoing $S_z = \hbar/2$ beam from the first machine going into an $\hat{x}$-machine. The outputs of this machine are –in analogy to the $\hat{z}$ machine– $S_x = \hbar/2$ and $S_x = -\hbar/2$. Classically an object with angular momentum along the z axis has no component of angular momentum along the x axis, these are orthogonal directions. But the result of the experiment indicates that quantum mechanically this is not true for spins. About half of the $S_z = \hbar/2$ atoms exit through the top $S_x = \hbar/2$ output, and the other half exit through the bottom $S_x = -\hbar/2$ output. Quantum mechanically, a state with a definite value of S_z has an amplitude along the state $S_x = \hbar/2$ as well as an amplitude along the state $S_x = -\hbar/2$.

In the third and bottom configuration the $S_z = \hbar/2$ beam from the first machine goes into the $\hat{x}$ machine and the top output is blocked so that we only have an $S_x = -\hbar/2$ output. That beam is

fed into a $\hat{z}$ type machine. One could speculate that the beam entering the third machine has both $S_x = -\hbar/2$ *and* $S_z = \hbar/2$, as it is composed of silver atoms that made it through both machines. If that were the case the third machine would let all atoms out the top output. This speculation is falsified by the result. There is no memory of the first filter: the particles out of the second machine do not anymore have $S_z = \hbar/2$. We find half of the particles make it out of the third machine with $S_z = \hbar/2$ and the other half with $S_z = -\hbar/2$. In the following section we discuss a mathematical framework consistent with the the results of the above thought experiments.

2 Spin one-half states and operators

The SG experiment suggests that the spin states of the electron can be described using two basis vectors (or kets):

$$|z;+\rangle \quad \text{and} \quad |z;-\rangle\,. \tag{2.1}$$

The first corresponds to an electron with $S_z = \frac{\hbar}{2}$. The z label indicates the component of the spin, and the $+$ the fact that the component of spin is positive. This state is also called 'spin up' along z. The second state corresponds to an electron with $S_z = -\frac{\hbar}{2}$, that is a 'spin down' along z. Mathematically, we have an operator $\hat{S}_z$ for which the above states are eigenstates, with opposite eigenvalues:

$$\begin{aligned} \hat{S}_z|z;+\rangle &= +\frac{\hbar}{2}\,|z;+\rangle \\ \hat{S}_z|z;-\rangle &= -\frac{\hbar}{2}\,|z;-\rangle\,. \end{aligned} \tag{2.2}$$

If we have two basis states, then the state space of electron spin is a two-dimensional *complex vector space*. Each vector in this vector space represents a possible state of the electron spin. We are not discussing other degrees of freedom of the electron, such as its position, momentum, or energy. The general vector in the two-dimensional space is an arbitrary linear combination of the basis states and thus takes the form

$$|\Psi\rangle \;=\; c_1|z;+\rangle \;+\; c_2|z;-\rangle\,, \quad \text{with} \;\; c_1, c_2 \in \mathbb{C} \tag{2.3}$$

It is customary to call the state $|z;+\rangle$ the *first* basis state and it denote by $|1\rangle$. The state $|z;-\rangle$ is called the *second* basis state and is denoted by $|2\rangle$. States are vectors in some vector space. In a two-dimensional vector space a vector is explicitly *represented* as a column vector with two components. The first basis vector is represented as $\begin{pmatrix}1\\0\end{pmatrix}$ and the second basis vector is represented as $\begin{pmatrix}0\\1\end{pmatrix}$. Thus we have the following names for states and their concrete representation as column vectors

$$\begin{aligned} |z:+\rangle &= |1\rangle \;\longleftrightarrow\; \begin{pmatrix}1\\0\end{pmatrix}\,, \\ |z:-\rangle &= |2\rangle \;\longleftrightarrow\; \begin{pmatrix}0\\1\end{pmatrix}\,. \end{aligned} \tag{2.4}$$

Using these options the state in (2.3) takes the possible forms

$$|\Psi\rangle \;=\; c_1|z;+\rangle \;+\; c_2|z;-\rangle \;=\; c_1|1\rangle + c_2|2\rangle \;\longleftrightarrow\; c_1\begin{pmatrix}1\\0\end{pmatrix} + c_2\begin{pmatrix}0\\1\end{pmatrix} \;=\; \begin{pmatrix}c_1\\c_2\end{pmatrix}. \tag{2.5}$$

As we mentioned before, the top experiment in Figure 5 suggests that we have an orthonormal basis. The state $|z;+\rangle$ entering the second machine must have zero overlap with $|z,-\rangle$ since no such down spins emerge. Moreover the overlap of $|z;+\rangle$ with itself must be one, as all states emerge from the second machine top output. We thus write

$$\langle z;-|z;+\rangle \;=\; 0\,, \quad \langle z;+|z;+\rangle \;=\; 1\,. \tag{2.6}$$

and similarly, we expect

$$\langle z;+|z;-\rangle \;=\; 0\,, \quad \langle z;-|z;-\rangle \;=\; 1\,. \tag{2.7}$$

Using the notation where the basis states are labeled as $|1\rangle$ and $|2\rangle$ we have the simpler form that summarizes the four equations above:

$$\langle i|j\rangle \;=\; \delta_{ij}\,, \quad i,j=1,2. \tag{2.8}$$

We have not yet made precise what we mean by the 'bras' so let us do so briefly. We define the basis 'bras' as the *row vectors* obtained by transposition and complex conjugation:

$$\langle 1| \longleftrightarrow\; (1,0)\,, \quad \langle 2| \longleftrightarrow\; (0,1)\,. \tag{2.9}$$

Given states $|\alpha\rangle$ and $|\beta\rangle$

$$\begin{aligned} |\alpha\rangle \;&=\; \alpha_1|1\rangle + \alpha_2|2\rangle \;\longleftrightarrow\; \begin{pmatrix}\alpha_1\\\alpha_2\end{pmatrix} \\ |\beta\rangle \;&=\; \beta_1|1\rangle + \beta_2|2\rangle \;\longleftrightarrow\; \begin{pmatrix}\beta_1\\\beta_2\end{pmatrix} \end{aligned} \tag{2.10}$$

we associate

$$\langle\alpha| \equiv \alpha_1^*\langle 1| + \alpha_2^*\langle 2| \;\longleftrightarrow\; (\alpha_1^*\,,\alpha_2^*) \tag{2.11}$$

and the 'bra-ket' inner product is defined as the 'obvious' matrix product of the row vector and column vector representatives:

$$\langle\alpha|\beta\rangle \;\equiv\; (\alpha_1^*\,,\alpha_2^*)\cdot\begin{pmatrix}\beta_1\\\beta_2\end{pmatrix} \;=\; \alpha_1^*\beta_1 + \alpha_2^*\beta_2\,. \tag{2.12}$$

Note that this definition is consistent with (2.8).

When we represent the states as two-component column vectors the operators that act on the states to give new states can be *represented* as two-by-two matrices. We can thus represent the operator $\hat{S}_z$ as a 2×2 matrix which we claim takes the form

$$\boxed{\hat{S}_z \;=\; \frac{\hbar}{2}\begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}.} \tag{2.13}$$

To test this, it suffices to verify that the matrix $\hat{S}_z$ acts on the column vectors that represent the basis states as expected from (2.2). Indeed,

$$\begin{aligned} \hat{S}_z|z;+\rangle &= +\frac{\hbar}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix} = +\frac{\hbar}{2}\begin{pmatrix}1\\0\end{pmatrix} = +\frac{\hbar}{2}|z;+\rangle \\ \hat{S}_z|z;-\rangle &= +\frac{\hbar}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix} = -\frac{\hbar}{2}\begin{pmatrix}0\\1\end{pmatrix} = -\frac{\hbar}{2}|z;-\rangle\,. \end{aligned} \tag{2.14}$$

In fact, the states $|1\rangle$ and $|2\rangle$, viewed as column vectors are the eigenstates of matrix $\hat{S}_z$.

There is nothing particular about the z axis. We could have started with a SG apparatus that measures spin along the x axis and we would have been led to an operator $\hat{S}_x$. Had we used the y axis we would have been led to the operator $\hat{S}_y$. Since spin represents angular momentum (albeit of intrinsic type), it is expected to have three components, just like orbital angular momentum has three components: $\hat{L}_x, \hat{L}_y$, and $\hat{L}_z$. These are all hermitian operators, written as products of coordinates and momenta in three-dimensional space. Writing $\hat{L}_x = \hat{L}_1, \hat{L}_y = \hat{L}_2$, and $\hat{L}_z = \hat{L}_3$, their commutation relations can be briefly stated as

$$\left[\hat{L}_i\,,\hat{L}_j\right] = i\hbar\epsilon_{ijk}\,\hat{L}_k \tag{2.15}$$

This is the famous algebra of angular momentum, repeated indices are summed over the values 1,2,3, and ϵ_{ijk} is the totally antisymmetric symbol with $\epsilon_{123} = +1$. Make sure that you understand this notation clearly, and can use it to see that it implies the relations

$$\begin{aligned} [\hat{L}_x\,,\hat{L}_y] &= i\hbar\hat{L}_z\,, \\ [\hat{L}_y\,,\hat{L}_z] &= i\hbar\hat{L}_x\,, \\ [\hat{L}_z\,,\hat{L}_x] &= i\hbar\hat{L}_y\,. \end{aligned} \tag{2.16}$$

While, for example, $\hat{L}_z = \hat{x}\hat{p}_y - \hat{y}\hat{p}_x$ is a hermitian operator written in terms of coordinates and momenta, we have no such construction for $\hat{S}_z$. The latter is a more abstract operator, it does not act on wavefunctions $\psi(\vec{x})$ but rather on the 2-component column vectors introduced above. The operator $\hat{S}_z$ is just a two-by-two *hermitian*[2] matrix with constant entries! If spin is a quantum mechanical angular momentum, we must have that the triplet of operators $\hat{S}_z, \hat{S}_x$, and $\hat{S}_y$ satisfy

$$\begin{aligned} [\hat{S}_x\,,\hat{S}_y] &= i\hbar\hat{S}_z\,, \\ [\hat{S}_y\,,\hat{S}_z] &= i\hbar\hat{S}_x\,, \\ [\hat{S}_z\,,\hat{S}_x] &= i\hbar\hat{S}_y\,, \end{aligned} \tag{2.17}$$

or, again using numerical subscripts for the components ($\hat{S}_1 = \hat{S}_x, \cdots$) we must have

$$\left[\hat{S}_i\,,\hat{S}_j\right] = i\hbar\epsilon_{ijk}\,\hat{S}_k\,. \tag{2.18}$$

[2]Hermitian means that the matrix is preserved by taking the operations of transposition and complex conjugation.

We can now try to figure out how the matrices for $\hat{S}_x$ and $\hat{S}_y$ must look, given that we know the matrix for $\hat{S}_z$. We have a few constraints. First the matrices must be hermitian, just like the angular momentum operators are. Two-by-two hermitian matrices take the form

$$\begin{pmatrix} 2c & a-ib \\ a+ib & 2d \end{pmatrix}, \text{ with } a,b,c,d \in \mathbb{R} \tag{2.19}$$

Indeed, you can easily see that transposing and complex conjugating gives exactly the same matrix. Since the two-by-two identity matrix commutes with every matrix, we can subtract from the above matrix any multiple of the identity without any loss of generality. Subtracting the identity matrix times $(c+d)$ we find the still hermitian matrix

$$\begin{pmatrix} c-d & a-ib \\ a+ib & d-c \end{pmatrix}, \text{ with } a,b,c,d \in \mathbb{R} \tag{2.20}$$

Since we are on the lookout for $\hat{S}_x$ and $\hat{S}_y$ we can subtract a matrix proportional to $\hat{S}_z$. Since $\hat{S}_z$ is diagonal with entries of same value but opposite signs, we can cancel the diagonal terms above and are left over with

$$\begin{pmatrix} 0 & a-ib \\ a+ib & 0 \end{pmatrix}, \text{ with } a,b \in \mathbb{R} \tag{2.21}$$

Thinking of the space of two-by-two hermitian matrices as a real vector space, the hermitian matrices given above can be associated to two basis "vectors" that are the matrices

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{2.22}$$

since multiplying the first by the real constant a and the second by the real constant b and adding gives us the matrix above. In fact, together with the identity matrix and the $\hat{S}_z$ matrix, with the $\hbar/2$ deleted,

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{2.23}$$

we got the complete set of four two-by-two matrices that viewed as basis vectors in a real vector space, can be used to build the most general hermitian two-by-two matrix by using real linear combinations.

Back to our problem, we are supposed to find $\hat{S}_x$ and $\hat{S}_y$ among the matrices in (2.22). The overall scale of the matrices can be fixed by the constraint that their eigenvalues be $\pm\hbar/2$, just like they are for $\hat{S}_z$. Let us give the eigenvalues (denoted by λ) and the associated normalized eigenvectors for these two matrices. Short computations (can you do them?) give

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}: \qquad \lambda = 1, \text{ for } \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \lambda = -1, \text{ for } \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \tag{2.24}$$

for the first matrix and

$$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}: \qquad \lambda = 1, \text{ for } \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}, \quad \lambda = -1, \text{ for } \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}, \tag{2.25}$$

for the second matrix. In case you are puzzled by the normalizations, note that a vector $\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ is normalized if $|c_1|^2 + |c_2|^2 = 1$. Since the eigenvalues of both matrices are ± 1, we tentatively identify

$$\hat{S}_x \;=\; \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \hat{S}_y \;=\; \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \tag{2.26}$$

which have, at least, the correct eigenvalues. But in fact, these also satisfy the commutation relations! Indeed, we check that, as desired,

$$\begin{aligned} [\hat{S}_x\,,\hat{S}_y] \;&=\; \frac{\hbar^2}{4}\Big(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \;-\; \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\Big) \\ &=\; \frac{\hbar^2}{4}\Big(\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \;-\; \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}\Big) \\ &=\; \frac{\hbar^2}{4}\begin{pmatrix} 2i & 0 \\ 0 & -2i \end{pmatrix} \;=\; i\hbar\frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \;=\; i\hbar\hat{S}_z\,. \end{aligned} \tag{2.27}$$

All in all we have

$$\boxed{\hat{S}_x \;=\; \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \hat{S}_y \;=\; \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \hat{S}_z \;=\; \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.} \tag{2.28}$$

Exercise. Verify that the above matrices satisfy the other two commutation relations in (2.17).

You could ask if we got the unique solution for $\hat{S}_x$ and $\hat{S}_y$ given the choice of $\hat{S}_z$? The answer is no, but it is not our fault, as illustrated by the following check:

Exercise. Check that the set of commutation relations of the spin operators are in fact preserved when we replace $\hat{S}_x \to -\hat{S}_y$ and $\hat{S}_y \to \hat{S}_x$.

The solution we gave is the one conventionally used by all physicists. Any other solution is physically equivalent to the one we gave (as will be explained in more detail after we develop more results). The solution defines the **Pauli matrices** σ_i by writing

$$\hat{S}_i \;=\; \frac{\hbar}{2}\,\sigma_i\,. \tag{2.29}$$

We then have that the Pauli matrices are

$$\boxed{\sigma_1 \;=\; \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 \;=\; \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 \;=\; \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.} \tag{2.30}$$

Let us describe the eigenstates of $\hat{S}_x$, which given (2.26) can be read from (2.24):

$$\hat{S}_x\,|x;\pm\rangle \;=\; \pm\,|x;\pm\rangle\,. \tag{2.31}$$

with

$$\boxed{\begin{aligned} |x;+\rangle \;&=\; \frac{1}{\sqrt{2}}|z;+\rangle + \frac{1}{\sqrt{2}}|z;-\rangle \;\longleftrightarrow\; \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \\ |x;-\rangle \;&=\; \frac{1}{\sqrt{2}}|z;+\rangle - \frac{1}{\sqrt{2}}|z;-\rangle \;\longleftrightarrow\; \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \end{aligned}} \tag{2.32}$$

Note that these states are orthogonal to each other. The above equations can be inverted to find

$$\begin{aligned} |z;+\rangle &= \frac{1}{\sqrt{2}}|x;+\rangle + \frac{1}{\sqrt{2}}|x;-\rangle \\ |z;-\rangle &= \frac{1}{\sqrt{2}}|x;+\rangle - \frac{1}{\sqrt{2}}|x;-\rangle \end{aligned} \tag{2.33}$$

These relations are consistent with the second experiment shown in Figure 5. The state $|z;+\rangle$ entering the second, $\hat{x}$-type SG apparatus, has equal probability to be found in $|x;+\rangle$ as it has probability to be found in $|x;-\rangle$. This is reflected in the first of the above relations, since we have the amplitudes

$$\langle x;+|z;+\rangle = \frac{1}{\sqrt{2}}\,, \quad \langle x;-|z;+\rangle = \frac{1}{\sqrt{2}}\,. \tag{2.34}$$

These probabilities, being equal to the norm squared of the amplitudes, are 1/2 in both cases. The relative minus sign on the second equation above is needed to make it orthogonal to the state on the first equation.

We can finally consider the eigenstates of $\hat{S}_y$ axis. We have

$$\hat{S}_y\,|y;\pm\rangle = \pm\frac{\hbar}{2}\,|y;\pm\rangle\,. \tag{2.35}$$

and using (2.25) we read

$$\boxed{\begin{aligned} |y;+\rangle &= \frac{1}{\sqrt{2}}|z;+\rangle + \frac{i}{\sqrt{2}}|z;-\rangle \;\longleftrightarrow\; \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ i\end{pmatrix}, \\ |y;-\rangle &= \frac{1}{\sqrt{2}}|z;+\rangle - \frac{i}{\sqrt{2}}|z;-\rangle \;\longleftrightarrow\; \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ -i\end{pmatrix}. \end{aligned}} \tag{2.36}$$

Note that this time the superposition of $|z;\pm\rangle$ states involves complex numbers (there would be no way to find y type states without them).

3 Properties of Pauli matrices and index notation

Since we know the commutation relations for the spin operators

$$\big[\hat{S}_i\,,\,\hat{S}_j\big] = i\hbar\,\epsilon_{ijk}\hat{S}_k\,, \tag{3.37}$$

and we have $S_i = \frac{\hbar}{2}\sigma_i$, it follows that

$$\frac{\hbar}{2}\frac{\hbar}{2}\big[\sigma_i\,,\,\sigma_j\big] = i\hbar\,\epsilon_{ijk}\frac{\hbar}{2}\sigma_k\,. \tag{3.38}$$

Cancelling the $\hbar$'s and some factors of two, we find

$$\boxed{[\sigma_i,\sigma_j] \;=\; 2i\,\epsilon_{ijk}\sigma_k\,.} \tag{3.39}$$

Another important property of the Pauli matrices is that they square to the identity matrix. This is best checked explicitly (do it!):

$$\boxed{(\sigma_1)^2 \;=\; (\sigma_2)^2 \;=\; (\sigma_3)^2 \;=\; \mathbf{1}\,.} \tag{3.40}$$

This property "explains" that the eigenvalues of each of the Pauli matrices could only be plus or minus one. Indeed, the eigenvalues of a matrix satisfy the algebraic equation that the matrix satisfies. Take for example a matrix M that satisfies the matrix equation

$$M^2 + \alpha M + \beta \mathbf{1} \;=\; 0 \tag{3.41}$$

Let v be an eigenvector of M with eigenvalue λ: $Mv = \lambda v$. Let the above equation act on v

$$M^2 v + \alpha M v + \beta \mathbf{1} v \;=\; 0 \quad \rightarrow \quad \lambda^2 v + \alpha\lambda v + \beta v = 0 \quad \rightarrow \quad (\lambda^2 + \alpha\lambda + \beta)v = 0\,, \tag{3.42}$$

and since $v \neq 0$ (by definition an eigenvector cannot be zero!) we conclude that $\lambda^2 + \alpha\lambda + \beta = 0$, as claimed. For the case of the Pauli matrices we have $(\sigma_i)^2 = 1$ and therefore the eigenvalues must satisfy $\lambda^2 = 1$. As a result, $\lambda = \pm 1$ are the only options.

We also note, by inspection, that the Pauli matrices have zero trace, namely, the sum of entries on the diagonal is zero:

$$\boxed{\mathrm{tr}(\sigma_i) \;=\; 0\,, \qquad i = 1, 2, 3.} \tag{3.43}$$

A fact from linear algebra is that the trace of a matrix is equal to the sum of its eigenvalues. So each Pauli matrix must have two eigenvalues that add up to zero. Since the eigenvalues can only be plus or minus one, we must have one of each. This shows that each of the Pauli matrices has a plus one and a minus one eigenvalue.

If you compute a commutator of Pauli matrices by hand you might notice a curious property. Take the commutator of σ_1 and σ_2:

$$[\sigma_1\,, \sigma_2\,] \;=\; \sigma_1\sigma_2 - \sigma_2\sigma_1\,. \tag{3.44}$$

The two contributions on the right hand side give

$$\begin{aligned} \sigma_1\sigma_2 \;&=\; \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \;=\; \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}\,, \\ \sigma_2\sigma_1 \;&=\; \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \;=\; \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}\,. \end{aligned} \tag{3.45}$$

The second contribution is minus the first, so that both terms contribute equally to the commutator! In other words,

$$\sigma_1\sigma_2 \;=\; -\,\sigma_2\sigma_1\,. \tag{3.46}$$

This equation is taken to mean that σ_1 and σ_2 *anticommute*. Just like we define the commutator of two operators X, Y by $[X, Y] \equiv XY - YX$, we define the **anticommutator**, denoted by curly brackets, by

$$\boxed{\text{Anticommutator:}\;\; \{X\,, Y\} \;\equiv\; XY + YX\,.} \tag{3.47}$$

In this language we have checked that

$$\{\sigma_1\,,\sigma_2\} \;=\; 0\,, \tag{3.48}$$

and the property $\sigma_1^2 = \mathbf{1}$, for example, can be rewritten as

$$\{\sigma_1\,,\sigma_1\} \;=\; 2\cdot\mathbf{1}\,. \tag{3.49}$$

In fact, as you can check (two cases to examine) that any two different Pauli matrices anticommute:

$$\{\sigma_i\,,\sigma_j\} \;=\; 0\,, \quad \text{for} \;\; i\neq j\,. \tag{3.50}$$

We can easily improve on this equation to make it work also when i is equal to j. We claim that

$$\boxed{\{\sigma_i\,,\sigma_j\} \;=\; 2\delta_{ij}\,\mathbf{1}\,.} \tag{3.51}$$

Indeed, when $i\neq j$ the right-hand side vanishes, as needed, and when i is equal to j, the right-hand side gives $2\cdot\mathbf{1}$, also as needed in view of (3.49) and its analogs for the other Pauli matrices.

Both the commutator and anti-commutator identities for the Pauli matrices can be summarized in a single equation. This is possible because, for any two operators X, Y we have

$$X\,Y \;=\; \tfrac{1}{2}\,\{X,Y\} \;+\; \tfrac{1}{2}\,[X,Y]\,, \tag{3.52}$$

as you should confirm by expansion. Applied to the product of two Pauli matrices and using our expressions for the commutator and anticommutator we get

$$\boxed{\sigma_i\sigma_j \;=\; \delta_{ij}\,\mathbf{1} \;+\; i\,\epsilon_{ijk}\,\sigma_k\,.} \tag{3.53}$$

This equation can be recast in vector notation. Denote by bold symbols three-component vectors, for example, $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$. Then the dot product

$$\mathbf{a}\cdot\mathbf{b} \;=\; a_1b_1 + a_2b_2 + a_3b_3 = a_ib_i \;=\; a_ib_j\,\delta_{ij}\,. \tag{3.54}$$

Note the use of the sum convention: repeated indices are summed over. Moreover, note that $b_j\delta_{ij} = b_i$ (can you see why?). We also have that

$$\mathbf{a}\cdot\mathbf{a} \;=\; |\mathbf{a}|^2\,. \tag{3.55}$$

Cross products use the epsilon symbol. Make sure you understand why

$$(\mathbf{a}\times\mathbf{b})_k \;=\; a_ib_j\,\epsilon_{ijk}\,. \tag{3.56}$$

We can also have triplets of operators, or matrices. For the Pauli matrices we denote

$$\boldsymbol{\sigma} \;\equiv\; (\sigma_1,\sigma_2,\sigma_3)\,. \tag{3.57}$$

We can construct a matrix by dot product of a vector $\mathbf{a}$ with the 'vector' $\boldsymbol{\sigma}$. We define

$$\mathbf{a}\cdot\boldsymbol{\sigma} \;\equiv\; a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3 \;=\; a_i\sigma_i\,. \tag{3.58}$$

Note that $\mathbf{a}\cdot\boldsymbol{\sigma}$ is just a single two-by-two matrix. Since the components of $\mathbf{a}$ are numbers, and numbers commute with matrices, this dot product is commutative: $\mathbf{a}\cdot\boldsymbol{\sigma} = \boldsymbol{\sigma}\cdot\mathbf{a}$. We are now ready to rewrite (3.53). Multiply this equation by $a_i b_j$ to get

$$\begin{aligned} a_i\sigma_i \; b_j\sigma_j &= a_i b_j \delta_{ij}\,\mathbf{1} \;+\; i\,(a_i b_j \epsilon_{ijk})\,\sigma_k \\ &= (\mathbf{a}\cdot\mathbf{b})\,\mathbf{1} \;+\; i\,(\mathbf{a}\times\mathbf{b})_k\,\sigma_k\,, \end{aligned} \tag{3.59}$$

so that, finally, we get the matrix equation

$$\boxed{(\mathbf{a}\cdot\boldsymbol{\sigma})(\mathbf{b}\cdot\boldsymbol{\sigma}) = (\mathbf{a}\cdot\mathbf{b})\,\mathbf{1} + i\,(\mathbf{a}\times\mathbf{b})\cdot\boldsymbol{\sigma}\,.} \tag{3.60}$$

As a simple application we take $\mathbf{b} = \mathbf{a}$. We then have $\mathbf{a}\cdot\mathbf{a} = |\mathbf{a}|^2$ and $\mathbf{a}\times\mathbf{a} = 0$, so that the above equation gives

$$(\mathbf{a}\cdot\boldsymbol{\sigma})^2 = |\mathbf{a}|^2\,\mathbf{1}\,. \tag{3.61}$$

When $\mathbf{a}$ is a unit vector this becomes

$$(\mathbf{n}\cdot\boldsymbol{\sigma})^2 = \mathbf{1}\,, \quad \mathbf{n} \text{ a unit vector.} \tag{3.62}$$

The epsilon symbol satisfies useful identities. One can show that the product of two epsilons with one index contracted is a sum of products of Kronecker deltas:

$$\epsilon_{ijk}\,\epsilon_{ipq} \;=\; \delta_{jp}\delta_{kq} \;-\; \delta_{jq}\delta_{kp}\,. \tag{3.63}$$

Its contraction (setting $p = j$) is also useful:

$$\epsilon_{ijk}\,\epsilon_{ijq} \;=\; 2\delta_{kq}\,. \tag{3.64}$$

The first of these two allows one to prove the familiar vector identity

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) \;=\; \mathbf{b}\,(\mathbf{a}\cdot\mathbf{c}) \;-\; (\mathbf{a}\cdot\mathbf{b})\,\mathbf{c}\,. \tag{3.65}$$

It will be useful later on to consider the dot and cross products of *operator* triplets. Given the operators $\mathbf{X} = (\hat{X}_1, \hat{X}_2, \hat{X}_3)$ and $\mathbf{Y} = (\hat{Y}_1, \hat{Y}_2, \hat{Y}_3)$ we define

$$\begin{aligned} \mathbf{X}\cdot\mathbf{Y} &\equiv \hat{X}_i\,\hat{Y}_i\,, \\ (\mathbf{X}\times\mathbf{Y})_i &\equiv \epsilon_{ijk}\,\hat{X}_j\,\hat{Y}_k\,. \end{aligned} \tag{3.66}$$

In these definitions the order of the operators on the right hand side is as in the left-hand side. This is important to keep track of, since the $\hat{X}_i$ and $\hat{Y}_j$ operators may not commute. The dot product of two operator triplets is not necessarily commutative, nor is the cross product necessarily antisymmetric.

4 Spin states in arbitrary direction

We consider here the description and analysis of spin states that point in arbitrary directions, as specified by a unit vector $\mathbf{n}$:

$$\mathbf{n} \;=\; (n_x, n_y, n_z) \;=\; (\sin\theta\cos\phi,\, \sin\theta\sin\phi,\, \cos\theta)\;. \tag{4.67}$$

Here θ and ϕ are the familiar polar and azimuthal angles. We view the spatial vector $\mathbf{n}$ as a triplet of numbers. Just like we did for $\boldsymbol{\sigma}$, we can define $\mathbf{S}$ as the triplet of operators

$$\mathbf{S} \;=\; (\hat{S}_x\,, \hat{S}_y\,, \hat{S}_z)\,. \tag{4.68}$$

Note that, in fact,

$$\mathbf{S} \;=\; \frac{\hbar}{2}\,\boldsymbol{\sigma}\,. \tag{4.69}$$

We can use $\mathbf{S}$ to obtain, by a *dot* product with $\mathbf{n}$ a spin operator $\hat{S}_{\mathbf{n}}$ that has a simple interpretation:

$$\hat{S}_{\mathbf{n}} \;\equiv\; \mathbf{n}\cdot\mathbf{S} \;\equiv\; n_x\hat{S}_x + n_y\hat{S}_y + n_z\hat{S}_z \;=\; \frac{\hbar}{2}\,\mathbf{n}\cdot\boldsymbol{\sigma}\,. \tag{4.70}$$

Note that $\hat{S}_{\mathbf{n}}$ is just an operator, or a hermitian matrix. We view $\hat{S}_{\mathbf{n}}$ as the spin operator in the direction of the unit vector $\mathbf{n}$. To convince you that this makes sense note that, for example, when $\mathbf{n}$ points along z, we have $(n_x, n_y, n_z) = (0, 0, 1)$ and $\hat{S}_{\mathbf{n}}$ becomes $\hat{S}_z$. The same holds, of course, for the x and y directions. Moreover, just like all the $\hat{S}_i$, the eigenvalues of $\hat{S}_{\mathbf{n}}$ are $\pm\hbar/2$. This is needed physically, since all directions are physically equivalent and those two values for spin must be the only allowed values for all directions. To see that this is true we first compute the square of the matrix $\hat{S}_{\mathbf{n}}$:

$$(\hat{S}_{\mathbf{n}})^2 \;=\; \left(\frac{\hbar}{2}\right)^2(\mathbf{n}\cdot\boldsymbol{\sigma})^2 \;=\; \left(\frac{\hbar}{2}\right)^2, \tag{4.71}$$

using (3.62). Moreover, since the Pauli matrices are traceless so is $\hat{S}_{\mathbf{n}}$:

$$\mathrm{tr}(\hat{S}_{\mathbf{n}}) \;= n_i\,\mathrm{tr}(\hat{S}_i) \;=\; n_i\,\frac{\hbar}{2}\,\mathrm{tr}(\sigma_i) \;=\; 0\,. \tag{4.72}$$

By the same argument we used for Pauli matrices, we conclude that the eigenvalues of $\hat{S}_{\mathbf{n}}$ are indeed $\pm\hbar/2$. For an arbitrary direction we can write the matrix $\hat{S}_{\mathbf{n}}$ explicitly:

$$\begin{aligned}
\hat{S}_{\mathbf{n}} \;&=\; \frac{\hbar}{2}\Big[n_x\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} + n_y\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix} + n_z\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\Big] \\
&=\; \frac{\hbar}{2}\begin{pmatrix} n_z & n_x - in_y \\ n_x + in_y & -n_z\end{pmatrix} \\
&=\; \frac{\hbar}{2}\begin{pmatrix} \cos\theta & \sin\theta e^{-i\phi} \\ \sin\theta e^{i\phi} & -\cos\theta\end{pmatrix}.
\end{aligned} \tag{4.73}$$

Since the eigenvalues of $\hat{S}_{\mathbf{n}}$ are $\pm\hbar/2$ the associated spin eigenstates, denoted as $|\mathbf{n};\pm\rangle$, satisfy

$$\hat{S}_{\mathbf{n}}|\mathbf{n};\pm\rangle \;=\; \pm\frac{\hbar}{2}|\mathbf{n};\pm\rangle\,. \tag{4.74}$$

The states $|\mathbf{n};+\rangle$ and $|\mathbf{n};-\rangle$ represent, respectively, a spin state that points up along $\mathbf{n}$, and a spin state that points down along $\mathbf{n}$. We can also find the eigenvalues of the matrix $\hat{S}_\mathbf{n}$ by direct computation. The eigenvalues are the roots of the equation $\det(\hat{S}_\mathbf{n} - \lambda\mathbf{1}) = 0$:

$$\det\begin{pmatrix} \frac{\hbar}{2}\cos\theta - \lambda & \frac{\hbar}{2}\sin\theta e^{-i\phi} \\ \frac{\hbar}{2}\sin\theta e^{i\phi} & -\frac{\hbar}{2}\cos\theta - \lambda \end{pmatrix} = \lambda^2 - \frac{\hbar^2}{4}(\cos^2\theta + \sin^2\theta) \;=\; \lambda^2 - \frac{\hbar^2}{4} \;=\; 0\,. \tag{4.75}$$

The eigenvalues are thus $\lambda = \pm\hbar/2$, as claimed. To find the eigenvector v associated with the eigenvalue λ we must solve the linear equation $(\hat{S}_\mathbf{n} - \lambda\mathbf{1})v = 0$. We denote by $|\mathbf{n};+\rangle$ the eigenvector associated with the eigenvalue $\hbar/2$. For this eigenvector we write the ansatz

$$|\mathbf{n};+\rangle = c_1|+\rangle + c_2|-\rangle \;=\; \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}\,, \tag{4.76}$$

where for notational simplicity $|\pm\rangle$ refer to the states $|z;\pm\rangle$. The eigenvector equation becomes $(\hat{S}_\mathbf{n} - \frac{\hbar}{2}\mathbf{1})|\mathbf{n};+\rangle = 0$ and explicitly reads

$$\frac{\hbar}{2}\begin{pmatrix} \cos\theta - 1 & \sin\theta e^{-i\phi} \\ \sin\theta e^{i\phi} & -\cos\theta - 1 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = 0\,. \tag{4.77}$$

Either equation gives the same relation between c_1 and c_2. The top equation, for example gives

$$c_2 = e^{i\phi}\,\frac{1-\cos\theta}{\sin\theta}\,c_1 = e^{i\phi}\,\frac{\sin\frac{\theta}{2}}{\cos\frac{\theta}{2}}\,c_1\,. \tag{4.78}$$

(Check that the second equation gives the same relation.) We want normalized states, and therefore

$$|c_1|^2 + |c_2|^2 = 1 \quad\to\quad |c_1|^2\left[1 + \frac{\sin^2\frac{\theta}{2}}{\cos^2\frac{\theta}{2}}\right] = 1 \quad\to\quad |c_1|^2 = \cos^2\frac{\theta}{2}\,. \tag{4.79}$$

Since the overall phase of the eigenstate is not observable we take the simplest option for c_1:

$$c_1 \;=\; \cos\tfrac{\theta}{2}\,, \qquad c_2 = \sin\tfrac{\theta}{2}\exp(i\phi)\;, \tag{4.80}$$

that is

$$|\mathbf{n};+\rangle \;=\; \cos\tfrac{\theta}{2}|+\rangle \;+\; \sin\tfrac{\theta}{2}e^{i\phi}|-\rangle\,. \tag{4.81}$$

As a quick check we see that for $\theta = 0$, which corresponds to a unit vector $\mathbf{n} = \mathbf{e}_3$ along the plus z direction we get $|\mathbf{e}_3;+\rangle = |+\rangle$. Note that even though ϕ is ambiguous when $\theta = 0$, this does not affect our answer, since the term with ϕ dependence vanishes. In the same way one can obtain the normalized eigenstate corresponding to $-\hbar/2$. A simple phase choice gives

$$|\mathbf{n};-\rangle \;=\; \sin\tfrac{\theta}{2}|+\rangle \;-\; \cos\tfrac{\theta}{2}e^{i\phi}|-\rangle\,. \tag{4.82}$$

If we again consider the $\theta = 0$ direction, this time the ambiguity of ϕ remains in the term that contains the $|z;-\rangle$ state. It is convenient to multiply this state by the phase $-e^{-i\phi}$. Doing this, the pair of

eigenstates read[3]

$$\boxed{\begin{array}{rcl} |\mathbf{n};+\rangle &=& \cos\tfrac{\theta}{2}\,|+\rangle \;+\; \sin\tfrac{\theta}{2}e^{i\phi}|-\rangle\,, \\ |\mathbf{n};-\rangle &=& -\sin\tfrac{\theta}{2}e^{-i\phi}|+\rangle \;+\; \cos\tfrac{\theta}{2}|-\rangle\,. \end{array}} \tag{4.83}$$

The vectors are normalized. Furthermore, they are orthogonal

$$\langle \mathbf{n};-|\mathbf{n};+\rangle = -\sin\tfrac{\theta}{2}e^{i\phi}\cos\tfrac{\theta}{2} + \cos\tfrac{\theta}{2}\sin\tfrac{\theta}{2}e^{i\phi} = 0\,. \tag{4.84}$$

Therefore, $|\mathbf{n};+\rangle$ and $|\mathbf{n};-\rangle$ are an orthonormal pair of states.

Let us verify that the $|\mathbf{n};\pm\rangle$ reduce to the known results as $\mathbf{n}$ points along the z, x, and y axes. Again, if $\mathbf{n} = (0,0,1) = \mathbf{e}_3$, we have $\theta = 0$, and hence

$$|\mathbf{e}_3;+\rangle = |+\rangle\,, \quad |\mathbf{e}_3;-\rangle = |-\rangle\,, \tag{4.85}$$

which are, as expected, the familiar eigenstates of $\hat{S}_z$. If we point along the x axis, $\mathbf{n} = (1,0,0) = \mathbf{e}_1$ which corresponds to $\theta = \pi/2$, $\phi = 0$. Hence

$$|\mathbf{e}_1;+\rangle = \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle) \;=\; |x;+\rangle\,, \quad |\mathbf{e}_1;-\rangle = \frac{1}{\sqrt{2}}(-|+\rangle + |-\rangle) \;=\; -|x;-\rangle\,, \tag{4.86}$$

where we compared with (2.32). Note that the second state came out with an overall minus sign. Since overall phases (or signs) are physically irrelevant, this is the expected answer: we got the eigenvectors of $\hat{S}_x$. Finally, if $\mathbf{n} = (0,1,0) = \mathbf{e}_2$, we have $\theta = \pi/2$, $\phi = \pi/2$ and hence, with $e^{\pm i\phi} = \pm i$, we have

$$|\mathbf{e}_2;+\rangle = \frac{1}{\sqrt{2}}(|+\rangle + i|-\rangle) = |y;+\rangle\,, \quad |\mathbf{e}_2;-\rangle = \frac{1}{\sqrt{2}}(i|+\rangle + |-\rangle) = i\frac{1}{\sqrt{2}}(|+\rangle - i|-\rangle) = i|y;-\rangle \tag{4.87}$$

which are, up to a phase for the second one, the eigenvectors of $\hat{S}_y$.

[3]The formula (4.83) works nicely at the north pole ($\theta = 0$), but at the south pole ($\theta = \pi$) the ϕ ambiguity shows up again. If one works near the south pole multiplying the results in (4.83) by suitable phases will do the job. The fact that no formula works well unambiguously through the full the sphere is not an accident.

LINEAR ALGEBRA: VECTOR SPACES AND OPERATORS

B. Zwiebach

October 21, 2013

Contents

1 Vector spaces and dimensionality

In quantum mechanics the state of a physical system is a *vector* in a *complex* vector space. Observables are linear operators, in fact, Hermitian operators acting on this complex vector space. The purpose of this chapter is to learn the basics of vector spaces, the structures that can be built on those spaces, and the operators that act on them.

Complex vector spaces are somewhat different from the more familiar real vector spaces. I would say they have more powerful properties. In order to understand more generally complex vector spaces it is useful to compare them often to their real dimensional friends. We will follow here the discussion of the book *Linear algebra done right*, by Sheldon Axler.

In a vector space one has vectors and numbers. We can add vectors to get vectors and we can multiply vectors by numbers to get vectors. If the numbers we use are real, we have a real vector space. If the numbers we use are complex, we have a complex vector space. More generally, the numbers we use belong to what is called in mathematics a 'field' and denoted by the letter $\mathbb{F}$. We will discuss just two cases, $\mathbb{F} = \mathbb{R}$, meaning that the numbers are real, and $\mathbb{F} = \mathbb{C}$, meaning that the numbers are complex.

The definition of a vector space is the same for $\mathbb{F}$ being $\mathbb{R}$ or $\mathbb{C}$. A vector space V is a set of vectors with an operation of **addition** $(+)$ that assigns an element $u + v \in V$ to each $u, v \in V$. This means that V is closed under addition. There is also a **scalar multiplication** by elements of $\mathbb{F}$, with $av \in V$

for any $a \in \mathbb{F}$ and $v \in V$. This means the space V is closed under multiplication by numbers. These operations must satisfy the following additional properties:

1. $u + v = v + u \in V$ for all $u, v \in V$ (addition is commutative).
2. $u + (v + w) = (u + v) + w$ and $(ab)u = a(bu)$ for any $u, v, w \in V$ and $a, b \in \mathbb{F}$ (associativity).
3. There is a vector $0 \in V$ such that $0 + u = u$ for all $u \in V$ (additive identity).
4. For each $v \in V$ there is a $u \in V$ such that $v + u = 0$ (additive inverse).
5. The element $1 \in \mathbb{F}$ satisfies $1v = v$ for all $v \in V$ (multiplicative identity).
6. $a(u+v) = au+av$ and $(a+b)v = av+bv$ for every $u, v \in V$ and $a, b \in \mathbb{F}$ (distributive property).

This definition is very efficient. Several familiar properties follow from it by short proofs (which we will not give, but are not complicated and you may try to produce):

- The additive identity is unique: any vector $0'$ that acts like 0 is actually equal to 0.
- $0v = 0$, for any $v \in V$, where the first zero is a number and the second one is a vector. This means that the number zero acts as expected when multiplying a vector.
- $a0 = 0$, for any $a \in \mathbb{F}$. Here both zeroes are vectors. This means that the zero vector multiplied by any number is still the zero vector.
- The additive inverse of any vector $v \in V$ is unique. It is denoted by $-v$ and in fact $-v = (-1)v$.

We must emphasize that while the numbers, in $\mathbb{F}$ are sometimes real or complex, we never speak of the vectors themselves as real or complex. A vector multiplied by a complex number is not said to be a complex vector, for example! The vectors in a real vector space are not themselves real, nor are the vectors in a complex vector space complex. We have the following examples of vector spaces:

1. The set of N-component vectors

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}, \quad a_i \in \mathbb{R}, \quad i = 1, 2, \ldots N. \tag{1.1}$$

   form a real vector space.

2. The set of $M \times N$ matrices with complex entries

$$\begin{pmatrix} a_{11} & \ldots & a_{1N} \\ a_{21} & \ldots & a_{2N} \\ \vdots & \vdots & \vdots \\ a_{M1} & \ldots & a_{MN} \end{pmatrix}, \qquad a_{ij} \in \mathbb{C}, \tag{1.2}$$

is a complex vector space. In here multiplication by a constant multiplies each entry of the matrix by the constant.

3. We can have matrices with complex entries that naturally form a real vector space. The space of two-by-two *hermitian* matrices define a *real* vector space. They do not form a complex vector space since multiplication of a hermitian matrix by a complex number ruins the hermiticity.

4. The set $\mathcal{P}(\mathbb{F})$ of polynomials $p(z)$. Here the variable $z \in \mathbb{F}$ and $p(z) \in \mathbb{F}$. Each polynomial $p(z)$ has coefficients $a_0, a_1, \ldots a_n$ also in $\mathbb{F}$:

$$p(z) = a_0 + a_1 z + a_2 z^2 + \ldots + a_n z^n \,. \tag{1.3}$$

   By definition, the integer n is finite but it can take any nonnegative value. Addition of polynomials works as expected and multiplication by a constant is also the obvious multiplication. The space $\mathcal{P}(\mathbb{F})$ of all polynomials so defined form a vector space over $\mathbb{F}$.

5. The set $\mathbb{F}^\infty$ of infinite sequences $(x_1, x_2, \ldots)$ of elements $x_i \in \mathbb{F}$. Here

$$\begin{aligned} (x_1, x_2, \ldots) + (y_1, y_2, \ldots) &= (x_1 + y_1, x_2 + y_2, \ldots) \\ a(x_1, x_2, \ldots) &= (ax_1, ax_2, \ldots) \quad a \in \mathbb{F} \,. \end{aligned} \tag{1.4}$$

   This is a vector space over $\mathbb{F}$.

6. The set of complex functions on an interval $x \in [0, L]$, form a vector space over $\mathbb{C}$.

To better understand a vector space one can try to figure out its possible subspaces. A **subspace** of a vector space V is a subset of V that is also a vector space. To verify that a subset U of V is a subspace you must check that U contains the vector 0, and that U is closed under addition and scalar multiplication.

Sometimes a vector space V can be described clearly in terms of collection $U_1, U_2, \ldots U_m$ of subspaces of V. We say that the space V is the **direct sum** of the subspaces $U_1, U_2, \ldots U_m$ and we write

$$V \;=\; U_1 \oplus U_2 \oplus \cdots \oplus U_m \tag{1.5}$$

if any vector in V can be written *uniquely* as the sum $u_1 + u_2 + \ldots + u_m$, where $u_i \in U_i$. To check uniqueness one can, alternatively, verify that the only way to write 0 as a sum $u_1 + u_2 + \ldots + u_m$ with $u_i \in U_i$ is by taking all u_i's equal to zero. For the case of two subspaces $V = U \oplus W$, it suffices to prove that any vector can be written as $u + w$ with $u \in U$ and $w \in W$ and that $U \cap W = 0$.

Given a vector space we can produce lists of vectors. A **list** $(v_1, v_2, \ldots, v_n)$ of vectors in V contains, by definition, a finite number of vectors. The number of vectors in the list is the length of the list. The **span** of a list of vectors $(v_1, v_2, \cdots v_n)$ in V, denoted as $\text{span}(v_1, v_2, \cdots, v_n)$, is the set of all linear combinations of these vectors

$$a_1 v_1 + a_2 v_2 + \ldots a_n v_n \,, \qquad a_i \in \mathbb{F} \tag{1.6}$$

A vector space V is spanned by a list $(v_1, v_2, \cdots v_n)$ if $V = \text{span}(v_1, v_2, \cdots v_n)$.

Now comes a very natural definition: A vector space V is said to be **finite dimensional** if it is spanned by some list of vectors in V. If V is not finite dimensional, it is **infinite dimensional**. In such case, no list of vectors from V can span V.

Let us show that the vector space of all polynomials $p(z)$ considered in Example 4 is an infinite dimensional vector space. Indeed, consider any list of polynomials. In this list there is a polynomial of maximum degree (recall the list is finite). Thus polynomials of higher degree are not in the span of the list. Since no list can span the space, it is infinite dimensional.

For example 1, consider the list of vectors $(e_1, e_2, \ldots e_N)$ with

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \; e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots \quad e_N = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \tag{1.7}$$

This list spans the space (the vector displayed is $a_1 e_1 + a_2 e_2 + \ldots a_N e_N$). This vector space is finite dimensional.

A list of vectors $(v_1, v_2, \ldots, v_n)$, with $v_i \in V$ is said to be **linearly independent** if the equation

$$a_1 v_1 + a_2 v_2 + \ldots + a_n v_n \; = \; 0\,, \tag{1.8}$$

only has the solution $a_1 = a_2 = \cdots = a_n = 0$. One can show that the length of any linearly independent list is shorter or equal to the length of any spanning list. This is reasonable, because spanning lists can be arbitrarily long (adding vectors to a spanning list gives still a spanning list), but a linearly independent list cannot be enlarged beyond a certain point.

Finally, we get to the concept of a basis for a vector space. A **basis** of V is a list of vectors in V that both spans V and it is linearly independent. Mathematicians easily prove that any finite dimensional vector space has a basis. Moreover, all bases of a finite dimensional vector space have the same length. The **dimension** of a finite-dimensional vector space is given by the length of any list of basis vectors. One can also show that for a finite dimensional vector space a list of vectors of length $\dim V$ is a basis if it is linearly independent list or if it is a spanning list.

For example 1 we see that the list $(e_1, e_2, \ldots e_N)$ in (1.7) is not only a spanning list but a linearly independent list (prove it!). Thus the dimensionality of this space is N.

For example 3, recall that the most general hermitian two-by-two matrix takes the form

$$\begin{pmatrix} a_0 + a_3 & a_1 - i a_2 \\ a_1 + i a_2 & a_0 - a_3 \end{pmatrix}, \qquad a_0, a_1, a_2, a_3 \in \mathbb{R}. \tag{1.9}$$

Now consider the following list of four 'vectors' $(\mathbf{1}, \sigma_1, \sigma_2, \sigma_3)$. All entries in this list are hermitian matrices, so this is a list of vectors in the space. Moreover they span the space since the most general hermitian matrix, as shown above, is simply $a_0 \mathbf{1} + a_1 \sigma_1 + a_2 \sigma_2 + a_3 \sigma_3$. The list is linearly independent

as $a_0\mathbf{1} + a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3 = 0$ implies that

$$\begin{pmatrix} a_0 + a_3 & a_1 - ia_2 \\ a_1 + ia_2 & a_0 - a_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \tag{1.10}$$

and you can quickly see that this implies a_0, a_1, a_2, and a_3 are zero. So the list is a basis and the space in question is a four-dimensional real vector space.

Exercise. Explain why the vector space in example 2 has dimension $M \cdot N$.

It seems pretty obvious that the vector space in example 5 is infinite dimensional, but it actually takes a bit of work to prove it.

2 Linear operators and matrices

A linear map refers in general to a certain kind of function from one vector space V to another vector space W. When the linear map takes the vector space V to itself, we call the linear map a linear operator. We will focus our attention on those operators. Let us then define a linear operator.

A **linear operator** T on a vector space V is a function that takes V to V with the properties:

1. $T(u + v) = Tu + Tv$, for all $u, v \in V$.
2. $T(au) = aTu$, for all $a \in \mathbb{F}$ and $u \in V$.

We call $\mathcal{L}(V)$ the set of all linear operators that act on V. This can be a very interesting set, as we will see below. Let us consider a few examples of linear operators.

1. Let V denote the space of real polynomials $p(x)$ of a real variable x with real coefficients. Here are two linear operators:
   - Let T denote differentiation: $Tp = p'$. This operator is linear because $(p_1 + p_2)' = p_1' + p_2'$ and $(ap)' = ap'$.
   - Let S denote multiplication by x: $Sp = xp$. S is also a linear operator.
2. In the space $\mathbb{F}^\infty$ of infinite sequences define the left-shift operator L by

   $$L(x_1, x_2, x_3, \ldots) = (x_2, x_3, \ldots). \tag{2.11}$$

   We lose the first entry, but that is perfectly consistent with linearity. We also have the right-shift operator R that acts as follows:

   $$R(x_1, x_2, \ldots) = (0, x_1, x_2, \ldots). \tag{2.12}$$

   Note that the first entry in the result is zero. It could not be any other number because the zero element (a sequence of all zeroes) should be mapped to itself (by linearity).

3. For any V, the zero map 0 such that $0v = 0$. This map is linear and maps all elements of V to the zero element.

4. For any V, the identity map I for which $Iv = v$ for all $v \in V$. This map leaves all vectors invariant.

Since operators on V can be added and can also be multiplied by numbers, the set $\mathcal{L}(V)$ introduced above is itself a vector space (the vectors being the operators!). Indeed for any two operators $T, S \in \mathcal{L}(V)$ we have the natural definition

$$\begin{aligned} (S+T)v &= Sv + Tv\,, \\ (aS)v &= a(Sv)\,. \end{aligned} \tag{2.13}$$

The additive identity in the vector space $\mathcal{L}(V)$ is the zero map of example 3.

In this vector space there is a surprising new structure: the vectors (the operators!) can be multiplied. There is a multiplication of linear operators that gives a linear operator. We just let one operator act first and the second later. So given $S, T \in \mathcal{L}(V)$ we define the operator ST as

$$(ST)v \equiv S(Tv) \tag{2.14}$$

You should convince yourself that ST is a linear operator. This product structure in the space of linear operators is associative: $S(TU) = (ST)U$, for S, T, U, linear operators. Moreover it has an identity element: the identity map of example 4. Most crucially this multiplication is, in general, *noncommutative*. We can check this using the two operators T and S of example 1 acting on the polynomial $p = x^n$. Since T differentiates and S multiplies by x we get

$$(TS)x^n = T(Sx^n) = T(x^{n+1}) = (n+1)x^n\,, \quad \text{while} \quad (ST)x^n = S(Tx^n) = S(nx^{n-1}) = nx^n\,. \tag{2.15}$$

We can quantify this failure of commutativity by writing the difference

$$(TS - ST)x^n = (n+1)x^n - nx^n = x^n = I\,x^n \tag{2.16}$$

where we inserted the identity operator at the last step. Since this relation is true for any x^n, it would also hold acting on any polynomial, namely on any element of the vector space. So we write

$$[\,T\,,S\,] \;=\; I\,. \tag{2.17}$$

where we introduced the commutator $[\cdot,\cdot]$ of two operators X, Y, defined as $[X,Y] \equiv XY - YX$.

The most basic features of an operator are captured by two simple concepts: its null space and its range. Given some linear operator T on V it is of interest to consider those elements of V that are mapped to the zero element. The **null space** (or kernel) of $T \in \mathcal{L}(V)$ is the subset of vectors in V that are mapped to zero by T:

$$\text{null}\; T \;=\; \{v \in V;\; Tv = 0\}\,. \tag{2.18}$$

Actually $\text{null}\, T$ is a *subspace* of V (The only nontrivial part of this proof is to show that $T(0) = 0$. This follows from $T(0) = T(0+0) = T(0) + T(0)$ and then adding to both sides of this equation the additive inverse to $T(0)$).

A linear operator $T : V \to V$ is said to be **injective** if $Tu = Tv$, with $u, v \in V$, implies $u = v$. An injective map is called a *one-to-one* map, because not two different elements can be mapped to the same one. In fact, physicist Sean Carroll has suggested that a better name would be *two-to-two* as injectivity really means that two different elements are mapped by T to two different elements! We leave for you as an exercise to prove the following important characterization of injective maps:

Exercise. Show that T is injective if and only if $\text{null}\, T = \{0\}$.

Given a linear operator T on V it is also of interest to consider the elements of V of the form Tv. The linear operator may not produce by its action all of the elements of V. We define the **range** of T as the image of V under the map T:

$$\text{range}\; T \;=\; \{Tv;\; v \in V\}\,. \tag{2.19}$$

Actually $\text{range}\, T$ is a *subspace* of V (can you prove it?). The linear operator T is said to be **surjective** if $\text{range}\, T = V$. That is, if the image of V under T is the complete V.

Since both the null space and the range of a linear operator $T : V \to V$ are subspaces of V, one can assign a dimension to them, and the following theorem is nontrivial:

$$\dim V \;=\; \dim(\text{null}\; T) \;+\; \dim(\text{range}\; T)\,. \tag{2.20}$$

Example. Describe the null space and range of the operator

$$T = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{2.21}$$

Let us now consider invertible linear operators. A linear operator $T \in \mathcal{L}(V)$ is **invertible** if there exists another linear operator $S \in \mathcal{L}(V)$ such that ST and TS are identity maps (written as I). The linear operator S is called the **inverse** of T. The inverse is actually unique. Say S and S' are inverses of T. Then we have

$$S \;=\; SI \;=\; S(TS') \;=\; (ST)S' \;=\; IS' \;=\; S'\,. \tag{2.22}$$

Note that we required the inverse S to be an inverse acting from the left and acting from the right. This is useful for infinite dimensional vector spaces. For finite-dimensional vector spaces one suffices; one can then show that $ST = I$ if and only if $TS = I$.

It is useful to have a good characterization of invertible linear operators. For a finite-dimensional vector space V the following three statements are equivalent!

$$\text{Finite dimension:} \quad \boxed{T \text{ is invertible}} \;\longleftrightarrow\; \boxed{T \text{ is injective}} \;\longleftrightarrow\; \boxed{T \text{ is surjective}} \tag{2.23}$$

For infinite dimensional vector spaces injectivity and surjectivity are not equivalent (each can fail independently). In that case invertibility is equivalent to injectivity plus surjectivity:

$$\text{Infinite dimension:} \quad \boxed{T \text{ is invertible}} \longleftrightarrow \boxed{T \text{ is injective and surjective}} \tag{2.24}$$

The left shift operator L is not injective (maps $(x_1, 0, \ldots)$ to zero) but it is surjective. The right shift operator is not surjective although it is injective.

Now we consider the **matrix associated to a linear operator** T that acts on a vector space V. This matrix will depend on the basis we choose for V. Let us declare that our basis is the list $(v_1, v_2, \ldots v_n)$. It is clear that the full knowledge of the action of T on V is encoded in the action of T on the basis vectors, that is on the values $(Tv_1, Tv_2, \ldots, Tv_n)$. Since Tv_j is in V, it can be written as a linear combination of basis vectors. We then have

$$\boxed{Tv_j \;=\; T_{1\,j}\, v_1 \,+\, T_{2\,j}\, v_2 \,+\, \ldots \,+ T_{n\,j}\, v_n\,,} \tag{2.25}$$

where we introduced the constants $T_{i,j}$ that are known if the operator T is known. As we will see, these are the entries form the matrix representation of the operator T in the chosen basis. The above relation can be written more briefly as

$$\boxed{Tv_j \;=\; \sum_{i=1}^{n} T_{ij}\, v_i\,.} \tag{2.26}$$

When we deal with different bases it can be useful to use notation where we replace

$$T_{ij} \quad \to \quad T_{ij}(\{v\})\,, \tag{2.27}$$

so that it makes clear that T is being represented using the v basis $(v_1, \ldots, v_n)$.

I want to make clear why (2.25) is reasonable before we show that it makes for a consistent association between operator multiplication and matrix multiplication. The left-hand side, where we have the action of the matrix for T on the j-th basis vector, can be viewed concretely as

$$Tv_j \quad \longleftrightarrow \quad \begin{pmatrix} T_{11} & \cdots & T_{1\,j} & \cdots & T_{1n} \\ T_{21} & \cdots & T_{2\,j} & \cdots & T_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ T_{n1} & \cdots & T_{n\,j} & \cdots & T_{nn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \; j\text{-th position} \tag{2.28}$$

where the column vector has zeroes everywhere except on the j-th entry. The product, by the usual rule of matrix multiplication is the column vector

$$\begin{pmatrix} T_{1\,j} \\ T_{2\,j} \\ \vdots \\ T_{n\,j} \end{pmatrix} \;=\; T_{1\,j} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + T_{2\,j} \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \;\ldots\; + T_{n\,j} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \quad \longleftrightarrow \quad T_{1\,j} v_1 + \ldots T_{n\,j} v_n\,. \tag{2.29}$$

which we identify with the right-hand side of (2.25). So (2.25) is reasonable.

Exercise. Verify that the matrix representation of the identity operator is a diagonal matrix with an entry of one at each element of the diagonal. This is true for any basis.

Let us now examine the product of two operators and their matrix representation. Consider the operator TS acting on v_j:

$$(TS)v_j \;=\; T(Sv_j) \;=\; T\sum_p S_{pj}v_p \;=\; \sum_p S_{pj}\; Tv_p \;=\; \sum_p S_{pj}\;\sum_i T_{ip}v_i \tag{2.30}$$

so that changing the order of the sums we find

$$(TS)v_j \;=\; \sum_i \Big(\sum_p T_{ip}S_{pj}\Big)\, v_i\,. \tag{2.31}$$

Using the identification implicit in (2.26) we see that the object in parenthesis is the i, j matrix element of the matrix that represents TS. Therefore we found

$$(TS)_{ij} \;=\; \sum_p T_{ip}S_{pj}\,, \tag{2.32}$$

which is precisely the right formula for matrix multiplication. In other words, the matrix that represents TS is the product of the matrix that represents T with the matrix that represents S, in that order.

Changing basis

While matrix representations are very useful for concrete visualization, they are basis dependent. It is a good idea to try to figure out if there are quantities that can be calculated using a matrix representation that are, nevertheless, guaranteed to be basis independent. One such quantity is the **trace** of the matrix representation of a linear operator. The trace is the sum of the matrix elements in the diagonal. Remarkably, that sum is the same independent of the basis used. Consider a linear operator T in $\mathcal{L}(V)$ and two sets of basis vectors $(v_1, \ldots, v_n)$ and $(u_1, \ldots, u_n)$ for V. Using the explicit notation (2.27) for the matrix representation we state this property as

$$\mathrm{tr}\, T(\{v\}) \;=\; \mathrm{tr}\, T(\{u\})\,. \tag{2.33}$$

We will establish this result below. On the other hand, if this trace is actually basis independent, there should be a way to define the trace of the linear operator T *without* using its matrix representation. This is actually possible, as we will see. Another basis independent quantity is the determinant of the matrix representation of T.

Let us then consider the effect of a change of basis on the matrix representation of an operator. Consider a vector space V and a change of basis from $(v_1, \ldots v_n)$ to $(u_1, \ldots u_n)$ defined by the linear operator A as follows:

$$A:\; v_k \to u_k,\; \text{for } k = 1, \ldots, n\,. \tag{2.34}$$

This can also be written as

$$Av_k = u_k \tag{2.35}$$

Since we know how A acts on every element of the basis we know, by linearity how it acts on any vector. The operator A is clearly *invertible* because, letting $B : u_k \to v_k$ or

$$Bu_k = v_k \,, \tag{2.36}$$

we have

$$\begin{aligned} BAv_k = B(Av_k) = Bu_k &= v_k \\ ABu_k = A(Bu_k) = Av_k &= u_k \,, \end{aligned} \tag{2.37}$$

showing that $BA = I$ and $AB = I$. Thus B is the inverse of A. Using the definition of matrix representation, the right-hand sides of the relations $u_k = Av_k$ and $v_k = Bu_k$ can be written so that the equations take the form

$$u_k = A_{jk}\, v_j \,, \qquad v_k = B_{jk}\, u_j \,, \tag{2.38}$$

where we used the convention that repeated indices are summed over. A_{ij} are the elements of the matrix representation of A in the v basis and B_{ij} are the elements of the matrix representation of B in the u basis. Replacing the second relation on the first, and then replacing the first on the second we get

$$\begin{aligned} u_k &= A_{jk}\, B_{ij}\, u_i = B_{ij} A_{jk}\, u_i \\ v_k &= B_{jk}\, A_{ij}\, v_i = A_{ij} B_{jk}\, v_i \end{aligned} \tag{2.39}$$

Since the u's and v's are basis vectors we must have

$$B_{ij} A_{jk} = \delta_{ik} \quad \text{and} \quad A_{ij} B_{jk} = \delta_{ik} \tag{2.40}$$

which means that the B matrix is the inverse of the A matrix. We have thus learned that

$$v_k = (A^{-1})_{jk}\, u_j \,. \tag{2.41}$$

We can now apply these preparatory results to the matrix representations of the operator T. We have, by definition,

$$Tv_k = T_{ik}(\{v\})\, v_i \,. \tag{2.42}$$

We now want to calculate T on u_k so that we can read the formula for the matrix T on the u basis:

$$Tu_k = T_{ik}(\{u\})\, u_i \,. \tag{2.43}$$

Computing the left-hand side, using the linearity of the operator T, we have

$$Tu_k = T(A_{jk} v_j) = A_{jk} T v_j = A_{jk} T_{pj}(\{v\})\, v_p \tag{2.44}$$

and using (2.41) we get

$$Tu_k = A_{jk}T_{pj}(\{v\})\,(A^{-1})_{ip}\,u_i \;=\; \Big((A^{-1})_{ip}\,T_{pj}(\{v\})\,A_{jk}\Big)u_i \;=\; \big(\,A^{-1}T(\{v\})A\big)_{ik}u_i\,. \tag{2.45}$$

Comparing with (2.43) we get

$$T_{ij}(\{u\}) \;=\; \big(\,A^{-1}T(\{v\})A\big)_{ij} \quad\to\quad \boxed{\;T(\{u\}) \;=\; A^{-1}T(\{v\})A\,.\;} \tag{2.46}$$

This is the result we wanted to obtain.

The trace of a matrix T_{ij} is given by T_{ii}, where sum over i is understood. To show that the trace of T is basis independent we write

$$\begin{aligned}\mathrm{tr}(T(\{u\})) \;&=\; T_{ii}(\{u\}) \;=\; (A^{-1})_{ij}T_{jk}(\{v\})A_{ki}\\ &= A_{ki}(A^{-1})_{ij}T_{jk}(\{v\})\\ &= \delta_{kj}T_{jk}(\{v\}) \;=\; T_{jj}(\{v\}) \;=\; \mathrm{tr}(T(\{v\}))\,.\end{aligned} \tag{2.47}$$

For the determinant we recall that $\det(AB) = (\det A)(\det B)$. Therefore $\det(A)\det(A^{-1}) = 1$. From (2.46) we then get

$$\det T(\{u\}) \;=\; \det(A^{-1})\det T(\{v\})\det A \;=\; \det T(\{v\})\,. \tag{2.48}$$

Thus the determinant of the matrix that represents a linear operator is independent of the basis used.

3 Eigenvalues and eigenvectors

In quantum mechanics we need to consider eigenvalues and eigenstates of hermitian operators acting on complex vector spaces. These operators are called observables and their eigenvalues represent possible results of a measurement. In order to acquire a better perspective on these matters, we consider the eigenvalue/eigenvector problem in more generality.

One way to understand the action of an operator $T \in \mathcal{L}(V)$ on a vector space V is to understand how it acts on subspaces of V, as those are smaller than V and thus possibly simpler to deal with. Let U denote a subspace of V. In general, the action of T may take elements of U outside U. We have a noteworthy situation if T acting on any element of U gives an element of U. In this case U is said to be **invariant** under T, and T is then a well-defined linear operator on U. A very interesting situation arises if a suitable list of invariant subspaces give the space V as a direct sum.

Of all subspaces, one-dimensional ones are the simplest. Given some vector $u \in V$ one can consider the one-dimensional subspace U spanned by u:

$$U = \{cu:\;\; c \in \mathbb{F}\}\,. \tag{3.49}$$

We can ask if the one-dimensional subspace U is left invariant by the operator T. For this Tu must be equal to a number times u, as this guarantees that $Tu \in U$. Calling the number λ, we write

$$T\,u \;=\; \lambda\,u\,. \tag{3.50}$$

This equation is so ubiquitous that names have been invented to label the objects involved. The number $\lambda \in \mathbb{F}$ is called an **eigenvalue** of the linear operator T if there is a **nonzero** vector $u \in V$ such that the equation above is satisfied. Suppose we find for some specific λ a nonzero vector u satisfying this equation. Then it follows that cu, for any $c \in \mathbb{F}$ also satisfies equation (3.50), so that the solution space of the equation includes the subspace U, which is now said to be an invariant subspace under T. It is convenient to call any vector that satisfies (3.50) for a given λ an **eigenvector** of T corresponding to λ. In doing so we are including the zero vector as a solution and thus as an eigenvector. It can often happen that for a given λ there are several linearly independent eigenvectors. In this case the invariant subspace associated with the eigenvalue λ is higher dimensional. The set of eigenvalues of T is called the **spectrum** of T.

Our equation above is equivalent to

$$(T - \lambda I)\,u = 0\,, \tag{3.51}$$

for some nonzero u. It is therefore the case that

$$\boxed{\lambda \;\text{ is an eigenvalue}} \quad \longleftrightarrow \quad \boxed{(T-\lambda I) \quad \text{not injective.}} \tag{3.52}$$

Using (2.23) we conclude that λ is an eigenvalue also means that $(T - \lambda I)$ is **not invertible**, and not surjective. We also note that

$$\text{Set of eigenvectors of } T \text{ corresponding to } \lambda \;=\; \text{null}\,(T - \lambda I)\,. \tag{3.53}$$

It should be emphasized that the eigenvalues of T and the invariant subspaces (or eigenvectors associated with fixed eigenvalues) are basis independent objects. Nowhere in our discussion we had to invoke the use of a basis, nor we had to use any matrix representation. Below, we will discuss the familiar calculation of eigenvalues and eigenvectors using a matrix representation of the operator T in some particular basis.

Let us consider some examples. Take a real three-dimensional vector space V (our space to great accuracy!). Consider the rotation operator T that rotates all vectors by a fixed angle small about the z axis. To find eigenvalues and eigenvectors we just think of the invariant subspaces. We must ask which are the vectors for which this rotation doesn't change their direction and effectively just multiplies them by a number? Only the vectors along the z-direction do not change direction upon this rotation. So the vector space spanned by $\mathbf{e}_z$ is the invariant subspace, or the space of eigenvectors. The eigenvectors are associated with the eigenvalue of one, as the vectors are not altered at all by the rotation.

Consider now the case where T is a rotation by ninety degrees on a two-dimensional *real* vector space V. Are there one-dimensional subspaces left invariant by T? No, **all** vectors are rotated, none remains pointing in the same direction. Thus there are **no eigenvalues**, nor, of course, eigenvectors. If you tried calculating the eigenvalues by the usual recipe, you will find complex numbers. A complex eigenvalue is meaningless in a real vector space.

Although we will not prove the following result, it follows from the facts we have introduced and no extra machinery. It is of interest being completely general and valid for both real and complex vector spaces:

Theorem: Let $T \in \mathcal{L}(V)$ and assume $\lambda_1, \ldots \lambda_n$ are distinct eigenvalues of T and $u_1, \ldots u_n$ are corresponding nonzero eigenvectors. Then $(u_1, \ldots u_n)$ are linearly independent.

Note that we cannot ask if the eigenvectors are orthogonal to each other as we have not yet introduced an inner product on the vector space V. In this theorem there may be more than one linearly independent eigenvector associated with some eigenvalues. In that case any one eigenvector will do. Since an n-dimensional vector space V does not have more than n linearly independent vectors, no linear operator on V can have more than n distinct eigenvalues.

We saw that some linear operators in real vector spaces can fail to have eigenvalues. Complex vector spaces are nicer. In fact, *every linear operator on a finite-dimensional complex vector space has at least one eigenvalue*. This is a fundamental result. It can be proven without using determinants with an elegant argument, but the proof using determinants is quite short.

When λ is an eigenvalue, we have seen that $T - \lambda I$ is not an invertible operator. This also means that using any basis, the matrix representative of $T - \lambda I$ is non-invertible. The condition of non-invertibility of a matrix is identical to the condition that its determinant vanish:

$$\det(T - \lambda \mathbf{1}) \;=\; 0\,. \tag{3.54}$$

This condition, in an N-dimensional vector space looks like

$$\det \begin{pmatrix} T_{11} - \lambda & T_{12} & \ldots & T_{1N} \\ T_{21} & T_{22} - \lambda & \ldots & T_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ T_{N1} & T_{N2} & \ldots & T_{NN} - \lambda \end{pmatrix} \;=\; 0\,. \tag{3.55}$$

The left-hand side is a polynomial $f(\lambda)$ in λ of degree N called the *characteristic polynomial*:

$$f(\lambda) \;=\; \det(T - \lambda \mathbf{1}) = (-\lambda)^N + b_{N-1}\lambda^{N-1} + \ldots b_1\lambda + b_0\,, \tag{3.56}$$

where the b_i are constants. We are interested in the equation $f(\lambda) = 0$, as this determines all possible eigenvalues. If we are working on real vector spaces, the constants b_i are real but there is no guarantee of real roots for $f(\lambda) = 0$. With complex vector spaces, the constants b_i will be complex, but a complex solution for $f(\lambda) = 0$ always exists. Indeed, over the complex numbers we can factor the polynomial $f(\lambda)$ as follows

$$f(\lambda) \;=\; (-1)^N(\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_N)\,, \tag{3.57}$$

where the notation does not preclude the possibility that some of the λ_i's may be equal. The λ_i's are the eigenvalues, since they lead to $f(\lambda) = 0$ for $\lambda = \lambda_i$. If all eigenvalues of T are different the spectrum of T is said to be *non-degenerate*. If an eigenvalue appears k times it is said to be a degenerate eigenvalue with of multiplicity k. Even in the most degenerate case we must have at least one eigenvalue. The eigenvectors exist because $(T - \lambda I)$ non-invertible means it is not injective, and therefore there are nonzero vectors that are mapped to zero by this operator.

4 Inner products

We have been able to go a long way without introducing extra structure on the vector spaces. We have considered linear operators, matrix representations, traces, invariant subspaces, eigenvalues and eigenvectors. It is now time to put some additional structure on the vector spaces. In this section we consider a function called an *inner product* that allows us to construct numbers from vectors. A vector space equipped with an inner product is called an inner-product space.

An **inner product** on a vector space V over $\mathbb{F}$ is a machine that takes an *ordered* pair of elements of V, that is, a first vector and a second vector, and yields a number in $\mathbb{F}$. In order to motivate the definition of an inner product we first discuss the familiar way in which we associate a length to a vector.

The length of a vector, or **norm** of a vector is a real number that is positive or zero, if the vector is the zero vector. In $\mathbb{R}^n$ a vector $a = (a_1, \ldots a_n)$ has norm $|a|$ defined by

$$|a| \;=\; \sqrt{a_1^2 + \ldots a_n^2} \tag{4.58}$$

Squaring this one may think of $|a|^2$ as the *dot product* of a with a:

$$|a|^2 = a \cdot a = a_1^2 + \ldots a_n^2 \tag{4.59}$$

Based on this the dot product of any two vectors a and b is defined by

$$a \cdot b \;=\; a_1 b_1 + \;\ldots\; + a_n b_n \,. \tag{4.60}$$

If we try to generalize this dot product we may require as needed properties the following

1. $a \cdot a \;\geq 0$, for all vectors a.
2. $a \cdot a = 0$ if and only if $a = 0$.
3. $a \cdot (b_1 + b_2) \;=\; a \cdot b_1 + a \cdot b_2$. Additivity in the second entry.
4. $a \cdot (\alpha\, b) \;=\; \alpha\, a \cdot b$, with α a number.
5. $a \cdot b \;=\; b \cdot a$.

Along with these axioms, the length $|a|$ of a vector a is the positive or zero number defined by relation

$$|a|^2 \;=\; a \cdot a\,. \tag{4.61}$$

These axioms are satisfied by the definition (4.60) but do not require it. A new dot product defined by $a \cdot b = c_1 a_1 b_1 + \ldots + c_n a_n b_n$, with $c_1, \ldots c_n$ positive constants, would do equally well! So whatever can be proven with these axioms holds true not only for the conventional dot product.

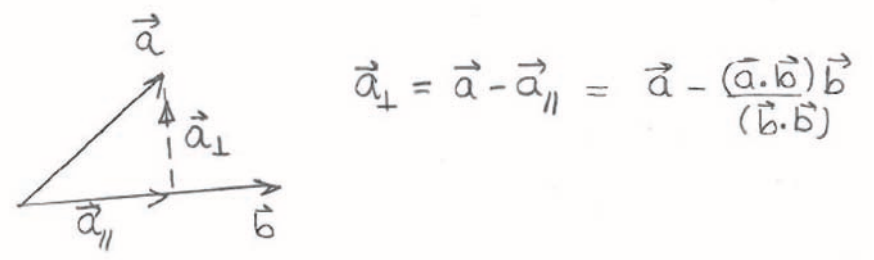
The above axioms guarantee that the Schwarz inequality holds:

$$|a \cdot b| \;\leq\; |\,a\,|\,|\,b\,|\,. \tag{4.62}$$

To prove this consider two (nonzero) vectors a and b and then consider the shortest vector joining the tip of a to the line defined by the direction of b (see the figure below). This is the vector $a_\perp$, given by

$$a_\perp \;\equiv\; a \;-\; \frac{a \cdot b}{b \cdot b}\; b\,. \tag{4.63}$$

The subscript $\perp$ is there because the vector is perpendicular to b, namely $a_\perp \cdot b = 0$, as you can quickly see. To write the above vector we subtracted from a the component of a parallel to b. Note that the vector $a_\perp$ is not changed as $b \to cb$; it does not depend on the overall length of b. Moreover, as it should, the vector $a_\perp$ is zero if and only if the vectors a and b are parallel. All this is only motivation, we could have just said "consider the following vector $a_\perp$".



Given axiom (1) we have that $a_\perp \cdot a_\perp \geq 0$ and therefore using (4.63)

$$a_\perp \cdot a_\perp \;=\; a \cdot a - \frac{(a \cdot b)^2}{b \cdot b} \;\geq\; 0\,. \tag{4.64}$$

Since b is not the zero vector we then have

$$(a \cdot b)^2 \;\leq\; (a \cdot a)(b \cdot b)\,. \tag{4.65}$$

Taking the square root of this relation we obtain the Schwarz inequality (4.62). The inequality becomes an equality only if $a_\perp = 0$ or, as discussed above, when $a = cb$ with c a real constant.

For complex vector spaces some modification is necessary. Recall that the length $|\gamma|$ of a complex number γ is given by $|\gamma| = \sqrt{\gamma^* \gamma}$, where the asterisk superscript denotes complex conjugation. It is

not hard to generalize this a bit. Let $z = (z_1, \ldots, z_n)$ be a vector in $\mathbb{C}^n$. Then the length of the vector $|z|$ is a real number greater than zero given by

$$|z| \;=\; \sqrt{z_1^* z_1 + \ldots + z_n^* z_n}\,. \tag{4.66}$$

We must use complex conjugates, denoted by the asterisk superscript, to produce a real number greater than or equal to zero. Squaring this we have

$$|z|^2 \;=\; z_1^* z_1 + \ldots + z_n^* z_n\,. \tag{4.67}$$

This suggests that for vectors $z = (z_1, \ldots, z_n)$ and $w = (w_1, \ldots, w_n)$ an inner product could be given by

$$w_1^* z_1 + \ldots + w_n^* z_n\,, \tag{4.68}$$

and we see that we are not treating the two vectors in an equivalent way. There is the first vector, in this case w whose components are conjugated and a second vector z whose components are not conjugated. If the order of vectors is reversed, we get for the inner product the complex conjugate of the original value. As it was mentioned at the beginning of the section, the inner product requires an ordered pair of vectors. It certainly does for complex vector spaces. Moreover, one can define an inner product in general in a way that applies both to complex and real vector spaces.

An **inner product** on a vector space V over $\mathbb{F}$ is a map from an ordered pair (u, v) of vectors in V to a number $\langle u, v\rangle$ in $\mathbb{F}$. The axioms for $\langle u, v\rangle$ are inspired by the axioms we listed for the dot product.

1. $\langle v\,, v\rangle \geq 0$, for all vectors $v \in V$.
2. $\langle v, v\rangle = 0$ if and only if $v = 0$.
3. $\langle u\,, v_1 + v_2\rangle \;=\; \langle u\,, v_1\rangle + \langle u\,, v_2\rangle$. Additivity in the second entry.
4. $\langle u\,, \alpha\, v\rangle \;=\; \alpha\,\langle u\,, v\rangle$, with $\alpha \in \mathbb{F}$. Homogeneity in the second entry.
5. $\langle u\,, v\rangle \;=\; \langle v\,, u\rangle^*$. Conjugate exchange symmetry.

This time the **norm** $|v|$ of a vector $v \in V$ is the positive or zero number defined by relation

$$|v|^2 \;=\; \langle v\,, v\rangle\,. \tag{4.69}$$

From the axioms above, the only major difference is in number five, where we find that the inner product is not symmetric. We know what complex conjugation is in $\mathbb{C}$. For the above axioms to apply to vector spaces over $\mathbb{R}$ we just define the obvious: complex conjugation of a real number is a real number. In a real vector space the $*$ conjugation does nothing and the inner product is strictly symmetric in its inputs.

A few comments. One can use (3) with $v_2 = 0$ to show that $\langle u, 0\rangle = 0$ for all $u \in V$, and thus, by (5) also $\langle 0, u\rangle = 0$. Properties (3) and (4) amount to full linearity in the second entry. It is important to note that additivity holds for the first entry as well:

$$\begin{aligned} \langle u_1 + u_2, v\rangle &= \langle v, u_1 + u_2\rangle^* \\ &= (\langle v, u_1\rangle + \langle v, u_2\rangle)^* \\ &= \langle v, u_1\rangle^* + \langle v, u_2\rangle^* \\ &= \langle u_1, v\rangle + \langle u_2, v\rangle\,. \end{aligned} \tag{4.70}$$

Homogeneity works differently on the first entry, however,

$$\begin{aligned} \langle \alpha\, u\,,\, v\rangle &= \langle v\,, \alpha\, u\rangle^* \\ &= (\alpha \langle v\,, u\rangle)^* \\ &= \alpha^*\, \langle u\,,\, v\rangle\,. \end{aligned} \tag{4.71}$$

Thus we get **conjugate homogeneity** on the first entry. This is a very important fact. Of course, for a real vector space conjugate homogeneity is the same as just plain homogeneity.

Two vectors $u, v \in V$ are said to be **orthogonal** if $\langle u, v\rangle = 0$. This, of course, means that $\langle v, u\rangle = 0$ as well. The zero vector is orthogonal to all vectors (including itself). Any vector orthogonal to all vectors in the vector space must be equal to zero. Indeed, if $x \in V$ is such that $\langle x, v\rangle = 0$ for all v, pick $v = x$, so that $\langle x, x\rangle = 0$ implies $x = 0$ by axiom 2. This property is sometimes stated as the **non-degeneracy** of the inner product. The "Pythagorean" identity holds for the norm-squared of orthogonal vectors in an inner-product vector space. As you can quickly verify,

$$|u + v|^2 \;=\; |u|^2 + |v|^2\,, \quad \text{for } u, v \in V, \text{ orthogonal vectors.} \tag{4.72}$$

The Schwarz inequality can be proven by an argument fairly analogous to the one we gave above for dot products. The result now reads

$$\boxed{\text{Schwarz Inequality:} \qquad |\langle u\,, v\rangle| \;\le\; |u|\,|v|\,.} \tag{4.73}$$

The inequality is saturated if and only if one vector is a multiple of the other. Note that in the left-hand side |...| denotes the norm of a complex number and on the right-hand side each |...| denotes the norm of a vector. You will prove this identity in a slightly different way in the homework. You will also consider there the *triangle inequality*

$$|u + v| \le |u| + |v|\,, \tag{4.74}$$

which is saturated when $u = cv$ for c a real, positive constant. Our definition (4.69) of norm on a vector space V is mathematically sound: a norm is required to satisfy the triangle inequality. Other properties are required: (i) $|v| \ge 0$ for all v, (ii) $|v| = 0$ if and only if $v = 0$, and (iii) $|cv| = |c||a|$ for c some constant. Our norm satisfies all of them.

A complex vector space with an inner product as we have defined is a *Hilbert space* if it is finite dimensional. If the vector space is infinite dimensional, an extra *completeness* requirement must be satisfied for the space to be a Hilbert space: all Cauchy sequences of vectors must converge to vectors in the space. An infinite sequence of vectors v_i, with $i = 1, 2, \ldots, \infty$ is a Cauchy sequence if for any $\epsilon > 0$ there is an N such that $|v_n - v_m| < \epsilon$ whenever $n, m > N$.

5 Orthonormal basis and orthogonal projectors

In an inner-product space we can demand that basis vectors have special properties. A list of vectors is said to be **orthonormal** if all vectors have norm one and are pairwise orthogonal. Consider a list $(e_1, \ldots, e_n)$ of orthonormal vectors in V. Orthonormality means that

$$\langle e_i, e_j \rangle \;=\; \delta_{ij}\,. \tag{5.75}$$

We also have a simple expression for the norm of $a_1 e_1 + \ldots + a_n e_n$, with $a_i \in \mathbb{F}$:

$$\begin{aligned} |a_1 e_1 + \ldots + a_n e_n|^2 \;&= \big\langle a_1 e_1 + \ldots + a_n e_n \,,\, a_1 e_1 + \ldots + a_n e_n \big\rangle \\ &= \langle a_1 e_1 \,, a_1 e_1 \rangle \;+ \ldots + \langle a_n e_n \,, a_n e_n \rangle \\ &= |a_1|^2 + \;\ldots\; + |a_n|^2\,. \end{aligned} \tag{5.76}$$

This result implies the somewhat nontrivial fact that *the vectors in any orthonormal list are linearly independent*. Indeed if $a_1 e_1 + \ldots + a_n e_n = 0$ then its norm is zero and so is $|a_1|^2 + \ldots + |a_n|^2$. This implies all $a_i = 0$, thus proving the claim.

An **orthonormal basis** of V is a list of orthonormal vectors that is also a basis for V. Let $(e_1, \ldots, e_n)$ denote an orthonormal basis. Then any vector v can be written as

$$v \;=\; a_1 e_1 + \ldots + a_n e_n\,, \tag{5.77}$$

for some constants a_i that can be calculated as follows

$$\langle e_i, v \rangle \;=\; \langle e_i \,, a_i e_i \rangle \;=\; a_i\,, \quad (\,i \text{ not summed}). \tag{5.78}$$

Therefore any vector v can be written as

$$v \;=\; \langle e_1, v \rangle\, e_1 + \;\ldots\; + \langle e_n \,, v \rangle \;=\; \langle e_i \,, v \rangle\, e_i\,. \tag{5.79}$$

To find an orthonormal basis on an inner product space V we just need to start with a basis and then use an algorithm to turn it into an orthogonal basis. In fact, a little more generally:
Gram-Schmidt: Given a list $(v_1, \ldots, v_n)$ of linearly independent vectors in V one can construct a list $(e_1, \ldots, e_n)$ of orthonormal vectors such that both lists span the same subspace of V.
The Gram-Schmidt algorithm goes as follows. You take e_1 to be v_1, normalized to have unit norm: $e_1 = v_1/|v_1|$. Then take $v_2 + \alpha e_1$ and fix the constant α so that this vector is orthogonal to e_1. The

answer is clearly $v_2 - \langle e_1, v_2\rangle e_1$. This vector, normalized by dividing it by its norm, is set equal to e_2. In fact we can write the general vector in a recursive fashion. If we know $e_1, e_2, \ldots, e_{j-1}$, we can write e_j as follows:

$$e_j \;=\; \frac{v_j - \langle e_1, v_j\rangle e_1 - \ldots - \langle e_{j-1}, v_j\rangle e_{j-1}}{|v_j - \langle e_1, v_j\rangle e_1 - \ldots - \langle e_{j-1}, v_j\rangle e_{j-1}|} \tag{5.80}$$

It should be clear to you by inspection that this vector is orthogonal to the vectors e_i with $i < j$ and has unit norm. The Gram-Schmidt procedure is quite practical.

With an inner product we can construct interesting subspaces of a vector space V. Consider a subset U of vectors in V (not necessarily a subspace). Then we can define a subspace $U^\perp$, called the **orthogonal complement** of U as the set of all vectors orthogonal to the vectors in U:

$$U^\perp \;=\; \{v \in V \,|\langle v, u\rangle = 0, \text{ for all } u \in U\}\,. \tag{5.81}$$

This is clearly a subspace of V. When U is a subspace, then U and $U^\perp$ actually give a direct sum decomposition of the full space:

Theorem: If U is a subspace of V, then $V = U \oplus U^\perp$.

Proof: This is a fundamental result and is not hard to prove. Let $(e_1, \ldots e_n)$ be an orthonormal basis for U. We can clearly write any vector v in V as

$$v \;=\; (\langle e_1, v\rangle e_1 + \ldots + \langle e_n, v\rangle e_n\,) \;+\; (\,v - \langle e_1, v\rangle e_1 - \ldots - \langle e_n, v\rangle e_n\,)\,. \tag{5.82}$$

On the right-hand side the first vector in parenthesis is clearly in U as it is written as a linear combination of U basis vectors. The second vector is clearly in $U^\perp$ as one can see that it is orthogonal to any vector in U. To complete the proof one must show that there is no vector except the zero vector in the intersection $U \cap U^\perp$ (recall the comments below (1.5)). Let $v \in U \cap U^\perp$. Then v is in U and in $U^\perp$ so it should satisfy $\langle v, v\rangle = 0$. But then $v = 0$, completing the proof.

Given this decomposition any vector $v \in V$ can be written uniquely as $v = u + w$ where $u \in U$ and $w \in U^\perp$. One can define a linear operator P_U, called the **orthogonal projection** of V onto U, that and that acting on v above gives the vector u. It is clear from this definition that: (i) the range of P_U is U. (ii) the null space of P_U is $U^\perp$, (iii) that P_U is not invertible and, (iv) acting on U, the operator P_U is the identity operator. The formula for the vector u can be read from (5.82)

$$P_U v \;=\; \langle e_1, v\rangle e_1 + \ldots + \langle e_n, v\rangle e_n\,. \tag{5.83}$$

It is a straightforward but a good exercise to verify that this formula is consistent with the fact that acting on U, the operator P_U is the identity operator. Thus if we act twice in succession with P_U on a vector, the second action has no effect as it is already acting on a vector in U. It follows from this that

$$P_U P_U = I\, P_U = P_U \quad \rightarrow \quad \boxed{P_U^2 \;=\; P_U\,.} \tag{5.84}$$

The eigenvalues and eigenvectors of P_U are easy to describe. Since all vectors in U are left invariant by the action of P_U, an orthonormal basis of U provides a set of orthonormal eigenvectors of P all with

eigenvalue one. If we choose on $U^\perp$ an orthonormal basis, that basis provides orthonormal eigenvectors of P all with eigenvalue zero.

In fact equation (5.84) implies that the eigenvalues of P_U can only be one or zero. T he eigenvalues of an operator satisfy whatever equation the operator satisfies (as shown by letting the equation act on a presumed eigenvector) thus $\lambda^2 = \lambda$ is needed, and this gives $\lambda(\lambda - 1) = 0$, and $\lambda = 0, 1$, as the only possibilities.

Consider a vector space $V = U \oplus U^\perp$ that is $(n + k)$-dimensional, where U is n-dimensional and $U^\perp$ is k-dimensional. Let $(e_1, \ldots, e_n)$ be an orthonormal basis for U and $(f_1, \ldots f_k)$ an orthonormal basis for $U^\perp$. We then see that the list of vectors $(g_1, \ldots g_{n+k})$ defined by

$$(g_1\,, \ldots, g_{n+k}) \;=\; (e_1, \ldots\,, e_n,\; f_1, \ldots f_k) \;\text{ is an orthonormal basis for }\; V. \tag{5.85}$$

Exercise: Use $P_U e_i = e_i$, for $i = 1, \ldots n$ and $P_U f_i = 0$, for $i = 1, \ldots, k$, to show that in the above basis the projector operator is represented by the diagonal matrix:

$$P_U \;=\; \mathrm{diag}\big(\; \underbrace{1, \ldots 1}_{n \text{ entries}}\;,\; \underbrace{0, \ldots, 0}_{k \text{ entries}}\;\big)\,. \tag{5.86}$$

We see that, as expected from its non-invertibility, $\det(P_U) = 0$. But more interestingly we see that the trace of the matrix P_U is n. Therefore

$$\mathrm{tr}\, P_U \;=\; \dim U\,. \tag{5.87}$$

The dimension of U is the **rank** of the projector P_U. Rank one projectors are the most common projectors. They project to one-dimensional subspaces of the vector space.

Projection operators are useful in quantum mechanics, where observables are described by operators. The effect of measuring an observable on a physical state vector is to turn this original vector instantaneously into another vector. This resulting vector is the orthogonal projection of the original vector down to some eigenspace of the operator associated with the observable.

6 Linear functionals and adjoint operators

When we consider a linear operator T on a vector space V that has an inner product, we can construct a related linear operator $T^\dagger$ on V called the **adjoint** of T. This is a very useful operator and is typically different from T. When the adjoint $T^\dagger$ happens to be equal to T, the operator is said to be *Hermitian*. To understand adjoints, we first need to develop the concept of a linear functional.

A **linear functional** ϕ on the vector space V is a linear map from V to the numbers $\mathbb{F}$: for $v \in V$, $\phi(v) \in \mathbb{F}$. A linear functional has the following two properties:

1. $\phi(v_1 + v_2) = \phi(v_1) + \phi(v_2)$, with $v_1, v_2 \in V$.

2. $\phi(av) = a\phi(v)$ for $v \in V$ and $a \in \mathbb{F}$.

As an example, consider the three-dimensional real vector space $\mathbb{R}^3$ with inner product equal to the familiar dot product. Writing a vector v as the triplet $v = (v_1, v_2, v_3)$, we take

$$\phi(v) \;=\; 3v_1 + 2v_2 - 4v_3\,. \tag{6.1}$$

Linearity is clear as the right-hand side features the components v_1, v_2, v_3 appearing linearly. We can use a vector $u = (3, 2, -4)$ to write the linear functional as an inner product. Indeed, one can readily see that

$$\phi(v) \;=\; \langle u, v\rangle. \tag{6.2}$$

This is no accident, in fact. We can prove that any linear functional $\phi(v)$ admits such representation with some suitable choice of vector u.

Theorem: Let ϕ be a linear functional on V. There is a unique vector $u \in V$ such that $\phi(v) = \langle u, v\rangle$ for all $v \in V$.

Proof: Consider an orthonormal basis, $(e_1, \ldots, e_n)$ and write the vector v as

$$v = \langle e_1, v\rangle e_1 + \ldots + \langle e_n, v\rangle e_n\,. \tag{6.3}$$

When ϕ acts on v we find, first by linearity and then by conjugate homogeneity

$$\begin{aligned}
\phi(v) \;&= \phi\big(\,\langle e_1, v\rangle e_1 + \ldots + \langle e_n, v\rangle e_n\big) \\
&= \langle e_1, v\rangle \phi(e_1) + \ldots + \langle e_n, v\rangle \phi(e_n) \\
&= \langle \phi(e_1)^* e_1, v\rangle + \ldots + \langle \phi(e_n)^* e_n\,, v\rangle \\
&= \big\langle \phi(e_1)^* e_1 + \ldots + \phi(e_n)^* e_n\,, v\big\rangle\,.
\end{aligned} \tag{6.4}$$

We have thus shown that, as claimed

$$\phi(v) = \langle u, v\rangle \quad \text{with} \quad u = \phi(e_1)^* e_1 + \ldots + \phi(e_n)^* e_n\,. \tag{6.5}$$

Next, we prove that this u is unique. If there exists another vector, u', that also gives the correct result for all v, then $\langle u', v\rangle = \langle u, v\rangle$, which implies $\langle u - u', v\rangle = 0$ for all v. Taking $v = u' - u$, we see that this shows $u' - u = 0$ or $u' = u$, proving uniqueness.[1]

We can modify a bit the notation when needed, to write

$$\phi_u(v) \;\equiv \langle u, v\rangle\,, \tag{6.6}$$

where the left-hand side makes it clear that this is a functional acting on v that depends on u.

We can now address the construction of the adjoint. Consider: $\phi(v) = \langle u, T\,v\rangle$, which is clearly a linear functional, whatever the operator T is. Since any linear functional can be written as $\langle w, v\rangle$, with some suitable vector w, we write

$$\langle u,\, Tv\rangle \;=\; \langle w\,, v\rangle\,, \tag{6.7}$$

[1] This theorem holds for infinite dimensional Hilbert spaces, for *continuous* linear functionals.

Of course, the vector w must depend on the vector u that appears on the left-hand side. Moreover, it must have something to do with the operator T, who does not appear anymore on the right-hand side. So we must look for some good notation here. We can think of w as a function of the vector u and thus write $w = T^\dagger u$ where $T^\dagger$ denotes a map (not obviously linear) from V to V. So, we think of $T^\dagger u$ as the vector obtained by acting with some function $T^\dagger$ on u. The above equation is written as

$$\langle u\,, Tv\rangle \;=\; \langle T^\dagger u\,, v\rangle\,, \tag{6.8}$$

Our next step is to show that, in fact, $T^\dagger$ is a linear operator on V. The operator $T^\dagger$ is called the **adjoint** of T. Consider

$$\langle u_1 + u_2, Tv\rangle \;=\; \langle T^\dagger(u_1+u_2), v\rangle\,, \tag{6.9}$$

and work on the left-hand side to get

$$\begin{aligned}\langle u_1 + u_2, Tv\rangle \;&=\; \langle u_1, Tv\rangle + \langle u_2, Tv\rangle \\ &=\; \langle T^\dagger u_1, v\rangle + \langle T^\dagger u_2, v\rangle \\ &=\; \left\langle T^\dagger u_1 + T^\dagger u_2\,, v\right\rangle.\end{aligned} \tag{6.10}$$

Comparing the right-hand sides of the last two equations we get the desired:

$$T^\dagger(u_1+u_2) \;=\; T^\dagger u_1 + T^\dagger u_2\,. \tag{6.11}$$

Having established linearity now we establish homogeneity. Consider

$$\langle\, au,\, Tv\rangle \;=\; \langle T^\dagger(au)\,,\, v\rangle\,. \tag{6.12}$$

The left hand side is

$$\langle\, au, Tv\rangle \;=\; a^\star\langle u, Tv\rangle \;=\; a^\star\langle T^\dagger u, v\rangle \;=\; \langle\, aT^\dagger u, v\rangle\,. \tag{6.13}$$

This time we conclude that

$$T^\dagger(au) \;=\; aT^\dagger u\,. \tag{6.14}$$

This concludes the proof that $T^\dagger$, so defined is a linear operator on V.

A couple of important properties are readily proven:

Claim: $(ST)^\dagger = T^\dagger S^\dagger$. We can show this as follows: $\langle u, STv\rangle = \langle S^\dagger u, Tv\rangle = \langle T^\dagger S^\dagger u, v\rangle$.

Claim: The adjoint of the adjoint is the original operator: $(S^\dagger)^\dagger = S$. We can show this as follows: $\langle u, S^\dagger v\rangle = \langle (S^\dagger)^\dagger u, v\rangle$. Now, additionally $\langle u, S^\dagger v\rangle = \langle S^\dagger v, u\rangle^* = \langle v, Su\rangle^* = \langle Su, v\rangle$. Comparing with the first result, we have shown that $(S^\dagger)^\dagger u = Su$, for any u, which proves the claim

Example: Let $v = (v_1, v_2, v_3)$, with $v_i \in \mathbb{C}$ denote a vector in the three-dimensional complex vector space, $\mathbb{C}^3$. Define a linear operator T that acts on v as follows:

$$T(v_1, v_2, v_3) \;=\; (\,0v_1 + 2v_2 + iv_3\,,\, v_1 - iv_2 + 0v_3\,,\, 3iv_1 + v_2 + 7v_3\,)\,. \tag{6.15}$$

Calculate the action of $T^\dagger$ on a vector. Give the matrix representations of T and $T^\dagger$ using the orthonormal basis $e_1 = (1,0,0), e_2 = (0,1,0), e_3 = (0,0,1)$. Assume the inner product is the standard on on $\mathbb{C}^3$.

Solution: We introduce a vector $u = (u_1, u_2, u_3)$ and will use the basic identity $\langle u, Tv\rangle = \langle T^\dagger u, v\rangle$. The left-hand side of the identity gives:

$$\langle\, u, Tv\,\rangle \;=\; u_1^*(2v_2 + iv_3) \;+\; u_2^*(v_1 - iv_2) \;+\; u_3^*(3iv_1 + v_2 + 7v_3)\,. \tag{6.16}$$

This is now rewritten by factoring the various v_i's

$$\langle\, u, Tv\,\rangle \;=\; (u_2^* + 3iu_3^*)v_1 + (2u_1^* - iu_2^* + u_3^*)v_2 + (iu_1^* + 7u_3^*)v_3\,. \tag{6.17}$$

Identifying the right-hand side with $\langle T^\dagger u, v\rangle$ we now deduce that

$$T^\dagger(u_1, u_2, u_3) \;=\; (\, u_2 - 3iu_3\,,\, 2u_1 + iu_2 + u_3\,,\, -iu_1 + 7u_3\,)\,. \tag{6.18}$$

This gives the action of $T^\dagger$. To find the matrix representation we begin with T. Using basis vectors, we have from (6.15)

$$Te_1 = T(1,0,0) = (0,1,3i) = e_2 + 3ie_3 = T_{11}e_1 + T_{21}e_2 + T_{31}e_3\,, \tag{6.19}$$

and deduce that $T_{11} = 0$, $T_{21} = 1$, $T_{31} = 3i$. This can be repeated, and the rule becomes clear quickly: the coefficients of v_i read left to right fit into the i-th column of the matrix. Thus, we have

$$T = \begin{pmatrix} 0 & 2 & i \\ 1 & -i & 0 \\ 3i & 1 & 7 \end{pmatrix} \qquad \text{and} \qquad T^\dagger = \begin{pmatrix} 0 & 1 & -3i \\ 2 & i & 1 \\ -i & 0 & 7 \end{pmatrix}\,. \tag{6.20}$$

These matrices are related: one is the transpose and complex conjugate of the other! This is not an accident.

Let us reframe this using matrix notation. Let $u = e_i$ and $v = e_j$ where e_i and e_j are orthonormal basis vectors. Then the definition $\langle u, Tv\rangle = \langle T^\dagger u, v\rangle$ can be written as

$$\begin{aligned} \langle T^\dagger e_i, e_j\rangle &= \langle e_i, Te_j\rangle \\ \langle T^\dagger_{ki} e_k, e_j\rangle &= \langle e_i, T_{kj} e_k\rangle \\ (T^\dagger_{ki})^* \delta_{kj} &= T_{jk}\delta_{ik} \\ (T^\dagger)^*_{ji} &= T_{ij} \end{aligned} \tag{6.21}$$

Relabeling i and j and taking the complex conjugate we find the familiar relation between a matrix and its adjoint:

$$(T^\dagger)_{ij} = (T_{ji})^*\,. \tag{6.22}$$

The adjoint matrix is the transpose and complex conjugate matrix only if we use an orthonormal basis. If we did not, in the equation above the use of $\langle e_i, e_j\rangle = \delta_{ij}$ would be replaced by $\langle e_i, e_j\rangle = g_{ij}$, where g_{ij} is some constant matrix that would appear in the rule for the construction of the adjoint matrix.

7 Hermitian and Unitary operators

Before we begin looking at special kinds of operators let us consider a very surprising fact about operators on complex vector spaces, as opposed to operators on real vector spaces.

Suppose we have an operator T that is such that for any vector $v \in V$ the following inner product vanishes

$$\langle\, v, Tv\rangle = 0 \quad \text{for all } v \in V. \tag{7.23}$$

What can we say about the operator T? The condition states that T is an operator that starting from a vector gives a vector orthogonal to the original one. In a two-dimensional real vector space, this is simply the operator that rotates any vector by ninety degrees! It is quite surprising and important that for *complex* vector spaces the result is very strong: any such operator T necessarily vanishes. This is a theorem:

Theorem: Let T be a linear operator in a **complex vector space** V**:**

$$\boxed{\text{If } \langle v\,, Tv\rangle = 0 \text{ for all } v \in V, \text{ then } T = 0.} \tag{7.24}$$

Proof: Any proof must be such that it fails to work for real vector space. Note that the result follows if we could prove that $\langle u, Tv\rangle = 0$, for all $u, v \in V$. Indeed, if this holds, then take $u = Tv$, then $\langle Tv, Tv\rangle = 0$ for all v implies that $Tv = 0$ for all v and therefore $T = 0$.
We will thus try to show that $\langle u\,, Tv\rangle = 0$ for all $u, v \in V$. All we know is that objects of the form $\langle \#, T\#\rangle$ vanish, whatever $\#$ is. So we must aim to form linear combinations of such terms in order to reproduce $\langle u\,, Tv\rangle$. We begin by trying the following

$$\langle u+v, T(u+v)\rangle \;-\; \langle u-v, T(u-v)\rangle \;=\; 2\langle u, Tv\rangle \;+\; 2\langle v, Tu\rangle\,. \tag{7.25}$$

We see that the "diagonal" term vanished, but instead of getting just $\langle u\,, Tv\rangle$ we also got $\langle v\,, Tu\rangle$. Here is where complex numbers help, we can get the same two terms but with opposite signs by trying,

$$\langle u+iv, T(u+iv)\rangle \;-\; \langle u-iv, T(u-iv)\rangle \;=\; 2i\langle u, Tv\rangle \;-\; 2i\,\langle v, Tu\rangle\,. \tag{7.26}$$

It follows from the last two relations that

$$\langle u\,, Tv\rangle \;=\; \frac{1}{4}\Big(\langle u+v, T(u+v)\rangle - \langle u-v, T(u-v)\rangle + \frac{1}{i}\langle u+iv, T(u+iv)\rangle - \frac{1}{i}\langle u-iv, T(u-iv)\rangle\Big)\,. \tag{7.27}$$

The condition $\langle v, Tv\rangle = 0$ for all v, implies that each term of the above right-hand side vanishes, thus showing that $\langle u\,, Tv\rangle = 0$ for all $u, v \in V$. As explained above this proves the result.

An operator T is said to be **Hermitian** if $T^\dagger = T$. Hermitian operators are pervasive in quantum mechanics. The above theorem in fact helps us discover Hermitian operators. It is familiar that the expectation value of a Hermitian operator, on any state, is real. It is also true, however, that any operator whose expectation value is real for all states must be Hermitian:

$$\boxed{T = T^\dagger \text{ if and only if } \langle v, Tv\rangle \in \mathbb{R} \text{ for all } v\,.} \tag{7.28}$$

To prove this first go from left to right. If $T = T^\dagger$

$$\langle v, Tv\rangle \;=\; \langle T^\dagger v, v\rangle \;=\; \langle Tv, v\rangle = \langle v, Tv\rangle^*\,, \tag{7.29}$$

showing that $\langle v, Tv\rangle$ is real. To go from right to left first note that the reality condition means that

$$\langle v, Tv\rangle \;=\; \langle Tv, v\rangle \;=\; \langle v, T^\dagger v\rangle\,, \tag{7.30}$$

where the last equality follows because $(T^\dagger)^\dagger = T$. Now the leftmost and rightmost terms can be combined to give $\langle v, (T - T^\dagger)v\rangle = 0$, which holding for all v implies, by the theorem, that $T = T^\dagger$.

We can prove two additional results of Hermitian operators rather easily. We have discussed earlier the fact that on a complex vector space any linear operator has at least one eigenvalue. Here we learn that the eigenvalues of a hermitian operator are real numbers. Moreover, while we have noted that eigenvectors corresponding to different eigenvalues are linearly independent, for Hermitian operators they are guaranteed to be orthogonal. Thus we have the following theorems
Theorem 1: The eigenvalues of Hermitian operators are real.
Theorem 2: Different eigenvalues of a Hermitian operator correspond to orthogonal eigenfunctions.

Proof 1: Let v be a nonzero eigenvector of the Hermitian operator T with eigenvalue λ: $Tv = \lambda v$. Taking the inner product with v we have that

$$\langle v, Tv\rangle = \langle v, \lambda v\rangle = \lambda\langle v, v\rangle\,. \tag{7.31}$$

Since T is hermitian, we can also evaluate $\langle v, Tv\rangle$ as follows

$$\langle v, Tv\rangle = \langle Tv, v\rangle = \langle \lambda v, v\rangle = \lambda^*\,\langle v, v\rangle\,. \tag{7.32}$$

The above equations give $(\lambda - \lambda^*)\langle v, v\rangle = 0$ and since v is not the zero vector, we conclude that $\lambda^* = \lambda$, showing the reality of λ.

Proof 2: Let v_1 and v_2 be eigenvectors of the operator T:

$$Tv_1 = \lambda_1 v_1, \qquad Tv_2 = \lambda_2 v_2\,, \tag{7.33}$$

with λ_1 and λ_2 real (previous theorem) and different from each other. Consider the inner product $\langle v_2, Tv_1\rangle$ and evaluate it in two different ways. First

$$\langle v_2, Tv_1\rangle = \langle v_2, \lambda_1 v_1\rangle = \lambda_1\langle v_2, v_1\rangle\,, \tag{7.34}$$

and second, using hermiticity of T,

$$\langle v_2, Tv_1\rangle = \langle Tv_2, v_1\rangle = \langle \lambda_2 v_2, v_1\rangle = \lambda_2\langle v_2, v_1\rangle\,. \tag{7.35}$$

From these two evaluations we conclude that

$$(\lambda_1 - \lambda_2)\langle v_1, v_2\rangle = 0 \tag{7.36}$$

and the assumption $\lambda_1 \neq \lambda_2$, leads to $\langle v_1, v_2\rangle = 0$, showing the orthogonality of the eigenvectors.

Let us now consider another important class of linear operators on a complex vector space, the so-called unitary operators. An operator $U \in \mathcal{L}(V)$ in a complex vector space V is said to be a **unitary operator** if it is surjective and does not change the magnitude of the vector it acts upon:

$$|Uu| = |u|\,,\ \text{for all } u \in V\,. \tag{7.37}$$

We tailored the definition to be useful even for infinite dimensional spaces. Note that U can only kill vectors of zero length, and since the only such vector is the zero vector, $\text{null}\, U = 0$, and U is injective. Since U is also assumed to be surjective, **a unitary operator U is always invertible.**

A simple example of a unitary operator is the operator λI with λ a complex number of unit-norm: $|\lambda| = 1$. Indeed $|\lambda Iu| = |\lambda u| = |\lambda||u| = |u|$ for all u. Moreover, the operator is clearly surjective.

For another useful characterization of unitary operators we begin by squaring (7.37)

$$\langle Uu, Uu\rangle = \langle u, u\rangle \tag{7.38}$$

By the definition of adjoint

$$\langle u, U^\dagger U\, u\rangle = \langle u, u\rangle \quad \rightarrow \quad \langle u\,,\, (U^\dagger U - I)u\rangle \;=\; 0 \;\;\text{for all}\;\; u\,. \tag{7.39}$$

So by our theorem $U^\dagger U = I$, and since U is invertible this means $U^\dagger$ is the inverse of U and we also have $UU^\dagger = I$:

$$\boxed{\;U^\dagger U \;=\; UU^\dagger \;=\; I\,.\;} \tag{7.40}$$

Unitary operators *preserve inner products* in the following sense

$$\langle\, Uu\,,\, Uv\rangle \;=\; \langle u\,, v\rangle\,. \tag{7.41}$$

This follows immediately by moving the second U to act on the first input and using $U^\dagger U = I$.

Assume the vector space V is finite dimensional and has an orthonormal basis $(e_1, \ldots e_n)$. Consider the new set of vectors $(f_1, \ldots, f_n)$ where the f's are obtained from the e's by the action of a unitary operator U:

$$f_i \;=\; U\, e_i\,. \tag{7.42}$$

This also means that $e_i = U^\dagger f_i$. We readily see that the f's are also a basis, because they are linearly independent: Acting on $a_1 f_1 + \ldots + a_n f_n = 0$ with $U^\dagger$ we find $a_1 e_1 + \ldots + a_n e_n = 0$, and thus $a_i = 0$. We now see that the new basis is also orthonormal:

$$\langle f_i \, , f_j \rangle \;=\; \langle \, U e_i \, , \, U e_j \rangle \;=\; \langle e_i \, , \, e_j \rangle \;=\; \delta_{ij} \, . \tag{7.43}$$

The matrix elements of U in the e-basis are

$$U_{ki} \;=\; \langle e_k \, , U e_i \rangle \, . \tag{7.44}$$

Let us compute the matrix elements U'_{ki} of U in the f-basis

$$U'_{ki} \;=\; \langle f_k \, , U f_i \rangle \;=\; \langle U e_k \, , U f_i \rangle \;=\; \langle e_k \, , f_i \rangle \;=\; \langle e_k \, , U e_i \rangle \;=\; U_{ki} \tag{7.45}$$

The matrix elements are the same! Can you find an explanation for this result?

DIRAC's BRA AND KET NOTATION

B. Zwiebach

October 7, 2013

Contents

1 From inner products to bra-kets

Dirac invented a useful alternative notation for inner products that leads to the concepts of bras and kets. The notation is sometimes more efficient than the conventional mathematical notation we have been using. It is also widely although not universally used. It all begins by writing the inner product differently. The rule is to turn inner products into bra-ket pairs as follows

$$\langle\, u\,, v\,\rangle \quad \longrightarrow \quad \langle u|\, v\rangle\,. \tag{1.1}$$

Instead of the inner product comma we simply put a vertical bar! We can translate our earlier discussion of inner products trivially. In order to make you familiar with the new look we do it.

We now write $\langle u|v\rangle \;=\; \langle v|u\rangle^*$, as well as $\langle v|v\rangle \geq 0$ for all v, while $\langle v|v\rangle = 0$ if and only if $v = 0$. We have linearity in the second argument

$$\langle u|c_1 v_1 + c_2 v_2\rangle \;=\; c_1\langle u|v_1\rangle + c_2\langle u|v_2\rangle\,, \tag{1.2}$$

for complex constants c_1 and c_2, but antilinearity in the first argument

$$\langle c_1 u_1 + c_2 u_2|v\rangle \;=\; c_1^*\,\langle u_1|v\rangle + c_2^*\,\langle u_2|v\rangle\,. \tag{1.3}$$

Two vectors u and v for which $\langle u|v\rangle = 0$ are orthogonal. For the norm: $|v|^2 \;=\; \langle v|v\rangle$. The Schwarz inequality, for any pair u and v of vectors reads $|\langle u|v\rangle| \leq |u|\,|v|$.

For a given physical situation, the inner product must be defined and should satisfy the axioms. Let us consider two examples:

1. Let $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ and $b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ be two vectors in a complex dimensional vector space of dimension two. We then define

$$\langle a|b\rangle \;\equiv\; a_1^* b_1 + a_2^* b_2\,. \tag{1.4}$$

   You should confirm the axioms are satisfied.

2. Consider the complex vector space of complex function $f(x) \in \mathbb{C}$ with $x \in [0, L]$. Given two such functions $f(x), g(x)$ we define

$$\langle f|g\rangle \;\equiv\; \int_0^L f^*(x) g(x) dx\,. \tag{1.5}$$

   The verification of the axioms is again quite straightforward.

A set of basis vectors $\{e_i\}$ labelled by the integers $i = 1, \ldots, n$ satisfying

$$\langle e_i|e_j\rangle \;=\; \delta_{ij}\,, \tag{1.6}$$

is orthonormal. An arbitrary vector can be written as a linear superposition of basis states:

$$v \;=\; \sum_i \alpha_i\, e_i\,, \tag{1.7}$$

We then see that the coefficients are determined by the inner product

$$\langle e_k|v\rangle \;=\; \langle e_k|\sum_i \alpha_i e_i\rangle \;=\; \sum_i \alpha_i \langle e_k|e_i\rangle \;=\; \alpha_k\,. \tag{1.8}$$

We can therefore write

$$v \;=\; \sum_i e_i \langle e_i|v\rangle\,. \tag{1.9}$$

To obtain now bras and kets, we reinterpret the inner product. We want to "split" the inner product into two ingredients

$$\langle u|v\rangle \quad\to\quad \langle u|\;\;|v\rangle\,. \tag{1.10}$$

Here $|v\rangle$ is called a **ket** and $\langle u|$ is called a **bra**. We will view the ket $|v\rangle$ just as another way to represent the vector v. This is a small subtlety with the notation: we think of $v \in V$ as a vector and also $|v\rangle \in V$ as a vector. It is as if we added some decoration $|\;\rangle$ around the vector v to make it clear by inspection that it is a vector, perhaps like the usual top arrows that are added in some cases. The label in the ket is a vector and the ket itself is that vector!

Bras are somewhat different objects. We say that bras belong to the space V^* dual to V. Elements of V^* are linear maps from V to $\mathbb{C}$. In conventional mathematical notation one has a $v \in V$ and a linear function $\phi \in V^*$ such that $\phi(v)$, which denotes the action of the function of the vector v, is a number. In the bracket notation we have the replacements

$$\begin{aligned} v \;&\to |v\rangle\,, \\ \phi \;&\to \langle u|\,, \\ \phi_u(v) \;&\to \langle u|v\rangle\,, \end{aligned} \tag{1.11}$$

where we used the notation in (6.6). Our bras are labelled by vectors: the object inside the $\langle\ |$ is a vector. But bras are *not* vectors. If kets are viewed as column vectors, then bras are viewed as row vectors. In this way a bra to the left of a ket makes sense: matrix multiplication of a row vector times a column vector gives a number. Indeed, for vectors

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} , \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \tag{1.12}$$

we had

$$\langle a|b\rangle \ = \ a_1^* b_1 + a_2^* b_2 + \ldots a_n^* b_n \tag{1.13}$$

Now we think of this as

$$\langle a| = \left(a_1^*, a_2^* \ldots, a_n^*\right) , \quad |b\rangle = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \tag{1.14}$$

and matrix multiplication gives us the desired answer

$$\langle a|b\rangle \ = \ \left(a_1^*, a_2^* \ldots, a_n^*\right) \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \ = \ a_1^* b_1 + a_2^* b_2 + \ldots a_n^* b_n \, . \tag{1.15}$$

Note that the bra labeled by the vector a is obtained by forming the row vector and complex conjugating the entries. More abstractly the bra $\langle u|$ labeled by the vector u is defined by its action on arbitrary vectors $|v\rangle$ as follows

$$\langle u| \ : \ |v\rangle \ \to \ \langle u|v\rangle \, . \tag{1.16}$$

As required by the definition, any linear map from V to $\mathbb{C}$ defines a bra, and the corresponding underlying vector. For example let v be a generic vector:

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} , \tag{1.17}$$

A linear map $f(v)$ that acting on a vector v gives a number is an expression of the form

$$f(v) \ = \ \alpha_1^* \, v_1 + \alpha_2^* \, v_2 + \ldots \alpha_n^* \, v_n \, . \tag{1.18}$$

It is a linear function of the components of the vector. The linear function is specified by the numbers α_i, and for convenience (and without loss of generality) we used their complex conjugates. Note that we need exactly n constants, so they can be used to assemble a row vector or a bra

$$\langle \alpha| \ = \ \left(\alpha_1^* \, , \alpha_2^* \, , \ldots \, , \alpha_n^*\right) \tag{1.19}$$

and the associated vector or ket

$$|\alpha\rangle \;=\; \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \tag{1.20}$$

Note that, by construction

$$f(v) = \langle\alpha|v\rangle\,. \tag{1.21}$$

This illustrates the point that (i) bras represent dual objects that act on vectors and (ii) bras are labelled by vectors.

Bras can be added and can be multiplied by complex numbers and there is a zero bra defined to give zero acting on any vector, so V^* is also a complex vector space. As a bra, the linear superposition

$$\langle\omega| \equiv \alpha\langle a| + \beta\langle b| \;\in\; V^*\,, \quad \alpha,\beta \in \mathbb{C}\,, \tag{1.22}$$

is defined to act on a vector (ket) $|c\rangle$ to give the number

$$\alpha\langle a|c\rangle + \beta\langle b|c\rangle\,. \tag{1.23}$$

For any vector $|v\rangle \in V$ there is a *unique* bra $\langle v| \in V^*$. If there would be another bra $\langle v'|$ it would have to act on arbitrary vectors $|w\rangle$ just like $\langle v|$:

$$\langle v'|w\rangle \;=\; \langle v|w\rangle \quad\to\quad \langle w|v\rangle - \langle w|v'\rangle \;=\; 0 \quad\to\quad \langle w|v-v'\rangle \;=\; 0\,. \tag{1.24}$$

In the first step we used complex conjugation and in the second step linearity. Now the vector $v - v'$ must have zero inner product with *any* vector w, so $v - v' = 0$ and $v = v'$.

We can now reconsider equation (1.3) and write an extra right-hand side

$$\langle\alpha_1 a_1 + \alpha_2 a_2|b\rangle \;=\; \alpha_1^*\langle a_1|b\rangle + \alpha_2^*\langle a_2|b\rangle \;=\; \left(\alpha_1^*\langle a_1| + \alpha_2^*\langle a_2|\right)|b\rangle \tag{1.25}$$

so that we conclude that the rules to pass from kets to bras include

$$\boxed{|v\rangle = \alpha_1|a_1\rangle + \alpha_2|a_2\rangle \quad\longleftrightarrow\quad \langle v| = \alpha_1^*\langle a_1| + \alpha_2^*\langle a_2|\,.} \tag{1.26}$$

For simplicity of notation we sometimes write kets with labels simpler than vectors. Let us reconsider the basis vectors $\{e_i\}$ discussed in (1.6). The ket $|e_i\rangle$ is simply called $|i\rangle$ and the orthonormal condition reads

$$\langle i|j\rangle \;=\; \delta_{ij}\,. \tag{1.27}$$

The expansion (1.7) of a vector now reads

$$|v\rangle \;=\; \sum_i |i\rangle\alpha_i\,, \tag{1.28}$$

As in (1.8) the expansion coefficients are $\alpha_k \;=\; \langle k|v\rangle$ so that

$$\boxed{|v\rangle \;=\; \sum_i |i\rangle\langle i|v\rangle\,.} \tag{1.29}$$

2 Operators revisited

Let T be an operator in a vector space V. This means that acting on vectors on V it gives vectors on V, something we write as

$$\Omega : V \to V\,. \tag{2.30}$$

We denote by $\Omega|a\rangle$ the vector obtained by acting with Ω on the vector $|a\rangle$:

$$|a\rangle \in V \;\to\; \Omega|a\rangle \in V\,. \tag{2.31}$$

The operator Ω is linear if additionally we have

$$\Omega\big(|a\rangle + |b\rangle\big) \;=\; \Omega|a\rangle + \Omega|b\rangle\,, \quad \text{and} \quad \Omega(\alpha|a\rangle) \;=\; \alpha\,\Omega|a\rangle\,. \tag{2.32}$$

When kets are labeled by vectors we sometimes write

$$|\Omega\, a\rangle \;\equiv\; \Omega|a\rangle, \tag{2.33}$$

It is useful to note that a linear operator on V is also a linear operator on V^*

$$\Omega : V^* \to V^*\,, \tag{2.34}$$

We write this as

$$\langle a| \;\to\; \langle a|\Omega \in V^*\,. \tag{2.35}$$

The object $\langle a|\Omega$ is defined to be the bra that acting on the ket $|b\rangle$ gives the number $\langle a|\Omega|b\rangle$.

We can write operators in terms of bras and kets, written in a suitable order. As an example of an operator consider a bra $\langle a|$ and a ket $|b\rangle$. We claim that the object

$$\Omega \;=\; |a\rangle\langle b|\,, \tag{2.36}$$

is naturally viewed as a linear operator on V and on V^*. Indeed, acting on a vector we let it act as the bra-ket notation suggests:

$$\Omega|v\rangle \;\equiv\; |a\rangle\,\langle b|v\rangle \sim |a\rangle\,, \quad \text{since } \langle b|v\rangle \text{ is a number.} \tag{2.37}$$

Acting on a bra it gives a bra:

$$\langle w|\Omega \;\equiv\; \langle w|a\rangle\,\langle b| \sim \langle b|\,, \quad \text{since } \langle w|a\rangle \text{ is a number.} \tag{2.38}$$

Let us now review the description of operators as matrices. The choice of basis is ours to make. For simplicity, however, we will usually consider orthonormal bases.

Consider therefore, two vectors expanded in an orthonormal basis $\{|i\rangle\}$:

$$|a\rangle \;=\; \sum_n |n\rangle a_n\,, \qquad |b\rangle \;=\; \sum_n |n\rangle b_n\,. \tag{2.39}$$

Assume $|b\rangle$ is obtained by the action of Ω on $|a\rangle$:

$$\Omega|a\rangle \;=\; |b\rangle \quad\to\quad \sum_n \Omega|n\rangle a_n \;=\; \sum_n |n\rangle b_n\,. \tag{2.40}$$

Acting on both sides of this vector equation with the bra $\langle m|$ we find

$$\sum_n \langle m|\Omega|n\rangle a_n \;=\; \sum_n \langle m|n\rangle b_n \;=\; b_m \tag{2.41}$$

We now define the 'matrix elements'

$$\boxed{\Omega_{mn} \;\equiv\; \langle m|\Omega|n\rangle\,.} \tag{2.42}$$

so that the above equation reads

$$\sum_n \Omega_{mn} a_n \;=\; b_m\,, \tag{2.43}$$

which is the matrix version of the original relation $\Omega|a\rangle \;=\; |b\rangle$. The chosen basis has allowed us to view the linear operator Ω as a matrix, also denoted as Ω, with matrix components Ω_{mn}:

$$\Omega \;\longleftrightarrow\; \begin{pmatrix} \Omega_{11} & \Omega_{12} & \dots & \dots & \Omega_{1N} \\ \Omega_{21} & \Omega_{22} & \dots & \dots & \Omega_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \Omega_{N1} & \Omega_{N2} & \dots & \dots & \Omega_{NN} \end{pmatrix}, \quad \text{with } \Omega_{ij} \;=\; \langle i|\Omega|j\rangle\,. \tag{2.44}$$

There is one additional claim. The operator itself can be written in terms of the matrix elements and basis bras and kets. We claim that

$$\boxed{\Omega \;=\; \sum_{m,n} |m\rangle \Omega_{mn} \langle n|\,.} \tag{2.45}$$

We can verify that this is correct by computing the matrix elements using it:

$$\langle m'|\Omega|n'\rangle \;=\; \sum_{m,n} \Omega_{mn} \langle m'|m\rangle\, \langle n|n'\rangle \;=\; \sum_{m,n} \Omega_{mn} \delta_{m'm}\, \delta_{nn'} \;=\; \Omega_{m'n'}\,, \tag{2.46}$$

as expected from the definition (2.42).

2.1 Projection Operators

Consider the familiar orthonormal basis $\{|i\rangle\}$ of V and choose one element $|m\rangle$ from the basis to form an operator P_m defined by

$$P_m \;\equiv\; |m\rangle\langle m|\,. \tag{2.47}$$

This operator maps any vector $|v\rangle \in V$ to a vector along $|,\rangle$. Indeed, acting on $|v\rangle$ it gives

$$P_m|v\rangle \;=\; |m\rangle\langle m|v\rangle \;\sim\; |m\rangle\,. \tag{2.48}$$

Comparing the above expression for P_m with (2.45) we see that in the chosen basis, P_n is represented by a matrix all of whose elements are zero, except for the (n,n) element $(P_n)_{nn}$ which is one:

$$P_n \quad \longleftrightarrow \quad \begin{pmatrix} 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \dots & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 \end{pmatrix} . \tag{2.49}$$

A hermitian operator P is said to be a *projection* operator if it satisfies the operator equation $PP = P$. This means that acting twice with a projection operator on a vector gives the same as acting once. The operator P_m is a projection operator since

$$P_m P_m \;=\; \big(|m\rangle\langle m|\big)\big(|m\rangle\langle m|\big) \;=\; |m\rangle\,\langle m|m\rangle\,\langle m| \;=\; |m\rangle\langle m|\,, \tag{2.50}$$

since $\langle m|m\rangle = 1$. The operator P_m is said to be a *rank one* projection operator since it projects to a one-dimensional subspace of V, the subspace generated by $|m\rangle$.

Using the basis vector $|m\rangle$ with $m \neq n$ we can define

$$P_{m,n} \;\equiv\; |m\rangle\langle m| + |n\rangle\langle n|\,. \tag{2.51}$$

Acting on any vector $|v\rangle \in V$, this operator gives us a vector in the subspace spanned by $|m\rangle$ and $|n\rangle$:

$$P_{m,n}|v\rangle \;=\; |m\rangle\,\langle m|v\rangle + |n\rangle\,\langle n|v\rangle\,. \tag{2.52}$$

Using the orthogonality of $|m\rangle$ and $|n\rangle$ we quickly find that $P_{m,n}P_{m,n} = P_{m,n}$ and therefore $P_{m,n}$ is a projector. It is a rank two projector, since it projects to a two-dimensional subspace of V, the subspace spanned by $|m\rangle$ and $|n\rangle$. Similarly, we can construct a rank three projector by adding an extra term $|k\rangle\langle k|$ with $k \neq m$ and $k \neq n$. If we include all basis vectors we would have the operator

$$P_{1,\dots,N} \;\equiv\; |1\rangle\langle 1| + |2\rangle\langle 2| + \dots + |N\rangle\langle N|\,. \tag{2.53}$$

As a matrix $P_{1,\dots,N}$ has a one on every element of the diagonal and a zero everywhere else. This is therefore the unit matrix, which represents the identity operator. Indeed we anticipated this in (1.29), and we thus write

$$\boxed{\mathbf{1} \;=\; \sum_i |i\rangle\langle i|\,.} \tag{2.54}$$

This is the completeness relation for the chosen orthonormal basis. This equation is sometimes called the 'resolution' of the identity.

Example. For the spin one-half system the unit operator can be written as a sum of two terms since the vector space is two dimensional. Using the orthonormal basis vectors $|+\rangle$ and $|-\rangle$ for spins along the positive and negative z directions, respectively, we have

$$\mathbf{1} \;=\; |+\rangle\langle +| \;+\; |-\rangle\langle -|\,. \tag{2.55}$$

Example. We can use the completeness relation to show that our formula (2.42) for matrix elements is consistent with matrix multiplication. Indeed for the product $\Omega_1\Omega_2$ of two operators we write

$$\begin{aligned}(\Omega_1\Omega_2)_{mn} &= \langle m|\Omega_1\Omega_2|n\rangle = \langle m|\Omega_1\,\mathbf{1}\,\Omega_2|n\rangle \\ &= \langle m|\Omega_1\Big(\sum_{k=1}^{N}|k\rangle\langle k|\Big)\Omega_2|n\rangle = \sum_{k=1}^{N}\langle m|\Omega_1|k\rangle\langle k|\Omega_2|n\rangle = \sum_{k=1}^{N}(\Omega_1)_{mk}(\Omega_2)_{kn}\,.\end{aligned} \tag{2.56}$$

This is the expected rule for the multiplication of the matrices corresponding to Ω_1 and Ω_2.

2.2 Adjoint of a linear operator

A linear operator Ω on V is defined by its action on the vectors in V. We have noted that Ω can also be viewed as a linear operator on the dual space V^*. We defined the linear operator $\Omega^\dagger$ associated with Ω. In general $\Omega^\dagger$ by

$$\langle\Omega^\dagger u|v\rangle = \langle u|\Omega v\rangle \tag{2.57}$$

Flipping the order on the left-hand side we get

$$\langle v|\Omega^\dagger u\rangle^* = \langle u|\Omega v\rangle \tag{2.58}$$

Complex conjugating, and writing the operators more explicitly

$$\boxed{\langle v|\Omega^\dagger|u\rangle = \langle u|\Omega|v\rangle^*\,,\quad \forall u, v\,.} \tag{2.59}$$

Flipping the two sides of (2.57) we also get

$$\langle v|\Omega^\dagger|u\rangle = \langle\Omega v|u\rangle \tag{2.60}$$

from which, taking the ket away, we learn that

$$\boxed{\langle v|\Omega^\dagger \equiv \langle\Omega v|\,.} \tag{2.61}$$

Another way to state the action of the operator $\Omega^\dagger$ is as follows. The linear operator Ω induces a map $|v\rangle \to |v'\rangle$ of vectors in V and, in fact, is defined by giving a complete list of these maps. The operator $\Omega^\dagger$ is defined as the one that induces the maps $\langle v| \to \langle v'|$ of the *corresponding* bras. Indeed,

$$\begin{aligned}|v'\rangle &= |\Omega v\rangle = \Omega|v\rangle\,, \\ \langle v'| &= \langle\Omega v| = \langle v|\Omega^\dagger\end{aligned} \tag{2.62}$$

The first line is just definitions. On the second line, the first equality is obtained by taking bras of the first equality on the first line. The second equality is just (2.61). We say it as

$$\boxed{\text{The bra associated with } \Omega|v\rangle \text{ is } \langle v|\Omega^\dagger\ .} \tag{2.63}$$

To see what hermiticity means at the level of matrix elements, we take u, v to be orthonormal basis vectors in (2.59)

$$\langle i|\Omega^\dagger|j\rangle \;=\; \langle j|\Omega|i\rangle^* \quad\to\quad (\Omega^\dagger)_{ij} \;=\; (\Omega_{ji})^* \,. \tag{2.64}$$

In matrix notation we have $\Omega^\dagger = (\Omega^t)^*$ where the superscript t denotes transposition.

Exercise. Show that $(\Omega_1\Omega_2)^\dagger = \Omega_2^\dagger\Omega_1^\dagger$ by taking matrix elements.
Exercise. Given an operator $\Omega = |a\rangle\langle b|$ for arbitrary vectors a, b, write a bra-ket expression for $\Omega^\dagger$.
Solution: Acting with Ω on $|v\rangle$ and then taking the dual gives

$$\Omega|v\rangle \;=\; |a\rangle\,\langle b|v\rangle \quad\to\; \langle v|\Omega^\dagger \;=\; \langle v|b\rangle\,\langle a|\,, \tag{2.65}$$

Since this equation is valid for any bra $\langle v|$ we read

$$\Omega^\dagger \;=\; |b\rangle\langle a|\,. \tag{2.66}$$

2.3 Hermitian and Unitary Operators

A linear operator Ω is said to be *hermitian* if it is equal to its adjoint:

$$\boxed{\text{Hermitian Operator:}\;\; \Omega^\dagger \;=\; \Omega\,.} \tag{2.67}$$

In quantum mechanics Hermitian operators are associated with observables. The eigenvalues of a Hermitian operator are the possible measured values of the observables. As we will show soon, the eigenvalues of a Hermitian operator are all real. An operator A is said to be *anti*-hermitian if $A^\dagger = -A$.

Exercise: Show that the commutator $[\Omega_1, \Omega_2]$ of two hermitian operators Ω_1 and Ω_2 is anti-hermitian.

There are a couple of equations that rewrite in useful ways the main property of Hermitian operators. Using $\Omega^\dagger = \Omega$ in (2.59) we find

$$\boxed{\text{If } \Omega \text{ is a Hermitian Operator:}\;\; \langle v|\Omega|u\rangle \;=\; \langle u|\Omega|v\rangle^*\,,\quad \forall u, v\,.} \tag{2.68}$$

It follows that the expectation value of a Hermitian operator in *any* state is real

$$\langle v|\Omega|v\rangle \;\text{ is real for any hermitian }\; \Omega\,. \tag{2.69}$$

Another neat form of the hermiticity condition is derived as follows:

$$\langle \Omega u|v\rangle \;=\; \langle u|\Omega^\dagger|v\rangle \;=\; \langle u|\Omega|v\rangle \;=\; \langle u|\Omega v\rangle\,, \tag{2.70}$$

so that all in all

$$\boxed{\text{Hermitian Operator:}\;\; \langle \Omega u|v\rangle \;=\; \langle u|\Omega v\rangle\,.} \tag{2.71}$$

In this expression we see that a hermitian operator moves freely from the bra to the ket (and viceversa).
Example: For wavefunction $f(x) \in \mathbb{C}$ we have written

$$\langle f|g\rangle \;=\; \int_{-\infty}^{\infty} (f(x))^* g(x) dx \tag{2.72}$$

For a Hermitian Ω we have $\langle \Omega f|g\rangle = \langle f|\Omega g\rangle$ or explicitly

$$\int_{-\infty}^{\infty} (\Omega f(x))^* g(x) dx \;=\; \int_{-\infty}^{\infty} (f(x))^* \Omega g(x) dx \tag{2.73}$$

Verify that the linear operator $\Omega = \frac{\hbar}{i}\frac{d}{dx}$ is hermitian when we restrict to functions that vanish at $\pm\infty$.

An operator U is said to be a unitary operator if $U^\dagger$ is an inverse for U, that is, $U^\dagger U$ and $UU^\dagger$ are both the identity operator:

$$\boxed{U \text{ is a unitary operator:} \quad U^\dagger U \;=\; UU^\dagger \;=\; \mathbf{1}} \tag{2.74}$$

In finite dimensional vector spaces $U^\dagger U = \mathbf{1}$ implies $UU^\dagger = 1$, but this is not always the case for infinite dimensional vector spaces. A key property of unitary operators is that they preserve the norm of states. Indeed, assume that $|\psi'\rangle$ is obtained by the action of U on $|\psi\rangle$:

$$|\psi'\rangle = U|\psi\rangle \tag{2.75}$$

Taking the dual we have

$$\langle\psi'| = \langle\psi|U^\dagger\,, \tag{2.76}$$

and therefore

$$\langle\psi'|\psi'\rangle \;=\; \langle\psi|U^\dagger U|\psi\rangle \;=\; \langle\psi|\psi\rangle\,, \tag{2.77}$$

showing that $|\psi\rangle$ and $U|\psi\rangle$ are states with the same norm. More generally

$$\boxed{\langle Ua|Ub\rangle \;=\; \langle a|U^\dagger U|b\rangle \;=\; \langle a|b\rangle\,.} \tag{2.78}$$

Another important property of unitary operators is that acting on an orthonormal basis they give another orthonormal basis. To show this consider the orthonormal basis

$$|a_1\rangle\,, |a_2\rangle\,, \ldots |a_N\rangle\,, \qquad \langle a_i|a_j\rangle = \delta_{ij} \tag{2.79}$$

Acting with U we get

$$|Ua_1\rangle\,, |Ua_2\rangle\,, \ldots |Ua_N\rangle\,, \tag{2.80}$$

To show that this is a basis we must prove that

$$\sum_i \beta_i |Ua_i\rangle \;=\; 0 \tag{2.81}$$

implies $\beta_i = 0$ for all i. Indeed, the above gives

$$\sum_i \beta_i |Ua_i\rangle \;=\; \sum_i \beta_i U|a_i\rangle \;=\; U\sum_i \beta_i |a_i\rangle \;=\; 0\,. \tag{2.82}$$

Acting with $U^\dagger$ from the left we find that $\sum_i \beta_i|a_i\rangle = 0$ and, since the $|a_i\rangle$ form a basis, we get $\beta_i = 0$ for all i, as desired. The new basis is orthonormal because

$$\langle Ua_i|Ua_j\rangle \;=\; \langle a_i|U^\dagger U|a_j\rangle = \langle a_i|a_j\rangle \;=\; \delta_{ij}\,. \tag{2.83}$$

It follows from the above that the operator U can be written as

$$U \;=\; \sum_{i=1}^{N} |Ua_i\rangle\langle a_i|\,, \tag{2.84}$$

since

$$U|a_j\rangle = \sum_{i=1}^{N} |Ua_i\rangle\langle a_i|a_j\rangle = |Ua_i\rangle\,. \tag{2.85}$$

In fact for *any* unitary operator U in a vector space V there exist orthonormal bases $\{|a_i\rangle\}$ and $\{|b_i\rangle\}$ such that U can be written as

$$U \;=\; \sum_{i=1}^{N} |b_i\rangle\langle a_i|\,. \tag{2.86}$$

Indeed, this is just a rewriting of (2.84), with $|a_i\rangle$ any orthonormal basis and $|b_i\rangle = |Ua_i\rangle$.
Exercise: Verify that U in (2.86) satisfies $U^\dagger U = UU^\dagger = \mathbf{1}$.
Exercise: Prove that $\langle a_i|U|a_j\rangle \;=\; \langle b_i|U|b_j\rangle\,.$

3 Non-denumerable basis

In this section we describe the use of bras and kets for the position and momentum states of a particle moving on the real line $x \in \mathbb{R}$.

Let us begin with position. We will introduce position states $|x\rangle$ where the label x in the ket is the value of the position. Since x is a continuous variable and we position states $|x\rangle$ for all values of x to form a basis, we are dealing with an infinite basis that is not possible to label as $|1\rangle, |2\rangle, \ldots$, it is a non-denumerable basis. So we have

$$\text{Basis states}: \quad |x\rangle\,, \quad \forall x \in \mathbb{R}\,. \tag{3.87}$$

Basis states with different values of x are different vectors in the state space (a complex vector space, as always in quantum mechanics). Note here that the label on the ket is not a vector! So $|ax\rangle \neq a|x\rangle$, for any real $a \neq 1$. In particular $|-x\rangle \neq |x\rangle$ unless $x = 0$. For quantum mechanics in three dimensions, we have position states $|\vec{x}\,\rangle$. Here the label is a vector in a three-dimensional real vector space (our space!) while the ket is a vector in the infinite dimensional complex vector space of states of the theory.

Again something like $|\vec{x}_1 + \vec{x_2}\,\rangle$ has nothing to do with $|\vec{x}_1\rangle + |\vec{x}_2\rangle$. The $|\ \rangle$ enclosing the label of the position eigenstates plays a crucial role: it helps us see that object lives in an infinite dimensional complex vector space.

The inner product must be defined, so we will take

$$\langle x|y\rangle \ = \ \delta(x-y)\,. \tag{3.88}$$

It follows that position states with different positions are orthogonal to each other. The norm of a position state is infinite: $\langle x|x\rangle = \delta(0) = \infty$, so these are not allowed states of particles. We visualize the state $|x\rangle$ as the state of a particle perfectly localized at x, but this is an idealization. We can easily construct normalizable states using superpositions of position states. We also have a completeness relation

$$\mathbf{1} \ = \ \int dx\,|x\rangle\langle x|\,. \tag{3.89}$$

This is consistent with our inner product above. Letting the above equation act on $|y\rangle$ we find an equality:

$$|y\rangle \ = \ \int dx\,|x\rangle\langle x|y\rangle \ = \ \int dx\,|x\rangle\,\delta(x-y) \ = \ |y\rangle\,. \tag{3.90}$$

The position operator $\hat{x}$ is defined by its action on the position states. Not surprisingly we let

$$\hat{x}\,|x\rangle \ = \ x\,|x\rangle\,, \tag{3.91}$$

thus declaring that $|x\rangle$ are $\hat{x}$ eigenstates with eigenvalue equal to the position x. We can also show that $\hat{x}$ is a Hermitian operator by checking that $\hat{x}^\dagger$ and $\hat{x}$ have the same matrix elements:

$$\langle x_1|\hat{x}^\dagger|x_2\rangle = \langle x_2|\hat{x}|x_1\rangle^* = [x_1\delta(x_1-x_2)]^* = x_2\delta(x_1-x_2) = \langle x_1|\hat{x}|x_2\rangle\,. \tag{3.92}$$

We thus conclude that $\hat{x}^\dagger = \hat{x}$ and the bra associated with (3.91) is

$$\langle x|\hat{x} = x\langle x|\,. \tag{3.93}$$

Given the state $|\psi\rangle$ of a particle, we define the associated position-state wavefunction $\psi(x)$ by

$$\psi(x) \equiv \langle x|\psi\rangle \ \in \mathbb{C}\,. \tag{3.94}$$

This is sensible: $\langle x|\psi\rangle$ is a number that depends on the value of x, thus a function of x. We can now do a number of basic computations. First we write any state as a superposition of position eigenstates, by inserting $\mathbf{1}$ as in the completeness relation

$$|\psi\rangle \ = \ \mathbf{1}|\psi\rangle \ = \ \int dx\,|x\rangle\,\langle x|\psi\rangle \ = \ \int dx\,|x\rangle\,\psi(x)\,. \tag{3.95}$$

As expected, $\psi(x)$ is the component of ψ along the state $|x\rangle$. Ovelap of states can also be written in position space:

$$\langle\phi|\psi\rangle \ = \ \int dx\,\langle\phi|x\rangle\,\langle x|\psi\rangle \ = \ \int dx\ \phi^*(x)\psi(x)\,. \tag{3.96}$$

Matrix elements involving $\hat{x}$ are also easily evaluated

$$\langle\phi|\hat{x}|\psi\rangle \;=\; \langle\phi|\hat{x}\mathbf{1}|\psi\rangle \;=\; \int dx\,\langle\phi|\hat{x}|x\rangle\,\langle x|\psi\rangle \;=\; \int dx\,\langle\phi|x\rangle\,x\,\langle x|\psi\rangle \;=\; \int dx\;\phi^*(x)\,x\,\psi(x)\,. \tag{3.97}$$

We now introduce momentum states $|p\rangle$ that are eigenstates of the momentum operator $\hat{p}$ in complete analogy to the position states

$$\begin{aligned}
&\text{Basis states}: \quad |p\rangle\,, \quad \forall p\in\mathbb{R}\,.\\
&\langle p'|p\rangle \;=\; \delta(p-p')\,,\\
&\mathbf{1} \;=\; \int dp\,|p\rangle\langle p|\,,\\
&\hat{p}\,|p\rangle \;=\; p\,|p\rangle
\end{aligned} \tag{3.98}$$

Just as for coordinate space we also have

$$\hat{p}^\dagger = \hat{p}\,, \quad \text{and} \qquad \langle p|\hat{p} = p\langle p|\,. \tag{3.99}$$

In order to relate the two bases we need the value of the overlap $\langle x|p\rangle$. Since we interpret this as the wavefunction for a particle with momentum p we have from (6.39) of Chapter 1 that

$$\langle x|p\rangle \;= \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}}\;. \tag{3.100}$$

The normalization was adjusted properly to be compatible with the completeness relations. Indeed, for example, consider the $\langle p'|p\rangle$ overlap and use the completeness in x to evaluate it

$$\langle p'|p\rangle \;=\; \int dx\langle p'|x\rangle\langle x|p\rangle \;=\; \frac{1}{2\pi\hbar}\int dx e^{i(p-p')x/\hbar} \;=\; \frac{1}{2\pi}\int du\,e^{i(p-p')u}\,, \tag{3.101}$$

where we let $u = x/\hbar$ in the last step. We claim that the last integral is precisely the integral representation of the delta function $\delta(p-p')$:

$$\frac{1}{2\pi}\int du\,e^{i(p-p')u} \;=\; \delta(p-p')\,. \tag{3.102}$$

This, then gives the correct value for the overlap $\langle p|p'\rangle$, as we claimed. The integral (3.102) can be justified using the fact that the functions

$$f_n(x) \equiv \frac{1}{\sqrt{L}}\exp\Big(\frac{2\pi i n x}{L}\Big)\,, \tag{3.103}$$

form a complete orthornormal set of functions over the interval $x\in[-L/2, L/2]$. Completeness then means that

$$\sum_{n\in\mathbb{Z}} f_n^*(x) f_n(x') \;=\; \delta(x-x')\,. \tag{3.104}$$

We thus have

$$\sum_{n\in\mathbb{Z}} \frac{1}{L}\exp\Big(2\pi i\,\frac{n}{L}(x-x')\Big) \;=\; \delta(x-x')\,. \tag{3.105}$$

In the limit as L goes to infinity the above sum can be written as an integral since the exponential is a very slowly varying function of $n \in \mathbb{Z}$. Since $\Delta n = 1$ with $u = 2\pi n/L$ we have $\Delta u = 2\pi/L \ll 1$ and then

$$\sum_{n\in\mathbb{Z}} \frac{1}{L} \exp\Big(2\pi i\, \frac{n}{L}(x-x')\Big) \;=\; \sum_u \frac{\Delta u}{2\pi} \exp\Big(i\, u(x-x')\Big) \;\to\; \frac{1}{2\pi}\int du e^{iu(x-x')}\,, \tag{3.106}$$

and back in (3.105) we have justified (3.102).

We can now ask: What is $\langle p|\psi\rangle$? We compute

$$\langle p|\psi\rangle \;=\; \int dx \langle p|x\rangle\langle x|\psi\rangle \;=\; \frac{1}{\sqrt{2\pi\hbar}}\int dx e^{-ipx/\hbar}\psi(x) \;=\; \tilde{\psi}(p)\,, \tag{3.107}$$

which is the Fourier transform of $\psi(x)$, as defined in (6.41) of Chapter 1. Thus the Fourier transform of $\psi(x)$ is the wavefunction in the momentum representation.

It is useful to know how to evaluate $\langle x|\hat{p}|\psi\rangle$. We do it by inserting a complete set of momentum states:

$$\langle x|\,\hat{p}\,|\psi\rangle \;=\; \int dp\, \langle x|p\rangle\langle p|\hat{p}\,|\psi\rangle \;=\; \int dp\, (p\langle x|p\rangle)\langle p|\psi\rangle \tag{3.108}$$

Now we notice that

$$p\langle x|p\rangle \;=\; \frac{\hbar}{i}\frac{d}{dx}\langle x|p\rangle \tag{3.109}$$

and thus

$$\langle x|\,\hat{p}\,|\psi\rangle \;=\; \int dp \Big(\frac{\hbar}{i}\frac{d}{dx}\langle x|p\rangle\Big)\langle p|\psi\rangle\,. \tag{3.110}$$

The derivative can be moved out of the integral, since no other part of the integrand depends on x:

$$\langle x|\,\hat{p}\,|\psi\rangle \;=\; \frac{\hbar}{i}\frac{d}{dx}\int dp\, \langle x|p\rangle\langle p|\psi\rangle \tag{3.111}$$

The completeness sum is now trivial and can be discarded to obtain

$$\boxed{\langle x|\,\hat{p}\,|\psi\rangle \;=\; \frac{\hbar}{i}\frac{d}{dx}\,\langle x|\psi\rangle \;=\; \frac{\hbar}{i}\frac{d}{dx}\,\psi(x)\,.} \tag{3.112}$$

Exercise. Show that

$$\langle p|\,\hat{x}\,|\psi\rangle \;=\; i\hbar\frac{d}{dp}\tilde{\psi}(p)\,. \tag{3.113}$$

UNCERTAINTY PRINCIPLE AND COMPATIBLE OBSERVABLES

B. Zwiebach

October 21, 2013

Contents

1 Uncertainty defined

As we know, observables are associated to Hermitian operators. Given one such operator A we can use it to measure some property of the physical system, as represented by a state Ψ. If the state is in an eigenstate of the operator A, we have no uncertainty in the value of the observable, which coincides with the eigenvalue corresponding to the eigenstate. We only have uncertainty in the value of the observable if the physical state is not an eigenstate of A, but rather a superposition of various eigenstates with different eigenvalues.

We want to define the **uncertainty** $\Delta A(\Psi)$ of the Hermitian operator A on the state Ψ. This uncertainty should vanish if and only if the state is an eigenstate of A. The uncertainty, moreover, should be a real number. In order to define such uncertainty we first recall that the expectation value of A on the state Ψ, assumed to be normalized, is given by

$$\langle A\rangle \;=\; \langle\Psi|A|\Psi\rangle \;=\; \langle\Psi, A\Psi\rangle\,. \tag{1.1}$$

The expectation $\langle A\rangle$ is guaranteed to be real since A is Hermitian. We then define the uncertainty as the norm of the vector obtained by acting with $(A-\langle A\rangle I)$ on the physical state (I is the identity operator):

$$\boxed{\Delta A(\Psi) \;\equiv\; \Big|\big(A-\langle A\rangle I\big)\Psi\Big|\,.} \tag{1.2}$$

The uncertainty, so defined is manifestly non-negative. If the uncertainty is zero, the vector inside the norm is zero and therefore:

$$\Delta A(\Psi) \;=\; 0 \quad \rightarrow \quad \big(A - \langle A\rangle I\big)\Psi \;=\; 0 \quad \rightarrow \quad A\,\Psi = \langle A\rangle \Psi\,, \tag{1.3}$$

and the last equation confirms that the state is indeed an eigenstate of A (note that $\langle A\rangle$ is a number). You should also note that $\langle A\rangle$ is indeed the eigenvalue, since taking the eigenvalue equation $A\Psi = \lambda\Psi$ and forming the inner product with another Ψ we get

$$\langle \Psi, A\Psi\rangle = \lambda\langle\Psi,\Psi\rangle \;=\; \lambda \quad \rightarrow \quad \lambda = \langle A\rangle. \tag{1.4}$$

Alternatively, if the state Ψ is an eigenstate, we now now that the eigenvalue if $\langle A\rangle$ and therefore the state $(A - \langle A\rangle I)\Psi$ vanishes and its norm is zero. We have therefore shown that

$$\boxed{\text{The uncertainty } \;\Delta A(\Psi) \text{ vanishes if and only if } \;\Psi \text{ is an eigenstate of } A\,.} \tag{1.5}$$

To compute the uncertainty one usually squares the expression in (1.2) so that

$$(\Delta A(\Psi))^2 \;=\; \big\langle\, \big(A - \langle A\rangle I\big)\Psi\,, \big(A - \langle A\rangle I\big)\Psi\,\big\rangle \tag{1.6}$$

Since the operator A is assumed to be Hermitian and consequently $\langle A\rangle$ is real, we have $(A - \langle A\rangle I)^\dagger = A - \langle A\rangle I$, and therefore we can move the operator on the first entry onto the second one to find

$$(\Delta A(\Psi))^2 \;=\; \big\langle\, \Psi\,, \big(A - \langle A\rangle I\big)^2\Psi\,\big\rangle\,. \tag{1.7}$$

While this is a reasonable form, we can simplify it further by expansion

$$(\Delta A(\Psi))^2 \;=\; \big\langle\, \Psi\,, \big(A^2 - 2\langle A\rangle A + \langle A\rangle^2 I\big)\Psi\,\big\rangle\,. \tag{1.8}$$

The last two term combine and we find

$$\boxed{(\Delta A(\Psi))^2 \;=\; \langle A^2\rangle - \langle A\rangle^2\,.} \tag{1.9}$$

Since the left-hand side is greater than or equal to zero, this incidentally shows that the expectation value of A^2 is larger than the expectation value of A, squared:

$$\boxed{\langle A^2\rangle \;\geq\; \langle A\rangle^2\,.} \tag{1.10}$$

An interesting geometrical interpretation of the uncertainty goes as follows. Consider the one-dimensional vector subspace U_Ψ generated by Ψ. Take the state $A\Psi$ and project it to the subspace U_Ψ. The projection, we claim is $\langle A\rangle\Psi$ and the part of $A\,\Psi$ in the orthogonal subspace $U_\Psi^\perp$ is a vector of norm equal to the uncertainty ΔA. Indeed the orthogonal projector P_{U_Ψ} is

$$P_{U_\Psi} \;=\; |\Psi\rangle\langle\Psi|\,, \tag{1.11}$$

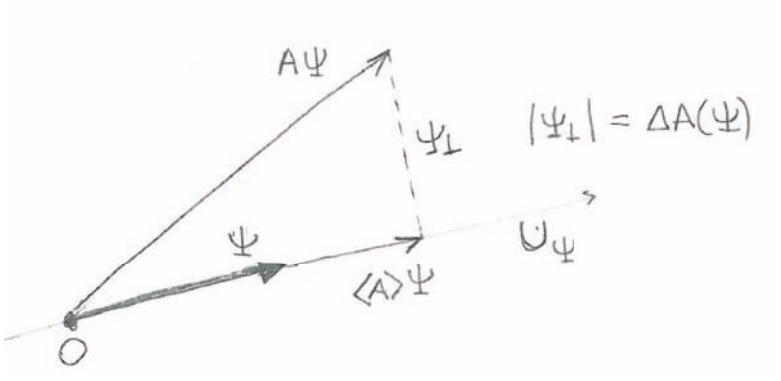


Figure 1: A state Ψ and the one-dimensional subspace U_Ψ generated by it. The projection of $A\Psi$ to U_Ψ is $\langle A\rangle\Psi$. The orthogonal complement $\Psi_\perp$ is a vector whose norm is the uncertainty $\Delta A(\Psi)$.

so that

$$P_{U_\Psi} A|\Psi\rangle \;=\; |\Psi\rangle\langle\Psi|A|\Psi\rangle \;=\; |\Psi\rangle\langle A\rangle\,. \tag{1.12}$$

Moreover, the vector $A|\Psi\rangle$ minus its projection must be a vector $|\Psi_\perp\rangle$ orthogonal to $|\Psi\rangle$

$$A|\Psi\rangle \;-\; \langle A\rangle|\Psi\rangle \;=\; |\Psi_\perp\rangle\,, \tag{1.13}$$

as is easily confirmed by taking the overlap with the bra Ψ. Since the norm of the above left-hand side is the uncertainty, we confirm that $\Delta A = |\Psi_\perp|$, as claimed. These results are illustrated in Figure 1.

2 The Uncertainty Principle

The uncertainty principle is an inequality that is satisfied by the product of the uncertainties of two Hermitian operators that fail to commute. Since the uncertainty of an operator on any given physical state is a number greater than or equal to zero, the product of uncertainties is also a real number greater than or equal to zero. The uncertainty inequality often gives us a lower bound for this product.

When the two operators in question commute, the uncertainty inequality gives no information.

Let us state the uncertainty inequality. Consider two Hermitian operators A and B and a physical state Ψ of the quantum system. Let ΔA and ΔB denote the uncertainties of A and B, respectively, in the state Ψ. Then we have

$$\boxed{(\Delta A)^2(\Delta B)^2 \;\geq\; \Big(\langle\Psi|\frac{1}{2i}[A,B]\big|\Psi\rangle\Big)^2.} \tag{2.14}$$

The left hand side is a real, non-negative number. For this to be consistent inequality, the right-hand side must also be a real number that is not negative. Since the right-hand side appears squared, the object inside the parenthesis must be real. This can only happen for all Ψ if the operator

$$\frac{1}{2i}[A,B] \tag{2.15}$$

is Hermitian. For this first note that the commutator of two Hermitian operators is *anti*-Hermitian:

$$[A,B]^\dagger = (AB)^\dagger - (BA)^\dagger = B^\dagger A^\dagger - A^\dagger B^\dagger - BA = -[A,B] \tag{2.16}$$

The presence of the i then makes the operator in (2.15) Hermitian. Note that the uncertainty inequality can also be written as

$$\boxed{\Delta A\,\Delta B \;\geq\; \left|\langle\Psi|\frac{1}{2i}[A,B]|\Psi\rangle\right|.} \tag{2.17}$$

where the bars on the right-hand side denote absolute value.

Before we prove the theorem, let's do the canonical example! Substuting $\hat{x}$ for A and $\hat{p}$ for B results in the position-momentum uncertainty relation you have certainly worked with:

$$(\Delta x)^2(\Delta p)^2 \geq \left(\langle\Psi|\frac{1}{2i}[\hat{x},\hat{p}]|\Psi\rangle\right)^2. \tag{2.18}$$

Since $[\hat{x},\hat{p}]/(2i) = \hbar/2$ we get

$$(\Delta x)^2(\Delta p)^2 \geq \frac{\hbar^2}{4} \quad\rightarrow\quad \Delta x\,\Delta p \;\geq\; \frac{\hbar}{2}. \tag{2.19}$$

We are interested in the proof of the uncertainty inequality for it gives the information that is needed to find the conditions that lead to saturation.

Proof. We define the following two states:

$$\begin{aligned} |f\rangle &\equiv (A - \langle A\rangle I)|\Psi\rangle \\ |g\rangle &\equiv (B - \langle B\rangle I)|\Psi\rangle\,. \end{aligned} \tag{2.20}$$

Note that by the definition (1.2) of uncertainty,

$$\begin{aligned} \langle f|f\rangle &= (\Delta A)^2\,, \\ \langle g|g\rangle &= (\Delta B)^2\,. \end{aligned} \tag{2.21}$$

The Schwarz inequality immediately furnishes us an inequality involving precisely the uncertainties

$$\langle f|f\rangle\langle g|g\rangle \geq |\langle f|g\rangle|^2\,, \tag{2.22}$$

and therefore we have

$$(\Delta A)^2(\Delta B)^2 \;\geq\; |\langle f|g\rangle|^2 \;=\; (\mathrm{Re}\langle f|g\rangle)^2 + (\mathrm{Im}\langle f|g\rangle)^2\,. \tag{2.23}$$

Writing $\check{A} = (A - \langle A\rangle I)$ and $\check{B} = (B - \langle B\rangle I)$, we now begin to compute the right-hand side:

$$\langle f|g\rangle = \langle\Psi|\check{A}\check{B}|\Psi\rangle \;=\; \langle\Psi|(A - \langle A\rangle I)(B - \langle B\rangle I)|\Psi\rangle \;=\; \langle\Psi|AB|\Psi\rangle - \langle A\rangle\langle B\rangle, \tag{2.24}$$

and since $|f\rangle$ and $|g\rangle$ go into each other as we exchange A and B,

$$\langle g|f\rangle = \langle\Psi|\check{A}\check{B}|\Psi\rangle = \langle\Psi|BA|\Psi\rangle - \langle B\rangle\langle A\rangle. \tag{2.25}$$

From the two equations above we find a nice expression for the imaginary part of $\langle f|g\rangle$:

$$\mathrm{Im}\langle f|g\rangle = \frac{1}{2i}(\langle f|g\rangle - \langle g|f\rangle) \;=\; \frac{1}{2i}\langle\Psi|[A,B]|\Psi\rangle\,. \tag{2.26}$$

For the real part the expression is not that simple, so it is best to leave it as the anticommutator of the checked operators:

$$\mathrm{Re}\langle f|g\rangle = \frac{1}{2}(\langle f|g\rangle + \langle g|f\rangle) \;=\; \frac{1}{2}\langle\Psi|\{\hat{A},\hat{B}\}|\Psi\rangle \tag{2.27}$$

Back in (2.23) we get

$$(\Delta A)^2(\Delta B)^2 \;\geq\; \Big(\langle\Psi|\frac{1}{2i}[A,B]|\Psi\rangle\Big)^2 + \Big(\langle\Psi|\frac{1}{2}\{\hat{A},\hat{B}\}|\Psi\rangle\Big)^2. \tag{2.28}$$

This can be viewed as the most complete form of the uncertainty inequality. It turns out, however, that the second term on the right hand side is seldom simple enough to be of use, and many times it can be made equal to zero for certain states. At any rate, the term is positive or zero so it can be dropped while preserving the inequality. This is often done, thus giving the celebrated form (2.14) that we have now established.

Now that we have a proven the uncertainty inequality, we can ask: What are the conditions for this inequality to be saturated? If the goal is to minimize uncertainties, under what conditions can we achieve the minimum possible product of uncertainties? As the proof shows, saturation is achieved under two conditions:

1. The Schwarz inequality is saturated. For this we need $|g\rangle = \beta|f\rangle$ where $\beta \in \mathbb{C}$.

2. $\mathrm{Re}(\langle f|g\rangle) = 0$, so that the last term in (2.28) vanishes. This means that $\langle f|g\rangle + \langle g|f\rangle = 0$.

Using $|g\rangle = \beta|f\rangle$ in Condition 2, we get

$$\langle f|g\rangle + \langle g|f\rangle \;=\; \beta\langle f|f\rangle + \beta^*\langle f|f\rangle \;=\; (\beta+\beta^*)\langle f|f\rangle = 0\,, \tag{2.29}$$

which requires $\beta+\beta^* = 0$ or that the real part of β vanish. It follows that β must be purely imaginary. So, $\beta = i\lambda$, with λ real, and therefore the uncertainty inequality will be saturated if and only if

$$|g\rangle \;=\; i\lambda|f\rangle, \quad \lambda\in\mathbb{R}\,. \tag{2.30}$$

More explicitly this requires

$$\boxed{\text{Saturation Condition:}\quad (B-\langle B\rangle\, I)|\Psi\rangle \;=\; i\lambda\,(A-\langle A\rangle I)|\Psi\rangle\,.} \tag{2.31}$$

This must be viewed as a condition for Ψ, given any two operators A and B. Moreover, note that $\langle A\rangle$ and $\langle B\rangle$ are Ψ dependent. What is λ, physically? Well, the norm of λ is actually fixed by the equation. Taking the norm of both sides we get

$$\Delta B = |\lambda|\Delta A \quad\to\quad |\lambda| \;=\; \frac{\Delta B}{\Delta A}. \tag{2.32}$$

The classic illustration of this saturation condition is worked out for the x,p uncertainty inequality $\Delta x\Delta p \geq \hbar/2$. You will find that gaussian wavefunctions satisfy the saturation condition.

3 The Energy-Time uncertainty

A more subtle form of the uncertainty relation deals with energy and time. The inequality is sometimes stated vaguely in the form $\Delta E \Delta t \gtrsim \hbar$. In here there is no problem in defining ΔE precisely, after all we have the Hamiltonian operator, and its uncertainty ΔH is a perfect candidate for the 'energy uncertainty'. The problem is time. Time is not an operator in quantum mechanics, it is a parameter, a real number used to describe the way systems change. Unless we define Δt in a precise way we cannot hope for a well-defined uncertainty relation.

We can try a rough, heuristic definition, in order to illustrate the spirit of the inequality. Consider a photon that is detected at some point in space, as a passing oscillatory wave of exact duration T. Even without quantum mechanical considerations we can ask the observer what was the angular frequency ω of the pulse. In order to answer our question the observer will attempt to count the number N of complete oscillations of the waveform that went through. Of course, this number N is given by T divided by the period $2\pi/\omega$ of the wave:

$$N \;=\; \frac{\omega\, T}{2\pi}\,. \tag{3.33}$$

The observer, however, will typically fail to count full waves, because as the pulse gets started from zero and later on dies off completely, the waveform will cease to follow the sinusoidal pattern. Thus we expect an uncertainty $\Delta N \gtrsim 1$. Given the above relation, this implies an uncertainty $\Delta\omega$ in the value of the angular frequency

$$\Delta\omega\, T \;\gtrsim\; 2\pi\,. \tag{3.34}$$

This is all still classical, the above identity is something electrical engineers are well aware of. It represents a limit on the ability to ascertain accurately the frequency of a wave that is observed for a limited amount of time. This becomes quantum mechanical if we speak of a single photon, whose energy is $E = \hbar\omega$. Then $\Delta E = \hbar\Delta\omega$, so that multiplying the above inequality by $\hbar$ we get

$$\Delta E\, T \;\gtrsim\; h\,. \tag{3.35}$$

In this uncertainty inequality T is the duration of the pulse. It is a reasonable relation but the presence of $\gtrsim$ betrays its lack of full precision.

We can find a precise energy/Q-ness uncertainty inequality by applying the general uncertainty inequality to the Hamiltonian H and another Hermitian operator Q, as did the distinguished Russian physicists L. Mandelstam and Tamm shortly after the formulation of the uncertainty principle. We would then have

$$\Delta H\, \Delta Q \;\geq\; \Big|\big\langle\Psi|\frac{1}{2i}[H,Q]\big|\Psi\big\rangle\Big|\,. \tag{3.36}$$

This starting point is interesting because the commutator $[H,Q]$ encodes something very physical about Q. Indeed, let us consider henceforth the case in which the operator Q has*no time dependence*. It could be, for example some function of $\hat{x}$ and $\hat{p}$, or for a spin-1/2 particle, the operator $|+\rangle\langle-|$. Such

operator Q can easily have time-dependent expectation values, but the time dependence originates from the time dependence of the states, not from the operator Q itself.

To explore the meaning of $[H, Q]$ we begin by computing the time-derivative of the expectation value of Q:

$$\frac{d}{dt}\langle Q\rangle \;=\; \frac{d}{dt}\big\langle \Psi\,, Q\Psi\big\rangle \;=\; \Big\langle \frac{\partial\Psi}{\partial t}\,, Q\Psi\Big\rangle + \Big\langle \Psi\,, Q\frac{\partial\Psi}{\partial t}\Big\rangle \tag{3.37}$$

where we did not have to differentiate Q as it is time-independent. At this point we can use the Schrödinger equation to find

$$\begin{aligned}\frac{d}{dt}\langle Q\rangle \;&=\; \Big\langle \frac{1}{i\hbar}\,H\Psi\,, Q\Psi\Big\rangle + \Big\langle \Psi\,, Q\frac{1}{i\hbar}\,H\Psi\Big\rangle \\ &=\; \frac{i}{\hbar}\Big(\big\langle\, H\Psi\,, Q\Psi\big\rangle - \big\langle \Psi\,, QH\Psi\big\rangle\Big) \\ &=\; \frac{i}{\hbar}\big\langle\, \Psi\,, (HQ-QH)\Psi\big\rangle \;=\; \frac{i}{\hbar}\big\langle\, \Psi\,, [H,Q]\Psi\big\rangle\end{aligned} \tag{3.38}$$

where we used the Hermiticity of the Hamiltonian. We have thus arrived at

$$\boxed{\frac{d}{dt}\langle Q\rangle \;=\; \frac{i}{\hbar}\big\langle\, [H,Q]\,\big\rangle \quad \text{for time-independent } Q\,.} \tag{3.39}$$

This is a very important result. Each time you see $[H, Q]$ you should think 'time derivative of $\langle Q\rangle$'. In classical mechanics one usually looks for conserved quantities, that is, functions of the dynamical variables that are time independent. In quantum mechanics a conserved operator is one whose expectation value is time independent. An operator Q is conserved if it commutes with the Hamiltonian!

With this result, the inequality (3.36) can be simplified. Indeed, using (3.39) we have

$$\Big|\Big\langle \frac{1}{2i}[H,Q]\,\Big\rangle\Big| \;=\; \Big|\; \frac{1}{2i}\,\frac{\hbar}{i}\,\frac{d\langle Q\rangle}{dt}\;\Big| \;=\; \frac{\hbar}{2}\,\Big|\frac{d\langle Q\rangle}{dt}\,\Big| \tag{3.40}$$

and therefore

$$\boxed{\Delta H\,\Delta Q \;\geq\; \frac{\hbar}{2}\,\Big|\frac{d\langle Q\rangle}{dt}\Big|, \quad \text{for time-independent } Q\,.} \tag{3.41}$$

This is a perfectly precise uncertainty inequality. The terms in it suggest a definition of a time Δt_Q

$$\Delta t_Q \;\equiv\; \frac{\Delta Q}{\Big|\frac{d\langle Q\rangle}{dt}\Big|}\,. \tag{3.42}$$

This quantity has units of time. It is the time it would take $\langle Q\rangle$ to change by ΔQ if both ΔQ and the velocity $\frac{d\langle Q\rangle}{dt}$ were time-independent. Since they are not necessarily so, we can view Δt_Q as the time for "appreciable" change in $\langle Q\rangle$. This is certainly so when $\langle Q\rangle$ and ΔQ are roughly of the same size. In terms of Δt_Q the uncertainty inequality reads

$$\Delta H \Delta t_Q \;\geq\; \frac{\hbar}{2}\,. \tag{3.43}$$

This is still a precise inequality, given that Δt_Q has a concrete definition in (3.42).

As you will consider in the homework, (3.41) can be used to derive an inequality for time $\Delta t_\perp$ that it takes for a system to become orthogonal to itself. If we call the initial state $\Psi(0)$, we call $\Delta t_\perp$ the smallest time for which $\langle \Psi(0), \Psi(\Delta t_\perp)\rangle = 0$. You will be able to show that

$$\Delta H\, \Delta t_\perp \;\geq\; \frac{h}{4}\,. \tag{3.44}$$

The speed in which a state can turn orthogonal depends on the energy uncertainty, and in quantum computation it plays a role in limiting the maximum possible speed of a computer for a fixed finite energy.

The uncertainty relation involves ΔH. It is natural to ask if this quantity is time dependent. As we show now, it is not, if the Hamiltonian is a time-independent operator. Indeed, if H is time independent, we can use H and H^2 for Q in (3.39) so that

$$\begin{aligned} \frac{d}{dt}\langle H\rangle \;&=\; \frac{i}{\hbar}\big\langle\, [H,H]\,\big\rangle \;=\; 0\,, \\ \frac{d}{dt}\langle H^2\rangle \;&=\; \frac{i}{\hbar}\big\langle\, [H,H^2]\,\big\rangle \;=\; 0\,. \end{aligned} \tag{3.45}$$

It then follows that

$$\frac{d}{dt}(\Delta H)^2 = \frac{d}{dt}\big(\,\langle H^2\rangle - \langle H\rangle^2\big) \;=\; 0\,. \tag{3.46}$$

showing that ΔH is a constant. So we have shown that

If H is time independent, the uncertainty ΔH is constant in time.	(3.47)

The concept of conservation of energy uncertainty can be used to understand some aspects of atomic decays. Consider, for illustration the hyperfine transition in the hydrogen atom. Due to the existence of proton spin and the electron spin, the ground state of hydrogen is fourfold degenerate, corresponding to the four possible combinations of spins (up-up, up-down, down-up, down-down). The magnetic interaction between the spins actually breaks this degeneracy and produces the so-called "hyperfine" splitting. This is a very tiny split: $\delta E = 5.88\times 10^{-6}$ev (compare with about 13.6 ev for the ground state energy). For a hyperfine atomic transition, the emitted photon carries the energy difference δE resulting in a wavelength of 21.1cm and a frequency $\nu = 1420.405751786(30)$MHz. The eleven significant digits of this frequency attest to the sharpness of the emission line. The issue of uncertainty arises because the excited state of the hyperfine splitting has a lifetime τ_H for decay to the ground state and emission of a photon. This lifetime is extremely long, in fact $\tau_H \sim 11$ million years ($= 3.4\times 10^{14}$ sec, recalling that a year is about $\pi\times 10^7$sec, accurate to better than 1%). This lifetime can be viewed as the time that takes some observable of the electron-proton system to change significantly (its total spin angular momentum, perhaps) so by the uncertainty principle it must be related to some energy uncertainty $\Delta E \sim \hbar/\tau_H \simeq 2\times 10^{-30}$ev. of the original excited state of the

hydrogen atom. Once the decay takes place the atom goes to the fully stable ground state, without any possible energy uncertainty. By the conservation of energy uncertainty, the photon must carry the uncertainty ΔE. But $\Delta E/\delta E \sim 3 \times 10^{-25}$, an absolutely infinitesimal effect on the photon. There is no broadening of the 21 cm line! That's one reason it is so useful in astronomy. For decays with much shorter lifetimes there can be an observable broadening of an emission line due to the energy-time uncertainty principle.

4 Lower bounds for ground state energies

You may recall that the variational principle could be used to find *upper* bounds on ground state energies. The uncertainty principle can be used to find *lower* bounds for the ground state energy of certain systems. use below the uncertainty principle in the form $\Delta x \Delta p \geq \hbar/2$ to find rigorous lower bounds for the ground state energy of one-dimensional Hamiltonians. This is best illustrated by an example.

Consider a particle in a one-dimensional quartic potential considered earlier

$$H \;=\; \frac{p^2}{2m} + \alpha\, x^4\,, \tag{4.48}$$

where $\alpha > 0$ is a constant with units of energy over length to the fourth power. Our goal is to find a *lower bound* for the ground state energy $\langle H \rangle_{gs}$. Taking the ground state expectation value of the Hamiltonian we have

$$\langle H \rangle_{gs} \;=\; \frac{\langle p^2 \rangle_{gs}}{2m} + \alpha\, \langle x^4 \rangle_{gs}\,, \tag{4.49}$$

Recalling that

$$(\Delta p)^2 = \langle p^2 \rangle - \langle p \rangle^2\,, \tag{4.50}$$

we see that

$$\langle p^2 \rangle \geq (\Delta p)^2\,, \tag{4.51}$$

for any state of the system. We should note however, that for the ground state (or any bound state) $\langle p \rangle = 0$ so that in fact

$$\langle p^2 \rangle_{gs} = (\Delta p)^2_{gs}\,, \tag{4.52}$$

From the inequality $\langle A^2 \rangle \geq \langle A \rangle^2$ we have

$$\langle x^4 \rangle \geq \langle x^2 \rangle^2\,. \tag{4.53}$$

Moreover, just like for momentum above, $(\Delta x)^2 = \langle x^2 \rangle - \langle x \rangle^2$ leads to

$$\langle x^2 \rangle \geq (\Delta x)^2\,, \tag{4.54}$$

so that

$$\langle x^4 \rangle \geq (\Delta x)^4\,, \tag{4.55}$$

for the expectation value on arbitrary states. Therefore

$$\langle H\rangle_{gs} \;=\; \frac{\langle p^2\rangle_{gs}}{2m} + \alpha\,\langle x^4\rangle_{gs} \geq \;\frac{(\Delta p_{gs})^2}{2m} + \alpha\,(\Delta x_{gs})^4 \tag{4.56}$$

From the uncertainty principle

$$\Delta x_{gs}\,\Delta p_{gs} \geq \frac{\hbar}{2} \quad\rightarrow\quad \Delta p_{gs} \geq \frac{\hbar}{2\Delta x_{gs}}\,. \tag{4.57}$$

Back to the value of $\langle H\rangle_{gs}$ we get

$$\langle H\rangle_{gs} \;\geq\; \frac{\hbar^2}{8m(\Delta x_{gs})^2} + \alpha\,(\Delta x_{gs})^4\,. \tag{4.58}$$

The quantity to the right of the inequality is a function of Δx_{gs}. This function has been plotted in Figure 2.

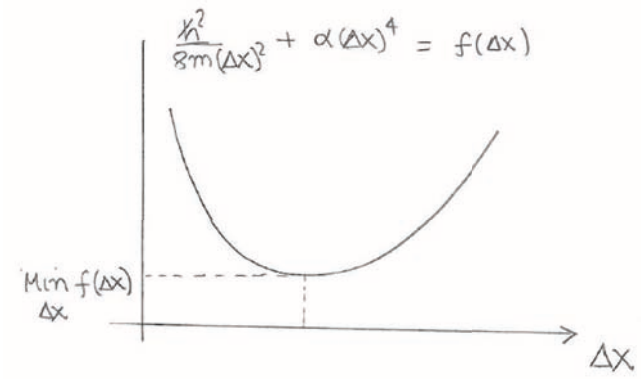


Figure 2: We have that $\langle H_{gs}\rangle \geq f(\Delta x_{gs})$ but we don't know the value of Δx_{gs}. As a result, we can only be certain that $\langle H_{gs}\rangle$ is greater than or equal to the *lowest* value the function $f(\Delta x_{gs})$ can take.

If we knew the value of Δx_{gs} we would immediately know that $\langle H\rangle_{gs}$ is bigger than the value taken by the right-hand side. This would be quite nice, since we want the highest possible lower bound. Since we don't know the value of Δx_{gs}, however, the only thing we can be sure of is that $\langle H\rangle_{gs}$ is bigger than the *lowest* value that can be taken by the expression to the right of the inequality as we vary Δx_{gs}:

$$\langle H\rangle_{gs} \;\geq\; \mathrm{Min}_{\Delta x}\Big(\,\frac{\hbar^2}{8m(\Delta x)^2} + \alpha\,(\Delta x)^4\Big)\,. \tag{4.59}$$

The minimization problem is straightforward. In fact

$$f(x) \;=\; \frac{A}{x^2} + Bx^4 \;\text{ is minimized for }\; x^2 = 2^{-1/3}\Big(\frac{A}{B}\Big)^{1/3} \;\text{ yielding }\; f \;=\; 2^{1/3}\frac{3}{2}\,(A^2B)^{1/3}\,. \tag{4.60}$$

Applied to (4.59) we obtaine

$$\langle H\rangle_{gs} \;\geq\; 2^{1/3}\,\frac{3}{8}\Big(\frac{\hbar^2\sqrt{\alpha}}{m}\Big)^{2/3} \;\simeq\; 0.4724\,\Big(\frac{\hbar^2\sqrt{\alpha}}{m}\Big)^{2/3}\,. \tag{4.61}$$

This is the final lower bound for the ground state energy. It is actually not too bad, for the ground state instead of the prefactor 0.4724, we have 0.668.

5 Diagonalization of Operators

When we have operators we wish to understand, it can be useful to find a basis on the vector space for which the operators are represented by matrices that take a simple form. Diagonal matrices are matrices where all non diagonal entries vanish. If we can find a set of basis vectors for which the matrix representing an operator is diagonal we say that the operator is **diagonalizable**.

If an operator T is diagonal in some basis $(u_1, \dots u_n)$ of the vector space V, its matrix takes the form $\text{diag}\,(\lambda_1, \dots \lambda_n)$, with constants λ_i, and we have

$$Tu_1 = \lambda_1 u_1\,, \quad \dots\,, \quad Tu_n = \lambda_n u_n\,. \tag{5.62}$$

The basis vectors are recognized as eigenvectors with eigenvalues given by the diagonal elements. It follows that *a matrix is diagonalizable if and only if it possesses a set of eigenvectors that span the vector space*. Recall that all operators T on complex vector spaces have at least one eigenvalue and thus at least a one eigenvector. But not even in complex vector spaces all operators have enough eigenvectors to span the space. Those operators cannot be diagonalized. The simplest example of such operator is provided by the two-by-two matrix

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\,. \tag{5.63}$$

The only eigenvalue of this matrix is $\lambda = 0$ and the associated eigenvector is $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Since a two-dimensional vector space cannot be spanned with one eigenvector, this matrix cannot be diagonalized. Having seen that the question of diagonalization of an operator is ultimately a question about its eigenvectors, we want to emphasize that the question can be formulated without referring to any basis. Bases, of course are useful, to express concretely

Suppose we have a vector space V and we have chosen a basis $(v_1, \dots, v_n)$ such that a linear operator has a matrix representation $T_{ij}(\{v\})$ that is not diagonal. As we learned before, if we change basis to a new one $(u_1, \dots, u_n)$ using a linear operator A such that

$$u_k \;=\; A\, v_k\,, \tag{5.64}$$

the matrix representation $T_{ij}(\{u\})$ of the operator in the new basis takes the form

$$T(\{u\}) = A^{-1} T(\{v\}) A \quad \text{or} \quad T_{ij}(\{u\}) \;=\; (A^{-1})_{ik} T_{kp}(\{v\})\, A_{pj}\,, \tag{5.65}$$

where the matrix A_{ij} is the representation of A in the original v-basis. The operator T is diagonalizable if there is an operator A such that $T_{ij}(\{u\})$ is diagonal.

There are two pictures of the diagonalization: One can consider the operator T and state that its matrix representation is diagonal when referred to the u basis obtained by acting with A on the original v basis. Alternatively, we can view the result as the existence of a related operator $A^{-1}TA$ that is diagonal in the *original* v basis. Indeed, $Tu_i = \lambda_i u_i$ (i not summed) implies that $TA\, v_i = \lambda_i A v_i$

and acting with A^{-1} that $(A^{-1}TA)\,v_i = \lambda_i v_i$, which confirms that $A^{-1}TA$ is represented by a diagonal matrix in the original v basis. Both viewpoints are valuable.

It is useful to note that the columns of the matrix A are in fact the eigenvectors of $T(\{v\})$. We see this as follows. Since the eigenvectors are the u_k we have

$$u_k = Av_k \quad \rightarrow \quad u_k = \sum_i A_{ik} v_i \,. \tag{5.66}$$

Using the original basis means v_i is represented by a column vector of zeroes with a single unit entry at the i-th position. We thus find

$$u_k = \begin{pmatrix} A_{1\,k} \\ \vdots \\ A_{nk} \end{pmatrix} \,. \tag{5.67}$$

confirming that the k-th column of A is the k-th eigenvector of T.

While not all operators on complex vector spaces can be diagonalized, the situation is much improved for Hermitian operators. Recall that T is Hermitian if $T = T^\dagger$. Hermitian operators can be diagonalized, and so can unitary operators. But even more is true: the operators take diagonal form in an orthonormal basis!

An operator M is said to be **unitarily** diagonalizable if there is an *orthonormal* basis in which its matrix representation is a diagonal matrix. That basis, therefore, is an orthonormal basis of eigenvectors. Starting with an arbitrary orthonormal basis $(e_1, \ldots, e_n)$ where the matrix representation of M is $M(\{e\})$, a unitary transformation of this basis produces the orthonormal basis in which the operator takes diagonal form. More explicitly, there is a unitary matrix U ($U^\dagger = U^{-1}$) and a diagonal matrix D_M such that

$$U^\dagger M(\{e\})\, U = D_M \,. \tag{5.68}$$

6 The Spectral Theorem

While we could prove, as most textbooks do, that Hermitian operators are unitarily diagonalizable, this result holds for a more general class of operators, called normal operators. The proof is not harder than the one for hermitian operators. An operator M is said to be **normal** if it commutes with its adjoint:

$$M \text{ is normal}: \quad [M^\dagger, M] = 0 \,. \tag{6.69}$$

Hermitian operators are clearly normal. So are anti-hermitian operators ($M^\dagger = -M$ is antihermitian). Unitary operators U are normal because both $U^\dagger U$ and $UU^\dagger$ are equal to the identity matrix and thus U and $U^\dagger$ commute.

Exercise. If an operator M is normal show that so is $V^\dagger M V$ where V is a unitary operator.

Lemma: Let w be an eigenvector of the normal operator M: $Mw = \lambda w$. Then w is also an eigenvector of $M^\dagger$ with complex conjugate eigenvalue:

$$M^\dagger w = \lambda^* w \,. \tag{6.70}$$

Proof: Define $u = (M^\dagger - \lambda^* I)w$. The result holds if u is the zero vector. To show this we compute the norm-squared of u:

$$|u|^2 \;=\; \langle u, u\rangle \;=\; \langle (M^\dagger - \lambda^* I)w \,, (M^\dagger - \lambda^* I)w\rangle \tag{6.71}$$

Using the adjoint property to move the operator in the first entry to the second entry:

$$|u|^2 \;=\; \langle w \,, (M - \lambda I)(M^\dagger - \lambda^* I)w\rangle \tag{6.72}$$

Since M and $M^\dagger$ commute, so do the two factors in parenthesis and therefore

$$|u|^2 \;=\; \langle w \,, (M^\dagger - \lambda^* I)(M - \lambda I)w\rangle \;=\; 0 \tag{6.73}$$

since $(M - \lambda I)$ kills w. It follows that $u = 0$ and therefore (6.70) holds. □

We can now state our main theorem, called the *spectral theorem*. It states that a matrix is unitarily diagonalizable if and only if it is normal. More to the point,

> **Spectral Theorem:** Let M be an operator in a complex vector space. The vector space has a orthonormal basis comprised of eigenvectors of M if and only if M is normal.

(6.74)

Proof. It is easy to show that unitarily diagonalizable implies normality. Indeed, from (5.68) and dropping the reference to the e-basis,

$$M = UD_M U^\dagger \;\text{ and therefore }\; M^\dagger = UD_M^\dagger U^\dagger \,.$$

We then get

$$M^\dagger M = UD_M^\dagger D_M U^\dagger \quad\text{and}\quad MM^\dagger = UD_M D_M^\dagger U^\dagger \,.$$

so that

$$[M^\dagger, M] = U(D_M^\dagger D_M - D_M D_M^\dagger)U^\dagger = 0 \,,$$

because any two diagonal matrices commute.

Now let us prove that M provides a basis of orthonormal eigenvectors. The proof is by induction. The result is clearly true for $\dim V = 1$. We assume that it holds for $(n-1)$-dimensional vector spaces and consider the case of n-dimensional V. Let M be an $n \times n$ matrix referred to the orthonormal basis $(|1\rangle, \ldots, |n\rangle)$ of V so that $M_{ij} = \langle i|M|j\rangle$. We know there is at least one eigenvalue λ_1 with a non-zero eigenvector $|x_1\rangle$ of unit norm:

$$M|x_1\rangle = \lambda_1|x_1\rangle \quad\text{and}\quad M^\dagger|x_1\rangle = \lambda_1^*|x_1\rangle \,, \tag{6.75}$$

in view of the Lemma. There is, we claim, a unitary matrix U_1 such that

$$|x_1\rangle = U_1|1\rangle \quad \rightarrow \quad U_1^\dagger|x_1\rangle = |1\rangle\,. \tag{6.76}$$

U_1 is not unique and can be constructed as follows: extend $|x_1\rangle$ to an orthonormal basis $|x_1\rangle, \ldots, |x_N\rangle$ using Gram-Schmidt. Then write $U_1 = \sum_i |x_i\rangle\langle i|$. Define now

$$M_1 \equiv U_1^\dagger M U_1\,. \tag{6.77}$$

M_1 is also normal and $M_1|1\rangle = U_1^\dagger M U_1|1\rangle = U_1^\dagger M|x_1\rangle = \lambda_1 U_1^\dagger|x_1\rangle = \lambda_1|1\rangle$, so that

$$M_1|1\rangle = \lambda_1|1\rangle\,. \tag{6.78}$$

Let us now examine the explicit form of the matrix M_1:

$$\langle j|M_1|1\rangle = \lambda_1\langle j|1\rangle = \lambda_1\delta_{i,j}\,, \tag{6.79}$$

which says that the first column of M_1 has zeroes in all entries except the first. Moreover

$$\langle 1|M_1|j\rangle = (\langle j|M_1^\dagger|1\rangle)^* = (\lambda_1^*\langle j|1\rangle)^* = \lambda_1\langle 1|j\rangle = \lambda_1\delta_{i,j}\,, \tag{6.80}$$

where we used $M_1^\dagger|1\rangle = \lambda_1^*|1\rangle$ which follows from the normality of M_1. It follows from the two last equations that M_1, in the original basis, takes the form

$$M_1 = \left(\begin{array}{c|ccc} \lambda_1 & 0 & \ldots & 0 \\ \hline 0 & & & \\ \vdots & & M' & \\ 0 & & & \end{array}\right)\,.$$

Since M_1 is normal, one can see that M' is a normal $(n-1)$-by-$(n-1)$ matrix. By the induction hypothesis M' can be unitarily diagonalized so that $U'^\dagger M' U'$ is diagonal for some $(n-1)$-by-$(n-1)$ unitary matrix U'. The matrix U' can be extended to an n-by-n unitary matrix $\hat{U}$ as follows

$$\hat{U} = \left(\begin{array}{c|ccc} 1 & 0 & \ldots & 0 \\ \hline 0 & & & \\ \vdots & & U' & \\ 0 & & & \end{array}\right)\,. \tag{6.81}$$

It follows that $\hat{U}^\dagger M_1 \hat{U} = \hat{U}^\dagger U_1^\dagger M\, U_1 \hat{U} = (U_1\hat{U})^\dagger M\,(U_1\hat{U})$ is diagonal, proving the desired result. $\square$.

Of course this theorem implies that Hermitian and unitary operators are unitarily diagonalizable. In other words the eigenvectors form an orthonormal basis. This is true whether or not there are degeneracies in the spectrum. The proof does not require discussion of this as a special case. If an eigenvalue of M is degenerate and appears k times, then there are k orthonormal eigenvectors associated with the corresponding k-dimensional M-invariant subspace of the vector space.

- - - - - - - - - - - - -

We conclude this section with a description of the general situation that we may encounter when diagonalizing a normal operator T. In general, we expect degeneracies in the eigenvalues so that each eigenvalue λ_k is repeated $d_k \geq 1$ times. An eigenvalue λ_k is degenerate if $d_k > 1$. It follows that V has T-invariant subspaces of different dimensionalities. Let U_k denote the T-invariant subspace of dimension $d_k \geq 1$ spanned by eigenvectors with eigenvalue λ_k:

$$U_k \;\equiv\; \{v \in V \,|\, T\,v = \lambda_k v\}\,, \quad \dim\, U_k = d_k\,. \tag{6.82}$$

By the spectral theorem U_k has a basis comprised by d_k orthonormal eigenvectors $(u_1^{(k)}\,,\;\ldots\;,\,u_{d_k}^{(k)})$. Note that while the addition of eigenvectors with different eigenvalues does not give eigenvectors, in the subspace U_k all vectors are eigenvectors with the same eigenvalue, and that's why addition makes sense, U_k as defined is a vector space, and adding eigenvectors in U_k gives eigenvectors. The full space V is decomposed as the direct sum of the invariant subspaces of T:

$$V = U_1 \oplus U_2 \oplus \ldots U_m\,, \quad \dim\, V = \sum_{i=1}^{m} d_i\,,\; m \geq 1\,. \tag{6.83}$$

All U_i subspaces are guaranteed to be orthogonal to each other. In fact the full list of eigenvectors is a list of orthonormal vectors that form a basis for V is conveniently ordered as follows:

$$(u_1^{(1)},\ldots\,,\,u_{d_1}^{(1)},\;\;\ldots\;\,,\;u_1^{(m)},\ldots,u_{d_m}^{(m)})\,. \tag{6.84}$$

The matrix T is manifestly diagonal in this basis because each vector above is an eigenvector of T and is orthogonal to all others. The matrix representation of T reads

$$T \;=\; \mathrm{diag}\,\big(\,\underbrace{\lambda_1,\;\ldots\,,\;\lambda_1}_{d_1\text{ times}}\,,\;\;\ldots\;\;,\underbrace{\lambda_m,\ldots\,,\,\lambda_m}_{d_m\text{ times}}\,\big) \tag{6.85}$$

This is is clear because the first d_1 vectors in the list are in U_1, the second d_2 vectors are in U_2, and so on and so forth until the last d_m vectors are in U_m.

If we had no degeneracies in the spectrum the basis (6.84) (with $d_i = 1$ for all i) would be rather unique if we require the matrix representation of T to be unchanged. Each vector could be multiplied by a phase. On the other hand, with degeneracies that the list (6.84) can be changed considerably without changing the matrix representation of T. Let V_k be a unitary operator on U_k, for each $k = 1,\ldots,m$. We claim that the following basis of eigenvectors leads to the same matrix T:

$$\big(\,V_1u_1^{(1)}\,,\ldots\,V_1u_{d_1}^{(1)},\;\;\ldots\;\,,\;V_mu_1^{(m)}\,,\;\ldots\;\,,V_mu_{d_m}^{(m)}\big)\,. \tag{6.86}$$

Indeed, this is still a collection of eigenvectors of T with each of them orthogonal to the rest. Moreover, the first d_1 vectors are in U_1, the second d_2 vectors are in U_2 and so on and so forth. More explicitly, for example, within U_k

$$\big\langle\; V_ku_i^{(k)}\,,\,T(V_ku_j^{(k)})\;\big\rangle \;=\; \lambda_k\,\langle\;V_ku_i^{(k)},\,V_ku_j^{(k)}\rangle \;=\; \lambda_k\langle\,u_i^{(k)},\,u_j^{(k)}\rangle \;=\; \lambda_k\delta_{ij} \tag{6.87}$$

showing that in the U_k subspace the matrix for T is still diagonal with al entries equal to λ_k.

7 Simultaneous Diagonalization of Hermitian Operators

We say that two operators S and T in a vector space V operators can be **simultaneously diagonalized** if there is some basis of V in which both the matrix representation of S and the matrix representation of T are diagonal. It then follows that each vector in this basis is an eigenvector of S *and* an eigenvector of T.

A necessary condition for simultaneous diagonalization is that the operators S and T commute. Indeed, if they can be simultaneously diagonalized there is a basis where both are diagonal and they manifestly commute. If the operators don't commute, this is a basis-independent statement and therefore a simultaneous diagonal presentation cannot exist. Since arbitrary linear operators S and T on a complex vector space cannot be diagonalized, the vanishing of $[S,T]$ does not guarantee simultaneous diagonalization. But if the operators are Hermitian it does, as we show now.

Theorem. If S and T are commuting Hermitian operators they can be simultaneously diagonalized.

Proof. The main complication is that degeneracies in the spectrum require an some discussion. Either both operators have degeneracies or one has no degeneracies. Without loss of generality we can assume that there are two cases to consider

(i) There is no degeneracy in the spectrum of T or,

(ii) Both T and S have degeneracies in their spectrum.

Consider case (i) first. Since T is non-degenerate there is a basis $(u_1, \dots u_n)$ of eigenvectors of T with different eigenvalues

$$Tu_i = \lambda_i u_i\,, \quad i \quad \text{not summed}\,, \quad \lambda_i \neq \lambda_j \;\text{ for }\; i \neq j\,. \tag{7.88}$$

We now want to understand what kind of vector is Su_i. For this we act with T on it

$$T(Su_i) = S(Tu_i) = S(\lambda_i u_i) = \lambda_i (S\,u_i)\,, \tag{7.89}$$

It follows that Su_i is also an eigenvector of T with eigenvalue λ_i, thus it must equal u_i, up to scale,

$$Su_i \;=\; \omega_i u_i\,, \tag{7.90}$$

showing that u_i is also an eigenvector of S, this time with eigenvalue ω_i. Thus any eigenvector of T is also an eigenvector of S, showing that these operators are simultaneously diagonalizable.

Now consider case (ii). Since T has degeneracies, as explained in the previous section, we have a decomposition of V in T-invariant subspaces U_k spanned by eigenvectors:

$$\begin{gathered} U_k \;\equiv\; \{u\,|Tu = \lambda_k u\}\,, \quad \dim U_k = d_k \qquad V \;=\; U_1 \oplus \dots U_m\,, \\ \text{orthonormal basis for } V: \qquad (u_1^{(1)}, \dots, u_{d_1}^{(1)}, \;\dots\;, u_1^{(m)}, \dots, u_{d_m}^{(m)})\,. \\ T \;=\; \text{diag}\,\big(\underbrace{\lambda_1, \dots, \lambda_1}_{d_1 \text{ times}}, \;\dots\; , \underbrace{\lambda_m, \dots, \lambda_m}_{d_m \text{ times}}\big) \quad \text{in this basis.} \end{gathered} \tag{7.91}$$

We also explained that the alternative orthonormal basis of V

$$\left(V_1 u_1^{(1)}, \dots V_1 u_{d_1}^{(1)}, \ \dots \ , \ V_m u_1^{(m)} \ , \ \dots \ , V_m u_{d_m}^{(m)} \right). \tag{7.92}$$

leads to the same matrix for T when each V_k is a unitary operator on U_k.

We now claim that the U_k are also S-invariant subspaces! To show this let $u \in U_k$ and examine the vector Su. We have

$$T(Su) = S(Tu) = \lambda_k Su \quad \rightarrow \quad Su \in U_k \,. \tag{7.93}$$

We use the subspaces U_k and the basis (7.91) to organize the matrix representation of S in blocks. It follows that this matrix must have *block-diagonal* form since each subspace is S-invariant and orthogonal to all other subspaces. We cannot guarantee, however, that S is diagonal within each square block because $Su_i^{(k)} \in U_k$ but we have no reason to believe that $Su_i^{(k)}$ points along $u_i^{(k)}$.

Since S restricted to each S-invariant subspace U_k is hermitian we can find an orthonormal basis of U_k in which the matrix S is diagonal. This new basis is unitarily related to the original basis $(u_1^{(k)}, \dots, u_{d_k}^{(k)})$ and thus takes the form $(V_k u_1^{(k)}, \dots, V_k u_{d_k}^{(k)})$ with V_k a unitary operator in U_k. Note that the eigenvalues of S in this block need not be degenerate. Doing this for each block, we find a basis of the form (7.92) in which S is diagonal. But T is still diagonal in this new basis, so both S and T have been simultaneously diagonalized. □

Remarks:

1. Note that the above proof gives an algorithmic way to produce the common list of eigenvectors. One diagonalizes one of the matrices and constructs the second matrix in the basis of eigenvectors of the first. These second matrix is block diagonal, where the blocks are organized by the degeneracies in the spectrum of the first matrix. One must then diagonalize within the blocks and is guaranteed that the new basis that works for the second matrix also works for the first.

2. If we had to simultaneously diagonalize three different commuting Hermitian operators S_1, S_2 and S_3, all of which have degenerate spectra, we would proceed as follows. We diagonalize S_1 and fix a basis in which S_1 is diagonal. In this basis we must find that S_2 and S_3 have exactly the same block structure. The corresponding block matrices are simply the matrix representations of S_2 and S_3 in each of the invariant spaces U_k appearing in the diagonalization of S_1. Since S_2 and S_3 commute, their restrictions to U_k commute. These restrictions can be diagonalized simultaneously, as guaranteed by our theorem which works for two matrices. The new basis in U_k that makes the restriction of S_2 and S_3 diagonal, will not disturb the diagonal form of S_1 in this block. This is repeated for each block, until we get a common basis of eigenvectors.

3. An inductive algorithm is clear. If we know how to simultaneously diagonalize n commuting Hermitian operators we can diagonalize $n+1$ of them, call them $S_1, \dots S_{n+1}$, as follows. We diagonalize S_1 and then consider the remaining n operators in the basis that makes S_1 diagonal.

We are guaranteed a common block structure for the n operators. The problem becomes one of simultaneous diagonalization of n commuting Hermitian block matrices, which is assumed known by the induction argument.

Corollary. If $\{S_1, \ldots, S_n\}$ is a set of mutually commuting Hermitian operators they can all be simultaneously diagonalized.

8 Complete Set of Commuting Observables

We have discussed the problem of finding eigenvectors and eigenvalues of a Hermitian operator S. This hermitian operator is thought as a quantum mechanical observable. The eigenvectors of S are physical states of the system in which the observable S can be measured without uncertainty. The result of the measurement is the eigenvalue associated with the eigenvector.

If the Hermitian operator S has a non-degenerate spectrum, all eigenvalues are different and we have a rather nice situation in which each eigenvector can be uniquely distinguished by labeling it with the corresponding eigenvalue of S. The physical quantity associated with the observable can be used to distinguish the various eigenstates. Moreover, these eigenstates provide an orthonormal basis for the full vector space. In this case the operator S provides a "complete set of commuting observables" or a CSCO, in short. The set here has just one observable, the operator S.

The situation is more nontrivial if the Hermitian operator S exhibits degeneracies in its spectrum. This means that V has an S-invariant subspace of dimension $d > 1$, spanned by orthonormal eigenvectors $(u_1, \ldots, u_d)$ all of which have S eigenvalue λ. This time, the eigenvalue of S does not allow us to distinguish or to label uniquely the basis eigenstates of the invariant subspace. Physically this is a deficient situation, as we have explicitly different states – the various u_i's – that we can't tell apart by the measurement of S alone. This time S does not provide a CSCO. Labeling eigenstates by the S eigenvalue does not suffice to distinguish them.

We are thus physically motivated to find another Hermitian operator T that is compatible with S. Two Hermitian operators are said to be **compatible observables** if they commute, since then we can find a basis of V comprised by simultaneous eigenvectors of the operators. These states can be labeled by two observables, namely, the two eigenvalues. If we are lucky, the basis eigenstates in each of the S-invariant subspaces of dimension higher than one can be organized into T eigenstates of different eigenvalues. In this case T breaks the spectral degeneracy of S and using T eigenvalues as well as S eigenvalues we can label uniquely a basis of orthonormal states of V. In this case we say that S and T form a CSCO.

We have now given enough motivation for a definition of a complete set of commuting observables. Consider a set of commuting observables, namely, a set $\{S_1, \ldots, S_k\}$ of Hermitian operators acting on a complex vector space V that represents the physical state-space of some quantum system. By the theorem in the previous section, we can find an orthonormal basis of vectors in V such that each vector

is an eigenstate of every operator in the set. Assume that each eigenstate in the basis is labeled by the eigenvalues of the S_i operators. The set $\{S_1, \ldots, S_k\}$ is said to be a **complete set of commuting observables** if no two states have the same labels.

It is a physically motivated assumption that for any physical quantum system there is a complete set of commuting observables, for otherwise there is no physical way to distinguish the various states that span the vector space. So in any physical problem we are urged to find such complete set, and we must include operators in such set until all degeneracies are broken. A CSCO need not be unique. Once we have a complete set of commuting observables, adding another observable causes no harm, although it is not necessary. Also, if (S_1, S_2) form a CSCO, so will $(S_1 + S_2, S_1 - S_2)$. Ideally, we want the smallest set of operators.

The first operator that is usually included in a CSCO is the Hamiltonian H. For bound state problems in one dimension, energy eigenstates are non-degenerate and thus the energy can be used to label uniquely the H-eigenstates. A simple example is the infinite square well. Another example is the one-dimensional harmonic oscillator. In such cases H forms the CSCO. If we have, however, a two-dimensional isotropic harmonic oscillator in the (x, y) plane, the Hamiltonian has degeneracies. At the first excited level we can have the first excited state of the x harmonic oscillator or, at the same energy, the first excited state of the y harmonic oscillator. We thus need another observable that can be used to distinguish these states. There are several options, as you will discuss in the homework.

QUANTUM DYNAMICS

B. Zwiebach
November 4, 2013

Contents

1 Harmonic oscillator

The harmonic oscillator is an ubiquitous and rich example of a quantum system. It is a solvable system and allows the exploration of quantum dynamics in detail as well as the study of quantum states with classical properties.

The harmonic oscillator is a system where the classical description suggests clearly the definition of the quantum system. Classically a harmonic oscillator is described by the position

$x(t)$ of a particle of mass m and its momentum $p(t)$. The energy E of a particle with position x and momentum p is given by

$$E \;=\; \frac{p^2}{2m} + \frac{1}{2} m\omega^2 \, x^2 \,. \tag{1.1}$$

Here the constant ω, with units of inverse time, is related to the period of oscillation T by $\omega = 2\pi/T$. In the simplest application, the classical harmonic oscillator arises when a mass m free to move along the x axis is attached to a spring with spring constant k. The restoring force $F = -kx$ acting on the mass then results in harmonic motion with angular frequency $\omega = \sqrt{k/m}$.

The quantum system is easily defined. Instead of position and momentum dynamical variables we have hermitian operators $\hat{x}$ and $\hat{p}$ with commutation relation

$$[\,\hat{x}\,,\hat{p}\,] \;=\; i\hbar\,\mathbf{1}\,. \tag{1.2}$$

To complete the definition of the system we need a Hamiltonian. Inspired by the classical energy function (1.1) above we *define*

$$\hat{H} \;\equiv\; \frac{\hat{p}^2}{2m} + \frac{1}{2} m\omega^2 \hat{x}^2 \,. \tag{1.3}$$

The state space $\mathcal{H}$ is the space of square-integrable complex valued functions of x. The system so defined is the *quantum* harmonic oscillator.

In order to solve the quantum system we attempt to 'factorize' the Hamiltonian. This means finding an operator V such that we can rewrite the Hamiltonian as $\hat{H} = V^\dagger V$. This is not exactly possible, but with a small modification it becomes possible. We can find a V for which

$$\boxed{\;\hat{H} \;=\; V^\dagger V \;+\; E_0\,\mathbf{1}\,,\;} \tag{1.4}$$

where E_0 is a constant with units of energy that multiplies the identity operator. This extra diagonal contribution does not complicate our task of finding the eigenstates of the Hamiltonian, nor their energies. This factorization allows us to show that any energy eigenstate must have energy greater than or equal to E_0. Indeed it follows from the above equation that

$$\langle\psi|\hat{H}|\psi\rangle \;=\; \langle\psi|V^\dagger V|\psi\rangle + E_0\langle\psi|\psi\rangle \;= \langle V\psi|V\psi\rangle + E_0\,, \tag{1.5}$$

Since any norm must be greater than or equal to zero, we have shown that

$$\boxed{\;\langle\psi|\hat{H}|\psi\rangle \;\geq\; E_0\,.\;} \tag{1.6}$$

For a normalized energy eigenstate $|E\rangle$ of energy E: $\hat{H}|E\rangle = E|E\rangle$, and the above inequality yields, as claimed

$$\langle E|\hat{H}|E\rangle = E \geq E_0\,. \tag{1.7}$$

To factorize the Hamiltonian we first rewrite it as

$$\hat{H} \;=\; \frac{1}{2}m\omega^2\Big(\hat{x}^2+\frac{\hat{p}^2}{m^2\omega^2}\Big)\,. \tag{1.8}$$

Motivated by the identity $a^2+b^2=(a-ib)(a+ib)$, holding for numbers a and b, we examine the product

$$\begin{aligned}\left(\hat{x}-\frac{i\hat{p}}{m\omega}\right)\left(\hat{x}+\frac{i\hat{p}}{m\omega}\right) \;&=\; \hat{x}^2+\frac{\hat{p}^2}{m^2\omega^2}+\frac{i}{m\omega}\big(\hat{x}\hat{p}-\hat{p}\hat{x}\big)\,,\\ &=\; \hat{x}^2+\frac{\hat{p}^2}{m^2\omega^2}-\frac{\hbar}{m\omega}\mathbf{1}\,,\end{aligned} \tag{1.9}$$

where the extra terms arise because $\hat{x}$ and $\hat{p}$, as opposed to numbers, do not commute. Letting

$$\begin{aligned}V \;&\equiv\; \hat{x}+\frac{i\hat{p}}{m\omega}\,,\\ V^\dagger \;&\equiv\; \hat{x}-\frac{i\hat{p}}{m\omega}\,,\end{aligned} \tag{1.10}$$

we rewrite (1.9) as

$$\hat{x}^2+\frac{\hat{p}^2}{m^2\omega^2} \;=\; V^\dagger V+\frac{\hbar}{m\omega}\mathbf{1}\,, \tag{1.11}$$

and therefore back in the Hamiltonian (1.8) we find,

$$\hat{H} \;=\; \frac{1}{2}m\omega^2\Big(V^\dagger V+\frac{\hbar}{m\omega}\mathbf{1}\Big) \;=\; \frac{1}{2}m\omega^2\,V^\dagger V \;+\; \frac{1}{2}\hbar\omega\mathbf{1}\,. \tag{1.12}$$

The constant E_0 defined in (1.4) is thus $\frac{1}{2}\hbar\omega$ and (1.6) implies that

$$\boxed{\langle\psi|\hat{H}|\psi\rangle \;\geq\; \frac{1}{2}\hbar\omega\,.} \tag{1.13}$$

This shows that $E\geq\frac{1}{2}\hbar\omega$ for any eigenstate of the oscillator.

It is convenient to scale the operators V and $V^\dagger$ so that they commute to give a simple, unit-free, constant. First we compute

$$\big[V\,,V^\dagger\big] \;=\; \Big[\hat{x}+\frac{i\hat{p}}{m\omega}\,,\hat{x}-\frac{i\hat{p}}{m\omega}\Big] \;=\; -\frac{i}{m\omega}[\hat{x}\,,\hat{p}]+\frac{i}{m\omega}[\hat{p},\hat{x}] \;=\; \frac{2\hbar}{m\omega}\mathbf{1}\,. \tag{1.14}$$

This suggests the definition of operators

$$\begin{aligned}\hat{a} \;&\equiv\; \sqrt{\frac{m\omega}{2\hbar}}\,V\,,\\ \hat{a}^\dagger \;&\equiv\; \sqrt{\frac{m\omega}{2\hbar}}\,V^\dagger\,.\end{aligned} \tag{1.15}$$

Due to the scaling we have

$$\boxed{\left[\,\hat{a}\,,\hat{a}^\dagger\,\right] = \mathbf{1}\,.} \tag{1.16}$$

From the above definitions we read the relations between $(\hat{a},\hat{a}^\dagger)$ and $(\hat{x}\,,\hat{p})$:

$$\boxed{\begin{aligned}\hat{a} &= \sqrt{\frac{m\omega}{2\hbar}}\left(\hat{x}+\frac{i\hat{p}}{m\omega}\right)\,,\\ \hat{a}^\dagger &= \sqrt{\frac{m\omega}{2\hbar}}\left(\hat{x}-\frac{i\hat{p}}{m\omega}\right)\,.\end{aligned}} \tag{1.17}$$

The inverse relations are many times useful as well,

$$\boxed{\begin{aligned}\hat{x} &= \sqrt{\frac{\hbar}{2m\omega}}(\hat{a}+\hat{a}^\dagger)\,,\\ \hat{p} &= i\sqrt{\frac{m\omega\hbar}{2}}(\hat{a}^\dagger-\hat{a})\,.\end{aligned}} \tag{1.18}$$

While neither $\hat{a}$ nor $\hat{a}^\dagger$ is hermitian (they are hermitian conjugates of each other), the above equations are consistent with the hermiticity of $\hat{x}$ and $\hat{p}$. We can now write the Hamiltonian in terms of the $\hat{a}$ and $\hat{a}^\dagger$ operators. Using (1.15) we have

$$V^\dagger V \;=\; \frac{2\hbar}{m\omega}\,\hat{a}^\dagger\hat{a}\,, \tag{1.19}$$

and therefore back in (1.12) we get

$$\boxed{\hat{H} \;=\; \hbar\omega\Big(\hat{a}^\dagger\hat{a}+\frac{1}{2}\Big) \;=\; \hbar\omega\Big(\hat{N}+\frac{1}{2}\Big)\,,\quad \hat{N} \;\equiv\; \hat{a}^\dagger\hat{a}\,.} \tag{1.20}$$

In here we have dropped the identity operator, which is usually understood. We have also introduced the *number* operator $\hat{N}$. This is, by construction, a hermitian operator and it is, up to a scale and an additive constant, equal to the Hamiltonian. An eigenstate of $\hat{H}$ is also an eigenstate of $\hat{N}$ and it follows from the above relation that the respective eigenvalues E and N are related by

$$E \;=\; \hbar\omega\left(N+\frac{1}{2}\right). \tag{1.21}$$

From the inequality (1.13) we have already shown that for any state

$$E \;\geq\; \frac{1}{2}\,\hbar\omega\,,\quad N\geq 0\,. \tag{1.22}$$

There cannot exist states with negative number. This can be confirmed directly. If $|\psi\rangle$ is a state of negative number we have

$$\hat{a}^\dagger\hat{a}|\psi\rangle = -\alpha^2|\psi\rangle\,,\quad \alpha>0\,. \tag{1.23}$$

Multiplying by the state bra $\langle\psi|$ and noticing that $\langle\psi|\hat{a}^\dagger\hat{a}|\psi\rangle = \langle\hat{a}\psi|\hat{a}\psi\rangle$ we get

$$\langle\hat{a}\psi|\hat{a}\psi\rangle = -\alpha^2\langle\psi|\psi\rangle\,. \tag{1.24}$$

This is a contradiction, for if $|\psi\rangle$ is not the zero vector, the right-hand side is negative, which cannot be since the left hand side is also a norm-squared and thus positive.

Exercise. Prove the following commutation relations

$$\begin{aligned} \big[\,\hat{H}\,,\,\hat{a}\,\big] &= -\hbar\omega\,\hat{a}\,, \\ \big[\,\hat{H}\,,\,\hat{a}^\dagger\,\big] &= +\hbar\omega\,\hat{a}^\dagger\,. \end{aligned} \tag{1.25}$$

To derive the spectrum of the oscillator we begin by assuming that one normalizable eigenstate $|E\rangle$ of energy E exists:

$$\hat{H}|E\rangle \;=\; E|E\rangle\,, \quad \langle E|E\rangle \;>\; 0\,. \tag{1.26}$$

Note that the state must have positive norm-squared, as indicated above. The state $|E\rangle$ also an eigenstate of the number operator, with eigenvalue N_E given by

$$\hat{N}|E\rangle \;=\; N_E|E\rangle\,, \quad \text{with} \;\; N_E \;=\; \frac{E}{\hbar\omega} - \frac{1}{2}\,. \tag{1.27}$$

We will now define two states

$$\begin{aligned} |E_+\rangle &= \hat{a}^\dagger|E\rangle\,, \\ |E_-\rangle &= \hat{a}\,|E\rangle\,. \end{aligned} \tag{1.28}$$

Let us assume, for the time being that both of these states exist – that is, they are not zero nor they are inconsistent by having negative norm-squared. We can then verify they are energy eigenstates

$$\begin{aligned} \hat{H}|E_+\rangle &= \hat{H}\hat{a}^\dagger|E\rangle = \big([\hat{H},\hat{a}^\dagger] + \hat{a}^\dagger\hat{H}\big)|E\rangle = (\hbar\omega + E)\,\hat{a}^\dagger|E\rangle = (E+\hbar\omega)|E_+\rangle\,, \\ \hat{H}|E_-\rangle &= \hat{H}\hat{a}\,|E\rangle = \big([\hat{H},\hat{a}\,] + \hat{a}\,\hat{H}\big)|E\rangle = (-\hbar\omega + E)\,\hat{a}\,|E\rangle = (E-\hbar\omega)|E_-\rangle\,, \end{aligned} \tag{1.29}$$

As we label the states with their energies, this shows that

$$\begin{aligned} E_+ &= E + \hbar\omega\,, \quad N_{E_+} = N_E + 1\,, \\ E_- &= E - \hbar\omega\,, \quad N_{E_-} = N_E - 1\,. \end{aligned} \tag{1.30}$$

We call $\hat{a}^\dagger$ the *creation* or *raising* operator because it adds energy $\hbar\omega$ to the eigenstate it acts on, or raises the number operator by one unit. We call $\hat{a}$ the *annihilation* or *lowering* operator because it subtracts energy $\hbar\omega$ to the eigenstate it acts on, or lowers the number operator by one unit. One more computation is needed: we must find the norm-squared of the $|E_\pm\rangle$ states:

$$\begin{aligned} \langle E_+|E_+\rangle &= \langle E|\hat{a}\hat{a}^\dagger|E\rangle = \langle E|(\hat{N}+1)|E\rangle = (N_E+1)\langle E|E\rangle\,, \\ \langle E_-|E_-\rangle &= \langle E|\hat{a}^\dagger\hat{a}|E\rangle = \langle E|\hat{N}|E\rangle = N_E\,\langle E|E\rangle\,. \end{aligned} \tag{1.31}$$

We can summarize this as

$$\begin{aligned} \langle \hat{a}^\dagger E | \hat{a}^\dagger E \rangle &= (N_E + 1) \langle E | E \rangle \,, \\ \langle \hat{a} E \,|\, \hat{a} E \,\rangle &= N_E \, \langle E | E \rangle \,. \end{aligned} \tag{1.32}$$

These equations tell us an interesting story. Since the state $|E\rangle$ is assumed to exist we must have $N_E \geq 0$ (see (1.22)) . We claim that as long as we act with $\hat{a}^\dagger$ on this state we do not obtain inconsistent states. Indeed the first equation above shows that norm-squared of $|\hat{a}^\dagger E\rangle$ is positive, as it should be. If we act again with $\hat{a}^\dagger$, since the number of $|\hat{a}^\dagger E\rangle$ is $N_E + 1$ we find

$$\langle \hat{a}^\dagger \hat{a}^\dagger E | \hat{a}^\dagger \hat{a}^\dagger E \rangle \;=\; (N_E + 2) \langle \hat{a}^\dagger E | \hat{a}^\dagger E \rangle \;=\; (N_E + 2)(N_E + 1) \langle E | E \rangle \,, \tag{1.33}$$

which is also positive. We *cannot* find an inconsistent negative norm-squared however many times we act with the raising operator.

The lowering operator, however, requires more care. Assume we have a state $|E\rangle$ with *integer* positive number N_E. The number eigenvalue goes down in steps of one unit each time we apply an $\hat{a}$ operator to the state. As long as the number of a state is positive, the *next* state having an extra $\hat{a}$ has positive norm-squared because of the relation $\langle \hat{a} E \,|\, \hat{a} E \,\rangle = N_E \, \langle E | E \rangle$. So no complication arises until we hit a state $|E'\rangle$ with number $N_{E'} = 0$, in which case it follows that

$$\langle \hat{a} E' \,|\, \hat{a} E' \,\rangle \;=\; N'_E \, \langle E' | E' \rangle = 0 \,. \tag{1.34}$$

Having zero norm, the state $|\hat{a} E'\rangle$ must be the zero vector and we cannot continue to apply lowering operators. We thus avoid inconsistency.

If the original $|E\rangle$ state has a positive *non-integer* number N_E we can lower the number by acting with $\hat{a}$'s until we get a state $|E'\rangle$ with number between zero and one. The next state $|\hat{a} E'\rangle$ has negative number and this is an inconsistency – as we showed before these cannot exist. This contradiction can only mean that the original assumptions cannot be true. So one of the following must be true

1. There is no state with non-integer positive number.

2. There is a state with non-integer positive number but the repeated application of $\hat{a}$ gives a vanishing state before we encounter states with negative number.

Option 2 actually cannot happen. For a state $|\psi\rangle$ of non-zero number $\hat{a}^\dagger \hat{a} |\psi\rangle \sim |\psi\rangle$ and therefore $\hat{a}$ cannot kill the state. We conclude that there are no states in the spectrum with non-integer number.

What are the energy eigenstates annihilated by $\hat{a}$? Assume there is such state $|E\rangle$:

$$\hat{a} \, |E\rangle \;=\; 0 \,. \tag{1.35}$$

Acting with $\hat{a}^\dagger$ we find $\hat{a}^\dagger a|E\rangle = \hat{N}|E\rangle = 0$, so such state must have zero number and thus lowest energy:

$$N_E = 0\,, \quad E = \frac{1}{2}\,\hbar\omega\,. \tag{1.36}$$

To show that the state annihilated by $\hat{a}$ exists and is unique we solve the differential equation implicit in (1.35). We act with a position bra to find

$$\langle x|\hat{a}\,|E\rangle \;=\; 0 \quad \to \quad \sqrt{\frac{m\omega}{2\hbar}}\,\langle x|\left(\hat{x} + \frac{i\hat{p}}{m\omega}\right)|E\rangle \;=\; 0\,. \tag{1.37}$$

The prefactor is irrelevant and we have, with $\psi_E(x) \equiv \langle x|E\rangle$,

$$\Big(x + \frac{\hbar}{m\omega}\frac{d}{dx}\Big)\psi_E(x) \;=\; 0 \quad \to \quad \frac{d\psi_E}{dx} \;=\; -\frac{m\omega}{\hbar}\,x\psi_E\,. \tag{1.38}$$

The solution of the first-order differential equation is unique (up to normalization)

$$\psi_E(x) \;=\; N_0\exp\Big(-\frac{m\omega}{2\hbar}\,x^2\Big)\,, \quad N_0 \;=\; \Big(\frac{m\omega}{\pi\hbar}\Big)^{1/4}\,. \tag{1.39}$$

We have found a single state annihilated by $\hat{a}$ and it has number zero. The $\psi_E(x)$ above is the normalized wavefunction for the ground state of the simple harmonic oscillator.

In the following we denote states as $|n\rangle$ where n is the eigenvalue of the number operator $\hat{N}$:

$$\hat{N}|n\rangle \;=\; n|n\rangle\,. \tag{1.40}$$

In this language the ground state is the non-degenerate state $|0\rangle$ (do not confuse this with the zero vector or a state of zero energy!). It is annihilated by $\hat{a}$:

$$\text{SHO ground state } |0\rangle: \quad \hat{a}\,|0\rangle \;=\; 0\,, \qquad \hat{N}|0\rangle \;=\; 0\,, \quad \hat{H}|0\rangle \;=\; \frac{1}{2}\hbar\omega|0\rangle\,. \tag{1.41}$$

The ground state wavefunction was determined above

$$\psi_0(x) = \langle x|0\rangle = \Big(\frac{m\omega}{\pi\hbar}\Big)^{1/4}\exp\Big(-\frac{m\omega}{2\hbar}x^2\Big)\;. \tag{1.42}$$

Excited states are obtained by the successive action of $\hat{a}^\dagger$ on the ground state. The first excited state is

$$|1\rangle \;\equiv\; \hat{a}^\dagger|0\rangle \tag{1.43}$$

This state has number equal to one. Indeed, since $\hat{N}$ kills the ground state,

$$\hat{N}\hat{a}^\dagger|0\rangle \;=\; [\hat{N},\hat{a}^\dagger]|0\rangle \;=\; \hat{a}^\dagger|0\rangle\,. \tag{1.44}$$

Moreover the state is properly normalized

$$\langle 1|1\rangle \;=\; \langle 0|\hat{a}\hat{a}^\dagger|0\rangle \;=\; \langle 0|[\hat{a}\,,\hat{a}^\dagger]|0\rangle \;=\; \langle 0|0\rangle = 1\,. \tag{1.45}$$

The next excited state is

$$|2\rangle \;=\; \frac{1}{\sqrt{2}}\,\hat{a}^\dagger\hat{a}^\dagger|0\rangle\,. \tag{1.46}$$

This state has number equal to two, as desired. The normalization is checked as follows:

$$\langle 2|2\rangle \;=\; \frac{1}{2}\langle 0|\,\hat{a}\hat{a}\,\hat{a}^\dagger\hat{a}^\dagger|0\rangle \;=\; \frac{1}{2}\langle 0|\,\hat{a}\,[\hat{a}\,,\hat{a}^\dagger\hat{a}^\dagger]\,|0\rangle \;=\; \frac{1}{2}\langle 0|\,\hat{a}\,(2\hat{a}^\dagger)|0\rangle \;=\; \langle 0|\,\hat{a}\,\hat{a}^\dagger|0\rangle \;=\; 1\,. \tag{1.47}$$

In order to get the general state it is useful to consider (1.32) in the new notation

$$\begin{aligned}\langle \hat{a}^\dagger n|\hat{a}^\dagger n\rangle \;&=\; (n+1)\langle n|n\rangle \;=\; n+1\,,\\ \langle \hat{a}n\,|\,\hat{a}n\,\rangle \;&=\; n\,\langle n|n\rangle \;=\; n\,.\end{aligned} \tag{1.48}$$

The first means that $\hat{a}^\dagger|n\rangle$ is a state of norm-squared $n+1$ and $\hat{a}|n\rangle$ is a state of norm-squared n. Since we know that $\hat{a}^\dagger|n\rangle \sim |n+1\rangle$ and $\hat{a}|n\rangle \sim |n-1\rangle$ we conclude that

$$\boxed{\begin{aligned}\hat{a}^\dagger|n\rangle \;&=\; \sqrt{n+1}\,|n+1\rangle\;,\\ \hat{a}\,|n\rangle \;&=\; \sqrt{n}\;\;|n-1\rangle\;.\end{aligned}} \tag{1.49}$$

The signs chosen for the square roots are consistent as you can check by using the two equations above to verify that $\hat{a}^\dagger\hat{a}\,|n\rangle = n|n\rangle$. From the top equation we have

$$|n\rangle \;=\; \frac{1}{\sqrt{n}}\,\hat{a}^\dagger|n-1\rangle\,. \tag{1.50}$$

Using that equation again for the rightmost ket, and then repeatedly, we find

$$\begin{aligned}|n\rangle \;&=\; \frac{1}{\sqrt{n}}\,\hat{a}^\dagger\frac{1}{\sqrt{n-1}}\hat{a}^\dagger|n-2\rangle \;=\; \frac{1}{\sqrt{n(n-1)}}\;(\hat{a}^\dagger)^2|n-2\rangle\\ &=\; \frac{1}{\sqrt{n(n-1)(n-2)}}\;(\hat{a}^\dagger)^3|n-3\rangle \;=\; \ldots\\ &=\; \frac{1}{\sqrt{n!}}\,(\hat{a}^\dagger)^n|0\rangle\,.\end{aligned} \tag{1.51}$$

It is a good exercise to verify explicitly that $\langle n|n\rangle = 1$. In summary, the energy eigenstates are an orthonormal basis

$$\boxed{|n\rangle \;=\; \frac{1}{\sqrt{n!}}(a^\dagger)^n|0\rangle\,,\qquad \langle m|n\rangle = \delta_{mn}\,.} \tag{1.52}$$

You can verify by explicit computation that $\langle m|n\rangle = 0$ for $m \neq n$, but you can be sure this is true because these are eigenstates of the hermitian operator $\hat{N}$ with different eigenvalues (recall that theorem?).

Their energies are given by

$$\boxed{H|n\rangle = E_n|n\rangle = \hbar\omega\Big(n+\frac{1}{2}\Big)|n\rangle\,,\quad \hat{N}|n\rangle = n|n\rangle\,.} \tag{1.53}$$

One can prove that there are no additional excited states. If there were, they would have to have integer number and thus be degenerate with some of the above states. It can be shown (homework) that any such degeneracy would imply a degeneracy of the ground state, something we have ruled out explicitly. Therefore we have shown that the state space has the direct sum decomposition into one-dimensional $\hat{N}$-invariant subspaces U_n:

$$\mathcal{H} = U_0 \oplus U_1 \oplus U_2 \oplus \cdots\,, \qquad U_n \equiv \{\alpha|n\rangle,\ \alpha\in\mathbb{C},\ \hat{N}|n\rangle = n|n\rangle\}\,. \tag{1.54}$$

The algebra of $\hat{a}$ and $\hat{a}^\dagger$ operators allows simple computation of expectation values. For example,

$$\begin{aligned}\langle n|\hat{x}|n\rangle &= \sqrt{\frac{\hbar}{2m\omega}}\langle n|(\hat{a}+\hat{a}^\dagger)|n\rangle = 0\,,\\ \langle n|\hat{p}|n\rangle &= i\sqrt{\frac{m\omega\hbar}{2}}\langle n|(\hat{a}^\dagger-\hat{a})|n\rangle = 0\,.\end{aligned} \tag{1.55}$$

In here we used that $\langle n|\hat{a}|n\rangle \sim \langle n|n-1\rangle = 0$ and $\langle n|\hat{a}^\dagger|n\rangle \sim \langle n|n+1\rangle = 0$. For the quadratic operators, both $\hat{a}\hat{a}$ and $\hat{a}^\dagger\hat{a}^\dagger$ have zero diagonal matrix elements and therefore

$$\begin{aligned}\langle n|\hat{x}^2|n\rangle &= \frac{\hbar}{2m\omega}\,\langle n|(\hat{a}+\hat{a}^\dagger)^2|n\rangle = \frac{\hbar}{2m\omega}\langle n|(\hat{a}\hat{a}^\dagger+\hat{a}^\dagger\hat{a})|n\rangle\,,\\ \langle n|\hat{p}^2|n\rangle &= -\frac{m\omega\hbar}{2}\langle n|(\hat{a}^\dagger-\hat{a})^2|n\rangle = \frac{m\omega\hbar}{2}\langle n|(\hat{a}^\dagger\hat{a}+\hat{a}\hat{a}^\dagger)|n\rangle\,.\end{aligned} \tag{1.56}$$

But $\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a} = 1 + \hat{N} + \hat{N} = 1 + 2\hat{N}$ so therefore

$$\begin{aligned}\langle n|\hat{x}^2|n\rangle &= \frac{\hbar}{2m\omega}\,\big(1+2n\,\big) = \frac{\hbar}{m\omega}\,\Big(n+\frac{1}{2}\Big)\,,\\ \langle n|\hat{p}^2|n\rangle &= \frac{m\omega\hbar}{2}\,\big(1+2n\,\big) = m\hbar\omega\Big(n+\frac{1}{2}\Big)\,.\end{aligned} \tag{1.57}$$

It follows that in the state $|n\rangle$ we have the uncertainties

$$\begin{aligned}(\Delta x)^2 &= \frac{\hbar}{m\omega}\,\Big(n+\frac{1}{2}\Big)\\ (\Delta p)^2 &= m\hbar\omega\Big(n+\frac{1}{2}\Big)\,.\end{aligned} \tag{1.58}$$

As a result

$$\text{On the state } |n\rangle: \quad \Delta x\,\Delta p = \hbar\Big(n+\frac{1}{2}\Big)\,. \tag{1.59}$$

Only for the ground state $n=0$ product of uncertainties saturates the lower bound given by the Heisenberg uncertainty principle.

2 Schrödinger dynamics

The state space of quantum mechanics –the Hilbert space $\mathcal{H}$ of states – is best thought as a space with time-independent basis vectors. There is no role for time in the definition of the state space $\mathcal{H}$. In the Schrödinger "picture" of the dynamics, the state that represents a quantum system depends on time. Time is viewed as a parameter: at different times the state of the system is represented by different states in the Hilbert space. We write the state vector as

$$|\Psi, t\rangle\,, \tag{2.1}$$

and it is a vector whose components along the basis vectors of $\mathcal{H}$ are time dependent. If we call those basis vectors $|u_i\rangle$, we have

$$|\Psi, t\rangle \;=\; \sum_i |u_i\rangle c_i(t)\,, \tag{2.2}$$

where the $c_i(t)$ are some functions of time. Since a state must be normalized, we can imagine $|\Psi, t\rangle$ as a unit vector whose tip, as a function of time, sweeps a trajectory in $\mathcal{H}$. We will first discuss the postulate of unitary time evolution and then show that the Schrödinger equation follows from it.

2.1 Unitary time evolution

We declare that for any quantum system there is a *unitary* operator $\mathcal{U}(t, t_0)$ such that for *any* state $|\Psi, t_0\rangle$ of the system at time t_0 the state at time t is obtained as

$$\boxed{|\Psi, t\rangle \;=\; \mathcal{U}(t, t_0)|\Psi, t_0\rangle\,, \qquad \forall\; t, t_0\,.} \tag{2.3}$$

It must be emphasized that the operator $\mathcal{U}$ generates time evolution for *any* possible state at time t_0 –it does *not* depend on the chosen state at time t_0. A physical system has a single operator $\mathcal{U}$ that generates the time evolution of all possible states. The above equation is valid for all times t, so t can be greater than, equal to, or less than t_0. As defined, the operator $\mathcal{U}$ is unique: if there is another operator $\mathcal{U}'$ that generates exactly the same evolution then $(\mathcal{U} - \mathcal{U}')|\Psi, t_0\rangle = 0$ and since the state $|\Psi, t_0\rangle$ is arbitrary we must have that the operator $\mathcal{U} - \mathcal{U}'$ vanishes, showing that $\mathcal{U} = \mathcal{U}'$.

The unitary property of $\mathcal{U}$ means that

$$(\mathcal{U}(t, t_0))^\dagger \mathcal{U}(t, t_0) \;=\; \mathbf{1}\,. \tag{2.4}$$

In order to avoid extra parenthesis, we will write

$$\mathcal{U}^\dagger(t, t_0) \;\equiv\; (\mathcal{U}(t, t_0))^\dagger\,, \tag{2.5}$$

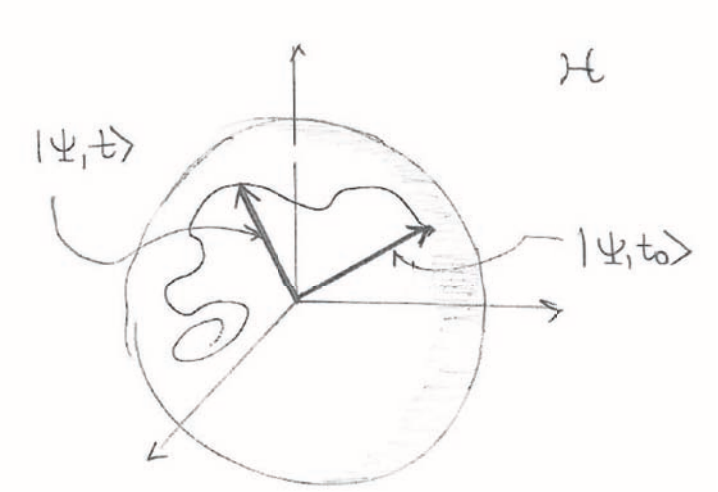


Figure 1: The initial state $|\Psi, t_0\rangle$ can be viewed as a vector in the complex vector space $\mathcal{H}$. As time goes by the vector moves, evolving by unitary transformations, so that its norm is preserved.

so that the unitarity property reads

$$\mathcal{U}^\dagger(t, t_0)\,\mathcal{U}(t, t_0) \;=\; \mathbf{1}\,. \tag{2.6}$$

Unitarity implies that the norm of the state is conserved[1]

$$\langle \Psi, t\,|\Psi, t\rangle \;=\; \langle \Psi, t_0\,|\mathcal{U}^\dagger(t, t_0)\mathcal{U}(t, t_0)|\Psi, t_0\rangle \;=\; \langle \Psi, t_0|\,\Psi, t_0\rangle\,. \tag{2.7}$$

This is illustrated in Figure 1.

We now make a series of comments on this postulate.

1. For time $t = t_0$, equation (2.3) gives no time evolution

$$|\Psi, t_0\rangle \;=\; \mathcal{U}(t_0, t_0)|\Psi, t_0\rangle\,. \tag{2.8}$$

Since this equality holds for *any* possible state at $t = t_0$ the unitary evolution operator must be the unit operator

$$\mathcal{U}(t_0, t_0) \;=\; \mathbf{1}\,, \quad \forall t_0\,. \tag{2.9}$$

2. Composition. Consider the evolution from t_0 to t_2 as a two-step procedure, from t_0 to t_1 and from t_1 to t_2:

$$|\Psi, t_2\rangle \;=\; \mathcal{U}(t_2, t_1)|\Psi, t_1\rangle \;=\; \mathcal{U}(t_2, t_1)\mathcal{U}(t_1, t_0)|\Psi, t_0\rangle\,. \tag{2.10}$$

It follows from this equation and $|\Psi, t_2\rangle \;=\; \mathcal{U}(t_2, t_0)|\Psi, t_0\rangle$ that

$$\mathcal{U}(t_2, t_0) \;=\; \mathcal{U}(t_2, t_1)\mathcal{U}(t_1, t_0)\,. \tag{2.11}$$

[1]We also recall that any operator that preserves the norm of arbitrary states is unitary.

3. Inverses. Consider (2.11) and set $t_2 = t_0$ and $t_1 = t$. Then using (2.9) we get

$$\mathbf{1} \;=\; \mathcal{U}(t_0,t)\,\mathcal{U}(t,t_0)\,. \tag{2.12}$$

Thus we have

$$\mathcal{U}(t_0,t) \;=\; (\mathcal{U}(t,t_0))^{-1} \;=\; (\mathcal{U}(t,t_0))^{\dagger}\,, \tag{2.13}$$

where the first relation follows from (2.12) and the second by unitarity. Again, declining to use parenthesis that are not really needed, we write

$$\boxed{\mathcal{U}(t_0,t) \;=\; \mathcal{U}^{-1}(t,t_0) \;=\; \mathcal{U}^{\dagger}(t,t_0)\,.} \tag{2.14}$$

Simply said, inverses or hermitian conjugation of $\mathcal{U}$ reverse the order of the time arguments.

2.2 Deriving the Schrödinger equation

The time evolution of states has been specified in terms of a unitary operator $\mathcal{U}$ assumed known. We now ask the 'reverse engineering' question. What kind of differential equation do the states satisfy for which the solution is unitary time evolution? The answer is simple and satisfying: a Schrödinger equation.

To obtain this result, we take the time derivative of (2.3) to find

$$\frac{\partial}{\partial t}|\Psi,t\rangle \;=\; \frac{\partial \mathcal{U}(t,t_0)}{\partial t}|\Psi,t_0\rangle\,. \tag{2.15}$$

We want the right hand side to involve the ket $|\Psi,t\rangle$ so we write

$$\frac{\partial}{\partial t}|\Psi,t\rangle \;=\; \frac{\partial \mathcal{U}(t,t_0)}{\partial t}\,\mathcal{U}(t_0,t)|\Psi,t\rangle\,. \tag{2.16}$$

Finally, it is convenient to have the same kind of $\mathcal{U}$ operator appearing, so we trade the order of times in the second $\mathcal{U}$ for a dagger:

$$\frac{\partial}{\partial t}|\Psi,t\rangle \;=\; \frac{\partial \mathcal{U}(t,t_0)}{\partial t}\,\mathcal{U}^{\dagger}(t,t_0)|\Psi,t\rangle\,. \tag{2.17}$$

This now looks like a differential equation for the state $|\Psi,t\rangle$. Let us introduce a name for the operator acting on the state in the right-hand side:

$$\frac{\partial}{\partial t}|\Psi,t\rangle \;=\; \Lambda(t,t_0)|\Psi,t\rangle\,, \tag{2.18}$$

where

$$\Lambda(t,t_0) \;\equiv\; \frac{\partial \mathcal{U}(t,t_0)}{\partial t}\,\mathcal{U}^{\dagger}(t,t_0)\,. \tag{2.19}$$

The operator Λ has units of inverse time. Note also that

$$\Lambda^\dagger(t,t_0) \;=\; \mathcal{U}(t,t_0)\,\frac{\partial \mathcal{U}^\dagger(t,t_0)}{\partial t}\,, \tag{2.20}$$

since the adjoint operation changes the order of operators and does not interfere with the time derivative.

We now want to prove two important facts about Λ:

1. $\Lambda(t,t_0)$ is antihermitian. To prove this begin with the equation

$$\mathcal{U}(t,t_0)\,\mathcal{U}^\dagger(t,t_0) \;=\; \mathbf{1}\,, \tag{2.21}$$

and take a derivative with respect to time to find,

$$\frac{\partial \mathcal{U}(t,t_0)}{\partial t}\,\mathcal{U}^\dagger(t,t_0) + \mathcal{U}(t,t_0)\,\frac{\partial \mathcal{U}^\dagger(t,t_0)}{\partial t} \;=\; 0\,. \tag{2.22}$$

Glancing at (2.19) and (2.20) we see that we got

$$\Lambda(t,t_0) + \Lambda^\dagger(t,t_0) \;=\; 0\,, \tag{2.23}$$

proving that $\Lambda(t,t_0)$ is indeed anti-hermitian.

2. $\Lambda(t,t_0)$ is actually independent of t_0. This is important because in the differential equation (2.17) t_0 appears nowhere except in Λ. To prove this independence we will show that $\Lambda(t,t_0)$ is actually equal to $\Lambda(t,t_1)$ for any other time t_1 different from t_0. So its value cannot depend on t_0. Or said differently, imagine $t_1 = t_0 + \epsilon$, then $\Lambda(t,t_0) = \Lambda(t,t_0+\epsilon)$ and as a result $\frac{\partial \Lambda(t,t_0)}{\partial t_0} \;=\; 0$. To prove the claim we begin with (2.19) and insert the unit operator in between the two factors

$$\begin{aligned}
\Lambda(t,t_0) \;&=\; \frac{\partial \mathcal{U}(t,t_0)}{\partial t}\,\mathcal{U}^\dagger(t,t_0) \\
&= \frac{\partial \mathcal{U}(t,t_0)}{\partial t}\Big(\mathcal{U}(t_0,t_1)\mathcal{U}^\dagger(t_0,t_1)\Big)\,\mathcal{U}^\dagger(t,t_0) \\
&= \frac{\partial}{\partial t}\Big(\mathcal{U}(t,t_0)\mathcal{U}(t_0,t_1)\Big)\mathcal{U}^\dagger(t_0,t_1)\,\mathcal{U}^\dagger(t,t_0) \\
&= \frac{\partial \mathcal{U}(t,t_1)}{\partial t}\,\mathcal{U}(t_1,t_0)\,\mathcal{U}(t_0,t) \;=\; \frac{\partial \mathcal{U}(t,t_1)}{\partial t}\,\mathcal{U}(t_1,t) \\
&= \frac{\partial \mathcal{U}(t,t_1)}{\partial t}\,\mathcal{U}^\dagger(t,t_1) \;=\; \Lambda(t,t_1)\,,
\end{aligned} \tag{2.24}$$

as we wanted to prove.

It follows that we can write $\Lambda(t) \equiv \Lambda(t, t_0)$, and thus equation (2.18) becomes

$$\frac{\partial}{\partial t}|\Psi, t\rangle \;=\; \Lambda(t)|\Psi, t\rangle\,. \tag{2.25}$$

We can define an operator $H(t)$ by multiplication of Λ by $i\hbar$:

$$\boxed{H(t) \;\equiv\; i\hbar\Lambda(t) \;=\; i\hbar\,\frac{\partial \mathcal{U}(t, t_0)}{\partial t}\,\mathcal{U}^\dagger(t, t_0)\,.} \tag{2.26}$$

Since Λ is antihermitian and has units of inverse time, $H(t)$ is a *hermitian* operator with units of energy. Multiplying (2.25) by $i\hbar$ we find the Schrödinger equation:

$$\boxed{\text{Schrödinger equation:}\quad i\hbar\frac{\partial}{\partial t}|\Psi, t\rangle \;=\; H(t)\,|\Psi, t\rangle\,.} \tag{2.27}$$

This is our main result. Unitary time evolution implies this equation. In this derivation the Hamiltonian is obtained from the knowledge of $\mathcal{U}$, as shown in (2.26). In most familiar situations, we know the Hamiltonian and wish to calculate the time evolution operator $\mathcal{U}$.

There are basically two reasons why the quantity $H(t)$ appearing in (2.27) is identified with the Hamiltonian. First, in quantum mechanics the momentum operator is given by $\hbar/i$ times the derivative with respect to a spatial coordinate. In special relativity energy corresponds to the time component of the momentum four-vector and thus it is reasonable to view it as an operator proportional to a time derivative. Second, we have used (2.27) to derive an equation for the time evolution of expectation values of observables. For an observable Q this took the form

$$\frac{d\langle Q\rangle}{dt} \;=\; \frac{1}{i\hbar}\langle[Q, H]\rangle \tag{2.28}$$

This equation is a natural generalization of the classical mechanics Hamiltonian equations and $H(t)$ plays a role analogous to that of the classical Hamiltonian. Indeed, in classical mechanics one has Poisson brackets $\{\cdot\,,\cdot\}_{pb}$ defined for functions of x and p by

$$\{A, B\}_{pb} \;=\; \frac{\partial A}{\partial x}\frac{\partial B}{\partial p} - \frac{\partial A}{\partial p}\frac{\partial B}{\partial x} \tag{2.29}$$

It then turns out that for any observable function $Q(x, p)$, its time derivative is given by taking the Poisson bracket of Q with the Hamiltonian:

$$\frac{dQ}{dt} \;=\; \{Q, H\}_{pb} \tag{2.30}$$

The similarity to (2.28) is quite striking. In fact, one can view commutators as essentially $\hbar$ times Poisson brackets

$$[A, B] \quad\Longleftrightarrow\quad i\hbar\,\{A, B\}_{pb} \tag{2.31}$$

Indeed $[x,p] = i\hbar$ while $\{x,p\}_{pb} = 1$. While these reasons justify our calling of H in the Schrödinger equation the Hamiltonian, ultimately we can say that any Hermitian operator with units of energy has the right to be called a Hamiltonian, regardless of any connection to a classical theory.

The Schrödinger wavefunction $\Psi(x,t)$ is defined by

$$\Psi(x,t) \;\equiv\; \langle x|\Psi,t\rangle\,. \tag{2.32}$$

If we hit (2.27) with the position state $\langle x|$ from the left we get

$$i\hbar\,\frac{\partial\Psi(x,t)}{\partial t} \;=\; \langle x|H(t)|\Psi,t\rangle\,. \tag{2.33}$$

If, moreover,

$$H(t) \;=\; \frac{\hat{p}^2}{2m} + V(\hat{x})\,, \tag{2.34}$$

then the equation becomes

$$i\hbar\frac{\partial\Psi(x,t)}{\partial t} \;=\; \Big(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x)\Big)\Psi(x,t)\,. \tag{2.35}$$

This is the familiar form of the Schrödinger equation for one-dimensional potentials.

2.3 Calculation of the unitary time evolution operator

The typical situation is one where the Hamiltonian $H(t)$ is known and we wish to calculate the unitary operator $\mathcal{U}$ that implements time evolution. From equation (2.26), multiplying by $\mathcal{U}(t,t_0)$ from the right gives

$$i\hbar\,\frac{\partial\mathcal{U}(t,t_0)}{\partial t} \;=\; H(t)\,\mathcal{U}(t,t_0)\,. \tag{2.36}$$

This is viewed as a differential equation for the *operator* $\mathcal{U}$. Note also that letting both sides of this equation act on $|\Psi,t_0\rangle$ gives us back the Schrödinger equation.

Since there is no possible confusion with the time derivatives, we do not need to write them as partial derivatives. Then the above equation takes the form

$$\frac{d\mathcal{U}}{dt} \;=\; -\frac{i}{\hbar}H(t)\,\mathcal{U}(t)\,. \tag{2.37}$$

If we view operators as matrices, this is a differential equation for the matrix $\mathcal{U}$. Solving this equation is in general quite difficult. We will consider three cases of increasing complexity.

Case 1. H is time independent. In this case, equation (2.37) is structurally of the form

$$\frac{d\mathcal{U}}{dt} \;=\; K\,\mathcal{U}(t)\,,\quad K \;=\; -\frac{i}{\hbar}\,H\,, \tag{2.38}$$

where $\mathcal{U}$ is a time dependent matrix, and K is a time-independent matrix. If the matrices were one-by-one, this reduces to the plain differential equation

$$\frac{du}{dt} = ku(t) \quad \rightarrow \quad u(t) \;=\; e^{kt}u(0)\,. \tag{2.39}$$

For the matrix case (2.38) we claim that

$$\mathcal{U}(t) \;=\; e^{tK}\,\mathcal{U}(0)\,. \tag{2.40}$$

Here we have the exponential of a matrix multiplied from the right by the matrix $\mathcal{U}$ at time equal zero. At $t = 0$ the ansatz gives the proper result, by construction. The exponential of a matrix is defined by the Taylor series

$$e^{tK} \;=\; 1 + tK + \frac{1}{2!}(tK)^2 + \frac{1}{3!}(tK)^3 + \cdots \;=\; \sum_{n=0}^{\infty} \frac{1}{n!}\, t^n K^n \tag{2.41}$$

Therefore it follows that the derivative takes the familiar simple form

$$\frac{d}{dt}e^{tK} \;=\; Ke^{tK} \;=\; e^{tK}K\,. \tag{2.42}$$

With this result we readily verify that (2.40) does solve (2.38):

$$\frac{d\mathcal{U}}{dt} \;=\; \frac{d}{dt}\left(e^{tK}\mathcal{U}(0)\right) \;=\; Ke^{tK}\mathcal{U}(0) \;=\; K\mathcal{U}(t)\,. \tag{2.43}$$

Using the explicit form of the matrix K the solution is therefore

$$\mathcal{U}(t, t_0) \;=\; e^{-\frac{i}{\hbar}Ht}\,\mathcal{U}_0\,, \tag{2.44}$$

where $\mathcal{U}_0$ is a constant matrix. Recalling that $\mathcal{U}(t_0, t_0) = \mathbf{1}$, we have $\mathcal{U}_0 = e^{iHt_0/\hbar}$ and therefore the full solution is

$$\boxed{\mathcal{U}(t, t_0) \;=\; \exp\Big[-\frac{i}{\hbar}H(t - t_0)\Big]\,, \quad \text{Time-independent } H\,.} \tag{2.45}$$

Exercise. Verify that the ansatz $\mathcal{U}(t) = \mathcal{U}(0)e^{tK}$, consistent for $t = 0$, would have not provided a solution of (2.38).

Case 2. $[H(t_1)\,, H(t_2)\,] = 0$ for all t_1, t_2. Here the Hamiltonian is time dependent but, despite this, the Hamiltonian at different times commute. One example is provided by the Hamiltonian for a spin in a magnetic field of time-dependent magnitude but constant direction.

We claim that the time evolution operator is now given by

$$\boxed{\mathcal{U}(t, t_0) \;=\; \exp\Big[-\frac{i}{\hbar}\int_{t_0}^{t} dt' H(t')\Big]\,.} \tag{2.46}$$

If the Hamiltonian is time independent, the above solution reduces correctly to (2.45). To prove that (2.46) solves the differential equation (2.37) we streamline notation by writing

$$R(t) \;\equiv\; -\frac{i}{\hbar}\int_{t_0}^{t} dt' H(t') \quad \to \quad R' \;=\; -\frac{i}{\hbar}\, H(t)\,, \tag{2.47}$$

where primes denote time derivatives. We claim that $R'(t)$ and $R(t)$ commute. Indeed

$$\big[R'(t)\,,R(t)\big] \;=\; \Big[-\frac{i}{\hbar}\,H(t)\,,-\frac{i}{\hbar}\int_{t_0}^{t} dt' H(t')\Big] \;=\; \Big(-\frac{i}{\hbar}\Big)^2\int_{t_0}^{t} dt'\,\big[H(t)\,,H(t')\big] \;=\; 0\,. \tag{2.48}$$

The claimed solution is

$$\mathcal{U} \;=\; \exp R(t) \;=\; 1 + R(t) + \frac{1}{2}R(t)R(t) + \frac{1}{3!}R(t)R(t)R(t) + \ldots \tag{2.49}$$

We have to take the time derivative of $\mathcal{U}$ and this time we do it slowly(!):

$$\begin{aligned}\frac{d}{dt}\mathcal{U} \;=\; \frac{d}{dt}\exp R \;&=\; R' + \frac{1}{2}(R'R + RR') + \frac{1}{3!}(R'RR + RR'R + RRR') + \ldots\,,\\ &=\; R' + R'R + \frac{1}{2!}R'RR + \ldots \;=\; R'\exp(R)\end{aligned} \tag{2.50}$$

The lesson here is that the derivative of $\exp R$ is simple if R' commutes with R. We have thus obtained

$$\frac{d}{dt}\mathcal{U} \;=\; -\frac{i}{\hbar}H(t)\,\mathcal{U}\,, \tag{2.51}$$

which is exactly what we wanted to show.

Case 3. $[H(t_1)\,,H(t_2)\,] \neq 0$. This is the most general situation and there is only a series solution. We write it here even though it will not be needed in our work. The solution for $\mathcal{U}$ is given by the so-called 'time-ordered' exponential, denoted by the symbol T in front of an exponential

$$\begin{aligned}\mathcal{U}(t,t_0) \;=\; \mathrm{T}\exp\Big[-\frac{i}{\hbar}\int_{t_0}^{t} dt' H(t')\Big] \;\equiv\; 1 &+ \Big(-\frac{i}{\hbar}\Big)\int_{t_0}^{t} dt_1 H(t_1)\\ &+ \Big(-\frac{i}{\hbar}\Big)^2\int_{t_0}^{t} dt_1 H(t_1)\int_{t_0}^{t_1} dt_2 H(t_2)\\ &+ \Big(-\frac{i}{\hbar}\Big)^3\int_{t_0}^{t} dt_1 H(t_1)\int_{t_0}^{t_1} dt_2 H(t_2)\int_{t_0}^{t_2} dt_3 H(t_3)\\ &+ \ldots\,.\end{aligned} \tag{2.52}$$

The term time-ordered refers to the fact that in the n-th term of the series we have a product $H(t_1)H(t_2)H(t_3)\ldots H(t_n)$ of *non-commuting* operators with integration ranges that force ordered times $t_1 \geq t_2 \geq t_3 \cdots \geq t_n$.

3 Heisenberg dynamics

The idea here is to confine the dynamical evolution to the operators. We will 'fold' the time dependence of the states into the operators. Since the objects we usually calculate are time-dependent expectation values of operators, this approach turns to be quite effective.

We will define time-dependent Heisenberg operators starting from Schrödinger operators. In fact, to any Schrödinger operator we can associate its corresponding Heisenberg operator. Schrödinger operators come in two types, time independent ones (like $\hat{x}, \hat{p}$) and time dependent ones (like Hamiltonians with time-dependent potentials). For each of those types of operators we will associate Heisenberg operators.

3.1 Heisenberg operators

Let us consider a Schrödinger operator $\hat{A}_S$, with the subscript S for Schrödinger. This operator may or may not have time dependence. We now examine a matrix element of $\hat{A}_S$ in between time dependent states $|\alpha, t\rangle$ and $|\beta, t\rangle$ and use the time-evolution operator to convert the states to time zero:

$$\langle \alpha, t|\hat{A}_S|\beta, t\rangle \;=\; \langle \alpha, 0|\,\mathcal{U}^\dagger(t,0)\,\hat{A}_S\,\mathcal{U}(t,0)\,|\beta, 0\rangle\,. \tag{3.1}$$

We simply define the Heisenberg operator $\hat{A}_H(t)$ associated with $\hat{A}_S$ as the object in between the time equal zero states:

$$\boxed{\hat{A}_H(t) \;\equiv\; \mathcal{U}^\dagger(t,0)\,\hat{A}_S\,\mathcal{U}(t,0)\,.} \tag{3.2}$$

Let us consider a number of important consequences of this definition.

1. At $t = 0$ the Heisenberg operator becomes equal to the Schrödinger operator:

$$\hat{A}_H(0) \;=\; \hat{A}_S\,. \tag{3.3}$$

The Heisenberg operator associated with the unit operator is the unit operator:

$$\mathbf{1}_H \;=\; \mathcal{U}^\dagger(t,0)\,\mathbf{1}\,\mathcal{U}(t,0) \;=\; \mathbf{1}\,. \tag{3.4}$$

2. The Heisenberg operator associated with the product of Schrödinger operators is equal to the product of the corresponding Heisenberg operators:

$$\hat{C}_S \;=\; \hat{A}_S\hat{B}_S \quad\rightarrow\quad \hat{C}_H(t) \;=\; \hat{A}_H(t)\hat{B}_H(t)\,. \tag{3.5}$$

Indeed,

$$\begin{aligned}\hat{C}_H(t) \;&= \mathcal{U}^\dagger(t,0)\,\hat{C}_S\,\mathcal{U}(t,0) \;=\; \mathcal{U}^\dagger(t,0)\,\hat{A}_S\hat{B}_S\,\mathcal{U}(t,0)\\ &= \hat{\mathcal{U}}^\dagger(t,0)\,\hat{A}_S\,\mathcal{U}(t,0)\mathcal{U}^\dagger(t,0)\,\hat{B}_S\,\mathcal{U}(t,0) \;=\; \hat{A}_H(t)\hat{B}_H(t)\,.\end{aligned} \tag{3.6}$$

3. It also follows from (3.5) that if we have a commutator of Schrödinger operators the corresponding Heisenberg operators satisfy the same commutation relations

$$[\hat{A}_S,\, \hat{B}_S\,] \;=\; C_S \quad \to \quad [\hat{A}_H(t),\, \hat{B}_H(t)\,] \;=\; \hat{C}_H(t)\,. \tag{3.7}$$

Since $\mathbf{1}_H = \mathbf{1}$, eqn. (3.7) implies that, for example,

$$[\hat{x}\,,\hat{p}\,] \;=\; i\hbar\,\mathbf{1} \quad \to \quad [\hat{x}_H(t)\,,\hat{p}_H(t)\,] \;=\; i\hbar\,\mathbf{1}\,. \tag{3.8}$$

4. Schrödinger and Heisenberg Hamiltonians. Assume we have a Schrödinger Hamiltonian that depends on some Schrödinger momenta and position operators $\hat{p}$ and $\hat{x}$, as in

$$H_S(\hat{p},\hat{x}\,;\,t)\,. \tag{3.9}$$

Since the $\hat{x}$ and $\hat{p}$ in H_S appear in products, property 2 implies that the associated Heisenberg Hamiltonian H_H takes the same form, with $\hat{x}$ and $\hat{p}$ replaced by their Heisenberg counterparts

$$H_H(t) \;=\; H_S(\hat{p}_H(t)\,,\hat{x}_H(t)\,;\,t)\,. \tag{3.10}$$

5. Equality of Hamiltonians. Under some circumstances the Heisenberg Hamiltonian is in fact equal to the Schrödinger Hamiltonian. Recall the definition

$$H_H(t) \;=\; \mathcal{U}^\dagger(t,0)\,H_S(t)\,\mathcal{U}(t,0)\,. \tag{3.11}$$

Assume now that $[H_S(t), H_S(t')] = 0$. Then (2.46) gives the time evolution operator

$$\mathcal{U}(t,0) \;=\; \exp\Big[-\frac{i}{\hbar}\int_0^t dt' H_S(t')\Big]\,. \tag{3.12}$$

Since the H_S at different times commute, $H_S(t)$ commutes both with $\mathcal{U}(t,0)$ and $\mathcal{U}^\dagger(t,0)$. Therefore the $H_S(t)$ can be moved, say to the right, in (3.11) giving us

$$H_H(t) \;=\; H_S(t)\,, \quad \text{when} \;\; [H_S(t), H_S(t')] = 0\,. \tag{3.13}$$

The meaning of this relation becomes clearer when we use (3.10) and (3.9) to write

$$H_S(\hat{p}_H(t)\,,\hat{x}_H(t)\,;\,t) \;=\; H_S(\hat{p}\,,\hat{x}\,;\,t)\,. \tag{3.14}$$

Operationally, this means that if we take $\hat{x}_H(t)$ and $\hat{p}_H(t)$ and plug them into the Hamiltonian (left-hand side), the result is as if we had simply plugged $\hat{x}$ and $\hat{p}$. We will confirm this for the case of the simple harmonic oscillator.

6. Equality of operators. If a Schrödinger operator A_S commutes with the Hamiltonian $H_S(t)$ for all times then A_S commutes with $\mathcal{U}(t,0)$ since this operator (even in the most complicated of cases) is built using $H_S(t)$. It follows that $A_H(t) = A_S$; the Heisenberg operator is equal to the Schrödinger operator.

7. Expectation values. Consider (3.1) and let $|\alpha, t\rangle = |\beta, t\rangle = |\Psi, t\rangle$. The matrix element now becomes an expectation value and we have:

$$\boxed{\langle \Psi, t|\hat{A}_S|\Psi, t\rangle \;=\; \langle \Psi, 0|\, \hat{A}_H(t)\, |\Psi, 0\rangle\,.} \tag{3.15}$$

With a little abuse of notation, we simply write this equation as

$$\boxed{\left\langle \hat{A}_S \right\rangle \;=\; \left\langle \hat{A}_H(t) \right\rangle.} \tag{3.16}$$

You should realize when writing such an equation that on the left hand side you compute the expectation value using the time-dependent state, while on the right-hand side you compute the expectation value using the state at time equal zero. If you prefer you can write out the equation as in (3.15) in case you think there is a possible confusion.

3.2 Heisenberg equation of motion

We can calculate the Heisenberg operator associated with a Schrödinger one using the definition (3.2). Alternatively, Heisenberg operators satisfy a differential equation: the Heisenberg equation of motion. This equation looks very much like the equations of motion of classical dynamical variables. So much so, that people trying to invent quantum theories sometimes begin with the equations of motion of some classical system and they postulate the existence of Heisenberg operators that satisfy similar equations. In that case they must also find a Heisenberg Hamiltonian and show that the equations of motion indeed arise in the quantum theory.

To determine the equation of motion of Heisenberg operators we will simply take time derivatives of the definition (3.2). For this purpose we recall (2.36) which we copy here using the subscript S for the Hamiltonian:

$$i\hbar\, \frac{\partial \mathcal{U}(t, t_0)}{\partial t} \;=\; H_S(t)\, \mathcal{U}(t, t_0)\,. \tag{3.17}$$

Taking the adjoint of this equation we find

$$i\hbar\, \frac{\partial \mathcal{U}^\dagger(t, t_0)}{\partial t} \;=\; -\mathcal{U}^\dagger(t, t_0)\, H_S(t)\,. \tag{3.18}$$

We can now calculate. Using (3.2) we find

$$\begin{aligned} i\hbar \frac{d}{dt} \hat{A}_H(t) \;=\; & \left(i\hbar \frac{\partial \mathcal{U}^\dagger}{\partial t}(t, 0) \right) \hat{A}_S(t)\, \mathcal{U}(t, 0) \\ & + \mathcal{U}^\dagger(t, 0)\, \hat{A}_S(t) \left(i\hbar \frac{\partial \mathcal{U}}{\partial t}(t, 0) \right) \\ & + \mathcal{U}^\dagger(t, 0)\, i\hbar \frac{\partial \hat{A}_S(t)}{\partial t}\, \mathcal{U}(t, 0) \end{aligned} \tag{3.19}$$

Using (3.17) and (3.18) we find

$$\begin{aligned} i\hbar\frac{d}{dt}\hat{A}_H(t) \;=\; & -\mathcal{U}^\dagger(t,0)\, H_S(t)\, \hat{A}_S(t)\, \mathcal{U}(t,0) \\ & +\mathcal{U}^\dagger(t,0)\, \hat{A}_S(t)\, H_S(t)\, \mathcal{U}(t,0) \\ & +\mathcal{U}^\dagger(t,0)\, i\hbar\frac{\partial \hat{A}_S(t)}{\partial t}\, \mathcal{U}(t,0) \end{aligned} \tag{3.20}$$

We now use (3.5) and recognize that in the last line we have the Heisenberg operator associated with the time derivative of $\hat{A}_S$:

$$i\hbar\frac{d}{dt}\hat{A}_H(t) \;=\; -H_H(t)\,\hat{A}_H(t) \;+\hat{A}_H(t)\,H_H(t) \;+i\hbar\Big(\frac{\partial \hat{A}_S(t)}{\partial t}\Big)_H \tag{3.21}$$

We now recognize a commutator on the right-hand side, so that our final result is

$$\boxed{\; i\hbar\,\frac{d\hat{A}_H(t)}{dt} \;=\; \big[\,\hat{A}_H(t)\,,\,H_H(t)\big] \;+\; i\hbar\,\Big(\frac{\partial \hat{A}_S(t)}{\partial t}\Big)_H . \;} \tag{3.22}$$

A few comments are in order.

1. Schrödinger operators without time dependence. If the operator $\hat{A}_S$ has no explicit time dependence then the last term in (3.22) vanishes and we have the simpler

$$i\hbar\,\frac{d\hat{A}_H(t)}{dt} \;=\; \big[\,\hat{A}_H(t)\,,\,H_H(t)\big]\,. \tag{3.23}$$

2. Time dependence of expectation values. Let A_S be a Schrödinger operator without time dependence. Let us now take the time derivative of the expectation value relation in (3.15):

$$\begin{aligned} i\hbar\frac{d}{dt}\langle\Psi,t|\hat{A}_S|\Psi,t\rangle \;=\; i\hbar\frac{d}{dt}\langle\Psi,0|\,\hat{A}_H(t)\,|\Psi,0\rangle \;&= \langle\Psi,0|\;i\hbar\,\frac{d\hat{A}_H(t)}{dt}\,|\Psi,0\rangle \\ &= \langle\Psi,0|\;\big[\,\hat{A}_H(t)\,,\,H_H(t)\big]\,|\Psi,0\rangle \end{aligned} \tag{3.24}$$

We write this as

$$\boxed{\; i\hbar\frac{d}{dt}\,\langle\,\hat{A}_H(t)\rangle \;=\; \big\langle\,\big[\,\hat{A}_H(t)\,,\,H_H(t)\big]\big\rangle\,. \;} \tag{3.25}$$

Notice that this equation takes exactly the same form in the Schrödinger picture (recall the comments below (3.16):

$$\boxed{\; i\hbar\frac{d}{dt}\,\langle\,\hat{A}_S\,\rangle \;=\; \big\langle\,\big[\,\hat{A}_S\,,\,H_S\big]\big\rangle\,. \;} \tag{3.26}$$

3. A time-independent operator $\hat{A}_S$ is said to be **conserved** if it commutes with the Hamiltonian:

$$\text{Conserved operator } \hat{A}_S\text{:} \quad [\,\hat{A}_S\,, H_S\,] \;=\; 0\,. \tag{3.27}$$

It then follows that $[\,\hat{A}_H(t)\,, H_H(t)\,] \;= 0$, and using (3.23) that

$$\frac{d\hat{A}_H(t)}{dt} \;= 0\,. \tag{3.28}$$

The Heisenberg operator is plain constant. Thus the expectation value of the operator is also constant. This is consistent with comment 6 in the previous section: $\hat{A}_H$ is in fact equal to $\hat{A}_S$!

3.3 Three examples.

Example 1. Part of the Homework. We just discuss here a few facts. Consider the Hamiltonian

$$H \;=\; \frac{\hat{p}^2}{2m} + V(\hat{x})\,, \tag{3.29}$$

where $V(x)$ is a potential. You will show that

$$\begin{aligned} \frac{d}{dt}\langle\hat{x}\rangle \;&=\; \frac{1}{m}\langle\hat{p}\rangle\,, \\ \frac{d}{dt}\langle\hat{p}\rangle \;&=\; -\Big\langle\frac{\partial V}{\partial\hat{x}}\Big\rangle\,. \end{aligned} \tag{3.30}$$

These two equations combined give

$$m\,\frac{d^2}{dt^2}\langle\hat{x}\rangle \;=\; -\Big\langle\frac{\partial V}{\partial\hat{x}}\Big\rangle\,. \tag{3.31}$$

This is the quantum analog of the classical equation

$$m\frac{d^2}{dt^2}x(t) \;=\; -\frac{\partial V}{\partial x}\,, \tag{3.32}$$

which describes the classical motion of a particle of mass m in a potential $V(x)$. Note that the force is $F = -\frac{\partial V}{\partial x}$.

Example 2. Harmonic oscillator. The Schrödinger Hamiltonian is

$$H_S \;=\; \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2\hat{x}^2\,, \tag{3.33}$$

and is time independent. Using (3.10) we note that the Heisenberg Hamiltonian takes the form

$$H_H(t) \;=\; \frac{\hat{p}_H^2(t)}{2m} + \frac{1}{2}m\omega^2\hat{x}_H^2(t)\,. \tag{3.34}$$

Consider now the Schrödinger operators $\hat{x}$ and $\hat{p}$. Using the Heisenberg equation of motion, we have for $\hat{x}$:

$$\begin{aligned} \frac{d}{dt}\,\hat{x}_H(t) \;&=\; \frac{1}{i\hbar}\big[\hat{x}_H(t), H_H(t)\big] \;=\; \frac{1}{i\hbar}\Big[\hat{x}_H(t), \frac{\hat{p}_H^2(t)}{2m}\Big] \\ &=\; \frac{1}{i\hbar}\,2\,\frac{\hat{p}_H(t)}{2m}\Big[\hat{x}_H(t), \hat{p}_H(t)\Big] \;=\; \frac{1}{i\hbar}\,\frac{\hat{p}_H(t)}{m}\,i\hbar \;=\; \frac{\hat{p}_H(t)}{m}\,, \end{aligned} \tag{3.35}$$

so that our first equation is

$$\boxed{\frac{d}{dt}\,\hat{x}_H(t) \;=\; \frac{\hat{p}_H(t)}{m}\,.} \tag{3.36}$$

For the momentum operator we get

$$\begin{aligned} \frac{d}{dt}\,\hat{p}_H(t) \;&=\; \frac{1}{i\hbar}\big[\hat{p}_H(t), H_H(t)\big] \;=\; \frac{1}{i\hbar}\Big[\hat{p}_H(t),\; \frac{1}{2}m\omega^2\,x_H^2(t)\Big] \\ &=\; \frac{1}{i\hbar}\,\frac{1}{2}m\omega^2\cdot 2(-i\hbar)\hat{x}_H(t) \;=\; -m\omega^2\,\hat{x}_H(t)\,, \end{aligned} \tag{3.37}$$

so our second equation is

$$\boxed{\frac{d}{dt}\,\hat{p}_H(t) \;=\; -m\omega^2\,\hat{x}_H(t)\,.} \tag{3.38}$$

Taking another time derivative of (3.36) and using (3.38) we get

$$\frac{d^2}{dt^2}\,\hat{x}_H(t) \;=\; -\omega^2\,\hat{x}_H(t)\,. \tag{3.39}$$

We now solve this differential equation. Being just an oscillator equation the solution is

$$\hat{x}_H(t) \;=\; \hat{A}\,\cos\omega t + \hat{B}\,\sin\omega t\,, \tag{3.40}$$

where $\hat{A}$ and $\hat{B}$ are time-independent operators to be determined by initial conditions. From (3.36) we can find the momentum operator

$$\hat{p}_H(t) \;=\; m\frac{d}{dt}\hat{x}_H(t) \;=\; -m\omega\hat{A}\sin\omega t + m\omega\hat{B}\cos\omega t\,. \tag{3.41}$$

At zero time the Heisenberg operators must equal the Schrödinger ones so,

$$\hat{x}_H(0) \;=\; \hat{A} \;=\; \hat{x}\,, \qquad \hat{p}_H(0) \;=\; m\omega\hat{B} \;=\hat{p}\,. \tag{3.42}$$

We have thus found that

$$\hat{A} \;=\hat{x}\,, \qquad \hat{B} \;=\; \frac{1}{m\omega}\,\hat{p}\,. \tag{3.43}$$

Finally, back in (3.40) and (3.41) we have our full solution for the Heisenberg operators of the SHO:

$$\boxed{\begin{aligned} \hat{x}_H(t) &= \hat{x}\,\cos\omega t + \frac{1}{m\omega}\,\hat{p}\,\sin\omega t\,, \\ \hat{p}_H(t) &= \hat{p}\cos\omega t - m\omega\,\hat{x}\,\sin\omega t\,. \end{aligned}} \tag{3.44}$$

Let us do a couple of small computations. Consider the energy eigenstate $|n\rangle$ of the harmonic oscillator:

$$|\psi, 0\rangle \;=\; |n\rangle\,. \tag{3.45}$$

We ask: What is the time-dependent expectation value of the x operator in this state? We compute

$$\langle\hat{x}\rangle \;=\; \langle\psi,t|\hat{x}|\psi,t\rangle \;=\; \langle\psi,0|\hat{x}_H(t)|\psi,0\rangle \;=\; \langle n|\hat{x}_H(t)|n\rangle\,. \tag{3.46}$$

Now we use the expression for $\hat{x}_H(t)$:

$$\langle\hat{x}\rangle \;=\; \langle n|\big(\hat{x}\,\cos\omega t + \frac{1}{m\omega}\,\hat{p}\,\sin\omega t\big)|n\rangle \;=\; \langle n|\hat{x}|n\rangle\,\cos\omega t + \;\langle n|\,\hat{p}\,|n\rangle\,\frac{1}{m\omega}\sin\omega t\,. \tag{3.47}$$

We now recall that $\langle n|\hat{x}|n\rangle = 0$ and $\langle n|\,\hat{p}\,|n\rangle = 0$. So as a result we find that on the energy eigenstate $|n\rangle$, the expectation value of x is zero at all times:

$$\langle\,\hat{x}\,\rangle \;=\; 0\,. \tag{3.48}$$

So energy eigenstates do not exhibit classical behavior (an oscillatory time-dependent $\langle\,\hat{x}\,\rangle$).

As a second calculation let us confirm that the Heisenberg Hamiltonian is time independent and in fact equal to the Schrödinger Hamiltonian. Starting with (3.34) and using (3.44) we have

$$\begin{aligned} H_H(t) =&\; \frac{\hat{p}_H^2(t)}{2m} + \frac{1}{2}m\omega^2\hat{x}_H^2(t) \\ =&\; \frac{1}{2m}\big(\hat{p}\cos\omega t - m\omega\,\hat{x}\,\sin\omega t\big)^2 + \frac{1}{2}m\omega^2\Big(\hat{x}\,\cos\omega t + \frac{1}{m\omega}\,\hat{p}\,\sin\omega t\Big)^2 \\ =&\; \frac{\cos^2\omega t}{2m}\hat{p}^2 + \frac{m^2\omega^2\sin^2\omega t}{2m}\,\hat{x}^2 - \frac{\omega}{2}\sin\omega t\cos\omega t(\hat{p}\hat{x}+\hat{x}\hat{p}) \\ &+ \frac{\sin^2\omega t}{2m}\hat{p}^2 + \frac{m\omega^2\cos^2\omega t}{2}\,\hat{x}^2 + \frac{\omega}{2}\cos\omega t\sin\omega t\,(\hat{x}\hat{p}+\hat{p}\hat{x}) \\ =&\; \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2\hat{x}^2\,. \end{aligned} \tag{3.49}$$

This is what we wanted to show.

Example 3. What are the Heisenberg operators corresponding to the simple harmonic oscillator creation and annihilation operators?

Given the Schrödinger operator $\hat{a}$, the Heisenberg operator would be denoted as $\hat{a}_H(t)$, but for simplicity we will just denote it as $\hat{a}(t)$. Since the harmonic oscillator Hamiltonian is time independent, we can use the definition

$$\hat{a}(t) \;\equiv\; e^{\frac{i}{\hbar}Ht}\,\hat{a}\,e^{-\frac{i}{\hbar}H\,t} \;=\; e^{i\omega\,t\hat{N}}\,a\,e^{-i\omega t\hat{N}}\,, \tag{3.50}$$

where we wrote $H = \hbar\omega(\hat{N} + \frac{1}{2})$ and noted that the additive constant has no effect on the commutator. A simple way to evaluate $a(t)$ goes through a differential equation. We take the time derivative of the above to find

$$\begin{aligned}\frac{d}{dt}\,\hat{a}(t) \;&=\; e^{i\omega\,t\hat{N}}(i\omega\hat{N})\,\hat{a}\,e^{-i\omega t\hat{N}} \;-\; e^{i\omega\,t\hat{N}}\,\hat{a}\,(i\omega\hat{N})e^{-i\omega t\hat{N}}\,,\\ &=\; i\omega\,e^{i\omega\,t\hat{N}}\,\big[\,\hat{N}\,,\hat{a}\,\big]e^{-i\omega t\hat{N}} \;=\; -i\omega\,e^{i\omega\,t\hat{N}}\;a\,e^{-i\omega t\hat{N}}\,.\end{aligned} \tag{3.51}$$

we recognize in final right-hand side the operator $a(t)$ so we have obtained the differential equation

$$\frac{d}{dt}\,\hat{a}(t) \;= -i\omega t\,\hat{a}(t)\,. \tag{3.52}$$

Since $\hat{a}(t=0) = \hat{a}$, the solution is

$$\hat{a}(t) \;=\; e^{-i\omega t}\,\hat{a}\,. \tag{3.53}$$

Together with the adjoint of this formula we have

$$\boxed{\begin{aligned}\hat{a}(t) \;&=\; e^{-i\omega t}\,\hat{a}\,.\\ \hat{a}^\dagger(t) \;&=\; e^{\,i\omega t}\;\hat{a}^\dagger\,.\end{aligned}} \tag{3.54}$$

The two equations above are our answer. As a check we consider the operator equation

$$\hat{x} \;=\; \sqrt{\frac{\hbar}{2m\omega}}\,(\hat{a} + \hat{a}^\dagger)\,, \tag{3.55}$$

whose Heisenberg version is

$$\hat{x}_H(t) \;=\; \sqrt{\frac{\hbar}{2m\omega}}\,(\hat{a}(t) + \hat{a}^\dagger(t)) \;=\; \sqrt{\frac{\hbar}{2m\omega}}\,(e^{-i\omega t}\hat{a} + e^{i\omega t}\hat{a}^\dagger)\,. \tag{3.56}$$

Expanding the exponentials, we recognize,

$$\begin{aligned}\hat{x}_H(t) \;&= \sqrt{\frac{\hbar}{2m\omega}}\,\Big((\hat{a} + \hat{a}^\dagger)\cos\omega t + i(\hat{a}^\dagger - \hat{a})\sin\omega t\Big)\,,\\ &= \hat{x}\cos\omega t + \frac{1}{m\omega}\,\hat{p}\,\sin\omega t\,,\end{aligned} \tag{3.57}$$

in agreement with (3.44).

4 Coherent states of the Harmonic oscillator

Coherent states are quantum states that exhibit some sort of classical behavior. We will introduce them and explore their properties. To begin our discussion we introduce translation operators.

4.1 Translation operator

Let us construct unitary *translation* operators T_{x_0} that acting on states moves them (or translates them) by a distance x_0, where x_0 is a real constant with units of length:

$$\boxed{\text{Translation operator:} \quad T_{x_0} \equiv e^{-\frac{i}{\hbar}\hat{p}\,x_0}\,.} \tag{4.1}$$

This operator is unitary because it is the exponential of an antihermitian operator ($\hat{p}$ is hermitian, and $i\hat{p}$ antihermitian). The multiplication of two such operators is simple:

$$T_{x_0}T_{y_0} \;=\; e^{-\frac{i}{\hbar}\hat{p}\,x_0}e^{-\frac{i}{\hbar}\hat{p}\,y_0} \;=\; e^{-\frac{i}{\hbar}\hat{p}\,(x_0+y_0)}\,, \tag{4.2}$$

since the exponents commute ($e^Ae^B = e^{A+B}$ if $[A,B]=0$). As a result

$$T_{x_0}T_{y_0} \;=\; T_{x_0+y_0}\,. \tag{4.3}$$

The translation operators form a group: the product of two translation is a translation. There is a unit element $T_0 = I$ corresponding to $x_0 = 0$, and each element T_{x_0} has an inverse T_{-x_0}. Note that the group multiplication rule is commutative.

It follows from the explicit definition of the translation operator that

$$(T_{x_0})^\dagger \;=\; e^{\frac{i}{\hbar}\hat{p}\,x_0} \;=\; e^{-\frac{i}{\hbar}\hat{p}\,(-x_0)} \;=\; T_{-x_0} \;=\; (T_{x_0})^{-1}\,. \tag{4.4}$$

confirming again that the operator is unitary. In the following we denote $(T_{x_0})^\dagger$ simply by $T^\dagger_{x_0}$.

We say that T_{x_0} translates by x_0 because of its action[2] on the operator $\hat{x}$ is as follows:

$$T^\dagger_{x_0}\,\hat{x}\,T_{x_0} \;=\; e^{\frac{i}{\hbar}\hat{p}\,x_0}\,\hat{x}\,e^{-\frac{i}{\hbar}\hat{p}\,x_0} \;=\; \hat{x} + \frac{i}{\hbar}\,[\hat{p},\hat{x}]x_0 \;=\; \hat{x} + x_0\,, \tag{4.5}$$

where we used the formula $e^ABe^{-A} = B + [A,B] + \ldots$ and the dots vanish in this case because $[A,B]$ is a number (check that you understand this!).

To see physically why the above is consistent with intuition, consider a state $|\psi\rangle$ and the expectation value of $\hat{x}$ on this state

$$\langle\hat{x}\rangle_\psi \;=\; \langle\psi|\,\hat{x}\,|\psi\rangle \tag{4.6}$$

[2]The action of a unitary operator $\mathcal{U}$ on an operator $\mathcal{O}$ is defined as $\mathcal{O} \to \mathcal{U}^\dagger\mathcal{O}\mathcal{U}$.

Now we ask: What is the expectation value of $\hat{x}$ on the state $T_{x_0}|\psi\rangle$? We find

$$\langle\hat{x}\rangle_{T_{x_0}\psi} \;=\; \langle\psi|T_{x_0}^\dagger\,\hat{x}\,T_{x_0}|\psi\rangle \tag{4.7}$$

The right-hand side explains why $T_{x_0}^\dagger\,\hat{x}\,T_{x_0}$ is the natural thing to compute! Indeed using our result for this

$$\langle\hat{x}\rangle_{T_{x_0}\psi} \;=\; \langle\psi|(\hat{x}+x_0)|\psi\rangle \;=\; \langle\hat{x}\rangle_{\psi} + x_0\,. \tag{4.8}$$

The expectation value of $\hat{x}$ on the displaced state is indeed equal to the expectation value of $\hat{x}$ in the original state plus x_0, confirming that *we should view $T_{x_0}|\psi\rangle$ as the state $|\psi\rangle$ displaced a distance x_0*.

As an example we look at position states. We claim that on position states the translation operator does what we expect:

$$\boxed{T_{x_0}|x_1\rangle \;=\; |x_1+x_0\rangle\,.} \tag{4.9}$$

We can prove (4.9) by acting on the above left-hand side an arbitrary momentum bra $\langle p|$:

$$\langle p|T_{x_0}|x_1\rangle \;=\; \langle p|e^{-\frac{i}{\hbar}\hat{p}x_0}|x_1\rangle \;=\; e^{-\frac{i}{\hbar}px_0}\frac{e^{-\frac{i}{\hbar}px_1}}{\sqrt{2\pi\hbar}} \;=\; \langle p|x_1+x_0\rangle\,, \tag{4.10}$$

proving the desired result, given that $\langle p|$ is arbitrary. It also follows from unitarity and (4.9) that

$$T_{x_0}^\dagger|x_1\rangle \;=\; T_{-x_0}|x_1\rangle \;=\; |x_1-x_0\rangle\,. \tag{4.11}$$

Taking the Hermitian conjugate we find

$$\boxed{\langle x_1|T_{x_0} \;=\; \langle x_1-x_0|\,.} \tag{4.12}$$

In terms of arbitrary states $|\psi\rangle$, we can also discuss the action of the translation operator by introducing the wavefunction $\psi(x) = \langle x|\psi\rangle$. Then the "translated" state $T_{x_0}|\psi\rangle$ has a wavefunction

$$\langle x|T_{x_0}|\psi\rangle \;=\; \langle x-x_0|\psi\rangle \;=\; \psi(x-x_0)\,. \tag{4.13}$$

Indeed, $\psi(x-x_0)$ is the function $\psi(x)$ translated by the distance $+x_0$. For example, the value that $\psi(x)$ takes at $x=0$ is taken by the function $\psi(x-x_0)$ at $x=x_0$.

4.2 Definition and basic properties of coherent states

We now finally introduce a coherent state $|\tilde{x}_0\rangle$ of the simple harmonic oscillator. The state is labeled by x_0 and the tilde is there to remind you that it is *not* a position state.[3] Here is the definition

$$\boxed{\text{Coherent state:} \quad |\tilde{x}_0\rangle \;\equiv\; T_{x_0}|0\rangle \;= e^{-\frac{i}{\hbar}\hat{p}x_0}|0\rangle\,,} \tag{4.14}$$

where $|0\rangle$ denotes the ground state of the oscillator. Do not confuse the coherent state with a position state. The coherent state is simply the translation of the ground state by a distance x_0. This state has no time dependence displayed, so it may be thought as the state of the system at $t = 0$. As t increases the state will evolve according to the Schrödinger equation, and we will be interested in this evolution, but not now. Note that the coherent state is well normalized

$$\langle\tilde{x}_0|\tilde{x}_0\rangle \;= \langle 0|T^\dagger_{x_0}T_{x_0}|0\rangle \;=\; \langle 0|0\rangle \;= 1\,. \tag{4.15}$$

This had to be so because T_{x_0} is unitary.

To begin with let us calculate the wavefunction associated to the state:

$$\psi_{x_0}(x) \;\equiv\; \langle x|\tilde{x}_0\rangle \;=\; \langle x|T_{x_0}|0\rangle \;=\; \langle x - x_0|0\rangle \;=\; \psi_0(x - x_0)\,, \tag{4.16}$$

where we used (4.12) and we denoted $\langle x|0\rangle = \psi_0(x)$, as the ground state wavefunction. So, as expected the wavefunction for the coherent state is just the ground state wavefunction displaced x_0 to the right. This is illustrated in Figure 2.

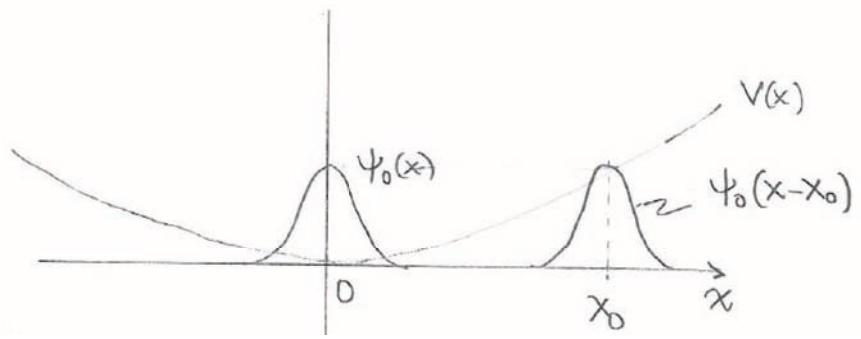


Figure 2: The ground state wavefunction $\psi_0(x)$ displaced to the right a distance x_0 is the wavefunction $\psi_0(x - x_0)$. The corresponding state, denoted as $|\tilde{x}_0\rangle$, is the simplest example of a coherent state.

Let us now do a few sample calculations to understand better these states.

[3]This is not great notation, but it is better than any alternative I have seen!

1. Calculate the expectation value of $\hat{x}$ in a coherent state.

$$\langle\tilde{x}_0|\hat{x}\,|\tilde{x}_0\rangle \;=\; \langle 0|T_{x_0}^\dagger\,\hat{x}\,T_{x_0}|0\rangle \;=\; \langle 0|(\hat{x}+x_0)|0\rangle\,, \tag{4.17}$$

where we used (4.5). Recalling now that $\langle 0|\hat{x}|0\rangle = 0$ we get

$$\langle\tilde{x}_0|\hat{x}\,|\tilde{x}_0\rangle \;=\; x_0\,. \tag{4.18}$$

Not that surprising! The position is essentially x_0.

2. Calculate the expectation value of $\hat{p}$ in a coherent state. Since $\hat{p}$ commutes with T_{x_0} we have

$$\langle\tilde{x}_0|\hat{p}\,|\tilde{x}_0\rangle \;=\; \langle 0|T_{x_0}^\dagger\,\hat{p}\,T_{x_0}|0\rangle \;=\; \langle 0|\,\hat{p}\,T_{x_0}^\dagger\,T_{x_0}|0\rangle \;=\; \langle 0|\,\hat{p}\,|0\rangle \;=\; 0\,, \tag{4.19}$$

The coherent state has no (initial) momentum. It has an initial position (as seen in 1. above)

3. Calculate the expectation value of the energy in a coherent state. Note that the coherent state is not an energy eigenstate (nor a position eigenstate, nor a momentum eigenstate!). With H the Hamiltonian we have

$$\langle\tilde{x}_0|H|x_0\rangle \;=\; \langle 0|T_{x_0}^\dagger H T_{x_0}|0\rangle\,. \tag{4.20}$$

We now compute

$$\begin{aligned} T_{x_0}^\dagger H T_{x_0} \;&=\; T_{x_0}^\dagger\Big(\frac{\hat{p}^2}{2m}+\frac{1}{2}m\omega^2\hat{x}^2\Big)T_{x_0} \;=\; \frac{\hat{p}^2}{2m}+\frac{1}{2}m\omega^2(\hat{x}+x_0)^2 \\ &=\; H + m\omega^2 x_0\hat{x}+\frac{1}{2}m\omega^2 x_0^2\,. \end{aligned} \tag{4.21}$$

where we recall that T_{x_0} commutes with $\hat{p}$ and used eqn. (4.5). Back in (4.20) we have

$$\langle\tilde{x}_0|H|x_0\rangle \;=\; \langle 0|H|0\rangle + m\omega^2 x_0\langle 0|\hat{x}|0\rangle + \frac{1}{2}m\omega^2 x_0^2\,. \tag{4.22}$$

Recalling that the ground state energy is $\hbar\omega/2$ and that in the ground state $\hat{x}$ has no expectation value we finally get

$$\langle\tilde{x}_0|H|x_0\rangle \;=\; \frac{1}{2}\hbar\omega + \frac{1}{2}m\omega^2 x_0^2\,. \tag{4.23}$$

This is reasonable: the total energy is the zero-point energy plus the potential energy of a particle at x_0. The coherent state $|\tilde{x}_0\rangle$ is the quantum version of a point particle on a spring held stretched to $x = x_0$.

4.3 Time evolution and uncertainties

Evolving the coherent states in time is a somewhat involved procedure that will be explained later. We can discuss time evolution quite easily using the Heisenberg picture, since we have already calculated in (3.44) the time-dependent Heisenberg operators $\hat{x}_H(t)$ and $\hat{p}_H(t)$.

If we have at time equal zero the coherent state $|\tilde{x}_0\rangle$ then at time t we write the time-evolved state as $|\tilde{x}_0, t\rangle$. We now ask what is the (time-dependent) expectation value of $\hat{x}$ on this state:

$$\langle\, \hat{x}\rangle(t) \;=\; \langle \tilde{x}_0, t|\, \hat{x}\, |\tilde{x}_0, t\rangle \;=\; \langle \tilde{x}_0|\, \hat{x}_H(t)\, |\tilde{x}_0\rangle\,. \tag{4.24}$$

Using (3.44) we get

$$\langle\, \hat{x}\rangle(t) \;=\; \langle \tilde{x}_0|\, \Big(\hat{x}\, \cos\omega t + \frac{1}{m\omega}\, \hat{p}\, \sin\omega t\Big)\, |\tilde{x}_0\rangle\,. \tag{4.25}$$

Finally, using (4.18) and (4.19) we get

$$\langle\, \hat{x}\rangle(t) \;=\; \langle \tilde{x}_0|\, \hat{x}_H(t)\, |\tilde{x}_0\rangle \;=\; x_0 \cos\omega t\,. \tag{4.26}$$

The expectation value of $\hat{x}$ is performing oscillatory motion! This confirms the classical interpretation of the coherent state. For the momentum the calculation is quite similar,

$$\langle\, \hat{p}\rangle(t) \;=\; \langle \tilde{x}_0|\, \hat{p}_H(t)\, |\tilde{x}_0\rangle \;=\; \langle \tilde{x}_0|\, \Big(\hat{p}\cos\omega t - m\omega\, \hat{x}\, \sin\omega t\Big)\, |\tilde{x}_0\rangle \tag{4.27}$$

and we thus find

$$\langle\, \hat{p}\rangle(t) \;=\; \langle \tilde{x}_0|\, \hat{p}_H(t)\, |\tilde{x}_0\rangle \;=\; -m\omega\, x_0\, \sin\omega t\,, \tag{4.28}$$

which is the expected result as it is equal to $m\frac{d}{dt}\langle\hat{x}\rangle(t)$.

We have seen that the harmonic oscillator ground state is a minimum uncertainty state. We will now discuss the extension of this fact to coherent states. We begin by calculating the uncertainties Δx and Δp in a coherent state at $t = 0$. We will see that the coherent state has minimum uncertainty for the product. Then we will calculate uncertainties of the coherent state as a function of time!

We have

$$\langle \tilde{x}_0|\hat{x}^2|\tilde{x}_0\rangle \;=\; \langle 0|T^\dagger_{x_0}\hat{x}^2 T_{x_0}|0\rangle \;=\; \langle 0|(\hat{x}+x_0)^2|0\rangle \;=\; \langle 0|\hat{x}^2|0\rangle + x_0^2\,. \tag{4.29}$$

The first term on the right-hand side was calculated in (1.58). We thus find

$$\langle \tilde{x}_0|\hat{x}^2|\tilde{x}_0\rangle \;=\; \frac{\hbar}{2m\omega} + x_0^2\,. \tag{4.30}$$

Since $\langle \tilde{x}_0|\hat{x}|\tilde{x}_0\rangle = x_0$ we find the uncertainty

$$(\Delta x)^2 \;=\; \langle \tilde{x}_0|\hat{x}^2|\tilde{x}_0\rangle - (\langle \tilde{x}_0|\hat{x}|\tilde{x}_0\rangle)^2 = \frac{\hbar}{2m\omega} + x_0^2 - x_0^2$$

$$\rightarrow \quad (\Delta x)^2 \;=\; \frac{\hbar}{2m\omega}\,, \quad \text{on the state } |\tilde{x}_0\rangle\,. \tag{4.31}$$

For the momentum operator we have, using (1.58),

$$\langle\tilde{x}_0|\hat{p}^2|\tilde{x}_0\rangle \;=\; \langle 0|T_{x_0}^\dagger \hat{p}^2 T_{x_0}|0\rangle \;=\; \langle 0|\hat{p}^2|0\rangle \;=\; \frac{m\hbar\omega}{2}\,. \tag{4.32}$$

Since $\langle\tilde{x}_0|\hat{p}|\tilde{x}_0\rangle = 0$, we have

$$(\Delta p)^2 \;=\; \frac{m\hbar\omega}{2}\,, \quad \text{on the state } |\tilde{x}_0\rangle\,. \tag{4.33}$$

As a result,

$$\Delta x \Delta p \;=\; \frac{\hbar}{2}\,, \quad \text{on the state } |\tilde{x}_0\rangle\,. \tag{4.34}$$

We see that the coherent state has minimum $\Delta x \Delta p$ at time equal zero. This is not surprising because at this time the state is just a displaced ground state.

For the time dependent situation we have

$$\begin{aligned}(\Delta x)^2(t) \;&=\; \langle\tilde{x}_0,t|\hat{x}^2|\tilde{x}_0,t\rangle - \big(\langle\tilde{x}_0,t|\hat{x}|\tilde{x}_0,t\rangle\big)^2 \\ &=\; \langle\tilde{x}_0|\hat{x}_H^2(t)|\tilde{x}_0\rangle - \big(\langle\tilde{x}_0|\hat{x}_H(t)|\tilde{x}_0\rangle\big)^2 \\ &=\; \langle\tilde{x}_0|\hat{x}_H^2(t)|\tilde{x}_0\rangle - x_0^2\cos^2\omega t\,,\end{aligned} \tag{4.35}$$

where we used the result in (4.26). The computation of the first term takes a few steps:

$$\begin{aligned}\langle\tilde{x}_0|\hat{x}_H^2(t)|\tilde{x}_0\rangle \;&=\; \langle\tilde{x}_0|\Big(\hat{x}\,\cos\omega t + \frac{1}{m\omega}\,\hat{p}\,\sin\omega t\Big)^2|\tilde{x}_0\rangle \\ &=\; \langle\tilde{x}_0|\hat{x}^2|\tilde{x}_0\rangle\cos^2\omega t + \langle\tilde{x}_0|\hat{p}^2|\tilde{x}_0\rangle\Big(\frac{\sin\omega t}{m\omega}\Big)^2 + \frac{\cos m\omega\sin m\omega}{m\omega}\langle\tilde{x}_0|\big(\hat{x}\hat{p}+\hat{p}\hat{x}\big)|\tilde{x}_0\rangle \\ &=\; \Big(\frac{\hbar}{2m\omega}+x_0^2\Big)\cos^2\omega t + \frac{m\hbar\omega}{2}\Big(\frac{\sin\omega t}{m\omega}\Big)^2 + \frac{\cos m\omega\sin m\omega}{m\omega}\langle\tilde{x}_0|\big(\hat{x}\hat{p}+\hat{p}\hat{x}\big)|\tilde{x}_0\rangle\,.\end{aligned}$$

We now show that the last expectation value vanishes:

$$\begin{aligned}\langle\tilde{x}_0|\big(\hat{x}\hat{p}+\hat{p}\hat{x}\big)|\tilde{x}_0\rangle \;&=\; \langle 0|\big((\hat{x}+x_0)\hat{p}+\hat{p}(\hat{x}+x_0)\big)|0\rangle \\ &=\; \langle 0|\big(\hat{x}\hat{p}+\hat{p}\hat{x}\big)|0\rangle \\ &=\; i\frac{\hbar}{2}\,\langle 0|\big((\hat{a}+\hat{a}^\dagger)(\hat{a}^\dagger-\hat{a})+(\hat{a}^\dagger-\hat{a})(\hat{a}+\hat{a}^\dagger)\big)|0\rangle \\ &=\; i\frac{\hbar}{2}\,\langle 0|\big(\hat{a}\hat{a}^\dagger+(-\hat{a})\hat{a}^\dagger\big)|0\rangle \;=\; 0\,.\end{aligned} \tag{4.36}$$

As a result,

$$\begin{aligned}\langle\tilde{x}_0|\hat{x}_H^2(t)|\tilde{x}_0\rangle \;&=\; \Big(\frac{\hbar}{2m\omega}+x_0^2\Big)\cos^2\omega t + \frac{m\hbar\omega}{2}\Big(\frac{\sin\omega t}{m\omega}\Big)^2 \\ &=\; \frac{\hbar}{2m\omega} + x_0^2\cos^2\omega t\,.\end{aligned} \tag{4.37}$$

Therefore, finally, back in (4.35) we get

$$(\Delta x)^2(t) \;=\; \frac{\hbar}{2m\omega}\,. \tag{4.38}$$

The uncertainty Δx does not change in time as the state evolves! This suggests, but does not yet prove, that the state does not change shape[4]. It is therefore useful to calculate the time-dependent uncertainty in the momentum:

$$\begin{aligned}(\Delta p)^2(t) \;&=\; \langle\tilde{x}_0,t|\hat{p}^2|\tilde{x}_0,t\rangle - \big(\langle\tilde{x}_0,t|\hat{p}|\tilde{x}_0,t\rangle\big)^2 \\ &=\; \langle\tilde{x}_0|\hat{p}_H^2(t)|\tilde{x}_0\rangle - \big(\langle\tilde{x}_0|\hat{p}_H(t)|\tilde{x}_0\rangle\big)^2 \\ &=\; \langle\tilde{x}_0|\hat{p}_H^2(t)|\tilde{x}_0\rangle - m^2\omega^2\,x_0^2\sin^2\omega t\,,\end{aligned} \tag{4.39}$$

where we used (4.28). The rest of the computation (recommended!) gives

$$\langle\tilde{x}_0|\hat{p}_H^2(t)|\tilde{x}_0\rangle \;=\; \frac{1}{2}\,m\hbar\omega \;+\; m^2\omega^2\,x_0^2\sin^2\omega t\,, \tag{4.40}$$

so that we have

$$(\Delta p)^2(t) \;=\; \frac{m\hbar\omega}{2}\,. \tag{4.41}$$

This together with (4.38) gives

$$\Delta x(t)\Delta p(t) \;=\; \frac{\hbar}{2}\,, \qquad \text{on the state } |\tilde{x}_0,t\rangle\,. \tag{4.42}$$

The coherent state remains a minimum $\Delta x\Delta p$ packet for all times. Since only gaussians have such minimum uncertainty, the state remains a gaussian for all times! Since Δx is constant the gaussian does not change shape. Thus the name *coherent state*, the state does not spread out in time, it just moves "coherently" without changing its shape.

In the harmonic oscillator there is a quantum length scale d that can be constructed from $\hbar, m$, and ω. This length scale appears, for example, in the uncertainty Δx in (4.38). We thus define

$$d \;\equiv\; \sqrt{\frac{\hbar}{m\omega}}\,, \tag{4.43}$$

and note that

$$\Delta x(t) \;=\; \frac{d}{\sqrt{2}}\,. \tag{4.44}$$

The length d is typically very small for a macroscopic oscillator. A coherent state with a large x_0 –large compared to d– is classical in the sense that the position uncertainty $\sim d$, is much smaller than the typical excursion x_0. Similarly, the momentum uncertainty

$$\Delta p(t) \;=\; m\omega\,\frac{d}{\sqrt{2}}\,. \tag{4.45}$$

[4]By this we mean that the shape of $|\psi(x,t)|^2$ does not change: at different times $|\psi(x,t)|^2$ and $|\psi(x,t')|^2$ differ just by an overall displacement in x.

is much smaller than the typical momentum $m\omega x_0$, by just the same factor $\sim d/x_0$.

Problem. Prove that

$$\frac{\Delta p(t)}{\sqrt{\overline{\langle \hat{p}^2\rangle}(t)}} \;=\; \frac{\Delta x(t)}{\sqrt{\overline{\langle \hat{x}^2\rangle}(t)}} \;=\; \frac{1}{\sqrt{1+\frac{x_0^2}{d^2}}} \tag{4.46}$$

where the overlines on the expectation values denote time average.

4.4 Coherent states in the energy basis

We can get an interesting expression for the coherent state $|\tilde{x}_0\rangle$ by rewriting the momentum operator in terms of creation and annihilation operators. From (1.18) we have

$$\hat{p} \;=\; i\sqrt{\frac{m\omega\hbar}{2}}\,(\hat{a}^\dagger - \hat{a}) \;=\; i\,\frac{\hbar}{\sqrt{2}\,d}\,(\hat{a}^\dagger - \hat{a})\,. \tag{4.47}$$

The final form is also nice to see that units work. We now have that the coherent state (4.14) is given by

$$|\tilde{x}_0\rangle \;=\; \exp\Big(-\frac{i}{\hbar}\,\hat{p}\,x_0\Big)|0\rangle \;=\; \exp\Big(\frac{x_0}{\sqrt{2}d}\,(\hat{a}^\dagger - \hat{a})\Big)|0\rangle\,. \tag{4.48}$$

Since $\hat{a}|0\rangle = 0$ the above formula admits simplification: we should be able to get rid of all the $\hat{a}$'s! We could do this if we could split the exponential into two exponentials, one with the $\hat{a}^\dagger$'s to the *left* of another one with the $\hat{a}$'s. The exponential with the $\hat{a}$'s would stand near the vacuum and give no contribution, as we will see below. For this purpose we recall the commutator identity

$$e^{X+Y} \;=\; e^X\,e^Y\,e^{-\frac{1}{2}[X,Y]}\,, \qquad \text{if } [X,Y] \text{ commutes with } X \text{ and with } Y. \tag{4.49}$$

Think of the term we are interested in as it appears in (4.48), and identify X and Y as

$$\exp\Big(\frac{x_0}{\sqrt{2}d}\hat{a}^\dagger - \frac{x_0}{\sqrt{2}d}\hat{a}\Big) \qquad \to \qquad X = \frac{x_0}{\sqrt{2}d}\hat{a}^\dagger\,,\;\; Y = -\frac{x_0}{\sqrt{2}d}\hat{a} \tag{4.50}$$

Then

$$[X,Y] \;=\; -\frac{x_0^2}{2d^2}\,[\,\hat{a}^\dagger\,,\hat{a}] \;=\; \frac{x_0^2}{2d^2} \tag{4.51}$$

and we find

$$\exp\Big(\frac{x_0}{\sqrt{2}d}\hat{a}^\dagger - \frac{x_0}{\sqrt{2}d}\hat{a}\Big) \;=\; \exp\Big(\frac{x_0}{\sqrt{2}d}\hat{a}^\dagger\Big)\,\exp\Big(-\frac{x_0}{\sqrt{2}d}\hat{a}\Big)\,\exp\Big(-\frac{1}{4}\frac{x_0^2}{d^2}\Big) \tag{4.52}$$

Since the last exponential is just a number, and $\exp(\gamma\hat{a})|0\rangle = |0\rangle$, for any γ, we have

$$\exp\Big(\frac{x_0}{\sqrt{2}d}\hat{a}^\dagger - \frac{x_0}{\sqrt{2}d}\hat{a}\Big)|0\rangle \;=\; \exp\Big(-\frac{1}{4}\frac{x_0^2}{d^2}\Big)\,\exp\Big(\frac{x_0}{\sqrt{2}d}\hat{a}^\dagger\Big)|0\rangle\,. \tag{4.53}$$

As a result, our coherent state in (4.48) becomes

$$\boxed{|\tilde{x}_0\rangle \;=\; \exp\Big(-\frac{i}{\hbar}\,\hat{p}\,x_0\Big)|0\rangle \;=\; \exp\Big(-\frac{1}{4}\frac{x_0^2}{d^2}\Big)\exp\Big(\frac{x_0}{\sqrt{2}d}\hat{a}^\dagger\Big)|0\rangle\,.} \tag{4.54}$$

While this form is quite nice to produce an expansion in energy eigenstates, the unit normalization of the state is not manifest anymore. Expanding the exponential with creation operators we get

$$\begin{aligned} |\tilde{x}_0\rangle \;&=\; \sum_{n=0}^{\infty}\exp\Big(-\frac{1}{4}\frac{x_0^2}{d^2}\Big)\cdot\frac{1}{n!}\Big(\frac{x_0}{\sqrt{2}d}\Big)^n(\hat{a}^\dagger)^n|0\rangle \\ &=\; \sum_{n=0}^{\infty}\exp\Big(-\frac{1}{4}\frac{x_0^2}{d^2}\Big)\cdot\frac{1}{\sqrt{n!}}\Big(\frac{x_0}{\sqrt{2}d}\Big)^n|n\rangle \end{aligned} \tag{4.55}$$

We thus have the desired expansion:

$$\boxed{|\tilde{x}_0\rangle \;=\; \sum_{n=0}^{\infty}c_n|n\rangle\,,\quad \text{with}\;\; c_n \;=\; \exp\Big(-\frac{1}{4}\frac{x_0^2}{d^2}\Big)\cdot\frac{1}{\sqrt{n!}}\Big(\frac{x_0}{\sqrt{2}d}\Big)^n.} \tag{4.56}$$

Since the probability to find the energy E_n is equal to c_n^2, we note that

$$c_n^2 \;=\; \exp\Big(-\frac{x_0^2}{2d^2}\Big)\cdot\frac{1}{n!}\Big(\frac{x_0^2}{2d^2}\Big)^n \tag{4.57}$$

If we define the quantity $\lambda(x_0,d)$ as

$$\boxed{\lambda \;\equiv\; \frac{x_0^2}{2d^2}\,,} \tag{4.58}$$

we can then see that

$$\boxed{c_n^2 \;=\; \frac{\lambda^n}{n!}\,e^{-\lambda}\,.} \tag{4.59}$$

The probability to measure an energy $E_n = \hbar\omega(n+\frac{1}{2})$ in the coherent state is c_n^2, so the c_n^2's must define a probability distribution for $n\in\mathbb{Z}$, parameterized by λ. This is in fact the familiar *Poisson distribution*. It is straightforward to verify that

$$\sum_{n=0}^{\infty}c_n^2 \;=\; e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!} \;=\; e^{-\lambda}e^{\lambda} \;=\; 1\,, \tag{4.60}$$

as it should be. The physical interpretation of λ can be obtained by computing the expectation value of n:[5]

$$\langle n\rangle \;\equiv\; \sum_{n=0}^{\infty} n c_n^2 \;=\; e^{-\lambda}\sum_{n=0}^{\infty} n\,\frac{\lambda^n}{n!} \;=\; e^{-\lambda}\sum_{n=0}^{\infty}\lambda\frac{d}{d\lambda}\frac{\lambda^n}{n!} \;=\; e^{-\lambda}\,\lambda\frac{d}{d\lambda}e^{\lambda} \;=\lambda\,. \tag{4.61}$$

Therefore λ equals the expected value $\langle n\rangle$. Note that $\langle n\rangle$ is just the expected value of the number operator $\hat{N}$ on the coherent state. Indeed,

$$\langle\tilde{x}_0|\hat{N}|\tilde{x}_0\rangle \;=\; \sum_{n,m} c_m c_n\langle m|\hat{N}|n\rangle \;=\; \sum_{n,m} c_m c_n\, n\delta_{m,n} \;=\; \sum_n n\, c_n^2 \;=\; \langle n\rangle\,. \tag{4.62}$$

It is also easy to verify (do it!) that

$$\langle n^2\rangle \;\equiv\; \sum_{n=0}^{\infty} n^2\, c_n^2 \;=\; \lambda^2+\lambda\,. \tag{4.63}$$

It then follows that

$$(\Delta n)^2 \;=\; \langle n^2\rangle - \langle n\rangle^2 \;=\; \lambda \qquad\to\qquad \Delta n \;=\; \sqrt{\lambda}\,. \tag{4.64}$$

In terms of energy we have $E=\hbar\omega(n+\frac{1}{2})$ so that

$$\langle E\rangle \;=\; \hbar\omega\Big(\langle n\rangle+\frac{1}{2}\Big) \;=\; \hbar\omega\Big(\lambda+\frac{1}{2}\Big)\,. \tag{4.65}$$

Similarly,

$$\langle E^2\rangle \;=\; \hbar^2\omega^2\Big\langle\Big(n+\frac{1}{2}\Big)^2\Big\rangle \;=\; \hbar^2\omega^2\Big\langle n^2+n+\frac{1}{4}\Big\rangle \;=\; \hbar^2\omega^2\Big(\lambda^2+\lambda+\lambda+\frac{1}{4}\Big)\,, \tag{4.66}$$

so that

$$\langle E^2\rangle \;=\; \hbar^2\omega^2\Big(\lambda^2+2\lambda+\frac{1}{4}\Big) \tag{4.67}$$

The energy uncertainty is thus obtained as

$$(\Delta E)^2 \;=\; \langle E^2\rangle - \langle E\rangle^2 \;=\; \hbar^2\omega^2\Big(\lambda^2+2\lambda+\frac{1}{4}-\Big(\lambda+\frac{1}{2}\Big)^2\Big) \;=\; \hbar^2\omega^2\,\lambda\,, \tag{4.68}$$

so that

$$\Delta E \;=\; \hbar\omega\sqrt{\lambda} \;=\; \hbar\omega\,\frac{x_0}{\sqrt{2}d}\,. \tag{4.69}$$

Note now the fundamental inequality, holding for $x_0/d\gg 1$,

$$\hbar\omega \quad\ll\quad \Delta E=\hbar\omega\,\frac{x_0}{\sqrt{2}d} \quad\ll\quad \langle\, E\,\rangle = \hbar\omega\Big(\frac{x_0}{\sqrt{2}d}\Big)^2+\frac{1}{2}\hbar\omega\,. \tag{4.70}$$

[5]Here we are thinking of n as a random variable of the probability distribution. In the quantum viewpoint $\langle n\rangle$ is simply the expectation value of the number operator.

We see that the uncertainty ΔE is big enough to contain about $\frac{x_0}{\sqrt{2}d}$ levels of the harmonic oscillator –a lot of levels. But even then, ΔE is about a factor $\frac{x_0}{\sqrt{2}d}$ smaller than the expected value $E\rangle$ of the energy. So, alternatively,

$$\frac{\Delta E}{\hbar\omega} \;\simeq\; \frac{x_0}{\sqrt{2}d} \;\simeq\; \frac{\langle E\rangle}{\Delta E}\,. \tag{4.71}$$

This is part of the semi-classical nature of coherent states.

Example of Poisson distribution. Consider a sample of radioactive material with $N_0 \gg 1$ atoms at $t = 0$. Assume that the half-lifetime of the material is τ_0, which means that the number $N(t)$ of atoms that have not yet decayed after time $t > 0$ is given by

$$N(t) = N_0 \exp(-t/\tau_0) \quad \to \quad \frac{dN}{dt}(t=0) = \frac{N_0}{\tau_0}\,.$$

It follows that in the time interval $t \in [0, \Delta t]$, with $\Delta t \ll \tau_0$ we expect a number of decays

$$\frac{N_0 \Delta t}{\tau_0} \equiv \lambda\,.$$

One can then show that the probability p_n to observe n decays during that same time interval Δt is (approximately) given by the Poisson distribution: $p_n = \frac{\lambda^n}{n!} e^{-\lambda}$.

4.5 General coherent states and time evolution

We wrote earlier coherent states using creation and annihilation operators:

$$|\tilde{x}_0\rangle \;=\; \exp\Big(-\frac{i}{\hbar}\,\hat{p}\,x_0\Big)|0\rangle \;=\; \exp\Big(\frac{x_0}{\sqrt{2}d}\,(\hat{a}^\dagger - \hat{a})\Big)|0\rangle\,. \tag{4.72}$$

Such coherent states can be written as

$$|\alpha\rangle \equiv e^{\alpha\,(\hat{a}^\dagger - \hat{a})}|0\rangle\,, \quad \text{with} \;\; \alpha \;=\; \frac{x_0}{\sqrt{2}d}\,. \tag{4.73}$$

This notation is not free of ambiguity: the label α in the coherent state above is now the coefficient of the factor $\hat{a}^\dagger - \hat{a}$ in the exponential. An obvious generalization is to take α to be a complex number: $\alpha \in \mathbb{C}$. This must be done with a little care, since the key property of the operator in the exponential (4.73) is that it is antihermitian (thus the exponential is unitary, as desired). We thus define

$$\boxed{|\alpha\rangle \;\equiv\; D(\alpha)|0\rangle \;\equiv\; \exp\big(\alpha\hat{a}^\dagger - \alpha^*\hat{a}\big)|0\rangle\,, \quad \text{with} \;\; \alpha \in \mathbb{C}\,.} \tag{4.74}$$

In this definition we introduced the unitary 'displacement' operator

$$D(\alpha) \equiv \exp\Big(\alpha\hat{a}^\dagger - \alpha^*\hat{a}\Big)\,. \tag{4.75}$$

Since $D(\alpha)$ is unitary it is clear that $\langle\alpha|\alpha\rangle = 1$.

The action of the annihilation operator on the states $|\alpha\rangle$ is quite interesting,

$$\begin{aligned}\hat{a}|\alpha\rangle &= \hat{a}\,e^{\alpha\hat{a}^\dagger-\alpha^*\hat{a}}|0\rangle \;=\; \big[\,\hat{a}\,,\,e^{\alpha\hat{a}^\dagger-\alpha^*\hat{a}}\big]|0\rangle \\ &= \big[\,\hat{a}\,,\,\alpha\hat{a}^\dagger-\alpha^*\hat{a}\,\big]e^{\alpha\hat{a}^\dagger-\alpha^*\hat{a}}|0\rangle \;=\; \alpha e^{\alpha\hat{a}^\dagger-\alpha^*\hat{a}}|0\rangle\,,\end{aligned} \tag{4.76}$$

so that we conclude that

$$\boxed{\;\hat{a}\,|\alpha\rangle \;=\; \alpha\,|\alpha\rangle\,.\;} \tag{4.77}$$

This result is kind of shocking: we have found eigenstates of the *non-hermitian* operator $\hat{a}$. Because $\hat{a}$ is not hermitian, our theorems about eigenstates and eigenvectors of hermitian operators do not apply. Thus, for example, the eigenvalues need not be real (they are not, in general $\alpha \in \mathbb{C}$), eigenvectors of different eigenvalue need not be orthogonal (they are not!) and the set of eigenvectors need not form a complete basis (coherent states actually give an overcomplete basis!).

Ordering the exponential in the state $|\alpha\rangle$ in (4.74) we find

$$|\alpha\rangle \;=\; e^{-\frac{1}{2}|\alpha|^2}e^{\alpha\hat{a}^\dagger}|0\rangle\,. \tag{4.78}$$

Exercise. Show that

$$\langle\beta|\alpha\rangle \;=\; \exp\Big(-\frac{1}{2}(|\alpha|^2+|\beta|^2)+\beta^*\alpha\Big)\,. \tag{4.79}$$

Hint: You may find it helpful to evaluate $e^{\beta^*\hat{a}+\alpha\hat{a}^\dagger}$ in two different ways using (4.49).

To find the physical interpretation of the complex number α we first note that when real, as in (4.73), α encodes the initial position x_0 of the coherent state (more precisely, it encodes the expectation value of $\hat{x}$ in the state at $t = 0$). For complex α, its real part is still related to the initial position:

$$\langle\alpha|\hat{x}|\alpha\rangle \;=\; \frac{d}{\sqrt{2}}\,\langle\alpha|(\hat{a}+\hat{a}^\dagger)|\alpha\rangle \;=\; \frac{d}{\sqrt{2}}(\alpha+\alpha^*) \;=\; d\sqrt{2}\,\mathrm{Re}(\alpha)\,, \tag{4.80}$$

where we used (1.18) and (4.77) both on bras and on kets. We have thus learned that

$$\mathrm{Re}(\alpha) \;=\; \frac{\langle\hat{x}\rangle}{\sqrt{2}\,d}\,. \tag{4.81}$$

It is natural to conjecture that the imaginary part of α is related to the momentum expectation value on the initial state. So we explore

$$\langle\alpha|\hat{p}|\alpha\rangle \;=\; \frac{i\hbar}{\sqrt{2}d}\,\langle\alpha|(\hat{a}^\dagger-\hat{a})|\alpha\rangle \;=\; -\frac{i\hbar}{\sqrt{2}d}(\alpha-\alpha^*) \;=\; -\frac{i\hbar}{\sqrt{2}d}(2i\mathrm{Im}(\alpha)) \;=\; \frac{\hbar\sqrt{2}}{d}\,\mathrm{Im}(\alpha)\,, \tag{4.82}$$

and learn that

$$\mathrm{Im}(\alpha) \;=\; \frac{\langle\hat{p}\rangle\, d}{\sqrt{2}\,\hbar}\,. \tag{4.83}$$

The identification of α in terms of expectation values of $\hat{x}$ and $\hat{p}$ is now complete:

$$\boxed{\alpha \;=\; \frac{\langle\hat{x}\rangle}{\sqrt{2}\,d} \;+\; i\,\frac{\langle\hat{p}\rangle\, d}{\sqrt{2}\,\hbar}\,.} \tag{4.84}$$

A calculation in the problem set shows that

$$\alpha\,\hat{a}^\dagger - \alpha^*\hat{a} \;=\; -\frac{i}{\hbar}\big(\hat{p}\,\langle x\rangle - \langle\hat{p}\rangle\,\hat{x}\big)\,, \tag{4.85}$$

affording yet another rewriting of the general coherent state (4.74), valid when α is defined as in (4.84):

$$\boxed{|\alpha\rangle \;=\; \exp\Big(-\frac{i\hat{p}\,\langle x\rangle}{\hbar} + \frac{i\langle\hat{p}\rangle\,\hat{x}}{\hbar}\Big)|0\rangle\,.} \tag{4.86}$$

In order to find the time evolution of the coherent state we can use a trick from the Heisenberg picture. We have using (4.74)

$$|\alpha,t\rangle \;\equiv\; e^{-i\frac{Ht}{\hbar}}|\alpha\rangle \;=\; e^{-i\frac{Ht}{\hbar}}e^{\alpha\hat{a}^\dagger-\alpha^*\hat{a}}|0\rangle \;=\; \Big(e^{-i\frac{Ht}{\hbar}}e^{\alpha\hat{a}^\dagger-\alpha^*\hat{a}}e^{i\frac{Ht}{\hbar}}\Big)e^{-i\frac{Ht}{\hbar}}|0\rangle \tag{4.87}$$

For a time independent Hamiltonian (as that of the SHO) and a Schrödinger operator $\mathcal{O}$, we have

$$\mathcal{O}_H(t) \;=\; e^{iHt/\hbar}\,\mathcal{O}\,e^{-iHt/\hbar} \tag{4.88}$$

and therefore with the opposite signs for the exponentials we get

$$e^{-iHt/\hbar}\,\mathcal{O}\,e^{iHt/\hbar} \;=\; \mathcal{O}_H(-t)\,. \tag{4.89}$$

Such a relation is also valid for any function of an operator:

$$e^{-iHt/\hbar}\,F(\mathcal{O})\,e^{iHt/\hbar} \;=\; F(\mathcal{O}_H(-t))\,. \tag{4.90}$$

as you can convince yourself is the case whenever $F(x)$ has a good Taylor expansion in powers of x. It then follows that back in (4.87) we have

$$|\alpha,t\rangle \;=\; \exp\Big(\alpha\hat{a}^\dagger(-t) - \alpha^*\hat{a}(-t)\Big)e^{-i\omega t/2}|0\rangle\,. \tag{4.91}$$

Recalling ((3.53)) that $\hat{a}(t) \;=\; e^{-i\omega t}\,\hat{a}$, and thus $\hat{a}^\dagger(t) = e^{i\omega t}\,\hat{a}^\dagger$, we find

$$|\alpha,t\rangle \;=\; e^{-i\omega t/2}\exp\Big(\alpha e^{-i\omega t}\hat{a}^\dagger - \alpha^* e^{i\omega t}\hat{a}\Big)|0\rangle\,. \tag{4.92}$$

Looking at the exponential we see that it is in fact the displacement operator with $\alpha \to \alpha e^{-i\omega t}$. As a result we have shown that

$$\boxed{|\alpha, t\rangle \;=\; e^{-i\omega t/2}|e^{-i\omega t}\alpha\rangle\,.} \tag{4.93}$$

This is how a coherent state $|\alpha\rangle$ evolves in time: up to an irrelevant phase, the state remains a coherent state with a time-varying parameter $e^{-i\omega t}\alpha$. In the complex α plane the state is represented by a vector that rotates in the clockwise direction with angular velocity ω. The α plane can be viewed as having a real axis that gives $\langle x\rangle$ (up to a proportionality constant) and an imaginary axis that gives $\langle \hat{p}\rangle$ (up to a proportionality constant). It is a phase space and the evolution of any state is represented by a circle. This is illustrated in Figure 3.

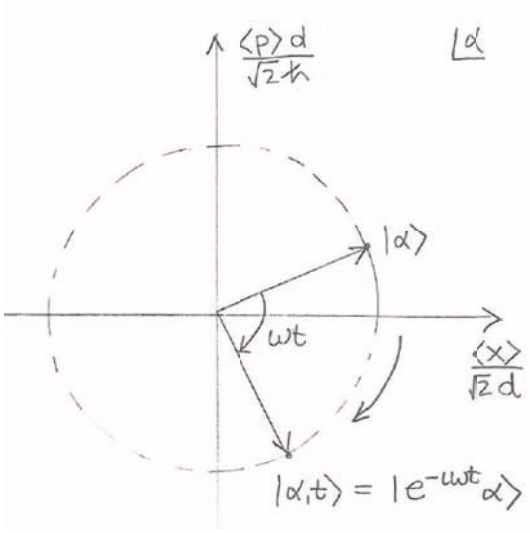


Figure 3: Time evolution of the coherent state $|\alpha\rangle$. The real and imaginary parts of α determine the expectation values $\langle x\rangle$ and $\langle p\rangle$ respectively. As time goes by the α parameter of the coherent state rotates clockwise with angular velocity ω.

An alternative, conventional, calculation of the time evolution begins by expanding the exponential in (4.78) to find:

$$|\alpha\rangle \;=\; e^{-\frac{1}{2}|\alpha|^2}\sum_{n=0}^{\infty}\frac{1}{\sqrt{n!}}\,\alpha^n|n\rangle\,. \tag{4.94}$$

The time-evolved state is then given by the action of $\exp(-iHt/\hbar)$:

$$\begin{aligned} |\alpha,t\rangle \;\equiv\; e^{-i\frac{Ht}{\hbar}}|\alpha\rangle \;&=\; e^{-\frac{1}{2}|\alpha|^2}\sum_{n=0}^{\infty}\frac{1}{\sqrt{n!}}\,\alpha^n e^{-i\hbar\omega(n+\frac{1}{2})t/\hbar}|n\rangle \\ &=\; e^{-\frac{1}{2}|\alpha|^2}\sum_{n=0}^{\infty}\frac{1}{\sqrt{n!}}\,\alpha^n e^{-i\omega tn}e^{-i\omega t/2}|n\rangle \\ &=\; e^{-i\omega t/2}\;\underbrace{e^{-\frac{1}{2}|\alpha|^2}\sum_{n=0}^{\infty}\frac{1}{\sqrt{n!}}\,(e^{-i\omega t}\alpha)^n|n\rangle}\,. \end{aligned} \tag{4.95}$$

Using (4.94) and noting that $|e^{-i\omega t}\alpha|^2 = |\alpha|^2$, we identify the terms under the brace as a coherent state $|\alpha e^{-i\omega t}\rangle$. This gives the earlier result (4.93).

In the coherent state $|\alpha\rangle$ the expectation value of $\hat{N}$ is easily calculated

$$\langle\hat{N}\rangle_\alpha \;=\; \langle\alpha|\hat{N}|\alpha\rangle \;=\; \langle\alpha|\hat{a}^\dagger\hat{a}|\alpha\rangle \;=\; \langle\alpha|\alpha^*\alpha|\alpha\rangle = |\alpha|^2\,. \tag{4.96}$$

To find the uncertainty ΔN we also compute

$$\begin{aligned} \langle\hat{N}^2\rangle_\alpha \;&=\; \langle\alpha|\,\hat{a}^\dagger\hat{a}\;\hat{a}^\dagger\hat{a}|\alpha\rangle \;=\; |\alpha|^2\langle\alpha|\,\hat{a}\;\hat{a}^\dagger|\alpha\rangle \\ &=\; |\alpha|^2\langle\alpha|\,[\hat{a}\,,\hat{a}^\dagger]+\hat{a}^\dagger\hat{a}|\alpha\rangle \;=\; |\alpha|^2\big(1+|\alpha|^2\big)\,. \end{aligned} \tag{4.97}$$

From these results we get

$$(\Delta N)^2 \;=\; \langle\hat{N}^2\rangle_\alpha - \langle\hat{N}\rangle_\alpha^2 \;=\; |\alpha|^2+|\alpha|^4-|\alpha|^4 = |\alpha|^2 \tag{4.98}$$

so that

$$\boxed{\Delta N \;=\; |\alpha|\,.} \tag{4.99}$$

in Figure 3 the magnitude of the rotating phasor is ΔN and the square of the magnitude is the expectation value $\langle\hat{N}\rangle_\alpha$.

We will soon discuss electromagnetic fields and waves as coherent states of photons. For such waves a number/phase uncertainty exists. A rough argument goes as follows. For a wave with N photons with frequency ω, the energy is $E = N\hbar\omega$ and the phase ϕ of the wave goes like $\phi = \omega t$. It follows that $\Delta E \sim \Delta N\hbar\omega$ and $\Delta\phi = \omega\Delta t$ (with the admittedly ambiguous meaning of Δt). Therefore

$$\Delta E\Delta t \geq \frac{\hbar}{2} \quad\to\quad \Delta N\hbar\omega\,\frac{\Delta\phi}{\omega} \geq \frac{\hbar}{2} \quad\to\quad \Delta N\Delta\phi \geq \frac{1}{2}\,. \tag{4.100}$$

A better intuition for this result follows from our coherent state $|\alpha\rangle$ for which we know that $\Delta N = |\alpha|$. The position and momentum uncertainties are the same as for the ground state:

$$\Delta x \;=\; \frac{d}{\sqrt{2}}\,,\quad \Delta p \;=\; \frac{\hbar}{d\sqrt{2}} \tag{4.101}$$

When we measure x on the state $|\alpha\rangle$ we expect to get a good fraction of values in a range Δx about the expected value $\langle x\rangle$ of x. This is, of course, just a rough estimate.

$$\text{Representative range for measured } x \;=\; [\langle x\rangle - \tfrac{1}{2}\Delta x\,,\langle x\rangle + \tfrac{1}{2}\Delta x] \tag{4.102}$$

Dividing by $\sqrt{2}d$ we have

$$\text{Representative range for measured } \frac{x}{\sqrt{2}d} \;=\; \Big[\frac{\langle x\rangle}{\sqrt{2}d} - \tfrac{1}{2}\,,\frac{\langle x\rangle}{\sqrt{2}d} + \tfrac{1}{2}\Big] \tag{4.103}$$

It follows that the position measurements, indicated on the horizontal axis of Figure 3, spread over a representative range of width one. Similarly, for momentum we have

$$\text{Representative range for measured } p \;=\; [\langle p\rangle - \tfrac{1}{2}\Delta p\,,\langle p\rangle + \tfrac{1}{2}\Delta p] \tag{4.104}$$

Multiplying by $d/(\sqrt{2}\hbar)$, we have

$$\text{Representative range for measured } \frac{p\,d}{\sqrt{2}\hbar} \;=\; \Big[\frac{\langle p\rangle\, d}{\sqrt{2}\hbar} - \tfrac{1}{2}\,,\frac{\langle p\rangle\, d}{\sqrt{2}\hbar} + \tfrac{1}{2}\Big] \tag{4.105}$$

It follows that the momentum measurements, indicated on the vertical axis of Figure 3, spread

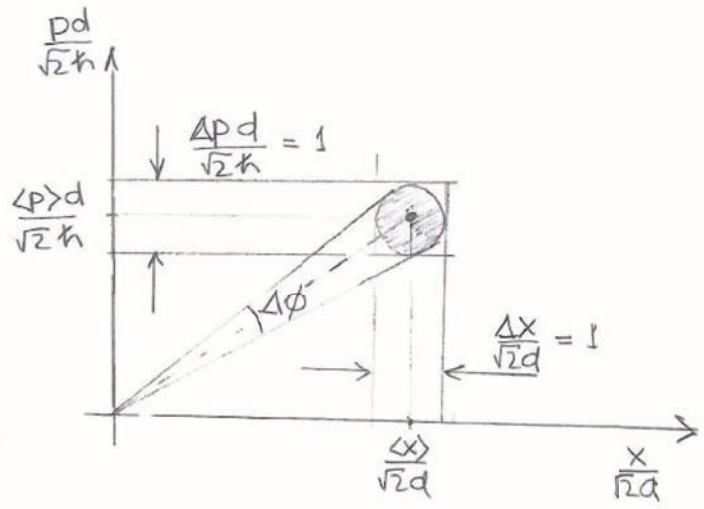


Figure 4: When doing measurements on $|\alpha\rangle$ the uncertainties on the value of α can be represented by a blob of unit diameter centered at α. The projections of this blob on the axes are, up to scale, the uncertainties Δx and Δp.

over a representative range of width one. We can thus reconsider the plot, this time indicating the ranges of values expected on the horizontal and vertical axes. Those ranges can be viewed as some kind of uncertainty in the value of α that we could find by measurements on the state

$|\alpha\rangle$. We draw a blob of unit diameter centered at α whose projections down along the axes reproduce the uncertainty ranges. This is shown in Figure 4. In the spirit of the discussion on time dependence, this blob must be imagined rotating with angular frequency ω. In such picture we have a phase ambiguity $\Delta\phi$, represented in the picture as the angle subtended by the uncertainty blob. Since the blob has diameter one and is centered at α, which is a distance $|\alpha|$ from the origin, we have

$$\Delta\phi \;\simeq\; \frac{1}{|\alpha|} \tag{4.106}$$

Recalling that $\Delta N = |\alpha|$ we finally obtain that for our coherent state

$$\Delta N \Delta\phi \;\simeq\; 1\,. \tag{4.107}$$

This is a familiar relation for coherent states of light. It then relates the uncertainty in the number of photons to the uncertainty in the phase of the wave.

5 Squeezed states

Squeezed states of the harmonic oscillator are states that are obtained by acting on the ground state with an exponential that includes terms quadratic in creation operators. They are the most general states for which $\Delta x \Delta p = \hbar/2$, thus achieving saturation of the uncertainty bound.

5.1 Squeezed vacuum states

One useful way to motivate the introduction of squeezed states is to consider the ground state of a harmonic oscillator Hamiltonian with mass and frequency parameters m_1 and ω_1, respectively:

$$H_1 \;=\; \frac{p^2}{2m_1} + \frac{1}{2}\, m_1 \omega_1^2 x^2\,. \tag{5.1}$$

Such ground state has uncertaintites Δx and Δp that follow from (1.58) :

$$\begin{aligned} \Delta x \;&=\; \sqrt{\frac{\hbar}{2m_1\omega_1}}\,, \\ \Delta p \;&=\; \sqrt{\frac{\hbar m_1 \omega_1}{2}}\,. \end{aligned} \tag{5.2}$$

Note that the product of uncertainties saturates the lower bound:

$$\Delta x \Delta p \;=\; \frac{\hbar}{2}\,. \tag{5.3}$$

Now we consider the following situation: suppose at time $t = 0^-$ the wavefunction is indeed that of the ground state of the oscillator. At $t = 0$, however, the oscillator parameters *change*

instantaneously from (m_1, ω_1) to some (m_2, ω_2) that define a second, different Hamiltonian:

$$H_2 \;=\; \frac{p^2}{2m_2} + \frac{1}{2}\, m_2 \omega_2^2 x^2 \,. \tag{5.4}$$

During this change the wavefunction is assumed not to change, so at $t = 0^+$ the wavefunction is still the same – the ground state of H_1. Since the Hamiltonian changed, however, the state of the system is no longer an energy eigenstate: the gaussian wavefunction that is a ground state for H_1 is *not* a ground state of H_2. In fact it is not an energy eigenstate of H_2 and its time evolution will be nontrivial. We will see that the ground state of H_1 is indeed a squeezed state of H_2.

Since the wavefunction does not change, at $t = 0^+$ the uncertainties in (5.2) do not change, and we can rewrite

$$\begin{aligned} \Delta x \;&=\; \sqrt{\frac{m_2\omega_2}{m_1\omega_1}}\,\sqrt{\frac{\hbar}{2m_2\omega_2}} \;=\; e^{-\gamma}\,\sqrt{\frac{\hbar}{2m_2\omega_2}}\;, \\ \Delta p \;&=\; \sqrt{\frac{m_1\omega_1}{m_2\omega_2}}\,\sqrt{\frac{\hbar m_2\omega_2}{2}} \;=\; e^{\gamma}\,\sqrt{\frac{\hbar m_2\omega_2}{2}}\;, \end{aligned} \tag{5.5}$$

where we defined the real constant γ by

$$e^{\gamma} \;\equiv\; \sqrt{\frac{m_1\omega_1}{m_2\omega_2}}\,. \tag{5.6}$$

We learn from (5.5) that at $t = 0^+$ the uncertainties, from the viewpoint of the second Hamiltonian, have been squeezed from the values they would take on the H_2 ground state: if $\gamma > 0$, the position uncertainty is reduced and the momentum uncertainty increased. Of course, the product still saturates the bound.

To work out the details of the state at $t = 0^+$ we need to relate the creation and annihilation operators of the two Hamiltonians. We note that the operators x and p have not been changed, we are not speaking about two oscillating particles, but rather a single one, with coordinate measured by the operator x and momentum measured by the operator p. We thus use the expression for x and p in terms of $a, a^\dagger$ (equation (1.18) to write

$$\begin{aligned} x \;&=\; \sqrt{\frac{\hbar}{2m_1\omega_1}}\;(\hat{a}_1 + \hat{a}_1^\dagger) \;=\; \sqrt{\frac{\hbar}{2m_2\omega_2}}\;(\hat{a}_2 + \hat{a}_2^\dagger) \\ p \;&=\; -i\sqrt{\frac{m_1\omega_1\hbar}{2}}(\hat{a}_1 - \hat{a}_1^\dagger) \;=\; -i\sqrt{\frac{m_2\omega_2\hbar}{2}}(\hat{a}_2 - \hat{a}_2^\dagger) \end{aligned} \tag{5.7}$$

Using the definition of e^γ we then have

$$\begin{aligned} \hat{a}_1 + \hat{a}_1^\dagger \;&=\; e^{\gamma}\;(\hat{a}_2 + \hat{a}_2^\dagger)\,, \\ \hat{a}_1 - \hat{a}_1^\dagger \;&=\; e^{-\gamma}\,(\hat{a}_2 - \hat{a}_2^\dagger)\,. \end{aligned} \tag{5.8}$$

Solving these equations for $(\hat{a}_1, \hat{a}_1^\dagger)$ in terms of $(\hat{a}_2, \hat{a}_2^\dagger)$ we find

$$\begin{aligned} \hat{a}_1 &= \hat{a}_2 \cosh\gamma + \hat{a}_2^\dagger \sinh\gamma\,, \\ \hat{a}_1^\dagger &= \hat{a}_2 \sinh\gamma + \hat{a}_2^\dagger \cosh\gamma\,. \end{aligned} \tag{5.9}$$

Note that the second equation is simply the adjoint of the first equation. The above relations are called *Bogoliubov* transformations. Notice that they preserve the commutation algebra. You can check that $[\hat{a}_1, \hat{a}_1^\dagger] = 1$ using (5.9) and the commutation relation of the $\hat{a}_2$ and $\hat{a}_2^\dagger$ operators. We can also obtain the second set of operators in terms of the first set by changing γ into $-\gamma$, as implied by the relations (5.8) :

$$\begin{aligned} \hat{a}_2 &= \hat{a}_1 \cosh\gamma - \hat{a}_1^\dagger \sinh\gamma\,, \\ \hat{a}_2^\dagger &= -\hat{a}_1 \sinh\gamma + \hat{a}_1^\dagger \cosh\gamma\,. \end{aligned} \tag{5.10}$$

We can now examine explicitly the question of the ground state. The initial state is the ground state of H_1 denoted as $|0\rangle_{(1)}$. Its defining property is that it is killed by a_1:

$$\hat{a}_1 |0\rangle_{(1)} = 0\,. \tag{5.11}$$

Using equation (5.9) we have

$$\big(\hat{a}_2 \cosh\gamma + \hat{a}_2^\dagger \sinh\gamma\big)|0\rangle_{(1)} = 0\,. \tag{5.12}$$

Solving this equation means finding some expression for $|0\rangle_{(1)}$ in terms of some combination of $a_2^\dagger$ operators acting on $|0\rangle_{(2)}$. We should be able to write the original ground-state wavefunction in terms of eigenfunctions of the second Hamiltonian, or equivalently, write the original state as a superposition of energy eigenstates of the second Hamiltonian. Since the original wavefunction is even in x, only states with even number of creation operators should enter in such an expansion. We thus expect a solution of the form

$$|0\rangle_{(1)} = c_0|0\rangle_{(2)} + c_2\, \hat{a}_2^\dagger\hat{a}_2^\dagger|0\rangle_{(2)} + c_4\, \hat{a}_2^\dagger\hat{a}_2^\dagger\hat{a}_2^\dagger\hat{a}_2^\dagger|0\rangle_{(2)} + \dots\,, \tag{5.13}$$

where the c_n's are coefficients to be determined. While we could proceed recursively, it is in fact possible to write an ansatz for the state and solve the problem directly.

We write an educated guess that uses the exponential of an expression quadratic in $\hat{a}_2^\dagger$:

$$|0\rangle_{(1)} = \mathcal{N}(\gamma)\, \exp\Big(-\frac{1}{2}\, f(\gamma)\, \hat{a}_2^\dagger\hat{a}_2^\dagger\Big)\, |0\rangle_{(2)}\,. \tag{5.14}$$

In here the functions $f(\gamma)$ and $\mathcal{N}(\gamma)$ are to be determined. Equation (5.12) gives

$$\big(\hat{a}_2 \cosh\gamma + \hat{a}_2^\dagger \sinh\gamma\big) \exp\Big(-\frac{1}{2}\, f(\gamma)\, \hat{a}_2^\dagger\hat{a}_2^\dagger\Big)\, |0\rangle_{(2)} = 0\,. \tag{5.15}$$

The action of $\hat{a}_2$ can be replaced by a commutator since it kills the vacuum $|0\rangle_{(2)}$:

$$\cosh\gamma\Big[\hat{a}_2\,,\,\exp\Big(-\frac{1}{2}\,f(\gamma)\,\hat{a}_2^\dagger\hat{a}_2^\dagger\Big)\Big]\,|0\rangle_{(2)}+\hat{a}_2^\dagger\sinh\gamma\exp\Big(-\frac{1}{2}\,f(\gamma)\,\hat{a}_2^\dagger\hat{a}_2^\dagger\Big)\,|0\rangle_{(2)}\;=\;0\,. \tag{5.16}$$

We can now apply the familiar $[A,e^B]=[A,B]e^B$ (if $[[A,B],B]=0$) to the evaluation of the commutator

$$\Big(\cosh\gamma\Big[\hat{a}_2\,,\,-\frac{1}{2}\,f(\gamma)\,\hat{a}_2^\dagger\hat{a}_2^\dagger\Big]\;+\hat{a}_2^\dagger\sinh\gamma\Big)\exp\Big(-\frac{1}{2}\,f(\gamma)\,\hat{a}_2^\dagger\hat{a}_2^\dagger\Big)\,|0\rangle_{(2)}\;=\;0\,. \tag{5.17}$$

Evaluating the remaining commutator gives

$$\Big(-\cosh\gamma\;f(\gamma)\;\;+\sinh\gamma\Big)\,\hat{a}_2^\dagger\exp\Big(-\frac{1}{2}\,f(\gamma)\,\hat{a}_2^\dagger\hat{a}_2^\dagger\Big)\,|0\rangle_{(2)}\;=\;0\,. \tag{5.18}$$

Since no annihilation operators remain, the equality requires that the pre factor in parenthesis be zero. This determines the function $f(\gamma)$:

$$f(\gamma)=\tanh\gamma\,, \tag{5.19}$$

and we therefore have

$$|0\rangle_{(1)}\;=\;\mathcal{N}(\gamma)\,\exp\Big(-\frac{1}{2}\,\tanh\gamma\,\hat{a}_2^\dagger\hat{a}_2^\dagger\Big)\,|0\rangle_{(2)}\,. \tag{5.20}$$

The normalization $\mathcal{N}$ is not determined by the above calculation. It is could be determined, for example, by the demand that the state on the right-hand side above have unit normalization, just like $|0\rangle_{(1)}$ does. This is not a simple calculation. A simpler way uses the overlap of the two sides of the above equation with ${}_{(2)}\langle 0|$. We find

$${}_{(2)}\langle 0|0\rangle_{(1)}\;=\;\mathcal{N}(\gamma)\,, \tag{5.21}$$

because on the right hand side we can expand the exponential and all oscillators give zero on account of ${}_{(2)}\langle 0|\hat{a}_2^\dagger=0$. Introducing a complete set of position states we get:

$$\mathcal{N}(\gamma)\;=\;\int_{-\infty}^{\infty}dx\,{}_{(2)}\langle 0|x\rangle\langle x|0\rangle_{(1)}\;=\;\int_{-\infty}^{\infty}dx\;(\psi_0^{(2)}(x))^*\,\psi_0^{(1)}(x)\,. \tag{5.22}$$

Using the expression (1.39) for the ground state wavefunctions

$$\begin{aligned}\mathcal{N}(\gamma)\;&=\Big(\frac{m_1\omega_1}{\pi\hbar}\Big)^{1/4}\Big(\frac{m_2\omega_2}{\pi\hbar}\Big)^{1/4}\int_{-\infty}^{\infty}dx\exp\Big(-\Big[\frac{m_1\omega_1+m_2\omega_2}{2\hbar}\Big]x^2\Big)\,,\\
&=\Big[\frac{\sqrt{m_1\omega_1m_2\omega_2}}{\pi\hbar}\Big]^{1/2}\frac{\sqrt{2\pi\hbar}}{\sqrt{m_1\omega_1+m_2\omega_2}}\;=\;\Big(\frac{1}{2}\,\frac{m_1\omega_1+m_2\omega_2}{\sqrt{m_1\omega_1m_2\omega_2}}\Big)^{-1/2}\\
&=\Big(\frac{1}{2}\,\Big[\sqrt{\frac{m_1\omega_1}{m_2\omega_2}}+\sqrt{\frac{m_2\omega_2}{m_1\omega_1}}\,\Big]\Big)^{-1/2}\;=\;\Big(\frac{1}{2}\,\big[\,e^{\gamma}+e^{-\gamma}\big]\Big)^{-1/2}\,,\end{aligned} \tag{5.23}$$

so that we finally have

$$\mathcal{N}(\gamma) \;=\; \frac{1}{\sqrt{\cosh\gamma}}\,. \tag{5.24}$$

All in all

$$\boxed{|0\rangle_{(1)} \;=\; \frac{1}{\sqrt{\cosh\gamma}}\,\exp\Big(-\frac{1}{2}\,\tanh\gamma\,\hat{a}_2^\dagger\hat{a}_2^\dagger\Big)\,|0\rangle_{(2)}\,.} \tag{5.25}$$

The state on the above right-hand side takes the form of an exponential of something quadratic in oscillators. It is a squeezed vacuum state of the second Hamiltonian.

Inspired by the discussion above we introduce squeezed states for an arbitrary harmonic oscillator Hamiltonian H with vacuum $|0\rangle$, parameters (m,ω) and operators $(a,a^\dagger)$. A a normalized squeezed vacuum state, denoted as $|0_\gamma\rangle$, thus takes the form

$$|0_\gamma\rangle \;\equiv\; \frac{1}{\sqrt{\cosh\gamma}}\exp\Big(-\frac{1}{2}\,\tanh\gamma\,\hat{a}^\dagger\hat{a}^\dagger\Big)\,|0\rangle\,. \tag{5.26}$$

For this state we have

$$\big(\hat{a}\,\cosh\gamma+\hat{a}^\dagger\,\sinh\gamma\big)|0_\gamma\rangle \;=\; 0\,. \tag{5.27}$$

For this squeezed vacuum state the x uncertainty follows directly from (5.5):

$$\Delta x \;=\; e^{-\gamma}\sqrt{\frac{\hbar}{2m\omega}}\,. \tag{5.28}$$

The above squeezed vacuum state can in fact be expressed in terms of a unitary operator $S(\gamma)$ acting on the vacuum. We claim that $|0_\gamma\rangle$ defined above is actually

$$|0_\gamma\rangle \;=\; S(\gamma)|0\rangle\,,\quad \text{with}\quad S(\gamma) \;=\; \exp\Big(-\frac{\gamma}{2}(\hat{a}^\dagger\hat{a}^\dagger-\hat{a}\hat{a})\Big)\,. \tag{5.29}$$

This claim implies that the following nontrivial identity holds:

$$\exp\Big(-\frac{\gamma}{2}(\hat{a}^\dagger\hat{a}^\dagger-\hat{a}\hat{a})\Big)|0\rangle \;=\; \frac{1}{\sqrt{\cosh\gamma}}\exp\Big(-\frac{1}{2}\,\tanh\gamma\,\hat{a}^\dagger\hat{a}^\dagger\Big)\,|0\rangle\,. \tag{5.30}$$

This equation takes a little effort to prove, but it is true.

5.2 More general squeezed states

In the limit $\gamma\to+\infty$ the state in (5.26) is completely squeezed in the x coordinate. It takes the form

$$|0_\infty\rangle \;\sim\; \exp\Big(-\frac{1}{2}\,\hat{a}^\dagger\hat{a}^\dagger\Big)\,|0\rangle\,, \tag{5.31}$$

where we have dropped the normalization constant, which is actually going to zero. We see that $(\hat{a}+\hat{a}^\dagger)|0_\infty\rangle=0$ by direct (quick) computation or by consideration of (5.27). This means

that the $\hat{x}$ operator kills this state. We conclude that the state must have a wavefunction proportional to $\delta(x)$. Alternatively, for $\gamma \to -\infty$ we have a state

$$|0_{-\infty}\rangle \;\sim\; \exp\Big(\frac{1}{2}\,\hat{a}^\dagger\hat{a}^\dagger\Big)\,|0\rangle\,, \tag{5.32}$$

This state is annihilated by $(\hat{a}-\hat{a}^\dagger)$, or equivalently, by the momentum operator. So it must be a state whose wavefunction in momentum space is $\delta(p)$ and in coordinate space is a constant! The right-hand side constructs the constant by superposition of Hermite polynomials times gaussians.

The above suggest that position states $|x\rangle$ (and momentum states $|p\rangle$) are squeezed states of the harmonic oscillator. Indeed, we can introduce the more general squeezed states

$$|x\rangle \;=\; \mathcal{N}\exp\Big(\sqrt{\frac{2m\omega}{\hbar}}\;x\,\hat{a}^\dagger \;-\; \frac{1}{2}\,\hat{a}^\dagger\hat{a}^\dagger\Big)|0\rangle\,. \tag{5.33}$$

A short calculation (do it!) shows that, indeed,

$$\hat{x}|x\rangle \;=\; \sqrt{\frac{\hbar}{2m\omega}}(\hat{a}+\hat{a}^\dagger)|x\rangle = x|x\rangle\,. \tag{5.34}$$

The normalization constant is x dependent and is quickly determined by contracting (5.33) with the ground state

$$\langle 0|x\rangle \;=\; \mathcal{N}\langle 0|\exp\Big(\sqrt{\frac{2m\omega}{\hbar}}\;x\,\hat{a}^\dagger \;-\; \frac{1}{2}\,\hat{a}^\dagger\hat{a}^\dagger\Big)|0\rangle \;=\; \mathcal{N}. \tag{5.35}$$

We thus conclude that the normalization constant is just the ground state wavefunction: $\mathcal{N} = \psi_0(x)$. Using (1.42) we finally have

$$|x\rangle \;=\; \Big(\frac{m\omega}{\pi\hbar}\Big)^{1/4}\exp\Big(-\frac{m\omega}{2\hbar}x^2\Big)\exp\Big(\sqrt{\frac{2m\omega}{\hbar}}\;x\,\hat{a}^\dagger \;-\; \frac{1}{2}\,\hat{a}^\dagger\hat{a}^\dagger\Big)|0\rangle\,. \tag{5.36}$$

Rather general squeezed states are obtained as follows. Recall that for coherent states we used the operator $D(\alpha)$ (D for displacement!) acting on the vacuum

$$D(\alpha) \;=\; \exp\big(\alpha\,\hat{a}^\dagger - \alpha^*\hat{a}\big)\,, \qquad |\alpha\rangle \;=\; D(\alpha)|0\rangle\,.$$

We can now introduce more general squeezed states $|\alpha,\gamma\rangle$ by first squeezing and then translating:

$$|\alpha,\gamma\rangle \;\equiv\; D(\alpha)S(\gamma)|0\rangle\,.$$

Note that $|0,\gamma\rangle = |0_\gamma\rangle$ and $|\alpha,0\rangle = |\alpha\rangle$.

5.3 Photon states

For a classical electromagnetic field the energy E is obtained by adding the contributions of the electric and magnetic field:

$$E \;=\; \frac{1}{2}\int d^3x\,\epsilon_0\left[\vec{E}^2(\vec{r},t)+c^2\vec{B}^2(\vec{r},t)\right]. \tag{5.37}$$

We consider a rectangular cavity of volume V with a single mode of the electromagnetic field, namely, a single frequency ω and corresponding wavenumber $k=\omega/c$. The electromagnetic fields form a standing wave in which electric and magnetic fields are out of phase. They can take the form

$$\begin{aligned} E_x(z,t) \;&=\; \sqrt{\frac{2}{V\epsilon_0}}\;\omega\,q(t)\sin kz\,, \\ cB_y(z,t) \;&=\; \sqrt{\frac{2}{V\epsilon_0}}\;\;p(t)\;\cos kz\;, \end{aligned} \tag{5.38}$$

The classical time-dependent functions $q(t)$ and $p(t)$ are to become in the quantum theory Heisenberg operators $\hat{q}(t)$ and $\hat{p}(t)$ with commutation relations $[\,\hat{q}(t)\,,\hat{p}(t)\,]\;=\;i\hbar$. A calculation of the energy E in (5.37) with the fields above gives[6]

$$E \;=\; \frac{1}{2}\big(p^2(t)+\omega^2\,q^2(t)\big) \tag{5.39}$$

There is some funny business here with units. The variables $q(t)$ and $p(t)$ do not have their familiar units, as you can see from the expression for the energy. Indeed one is missing a quantity with units of mass that divides the p^2 contribution and multiplies the q^2 contribution. One can see that p has units of $\sqrt{E}$ and q has units of $T\sqrt{E}$. Still, the product of q and p has the units of $\hbar$, which is useful. Since photons are massless particles there is no quantity with units of mass that we can use. Note that the dynamical variable $q(t)$ is not a position, it is essentially the electric field. The dynamical variable $p(t)$ is not a momentum, it is essentially the magnetic field.

The quantum theory of this EM field uses the structure implied by the above classical results. From the energy above we are let to *postulate* a Hamiltonian

$$H \;=\; \frac{1}{2}\big(\hat{p}^2+\omega^2\,\hat{q}^2\big)\,, \tag{5.40}$$

with Schrödinger operators $\hat{q}$ and $\hat{p}$ (and associated Heisenberg operators $\hat{q}(t)$ and $\hat{p}(t)$) that satisfy $[\hat{q},\hat{p}]=i\hbar$. As soon as we declare that the classical variables $q(t)$ and $p(t)$ are to become operators, we have the implication that the electric and magnetic fields in (5.38) will become

[6] If you wish to do the computation just recall that over the volume the average of $\sin^2 kz$ or $\cos^2 kz$ is $1/2$.

field operators, that is to say, space and time-dependent operators (more below!). This oscillator is our familiar SHO, but with m set equal to one, which is allowed given the unusual units of $\hat{q}$ and $\hat{p}$. With the familiar (1.17) and $m = 1$ we have

$$\hat{a} = \frac{1}{\sqrt{2\hbar\omega}}\,(\omega\hat{q} + i\hat{p}) \;\;, \qquad \hat{a}^\dagger = \frac{1}{\sqrt{2\hbar\omega}}\,(\omega\hat{q} - i\hat{p}) \;, \qquad [\hat{a}\,,\,\hat{a}^\dagger] \;=\; 1\,. \tag{5.41}$$

It follows that

$$\hbar\omega\,\hat{a}^\dagger\hat{a} \;=\; \frac{1}{2}\,(\omega\hat{q} - i\hat{p})\,(\omega\hat{q} + i\hat{p}) \;=\; \frac{1}{2}\big(\hat{p}^2 + \omega^2\hat{q}^2 + i\omega[\hat{q},\hat{p}]\big) \;=\; \frac{1}{2}\big(\hat{p}^2 + \omega^2\hat{q}^2 - \hbar\omega\big) \tag{5.42}$$

and comparing with (5.40) this gives the Hamiltonian

$$H \;=\; \hbar\omega\Big(\hat{a}^\dagger\hat{a} + \frac{1}{2}\Big) \;=\; \hbar\omega\big(\hat{N} + \frac{1}{2}\big)\,. \tag{5.43}$$

This was the expected answer as this formula does not depend on m and thus our setting it to one should have no import. At this point we got photons! A quantum state of the electromagnetic field is a photon state, which is just a state of the harmonic oscillator Hamiltonian above. In the number basis the state $|n\rangle$ with number eigenvalue n, has energy $\hbar\omega(n + \frac{1}{2})$ which is, up to the zero-point energy $\hbar\omega/2$, the energy of n photons each of energy $\hbar\omega$.

For more intuition we now consider the electromagnetic field operator, focusing on the electric field operator. For this we first note that

$$\hat{q} \;=\; \sqrt{\frac{\hbar}{2\omega}}\;(\hat{a} + \hat{a}^\dagger)\,, \tag{5.44}$$

and the corresponding Heisenberg operator is, using (3.53) and (3.54),

$$\hat{q}(t) \;=\; \sqrt{\frac{\hbar}{2\omega}}\;(\hat{a}e^{-i\omega t} + \hat{a}^\dagger e^{i\omega t})\,. \tag{5.45}$$

In quantum field theory –which is what we are doing here– the electric field becomes a Hermitian operator. Its form is obtained by substituting (5.45) into (5.38):

$$\boxed{\hat{E}_x(z,t) \;=\; \mathcal{E}_0\,\big(\hat{a}e^{-i\omega t} + \hat{a}^\dagger e^{i\omega t}\big)\sin kz\,, \qquad \mathcal{E}_0 \;=\; \sqrt{\frac{\hbar\omega}{\epsilon_0 V}}\,.} \tag{5.46}$$

This is a field operator in the sense that it is an operator that depends on time and on space (z in this case). The constant $\mathcal{E}_0$ is sometimes called the electric field of a photon.

A classical electric field can be identified as the expectation value of the electric field operator in the given photon state. We immediately see that in the energy eigenstate $|n\rangle$ the expectation value of $\hat{E}_x$ takes the form

$$\langle\hat{E}_x(z,t)\rangle \;=\; \mathcal{E}_0\,\big(\langle n|\hat{a}|n\rangle e^{-i\omega t} + \langle n|\hat{a}^\dagger|n\rangle e^{i\omega t}\big)\sin kz \;\;=\;\; 0\,, \tag{5.47}$$

since the matrix elements on the right hand side are zero. Thus the energy eigenstates of the photon field do not correspond to classical electromagnetic fields. Consider now the expectation value of the field in a coherent state $|\alpha\rangle$ with $\alpha \in \mathbb{C}$. This time we get

$$\hat{E}_x(z,t)\rangle \;=\; \mathcal{E}_0 \quad \alpha|\hat{a}|\alpha\rangle e^{-i\omega t} + \;\alpha|\hat{a}^\dagger|\alpha\rangle e^{i\omega t}\big)\sin kz\,. \tag{5.48}$$

Since $\hat{a}|\alpha\rangle = \alpha|\alpha\rangle$ we get

$$\hat{E}_x(z,t)\rangle \;=\; \mathcal{E}_0 \quad \alpha\, e^{-i\omega t} + \alpha^*\, e^{i\omega t}\big)\sin kz\,. \tag{5.49}$$

This now looks like a familiar standing wave! If we set $\alpha = |\alpha|e^{i\theta}$, we have

$$\hat{E}_x(z,t)\rangle \;=\; 2\mathcal{E}_0\,\mathrm{Re}(\alpha e^{-i\omega t})\sin kz \;\;=\;\; 2\mathcal{E}_0\,|\alpha|\,\cos(\omega t-\theta)\sin kz. \tag{5.50}$$

The coherent photon states are the ones that have a nice classical limit with classical electric fields. The standing wave in (5.50) corresponds to a state $|\alpha\rangle$ where the expectation value of the number operator $\hat{N}$ is $|\alpha|^2$. This is the expected number of photons in the state. It follows that the expectation value of the energy is

$$H\rangle_\alpha \;=\; \hbar\omega\,|\alpha|^2 + \frac{1}{2}\hbar\omega\,. \tag{5.51}$$

Up to the zero-point energy, the expected value of the energy is equal to the number of photons times $\hbar\omega$.

TWO STATE SYSTEMS

B. Zwiebach
November 15, 2013

Contents

1 Introduction

A two-state system does not just have two states! It has two *basis* states, namely the state space is a two-dimensional complex vector space. For such a state space the Hamiltonian can be viewed as the most general Hermitian 2×2 matrix. In the case when the Hamiltonian is time-independent, this Hermitian matrix is characterized by four real numbers.

Two-state systems are idealizations that are valid when other degrees of freedom are ignored. A spin one-half particle is a two-state system with regards to spin, but being a particle, it may move and thus has position, or momentum degrees of freedom that imply a much larger, higher dimensional state space. Only if we ignore these degrees of freedom – perhaps because the particle is at rest – we can speak of a two-state system.

We will study here two specific examples of two-state systems. The first will be the ammonia molecule, which exhibits curious oscillations. The second will be spin one-half particles as used in nuclear magnetic resonance.

The mathematics of two-state systems is always the same. In fact any two-state system can be visualized as a spin system, and this will sometimes be quite useful.

2 Spin precession in a magnetic field

Let us first recall our earlier discussion of magnetic dipole moments. Classically we had the following relation valid for a charged particle

$$\boldsymbol{\mu} \;=\; \frac{q}{2m}\mathbf{S} \tag{2.1}$$

where $\boldsymbol{\mu}$ is the dipole moment, q is the charge of the particle, m its mass, and $\mathbf{S}$ is its angular momentum, arising from its spinning. In the quantum world this equation gets modified by a constant unit-free factor g, different for each particle:

$$\boldsymbol{\mu} \;=\; g\,\frac{q}{2m}\mathbf{S} \;=\; g\frac{q\hbar}{2m}\frac{\mathbf{S}}{\hbar}\,. \tag{2.2}$$

Here $q\hbar/(2m)$ has the units of dipole moment. If we consider electrons and protons the following definitions are thus natural:

$$\begin{aligned} \text{Born magneton:}\quad \mu_B \;&=\; \frac{e\hbar}{2m_e} \;=\; 5.78\times 10^{-11}\,\frac{\text{MeV}}{\text{Tesla}}\,, \\ \text{Nuclear magneton:}\quad \mu_N \;&=\; \frac{e\hbar}{2m_p} \;=\; 3.15\times 10^{-14}\,\frac{\text{MeV}}{\text{Tesla}}\,. \end{aligned} \tag{2.3}$$

Note that the nuclear magneton is about two-thousand times smaller than the Bohr magneton. Nuclear magnetic dipole moments are much smaller that that of the electron! Including the g constant we have the following results. For an electron $g = 2$ and since the electron charge is negative we get

$$\boldsymbol{\mu}_e \;=\; -2\,\mu_B\,\frac{\mathbf{S}}{\hbar}\,. \tag{2.4}$$

The dipole moment and the angular momentum are antiparallel. For a proton, the experimental result is

$$\boldsymbol{\mu}_p \;=\; 2.79\,\mu_N\,\frac{\mathbf{S}}{\hbar}\,. \tag{2.5}$$

The neutron is neutral, so one would expect no magnetic dipole moment. But the neutron is not elementary: it is made by electrically charged quarks. A dipole moment is thus possible, depending on the way quarks are distributed. Indeed, experimentally,

$$\boldsymbol{\mu}_n \;=\; -1.91\,\mu_N\,\frac{\mathbf{S}}{\hbar}\,. \tag{2.6}$$

Somehow the negative charge beats the positive charge in its contribution to the dipole moment of the neutron.

For notational convenience we introduce the constant γ from

$$\boxed{\boldsymbol{\mu} \;=\; \gamma\,\mathbf{S}\,,\quad \text{with}\;\; \gamma \;=\; \frac{gq}{2m}\,.} \tag{2.7}$$

If we insert the particle in a magnetic field $\mathbf{B}$, the Hamiltonian H_S for the spin system is

$$\boxed{H_S \;=\; -\boldsymbol{\mu}\cdot\mathbf{B} \;=\; -\gamma\,\mathbf{B}\cdot\mathbf{S} \;=\; -\gamma\,(B_x\hat{S}_x + B_y\hat{S}_y + B_z\hat{S}_z)\,.} \tag{2.8}$$

If, for example we have a magnetic field $\vec{B} = B\hat{z}$ along the z axis, the Hamiltonian is

$$H_S \;=\; -\gamma B\,\hat{S}_z \tag{2.9}$$

The associated time evolution unitary operator is

$$\mathcal{U}(t,0) \;=\; \exp\Big(-\frac{iH_S t}{\hbar}\Big) \;=\; \exp\Big(-\frac{i(-\gamma B t)\hat{S}_z}{\hbar}\Big) \tag{2.10}$$

We now recall a result that was motivated in the homework. You examined a unitary operator $R_{\mathbf{n}}(\alpha)$ defined by a unit vector $\mathbf{n}$ and an angle α, and given by

$$R_{\mathbf{n}}(\alpha) \;=\; \exp\Big(-\frac{i\alpha\hat{S}_{\mathbf{n}}}{\hbar}\Big)\,, \quad \text{with} \quad \hat{S}_{\mathbf{n}} \equiv \mathbf{n}\cdot\mathbf{S}\,. \tag{2.11}$$

You found evidence that when acting on a spin state, this operator rotates it by an angle α about the axis defined by the vector $\mathbf{n}$. If we now compare (2.11) and (2.10) we conclude that $\mathcal{U}(t,0)$ should generate a rotation by the angle $(-\gamma B t)$ about the z-axis. We now confirm this explicitly.

Consider a spin pointing at time equal zero along the direction specified by the angles (θ_0, ϕ_0):

$$|\Psi,0\rangle \;=\; \cos\frac{\theta_0}{2}|+\rangle + \sin\frac{\theta_0}{2}e^{i\phi_0}|-\rangle \tag{2.12}$$

Given the Hamiltonian $H_S \;=\; -\gamma B\,\hat{S}_z$ in (2.9) we have

$$H_S|\pm\rangle \;=\; \mp\frac{\gamma B\hbar}{2}\,. \tag{2.13}$$

Then we have

$$\begin{aligned} |\Psi,t\rangle \;&=\; e^{-iH_S t/\hbar}|\Psi,0\rangle \;=\; e^{-iH_S t/\hbar}\Big(\cos\frac{\theta_0}{2}|+\rangle + \sin\frac{\theta_0}{2}e^{i\phi_0}|-\rangle\Big) \\ &=\; \cos\frac{\theta_0}{2}e^{-iH_S t/\hbar}|+\rangle + \sin\frac{\theta_0}{2}e^{i\phi_0}e^{-iH_S t/\hbar}|-\rangle \\ &=\; \cos\frac{\theta_0}{2}e^{+i\gamma B t/2}|+\rangle + \sin\frac{\theta_0}{2}e^{i\phi_0}e^{-i\gamma B t/2}|-\rangle \end{aligned} \tag{2.14}$$

using (2.13). To recognize the resulting state it is convenient to factor out the phase that multiplies the $|+\rangle$ state:

$$|\Psi,t\rangle \;=\; e^{+i\gamma B t/2}\Big(\cos\frac{\theta_0}{2}|+\rangle + \sin\frac{\theta_0}{2}e^{i(\phi_0-\gamma B t)}|-\rangle\Big). \tag{2.15}$$

Since the overall phase is not relevant we can now recognize the spin state as the state corresponding to the vector $\vec{n}(t)$ defined by angles

$$\begin{aligned} \theta(t) &= \theta_0\,, \\ \phi(t) &= \phi_0 - \gamma B t\,. \end{aligned} \tag{2.16}$$

Keeping θ constant while changing ϕ indeed corresponds to a rotation about the z axis and, after time t, the spin has rotated an angle $(-\gamma B t)$ as claimed above.

In fact spin states in a magnetic field precess in exactly the same way that magnetic dipoles in classical electromagnetism precess. The main fact from electromagnetic theory that we need is that in a magnetic field a dipole moment experiences a torque $\boldsymbol{\tau}$ given by

$$\boldsymbol{\tau} \;=\; \boldsymbol{\mu} \times \mathbf{B}\,. \tag{2.17}$$

Then the familiar mechanics equation for the rate of change of angular momentum being equal to the torque gives

$$\frac{d\mathbf{S}}{dt} \;=\; \boldsymbol{\tau} \;=\; \boldsymbol{\mu} \times \mathbf{B} \;=\; \gamma \mathbf{S} \times \mathbf{B}\,, \tag{2.18}$$

which we write as

$$\frac{d\mathbf{S}}{dt} \;=\; -\gamma \mathbf{B} \times \mathbf{S}\,. \tag{2.19}$$

We recognize that this equation states that the time dependent vector is rotating with angular velocity $\vec{\omega}_L$ given by

$$\boxed{\boldsymbol{\omega}_L \;=\; -\gamma \mathbf{B}\,.} \tag{2.20}$$

This is the so-called Larmor frequency. Indeed, this identification is standard in mechanics. A vector $\mathbf{v}$ rotating with angular velocity $\boldsymbol{\omega}$ satisfies the differential equation

$$\frac{d\mathbf{v}}{dt} \;=\; \boldsymbol{\omega} \times \mathbf{v}\,. \tag{2.21}$$

You can convince yourself of this with the help of a simple picture (see Figure 1) . Also note that the differential equation shows that the derivative of $\mathbf{v}$, given by the right-hand side, is orthogonal to $\mathbf{v}$ because the cross product involves $\mathbf{v}$. This is as it should when the vector $\mathbf{v}$ is rotated. Indeed, show that the above differential equation implies that $\frac{d}{dt}\mathbf{v} \cdot \mathbf{v} = 0$, so that the length of $\mathbf{v}$ is unchanged.

The Hamiltonian of a general spin in a magnetic field (2.8) is then

$$\boxed{H_S \;=\; -\boldsymbol{\mu} \cdot \mathbf{B} \;=\; -\gamma \mathbf{B} \cdot \mathbf{S} \;=\; \boldsymbol{\omega}_L \cdot \mathbf{S}\,.} \tag{2.22}$$

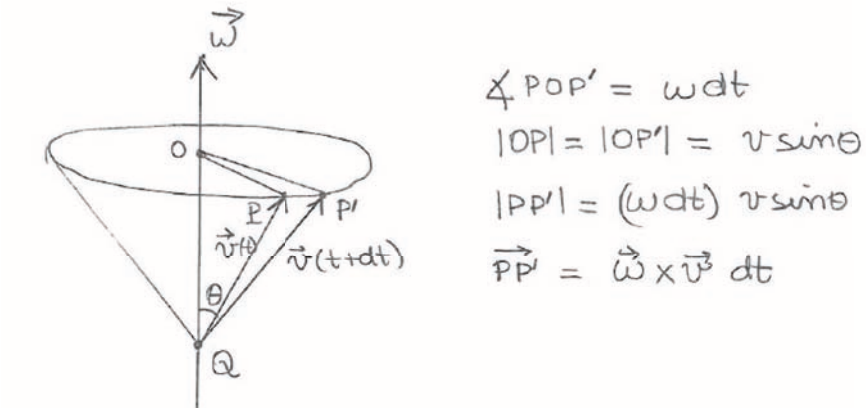


Figure 1: The vector $\mathbf{v}(t)$ and an instant later the vector $\mathbf{v}(t+dt)$. The angular velocity vector $\boldsymbol{\omega}$ is along the axis and $\mathbf{v}$ rotates about is origin Q. At all times the vector $\mathbf{v}$ and $\boldsymbol{\omega}$ make an angle θ. The calculations to the right should convince you that (2.21) is correct.

For time independent magnetic fields $\boldsymbol{\omega}_L$ is also time independent and the evolution operator is

$$\mathcal{U}(t,0) \;=\; \exp\bigl(-iH_S t/\hbar\bigr) \;=\; \exp\Bigl(-i\,\frac{\boldsymbol{\omega}_L\cdot\mathbf{S}}{\hbar}\,t\Bigr)\,. \tag{2.23}$$

If we write

$$\boldsymbol{\omega}_L \;=\; \omega_L\,\mathbf{n}\,,\quad \mathbf{n}\cdot\mathbf{n}=1\,, \tag{2.24}$$

we have

$$\mathcal{U}(t,0) \;=\; \exp\Bigl(-i\,\frac{\omega_L t\hat{S}_{\mathbf{n}}}{\hbar}\;\Bigr) \;=\; R_{\mathbf{n}}(\omega_L t)\,, \tag{2.25}$$

where we compared with (2.11). The time evolution operator $\mathcal{U}(t,0)$ rotates the spin states by the angle $\omega_L t$ about the $\mathbf{n}$ axis. In other words

With $H_S \;=\; \boldsymbol{\omega}_L\cdot\mathbf{S}$ spin states precess with angular velocity $\boldsymbol{\omega}_L\,.$	(2.26)

3 The general two-state system viewed as a spin system

The most general time-independent Hamiltonian for a two-state system is a hermitian operator represented by the most general hermitian two-by-two matrix H. In doing so we are using some orthonomal basis $\{|1\rangle,|2\rangle\}$. In any such basis the matrix can be characterized by four real constants $g_0,g_1,g_2,g_3\in\mathbb{R}$ as follows:

$$H \;=\; \begin{pmatrix} g_0+g_3 & g_1-ig_2 \\ g_1+ig_2 & g_0-g_3 \end{pmatrix} \;=\; g_0\mathbf{1}+g_1\sigma_1+g_2\sigma_2+g_3\sigma_3\,. \tag{3.27}$$

On the right-hand side we wrote the matrix as a sum of matrices, where $\mathbf{1}$ and the Pauli matrices σ_i, $i = 1, 2, 3$ are hermitian. We view (g_1, g_2, g_3) as a vector $\mathbf{g}$ and then define

$$\mathbf{g} \cdot \boldsymbol{\sigma} \;\equiv\; g_1\sigma_2 + g_2\sigma_2 + g_3\sigma_3 \,. \tag{3.28}$$

In this notation

$$\boxed{H \;=\; g_0\mathbf{1} + \mathbf{g} \cdot \boldsymbol{\sigma} \,.} \tag{3.29}$$

It is again convenient to introduce the magnitude g and the direction $\mathbf{n}$ of $\mathbf{g}$:

$$\mathbf{g} \;=\; g\,\mathbf{n} \,, \quad \mathbf{n}\cdot\mathbf{n} = 1 \,, \quad g \;=\; \sqrt{g_1^2 + g_2^2 + g_3^2} \,. \tag{3.30}$$

Now the Hamiltonian reads

$$H \;=\; g_0\mathbf{1} + g\,\mathbf{n} \cdot \boldsymbol{\sigma} \,. \tag{3.31}$$

Recall now that the spin states $|\mathbf{n}; \pm\rangle$ are eigenstates of $\mathbf{n} \cdot \boldsymbol{\sigma}$

$$\mathbf{n} \cdot \boldsymbol{\sigma}\, |\mathbf{n}; \pm\rangle \;=\; \pm|\mathbf{n}; \pm\rangle \,. \tag{3.32}$$

In writing the spin states, however, you must recall that what we call the z-up and z-down states are just the first and second basis states: $|+\rangle = |1\rangle$ and $|-\rangle = |2\rangle$. With this noted, the spin states $|\mathbf{n}; \pm\rangle$ are indeed the eigenstates of H, since using the last two equations above we have

$$H|\mathbf{n}; \pm\rangle \;=\; (g_0 \pm g)|\mathbf{n}; \pm\rangle \,. \tag{3.33}$$

This also shows that the energy eigenvalues are $g_0 \pm g$. In summary:

$$\text{Spectrum:} \;\; |\mathbf{n}; +\rangle \;\; \text{with energy } g_0 + g \,, \qquad |\mathbf{n}; -\rangle \;\; \text{with energy } g_0 - g \,. \tag{3.34}$$

Thus our spin states allow us to write the general solution for the spectrum of the Hamiltonian (again, writing $|+\rangle = |1\rangle$ and $|-\rangle = |2\rangle$). Clearly the $|1\rangle$ and $|2\rangle$ states will generally have nothing to do with spin states. They are the basis states of any two-state system.

To understand the time evolution of states with the Hamiltonian H, we first rewrite H in terms of the spin operators, instead of the Pauli matrices, recalling that $\mathbf{S} = \frac{\hbar}{2}\boldsymbol{\sigma}$. Using (3.29) we find

$$H \;=\; g_0\mathbf{1} \,+\, \frac{2}{\hbar}\,\mathbf{g} \cdot \mathbf{S} \,. \tag{3.35}$$

Comparison with the spin Hamiltonian $H_S = \boldsymbol{\omega}_L \cdot \mathbf{S}$ shows that in the system described by H the states precess with angular velocity $\boldsymbol{\omega}$ given by

$$\boxed{\boldsymbol{\omega} \;=\; \frac{2}{\hbar}\,\mathbf{g} \,.} \tag{3.36}$$

Note that the part $g_0\mathbf{1}$ of the Hamiltonian H does not rotate states during time evolution; it simply multiplies states by the time-dependent phase $\exp(-ig_0t/\hbar)$.

Operationally, if H is known, the vector $\boldsymbol{\omega}$ above is immediately calculable. And given a normalized state $\alpha|1\rangle + \beta|2\rangle$ of the system ($|\alpha|^2 + |\beta|^2 = 1$), we can identify the corresponding spin state $|\mathbf{n}; +\rangle = \alpha|+\rangle + \beta|-\rangle$. The time evolution of the spin state is due to Larmor precession and is intuitively understood. With this result, the time evolution of the state in the original system is simply obtained by letting $|+\rangle \to |1\rangle$ and $|-\rangle \to |2\rangle$ in the precessing spin state.

4 The ammonia molecule as a two-state system

The ammonia molecule NH_3 is composed of four atoms, one nitrogen and three hydrogen. Ammonia is naturally a gas, without color, but with a pungent odor. It is mostly used for fertilizers, and also for cleaning products and pharmaceuticals.

The ammonia molecule takes the shape of a flattened tetrahedron. If we imagine the three hydrogen atoms forming an equilateral triangle at the base, the nitrogen atom sits atop. The angle formed between any two lines joining the nitrogen to the hydrogen is about 108° – this indeed corresponds to a flattened tetrahedron since a regular tetrahedron would have a 60° angle. If the nitrogen was pushed all the way down to the base, the angle would be 120°.

The ammonia molecule has electronic excitations, vibrational excitations and rotational excitations. Those must largely be ignored in the two-state description of the molecule. The two states arise from transitions in which the nitrogen atom flips from being above the fixed hydrogen plane to being below the hydrogen plane. Since such a flip could be mimicked by a full rotation of the molecule, we can describe the transition more physically by considering the molecule spinning about the axis perpendicular to the hydrogen plane, with the N up, where up is the direction of the angular momentum. The transition would have the N down, or against the angular momentum of the rotating molecule.

More briefly, the two states are: nitrogen up, or nitrogen down. Both are classically stable configurations separated by a potential energy barrier. In classical mechanics these are the two options and they are degenerate in energy.

As long as the energy barrier is not infinite, in quantum mechanics the degeneracy is broken. This, of course is familiar. We can roughly represent the potential experienced by the nitrogen atom as the potential $V(z)$ in figure 3, where the two equilibrium positions of the nitrogen are at $\pm z_0$ and they are separated by a large barrier. In such a potential the ground state, which is symmetric, and the first excited state, which is antisymmetric, are almost degenerate in energy when the barrier is high. If the potential barrier was infinite, the two possible eigenstates would be the nitrogen wavefunction localized about z_0 and the nitrogen wavefunction localized about $-z_0$. Moreover, those states would be degenerate in energy. But with a large but finite barrier the ground state is represented by a wavefunction $\psi_g(z)$ even in z, as shown below the potential.

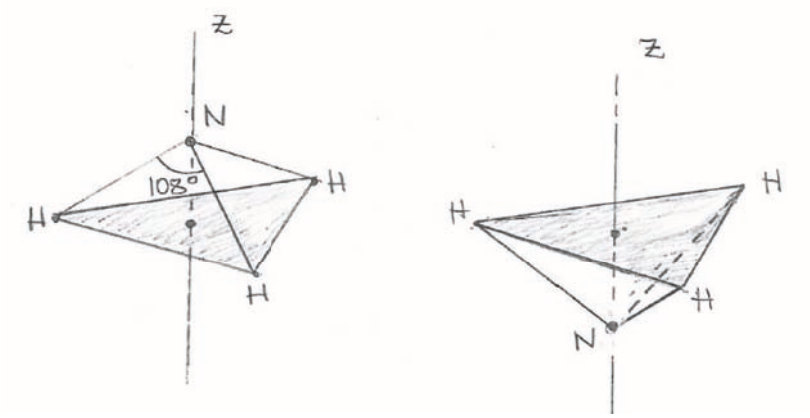


Figure 2: The ammonia molecule looks like a flattened tetrahedron. The nitrogen atom can be up or down with respect to the plane defined by the three hydrogen atoms. These are the two states of the ammonia molecule.

This even wavefunction is roughly the superposition, with the same sign, of the two localized wavefunctions. The next excited state $\psi_e(z)$ is odd in z and is roughly the superposition, this time with opposite signs, of the two localized wavefunctions.

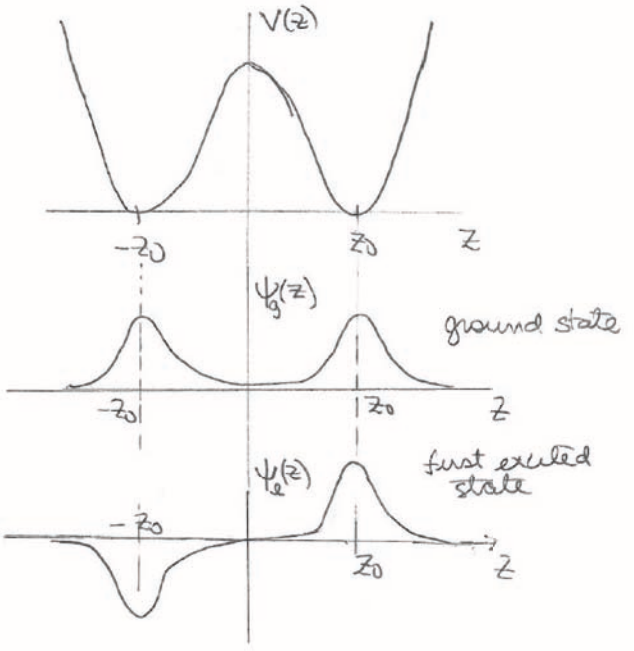


Figure 3: The potential $V(z)$ experienced by the nitrogen atom. There are two classically stable positions $\pm z_0$. The ground state and (first) excited state wavefunctions $\psi_g(z)$ and $\psi_e(z)$ are sketched below the potential.

Let us attempt a quantitative description of the situation. Let us label the two possible states:

$$|\uparrow\rangle \text{ is nitrogen up}\,, \qquad |\downarrow\rangle \text{ is nitrogen down} \tag{4.37}$$

We can associate to $|\uparrow\rangle$ a (positive) wavefunction localized around z_0 and to $|\downarrow\rangle$ a (positive) wavefunction localized around $-z_0$. Suppose the energy barrier is infinite. In this case the two states above must be energy eigenstates with the same energy E_0:

$$\begin{aligned} H|\uparrow\rangle &= E_0\,|\uparrow\rangle\,, \\ H|\downarrow\rangle &= E_0\,|\downarrow\rangle\,. \end{aligned} \tag{4.38}$$

The energy E_0 is arbitrary and will not play an important role. Choosing a basis

$$|1\rangle \equiv |\uparrow\rangle, \quad |2\rangle \equiv |\downarrow\rangle\,, \tag{4.39}$$

the Hamiltonian, in this basis takes the form of the two-by-two matrix

$$H \;=\; \begin{pmatrix} E_0 & 0 \\ 0 & E_0 \end{pmatrix}\,. \tag{4.40}$$

The ability to tunnel must correspond to off-diagonal elements in the Hamiltonian matrix – there is no other option, in fact! So we must have a nonzero $H_{12} = \langle 1|H|2\rangle \neq 0$. Since the Hamiltonian must be hermitian, we must have $H_{12} = H_{21}^*$. For the time being we will take the off-diagonal elements to be real and therefore:

$$H_{12} \;=\; H_{21} \;=\; -\Delta\,, \quad \Delta > 0\,. \tag{4.41}$$

The sign of the real constant Δ is conventional. We could change it by a change of basis in which we let, for example $|2\rangle \to -|2\rangle$. Our choice will be convenient. The full Hamiltonian is now

$$H \;=\; \begin{pmatrix} E_0 & -\Delta \\ -\Delta & E_0 \end{pmatrix} \;=\; E_0\,\mathbf{1} \;-\; \Delta\,\sigma_1\,, \tag{4.42}$$

where in the last step we wrote the matrix as a sum of a real number times the two-by-two identity matrix plus another real number times the first Pauli matrix. Both the identity matrix and the Pauli matrix are hermitian, consistent with having a hermitian Hamiltonian. The eigenvalues of the Hamiltonian follow from the equation

$$\begin{vmatrix} E_0-\lambda & -\Delta \\ -\Delta & E_0-\lambda \end{vmatrix} \;=\; 0 \quad\to\quad (E_0-\lambda)^2 = \Delta^2 \quad\to\quad \lambda_\pm \;=\; E_0 \pm \Delta\,. \tag{4.43}$$

The eigenstates corresponding to these eigenvalues are

$$\begin{aligned} |G\rangle &= \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}\big(|\uparrow\rangle + |\downarrow\rangle\big)\,, \quad E = E_0 - \Delta\,, \;\; \text{Ground state} \\ |E\rangle &= \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}}\big(|\uparrow\rangle - |\downarrow\rangle\big)\,, \quad E = E_0 + \Delta\,, \;\; \text{Excited state} \end{aligned} \tag{4.44}$$

Here $|G\rangle$ is the ground state (thus the G) and $|E\rangle$ is the excited state (thus the E). Our sign choices make the correspondence of the states with the wavefunction in figure 3 clear. The ground state $|G\rangle$ is a superposition, with the same sign, of the two localized (positive) wavefunctions. The excited state $|E\rangle$ has the wavefunction localized at z_0 (corresponding to $|1\rangle$ with) a plus sign and the wavefunction localized at $-z_0$ (corresponding to $|2\rangle$) with at negative sign. Note that in the notation of section 3, the Hamiltonian in (4.42) corresponds to $\mathbf{g} = -\Delta\mathbf{x}$ and therefore $\mathbf{n} = -\mathbf{x}$ and $g = \Delta$. The excited state corresponds to the spin state $\frac{1}{\sqrt{2}}(|+\rangle - |-\rangle)$, which points in the $-\mathbf{x}$ direction. The ground state corresponds to the spin state $\frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)$ which points in the $+\mathbf{x}$ direction.

The energy difference between these two eigenstates is 2Δ, which for the ammonia molecule takes the value

$$2\Delta \;=\; 0.9872 \times 10^{-4}\text{ev}\,. \tag{4.45}$$

Setting this energy equal to the energy of a photon that may be emitted in such a transition we find

$$2\Delta \;=\; \hbar\omega \;=\; h\nu\,, \quad \nu \;=\; 23.870 \times 10^{9}\text{Hz} \;=\; 23.870\,\text{GHz}\,, \tag{4.46}$$

corresponding to a wavelength λ of

$$\lambda \;=\; 1.2559\,\text{cm}. \tag{4.47}$$

Let us consider the time evolution of an ammonia molecule that at $t = 0$ is in the state $|\uparrow\rangle$. Using (4.44) we express the initial state in terms of energy eigenstates:

$$|\Psi, 0\rangle \;=\; |\uparrow\rangle \;=\; \frac{1}{\sqrt{2}}\big(|G\rangle + |E\rangle\big)\,. \tag{4.48}$$

The time-evolved state can now be readily written:

$$|\Psi, t\rangle \;=\; \frac{1}{\sqrt{2}}\Big(e^{-i(E-\Delta)t/\hbar}|G\rangle + e^{-i(E+\Delta)t/\hbar}|E\rangle\Big)\,, \tag{4.49}$$

and we now rewrite it in terms of the $|\uparrow\rangle, |\downarrow\rangle$ states as follows

$$\begin{aligned}
|\Psi, t\rangle \;&=\; \frac{1}{\sqrt{2}}\Big(e^{-i(E-\Delta)t/\hbar}\frac{1}{\sqrt{2}}\big(|\uparrow\rangle + |\downarrow\rangle\big) + e^{-i(E+\Delta)t/\hbar}\frac{1}{\sqrt{2}}\big(|\uparrow\rangle - |\downarrow\rangle\big)\Big)\,,\\
&=\; \frac{1}{2}\,e^{-iEt/\hbar}\Big(e^{i\Delta t/\hbar}\big(|\uparrow\rangle + |\downarrow\rangle\big) + e^{-i\Delta t/\hbar}\big(|\uparrow\rangle - |\downarrow\rangle\big)\Big)\,,\\
&=\; e^{-iEt/\hbar}\Big[\cos\Big(\frac{\Delta\, t}{\hbar}\Big)|\uparrow\rangle \;+\; i\sin\Big(\frac{\Delta\, t}{\hbar}\Big)|\downarrow\rangle\Big]\,.
\end{aligned} \tag{4.50}$$

The above time-dependent state oscillates from $|\uparrow\rangle$ to $|\downarrow\rangle$ with angular frequency $\omega = \Delta/\hbar \simeq 23\,\text{GHz}$. The probabilities $P_{\uparrow\downarrow}(t)$ that the state is found with nitrogen up or down are simply

$$\begin{aligned}
P_\uparrow(t) \;&=\; |\langle\uparrow|\Psi, t\rangle|^2 \;=\; \cos^2\Big(\frac{\Delta\, t}{\hbar}\Big)\,,\\
P_\downarrow(t) \;&=\; |\langle\downarrow|\Psi, t\rangle|^2 \;=\; \sin^2\Big(\frac{\Delta\, t}{\hbar}\Big)\,.
\end{aligned} \tag{4.51}$$

A plot of these is shown in Figure 4

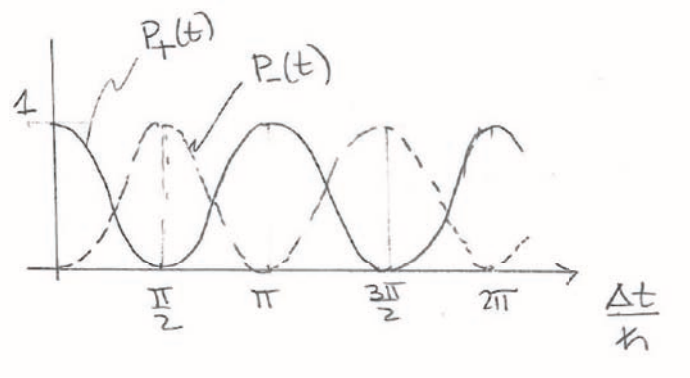


Figure 4: If the nitrogen atom starts in the state $|\uparrow\rangle$ at $t = 0$ (up position) then the probability $P_\uparrow(t)$ (shown as a continuous line) oscillates between one and zero. The probability $P_\downarrow(t)$ to be found in the state $|\downarrow\rangle$ is shown in dashed lines. Of course, $P_\uparrow(t) + P_\downarrow(t) = 1$. (Figure needs updating to change $\pm$ to $\uparrow\downarrow$.)

5 Ammonia molecule in an electric field

Let us now consider the electrostatic properties of the ammonia molecule. The electrons tend to cluster towards the nitrogen, leaving the nitrogen vertex slightly negative and the hydrogen plane slightly positive. As a result we get an electric dipole moment $\boldsymbol{\mu}$ that points down – when the nitrogen is up. The energy E of a dipole in an electric field $\boldsymbol{\mathcal{E}}$ is

$$E \;=\; -\boldsymbol{\mu}\cdot\boldsymbol{\mathcal{E}}\,. \tag{5.1}$$

With the electric field $\boldsymbol{\mathcal{E}} = \mathcal{E}\mathbf{z}$ along the z axis, and $\boldsymbol{\mu} = -\mu\mathbf{z}$, with $\mu > 0$, the state $|\uparrow\rangle$ with nitrogen up gets an extra positive contribution to the energy equal to $\mu\mathcal{E}$, while the $|\downarrow\rangle$ state gets the extra piece $-\mu\mathcal{E}$. The new Hamiltonian, including the effects of the electric field is then

$$H \;=\; \begin{pmatrix} E_0+\mu\mathcal{E} & -\Delta \\ -\Delta & E_0-\mu\mathcal{E} \end{pmatrix} \;=\; E_0\,\mathbf{1} \;-\; \Delta\,\sigma_1 \;+\mu\mathcal{E}\,\sigma_3\,, \tag{5.2}$$

This corresponds to $g = \sqrt{(\mu\mathcal{E})^2+\Delta^2}$ and therefore the energy eigenvalues are

$$\begin{aligned} E_E(\mathcal{E}) \;&= E_0 + \sqrt{\mu^2\mathcal{E}^2+\Delta^2}\,, \\ E_G(\mathcal{E}) \;&= E_0 - \sqrt{\mu^2\mathcal{E}^2+\Delta^2}\,, \end{aligned} \tag{5.3}$$

where we added the subscripts E for excited and G for ground to identify the energies as those of the excited and ground states when $\mathcal{E} = 0$. For small $\mathcal{E}$, or more precisely, small $\mu\mathcal{E}/\Delta$, we

have

$$\begin{aligned} E_E(\mathcal{E}) &\simeq E_0 + \Delta + \frac{\mu^2\mathcal{E}^2}{2\Delta} + \mathcal{O}(\mathcal{E}^4)\,, \\ E_G(\mathcal{E}) &\simeq E_0 - \Delta - \frac{\mu^2\mathcal{E}^2}{2\Delta} + \mathcal{O}(\mathcal{E}^4)\,, \end{aligned} \tag{5.4}$$

while for large $\mu\mathcal{E}$

$$\begin{aligned} E_E(\mathcal{E}) &\simeq E_0 + \mu\mathcal{E} + \mathcal{O}(1/\mathcal{E})\,, \\ E_G(\mathcal{E}) &\simeq E_0 - \mu\mathcal{E} + \mathcal{O}(1/\mathcal{E})\,. \end{aligned} \tag{5.5}$$

A plot of the energies is shown in Figure 5.

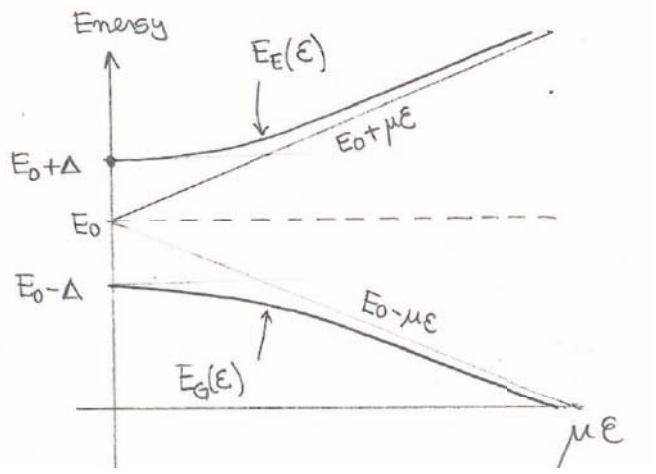


Figure 5: The energy levels of the two states of the ammonia molecule as a function of the magnitude $\mathcal{E}$ of electric field. The higher energy state, with energy $E_E(\mathcal{E})$ coincides with $|E\rangle$ when $\mathcal{E} = 0$. The the lower energy state, with energy $E_G(\mathcal{E})$ coincides with $|G\rangle$ when $\mathcal{E} = 0$.

This analysis gives us a way to split a beam of ammonia molecules into two beams, one with molecules in the state $|G\rangle$ and one with molecules in the state $|E\rangle$. As shown in the figure we have a beam entering a region with a spatially dependent electric field. The electric field gradient points up: the magnitude of the field is larger above than below. In a practical device $\mu\mathcal{E} \ll \Delta$ and we can use (5.4). A molecule in the $|E\rangle$ state will tend to go to the region of lower $|\mathcal{E}$ as this is the region of low energy. Similarly a molecule in the $|G\rangle$ state will tend to go to the region of larger $|\mathcal{E}|$. Thus this device acts as a beam splitter.

The idea now is build a resonant electromagnetic cavity tuned to the frequency of 23.87 GHz and with very small losses (a high Q cavity). On one end, through a small hole, we let in a beam of ammonia molecules in the $|E\rangle$ state. These molecules exit the cavity through another

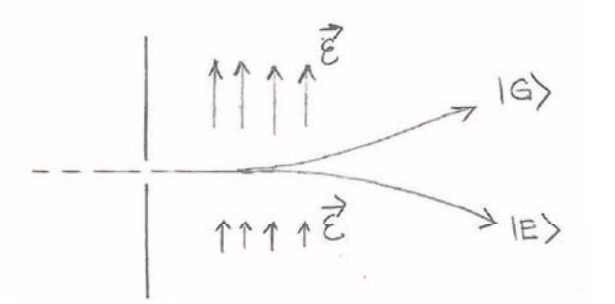


Figure 6: If a beam of ammonia molecules is exposed to an electric field with a strong gradient, molecules in the ground state $|G\rangle$ are deflected towards the stronger field (up) while molecules in the excited state $|E\rangle$ are deflected towards the weaker field (down).

hole on the opposite side (see Figure 7). If the design is done right, they exit on the ground state $|G\rangle$ thus having yielded an energy $2\Delta = \hbar\omega_0$ to the cavity. The ammonia molecules in the cavity interact with a spontaneously created electric field $\mathcal{E}$ that oscillates with the resonant frequency. The interaction with such field induces the transition $|E\rangle \to |G\rangle$. This transition also feeds energy into the field. We want to understand this transition.

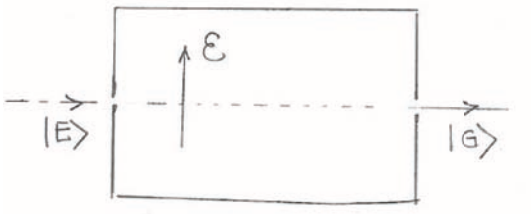


Figure 7: A resonant cavity tuned for 23.87GHz. A beam of ammonia molecules in the excited state $|E\rangle$ enter from the left. If properly designed, the molecules exit the cavity from the right on the ground state $|G\rangle$. In this process each molecule adds energy 2Δ to the electromagnetic field in the cavity.

The mathematics of the transition is clearer if we express the Hamiltonian in the primed basis

$$|1'\rangle \equiv |E\rangle\,, \quad |2'\rangle \equiv |G\rangle\,. \tag{5.6}$$

instead of the basis $|1\rangle = |\uparrow\rangle, |2\rangle = |\downarrow\rangle$ used to describe the Hamiltonian in (5.2). We can use

(4.44) to calculate the new matrix elements. For example

$$\begin{aligned}\langle E|H|E\rangle &= \frac{1}{2}\big(\langle\uparrow| - \langle\downarrow|\big)H\big(|\uparrow\rangle - |\downarrow\rangle\big)\\ &= \frac{1}{2}\big(\langle\uparrow|H|\uparrow\rangle - \langle\uparrow|H|\downarrow\rangle - \langle\downarrow|H|\uparrow\rangle + \langle\downarrow|H|\downarrow\rangle\big)\\ &= \frac{1}{2}\big(E_0 + \mu\mathcal{E} - (-\Delta) - (-\Delta) + E_0 - \mu\mathcal{E}\big)\\ &= E_0 + \Delta\,,\end{aligned} \tag{5.7}$$

and

$$\begin{aligned}\langle E|H|G\rangle &= \frac{1}{2}\big(\langle\uparrow| - \langle\downarrow|\big)H\big(|+\rangle + |\downarrow\rangle\big)\\ &= \frac{1}{2}\big(\langle+|H|\uparrow\rangle + \langle\uparrow|H|\downarrow\rangle - \langle\downarrow|H|\uparrow\rangle - \langle\downarrow|H|\downarrow\rangle\big)\\ &= \frac{1}{2}\big(E_0 + \mu\mathcal{E} + (-\Delta) - (-\Delta) - (E_0 - \mu\mathcal{E})\big)\\ &= \mu\mathcal{E}\,.\end{aligned} \tag{5.8}$$

and similarly $\langle G|H|G\rangle = E_0 - \Delta$. All in all the Hamiltonian in the new basis is given by

$$H = \begin{pmatrix} E_0+\Delta & \mu\mathcal{E} \\ \mu\mathcal{E} & E_0-\Delta \end{pmatrix} \quad \text{In the } |1'\rangle = |E\rangle,\ |2'\rangle = |G\rangle \text{ basis.} \tag{5.9}$$

We then write the wavefunction in terms of the amplitudes C_E and C_G to be in the $|E\rangle$ or $|G\rangle$ states respectively,

$$|\Psi\rangle = \begin{pmatrix} C_E(t) \\ C_G(t) \end{pmatrix}. \tag{5.10}$$

The constant energy E_0 is not relevant – it can be set to any value and we choose the value zero. Doing so, the Schrödinger equation takes the form

$$i\hbar\frac{d}{dt}\begin{pmatrix} C_E(t) \\ C_G(t) \end{pmatrix} = \begin{pmatrix} \Delta & \mu\mathcal{E} \\ \mu\mathcal{E} & -\Delta \end{pmatrix}\begin{pmatrix} C_E(t) \\ C_G(t) \end{pmatrix}. \tag{5.11}$$

A strategy to solve this equation is to imagine that $\mu\mathcal{E}$ is very small compared to Δ so that

$$\begin{pmatrix} C_E(t) \\ C_G(t) \end{pmatrix} = \begin{pmatrix} e^{-i\Delta t/\hbar}\ \beta_E(t) \\ e^{+i\Delta t/\hbar}\ \beta_G(t) \end{pmatrix}, \tag{5.12}$$

would be an exact solution with time-independent β_E and β_G if $\mu\mathcal{E} = 0$. When $\mu\mathcal{E}$ is small, we can expect solutions with β_E and β_G slowly varying in time (compared to the frequency $\Delta/\hbar$ of the phases we have brought out to the open. We now substitute into (5.11), with the result (do the algebra!) with several terms canceling and

$$i\hbar\frac{d}{dt}\begin{pmatrix} \beta_E(t) \\ \beta_G(t) \end{pmatrix} = \begin{pmatrix} 0 & e^{i\omega_0 t}\mu\mathcal{E} \\ e^{-i\omega_0 t}\mu\mathcal{E} & 0 \end{pmatrix}\begin{pmatrix} \beta_E(t) \\ \beta_G(t) \end{pmatrix}, \quad \omega_0 \equiv \frac{2\Delta}{\hbar} \tag{5.13}$$

where we defined ω_0 as the frequency of a photon associated to the transition $|E\rangle \to |G\rangle$. This frequency is the resonant frequency of the cavity to be used. We now assume that in the cavity the electric field $\mathcal{E}$ is at resonance so that

$$\mathcal{E}(t) \;=\; 2\mathcal{E}_0 \cos\omega_0 t \;=\; \mathcal{E}_0(e^{i\omega t} + e^{-i\omega t}) \tag{5.14}$$

so that

$$\begin{aligned} e^{i\omega_0 t}\mu\mathcal{E} \;&=\; \mu\mathcal{E}_0(1 + e^{2i\omega_0 t}) \\ e^{-i\omega_0 t}\mu\mathcal{E} \;&=\; \mu\mathcal{E}_0(1 + e^{-2i\omega_0 t}) \end{aligned} \tag{5.15}$$

We can now go back to the differential equation which gives

$$\begin{aligned} i\,\dot\beta_E(t) \;&=\; \frac{\mu\mathcal{E}_0}{\hbar}(1 + e^{2i\omega_0 t})\beta_G(t) \\ i\,\dot\beta_G(t) \;&=\; \frac{\mu\mathcal{E}_0}{\hbar}(1 + e^{-2i\omega_0 t})\beta_E(t)\,. \end{aligned} \tag{5.16}$$

With $\mu\mathcal{E}_0$ small, the rates of change of β_E and β_G will necessarily be small, as $\mu\mathcal{E}_0$ appears multiplicatively on the right-hand side. Thus β_E and β_G are essentially constant during the period of oscillation of the exponential terms $e^{\pm 2i\omega_0 t}$. Since these exponentials have zero time-averaged values, they can be dropped. Thus, we get

$$\begin{aligned} \dot\beta_E(t) \;&=\; -\,i\,\frac{\mu\mathcal{E}_0}{\hbar}\beta_G(t) \\ \dot\beta_G(t) \;&=\; -\,i\,\frac{\mu\mathcal{E}_0}{\hbar}\beta_E(t)\,. \end{aligned} \tag{5.17}$$

Taking another time derivative of the top equation we find;

$$\ddot\beta_E(t) \;=\; -\Big(\frac{\mu\mathcal{E}_0}{\hbar}\Big)^2 \beta_E(t) \tag{5.18}$$

This has the simple solution in terms of sines and cosines. If we assume that the molecule at time equal zero is indeed in the state $|E\rangle$ we then write

$$\beta_E(t) \;=\; \cos\Big(\frac{\mu\mathcal{E}_0}{\hbar}t\Big) \quad\to\quad \beta_G(t) \;=\; -i\sin\Big(\frac{\mu\mathcal{E}_0}{\hbar}t\Big)\,. \tag{5.19}$$

The time dependent probability $P_E(t)$ to be in the $|E\rangle$ state is then

$$P_E(t) = |C_E(t)|^2 \;=\; \big|e^{-i\Delta t/\hbar}\beta_E(t)\big| \;=\; \cos^2\Big(\frac{\mu\mathcal{E}_0}{\hbar}t\Big)\,. \tag{5.20}$$

This is our desired result. The molecule that enters the cavity in the state $|E\rangle$ will leave the cavity in the state $|G\rangle$ if the travel time T is such that the probability $P_E(T)$ to be in $|E\rangle$ vanishes. For this we need

$$\cos\Big(\frac{\mu\mathcal{E}_0}{\hbar}T\Big) \;=\; 0 \quad\to\quad \frac{\mu\mathcal{E}_0}{\hbar}T \;=\; \frac{\pi}{2}, \frac{3\pi}{2}, \ldots\,. \tag{5.21}$$

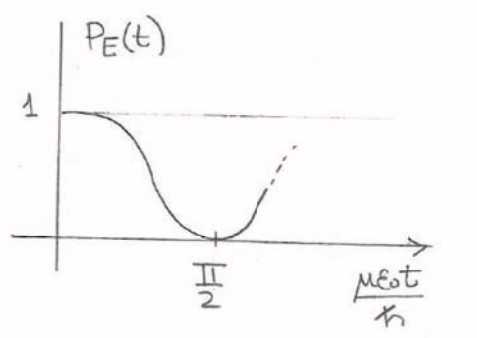


Figure 8: An ammonia molecule enters the resonant cavity at $t = 0$ in the excited state $|E\rangle$. We show probability $P_E(t)$ for the ammonia molecule to be in the excited $|E\rangle$ at time t. If the molecule exits at a time T for which $P_E(T) = 0$, the molecule will be in the ground state $|G\rangle$, as desired.

See figure 8.

If the velocity of the ammonia molecules is adjusted for this to happen, each molecule gives energy 2Δ to the cavity's electromagnetic field. The cavity's EM field, by producing the transition $|E\rangle \to |G\rangle$ of the traveling molecules, stimulates the emission of radiation. Moreover the energy released is in a field with the same configuration and frequency as the stimulating EM field. The molecules thus help build a *coherent* field in the cavity. Such a cavity is then a MASER, an acronym that stands for Microwave Amplification by Stimulated Emission of Radiation. The molecules are uncharged and therefore their motions produce no unwanted EM signals – this is in contrast to electrons in vacuum amplifiers, which produce *shot* noise.

Charles H. Townes, James P. Gordon, and H. J. Zeiger built the first ammonia maser working at Columbia University in 1953. As stated by Charles H. Townes in his Nobel lecture on December 11, 1964, "masers yield the most perfect amplification allowed by the uncertainty principle". For such EM waves the uncertainty principle can be written as

$$\Delta n \, \Delta\phi \geq \frac{1}{2} \, ,$$

where Δn is the uncertainty in the number of photons in the field and $\Delta\phi$ is the phase uncertainty of the wave, in radians. For a coherent field $\Delta n = \sqrt{\bar{n}}$, with $\bar{n}$ the expected number of photons. The saturation of the uncertainty principle leads to a phase uncertainty

$$\Delta\phi \;=\; \frac{1}{2\sqrt{\bar{n}}} \, , \tag{5.22}$$

that for any realistic $\bar{n}$ is fantastically small.

6 Nuclear Magnetic Resonance

The problem we want to discuss is that of a spin in a time dependent magnetic field. This magnetic field has a time-independent z component and a circularly polarized field representing a magnetic field rotating on the (x, y) plane. More concretely, we have

$$\mathbf{B}(t) \;=\; B_0\,\mathbf{z} + B_1\big(\mathbf{x}\cos\omega t - \mathbf{y}\sin\omega t\big)\,. \tag{6.23}$$

Typically, the constant z-component B_0 is larger than B_1, the magnitude of the RF (radio-frequency) signal. The time dependent part of the field points along the x axis at $t = 0$ and is rotating with angular velocity ω in the clockwise direction of the (x, y) plane. The spin Hamiltonian is

$$H_S(t) \;=\; -\gamma\,\mathbf{B}(t)\cdot\mathbf{S} \;=\; -\gamma\Big(B_0\hat{S}_z + B_1\big(\cos\omega t\,\hat{S}_x - \sin\omega t\hat{S}_y\big)\Big)\,. \tag{6.24}$$

Not only is this Hamiltonian time dependent, but the Hamiltonian at different times do not commute. So this is a nontrivial time evolution problem!

We attempt to simplify the problem by considering a frame of reference that rotates with the magnetic field. For this, imagine first the case when $H_S = 0$ because the magnetic field is zero. With no magnetic field spin states would simply be static. What would the Hamiltonian be in the frame rotating about the z-axis with angular frequency ω, just like the magnetic field above? In that frame, the spin states that are fixed in the original frame would be seen to rotate with positive angular velocity ω about the z direction. There must be a Hamiltonian that does have that effect. Since the unitary operator $\mathcal{U}$ that generates this rotation is

$$\mathcal{U}(t) \;=\; \exp\Big(-\frac{i\omega t\hat{S}_z}{\hbar}\Big) \quad\to\quad H_{\mathcal{U}} \;=\; \omega\,\hat{S}_z\,, \tag{6.25}$$

where the expression $H_{\mathcal{U}}$ for the Hamiltonian in the rotating frame is read from the relation $\mathcal{U} = \exp(-iH_{\mathcal{U}}t/\hbar)$. For the original case, when the original Hamiltonian in the static frame is H_S, we will use the above operator $\mathcal{U}$ to define a new rotating-frame state $|\Psi_R\rangle$ as follows

$$|\Psi_R, t\rangle \;\equiv\; \mathcal{U}(t)|\Psi, t\rangle\,. \tag{6.26}$$

If we knew $|\Psi_R, t\rangle$ we would know $|\Psi, t\rangle$. We wish to find out if the Schrödinger equation for $|\Psi_R\rangle$ becomes simpler. For this we must determine the corresponding Hamiltonian H_R. One quick way to do this is to note that the above equation implies that

$$|\Psi_R, t\rangle \;\equiv\; \mathcal{U}(t)\mathcal{U}_S(t)|\Psi, 0\rangle\,. \tag{6.27}$$

where we have introduced the unitary operator $\mathcal{U}_S(t)$ associated with the Hamiltonian $H_S(t)$ that evolves $|\Psi\rangle$ in time. Since the Hamiltonian associated to an arbitrary unitary time-evolution operator $\mathcal{U}$ is $i\hbar(\partial_t\mathcal{U})\mathcal{U}^\dagger$ (if you don't recall this, derive it from a Schrödinger equation)

we have

$$\begin{aligned} H_R &= i\hbar\,\partial_t(\mathcal{U}\mathcal{U}_S)\,\mathcal{U}_S^\dagger\mathcal{U}^\dagger \\ &= i\hbar\,(\partial_t\mathcal{U})\mathcal{U}^\dagger + \mathcal{U}\; i\hbar\,(\partial_t\mathcal{U}_S)\mathcal{U}_S^\dagger\mathcal{U}^\dagger \\ \rightarrow \quad H_R &= H_{\mathcal{U}} + \mathcal{U}H_S\mathcal{U}^\dagger\,, \end{aligned} \tag{6.28}$$

where $H_{\mathcal{U}}$ is the Hamiltonian associated to $\mathcal{U}$. Since $\mathcal{U}$ is the one above, we have $H_{\mathcal{U}} = \omega\hat{S}_z$ and therefore,

$$\begin{aligned} H_R &= \omega\hat{S}_z + \exp\Big(-\frac{i\omega t\hat{S}_z}{\hbar}\Big)\Big[-\gamma\Big(B_0\hat{S}_z + B_1\big(\cos\omega t\,\hat{S}_x - \sin\omega t\hat{S}_y\big)\Big)\Big]\exp\Big(\frac{i\omega t\hat{S}_z}{\hbar}\Big) \\ &= (-\gamma B_0+\omega)\hat{S}_z - \gamma B_1\;\underbrace{\exp\Big(-\frac{i\omega t\hat{S}_z}{\hbar}\Big)\big(\cos\omega t\,\hat{S}_x - \sin\omega t\hat{S}_y\big)\exp\Big(\frac{i\omega t\hat{S}_z}{\hbar}\Big)}_{\hat{M}(t)} \end{aligned} \tag{6.29}$$

and the big question is what is $\hat{M}(t)$. We can proceed by calculating the time-derivative of $\hat{M}$:

$$\begin{aligned} \partial_t\hat{M} &= e^{-\frac{i\omega t\hat{S}_z}{\hbar}}\Big(-\frac{i\omega}{\hbar}\left[\hat{S}_z\,,\cos\omega t\,\hat{S}_x - \sin\omega t\hat{S}_y\right] + \left[-\omega\sin\omega t\,\hat{S}_x - \omega\cos\omega t\,\hat{S}_y\right]\Big)e^{\frac{i\omega t\hat{S}_z}{\hbar}} \\ &= e^{-\frac{i\omega t\hat{S}_z}{\hbar}}\Big(-\frac{i\omega}{\hbar}\left[i\hbar\cos\omega t\,\hat{S}_y + i\hbar\sin\omega t\hat{S}_x\right] + \left[-\omega\sin\omega t\,\hat{S}_x - \omega\cos\omega t\,\hat{S}_y\right]\Big)e^{\frac{i\omega t\hat{S}_z}{\hbar}} \\ &= e^{-\frac{i\omega t\hat{S}_z}{\hbar}}\Big(\omega\cos\omega t\,\hat{S}_y + \omega\sin\omega t\hat{S}_x\; - \omega\sin\omega t\,\hat{S}_x - \omega\cos\omega t\,\hat{S}_y\Big)e^{\frac{i\omega t\hat{S}_z}{\hbar}} \\ &= 0\,. \end{aligned} \tag{6.30}$$

This is very good news. Since $\hat{M}$ has no time dependence, we can evaluate it at any time. The simplest time is $t = 0$ and we thus find that

$$\hat{M}(t) = \hat{S}_x\,. \tag{6.31}$$

As a result we have a totally time-independent H_R:

$$\begin{aligned} H_R &= (-\gamma B_0+\omega)\hat{S}_z - \gamma B_1\hat{S}_x \\ &= -\gamma\Big[\big(B_0 - \frac{\omega}{\gamma}\big)\hat{S}_z + B_1\hat{S}_x\Big] \\ &= -\gamma\Big[B_0\big(1 - \frac{\omega}{\omega_0}\big)\hat{S}_z + B_1\hat{S}_x\Big], \end{aligned} \tag{6.32}$$

using $\omega_0 = \gamma B_0$ for the Larmor frequency associated with the constant component of the field. We thus have a Hamiltonian H_R that can be associated with a magnetic field $\mathbf{B}_R$ given by Setting

$$\boxed{H_R \;=\; -\gamma\,\mathbf{B}_R\cdot\mathbf{S}\,, \quad \rightarrow \quad \mathbf{B}_R \;=\; B_1\mathbf{x} + B_0\big(1 - \frac{\omega}{\omega_0}\big)\mathbf{z}\,.} \tag{6.33}$$

The full solution for the state is obtained beginning with (6.26) and (6.25):

$$|\Psi, t\rangle \;=\; \mathcal{U}^{\dagger}(t)|\Psi_R, t\rangle = \;\exp\Big[\frac{i\omega t\hat{S}_z}{\hbar}\Big]|\Psi_R, t\rangle\,. \tag{6.34}$$

Since H_R it a time-independent Hamiltonian, we have that the full time evolution is

$$|\Psi, t\rangle \;=\; \exp\Big[\frac{i\omega t\hat{S}_z}{\hbar}\Big]\exp\Big[-i\frac{(-\gamma\mathbf{B}_R\cdot\mathbf{S}\,)\,t}{\hbar}\Big]|\Psi, 0\rangle \tag{6.35}$$

where the solution for $|\Psi_R, t\rangle$ is simply the time evolution from the $t = 0$ state generated by the Hamiltonian H_R. We have thus found that

$$\boxed{|\Psi, t\rangle \;=\; \exp\Big[\frac{i\omega t\hat{S}_z}{\hbar}\Big]\exp\Big[i\frac{\gamma\mathbf{B}_R\cdot\mathbf{S}\,t}{\hbar}\Big]|\Psi, 0\rangle\,.} \tag{6.36}$$

Exercise. Verify that for $B_1 = 0$ the above solution reduces to the one describing precession about the z-axis.

In the applications to be discussed below we always have

$$B_1 \ll B_0\,. \tag{6.37}$$

Consider now the evolution of a spin that initially points in the positive z direction. We look at two cases:

- $\omega \ll \omega_0$. In this case from (6.33) we have

$$\mathbf{B}_R \;\simeq\; B_0\mathbf{z} + B_1\,\mathbf{x}\,. \tag{6.38}$$

  This is a field mostly along the z axis, but tipped a little towards the x axis. The right-most exponential in (6.36) makes the spin precess about the direction $\mathbf{B}_R$ quite quickly, for $|\mathbf{B}_R| \sim B_0$ and thus the angular rate of precession is pretty much ω_0. The next exponential in (6.36) induces a rotation about the z-axis with smaller angular velocity ω. This is shown in the figure.

- $\omega = \omega_0$. In this case from (6.33) we have

$$\mathbf{B}_R \;=\; B_1\,\mathbf{x}\,. \tag{6.39}$$

  In this case the right-most exponential in (6.36) makes the spin precess about the x axis it go from the z axis towards the y axis with angular velocity $\omega_1 = \gamma B_1$. If we time the RF signal to last a time T such that

$$\omega_1 T = \frac{\pi}{2} \quad\to\quad T = \frac{\pi}{2\gamma B_1}\,, \tag{6.40}$$

the state $|\Psi_R, T\rangle$ points along the y axis. The effect of the other exponential in (6.36) is just to rotate the spin on the (x, y) plane. We have

$$|\Psi, t\rangle \;=\; \exp\Big[\frac{i\omega t \hat{S}_z}{\hbar}\Big]|\Psi_R, t\rangle\,, \quad t < T, \tag{6.41}$$

and if the RF pulse turns off after time T,

$$|\Psi, t\rangle \;=\; \exp\Big[\frac{i\omega t \hat{S}_z}{\hbar}\Big]|\Psi_R, T\rangle\,, \quad t > T, \tag{6.42}$$

The state $|\Psi, t\rangle$ can be visualized as a spin that is slowly rotating with angular velocity ω_1 from the z axis towards the plane, while rapidly rotating around the z axis with angular velocity ω. As a result the tip of the spin is performing a spiral motion on the surface of a hemisphere. By the time the polar angle reaches $\pi/2$ the RF signal turns off and the spin now just rotates within the (x, y) plane. This is called a 90° pulse.

Magnetic Resonance Imaging Based on work on nuclear magnetic resonance by Edward Purcell, who worked at MIT's radiation laboratory, and Felix Bloch. They got the Nobel prize for this work in 1952.

The NMR work led to the development of the technique called MRI for Magnetic Resonance Imaging. First studies in humans was done in 1977. The new perspective, as compared to X-rays, was that MRI allows one to distinguish various soft tissues.

The human body is mostly composed of water molecules. In those we have many hydrogen atoms, whose nuclei are protons and are the main players through their magnetic dipole moments (nuclear dipoles).

With a given external and large constant magnetic field B_0, at a given temperature, there is a net alignment of nuclear spins along B_0. This is called the "longitudinal magnetization". For all intents and purposes we have a net number of spins in play.

We apply a 90° pulse so that we get the spins to rotate with Larmor frequency ω_0 in the (x, y) plane. These rotating dipoles produce an oscillating magnetic field which is a signal that can be picked up by a receiver. The magnitude of the signal is proportional to the proton density. This is the first piece of information and allows differentiation of tissues.

The above signal from the rotation of the spins decays with a time constant T_2 that can be measured and is typically much smaller than a second. This decay is attributed to interactions between the spins. A T_2 weighted image allows doctors to detect abnormal accumulation of fluids (edema).

There is another time constant T_1 (of order one second) that controls the time to regain the longitudinal magnetization. This effect is due to the spins interacting with the rest of

the lattice of atoms. White matter, grey matter, and cerebrospinal fluids are distinguished by different T_1 constants (they have about the same proton density).

MRI's commonly include the use of contrast agents, which are substances that shorten the time constant T_1 and are usually administered by injection into the blood stream. The contrast agent (gadolinium) can accumulate at organs or locations where information is valuable. For a number of substances one can use the MRI apparatus to determine their (T_1, T_2) constants and build a table of data. This table can then be used as an aid to identify the results of MRI's.

The typical MRI machine has a B_0 of about 2T (two tesla or 20,000 gauss). This requires a superconducting magnet with cooling by liquid helium. For people with claustrophobia there are "open" MRI scanners that work with lower magnetic fields. In addition there are about three *gradient* magnets, each of about 200 gauss. They change locally the value of B_0 and thus provide spatial resolution as the Larmor frequency becomes spatially dependent. One can thus attain spatial resolutions of about half a millimiter! MRI's are considered safe, as there is no evidence of biological harm caused by very large static magnetic fields.

MULTIPARTICLE STATES AND TENSOR PRODUCTS

B. Zwiebach

December 14, 2013

Contents

1 Introduction to the Tensor Product

In this section, we develop the tools needed to describe a system that contains more than one particle. Most of the required ideas appear when we consider systems with two particles. We will assume the particles are distinguishable; for indistinguishable particles quantum mechanics imposes some additional constraints on the allowed set of states. We will study those constraints later in the course (or in 8.06!) The tools we are about to develop will be needed to understand addition of angular momenta. In that problem one is adding the angular momenta of the two or more particles in the system.

Consider then two particles. Below is a description of the quantum mechanics and family of operators associated with each particle:

- Particle 1: its quantum mechanics is described by a complex vector space V. It has associated operators $T_1, T_2, ...$
- Particle 2: its quantum mechanics is described by a complex vector space W. It has associated operators $S_1, S_2, ...$

This list of operators for each particle may include some or many of the operators you are already familiar with: position, momentum, spin, Hamiltonians, projectors, etc.

Once we have two particles, the two of them together form our system. We are after the description of quantum states of this two-particle system. On first thought, we may think that any state of this system should be described by giving the state $v \in V$ of the first particle and the state $w \in W$ of the second particle. This information could be represented by the ordered list (v, w) where the first item

is the state of the first particle and the second item the state of the second particle. This *is* a state of the two-particle system, but it is far from being the general state of the two-particle system. It misses remarkable new possibilities, as we shall soon see.

We thus introduce a new notation. Instead of representing the state of the two-particle system with particle one in v and particle two in w as (v, w), we will represent it as $v \otimes w$. This element $v \otimes w$ will be viewed as a vector in a new vector space $V \otimes W$ that will carry the description of the quantum states of the system of two particles. This $\otimes$ operation is called the "tensor product." In this case we have two vector spaces over $\mathbb{C}$ and the tensor product $V \otimes W$ is a new complex vector space:

$$v \otimes w \ \in \ V \otimes W \quad \text{when} \quad v \in V,\ w \in W\,. \tag{1.1}$$

In $v \otimes w$ there is no multiplication to be carried out, we are just placing one vector to the left of $\otimes$ and another to the right of $\otimes$.

We have only described some elements of $V \otimes W$, not quite given its definition yet.[1] We now explain two physically motivated rules that define the tensor product completely.

1. If the vector representing the state of the first particle is scaled by a complex number this is equivalent to scaling the state of the two particles. The same for the second particle. So we declare

$$\boxed{(av) \otimes w \ = \ v \otimes (aw) \ = \ a\ (v \otimes w), \qquad a \in \mathbb{C}\,.} \tag{1.2}$$

2. If the state of the first particle is a superposition of two states, the state of the two-particle system is also a superposition. We thus demand distributive properties for the tensor product:

$$\boxed{\begin{aligned} (v_1 + v_2) \otimes w \ &= v_1 \otimes w + v_2 \otimes w\,, \\ v \otimes (w_1 + w_2) \ &= v \otimes w_1 + v \otimes w_2\,. \end{aligned}} \tag{1.3}$$

The tensor product $V \otimes W$ is thus defined to be the vector space whose elements are (complex) linear combinations of elements of the form $v \otimes w$, with $v \in V, w \in W$, with the above rules for manipulation. The tensor product $V \otimes W$ is the complex vector space of states of the two-particle system!

Comments

1. The vector $0 \in V \otimes W$ is equal to $0 \otimes w$ or $v \otimes 0$. Indeed, by the first property above, with $a = 0$, we have $av = 0$ (rhs a vector) and $0 \otimes w = 0(0 \otimes w) = 0$

2. Let $v_1, v_2 \in V$ and $w_1, w_2 \in W$. A vector in $V \otimes W$ constructed by superposition is

$$\alpha_1(v_1 \otimes w_1) + \alpha_2(v_2 \otimes w_2) \in V \otimes W \tag{1.4}$$

[1] If we just left it like this, we would have defined the direct product of vector spaces.

This shows clearly that a general state of the two-particle system cannot be described by stating the state of the first particle and the state of the second particle. The above superpositions give rise to entangled states. An entangled state of the two particles is one that, roughly, cannot be disentangled into separate states of each of the particles. We will make this precise soon.

If $(e_1, \ldots, e_n)$ is a basis of V and $(f_1, \ldots, f_m)$ is a basis of W, then the set of elements $e_i \otimes f_j$ where $i = 1, \ldots, n$ and $f = 1, \ldots, m$ forms a basis for $V \otimes W$. It is simple to see these span the space since for any $v \otimes w$ we have $v = \sum_i v_i e_i$ and $w = \sum_j w_j f_j$ so that

$$v \otimes w \;=\; \big(\sum_i v_i e_i\big) \otimes \big(\sum_j w_j f_j\big) \;=\; \sum_{i,j} v_i w_j \, e_i \otimes f_j \,. \tag{1.5}$$

Given this, we see that the basis also spans linear superpositions of elements of the form $v \otimes w$, thus general elements of $V \otimes W$. With $n \cdot m$ basis vectors, the dimensionality of $V \otimes W$ is equal to the *product* of the dimensionalities of V and W:

$$\dim(V \otimes W) = \dim(V) \times \dim(W) \,. \tag{1.6}$$

Dimensions are multiplied (not added) in a tensor product.

How do we construct operators that act in the vector space $V \otimes W$? Let T be an operator in V and S be an operator in W. In other words, $T \in \mathcal{L}(V)$ and $S \in \mathcal{L}(W)$. We can then construct an operator $T \otimes S$

$$T \otimes S \;\in\; \mathcal{L}(V \otimes W) \tag{1.7}$$

defined to act as follows:

$$T \otimes S \;(v \otimes w) \;\equiv\; Tv \otimes \, Sw \,. \tag{1.8}$$

This is the only 'natural' option: we let T act on the vector it knows how to act, and S act on the vector it knows how to act.

Suppose that we want the operator $T \in \mathcal{L}(V)$ that acts on the first particle to act on the tensor product $V \otimes W$, even though we have not supplied an operator S to act on the W part. For this we upgrade the operator from one that acts on a single vector space to one, given by $T \otimes 1$, that acts on the tensor product:

$$T \in \mathcal{L}(V) \quad \to \quad T \otimes \mathbf{1} \in \mathcal{L}(V \otimes W) \,, \qquad T \otimes \mathbf{1} \,(v \otimes w) \equiv Tv \otimes w \,. \tag{1.9}$$

Similarly, an operator S belonging to $\mathcal{L}(W)$ is upgraded to $\mathbf{1} \otimes S$ to act on the tensor product. A basic result is that upgraded operators of the first particle **commute** with upgraded operators of the second particle. Indeed,

$$\begin{aligned} (T \otimes \mathbf{1}) \cdot (\mathbf{1} \otimes S) \;(v \otimes w) \;&=\; (T \otimes \mathbf{1})(v \otimes Sw) \;=\; Tv \otimes Sw \\ (1 \otimes S) \cdot (T \otimes \mathbf{1}) \;(v \otimes w) \;&= (\mathbf{1} \otimes S) \;(Tv \otimes w) \;=\; Tv \otimes Sw \,. \end{aligned} \tag{1.10}$$

and therefore

$$\big[\, T \otimes \mathbf{1}\,,\, \mathbf{1} \otimes S \,\big] \;=\; 0\,. \tag{1.11}$$

Given a system of two particles we can construct a simple total Hamiltonian H_T (describing no interactions) by upgrading each of the Hamiltonians H_1 and H_2 and adding them:

$$H_T \;\equiv\; H_1 \otimes 1 + 1 \otimes H_2 \tag{1.12}$$

Exercise. Convince yourself that

$$\exp\Big(-\frac{iH_T t}{\hbar}\Big) \;=\; \exp\Big(-\frac{iH_1 t}{\hbar}\Big) \otimes \exp\Big(-\frac{iH_2 t}{\hbar}\Big) \tag{1.13}$$

We turn now to a famous example at the basis of adding angular momenta.

Example 1: We have two spin-1/2 particles, and describe the first's state space V_1 with basis states $|+\rangle_1$ and $|-\rangle_1$ and the second's state space V_2 with basis states $|+\rangle_2$ and $|-\rangle_2$. The tensor product $V_1 \otimes V_2$ has four basis vectors:

$$|+\rangle_1 \otimes |+\rangle_2; \quad |+\rangle_1 \otimes |-\rangle_2; \quad |-\rangle_1 \otimes |+\rangle_2; \quad |-\rangle_1 \otimes |-\rangle_2 \tag{1.14}$$

If we follow the convention that the first ket corresponds to particle one and the second ket corresponds to particle two, the notation is simpler. The most general state of the two-particle system is a linear superposition of the four basis states:

$$|\Psi\rangle \;=\; \alpha_1|+\rangle_1 \otimes |+\rangle_2 \;+\; \alpha_2|+\rangle_1 \otimes |-\rangle_2 \;+\; \alpha_3|-\rangle_1 \otimes |+\rangle_2 \;+\; \alpha_4|-\rangle_1 \otimes |-\rangle_2\,. \tag{1.15}$$

Example 2: We now want to act on this state with the *total* z-component of angular momentum. Naively, this would be the sum of the z-components of each individual particle. However, we know better at this point - summing the two angular momenta really means constructing a new operator in the tensor product vector space:

$$S_z^T \;=\; S_z^{(1)} \otimes \mathbf{1} \;+\; \mathbf{1} \otimes S_z^{(2)}\,. \tag{1.16}$$

Performing the calculation in two parts,

$$\begin{aligned}
(S_z^{(1)} \otimes \mathbf{1})|\Psi\rangle \;&=\; \alpha_1 S_z|+\rangle_1 \otimes |+\rangle_2 \,+\, \alpha_2 S_z|+\rangle_1 \otimes |-\rangle_2 \,+\, \alpha_3 S_z|-\rangle_1 \otimes |+\rangle_2 \,+\, \alpha_4 S_z|-\rangle_1 \otimes |-\rangle_2 \\
&=\; \frac{\hbar}{2}\Big(\alpha_1|+\rangle_1 \otimes |+\rangle_2 \,+\, \alpha_2|+\rangle_1 \otimes |-\rangle_2 \,-\, \alpha_3|-\rangle_1 \otimes |+\rangle_2 \,-\, \alpha_4|-\rangle_1 \otimes |-\rangle_2\Big) \\
(1 \otimes S_z^{(2)})|\Psi\rangle \;&=\; \alpha_1|+\rangle_1 \otimes S_z|+\rangle_2 \,+\, \alpha_2|+\rangle_1 \otimes S_z|-\rangle_2 \,+\, \alpha_3|-\rangle_1 \otimes S_z|+\rangle_2 \,+\, \alpha_4|-\rangle_1 \otimes S_z|-\rangle_2 \\
&=\; \frac{\hbar}{2}\Big(\alpha_1|+\rangle_1 \otimes |+\rangle_2 \,-\, \alpha_2|+\rangle_1 \otimes |-\rangle_2 \,+\, \alpha_3|-\rangle_1 \otimes |+\rangle_2 \,-\, \alpha_4|-\rangle_1 \otimes |-\rangle_2\Big)
\end{aligned} \tag{1.17}$$

Adding these together, we have:

$$S_z^T|\Psi\rangle = \hbar\left(\alpha_1|+\rangle_1\otimes|+\rangle_2 \;-\; \alpha_4|-\rangle_1\otimes|-\rangle_2\right) \tag{1.18}$$

One can derive this result quickly by noting that since $S_z^{(1)}$ is diagonal in the first basis and $S_z^{(2)}$ is diagonal in the second basis, the total S_z is diagonal in the tensor space basis and its eigenvalue acting on a tensor state is the sum of the S_z eigenvalues for particle one and particle two. Thus,

$$\begin{aligned}
S_z^T|+\rangle\otimes|+\rangle &= \Big(\frac{\hbar}{2}+\frac{\hbar}{2}\Big)|+\rangle\otimes|+\rangle \;=\; \hbar|+\rangle\otimes|+\rangle \\
S_z^T|+\rangle\otimes|-\rangle &= \Big(\frac{\hbar}{2}-\frac{\hbar}{2}\Big)|+\rangle\otimes|+\rangle \;=\; 0 \\
S_z^T|-\rangle\otimes|+\rangle &= \Big(-\frac{\hbar}{2}+\frac{\hbar}{2}\Big)|+\rangle\otimes|+\rangle \;=\; 0 \\
S_z^T|-\rangle\otimes|-\rangle &= \Big(-\frac{\hbar}{2}-\frac{\hbar}{2}\Big)|-\rangle\otimes|-\rangle \;=\; -\hbar|-\rangle\otimes|-\rangle
\end{aligned} \tag{1.19}$$

The result in (1.18) follows quickly from the four relations above. Suppose we are only interested in states that have zero S_z^T. This requires

$$\alpha_1=\alpha_4=0 \quad\rightarrow\quad |\Psi\rangle=\alpha_2|+\rangle\otimes|-\rangle+\alpha_3|-\rangle\otimes|+\rangle \tag{1.20}$$

Example 3: Calculate the total x-component S_x^T of spin angular momentum on the above states with zero S_z^T. Recalling that

$$S_x|+\rangle=\frac{\hbar}{2}|-\rangle\,,\qquad S_x|-\rangle=\frac{\hbar}{2}|+\rangle \tag{1.21}$$

and writing

$$S_x^T = S_x\otimes 1 + 1\otimes S_x \tag{1.22}$$

the calculation proceeds as follows:

$$\begin{aligned}
S_x^T|+\rangle\otimes|-\rangle &= S_x|+\rangle\otimes|-\rangle \;+\; |+\rangle\otimes S_x|-\rangle \;=\; \frac{\hbar}{2}\big(|-\rangle\otimes|-\rangle \;+\; |+\rangle\otimes|+\rangle\big) \\
S_x^T|-\rangle\otimes|+\rangle &= S_x|-\rangle\otimes|+\rangle \;+\; |-\rangle\otimes S_x|+\rangle \;=\; \frac{\hbar}{2}\big(|+\rangle\otimes|+\rangle \;+\; |-\rangle\otimes|-\rangle\big)
\end{aligned} \tag{1.23}$$

Therefore

$$\begin{aligned}
S_x^T|\Psi\rangle &= \alpha_2\frac{\hbar}{2}\big(|-\rangle\otimes|-\rangle+|+\rangle\otimes|+\rangle\big)+\alpha_3\frac{\hbar}{2}\big(|+\rangle\otimes|+\rangle+|-\rangle\otimes|-\rangle\big) \\
&= \frac{\hbar}{2}(\alpha_2+\alpha_3)\big(|+\rangle\otimes|+\rangle+|-\rangle\otimes|-\rangle\big)
\end{aligned} \tag{1.24}$$

If we demand that S_x^T also be zero on the state we now find $\alpha_2=-\alpha_3$. Thus, the following state has zero S_x^T, S_z^T:

$$|\Psi\rangle \;=\; \alpha\left(|+\rangle\otimes|-\rangle-|-\rangle\otimes|+\rangle\right). \tag{1.25}$$

Exercise: Verify that $S_y^T|\Psi\rangle=0$. Thus we say that the state has total spin angular momentum zero.

We now consider the definition of an **inner product** in $V \otimes W$. To do this we simply give state how the most general inner product is computed using a basis $\{e_i \otimes f_j\}$ for the tensor product, with $\{e_i\}$ and $\{f_i\}$ *orthonormal* bases for V and W. We begin by declaring that

$$\langle e_i \otimes f_j \, , \, e_p \otimes f_q \rangle \, \equiv \, \delta_{ip}\delta_{jq} \, . \tag{1.26}$$

This makes the basis $\{e_i \otimes f_j\}$ orthonormal. In addition, we must declare that with vectors $X, Y, Z \in V \otimes W$ and a complex constant a the following axioms hold:

$$\begin{aligned} \langle X + Y \, , \, Z \rangle \; &= \; \langle X, \, Z \rangle + \langle Y, \, Z \rangle \, , \\ \langle X \, , Y + Z \rangle \; &= \; \langle X \, , Y \rangle + \langle X \, , Z \rangle \, , \\ \langle X, aY \rangle \; &= \; a \langle X, Y \rangle \\ \langle aX, Y \rangle \; &= \; a^* \langle X, Y \rangle \end{aligned} \tag{1.27}$$

This is a complete definition of the inner product in the tensor space: we can compute the inner product of any two vectors in $V \otimes W$ using the chosen basis and the above distributive rules. Indeed, using these properties we can show that

$$\langle v \otimes w \, , \, \tilde{v} \otimes \tilde{w} \rangle \, = \, \langle v \, , \, \tilde{v} \rangle \, \langle w \, , \tilde{w} \rangle \, , \tag{1.28}$$

where the inner products on the right-hand side are those in V and in W, making it clear that the inner product in $V \otimes W$ arises from the inner products in V and W. To prove this relation we begin by writing

$$\begin{aligned} v \; &= \; \sum_i v_i e_i \, , \quad w \; = \; \sum_j w_j f_j \, , \\ \tilde{v} \; &= \; \sum_p \tilde{v}_p e_p \, , \quad \tilde{w} \; = \; \sum_q \tilde{w}_q f_q \, . \end{aligned} \tag{1.29}$$

Since the basis vectors in V and W are orthonormal we find that

$$\langle v, \tilde{v} \rangle \; = \; \sum_i v_i^* \tilde{v}_i \, , \quad \langle w, \tilde{w} \rangle \; = \; \sum_j w_j^* \tilde{w}_j \, . \tag{1.30}$$

Now evaluating the left-hand side of (1.28)

$$\begin{aligned} \big\langle v \otimes w \, , \, \tilde{v} \otimes \tilde{w} \big\rangle \; &= \; \Big\langle \sum_i v_i e_i \otimes \sum_j w_j f_j \, , \, \sum_p \tilde{v}_p e_p \otimes \sum_q \tilde{w}_q f_q \Big\rangle \\ &= \; \sum_{i,j,p,q} \Big\langle v_i w_j \, e_i \otimes f_j \, , \, \tilde{v}_p \tilde{w}_q \, e_p \otimes f_q \Big\rangle \\ &= \; \sum_{i,j,p,q} v_i^* w_j^* \, \tilde{v}_p \tilde{w}_q \big\langle \, e_i \otimes f_j \, , \, e_p \otimes f_q \big\rangle \\ &= \; \sum_{i,j,p,q} v_i^* w_j^* \, \tilde{v}_p \tilde{w}_q \, \delta_{ip} \, \delta_{jq} \; = \; \sum_i v_i^* \tilde{v}_i \sum_j w_j^* \, \tilde{w}_j \\ &= \langle v, \tilde{v} \rangle \, \langle w, \, \tilde{w} \rangle \, . \end{aligned} \tag{1.31}$$

The verification that the inner-product on $V \otimes W$ satisfies the remaining axioms is left as a good practice for you. Assume below that $X, Y \in V \otimes W$. For both exercises above simply write the most general vector, as $X = \sum_{ij} x_{ij}\, e_i \otimes f_j$ and proceed.
Exercise: Show that $\langle X, X\rangle \geq 0$, and $\langle X, X\rangle = 0$ if and only if $X = 0$.
Exercise: Show that $\langle X, Y\rangle = \langle Y, X\rangle^*$.

Many times it is convenient to use bra-ket notation for inner products in the tensor product. We write

$$\begin{aligned} |v \otimes w\rangle &= |v\rangle_1 \otimes |w\rangle_2 \\ \langle v \otimes w| &= {}_1\langle v| \otimes {}_2\langle w| \,. \end{aligned} \tag{1.32}$$

Notice that both on bras and kets we write the state of particle one to the left of the state of particle two. We then write (1.28) as

$$\langle v \otimes w|\tilde{v} \otimes \tilde{w}\rangle \;=\; \big({}_1\langle v| \otimes {}_2\langle w|\big)\,\big(|\tilde{v}\rangle_1 \otimes |\tilde{w}\rangle_2\,\big) = \langle v|\tilde{v}\rangle\,\langle w|\tilde{w}\rangle\,. \tag{1.33}$$

Back to our example with spin states, our four basis vectors $|+\rangle_1 \otimes |+\rangle_2$, $|+\rangle_1 \otimes |-\rangle_2$, $|-\rangle_1 \otimes |+\rangle_2$, and $|-\rangle_1 \otimes |-\rangle_2$ are orthonormal. We had the un-normalized state in (1.25) given by

$$|\Psi\rangle = \alpha\,\Big(\,|+\rangle_1 \otimes |-\rangle_2 \;-\; |-\rangle_1 \otimes |+\rangle_2\,\Big)\,. \tag{1.34}$$

The associated bra is then

$$\langle\Psi| = \alpha^*\,\Big(\,{}_1\langle +| \otimes {}_2\langle -| \;-\; {}_1\langle -| \otimes {}_2\langle +|\,\Big)\,. \tag{1.35}$$

We then have

$$\begin{aligned} \langle\Psi|\Psi\rangle &= \alpha\alpha^*\,\Big({}_1\langle +| \otimes {}_2\langle -| \;-\; {}_1\langle -| \otimes {}_2\langle +|\Big)\Big(|+\rangle_1 \otimes |-\rangle_2 \;-\; |-\rangle_1 \otimes |+\rangle_2\,\Big) \\ &= \alpha\alpha^*\,\Big({}_1\langle +| \otimes {}_2\langle -||+\rangle_1 \otimes |-\rangle_2 \;+\; {}_1\langle -| \otimes {}_2\langle +|\,|-\rangle_1 \otimes |+\rangle_2\Big) \end{aligned} \tag{1.36}$$

since only terms where the spin states are the same for the first particle and for the second particle survive. We thus have, for normalization,

$$\langle\Psi|\Psi\rangle \;=\; |\alpha|^2(1+1) = 2|\alpha|^2 = 1\,, \quad \to \quad \alpha = \frac{1}{\sqrt{2}}\,. \tag{1.37}$$

The normalized state with zero total angular momentum is then

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\,\Big(|+\rangle_1 \otimes |-\rangle_2 - |-\rangle_1 \otimes |+\rangle_2\,\Big)\,. \tag{1.38}$$

2 Entangled States

You have learned that $V \otimes W$ includes states $\Psi = \sum_i \alpha_i \, v_i \otimes w_i$ obtained by linear superposition of simpler states of the form $v_i \otimes w_i$. If handed such a Ψ, you might want to know whether you can write it as a single term $v_* \otimes w_*$ for some $v_* \in V$ and $w_* \in W$. If so, you are able to describe the state of the particles in Ψ independently: particle one is in state v_* and particle two in state w_*. We then say that in the state Ψ the particles are *not entangled*. If no such v_* and w_* exist, we say that in the state $\Psi \in V \otimes W$ the particles are entangled or equivalently, that Ψ is an entangled stated of the two particles. Entanglement is a basis-independent property.

It is simplest to illustrate this using two-dimensional complex vector spaces V and W, like the ones we use for spin one-half. Let V have a basis e_1, e_2 and W have a basis f_1, f_2. Then, the most general state you can write is the following:

$$\Psi_A \;=\; a_{11}\, e_1 \otimes f_1 + a_{12}\, e_1 \otimes f_2 + a_{21}\, e_2 \otimes f_1 + a_{22}\, e_2 \otimes f_2 \,. \tag{2.39}$$

This state is encoded by a matrix A of coefficients

$$A \;=\; \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} . \tag{2.40}$$

The state is *not entagled* if there exist constants a_1, a_2, b_1, b_2 such that

$$a_{11}\, e_1 \otimes f_1 + a_{12}\, e_1 \otimes f_2 + a_{21}\, e_2 \otimes f_1 + a_{22}\, e_2 \otimes f_2 \;=\; (a_1 e_1 + a_2 e_2) \otimes (b_1 f_1 + b_2 f_2) \,. \tag{2.41}$$

Note that these four unknown constants are not uniquely determined: we can, for example, multiply a_1 and a_2 by some constant $c \neq 0$ and divide b_1 and b_2 by c, to obtain a different solution. Indeed $v \otimes w = (cv) \otimes (w/c)$ for any $c \neq 0$. Using the distributive laws for $\otimes$ to expand the right-hand side of (2.41) and recalling that $e_i \otimes f_j$ are basis vectors in the tensor product, we see that the equality requires the following four relations:

$$\begin{aligned} a_{11} &= a_1 b_1 \\ a_{12} &= a_1 b_2 \\ a_{21} &= a_2 b_1 \\ a_{22} &= a_2 b_2 \end{aligned} \tag{2.42}$$

Combining these four expressions leaves us with a consistency condition:

$$a_{11}a_{22} - a_{12}a_{21} = a_1 b_1 a_2 b_2 - a_1 b_2 a_2 b_1 \;=\; 0 \quad \rightarrow \quad \det A = 0 \,. \tag{2.43}$$

In other words, if Ψ_A is not entangled the determinant of the matrix A must be zero. We can in fact show that $\det A = 0$ implies that Ψ_A is not entangled. To do this we simply have to present a solution for the equations above under the condition $\det A = 0$.

Assume first that $a_{11} = 0$. Then $\det A = 0$ implies $a_{12}a_{21} = 0$. If $a_{12} = 0$ then

$$\Psi_A = a_{21} e_2 \otimes f_1 + a_{22} e_2 \otimes f_2 = e_2 \otimes (a_{21} f_1 + a_{22} f_2) \tag{2.44}$$

and the state is indeed not entangled. If $a_{21} = 0$ then

$$\Psi_A = a_{12} e_1 \otimes f_2 + a_{22} e_2 \otimes f_2 = (a_{12} e_1 + a_{22} e_2) \otimes f_2 \tag{2.45}$$

and again, the state is not entangled. Thus, we can solve all equations when $a_{11} = 0$. Now assuming $a_{11} \neq 0$ we can take

$$a_1 = \sqrt{a_{11}}\,, \qquad b_1 = \sqrt{a_{11}}\,, \tag{2.46}$$

to solve the first equation in (2.42). The second and third equations allow us to solve for b_2 and a_2

$$b_2 = \frac{a_{12}}{\sqrt{a_{11}}}\,, \quad a_2 = \frac{a_{21}}{\sqrt{a_{11}}} \tag{2.47}$$

The fourth equation is then automatically satisfied as

$$a_2 b_2 \;=\; \frac{a_{12}a_{21}}{a_{11}} \;=\; \frac{a_{11}a_{22}}{a_{11}} \;=\; a_{22} \tag{2.48}$$

using the vanishing determinant condition. We have thus solved the system of equations and we can write

$$\Psi_A \;=\; \left(\sqrt{a_{11}} e_1 + \frac{a_{21}}{\sqrt{a_{11}}} e_2\right) \otimes \left(\sqrt{a_{11}} f_1 + \frac{a_{12}}{\sqrt{a_{11}}} f_2\right) \quad \text{if} \quad \det A = 0\,. \tag{2.49}$$

We have thus proved that Ψ_A is entangled if and only if $\det A \neq 0$. For vector spaces of dimensions different than two the conditions for entanglement take a different form. Schrödinger called "entanglement" the essential feature of quantum mechanics.

Example: Consider our state of zero total spin angular momentum:

$$|\Phi\rangle_A \equiv \frac{1}{\sqrt{2}}\Big(|+\rangle_1 \otimes |+\rangle_2 \;-\; |-\rangle_1 \otimes |-\rangle_2\Big) \tag{2.50}$$

If we have the basis vectors $|e_1\rangle = |+\rangle_1, |e_2\rangle = |-\rangle_1$ and $|f_1\rangle = |+\rangle_2, |f_2\rangle = |-\rangle_2$ we see that the state is described by the matrix

$$A \;=\; \begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & -1/\sqrt{2} \end{pmatrix} \tag{2.51}$$

Since the determinant of this matrix is not zero, the state is entangled.

3 Bell basis states

Bell states are a set of entangled basis vectors. Take $V_1 \otimes V_2$, with V_1 and V_2 both the two-dimensional complex vector space of spin-1/2 particles. For brevity of notation we will leave out the 1 and 2

subscripts on the states and the $\otimes$ in between the states; it is always understood that in $V_1 \otimes V_2$ the state of V_1 appears to the left of the state of V_2. Consider now the state

$$|\Phi_0\rangle \;\equiv\; \frac{1}{\sqrt{2}}\Big(|+\rangle|+\rangle \,+\, |-\rangle|-\rangle\Big)\,. \tag{3.52}$$

This is clearly an entangled state: its associated matrix is diagonal with equal entries of $1/\sqrt{2}$ and thus non-zero determinant. Moreover this state is unit normalized

$$\langle\Phi_0|\Phi_0\rangle = 1\,. \tag{3.53}$$

We can use this state as the first of our basis vectors for $V_1 \otimes V_2$. Since this tensor product is four-dimensional we need three more entangled basis states. Here they are:

$$|\Phi_i\rangle \;\equiv\; (\mathbf{1}\otimes\sigma_i)|\Phi_0\rangle\,,\;\; i=1,2,3. \tag{3.54}$$

We will explicitly see below that these states are entangled, but this property is clear from the definition. If $|\Psi_i\rangle$ is not entangled, it would follow that that $\mathbf{1}\otimes\sigma_i|\Psi_i\rangle$ (i not summed) is not entangled either (do you see why?). But using $\sigma_i^2 = 1$, we see that this last state is in fact $|\Phi_0\rangle$, which is entangled. This contradiction shows that $|\Phi_i\rangle$ must be entangled. It is also manifest from the definition that the $|\Phi_i\rangle$ states are unit normalized.

Let us look at the form of $|\Phi_1\rangle$:

$$\begin{aligned} |\Phi_1\rangle \;&=\; (\mathbf{1}\otimes\sigma_1)\frac{1}{\sqrt{2}}\Big(|+\rangle|+\rangle \,+\, |-\rangle|-\rangle\Big) \;=\; \frac{1}{\sqrt{2}}\Big(|+\rangle\sigma_1|+\rangle \,+\, |-\rangle\sigma_1|-\rangle\Big) \\ &=\; \frac{1}{\sqrt{2}}\Big(|+\rangle|-\rangle \,+\, |-\rangle|+\rangle\Big)\,. \end{aligned} \tag{3.55}$$

The state is clearly entangled. By analogous calculations we obtain the full list of Bell states

$$\begin{aligned} |\Phi_0\rangle \;&=\; \mathbf{1}\otimes\mathbf{1}\,|\Phi_0\rangle \;=\; \frac{1}{\sqrt{2}}\Big(|+\rangle|+\rangle + |-\rangle|-\rangle\Big) \\ |\Phi_1\rangle \;&=\; \mathbf{1}\otimes\sigma_1|\Phi_0\rangle \;=\; \frac{1}{\sqrt{2}}\Big(|+\rangle|-\rangle + |-\rangle|+\rangle\Big) \\ |\Phi_2\rangle \;&=\; \mathbf{1}\otimes\sigma_2|\Phi_0\rangle \;=\; \frac{i}{\sqrt{2}}\Big(|+\rangle|-\rangle - |-\rangle|+\rangle\Big) \\ |\Phi_3\rangle \;&=\; \mathbf{1}\otimes\sigma_3|\Phi_0\rangle \;=\; \frac{1}{\sqrt{2}}\Big(|+\rangle|+\rangle - |-\rangle|-\rangle\Big)\,. \end{aligned} \tag{3.56}$$

By inspection we can confirm that Φ_0 is orthogonal to the other three: $\langle\Phi_0|\Phi_i\rangle = 0$. It is not much work either to see that the basis is in fact orthonormal. But a calculation is kind of fun:

$$\begin{aligned} \langle\Phi_i|\Phi_j\rangle \;&=\; \langle\Phi_0|(\mathbf{1}\otimes\sigma_i)(\mathbf{1}\otimes\sigma_j)|\Phi_0\rangle \\ &=\; \langle\Phi_0|\mathbf{1}\otimes\sigma_i\sigma_j|\Phi_0\rangle \\ &=\; \langle\Phi_0|\mathbf{1}\otimes\big(\mathbf{1}\delta_{ij} + i\epsilon_{ijk}\sigma_k\big)|\Phi_0\rangle \\ &=\; \delta_{ij}\langle\Phi_0|\mathbf{1}\otimes\mathbf{1}|\Phi_0\rangle + i\epsilon_{ijk}\langle\Phi_0|\mathbf{1}\otimes\sigma_k|\Phi_0\rangle \\ &=\; \delta_{ij}\langle\Phi_0|\Phi_0\rangle + i\epsilon_{ijk}\langle\Phi_0|\Phi_k\rangle \;=\; \delta_{ij}\,, \end{aligned} \tag{3.57}$$

as we wanted to show. Indeed, we have an orthonormal basis of entangled states.

We can solve for the old, non-entangled basis states in terms of the Bell states. We quickly find from (3.56)

$$\begin{aligned} |+\rangle|+\rangle &= \frac{1}{\sqrt{2}}\left(|\Phi_0\rangle + |\Phi_3\rangle\right) \\ |-\rangle|-\rangle &= \frac{1}{\sqrt{2}}\left(|\Phi_0\rangle - |\Phi_3\rangle\right) \\ |+\rangle|-\rangle &= \frac{1}{\sqrt{2}}\left(|\Phi_1\rangle - i|\Phi_2\rangle\right) \\ |-\rangle|+\rangle &= \frac{1}{\sqrt{2}}\left(|\Phi_1\rangle + i|\Phi_2\rangle\right). \end{aligned} \tag{3.58}$$

Introducing labels A and B for the two spaces in a tensor product $V_A \otimes V_B$ we rewrite the above equations as

$$\begin{aligned} |+\rangle_A|+\rangle_B &= \frac{1}{\sqrt{2}}\left(|\Phi_0\rangle_{AB} + |\Phi_3\rangle_{AB}\right) \\ |-\rangle_A|-\rangle_B &= \frac{1}{\sqrt{2}}\left(|\Phi_0\rangle_{AB} - |\Phi_3\rangle_{AB}\right) \\ |+\rangle_A|-\rangle_B &= \frac{1}{\sqrt{2}}\left(|\Phi_1\rangle_{AB} - i|\Phi_2\rangle_{AB}\right) \\ |-\rangle_A|+\rangle_B &= \frac{1}{\sqrt{2}}\left(|\Phi_1\rangle_{AB} + i|\Phi_2\rangle_{AB}\right), \end{aligned} \tag{3.59}$$

where $|\Phi_i\rangle_{AB}$ are the Bell states we defined above with tensor products in which the first state is in V_A and the second state is in V_B.

These basis states form the Bell basis. You could do an experiment to determine the probability of an arbitrary state being along any of the basis states in this orthonormal basis. You can use the experiment to detect which basis state the state is in. The state is, of course, a superposition of basis states, but during measurement will collapse into one of them with some probability. The Stern Gerlach device was an example of a device that allowed you to collapse a state into one basis state or another. This basis is more general, as it is not simply for two-state systems.
We conclude by presenting three facts.

1. Measuring in a basis. Given an orthonormal basis $|e_1\rangle, ..., |e_n\rangle$ we can measure a state $|\Psi\rangle$ along this basis and obtain that the probability $P(i)$ to be in the state $|i\rangle$ is $|\langle e_i|\Psi\rangle|^2$. After measurement the state will be in one of the states $|e_i\rangle$. This is exactly how it worked for the Stern-Gerlach experiment which, oriented about $\mathbf{z}$ amount to a measurement in the basis $|+\rangle, |-\rangle$.

   As another example, if we have a state with two particles A, B, we may choose the four Bell states as our orthonormal basis for the measurement. If so, after measurement the state will be in one of the Bell states $|\Phi_i\rangle_{AB}$, with probability given by the squared overlap $|\langle\Phi_i|_{AB}|\Psi\rangle|^2$.

2. Partial measurement. Suppose we have a general (entangled) state $\Psi \in V \otimes W$ of two particles. The observer Alice has access to both particles but decides to measure only the first particle along the basis $|e_1\rangle, \ldots, |e_n\rangle$ of V. How is this analyzed? As a first step we use that basis to write the state Ψ in the form

$$\Psi \;=\; \sum_i |e_i\rangle \otimes |w_i\rangle\,, \tag{3.60}$$

for some calculable vectors $|w_i\rangle$. As a second step we normalize the states $|w_i\rangle$:

$$\Psi \;=\; \sum_i \sqrt{\langle w_i|w_i\rangle}\; |e_i\rangle \otimes \frac{|w_i\rangle}{\sqrt{\langle w_i|w_i\rangle}}\,, \tag{3.61}$$

We claim that Alice will find the first particle to be in the state $|i\rangle$ with probability $\langle w_i|w_i\rangle$. After the measurement, the state of the particles will be

$$|e_i\rangle \otimes \frac{|w_i\rangle}{\sqrt{\langle w_i|w_i\rangle}}\,, \quad \text{for some value of } i\,. \tag{3.62}$$

(A justification of this answer was given in recitations.) You probably have used this rule before. As an example, suppose we have the entangled state of total spin zero:

$$|\Psi\rangle \;=\; \frac{1}{\sqrt{2}}\Big(|+\rangle_1 \otimes |-\rangle_2 \;-\; |-\rangle_1 \otimes |+\rangle_2\Big) \tag{3.63}$$

If we measure the first particle along the $|+\rangle_1, |+\rangle_2$ basis we find

$$\begin{array}{l} \text{Probability that the first particle is in } |+\rangle = \tfrac{1}{2}\,. \;\; \text{State after measurement: } |+\rangle_1 \otimes |-\rangle_2 \\ \text{Probability that the first particle is in } |-\rangle = \tfrac{1}{2}\,. \;\; \text{State after measurement: } |-\rangle_1 \otimes |+\rangle_2 \end{array} \tag{3.64}$$

It follows that after the measurement of the first particle, a measurement of the second particle will show that its spin is always opposite to the spin of the first particle.

As a more nontrivial example, consider now the state of three particles A, B, C which live in $V_A \otimes V_B \otimes V_C$, which contains states of the type $v \otimes w \otimes u$ with $v \in V_A, w \in V_B, u \in V_C$, and their linear combinations. To analyze what happens if Alice decides to do a Bell measurement of particles A, B, the state Ψ of the system must be written in the form

$$\Psi \;=\; |\Phi_0\rangle_{AB} \otimes |u_0\rangle_C + |\Phi_1\rangle_{AB} \otimes |u_1\rangle_C + |\Phi_2\rangle_{AB} \otimes |u_2\rangle_C + |\Phi_3\rangle_{AB} \otimes |u_3\rangle_C \tag{3.65}$$

After the measurement, the state of the particles A, B will be one of the Bell states $|\Phi_\mu\rangle_{AB}$ with $\mu = 0, 1, 2, 3$. We have

$$\begin{array}{l} \text{Probability that } (A,B) \text{ is in } |\Phi_\mu\rangle_{AB} \;=\; \langle u_\mu|u_\mu\rangle\,, \\ \text{State after measurement is } |\Phi_\mu\rangle_{AB} \otimes \dfrac{|u_\mu\rangle_C}{\sqrt{\langle u_\mu|u_\mu\rangle}} \text{ for some } \mu \in \{0,1,2,3\}\,. \end{array} \tag{3.66}$$

3. The action of the Pauli matrices on spin states can be realized as time evolution via some Hamiltonian. Note first that the Pauli matrices are unitary because they are Hermitian and square to the identity. Multiplying a state by σ_1 is thus acting with a unitary operator and unitary operators generate allowed time evolution. Thus, there is a Hamiltonian that applied to a system over some length of time will turn any spin state $|\Psi\rangle$ into $\sigma_i|\Psi\rangle$. In practice, this Hamiltonian would correspond to some device with a magnetic field of some determined magnitude and direction that acts for a few picoseconds and evolves spin states in time. We can check, for example, that any Pauli matrix can be written as the exponential of i times a Hermitian matrix (which would be proportional to the Hamiltonian):

$$e^{i\frac{\pi}{2}(-1+\sigma_i)} = e^{-i\frac{\pi}{2}}e^{i\frac{\pi}{2}\sigma_i} = (-i)(i\sigma_i) = \sigma_i \tag{3.67}$$

4 Quantum Teleportation

Classically, teleportation is impossible: there is no classical basis for dematerializing an object and recreating it somewhere else. In 1993, a group of scientists (Bennet, Brassand, Crépeau, Jozsa, Peres, and Wooters) discovered that teleportation *is* possible in quantum mechanics.

Imagine that Alice has a quantum state: the state of a 1/2 particle. The state is:

$$|\Psi\rangle_C \;=\; \alpha|+\rangle_C + \beta|-\rangle_C\,, \tag{4.68}$$

where $\alpha, \beta \in \mathbb{C}$ and the letter C denotes the state space V_C of this C particle to be teleported. Her goal is to teleport this state - called a "quantum bit," or *qubit* - to Bob, who is far away.

The quantum "no-cloning" principle prevents Alice from simply creating a copy of the state and sending it to Bob. In other words, it is impossible to create a copy of a quantum mechanical state. Measuring the state and telling Bob about the result is no option either: if Alice measures the state with some Stern-Gerlach apparatus, the spin will just point up or point down. What has she learned? Almost nothing. Only with many copies of the state she would be able to learn about the values of α and β. Having just one particle she is unable to measure α and β and send those values to Bob. Of course, it may be that for some reason Alice knows the values of α and β. In some cases she could transmit that information to Bob to recreate the state. But it could also be that α, for example is some transcendental number 0.178573675623..... with no discernible rhyme or reason, and she would need an infinite amount of information to send to Bob this value.

Here is a diagram of how Alice can will teleport the information:

The key tool Alice and Bob use is an entangled states of two particles A and B, in which Alice has access to particle A and Bob has access to particle B. One pair (A, B) of entangled particles will allow Alice to teleport the state C of one particle. To teleport a full person from one place to another, we would have to have an enormous reservoir of entangled pairs, one pair needed to teleport each quantum state of particles in the body of that person. This clearly remains science-fiction.

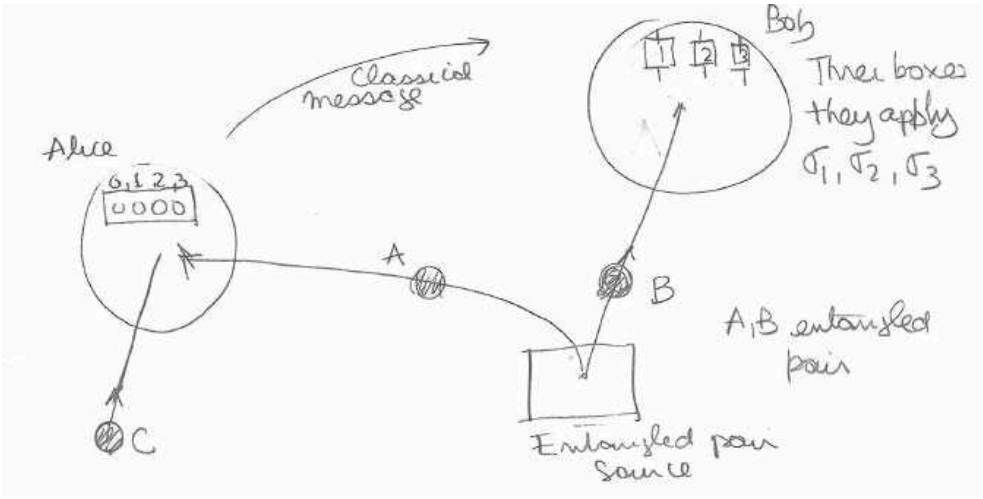


Figure 1: Alice has access to spin state C, to be teleported and to spin state A which is entangled with spin state B. Bob has spin state B. Alice will perform a Bell measurement on states A and C. After she measures, Bob's state will turn into the desired state, up to a simple unitary transformation.

Alice has a console with four lights, labeled with $\mu = 0, 1, 2, 3$. She will do a Bell measurement involving particle A (of the shared entangled pair) and particle C (the one to be teleported). As she does so, one of her four lights will blink: If it is th μ-th light it is because she ended with the Bell state $|\Phi_\mu\rangle_{AC}$. Bob, who is in possession of the particle B, has a console with four boxes which generate unitary transformations (via some Hamiltonians applied for a fixed time). The first box, labeled $\mu = 0$ does nothing to the state. The i-th box (with $i = 1, 2, 3$) applies the operator σ_i. Alice communicates to Bob that the μ-th light blinked. Then Bob submits particle B to the μ-th box and out comes, we claim, the teleported state $|\Psi\rangle_B$ as the state of particle B.

Let us prove this mathematically. Let the entangled shared pair, with A at Alice and B at Bob, be the first Bell basis state:

$$|\Phi_0\rangle_{AB} = \frac{1}{\sqrt{2}}\big(|+\rangle_A|+\rangle_B + |-\rangle_A|-\rangle_B\big)\,. \tag{4.69}$$

The total state of our three particles, A, B, C is therefore:

$$\begin{aligned} |\Phi_0\rangle_{AB} \otimes |\Psi\rangle_C &= |\Phi_0\rangle_{AB} \otimes (\alpha|+\rangle_C + \beta|-\rangle_C) \\ &= \frac{1}{\sqrt{2}}\Big(|+\rangle_A|+\rangle_B + |-\rangle_A|-\rangle_B\Big) \otimes \big(\alpha|+\rangle_C + \beta|-\rangle_C\big)\,. \end{aligned} \tag{4.70}$$

Expanding out and reordering the states to have A followed by C and then by B we have

$$\begin{aligned} |\Phi_0\rangle_{AB} \otimes |\Psi\rangle_C = \frac{1}{\sqrt{2}}\Big(& \alpha \underbrace{|+\rangle_A|+\rangle_C}|+\rangle_B + \beta \underbrace{|+\rangle_A|-\rangle_C}|+\rangle_B \\ & + \alpha \underbrace{|-\rangle_A|+\rangle_C}|-\rangle_B + \beta \underbrace{|-\rangle_A|-\rangle_C}|-\rangle_B\Big)\,. \end{aligned} \tag{4.71}$$

Note that as long as we label the states, the order in which we write them does not matter! We now write these basis states with braces in the Bell basis using (3.59). We find

$$\begin{aligned} |\Phi_0\rangle_{AB}\otimes|\Psi\rangle_C \;=\; & \tfrac{1}{2}\Big(|\Phi_0\rangle_{AC}+|\Phi_3\rangle_{AC}\Big)\;\alpha|+\rangle_B+\tfrac{1}{2}\Big(|\Phi_1\rangle_{AC}-i|\Phi_2\rangle_{AC}\Big)\;\beta|+\rangle_B \\ & +\tfrac{1}{2}\Big(|\Phi_1\rangle_{AC}+i|\Phi_2\rangle_{AC}\Big)\;\alpha|-\rangle_B+\tfrac{1}{2}\Big(|\Phi_0\rangle_{AC}-|\Phi_3\rangle_{AC}\Big)\;\beta|-\rangle_B\,. \end{aligned} \tag{4.72}$$

Collecting the Bell states we find

$$\begin{aligned} |\Phi_0\rangle_{AB}\otimes|\Psi\rangle_C \;=\; & \tfrac{1}{2}|\Phi_0\rangle_{AC}\left(\alpha|+\rangle_B\;+\;\beta|-\rangle_B\right)\;+\;\tfrac{1}{2}|\Phi_1\rangle_{AC}\left(\alpha|-\rangle_B\;+\;\beta|+\rangle_B\right) \\ & +\tfrac{1}{2}|\Phi_2\rangle_{AC}\left(i\alpha|-\rangle_B\;-\;i\beta|+\rangle_B\right)\;+\;\tfrac{1}{2}|\Phi_3\rangle_{AC}\left(\alpha|+\rangle_B\;-\;\beta|-\rangle_B\right). \end{aligned} \tag{4.73}$$

We can then see that in fact we got

$$\begin{aligned} |\Phi_0\rangle_{AB}\otimes|\Psi\rangle_C \;=\; & \tfrac{1}{2}|\Phi_0\rangle_{AC}\otimes|\Psi\rangle_B\;+\;\tfrac{1}{2}|\Phi_1\rangle_{AC}\otimes\sigma_1|\Psi\rangle_B \\ & +\tfrac{1}{2}|\Phi_2\rangle_{AC}\otimes\sigma_2|\Psi\rangle_B+\tfrac{1}{2}|\Phi_3\rangle_{AC}\otimes\sigma_3|\Psi\rangle_B\,. \end{aligned} \tag{4.74}$$

The above right-hand side allows us to understand what happens when Alice measures the state of (A,C) in the Bell basis. If she measures:

- $|\Phi_0\rangle_{AC}$, then the B state becomes $|\Psi\rangle_B$,
- $|\Phi_1\rangle_{AC}$, then the B state becomes $\sigma_1|\Psi\rangle_B$,
- $|\Phi_2\rangle_{AC}$, then the B state becomes $\sigma_2|\Psi\rangle_B$,
- $|\Phi_3\rangle_{AC}$, then the B state becomes $\sigma_3|\Psi\rangle_B$.

If Alice got $|\Phi_0\rangle_{AC}$ then Bob is in the possession of the teleported state and has to do nothing. If Alice gets $|\Phi_i\rangle_{AC}$, Bob's particle is goes into the state $\sigma_i|\Psi\rangle_B$. Bob applies the i-th box, which multiplies his state by σ_i giving him the desired state $|\Psi\rangle_B$. The teleporting is thus complete!

Note that Alice is left with one of the Bell states $|\Phi_\mu\rangle_{AC}$ which has no information whatsoever about the constants α and β that defined the state to be teleported. Thus the process did not create a copy of the state. The original state is destroyed in the process of teleportation.

It is noteworthy that all the mathematical work above led to the key result (4.74), which is neatly summarized as the following identity valid for arbitrary states $|\Psi\rangle$:

$$\boxed{|\Phi_0\rangle_{AB}\otimes|\Psi\rangle_C \;=\; \frac{1}{2}\sum_{i=0}^{3}|\Phi_i\rangle_{AC}\otimes\sigma_i|\Psi\rangle_B\,.} \tag{4.75}$$

This is an identity for a state of three particles. It expresses the tensor product of an entangled state of the first two particles, times a third, as a sum of products that involve entangled states of the first and third particle times a state of the second particle.

5 EPR and Bell Inequalities

In this section we begin by studying some properties of the singlet state of two particles of spin-1/2. We then turn to the claims of Einstein, Podolsky, and Rosen (EPR) concerning quantum mechanics. Finally, we discuss the so-called Bell inequalities that would follow if EPR were right. Of course, quantum mechanics violates these inequalities, and experiment shows that the inequalities are indeed violated. EPR were wrong.

We have been talking about the singlet state of two spin-1/2 particles. This state emerges, for example, in particle decays. The neutral η_0 meson (of rest mass 547 MeV) sometimes decays into two oppositely charged muons

$$\eta_0 \to \mu^+ + \mu^- \,. \tag{5.1}$$

The meson is a spinless particle and being at rest has zero orbital angular momentum. As a result it has zero total angular momentum. As it decays, the final state of the two muons must have zero total angular momentum as well. If the state of the two muons has zero orbital angular momentum, it must also have zero total spin angular momentum. The two muons flying away from each other with zero orbital angular momentum are in a singlet state. This state takes the form

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\big(|+\rangle_1|-\rangle_2 - |-\rangle_1|+\rangle_2\big)\,. \tag{5.2}$$

This singlet state is rotational invariant and therefore it is actually the same for whatever choice of direction$\mathbf{n}$ to define a basis of spin states:

$$|\Psi\rangle \;=\; \frac{1}{\sqrt{2}}\big(|\mathbf{n};+\rangle_1|\mathbf{n};-\rangle_2 - |\mathbf{n};-\rangle_1|\mathbf{n};+\rangle_2\big)\,. \tag{5.3}$$

We now ask: In this singlet, what is the probability $P(\mathbf{a},\mathbf{b})$ that the first particle is in the state $|\mathbf{a};+\rangle$ and the second particle is in the state $|\mathbf{b};+\rangle$, with $\mathbf{a}$ and $\mathbf{b}$ two arbitrarily chosen unit vectors? To help ourselves, we write the singlet state using the first vector

$$|\Psi\rangle \;=\; \frac{1}{\sqrt{2}}\big(|\mathbf{a};+\rangle_1|\mathbf{a};-\rangle_2 - |\mathbf{a};-\rangle_1|\mathbf{a};+\rangle_2\big)\,. \tag{5.4}$$

By definition, the probability we want is

$$P(\mathbf{a},\mathbf{b}) \;=\; \Big|{}_1\langle\mathbf{a};+|_2\langle\mathbf{b};+|\Psi\rangle\Big|^2 \tag{5.5}$$

Only the first term in (5.4) contributes and we get

$$P(\mathbf{a},\mathbf{b}) \;=\; \frac{1}{2}\big|\langle\mathbf{b};+|\mathbf{a};-\rangle\big|^2 \tag{5.6}$$

We recall that the overlap-squared between two spin states is given by the cosine-squared of half the angle in between them. Using figure 2 we see that the angle between $\mathbf{b}$ and $-\mathbf{a}$ is $\pi - \theta_{ab}$, where θ_{ab} is the angle between $\mathbf{b}$ and $\mathbf{a}$. Therefore

$$P(\mathbf{a},\mathbf{b}) \;=\; \tfrac{1}{2}\cos^2\big(\tfrac{1}{2}(\pi-\theta_{ab})\big) \tag{5.7}$$

Our final result is therefore

$$\boxed{P(\mathbf{a},\mathbf{b}) \;=\; \tfrac{1}{2}\sin^2\!\left(\tfrac{1}{2}\theta_{ab}\right).} \tag{5.8}$$

As a simple consistency check, if $\mathbf{b} = -\mathbf{a}$ then $\theta_{ab} = \pi$ and $P(\mathbf{a}, -\mathbf{a}) = 1/2$ which is what we expect.

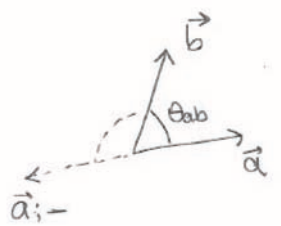


Figure 2: Directions associated with the vectors **a** and **b**.

If we measure about orthogonal vectors, like the unit vectors $\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ we get

$$P(\hat{\mathbf{z}}, \hat{\mathbf{x}}) = \tfrac{1}{2}\sin^2 45^\circ = \tfrac{1}{2}\cdot\tfrac{1}{2} = \tfrac{1}{4}\,. \tag{5.9}$$

The key statement of Einstein, Podolsky and Rosen (EPR) is the claim for **local realism**. This is posed as two properties of measurement:

1. The result of a measurement corresponds to some element of reality. If a measurement of an observable gives a value, that value was a property of the state.

2. The result of a measurement at one point cannot depend on whatever action takes place at a far away point at the same time.

Both properties seem quite plausible at first thought. The first, we are by now accustomed, is violated in Quantum Mechanics, where measurement involves collapse of the wavefunction, so that the result was not pre-ordained and does not correspond to a unequivocal property of the system. The violation of the second is perhaps equally disturbing, given our intuition that simultaneous spatially separated events can't affect each other. There is something non-local about quantum mechanics.

According to EPR the so called entangled pairs are just pairs of particles that have definite spins. They point out that the results of quantum mechanical measurements are reproduced if our large ensemble of pairs has the following distribution of states:

- In 50% of pairs, particle 1 has spin along $\hat{\mathbf{z}}$ and particle 2 has spin along $-\hat{\mathbf{z}}$,

- In 50% of pairs, particle 1 has spin along $-\hat{\mathbf{z}}$ and particle 2 has spin along $\hat{\mathbf{z}}$.

This would explain the perfect correlations and is consistent, for example, with $P(\hat{\mathbf{z}}, -\hat{\mathbf{z}}) = 1/2$, which we obtained quantum mechanically.

The challenge for EPR is to keep reproducing the results of more complicated measurements. Suppose each of the two observers can measure spin along two possible axes: the x and z axes. They measure once, in any of these two directions. EPR then state that in any pair each particle has a definite state of spin in these two directions. For example, a particle of type $(\hat{\mathbf{z}}, -\hat{\mathbf{x}})$ is one that if measured along z gives a plus $\hbar/2$ and if measured along x gives $-\hbar/2$. We do not do simultaneous measurements or subsequent measurements on each particle. EPR then claim that the observed quantum mechanical results are matched if our ensemble of pairs have the following properties

- 25% of pairs have particle 1 in $(\hat{\mathbf{z}}, \hat{\mathbf{x}})$ and particle 2 in $(-\hat{\mathbf{z}}, -\hat{\mathbf{x}})$
- 25% of pairs have particle 1 in $(\hat{\mathbf{z}}, -\hat{\mathbf{x}})$ and particle 2 in $(-\hat{\mathbf{z}}, \hat{\mathbf{x}})$
- 25% of pairs have particle 1 in $(-\hat{\mathbf{z}}, \hat{\mathbf{x}})$ and particle 2 in $(\hat{\mathbf{z}}, -\hat{\mathbf{x}})$
- 25% of pairs have particle 1 in $(-\hat{\mathbf{z}}, -\hat{\mathbf{x}})$ and particle 2 in $(\hat{\mathbf{z}}, \hat{\mathbf{x}})$

First note the complete correlations: particles one and two have opposite spins in each possible direction. This is, of course, needed to match the quantum mechanical singlets. We can ask what is $P(\hat{\mathbf{z}}, -\hat{\mathbf{z}})$, the probability that particle one is along $\hat{\mathbf{z}}$ and particle two along $-\hat{\mathbf{z}}$. The first two cases above apply, and thus this probability is $1/2$, consistent with quantum mechanics. We can also ask for $P(\hat{\mathbf{z}}, \hat{\mathbf{x}})$. This time only the second case applies giving us a probability of $1/4$ as we obtained earlier in (5.9). The quantum mechanical answers indeed arise.

The insight of Bell was that with Stern-Gerlach apparatuses that could measure in three directions one gets in trouble. Suppose each observer can measure along any one of the three vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$. Again, each particle is just measured once. Let us assume that we have a large number N of pairs that, following EPR, contain particles with well-defined spins on these three directions. A particle of type $(\mathbf{a}, -\mathbf{b}, \mathbf{c})$, for example, if measured along $\mathbf{a}$ would give $\hbar/2$, if measured along $\mathbf{b}$ would give $-\hbar/2$ and if measured along $\mathbf{c}$ would give $\hbar/2$. The following distribution is given:

Populations	**Particle 1**	**Particle 2**
N_1	$(\ \mathbf{a},\ \mathbf{b},\ \mathbf{c})$	$(-\mathbf{a}, -\mathbf{b}, -\mathbf{c})$
N_2	$(\ \mathbf{a},\ \mathbf{b}, -\mathbf{c})$	$(-\mathbf{a}, -\mathbf{b},\ \mathbf{c})$
N_3	$(\ \mathbf{a}, -\mathbf{b},\ \mathbf{c})$	$(-\mathbf{a},\ \mathbf{b}, -\mathbf{c})$
N_4	$(\ \mathbf{a}, -\mathbf{b}, -\mathbf{c})$	$(-\mathbf{a},\ \mathbf{b},\ \mathbf{c})$
N_5	$(-\mathbf{a},\ \mathbf{b},\ \mathbf{c})$	$(\ \mathbf{a}, -\mathbf{b}, -\mathbf{c})$
N_6	$(-\mathbf{a},\ \mathbf{b}, -\mathbf{c})$	$(\ \mathbf{a}, -\mathbf{b},\ \mathbf{c})$
N_7	$(-\mathbf{a}, -\mathbf{b},\ \mathbf{c})$	$(\ \mathbf{a},\ \mathbf{b}, -\mathbf{c})$
N_8	$(-\mathbf{a}, -\mathbf{b}, -\mathbf{c})$	$(\ \mathbf{a},\ \mathbf{b},\ \mathbf{c})$

As required, all spins are correlated in particles one and two. We also have $N = \sum_{i=1}^{8} N_i$. We now record the following probabilities that follow by inspection of the table:

$$P(\mathbf{a},\mathbf{b}) = \frac{N_3 + N_4}{N}\,, \quad P(\mathbf{a},\mathbf{c}) = \frac{N_2 + N_4}{N}\,, \quad P(\mathbf{c},\mathbf{b}) = \frac{N_3 + N_7}{N}\,. \tag{5.10}$$

Consider now the trivially correct inequality:

$$N_3 + N_4 \;\leq\; N_3 + N_7 \;+\; N_2 + N_4\,, \tag{5.11}$$

that on account of (5.10) implies the **Bell inequality**

$$\boxed{P(\mathbf{a},\mathbf{b}) \;\leq\; P(\mathbf{a},\mathbf{c}) + P(\mathbf{c},\mathbf{b})\,.} \tag{5.12}$$

If true quantum mechanically, given (5.8) we would have

$$\tfrac{1}{2}\sin^2\tfrac{1}{2}\theta_{ab} \;\leq\; \tfrac{1}{2}\sin^2\tfrac{1}{2}\theta_{ac} \;+\tfrac{1}{2}\sin^2\tfrac{1}{2}\theta_{cb}\,. \tag{5.13}$$

But this is violated for many choices of angles. Take, for example, the planar configuration in Fig. 3:

$$\theta_{ab} = 2\theta\,, \quad \theta_{ac} = \theta_{cb} = \theta\,. \tag{5.14}$$

For this situation, the inequality becomes

$$\tfrac{1}{2}\sin^2\theta \leq \sin^2\tfrac{1}{2}\theta\,. \tag{5.15}$$

This fails for sufficiently small θ: $\frac{1}{2}\theta^2 \leq \frac{\theta^2}{4}$ is just plain wrong. In fact, the inequality goes wrong for any $\theta < \frac{\pi}{2}$. Experimental results have confirmed that Bell inequalities are violated and thus the original claim of local realism by EPR is wrong.

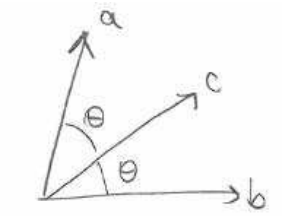


Figure 3: Special configuration for vectors $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$.

ANGULAR MOMENTUM

B. Zwiebach
December 16, 2013

Contents

1 Orbital angular momentum and central potentials

Classically the angular momentum vector $\vec{L}$ is defined as the cross-product of the position vector $\vec{r}$ and the momentum vector $\vec{p}$:

$$\vec{L} \;=\; \vec{r}\times\vec{p}. \tag{1.1}$$

In cartesian components, this equation reads

$$\begin{aligned} L_x &= yp_z - zp_y\,, \\ L_y &= zp_x - xp_z\,, \\ L_z &= xp_y - yp_x\,. \end{aligned} \tag{1.2}$$

In quantum mechanics the classical vectors $\vec{r}$, $\vec{p}$ and $\vec{L}$ become operators. More precisely, they give us triplets of operators:

$$\begin{aligned} \vec{r} &\rightarrow (\,\hat{x}\,,\,\hat{y}\,,\,\hat{z}\,)\,, \\ \vec{p} &\rightarrow (\,\hat{p}_x\,,\,\hat{p}_y\,,\,\hat{p}_z\,)\,, \\ \vec{L} &\rightarrow (\,\hat{L}_x\,,\,\hat{L}_y\,,\,\hat{L}_z\,)\,. \end{aligned} \tag{1.3}$$

When we want more uniform notation, instead of x, y, and z labels we use $1, 2$ and 3 labels:

$$\begin{aligned} (\,\hat{x}_1\,,\,\hat{x}_2\,,\,\hat{x}_3\,) &\equiv (\,\hat{x}\,,\,\hat{y}\,,\,\hat{z}\,)\,, \\ (\,\hat{p}_1\,,\,\hat{p}_2\,,\,\hat{p}_3\,) &\equiv (\,\hat{p}_x\,,\,\hat{p}_y\,,\,\hat{p}_z\,)\,, \\ (\,\hat{L}_1\,,\,\hat{L}_2\,,\,\hat{L}_3\,) &\equiv (\,\hat{L}_x\,,\,\hat{L}_y\,,\,\hat{L}_z\,)\,. \end{aligned} \tag{1.4}$$

The basic canonical commutation relations then are easily summarized as

$$\left[\,\hat{x}_i\,,\hat{p}_j\,\right] \;=\; i\hbar\,\delta_{ij}\,, \quad \left[\,\hat{x}_i\,,\hat{x}_j\,\right] \;=\; 0\,, \quad \left[\,\hat{p}_i\,,\hat{p}_j\,\right] \;=\; 0\,. \tag{1.5}$$

Thus, for example, $\hat{x}$ commutes with $\hat{y}, \hat{z}, \hat{p}_y$ and $\hat{p}_z$, but fails to commute with $\hat{p}_x$. In view of (1.2) and (1.3) it is natural to define the angular momentum *operators* by

$$\begin{aligned} \hat{L}_x &\equiv \hat{y}\,\hat{p}_z - \hat{z}\,\hat{p}_y\,, \\ \hat{L}_y &\equiv \hat{z}\,\hat{p}_x - \hat{x}\,\hat{p}_z\,, \\ \hat{L}_z &\equiv \hat{x}\,\hat{p}_y - \hat{y}\,\hat{p}_x\,. \end{aligned} \tag{1.6}$$

Note that these equations are free of ordering ambiguities: each product involves a coordinate and a momentum that commute! In terms of numbered operators

$$\begin{aligned} \hat{L}_1 &\equiv \hat{x}_2\,\hat{p}_3 - \hat{x}_3\,\hat{p}_2\,, \\ \hat{L}_2 &\equiv \hat{x}_3\,\hat{p}_1 - \hat{x}_1\,\hat{p}_3\,, \\ \hat{L}_3 &\equiv \hat{x}_1\,\hat{p}_2 - \hat{x}_2\,\hat{p}_1\,. \end{aligned} \tag{1.7}$$

Note that the angular momentum operators are Hermitian, since $\hat{x}_i$ and $\hat{p}_i$ are and the products can be reordered without cost:

$$\hat{L}_i^\dagger \;=\; \hat{L}_i\,. \tag{1.8}$$

1.1 Quantum mechanical vector identities

We will write triplets of operators as boldfaced vectors, each element of the triplet multiplied by a unit basis vector, just like we do for ordinary vectors. Thus, for example, we have

$$\begin{aligned} \mathbf{r} &\equiv \hat{x}_1\,\vec{e}_1 + \hat{x}_2\,\vec{e}_2 + \hat{x}_3\,\vec{e}_3\,, \\ \mathbf{p} &\equiv \hat{p}_1\,\vec{e}_1 + \hat{p}_2\,\vec{e}_2 + \hat{p}_3\,\vec{e}_3\,, \\ \mathbf{L} &\equiv \hat{L}_1\,\vec{e}_1 + \hat{L}_2\,\vec{e}_2 + \hat{L}_3\,\vec{e}_3\,. \end{aligned} \tag{1.9}$$

These boldface objects are a bit unusual. They are vectors whose components happen to be operators! Moreover, the basis vectors $\vec{e}_i$ must be declared to commute with any of the operators. The boldface objects are useful whenever we want to use the dot products and cross products of three-dimensional space.

Let us, for generality consider vectors $\mathbf{a}$ and $\mathbf{b}$

$$\begin{aligned} \mathbf{a} &\equiv a_1\,\vec{e}_1 + a_2\,\vec{e}_2 + a_3\,\vec{e}_3\,, \\ \mathbf{b} &\equiv b_1\,\vec{e}_1 + b_2\,\vec{e}_2 + b_3\,\vec{e}_3\,, \end{aligned} \tag{1.10}$$

and we will assume that the a_i's and b_j's are operators that do not commute. The following are then standard definitions:

$$\begin{aligned} \mathbf{a}\cdot\mathbf{b} &\equiv a_i\,b_i\,, \\ (\mathbf{a}\times\mathbf{b})_i &\equiv \epsilon_{ijk}\,a_j\,b_k\,. \end{aligned} \tag{1.11}$$

The order of the operators in the above right-hand sides cannot be changed; it was chosen conveniently, to be the same as the order of the operators on the left-hand sides. We also define,

$$\mathbf{a}^2 \equiv \mathbf{a}\cdot\mathbf{a}\,. \tag{1.12}$$

Since the operators do not commute, familiar properties of vector analysis do not hold. For example $\mathbf{a}\cdot\mathbf{b}$ is not equal to $\mathbf{b}\cdot\mathbf{a}$. Indeed,

$$\mathbf{a}\cdot\mathbf{b} = a_i\,b_i = [\,a_i\,,\,b_i\,] + b_i\,a_i\,, \tag{1.13}$$

so that

$$\boxed{\mathbf{a}\cdot\mathbf{b} = \mathbf{b}\cdot\mathbf{a} + [\,a_i\,,\,b_i\,]\,.} \tag{1.14}$$

As an application we have

$$\mathbf{r}\cdot\mathbf{p} = \mathbf{p}\cdot\mathbf{r} + [\,\hat{x}_i\,,\,\hat{p}_i\,]\,, \tag{1.15}$$

The right-most commutator gives $i\hbar\,\delta_{ii} = 3i\hbar$ so that we have the amusing three-dimensional identity

$$\boxed{\mathbf{r}\cdot\mathbf{p} = \mathbf{p}\cdot\mathbf{r} + 3\,i\hbar\,.} \tag{1.16}$$

For cross products we typically have $\mathbf{a}\times\mathbf{b} \neq -\mathbf{b}\times\mathbf{a}$. Indeed,

$$\begin{aligned} (\mathbf{a}\times\mathbf{b})_i &= \epsilon_{ijk}\,a_j\,b_k = \epsilon_{ijk}\,\big([a_j\,,\,b_k\,] + b_k\,a_j\big) \\ &= -\,\epsilon_{ikj}\;b_k\,a_j + \epsilon_{ijk}\,[a_j\,,\,b_k\,] \end{aligned} \tag{1.17}$$

where we flipped the k, j indices in one of the epsilon tensors in order to identify a cross product. Indeed, we have now

$$\boxed{(\mathbf{a}\times\mathbf{b})_i = -(\mathbf{b}\times\mathbf{a})_i + \epsilon_{ijk}\,[a_j\,,\,b_k\,]\,.} \tag{1.18}$$

The simplest example of the use of this identity is one where we use $\mathbf{r}$ and $\mathbf{p}$. Certainly

$$\mathbf{r}\times\mathbf{r} \;=\; 0\,, \quad \text{and} \quad \mathbf{p}\times\mathbf{p} \;=\; 0\,, \tag{1.19}$$

and more nontrivially,

$$(\mathbf{r}\times\mathbf{p})_i \;=\; -(\mathbf{p}\times\mathbf{r})_i + \epsilon_{ijk}\,[\,\hat{x}_j\,,\,\hat{p}_k\,]\,. \tag{1.20}$$

The last term vanishes for it is equal to $i\hbar\,\epsilon_{ijk}\delta_{jk} = 0$ (the epsilon symbol is antisymmetric in j,k while the delta is symmetric in j,k, resulting in a zero result). We therefore have, quantum mechanically,

$$\boxed{\mathbf{r}\times\mathbf{p} \;=\; -\,\mathbf{p}\times\mathbf{r}\,.} \tag{1.21}$$

Thus $\mathbf{r}$ and $\mathbf{p}$ can be moved across in the cross product but not in the dot product.

Exercise 1. Prove the following identities for Hermitian conjugation

$$\begin{aligned}(\mathbf{a}\cdot\mathbf{b})^\dagger \;&=\; \mathbf{b}^\dagger\cdot\mathbf{a}^\dagger\,,\\ (\mathbf{a}\times\mathbf{b})^\dagger \;&=\; -\;\mathbf{b}^\dagger\times\mathbf{a}^\dagger\,.\end{aligned} \tag{1.22}$$

Our definition of the angular momentum operators in (1.7) and the notation developed above imply that we have

$$\boxed{\mathbf{L} \;=\; \mathbf{r}\times\mathbf{p} \;=\; -\mathbf{p}\times\mathbf{r}\,.} \tag{1.23}$$

Indeed, given the definition of the product, we have

$$\boxed{\hat{L}_i \;=\; \epsilon_{ijk}\,\hat{x}_j\,\hat{p}_k\,.} \tag{1.24}$$

If you write out what this means for $i = 1,2,3$ (do it!) you will recover the expressions in (1.7). The angular operator is Hermitian. Indeed, using (1.22) and recalling that $\mathbf{r}$ and $\mathbf{p}$ are Hermitian we have

$$\mathbf{L}^\dagger \;=\; (\mathbf{r}\times\mathbf{p})^\dagger \;=\; -\mathbf{p}^\dagger\times\mathbf{r}^\dagger \;=\; -\mathbf{p}\times\mathbf{r} \;=\; \mathbf{L}\,. \tag{1.25}$$

The use of vector notation implies that, for example,

$$\mathbf{L}^2 \;=\; \mathbf{L}\cdot\mathbf{L} \;=\; \hat{L}_1\hat{L}_1 + \hat{L}_2\hat{L}_2 + \hat{L}_3\hat{L}_3 \;=\; \hat{L}_i\hat{L}_i\,. \tag{1.26}$$

The classical angular momentum operator is orthogonal to both $\vec{r}$ and $\vec{p}$ as it is built from the cross product of these two vectors. Happily, these properties also hold for the quantum angular momentum. Take for example the dot product of $\mathbf{r}$ with $\mathbf{L}$ to get

$$\mathbf{r}\cdot\mathbf{L} \;=\; \hat{x}_i\,\hat{L}_i \;=\; \hat{x}_i\epsilon_{ijk}\,\hat{x}_j\,\hat{p}_k \;=\; \epsilon_{ijk}\,\hat{x}_i\,\hat{x}_j\,\hat{p}_k \;=0\,. \tag{1.27}$$

The last expression is zero because the $\hat{x}$'s commute and thus form an object symmetric in i, j, while the epsilon symbol is antisymmetric in i, j. Similarly,

$$\mathbf{p}\cdot\mathbf{L} \;=\; \hat{p}_i\,\hat{L}_i \;=\; -\hat{p}_i(\mathbf{p}\times\mathbf{r})_i \;=\; -\hat{p}_i\epsilon_{ijk}\,\hat{p}_j\,\hat{x}_k \;=\; -\epsilon_{ijk}\,\hat{p}_i\,\hat{p}_j\,\hat{x}_k \;=0\,. \tag{1.28}$$

In summary:

$$\boxed{\;\mathbf{r}\cdot\mathbf{L} \;=\; \mathbf{p}\cdot\mathbf{L} \;=\; 0\,.\;} \tag{1.29}$$

In manipulating vector products the following identity is quite useful

$$\epsilon_{ijk}\,\epsilon_{ipq} \;=\; \delta_{jp}\delta_{kq} \;-\; \delta_{jq}\delta_{kp}\,. \tag{1.30}$$

Its contraction is also needed sometimes:

$$\epsilon_{ijk}\,\epsilon_{ijq} \;=\; 2\delta_{kq}\,. \tag{1.31}$$

For triple products we find

$$\begin{aligned}
\big[\mathbf{a}\times(\mathbf{b}\times\mathbf{c})\big]_k \;&=\; \epsilon_{kji}\,a_j\,(\mathbf{b}\times\mathbf{c})_i \\
&=\; \epsilon_{kji}\,\epsilon_{ipq}\;a_j\,b_p\,c_q \\
&=\; -\,\epsilon_{ijk}\,\epsilon_{ipq}\;a_j\,b_p\,c_q \\
&=\; -\,\big(\delta_{jp}\delta_{kq} \;-\; \delta_{jq}\delta_{kp}\big)a_j\,b_p\,c_q \\
&=\; a_j\,b_k\,c_j \;-\; a_j\,b_j\,c_k \\
&=\; [\,a_j\,,\,b_k\,]\,c_j \;+\; b_k\,a_j\;c_j \;-\; a_j\,b_j\,c_k \\
&=\; [\,a_j\,,\,b_k\,]\,c_j \;+\; b_k\,(\mathbf{a}\cdot\mathbf{c}) \;-\; (\mathbf{a}\cdot\mathbf{b})\,c_k
\end{aligned} \tag{1.32}$$

We can write this as

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) \;=\; \mathbf{b}\,(\mathbf{a}\cdot\mathbf{c}) \;-\; (\mathbf{a}\cdot\mathbf{b})\,\mathbf{c} \;+\; [\,a_j\,,\,\mathbf{b}\,]\,c_j\,. \tag{1.33}$$

The first two terms are familiar from vector analysis, the last term is quantum mechanical.

Another familiar relation from vector analysis is the classical

$$(\vec{a}\times\vec{b}\,)^2 \;\equiv\; (\vec{a}\times\vec{b}\,)\cdot(\vec{a}\times\vec{b}\,) \;=\; \vec{a}^{\,2}\,\vec{b}^{\,2} \;-\; (\vec{a}\cdot\vec{b}\,)^2\,. \tag{1.34}$$

In deriving such an equation you assume that the vector components are commuting numbers, not operators. If we have vector operators additional terms arise.

Exercise 2. Show that

$$\begin{aligned}
(\mathbf{a}\times\mathbf{b})^2 \;=\;& \quad \mathbf{a}^2\,\mathbf{b}^2 - (\mathbf{a}\cdot\mathbf{b})^2 \\
&- a_j\,[a_j,b_k]\,b_k + a_j\,[a_k,b_k]\,b_j - a_j\,[a_k,b_j]\,b_k \;-\; a_j a_k\,[b_k,b_j]\,,
\end{aligned} \tag{1.35}$$

and verify that this yields

$$(\mathbf{a}\times\mathbf{b})^2 \;=\; \mathbf{a}^2\,\mathbf{b}^2-(\mathbf{a}\cdot\mathbf{b})^2+\gamma\,\mathbf{a}\cdot\mathbf{b}\,,\qquad \text{when}\quad [a_i,b_j]=\gamma\,\delta_{ij}\,,\quad \gamma\in\mathbb{C}\,,\quad [b_i,b_j]=0\,. \tag{1.36}$$

As an application we calculate $\mathbf{L}^2$

$$\mathbf{L}^2 \;=\; (\mathbf{r}\times\mathbf{p})^2\,, \tag{1.37}$$

equation (1.36) can be applied with $\mathbf{a}=\mathbf{r}$ and $\mathbf{b}=\mathbf{p}$. Since $[a_i,b_j]=[\hat{x}_i,\hat{p}_j]=i\hbar\,\delta_{ij}$ we read that $\gamma=i\hbar$, so that

$$\boxed{\mathbf{L}^2 \;=\; \mathbf{r}^2\,\mathbf{p}^2-(\mathbf{r}\cdot\mathbf{p})^2+\,i\hbar\,\mathbf{r}\cdot\mathbf{p}\,.} \tag{1.38}$$

Another useful and simple identity is the following

$$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) \;=\; (\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}\,, \tag{1.39}$$

as you should confirm in a one-line computation. In commuting vector analysis this triple product is known to be cyclically symmetric. Note, that in the above no operator has been moved across each other –that's why it holds.

1.2 Properties of angular momentum

A key property of the angular momentum operators is their commutation relations with the $\hat{x}_i$ and $\hat{p}_i$ operators. You should verify that

$$\boxed{\begin{aligned}[\,\hat{L}_i\,,\,\hat{x}_j\,] \;&=\; i\hbar\,\epsilon_{ijk}\,\hat{x}_k\,,\\ [\,\hat{L}_i\,,\,\hat{p}_j\,] \;&=\; i\hbar\,\epsilon_{ijk}\,\hat{p}_k\,.\end{aligned}} \tag{1.40}$$

We say that these equations mean that $\mathbf{r}$ and $\mathbf{p}$ are vectors *under* rotations.

Exercise 3. Use the above relations and (1.18) to show that

$$\mathbf{p}\times\mathbf{L} \;=\; -\,\mathbf{L}\times\mathbf{p} \;+\; 2i\hbar\,\mathbf{p}\,. \tag{1.41}$$

Hermitization is the process by which we construct a Hermitian operator starting from a non-Hermitian one. Say Ω is not hermitian, its Hermitization Ω_h is defined to be

$$\Omega_h \;\equiv\; \frac{1}{2}(\Omega+\Omega^\dagger)\,. \tag{1.42}$$

Exercise 4. Show that the Hermitization of $\mathbf{p}\times\mathbf{L}$ is

$$(\mathbf{p}\times\mathbf{L})_h \;=\; \frac{1}{2}\big(\mathbf{p}\times\mathbf{L}\;-\;\mathbf{L}\times\mathbf{p}\big) \;=\; \mathbf{p}\times\mathbf{L}\;-\;i\hbar\,\mathbf{p}\,. \tag{1.43}$$

Inspired by the behavior of $\mathbf{r}$ and $\mathbf{p}$ under rotations, we declare that an operator $\mathbf{u}$ defined by the triplet $(\hat{u}_1, \hat{u}_2, \hat{u}_3)$ is a *vector under rotations* if

$$[\,\hat{L}_i\,,\,\hat{u}_j\,] \;=\; i\hbar\,\epsilon_{ijk}\,\hat{u}_k\,. \tag{1.44}$$

Exercise 5. Use the above commutator to show that with $\mathbf{n}$ a constant vector,

$$\big[\,\mathbf{n}\cdot\mathbf{L}\,,\,\mathbf{u}\,\big] \;=\; -i\hbar\;\mathbf{n}\times\mathbf{u}\,. \tag{1.45}$$

Given two operators $\mathbf{u}$ and $\mathbf{v}$ that are vectors under rotations you will show that their dot product is a scalar –it commutes with all $\hat{L}_i$– and their cross product is a vector:

$$\boxed{\begin{aligned}[\,\hat{L}_i\,,\mathbf{u}\cdot\mathbf{v}\,] \;&=\; 0\,,\\ [\,\hat{L}_i\,,(\mathbf{u}\times\mathbf{v})_j\,] \;&=\; i\hbar\,\epsilon_{ijk}\,(\mathbf{u}\times\mathbf{v})_k\,.\end{aligned}} \tag{1.46}$$

Exercise 6. Prove the above equations.

A number of useful commutator identities follow from (1.46). Most importantly, from the second one, taking $\mathbf{u}=\mathbf{r}$ and $\mathbf{v}=\mathbf{p}$ we get

$$[\,\hat{L}_i\,,(\mathbf{r}\times\mathbf{p})_j\,] \;=\; i\hbar\,\epsilon_{ijk}\,(\mathbf{r}\times\mathbf{p})_k\,, \tag{1.47}$$

which gives the celebrated Lie algebra of angular momentum[1]

$$\boxed{[\,\hat{L}_i\,,\hat{L}_j\,] \;=\; i\hbar\,\epsilon_{ijk}\,\hat{L}_k\,.} \tag{1.48}$$

More explicitly,

$$\begin{aligned}\big[\,\hat{L}_x\,,\,\hat{L}_y\,\big] \;&=\; i\hbar\,\hat{L}_z\,,\\ \big[\,\hat{L}_y\,,\,\hat{L}_z\,\big] \;&=\; i\hbar\,\hat{L}_x\,,\\ \big[\,\hat{L}_z\,,\,\hat{L}_x\,\big] \;&=\; i\hbar\,\hat{L}_y\,.\end{aligned} \tag{1.49}$$

Note the cyclic nature of these equations: take the first and cycle indices $(x\to y\to z\to x)$ to obtain the other two. Another set of examples follows from the first identity in (1.46) using our list of vector operators. For example

$$[\,\hat{L}_i\,,\mathbf{r}^2\,] \;=\; [\,\hat{L}_i\,,\mathbf{p}^2\,] \;=\; [\,\hat{L}_i\,,\mathbf{r}\cdot\mathbf{p}\,] \;=\; 0\,, \tag{1.50}$$

[1]We do not know who celebrated it because we were not invited.

and, very importantly,

$$[\,\hat{L}_i\,,\,\mathbf{L}^2\,]\;=\;0\,. \tag{1.51}$$

This equation is the reason the operator $\mathbf{L}^2$ plays a very important role in the study of central potentials. $\mathbf{L}^2$ will feature as one of the operators in complete sets of commuting observables. An operator, such as $\mathbf{L}^2$, that commutes with all the angular momentum operators is called a "Casimir" of the algebra of angular momentum. Note that the validity of (1.51) just uses the algebra of the $\hat{L}_i$ operators not, for example, how they are built from $\mathbf{r}$ and $\mathbf{p}$.

Exercise 7. Use (1.18) and the algebra of $\hat{L}$ operators to show that

$$\mathbf{L}\times\mathbf{L}\;=\;i\hbar\,\mathbf{L}\,. \tag{1.52}$$

This is a very elegant way to express the algebra of angular momentum. In fact, we can show that it is totally equivalent to (1.48). Thus we write

$$\boxed{\mathbf{L}\times\mathbf{L}\;=\;i\hbar\,\mathbf{L}\quad\Longleftrightarrow\quad[\,\hat{L}_i\,,\hat{L}_j\,]\;=\;i\hbar\,\epsilon_{ijk}\,\hat{L}_k\,.} \tag{1.53}$$

Commutation relations of the form

$$[\,a_i\,,b_j\,]\;=\;\epsilon_{ijk}\,c_k\,, \tag{1.54}$$

admit a natural rewriting in terms of cross products. From (1.18)

$$(\mathbf{a}\times\mathbf{b})_i\;+\;(\mathbf{b}\times\mathbf{a})_i\;=\;\epsilon_{ijk}\,[a_j\,,\,b_k\,]\;=\;\epsilon_{ijk}\epsilon_{jkp}c_p\;=\;2\,c_i\,. \tag{1.55}$$

This means that

$$\boxed{[\,a_i\,,b_j\,]\;=\;\epsilon_{ijk}\,c_k\quad\rightarrow\quad\mathbf{a}\times\mathbf{b}\;+\;\mathbf{b}\times\mathbf{a}\;=\;2\,\mathbf{c}\,.} \tag{1.56}$$

The arrow does *not* work in the reverse direction. One finds $[\,a_i\,,b_j\,]\;=\;\epsilon_{ijk}\,c_k+s_{ij}$ where $s_{ij}=s_{ji}$ is arbitrary and is *not* determined. If the arrow could be reversed, $\mathbf{a}\times\mathbf{b}\;+\;\mathbf{b}\times\mathbf{a}=0$ would imply that $\mathbf{a}$ and $\mathbf{b}$ commute. We have, however, a familiar example where this does not happen: while $\mathbf{r}\times\mathbf{p}+\mathbf{p}\times\mathbf{r}=0$ (see (1.23)), the operators $\mathbf{r}$ and $\mathbf{p}$ don't commute.

For a vector $\mathbf{u}$ under rotations, equation (1.56) becomes

$$\mathbf{L}\times\mathbf{u}\;+\;\mathbf{u}\times\mathbf{L}\;=\;2\,i\hbar\,\mathbf{u}\,. \tag{1.57}$$

1.3 The central potential Hamiltonian

Angular momentum plays a crucial role in the study of three-dimensional central potential problems. Those are problems where the Hamiltonian describes a particle moving in a potential $V(r)$ that depends just on r, the distance of the particle to the chosen origin. The Hamiltonian takes the form

$$H \;=\; \frac{\mathbf{p}^2}{2m} + V(r)\,. \tag{1.58}$$

When writing the Schrödinger equation in position space we identify

$$\mathbf{p} \;=\; \frac{\hbar}{i}\,\nabla\,, \tag{1.59}$$

and therefore

$$\mathbf{p}^2 \;=\; -\,\hbar^2\,\nabla^2\,, \tag{1.60}$$

where ∇^2 denotes the Laplacian operator –a second order differential operator. In spherical coordinates the Laplacian is well known and gives us

$$\mathbf{p}^2 \;=\; -\,\hbar^2\left[\;\frac{1}{r}\,\frac{\partial^2}{\partial r^2}\,r \;+\; \frac{1}{r^2}\Big(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\,\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\Big)\;\right]. \tag{1.61}$$

Our goal is to relate the "angular" part of the above differential operator to angular momentum operators. This will be done by calculating $\mathbf{L}^2$ and relating it to $\mathbf{p}^2$. Since we had from (1.38)

$$\mathbf{L}^2 \;=\; \mathbf{r}^2\,\mathbf{p}^2 - (\mathbf{r}\cdot\mathbf{p})^2 + \,i\hbar\,\mathbf{r}\cdot\mathbf{p}\,, \tag{1.62}$$

We solve for $\mathbf{p}^2$ to get

$$\mathbf{p}^2 \;=\; \frac{1}{\mathbf{r}^2}\left[(\mathbf{r}\cdot\mathbf{p})^2 - \,i\hbar\,\mathbf{r}\cdot\mathbf{p} + \mathbf{L}^2\right]. \tag{1.63}$$

Let us now consider the above equation in coordinate space, where $\mathbf{p}$ is a gradient. We then have:

$$\mathbf{r}\cdot\mathbf{p} \;=\; \frac{\hbar}{i}\;r\frac{\partial}{\partial r}\,, \tag{1.64}$$

and thus

$$(\mathbf{r}\cdot\mathbf{p})^2 - \,i\hbar\,\mathbf{r}\cdot\mathbf{p} \;=\; -\hbar^2\left(r\frac{\partial}{\partial r}r\frac{\partial}{\partial r} + r\frac{\partial}{\partial r}\right) \;=\; -\hbar^2\left(r^2\frac{\partial^2}{\partial r^2} + 2r\frac{\partial}{\partial r}\right). \tag{1.65}$$

It then follows that

$$\frac{1}{\mathbf{r}^2}\left[(\mathbf{r}\cdot\mathbf{p})^2 - \,i\hbar\,\mathbf{r}\cdot\mathbf{p}\right] \;=\; -\hbar^2\left(\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r}\right) \;=\; -\hbar^2\,\frac{1}{r}\,\frac{\partial^2}{\partial r^2}\;r\,, \tag{1.66}$$

where the last step is readily checked by explicit expansion. Back in (1.63) we get

$$\mathbf{p}^2 \;=\; -\hbar^2\,\frac{1}{r}\,\frac{\partial^2}{\partial r^2}\;r \;+\; \frac{1}{r^2}\,\mathbf{L}^2\,. \tag{1.67}$$

Comparing with (1.61) we identify $\mathbf{L}^2$ as the operator

$$\boxed{\mathbf{L}^2 \;=\; -\hbar^2\Big(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\,\sin\theta\frac{\partial}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\Big)\,.} \tag{1.68}$$

Note that the units are fully carried by the $\hbar^2$ in front and that the differential operator is completely angular: it has no radial dependence. Given our expression (1.67) for $\mathbf{p}^2$ we can now rewrite the three-dimensional Hamiltonian as

$$\boxed{H \;=\; \frac{\mathbf{p}^{\,2}}{2m}+V(r) \;=\; -\frac{\hbar^2}{2m}\,\frac{1}{r}\,\frac{\partial^2}{\partial r^2}\,r+\frac{1}{2mr^2}\mathbf{L}^2+V(r)\,.} \tag{1.69}$$

A key property of central potential problems is that the angular momentum operators commute with the Hamiltonian

$$\text{Central potential Hamiltonians:}\quad [\,\hat{L}_i\,,H\,] \;=\; 0\,. \tag{1.70}$$

We have seen that $\hat{L}_i$ commutes with $\mathbf{p}^2$ so it is only needed to show that the $\hat{L}_i$ commute with $V(r)$. This is eminently reasonable, for $\hat{L}_i$ commutes with $\mathbf{r}^2=r^2$, so one would expect it to commute with any function of $\sqrt{\mathbf{r}^2}=r$. In the problem set you will consider this question and develop a formal argument that confirms the expectation.

The above commutator implies that the $\hat{L}_i$ operators are conserved in central potentials. Indeed

$$i\hbar\frac{d}{dt}\,\langle\hat{L}_i\rangle \;=\; \langle\,[\,\hat{L}_i\,,H\,]\,\rangle \;=\; 0\,. \tag{1.71}$$

We can now consider the issue of complete sets of commuting observables. The list of operators that we have is

$$H,\quad \hat{x}_1,\hat{x}_2,\hat{x}_3,\quad \hat{p}_1,\hat{p}_2,\hat{p}_3,\quad \hat{L}_1,\hat{L}_2,\hat{L}_3\,,\quad \mathbf{r}^2,\ \mathbf{p}^2,\ \mathbf{r}\cdot\mathbf{p}\,,\ \mathbf{L}^2\,,\ldots \tag{1.72}$$

where, for the time being, we included all operators up to squares of coordinates, momenta, and angular momenta. Since we want to understand the spectrum of the Hamiltonian, one of the labels of states will be the energy and thus H must be in the list of commuting observables. Because of the potential $V(r)$ none of the $\hat{p}_i$ operators commute the Hamiltonian. Because of the $\mathbf{p}^2$ term in the Hamiltonian none of the $\hat{x}_i$ commute with the Hamiltonian. Nor will $\mathbf{r}^2,\mathbf{p}^2$ and $\mathbf{r}\cdot\mathbf{p}$. The list is thus reduced to

$$H,\quad \hat{L}_1,\hat{L}_2,\hat{L}_3\,,\ \mathbf{L}^2\,. \tag{1.73}$$

where there are no dots anymore, since without $\hat{x}_i$ or $\hat{p}_i$ there are no other operators to build (recall also that $\mathbf{L}\times\mathbf{L}=i\hbar\mathbf{L}$ and thus it is not new). All the operators in the list commute

with H: the $\hat{L}_i$ as discussed in (1.70), and $\mathbf{L}^2$ because, after all, it is built from $\hat{L}_i$. But all the operators do not commute with each other. From the $\hat{L}_i$ we can only pick at most one, for then the other two necessarily do not commute with the chosen one. Happily we can also keep $\mathbf{L}^2$ because of its Casimir property (1.51). Conventionally, everybody chooses $\hat{L}_3 = \hat{L}_z$ as one element of the set of commuting observables. Thus we have

$$\boxed{\text{Commuting observables:} \quad H\,,\ \hat{L}_z\,,\ \mathbf{L}^2\,.} \tag{1.74}$$

We can wonder if this set is complete in the sense that all energy eigenstates are uniquely labelled by the eigenvalues of the above operators. The answer is yes, for the bound state spectrum of a particle that has no other degrees of freedom (say, no spin).

2 Algebraic theory of angular momentum

Hermitian operators $\hat{J}_x, \hat{J}_y, \hat{J}_z$ are said to satisfy the algebra of **angular momentum** if the following commutation relations:

$$[\hat{J}_i\,,\ \hat{J}_j] \ = \ i\hbar\,\epsilon_{ijk}\,\hat{J}_k\,. \tag{2.1}$$

More explicitly, in components

$$\begin{aligned} [\hat{J}_x, \hat{J}_y] \ &= \ i\hbar\hat{J}_z \\ [\hat{J}_y, \hat{J}_z] \ &= \ i\hbar\hat{J}_x \\ [\hat{J}_z, \hat{J}_x] \ &= \ i\hbar\hat{J}_y\,. \end{aligned} \tag{2.2}$$

The $\hat{J}_i$ operators could be $\hat{L}_i$, or $\hat{S}_i$ or something else! Will only use this algebra and the Hermiticity of the operators. From this algebra it also follows that

$$[\hat{J}_i\,, \mathbf{J}^2\,] \ = \ 0\,. \tag{2.3}$$

This can be checked explicitly, but our proof of the analogous result (1.51): $[\hat{L}_i, \mathbf{L}^2] = 0$ only used the algebra of the operators $\hat{L}_i$, so this also holds for the $\hat{J}_i$, which satisfy the same algebra. It is not convenient to define

$$\boxed{\begin{aligned} \hat{J}_+ \ &\equiv \ \hat{J}_x + i\hat{J}_y\,, \\ \hat{J}_- \ &\equiv \ \hat{J}_x - i\hat{J}_y\,, \end{aligned}} \tag{2.4}$$

such that the two operators are Hermitian conjugates of each other:

$$(\hat{J}_+)^\dagger \ = \ \hat{J}_-\,. \tag{2.5}$$

Note that both $\hat{J}_x$ and $\hat{J}_y$ can be solved for in terms of $\hat{J}_+$ and $\hat{J}_-$. It is useful to compute the algebra of the operators $\hat{J}_+, \hat{J}_-$, and $\hat{J}_z$. We begin by computing the product $\hat{J}_+\hat{J}_-$:

$$\hat{J}_+\hat{J}_- = \hat{J}_x^2 + \hat{J}_y^2 - i[\hat{J}_x\,,\,\hat{J}_y] = \hat{J}_x^2 + \hat{J}_y^2 + \hbar\hat{J}_z\,. \tag{2.6}$$

Together with the product in the opposite order we have

$$\begin{aligned} \hat{J}_+\hat{J}_- &= \hat{J}_x^2 + \hat{J}_y^2 + \hbar\hat{J}_z\,, \\ \hat{J}_-\hat{J}_+ &= \hat{J}_x^2 + \hat{J}_y^2 - \hbar\hat{J}_z\,. \end{aligned} \tag{2.7}$$

From these two we can quickly get the commutator:

$$[\hat{J}_+\,,\,\hat{J}_-] \;=\; 2\hbar\hat{J}_z\,. \tag{2.8}$$

Moreover, we obtain two expressions for $\hat{J}_x^2 + \hat{J}_y^2$

$$\hat{J}_x^2 + \hat{J}_y^2 \;=\; \hat{J}_+\hat{J}_- - \hbar\hat{J}_z \;=\; \hat{J}_-\hat{J}_+ + \hbar\hat{J}_z\,. \tag{2.9}$$

Adding $\hat{J}_z^2$ to both sides of the equation we find

$$\boxed{\mathbf{J}^2 \;=\; \hat{J}_+\hat{J}_- + \hat{J}_z^2 - \hbar\hat{J}_z \;=\; \hat{J}_-\hat{J}_+ + \hat{J}_z^2 + \hbar\hat{J}_z\,.} \tag{2.10}$$

Of course, since $\hat{J}_i$ and $\mathbf{J}^2$ commute, we also have

$$[\,\hat{J}_\pm\,,\,\mathbf{J}^2\,] \;=\; 0\,. \tag{2.11}$$

We finally have to compute the commutator of $\hat{J}_\pm$ with $\hat{J}_z$. This is quickly done:

$$[\hat{J}_z, \hat{J}_+] \;=\; [\hat{J}_z, \hat{J}_x] + i[\hat{J}_z\,,\,\hat{J}_y] = \; i\hbar\hat{J}_y + i(-i\hbar\hat{J}_x) \;=\; \hbar(\hat{J}_x + i\hat{J}_y) \;=\; \hbar\hat{J}_+\,. \tag{2.12}$$

Similarly, $[\hat{J}_z, \hat{J}_-] = -\hbar\hat{L}_-$ and therefore, all in all

$$\boxed{[\hat{J}_z, \hat{J}_\pm\,] \;=\; \pm\hbar\,\hat{J}_\pm\,.} \tag{2.13}$$

This is similar to our harmonic oscillator commutators $[\hat{N}, \hat{a}^\dagger] = \hat{a}^\dagger$ and $[N, \hat{a}] = -\hat{a}$, if we identify $\hat{N}$ with $\hat{J}_z$, $\hat{a}^\dagger$ with $\hat{J}_+$ and $\hat{a}$ with $\hat{J}_-$. In the oscillator case we learned from these that, acting on states, $\hat{a}^\dagger$ raises the $\hat{N}$ eigenvalue by one unit while $\hat{a}$ decreases it by one unit. As we will see, $\hat{J}_+$ adds $\hbar$ to the $\hat{J}_z$ eigenvalue and $\hat{J}_-$ subtracts $\hbar$ to the $\hat{J}_z$ eigenvalue.

Since $\mathbf{J}^2$ and $\hat{J}_z$ are hermitian and commute, they can be simultaneously diagonalized. In fact, there are no more operators in the angular momentum algebra can be added to this list of simultaneously diagonalizable operators. The common eigenstates form an orthonormal basis

for the relevant vector space. We thus introduce eigenstates $|j,m\rangle$, with $j,m \in \mathbb{R}$, where the first label relates to the $\mathbf{J}^2$ eigenvalue and the second label to the $\hat{J}_z$ eigenvalue:

$$\begin{aligned} \mathbf{J}^2|j,m\rangle &= \hbar^2\, j(j+1)\,|j,m\rangle\,, \\ \hat{J}_z\,|j,m\rangle &= \hbar m\,|j,m\rangle\,. \end{aligned} \tag{2.14}$$

The orthonormality of states implies that

$$\langle j',m'|j,m\rangle = \delta_{j',j}\delta_{m',m}\,, \tag{2.15}$$

where we assumed that we will not have to deal with continuous values of j,m that would require delta function normalization. This will be confirmed below. Since j and m are real, the eigenvalues of the hermitian operators are real, as they have to be. The first line shows that the eigenvalue of $\mathbf{J}^2$ is defined to be $\hbar^2 j(j+1)$. This can seem curious: why not $\hbar^2 j^2$? The answer is convenience, as we will see below. Alternatively, if we know $\hbar^2 j(j+1)$, how do we get j? For this first note that $\hbar^2 j(j+1)$ must be non-negative:

$$\hbar^2\, j(j+1) = \langle j,m|\mathbf{J}^2|j,m\rangle = \sum_{i=1}^{3}\langle j,m|\hat{J}_i\hat{J}_i|j,m\rangle = \sum_{i=1}^{3}||\hat{J}_i|j,m\rangle||^2 \geq 0\,, \tag{2.16}$$

where in the first step we used the eigenvalue definition and orthonormality. Therefore the condition

$$j(j+1) \geq 0 \tag{2.17}$$

is the only a priori condition on the values of j. Since what matters is the eigenvalue of $\mathbf{J}^2$ we can use any of the two j's that give a particular value of $j(j+1)$. As shown in the figure below, the positivity of $j(j+1)$ requires $j \geq 0$ or $j \leq -1$. We can simply use $j \geq 0$.

$$\text{States are labeled as } \; |j,m\rangle \; \text{ with } \; j \geq 0\,. \tag{2.18}$$

You should not think that there are two different states, with two different j's associated with the eigenvalue $\hbar j(j+1)$. It is just one state, that we are labeling in an unusual way. Of course, a theory may end up having more than one state with the same $\mathbf{J}^2$ eigenvalue. In that case we will have more than one state with the same $j > 0$.

Let us now investigate what the operators $\hat{J}_\pm$ do when acting on the above eigenstates. Since they commute with $\mathbf{J}^2$, the operators J_+ or J_- do not change the j value of a state:

$$\mathbf{J}^2(\hat{J}_\pm|j,m\rangle) = \hat{J}_\pm\mathbf{J}^2|j,m\rangle = \hbar j(j+1)(\hat{J}_\pm|j,m\rangle)\,, \tag{2.19}$$

so that we must have

$$J_\pm|j,m\rangle \propto |j,m'\rangle\,, \quad \text{for some } m'\,. \tag{2.20}$$

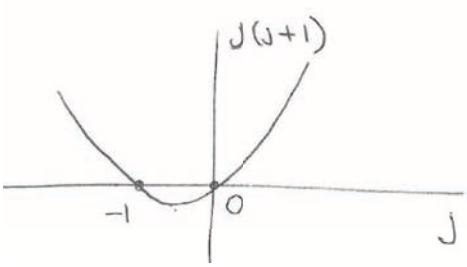


Figure 1: Since $j(j+1) \geq 0$ for consistency, we can label the states $|j,m\rangle$ using $j \geq 0$.

On the other hand, as anticipated above, the $\hat{J}_\pm$ operators change the value of m:

$$\begin{aligned} \hat{J}_z(\hat{J}_\pm|j,m\rangle) &= ([\hat{J}_z, J_\pm] + J_\pm \hat{J}_z)|j,m\rangle \\ &= (\pm\hbar J_\pm + \hbar m J_\pm)|j,m\rangle \\ &= \hbar(m \pm 1)(J_\pm|j,m\rangle)\,, \end{aligned} \tag{2.21}$$

from which we learn that

$$\hat{J}_\pm|j,m\rangle \;=\; C_\pm(j,m)|j,m\pm 1\rangle\,, \tag{2.22}$$

where $C_\pm(j,m)$ is a constant to be determined. Indeed, we see that $\hat{J}_+$ raised the m eigenvalue by one unit while $\hat{J}_-$ decreases the m eigenvalue by one unit. To determine $C_\pm(j,m)$ we first take the adjoint of the above equation

$$\langle j,m|\hat{J}_\mp \;=\; \langle j,m\pm 1|C_\pm(j,m)^*\,, \tag{2.23}$$

and then form the overlap

$$\langle j,m|J_\mp J_\pm|j,m\rangle = |C_\pm(j,m)|^2\,. \tag{2.24}$$

To evaluate the left-hand side use (2.10) in the form

$$\hat{J}_\mp\hat{J}_\pm \;=\; \mathbf{J}^2 - \hat{J}_z^2 \mp \hbar\hat{J}_z \tag{2.25}$$

as well as $\langle j,m|j,m\rangle = 1$:

$$|C_\pm(j,m)|^2 \;=\; \langle j,m|\big(\mathbf{J}^2 - \hat{J}_z^2 \mp \hbar\hat{J}_z\big)|j,m\rangle \;=\; \hbar^2 j(j+1) - \hbar^2 m^2 \mp \hbar^2 m\,. \tag{2.26}$$

We have thus found that

$$\boxed{|C_\pm(j,m)|^2 \;=\; \hbar^2\left(j(j+1) - m(m\pm 1)\right) \;=\; ||\hat{J}_\pm|j,m\rangle||^2\,.} \tag{2.27}$$

Here we learn a few things. If we start with a consistent state $|j,m\rangle$ of norm one (as assumed above), the states $\hat{J}_\pm|j,m\rangle \sim |j,m\pm 1\rangle$ created by the action of $\hat{J}_\pm$ on $|j,m\rangle$ are inconsistent if

the middle expression in the above relation is negative. This is because that middle expression is in fact the norm-squared of $\hat{J}_\pm|j,m\rangle$. Assuming that middle expression is positive (or zero) we can take $C_\pm(j,m)$ real and equal to the positive square root

$$C_\pm(j,m) \;=\; \hbar\sqrt{j(j+1)-m(m\pm1)}\,. \tag{2.28}$$

We have thus obtained

$$\boxed{J_\pm|j,m\rangle \;=\; \hbar\sqrt{j(j+1)-m(m\pm1)}\;|j,m\pm1\rangle\,.} \tag{2.29}$$

Given a consistent state $|j,m\rangle$, how far can we raise or lower the value of m? Our classical intuition is that $|\hat{J}_z| \le |\mathbf{J}|$. So we should get something like $|m| \lesssim \sqrt{j(j+1)}$.

Consider this in two steps:

1. For the raised state to be consistent we must have $||J_+|j,m\rangle||^2 \ge 0$ and therefore

$$j(j+1)-m(m+1)\ge 0 \quad\rightarrow\quad m(m+1)\le j(j+1) \tag{2.30}$$

The solution to the inequality is given in figure 2:

$$-j-1 \;\le m\; \le j\,. \tag{2.31}$$

Had we not chosen $\hbar^2 j(j+1)$ to be the eigenvalue of $\mathbf{J}^2$ this inequality would not have had a simple solution.

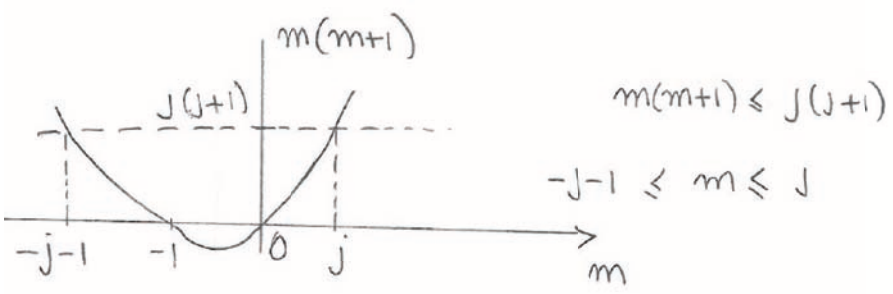


Figure 2: Solving the inequality $m(m+1)\le j(j+1)$.

Since we are raising m we can focus on the troubles that raising can give given that $m \le j$. Assume $m = j-\beta$ with $0<\beta<1$ so that the inequality (2.31) is satisfied and m is less than one unit below j. Then the raising once gives us a state with $m' = m+1 > j$ and since the inequality is now violated raising one more time would then give an inconsistent

state. To prevent such inconsistency the process of raising must terminate: there must be a state that raising gives no state (the zero state). That indeed happens only if $m = j$ since then $C_+(j, j) = 0$

$$\hat{J}_+|j, j\rangle \;=\; 0\,. \tag{2.32}$$

2. For the lowered state to be consistent we must have $||\hat{J}_-|j, m\rangle||^2 \geq 0$ and therefore

$$j(j+1) - m(m-1) \geq 0 \quad \rightarrow \quad m(m-1) \leq\; j(j+1) \tag{2.33}$$

The solution to this inequality is obtained using figure 3 and gives

$$-j \;\leq m \;\leq j+1\,. \tag{2.34}$$

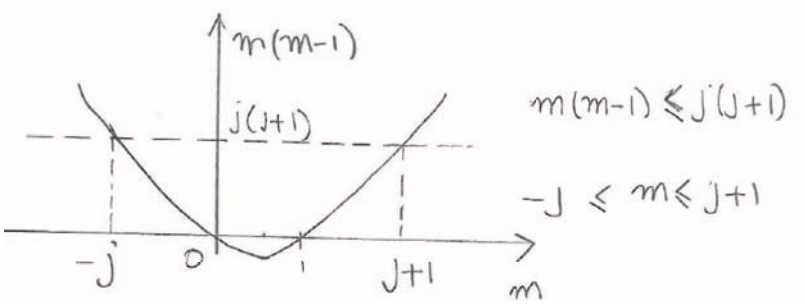


Figure 3: Solving the inequality $m(m-1) \leq j(j-1)$.

This time we can focus here on $m \geq -j$ and the complications due to lowering. Assume $m = -j + \beta$ with $0 < \beta < 1$ so that the constraint (2.34) is satisfied and m is less than one unit above $-j$. Then lowing once gives us a state with $m' = m - 1 < -j$ and since the inequality is now violated lowering one more time would then give an inconsistent state. To prevent such inconsistency we need that lowering terminates on some state for which lowering gives no state (the zero state). That indeed happens only if $m = -j$ since then $C_-(j, -j) = 0$

$$\hat{J}_-|j, -j\rangle \;=\; 0\,. \tag{2.35}$$

The above analysis shows that for consistency a multiplet of states with some given fixed j must be such that the m values must include $-j$ and $+j$. Since m is increased or decreased by integer steps via the $\hat{J}_\pm$ operators, the distance $2j$ between j and $-j$ must be an integer:

$$\boxed{2j \in \mathbb{Z} \quad \rightarrow \quad j \in \mathbb{Z}/2\,, \quad \rightarrow \quad j = 0,\, \tfrac{1}{2},\, 1,\, \tfrac{3}{2},\, 2, \ldots\,.} \tag{2.36}$$

This is the fundamental quantization of angular momentum. Angular momentum can be integral or half-integral. For any allowed value of j the m values will be $j, j-1, \ldots, -j$. Thus the multiplet with angular momentum j has the following $2j+1$ states

$$\begin{aligned} &|j,j\rangle, \\ &|j,j-1\rangle, \\ &\quad\vdots \\ &|j,-j\rangle\,. \end{aligned} \tag{2.37}$$

For $j=0$ there is just one state, the **singlet** with $m=0$: $|0,0\rangle$.
For $j=\frac{1}{2}$ we have two states:

$$\begin{aligned} &|\tfrac{1}{2},\tfrac{1}{2}\rangle\,, \\ &|\tfrac{1}{2},-\tfrac{1}{2}\rangle\,. \end{aligned} \tag{2.38}$$

These are the states of a spin-1/2 particle, when we identify the angular momentum $\mathbf{J}$ with the spin angular momentum operator $\mathbf{S}$. The top state has $\hat{S}_z = \hbar/2$ and the lower state has $\hat{S}_z = -\hbar/2$. These are our conventional $|+\rangle$ and $|-\rangle$ states, respectively.
For $j=1$ we have three states :

$$\begin{aligned} &|1,1\rangle\,, \\ &|1,0\rangle\,, \\ &|1,-1\rangle\,. \end{aligned} \tag{2.39}$$

For $j=3/2$ we have four states:

$$\begin{aligned} &|\tfrac{3}{2},\tfrac{3}{2}\rangle \\ &|\tfrac{3}{2},\tfrac{1}{2}\rangle \\ &|\tfrac{3}{2},-\tfrac{1}{2}\rangle \\ &|\tfrac{3}{2},-\tfrac{3}{2}\rangle \end{aligned} \tag{2.40}$$

One last one! For $j=2$ we have five states:

$$\begin{aligned} &|2,2\rangle\,, \\ &|2,1\rangle\,, \\ &|2,0\rangle\,, \\ &|2,-1\rangle\,, \\ &|2,-2\rangle\,. \end{aligned} \tag{2.41}$$

On any state of a multiplet with angular momentum j the eigenvalue J^2 of $\mathbf{J}^2$ is

$$J^2 \;=\; \hbar^2 j(j+1) \quad\rightarrow\quad \frac{1}{\hbar}\,J \;=\; \sqrt{j(j+1)} \tag{2.42}$$

In the limit as j is large

$$\frac{1}{\hbar}\,J \;=\; j\sqrt{1+\frac{1}{j}} \;\simeq\; j+\frac{1}{2}+\mathcal{O}(1/j)\,. \tag{2.43}$$

So for large j the angular momentum is roughly $J \simeq j$.

3 Comments on spherical harmonics

We have constructed the $\mathbf{L}^2$ operator as a differential operator in position space, with coordinates r, θ, ϕ. The operator happens to depend only on θ and ϕ and takes the form (1.68)

$$\mathcal{L}^2 \;=\; -\,\hbar^2\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\,\sin\theta\frac{\partial}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right). \tag{3.1}$$

where we denoted it with a calligraphic symbol to make it clear we are talking about a differential operator. We also have, with the same notation

$$\hat{\mathcal{L}}_z \;=\; \frac{\hbar}{i}\Big(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\Big)\,. \tag{3.2}$$

A short calculation, passing to spherical coordinates (do it!) shows that

$$\hat{\mathcal{L}}_z \;=\; \frac{\hbar}{i}\,\frac{\partial}{\partial\phi}\,. \tag{3.3}$$

Finally, a longer calculation shows that

$$\mathcal{L}_\pm \;=\; \hbar e^{\pm i\phi}\Big(i\cot\theta\frac{\partial}{\partial\phi}\pm\frac{\partial}{\partial\theta}\Big)\,. \tag{3.4}$$

Recall now how things work for coordinate representations. For a single coordinate x we had that the operator $\hat{p}$ can be taken out of the matrix element as the differential operator $\mathfrak{p}$:

$$\langle x|\hat{p}|\psi\rangle \;=\; \mathfrak{p}\langle x|\psi\rangle\,,\quad \text{where}\;\; \mathfrak{p} \;=\; \frac{\hbar}{i}\frac{\partial}{\partial x} \tag{3.5}$$

We will let $\langle\theta\phi|$ denote position states on the unit sphere and the spherical harmonic $Y_{\ell m}$ will be viewed as the wavefunction for the state $|\ell m\rangle$ so that

$$Y_{\ell m}(\theta,\phi) \;\equiv\; \langle\theta\phi|\ell,m\rangle\,. \tag{3.6}$$

Consider now (2.14) in the form

$$\begin{aligned}\mathbf{L}^2|\ell,m\rangle \;&=\; \hbar^2\,\ell(\ell+1)\,|\ell,m\rangle\,,\\ \hat{L}_z\,|\ell,m\rangle \;&=\; \hbar m\,|\ell,m\rangle\,.\end{aligned} \tag{3.7}$$

Letting the bra $\langle\theta\phi|$ act on them we have

$$\begin{aligned}\langle\theta\phi|\mathbf{L}^2|\ell,m\rangle &= \hbar^2\,\ell(\ell+1)\,\langle\theta\phi|\ell,m\rangle\,,\\ \langle\theta\phi|\hat{L}_z\,|\ell,m\rangle &= \hbar m\,\langle\theta\phi|\ell,m\rangle\,.\end{aligned} \tag{3.8}$$

Using the analog of (3.5) for our operators we have

$$\begin{aligned}\mathcal{L}^2\langle\theta\phi|\ell,m\rangle &= \hbar^2\,\ell(\ell+1)\,\langle\theta\phi|\ell,m\rangle\,,\\ \hat{\mathcal{L}}_z\langle\theta\phi|\ell,m\rangle &= \hbar m\,\langle\theta\phi|\ell,m\rangle\,.\end{aligned} \tag{3.9}$$

These are equivalent to

$$\boxed{\begin{aligned}\mathcal{L}^2\,Y_{\ell m}(\theta,\phi) &= \hbar^2\,\ell(\ell+1)\,Y_{\ell m}(\theta,\phi)\,,\\ \hat{\mathcal{L}}_z\,Y_{\ell m}(\theta,\phi) &= \hbar m\,Y_{\ell m}(\theta,\phi)\,.\end{aligned}} \tag{3.10}$$

where $\mathcal{L}^2$ and $\hat{\mathcal{L}}_z$ are the coordinate representation operators for $\mathbf{L}^2$ and $\hat{L}_z$ respectively.

On the unit sphere the measure of integration is $\sin\theta d\theta d\phi$ so we postulate that the completeness relation for the $|\theta\phi\rangle$ position states reads

$$\int_0^\pi d\theta\sin\theta\int_0^{2\pi}d\phi\,|\theta\phi\rangle\langle\theta\phi| \;=\; \mathbf{1}\,. \tag{3.11}$$

The integral will be written more briefly as

$$\int d\Omega\,|\theta\phi\rangle\langle\theta\phi| \;=\; \mathbf{1} \tag{3.12}$$

where

$$\int d\Omega \;=\; \int_0^\pi d\theta\sin\theta\int_0^{2\pi}d\phi \;=\; -\int_1^{-1}d(\cos\theta)\int_0^{2\pi}d\phi \;=\; \int_{-1}^{1}d(\cos\theta)\int_0^{2\pi}d\phi\,. \tag{3.13}$$

Our orthogonality relation

$$\langle\ell',m'|\ell,m\rangle \;=\; \delta_{\ell',l}\delta_{m',m}\,, \tag{3.14}$$

gives, by including a complete set of position states

$$\int d\Omega\,\langle\ell',m'|\theta\phi\rangle\langle\theta\phi|\ell,m\rangle \;=\; \delta_{\ell',l}\delta_{m',m}\,. \tag{3.15}$$

This gives the familiar orthogonality property of the spherical harmonics:

$$\int d\Omega\;Y^*_{\ell'm'}(\theta,\phi)\,Y_{\ell m}(\theta,\phi) \;=\; \delta_{\ell',l}\delta_{m',m}\,. \tag{3.16}$$

Note that the equation

$$\hat{\mathcal{L}}_z Y_{\ell m} \;=\; \hbar m\,Y_{\ell m} \tag{3.17}$$

together with (3.3) implies that

$$Y_{\ell m}(\theta,\phi) \;=\; P_{\ell m}(\theta)\, e^{im\phi}\,. \tag{3.18}$$

The ϕ dependence of the spherical harmonics is very simple indeed!

One can show that for spherical harmonics, which are related to **orbital** angular momentum, one can only have integer ℓ. While j can be half-integral, any attempt to define spherical harmonics for half-integral ℓ fails. You will indeed show in the homework that this is necessarily the case.

4 The radial equation

Recall that from (1.69) we have

$$H \;=\; \frac{\mathbf{p}^2}{2m} + V(r) \;=\; -\frac{\hbar^2}{2m}\,\frac{1}{r}\frac{\partial^2}{\partial r^2}\,r + \frac{1}{2mr^2}\mathcal{L}^2 + V(r) \tag{4.1}$$

where we used the differential operator realization $\mathcal{L}^2$ of the operator $\mathbf{L}^2$. The Schrödinger equation will be solved using the following ansatz for energy eigenstates

$$\Psi_{E\ell m}(\mathbf{x}) \;=\; f_{E\ell m}(r)\, Y_{\ell m}(\theta,\phi) \tag{4.2}$$

We have the product of a radial function $f_{E\ell m}(r)$ times an angular function $Y_{\ell m}$ which is an eigenstate of $\mathbf{L^2}$ and of $\hat{L}_z$:

$$\mathcal{L}^2 Y_{\ell m} \;=\; \hbar^2\ell(\ell+1)Y_{\ell m}\,, \qquad \hat{\mathcal{L}}_z Y_{\ell m} \;=\; \hbar m\, Y_{\ell m} \tag{4.3}$$

Plugging this into the Schrödinger equation $H\Psi = E\Psi$, the $Y_{\ell m}$ dependence can be cancelled out and we get

$$-\frac{\hbar^2}{2m}\,\frac{1}{r}\frac{d^2}{dr^2}\,(r f_{E\ell m}) + \frac{\hbar^2\ell(\ell+1)}{2mr^2} f_{E\ell m} + V(r) f_{E\ell m} \;=\; E f_{E\ell m} \tag{4.4}$$

We note that this equation does not depend on the quantum number m (do not confuse this with the mass m!) Therefore the label m is not needed in the radial function and we let $f_{E\ell m} \to f_{E\ell}$ so that we have

$$\Psi_{E\ell m}(\mathbf{x}) \;=\; f_{E\ell}(r)\, Y_{\ell m}(\theta,\phi) \tag{4.5}$$

and the differential equation, multiplying by r becomes

$$-\frac{\hbar^2}{2m}\,\frac{d^2}{dr^2}\,(r f_{E\ell}) + \Big(V(r) + \frac{\hbar^2\ell(\ell+1)}{2mr^2}\Big)(r f_{E\ell}) \;=\; E\,(r f_{E\ell}) \tag{4.6}$$

This suggests writing introducing a modified radial function $u_{E\ell}(r)$ by

$$f_{E\ell}(r) \;=\; \frac{u_{E\ell}(r)}{r} \tag{4.7}$$

so that we have

$$\boxed{\Psi_{E\ell m}(\mathbf{x}) \;=\; \frac{u_{E\ell}(r)}{r}\, Y_{\ell m}(\theta,\phi)\,,} \tag{4.8}$$

with radial equation

$$-\frac{\hbar^2}{2m}\,\frac{d^2u_{E\ell}}{dr^2} \;+\; V_{\text{eff}}(r)u_{E\ell} \;=\; E\,u_{E\ell}\,, \tag{4.9}$$

where the effective potential V_{eff} constructed by adding to the potential $V(r)$ the centrifugal barrier term proportional to $\mathbf{L}^2$:

$$V_{\text{eff}}(r) \;\equiv\; V(r) + \frac{\hbar^2\ell(\ell+1)}{2mr^2} \tag{4.10}$$

This is like a one-dimensional Schrödinger equation in the variable r, but as opposed to our usual problems with $x \in (-\infty,\infty)$, the radius $r \in [0,\infty]$ and we will need some special care for $r = 0$.

The normalization of our wavefunctions proceeds as follows

$$\int d^3\mathbf{x}\,|\Psi_{E\ell m}(\mathbf{x})|^2 \;=\; 1 \tag{4.11}$$

This gives

$$\int r^2dr\,d\Omega\;\frac{|u_{E\ell}(r)|^2}{r^2}\;Y^*_{\ell m}(\Omega)Y_{\ell m}(\Omega) \;=\; 1 \tag{4.12}$$

the angular integral gives one and we get

$$\int_0^\infty dr\,|u_{E\ell}(r)|^2 \;=\; 1 \tag{4.13}$$

a rather natural result for the function $u_{E\ell}$ that plays the role of radial wavefunction.

Behavior of solutions as $r \to 0$. We claim that

$$\lim_{r\to 0} u_{E\ell}(r) \;=\; 0\,. \tag{4.14}$$

This requirement does not arise from normalization: as you can see in (4.13) a finite $u_{E\ell}$ at $r = 0$ would cause no trouble. Imagine taking a solution u_{E0} with $\ell = 0$ that approaches a constant as $r \to 0$:

$$\lim_{r\to 0} u_{E0}(r) \;=\; c \neq 0\,. \tag{4.15}$$

The full solution Ψ near the origin would then take the form

$$\Psi(\mathbf{x}) \;\simeq\; \frac{c}{r}\,Y_{00} \;=\; \frac{c'}{r}\,, \tag{4.16}$$

since Y_{00} is simply a constant. The problem with this wavefunction is that it simply *does not solve* the Schrödinger equation! You may remember from electromagnetism that the Laplacian of the $1/r$ function has a delta function at the origin, so that as a result

$$\nabla^2 \Psi(\mathbf{x}) \;=\; -4\pi c' \delta(\mathbf{x})\,. \tag{4.17}$$

Since the Laplacian is part of the Hamiltonian, this delta function must be cancelled by some other contribution, but there is none, since the potential $V(r)$ does not have delta functions[2].

We can learn more about the behavior of the radial solution under the reasonable assumption that the *centrifugal barrier dominates the potential as* $r \to 0$. In this case the most singular terms of the radial differential equation must cancel each other out, leaving less singular terms that we can ignore in this leading term calculation. So we set:

$$-\frac{\hbar^2}{2m}\frac{d^2 u_{E\ell}}{dr^2} + \frac{\hbar^2\,\ell(\ell+1)}{2mr^2} u_{E\ell} \;=\; 0\,, \qquad \text{as } r \to 0\,. \tag{4.18}$$

or equivalently

$$\frac{d^2 u_{E\ell}}{dr^2} \;=\; \frac{\ell(\ell+1)}{r^2} u_{E\ell}\,, \tag{4.19}$$

The solutions of this can be taken to be $u_{E\ell} = r^s$ with s a constant to be determined. We then find

$$s(s-1) \;=\; \ell(\ell+1) \quad \to \quad s = \ell+1,\; s = -\ell \tag{4.20}$$

thus leading to two possible behaviors near $r = 0$:

$$u_{E\ell} \;\sim\; r^{\ell+1}\,, \qquad u_{E\ell} \sim \;\frac{1}{r^\ell}\,. \tag{4.21}$$

For $\ell = 0$ the second behavior was shown to be inconsistent with the Schrödinger equation at $r = 0$ (because of a delta function). For $\ell > 0$ the second behavior is not consistent with normalization. Therefore we have established that

$$\boxed{u_{E\ell} \;\sim c\, r^{\ell+1}\,, \quad \text{as } r \to 0\,.} \tag{4.22}$$

Note that the full radial dependence is obtained by dividing by r, so that

$$f_{E\ell} \;\sim\; c\, r^\ell\,, \tag{4.23}$$

This allows for a constant non-zero wavefunction at the origin only for $\ell = 0$. Only for $\ell = 0$ a particle can be at the origin. For $\ell \neq 0$ the angular momentum "barrier" prevents the particle from reaching the origin.

[2]Delta function potentials in more than one dimension are very singular and require regulation.

Behavior of solutions as $r \to \infty$. Again, we can make some definite statements once we assume some properties of the potential. Let us consider the case when the potential $V(r)$ vanishes beyond some radius or at least decays fast enough as the radius grows without bound

$$V(r) \;=\; 0\,,\ \text{for } r > r_0\,,\quad \text{or}\quad \lim_{r\to\infty} rV(r) \;=\; 0\,. \tag{4.24}$$

Curiously, the above assumptions are violated for the $1/r$ potential of the Hydrogen atom (an extended discussion of related issues can be found in Shankar around page 344). Under these assumptions we ignore the effective potential completely (including the centrifugal barrier) and the equation becomes

$$-\frac{\hbar^2}{2m}\frac{d^2u_{E\ell}}{dr^2} \;=\; Eu_{E\ell}(r)\,. \tag{4.25}$$

The equation is the familiar

$$\frac{d^2u_{E\ell}}{dr^2} \;=\; -\,\frac{2mE}{\hbar^2}\,u_{E\ell}\,. \tag{4.26}$$

The resulting $r \to \infty$ behavior follows immediately

$$\begin{aligned} E<0\,,\qquad & u_{E\ell} \;\sim\; \exp\Bigl(-\sqrt{\frac{2m|E|}{\hbar^2}}\,r\Bigr)\,,\\ E>0\,,\qquad & u_{E\ell} \;\sim\; \exp\bigl(\pm ikr\bigr)\,,\quad k=\sqrt{\frac{2mE}{\hbar^2}}\,. \end{aligned} \tag{4.27}$$

The first behavior, for $E < 0$ is typical of bound states. For $E > 0$ we have a continuous spectrum with degenerate solutions (hence the $\pm$). Having understood the behavior of solutions near $r = 0$ and for $r \to \infty$ this allows for qualitative plots of radial solutions.

The discrete spectrum is organized as follows. We have energy eigenstates for all values of ℓ. In fact for each value of ℓ the potential V_{eff} in the radial equation is different. So this equation must be solved for $\ell = 0, 1, \ldots$. For each fixed ℓ we have a one-dimensional problem, so we have no degeneracies in the bound state spectrum. We have a set of allowed values of energies that depend on ℓ and are numbered using an integer $n = 1, 2 \ldots$. For each allowed energy $E_{n\ell}$ we have a single radial solution $u_{n\ell}$.

$$\text{Fixed } \ell,\ \text{Energies: } E_{n\ell}\,,\quad \text{Radial function: } u_{n\ell}\,,\quad n = 1, 2, \ldots \tag{4.28}$$

Of course each solution $u_{n\ell}$ for the radial equation represents $2\ell+1$ degenerate solutions to the Schrödinger equation corresponding to the possible values of $\hat{L}_z$ in the range $(-\ell\hbar, \ell\hbar)$. Note that n has replaced the label E in the radial solution, and the energies have now been labeled. This is illustrated in the diagram of Figure 4, where each solution of the radial equation is shown as a short line atop an ℓ label on the horizontal axis. This is the spectral diagram for the central-potential Hamiltonian. Each line of a given ℓ represents the $(2\ell + 1)$ degenerate states obtained with $m = -\ell, \ldots, \ell$. Because the bound state spectrum of a one-dimensional potential

is non-degenerate, our radial equation can't have any degeneracies for any fixed ℓ. Thus the lines on the diagram are single lines! Of course, other types of degeneracies of the spectrum can exist: some states having different values of ℓ may have the same energy. In other words, the states may match across columns on the figure. Finally, note that since the potential becomes more positive as ℓ is increased, the lowest energy state occurs for $\ell = 0$ and the energy $E_{1,\ell}$ of the lowest state for a given ℓ increases as we increase ℓ.

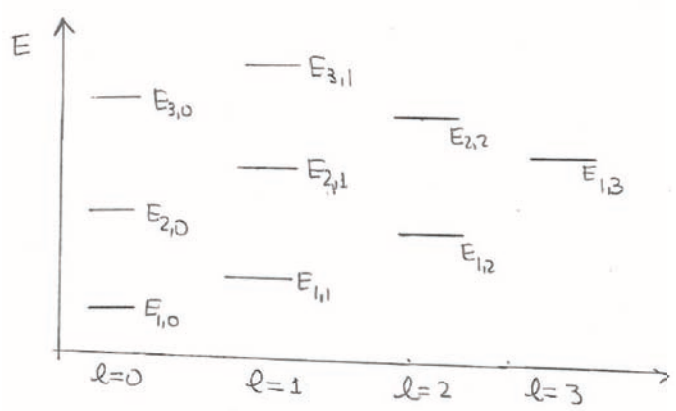


Figure 4: The generic discrete spectrum of a central-potential Hamiltonian, showing the angular momentum ℓ multiplets and their energies.

5 The free particle and the infinite spherical well

5.1 Free particle

It may sound funny at first, but it is interesting to find the radial solutions that correspond to a free particle! A particle that moves in $V(r) = 0$. This amounts to a very different description of the energy eigenstates. In cartesian coordinates we would write solutions as momentum eigenstates, for all values of the momentum. To label such solutions we could use three labels: the components of the momentum. Alternatively, we can use the energy and the direction defined by the momentum, which uses two labels. Here the solutions will be labeled by the energy and (ℓ, m), the usual two integers that describe the angular dependence (of course, ℓ affects the radial dependence too). The radial equation is

$$-\frac{\hbar^2}{2m}\frac{d^2u_{E\ell}}{dr^2} + \frac{\hbar^2}{2m}\frac{\ell(\ell+1)}{r^2}u_{E\ell} \;=\; Eu_{E\ell} \tag{5.29}$$

which is, equivalently

$$-\frac{d^2u_{E\ell}}{dr^2} + \frac{\ell(\ell+1)}{r^2}u_{E\ell} \;=\; k^2u_{E\ell}\,, \qquad k \equiv \sqrt{\frac{2mE}{\hbar^2}} \tag{5.30}$$

In this equation there is no quantization of the energy. Indeed we can redefine the radial coordinate in a way that the energy does not appear, namely, k does not appear. Letting $\rho = kr$ the equation becomes

$$-\frac{d^2 u_{E\ell}}{d\rho^2} + \frac{\ell(\ell+1)}{\rho^2} u_{E\ell} \;=\; u_{E\ell}\,, \tag{5.31}$$

The solution of this differential equation with regular behavior at the origin is $u_{E\ell} = c\rho j_\ell(\rho)$, where c is an arbitrary constant. This means that we can take

$$u_{E\ell} \;=\; r j_\ell(kr)\,. \tag{5.32}$$

Here the $j_\ell(x)$ are the spherical Bessel functions. All in all we have

$$\text{Free particle:} \qquad \Psi_{E\ell m}(\mathbf{x}) \;=\; j_\ell(kr)\, Y_{lm}(\theta,\phi)\,. \tag{5.33}$$

The spherical Bessel functions have the following behavior

$$x\, j_\ell(x) \sim \frac{x^{\ell+1}}{(2\ell+1)!!}\,,\ \text{as}\ \ x\to 0\,, \qquad \text{and} \qquad x\, j_\ell(x) \sim \sin\Big(x - \frac{\ell\pi}{2}\Big)\ \text{as}\ \ x\to\infty\,. \tag{5.34}$$

which implies the correct behavior for $u_{E\ell}$ as $r\to 0$ and $r\to\infty$. Indeed, for $r\to\infty$ we have

$$\text{Free particle}: \qquad u_{E\ell} \sim \sin\Big(kr - \frac{\ell\pi}{2}\Big)\,,\ \ \text{as}\ \ r\to\infty\,. \tag{5.35}$$

Whenever the potential is not zero, but vanishes beyond some radius, the solutions, for $r\to\infty$ take the form

$$u_{E\ell} \sim \sin\Big(kr - \frac{\ell\pi}{2} + \delta_\ell(E)\Big)\,,\ \ \text{as}\ \ r\to\infty\,. \tag{5.36}$$

Here $\delta_\ell(E)$ is called the **phase shift** and by definition vanishes if there is no potential. The form of the solution above is consistent with our general behavior, as this is a superposition of the two solutions available in (4.27) for $E>0$. The phase shift contains all the information about a potential $V(r)$ available to someone probing the potential from far away by scattering particles off of it.

5.2 The infinite spherical well

An infinite spherical well of radius a is a potential that forces the particle to be within the sphere $r=a$. The potential is zero for $r\le a$ and it is infinite for $r>a$.

$$V(r) \;=\; \begin{cases} 0\,,\ \text{if}\ \ r\le a \\ \infty\,,\ \text{if}\ r>a \end{cases} \tag{5.37}$$

The Schrödinger radial equation is the same as the one for the free particle

$$-\frac{d^2 u_{E\ell}}{d\rho^2} + \frac{\ell(\ell+1)}{\rho^2} u_{E\ell} \;=\; u_{E\ell}\,, \quad \rho = kr \tag{5.38}$$

where k again encodes the energy E, which is greater than zero. It follows that the solutions are the ones we had before, with spherical Bessel functions, but this time quantization of the energy arises because the wavefunctions must vanish for $r = a$.

Let us do the case $\ell = 0$ without resorting to the Bessel functions. The above equation becomes

$$-\frac{d^2 u_{E,0}}{d\rho^2} \;=\; u_{E,0} \quad \to \quad u_{E,0} \;=\; A\sin\rho + B\cos\rho\,. \tag{5.39}$$

Since the solution must vanish at $r = 0$ we must choose the sin function:

$$u_{E,0}(r) \;=\; \sin kr\,. \tag{5.40}$$

Since this must vanish for $r = a$ we have that k must take values k_n with

$$k_n a = n\pi\,, \quad \text{for} \quad n = 1, 2, \ldots \infty\,. \tag{5.41}$$

Those values of k_n correspond to energies $E_{n,0}$ where the n indexes the solutions and the 0 represents $\ell = 0$:

$$E_{n,0} \;=\; \frac{\hbar^2 k_n^2}{2m} \;=\; \frac{\hbar^2}{2ma^2}(k_n a)^2 \;=\; \frac{\hbar^2}{2ma^2}\; n^2\pi^2\,. \tag{5.42}$$

Note that $\frac{\hbar^2}{2ma^2}$ is the natural energy scale for this problem and therefore it is convenient to define the unit-free scaled energies $\mathcal{E}_{n,\ell}$ by dividing $E_{n,\ell}$ by the natural energy by

$$\mathcal{E}_{n,\ell} \;\equiv\; \frac{2ma^2}{\hbar^2} E_{n,\ell}\,. \tag{5.43}$$

It follows that the 'energies' for $\ell = 0$ are

$$\mathcal{E}_{n,0} \;=\; n^2\pi^2\,, \qquad u_{n,0} \;=\; \sin\Big(\frac{n\pi r}{a}\Big)\,. \tag{5.44}$$

We have

$$\mathcal{E}_{1,0} \;\simeq\; 9.8696\,, \quad \mathcal{E}_{2,0} \;\simeq\; 39.478\,, \quad \mathcal{E}_{3,0} \;\simeq\; 88.826\,, \tag{5.45}$$

Let us now do $\ell = 1$. Here the solutions are $u_{E,1} = r\, j_1(kr)$. This Bessel function is

$$j_1(\rho) \;=\; \frac{\sin\rho}{\rho^2} - \frac{\cos\rho}{\rho} \tag{5.46}$$

The zeroes of $j_1(\rho)$ occur for $\tan\rho = \rho$. Of course, we are not interested in the zero at $\rho = 0$. You can check that the first three zeroes occur for $4.4934, 7.7252, 10.904$. For higher values of ℓ it becomes a bit more complicated but there are tables of zeroes on the web.

There is notation in which the nontrivial zeroes are denoted by $z_{n,\ell}$ where

$$z_{n,\ell} \;\text{ is the } n\text{-th zero of } j_\ell :\; j_\ell(z_{n,\ell}) = 0\,. \tag{5.47}$$

The vanishing condition at $r = a$ quantizes k so that

$$k_{n,\ell}\, a \;=\; z_{n,\ell} \tag{5.48}$$

and the energies

$$E_{n,\ell} \;=\; \frac{\hbar^2}{2ma^2}\,(k_{n,\ell}a)^2 \quad \rightarrow \quad \mathcal{E}_{n,\ell} \;=\; z_{n,\ell}^2 \tag{5.49}$$

We have

$$\begin{aligned} &z_{1,1} = 4.4934\,, \quad z_{2,1} = 7.7252\,, \quad z_{3,1} = 10.904 \\ &z_{1,2} = 5.7634\,, \quad z_{2,2} = 9.095\,, \\ &z_{1,3} = 6.9879\,, \quad z_{2,3} = 10.417\,. \end{aligned} \tag{5.50}$$

which give us

$$\begin{aligned} &\mathcal{E}_{1,1} = 20.191\,, \quad \mathcal{E}_{2,1} = 59.679\,, \quad \mathcal{E}_{3,1} = 118.89 \\ &\mathcal{E}_{1,2} = 33.217\,, \quad \mathcal{E}_{2,2} = 82.719\,, \\ &\mathcal{E}_{1,3} = 48.83\,, \quad \mathcal{E}_{2,3} = 108.51\,. \end{aligned} \tag{5.51}$$

The main point to be made is that there are no accidental degeneracies: the energies for different values of ℓ never coincide. More explicitly, with $\ell \neq \ell'$ we have that $\mathcal{E}_{n,\ell} \neq \mathcal{E}_{n',\ell'}$ for any choices of n and n'. This is illustrated in figure 5.

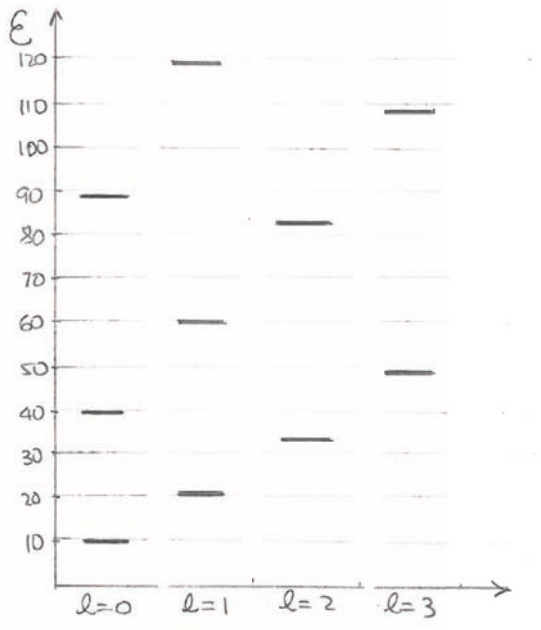


Figure 5: The spectrum of the infinite spherical square well. There are no accidental degeneracies.

6 The three-dimensional isotropic oscillator

The potential of the 3D isotropic harmonic oscillator is as follows:

$$V = \frac{1}{2}m\omega^2(x^2+y^2+z^2) = \frac{1}{2}m\omega^2 r^2\,. \tag{6.52}$$

As we will see, the spectrum for this quantum mechanical system has degeneracies, that are explained by the existence of some **hidden symmetry**, a symmetry that is not obvious from the start. Thus in some ways this quantum 3D oscillator is a lot more symmetric than the infinite spherical well.

As you know, for the 3D oscillator we can use creation and annihilation operators $\hat{a}_x^\dagger, \hat{a}_y^\dagger, \hat{a}_z^\dagger$ and $\hat{a}_x, \hat{a}_y, \hat{a}_z$ associated with 1D oscillators in the x, y, and z directions. The Hamiltonian then takes the form:

$$H = \hbar\omega\big(\hat{N}_1+\hat{N}_2+\hat{N}_3+\tfrac{3}{2}\big) \;=\; \hbar\omega\big(\hat{N}+\tfrac{3}{2}\big)\,. \tag{6.53}$$

where we defined $\hat{N} \equiv \hat{N}_1+\hat{N}_2+\hat{N}_3$.

We now want to explain how tensor products are relevant to the 3D oscillator. We have discussed tensor products before to describe two particles, each associated with a vector space and the combined system associated with the tensor product of vector spaces. But tensor products are also relevant to single particles, if they have degrees of freedom that live in different spaces, or more than one set of attributes, each of which described by states in some vector space. For example, if a spin 1/2 particle can move, the relevant states live in the tensor product of momentum space and the 2-dimensional complex vector space of spin. States are obtained by superposition of basic states of the form $|p\rangle \otimes (\alpha|+\rangle + \beta|-\rangle)$

For the 3D oscillator, the Hamiltonian is the sum of commuting Hamiltonians of 1D oscillators for the x, y, and z directions. Thus the general states are obtained by tensoring the state spaces $\mathcal{H}_x, \mathcal{H}_y$, and $\mathcal{H}_z$ of the three independent oscillators. It is a single particle oscillating, but the description of what it is doing entails saying what is doing in each of the independent directions. Thus we write

$$\mathcal{H}_{3D} \;=\; \mathcal{H}_x \otimes \mathcal{H}_y \otimes \mathcal{H}_z\,. \tag{6.54}$$

Instead of this tensor product reflecting the behavior of three different particles, this tensor product allows us to describe the behavior of one particle in three different directions. The vacuum state $|0\rangle$ of the 3D oscillator can be viewed as

$$|0\rangle \;\equiv\; |0\rangle_x \otimes |0\rangle_y \otimes |0\rangle_z\,. \tag{6.55}$$

The associated wavefunction is

$$\Psi(x,y,z) \;=\; \langle x| \otimes \langle y| \otimes \langle z|\,|0\rangle \;=\; \langle x|0\rangle_x \langle y|0\rangle_y \langle z|0\rangle_z \;=\; \psi_0(x)\psi_0(y)\psi_0(z)\,. \tag{6.56}$$

where ψ_0 is the ground state wavefunction of the 1D oscillator. This is the expected answer. Recalling the form of (non-normalized) basis states for $\mathcal{H}_x, \mathcal{H}_y$, and $\mathcal{H}_z$:

$$\begin{aligned} &\text{basis states for } \mathcal{H}_x: \ (\hat{a}_x^\dagger)^{n_x}|0\rangle_x\,, \ n_x = 0, 1, \ldots \\ &\text{basis states for } \mathcal{H}_y: \ (\hat{a}_y^\dagger)^{n_y}|0\rangle_y\,, \ n_y = 0, 1, \ldots \\ &\text{basis states for } \mathcal{H}_z: \ (\hat{a}_z^\dagger)^{n_z}|0\rangle_z\,, \ n_z = 0, 1, \ldots \end{aligned} \tag{6.57}$$

We then have that the basis states for the 3D state space are

$$\text{basis states of } \mathcal{H}_{3D}: \ (\hat{a}_x^\dagger)^{n_x}|0\rangle_x \otimes (\hat{a}_y^\dagger)^{n_y}|0\rangle_y \otimes (\hat{a}_z^\dagger)^{n_z}|0\rangle_z\,, \quad n_x, n_y, n_z \in \{0, 1, \ldots\} \tag{6.58}$$

This is what we would expect intuitively, we simply pile arbitrary numbers of $\hat{a}_x^\dagger, \hat{a}_y^\dagger$, and $\hat{a}_z^\dagger$ on the vacuum. It is this multiplicative structure that is the signature of tensor products. Having understood the above, for brevity we write such basis states simply as

$$(\hat{a}_x^\dagger)^{n_x}(\hat{a}_y^\dagger)^{n_y}(\hat{a}_z^\dagger)^{n_z}|0\rangle\,. \tag{6.59}$$

Each of the states in (6.58) has a wavefunction that is the product of x, y, and z-dependent wavefunctions. Once we form superpositions of such states, the total wavefunction cannot any longer be factorized into x, y, and z-dependent wavefunctions. The x, y, and z-dependences become 'entangled'. These are precisely the analogs of entangled states of three particles.

We are ready to begin constructing the individual states of the 3D isotropic harmonic oscillator system. The key property is that the states must organize themselves into representations of angular momentum. Since angular momentum commutes with the Hamiltonian, angular momentum multiplets represent degenerate states.

We already built the ground state, which is a single state with $\hat{N}$ eigenvalue $N = 0$. All other states have higher energies, so this state must be, by itself a representation of angular momentum. It can only be the singlet $\ell = 0$. Thus we have

$$N = 0\,, \ E = \tfrac{3}{2}\,\hbar\omega\,, \quad |0\rangle \ \leftrightarrow \ \ell = 0\,. \tag{6.60}$$

The states with $N = 1$ have $E = (5/2)\hbar\omega$ and are

$$\hat{a}_x^\dagger|0\rangle\,, \ \hat{a}_y^\dagger|0\rangle\,, \hat{a}_z^\dagger|0\rangle \tag{6.61}$$

These three states fit precisely into an $\ell = 1$ multiplet (a triplet). There is no other possibility, in fact: any higher ℓ multiplet has too many states and we only have 3 degenerate ones. Moreover, we cannot have three singlets, this is a degeneracy inconsistent with the lack of degeneracy for 1D bound states (as discussed earlier). The $\ell = 0$ ground state and the $\ell = 1$ triplet at the first excited level are indicated in Figure 7.

Let us proceed now with the states at $N = 2$ or $E = (7/2)\hbar\omega$. These are, the following six states:

$$(\hat{a}_x^\dagger)^2|0\rangle\,,\ (\hat{a}_y^\dagger)^2|0\rangle\,,\ (\hat{a}_z^\dagger)^2|0\rangle\,,\ \hat{a}_x^\dagger\hat{a}_y^\dagger|0\rangle\,,\ \hat{a}_x^\dagger\hat{a}_z^\dagger|0\rangle\,,\ \hat{a}_y^\dagger\hat{a}_z^\dagger|0\rangle\,. \tag{6.62}$$

To help ourselves in trying to find the angular momentum multiplets recall that that the number of states # for a given ℓ are

ℓ	#
0	1
1	3
2	5
3	7
4	9
5	11
6	13
7	15

Since we cannot use the triplet twice, the only way to get six states is having five from $\ell = 2$ and one from $\ell = 0$. Thus

$$\text{Six } N = 2 \text{ states}: \quad (\ell = 2) \oplus (\ell = 0)\,. \tag{6.63}$$

Note that here we use the direct sum (not the tensor product!) the six states define a six dimensional vector space spanned by five vectors in $\ell = 0$ and one vector in $\ell = 0$. Had we used a tensor product we would just have 5 vectors.

Let us continue to figure out the pattern. At $N = 3$ with $E = (9/2)\hbar\omega$ we actually have 10 states (count them!) It would seem now that there are two possibilities for multiplets

$$(\ell = 3) \oplus (\ell = 1) \quad \text{or} \quad (\ell = 4) \oplus (\ell = 0) \tag{6.64}$$

We can argue that the second possibility cannot be. The problem with it is that the $\ell = 3$ multiplet, which has not appeared yet, would not arise at this level. If it would arise later, it would do so at a higher energy, and we would have the lowest $\ell = 3$ multiplet above the lowest $\ell = 4$ multiplet, which is not possible. You may think that perhaps $\ell = 3$ multiplets never appear and the inconsistency is avoided, but this is not true. At any rate we will give below a more rigorous argument. The conclusion, however is that

$$\text{Ten } N = 3 \text{ states}: \quad (\ell = 3) \oplus (\ell = 1)\,. \tag{6.65}$$

Let us do the next level! At $N = 4$ we find 15 states. Instead of writing them out let us count them without listing them. In fact, we can easily do the general case of arbitrary integer $N \geq 1$. The states we are looking for are of the form

$$(\hat{a}_x^\dagger)^{n_x}(\hat{a}_y^\dagger)^{n_y}(\hat{a}_z^\dagger)^{n_z}|0\rangle\,, \quad \text{with} \quad n_x + n_y + n_z = N\,. \tag{6.66}$$

We need to count how many different solutions there are to $n_x+n_y+n_z = N$, with $n_x, n_y, n_z \geq 0$. This is the number of states $\#(N)$ at level N. To visualize this think of $n_x + n_y + n_z = N$ as the equation for a plane in three-dimensional space with axes n_x, n_y, n_z. Since no integer can be negative, we are looking for points with integer coordinates in the region of the plane that lies on the positive octant, as shown in Figure 6. Starting at one of the three corners, say $(n_x, n_y, n_z) = (N, 0, 0)$ we have one point, then moving towards the origin we encounter two points, then three, and so on until we find $N+1$ points on the (n_y, n_z) plane. Thus, the number of states $\#(N)$ for number N is

$$\#(N) = 1 + 2 + \ldots + (N+1) \;=\; \frac{(N+1)(N+2)}{2} \tag{6.67}$$

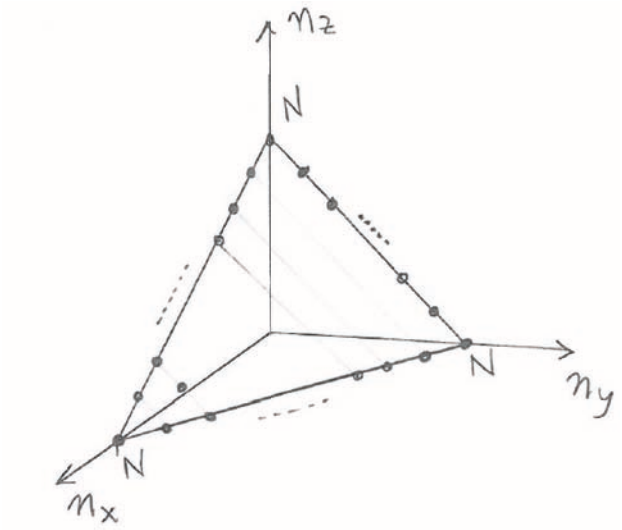


Figure 6: Counting the number of degenerate states with number N in the 3D simple harmonic oscillator.

Back to the $N = 4$ level, $\#(4)$=15. We rule out a single $\ell = 7$ multiplet since states with $\ell = 4, 5, 6$ have not appeared yet. By this logic the highest ℓ multiplet for $N = 4$ must be the lowest that has not appeared yet, thus $\ell = 4$, with 9 states. The remaining six must appear as $\ell = 2$ plus $\ell = 0$. Thus, we have

$$15 \; N = 4 \text{ states}: \quad (\ell = 4) \oplus (\ell = 2) \oplus (\ell = 0)\,. \tag{6.68}$$

Thus we see that ℓ jumps by steps of two, starting from the maximal ℓ. This is in fact the rule. It is quickly confirmed for the $\#(5)$=21 states with $N = 5$ would arise from $(\ell = 5) \oplus (\ell = 3) \oplus (\ell = 1)$. All this is shown in Figure 7.

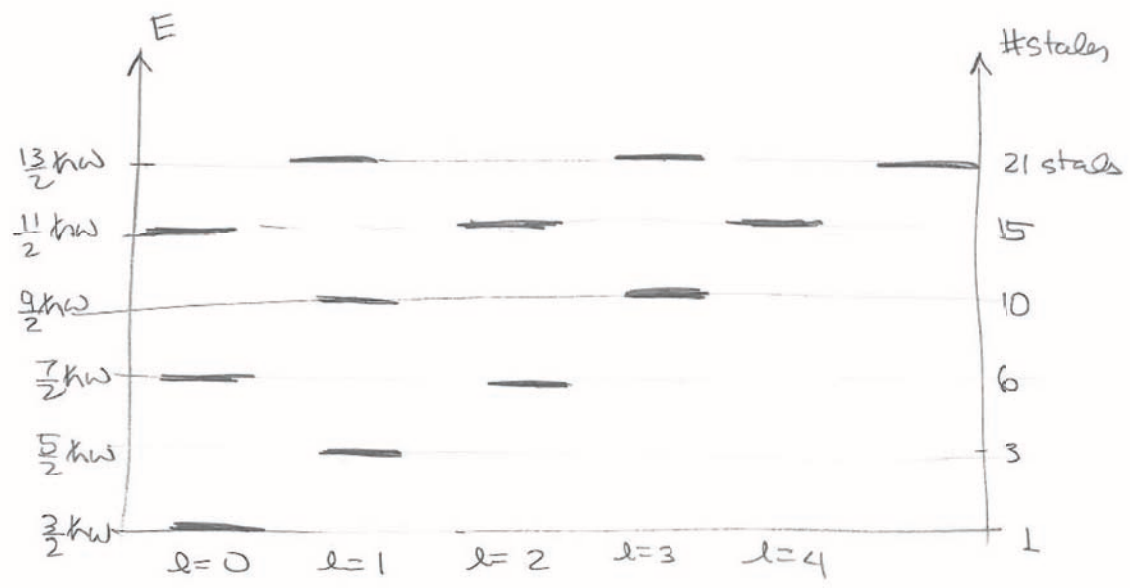


Figure 7: Spectral diagram for angular momentum multiplets in the 3D isotropic harmonic oscillator.

Some of the structure of angular momentum multiplets can be seen more explicitly using the $\hat{a}_L$ and $\hat{a}_R$ operators introduced for the 2D harmonic oscillator:

$$\hat{a}_L \;=\; \frac{1}{\sqrt{2}}(\hat{a}_x + i\hat{a}_y)\,, \quad \hat{a}_R \;=\; \frac{1}{\sqrt{2}}(\hat{a}_x - i\hat{a}_y)\,. \tag{6.69}$$

L and R objects commute with each other and we have $[\hat{a}_L, \hat{a}_L^\dagger] = [\hat{a}_R, \hat{a}_R^\dagger] = 1$. With number operators $\hat{N}_R = \hat{a}_R^\dagger \hat{a}_R$ and $\hat{N}_L = \hat{a}_L^\dagger \hat{a}_L$ we then have $H = \hbar\omega(\hat{N}_R + \hat{N}_L + \hat{N}_z + \frac{3}{2})$ and, more importantly, the z component $\hat{L}_z$ of angular momentum takes the simple form

$$\hat{L}_z = \hbar(\hat{N}_R - \hat{N}_L)\,. \tag{6.70}$$

Note that $\hat{a}_z$ carries no z-component of angular momentum. States are now build acting with arbitrary numbers of $\hat{a}_L^\dagger, \hat{a}_R^\dagger$ and $\hat{a}_z^\dagger$ operators on the vacuum. The $N = 1$ states are then presented as

$$\hat{a}_R^\dagger|0\rangle\,,\ \hat{a}_z^\dagger|0\rangle\,,\ \hat{a}_L^\dagger|0\rangle\,. \tag{6.71}$$

We see that the first state has $L_z = \hbar$, the second $L_z = 0$ and the third $\hat{L}_z = -\hbar$, exactly the three expected values of the $\ell = 1$ multiplet identified before. For number $N = 2$ the state with highest L_z is $(\hat{a}_R^\dagger)^2|0\rangle$ and it has $L_z = 2\hbar$. This shows that the highest ℓ multiplet is $\ell = 2$. For arbitrary positive integer number N, the state with highest L_z is $(\hat{a}_R^\dagger)^N|0\rangle$ and it has $L_z = \hbar N$. This shows we must have an $\ell = N$ multiplet. This is in fact what we got before! We can also understand the reason for the jump of two units from the top state of the multiplet. Consider

the above state with maximal $\hat{L}_z/\hbar$ equal to N and then the states with one and two units less of $\hat{L}_z/\hbar$:

$$\begin{aligned}
\hat{L}_z/\hbar = N \quad &: \qquad (\hat{a}_R^\dagger)^N|0\rangle \\
\hat{L}_z/\hbar = N-1 : \quad &: \qquad (\hat{a}_R^\dagger)^{N-1}\hat{a}_z^\dagger|0\rangle \\
\hat{L}_z/\hbar = N-2 : \quad &: \qquad (\hat{a}_R^\dagger)^{N-2}(\hat{a}_z^\dagger)^2|0\rangle\,, \quad (\hat{a}_R^\dagger)^{N-1}\,\hat{a}_L^\dagger|0\rangle
\end{aligned} \tag{6.72}$$

While there is only one state with one unit less of $\hat{L}_z/\hbar$ there are two states with two units less. One linear combination of these two states must belong to the $\ell = N$ multiplet, but the other linear combination must be the top state of an $\ell = N-2$ multiplet! This is the reason for the jump of two units.

For arbitrary N we can see why $\#(N)$ can be reproduced by ℓ multiplets skipping by two

$$\begin{aligned}
N \text{ odd}: \quad \#(N) &= \underbrace{1+2}_{\ell=1} + \underbrace{3+4}_{\ell=3} + \underbrace{5+6}_{\ell=5} + \underbrace{7+8}_{\ell=7} + \ldots + \underbrace{N+(N+1)}_{\ell=N} \\
N \text{ even}: \quad \#(N) &= \underbrace{1}_{\ell=0} + \underbrace{2+3}_{\ell=2} + \underbrace{4+5}_{\ell=4} + \underbrace{6+7}_{\ell=6} + \ldots + \underbrace{N+(N+1)}_{\ell=N}
\end{aligned} \tag{6.73}$$

The accidental degeneracy is "explained" if we identify an operator that commutes with the Hamiltonian (a symmetry) and connects the various ℓ multiplets that appear for a fixed number N. One such operator is

$$K \;\equiv\; \hat{a}_R^\dagger \hat{a}_L\,. \tag{6.74}$$

You can check it commutes with the Hamiltonian, and with a bit more work, that acting on the top state of the $\ell = N-2$ multiplet it gives the top state of the $\ell = N$ multiplet.

7 Hydrogen atom and Runge-Lenz vector

The hydrogen atom Hamiltonian is

$$H \;=\; \frac{\mathbf{p}^2}{2m} - \frac{e^2}{r}\,. \tag{7.75}$$

The natural length scale here is the Bohr radius a_0, which is the unique length that can be built using the constants in this Hamiltonian: $\hbar, m$, and e^2. We determine a_0 by setting $p \sim \hbar/a_0$ and equating magnitudes of kinetic and potential terms, ignoring numerical factors:

$$\frac{\hbar^2}{m a_0^2} \;=\; \frac{e^2}{a_0} \quad \to \quad a_0 \;=\; \frac{\hbar^2}{me^2} \;\simeq\; 0.529\mathring{A}\,. \tag{7.76}$$

Note that if the charge of the electron e^2 is decreased, the attraction force decreases and, correctly, the Bohr radius increases. The Bohr radius is the length scale of the hydrogen atom.

A natural energy scale E_0 is

$$E_0 \;=\; \frac{e^2}{a_0} \;=\; \frac{e^4 m}{\hbar^2} \;=\; \Big(\frac{e^2}{\hbar c}\Big)^2 mc^2 \;=\; \alpha^2(mc^2) \tag{7.77}$$

where we see the appearance of the fine-structure constant α that, in cgs units, takes the form

$$\alpha \;\equiv \frac{e^2}{\hbar c} \;\simeq\; \frac{1}{137}\,. \tag{7.78}$$

We thus see that the natural energy scale of the hydrogen atom is about $\alpha^2 \simeq 1/18770$ smaller than the rest energy of the electron. This gives about $E_0 = 27.2\text{eV}$. In fact $-E_0/2 = -13.6\text{eV}$ is the bound state energy of the electron in the ground state of the hydrogen atom.

One curious way to approach the calculation of the ground state energy and ground state wavefunction is to factorize the Hamiltonian. One can show that

$$H \;=\; \gamma \;+\; \frac{1}{2m}\sum_{k=1}^{3}\Big(\hat{p}_k + i\beta\frac{\hat{x}_k}{r}\Big)\Big(\hat{p}_k - i\beta\frac{\hat{x}_k}{r}\Big) \tag{7.79}$$

for suitable constants β and γ that you can calculate. The ground state $|\Psi_0\rangle$ is then the state for which

$$\Big(\hat{p}_k - i\beta\frac{\hat{x}_k}{r}\Big)|\Psi_0\rangle \;=\; 0\,. \tag{7.80}$$

The spectrum of the hydrogen atom is described in Figure 8. The energy levels are $E_{\nu\ell}$, where we used $\nu = 1, 2, \ldots$, instead of n to label the various solutions for a given ℓ. This is because the label n is reserved for what is called the "principal quantum number". The degeneracy of the system is such that multiplets with equal $n \equiv \nu + \ell$ have the same energy, as you can see in the figure. Thus, for example, $E_{2,0} = E_{1,1}$, which is to say that the first excited solution for $\ell = 0$ has the same energy as the lowest energy solution for $\ell = 1$. It is also important to note that for any fixed value of n the allowed values of ℓ are

$$\ell = 0, 1, \ldots, n-1 \tag{7.81}$$

Finally, the energies are given by

$$E_{\nu\ell} \;=\; -\frac{e^2}{2a_0}\,\frac{1}{(\nu+\ell)^2}\,, \quad n \equiv \nu + \ell\,. \tag{7.82}$$

The large amount of degeneracy in this spectrum asks for an explanation. The hydrogen Hamiltonian has in fact some hidden symmetry. It has to do with the so-called Runge-Lenz vector. In the following we discuss the classical origin of this conserved vector quantity.

Imagine we have an energy functional

$$E \;=\; \frac{\mathbf{p}^2}{2m} + V(r) \tag{7.83}$$

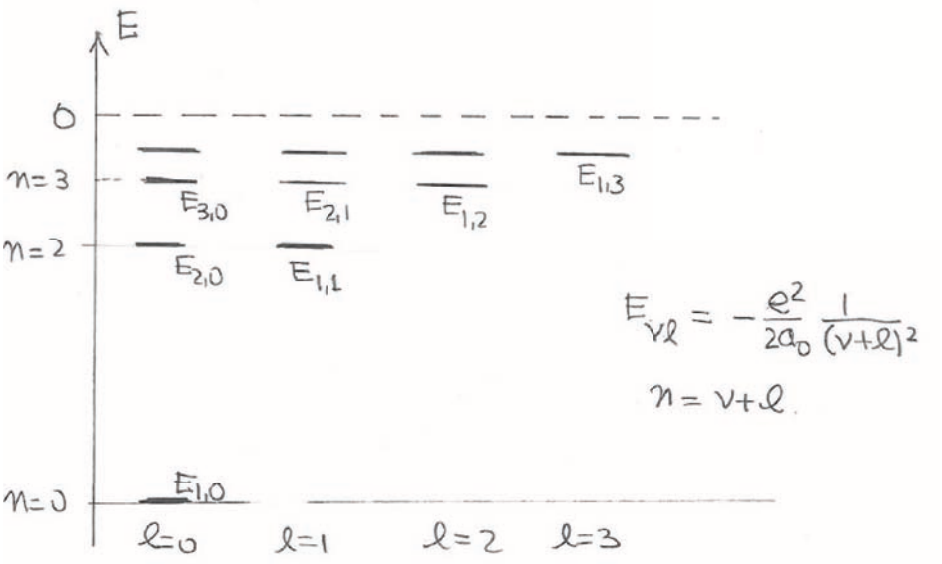


Figure 8: Spectrum of angular momentum multiplets for the hydrogen atom. Here $E_{\nu\ell}$ with $\nu = 1, 2, \ldots$, denotes the energy of the ν-th solution for any fixed ℓ. States with equal values of $n \equiv \nu + \ell$ are degenerate. For any fixed n, the values of ℓ run from zero to $n-1$. Correction: the $n = 0$ in the figure should be $n = 1$.

then the force on the particle moving in this potential is

$$\mathbf{F} \;=\; -\nabla V \;=\; -V'(r)\frac{\mathbf{r}}{r}\,, \tag{7.84}$$

where primes denote derivatives with respect to the argument. Newton's equation is

$$\frac{d\mathbf{p}}{dt} \;=\; -V'(r)\frac{\mathbf{r}}{r} \tag{7.85}$$

and it is simple to show (do it!) that in this central potential the angular momentum is conserved

$$\frac{d\mathbf{L}}{dt} \;=\; 0\,. \tag{7.86}$$

We now calculate (all classically) the time derivative of $\mathbf{p} \times \mathbf{L}$:

$$\begin{aligned}\frac{d}{dt}(\mathbf{p}\times\mathbf{L}) \;&=\; \frac{d\mathbf{p}}{dt}\times\mathbf{L} \;=\; -\frac{V'(r)}{r}\,\mathbf{r}\times(\mathbf{r}\times\mathbf{p}) \\ &=\; -\frac{mV'(r)}{r}\,\mathbf{r}\times(\mathbf{r}\times\dot{\mathbf{r}}) \\ &=\; -\frac{mV'(r)}{r}\left[\mathbf{r}(\mathbf{r}\cdot\dot{\mathbf{r}}) - \dot{\mathbf{r}}\,r^2\right].\end{aligned} \tag{7.87}$$

We now note that

$$\mathbf{r}\cdot\dot{\mathbf{r}} \;=\; \frac{1}{2}\frac{d}{dt}(\mathbf{r}\cdot\mathbf{r}) = \;\frac{1}{2}\frac{d}{dt}r^2 \;=\; r\dot{r}\,. \tag{7.88}$$

Using this

$$\begin{aligned}\frac{d}{dt}(\mathbf{p}\times\mathbf{L}) &= -\frac{mV'(r)}{r}\left[\mathbf{r}\,r\dot{r}-\dot{\mathbf{r}}\,r^2\right] = mV'(r)r^2\left[\frac{\dot{\mathbf{r}}}{r}-\frac{\mathbf{r}\,\dot{r}}{r^2}\right]\\ &= mV'(r)r^2\,\frac{d}{dt}\Big(\frac{\mathbf{r}}{r}\Big)\end{aligned} \tag{7.89}$$

Because of the factor $V'(r)r^2$, the right-hand side fails to be a total time derivative. But if we focus on potentials for which this factor is a constant we will get a conservation law. So, assume

$$V'(r)\,r^2 = \gamma\,, \tag{7.90}$$

for some constant γ. Then

$$\frac{d}{dt}(\mathbf{p}\times\mathbf{L}) = m\gamma\,\frac{d}{dt}\Big(\frac{\mathbf{r}}{r}\Big) \quad\to\quad \frac{d}{dt}\Big(\mathbf{p}\times\mathbf{L}-m\gamma\,\frac{\mathbf{r}}{r}\Big)=0 \tag{7.91}$$

We got a conservation law: that complicated vector inside the parenthesis is constant in time! Back to (7.90) we have

$$\frac{dV}{dr} = \frac{\gamma}{r^2} \quad\to\quad V(r)=-\frac{\gamma}{r}+c_0\,. \tag{7.92}$$

This is the most general potential for which we get a conservation law. For $c_0=0$ and $\gamma=e^2$ we have the hydrogen atom potential

$$V(r) = -\frac{e^2}{r}\,, \tag{7.93}$$

so we have

$$\frac{d}{dt}\Big(\mathbf{p}\times\mathbf{L}-me^2\,\frac{\mathbf{r}}{r}\Big)=0\,. \tag{7.94}$$

Factoring a constant we obtain the unit-free conserved **Runge-Lenz** vector $\mathbf{R}$:

$$\boxed{\mathbf{R} \equiv \frac{1}{me^2}\,\mathbf{p}\times\mathbf{L}-\frac{\mathbf{r}}{r}\,,\qquad \frac{d\mathbf{R}}{dt} = 0\,.} \tag{7.95}$$

The conservation of the Runge-Lenz vector is a property of inverse squared central forces. The second vector in $\mathbf{R}$ is simply minus the unit radial vector.

To understand the Runge-Lenz vector, we first examine its value for a circular orbit, as shown in figure 9. The vector $\mathbf{L}$ is out of the page and $\mathbf{p}\times\mathbf{L}$ points radially outward. The vector $\mathbf{R}$ is thus a competition between the outward radial first term and the inner radial second term. If these two terms would not cancel, the result would be a radial vector (outwards or inwards) but in any case, not conserved, as it rotates with the particle. Happily, the two terms cancel. Indeed for a circular orbit

$$m\frac{v^2}{r} = \frac{e^2}{r^2} \quad\to\quad \frac{m^2v^2r}{me^2} = 1 \quad\to\quad \frac{(mv)(mvr)}{me^2} = 1 \quad\to\quad \frac{pL}{me^2}=1\,, \tag{7.96}$$

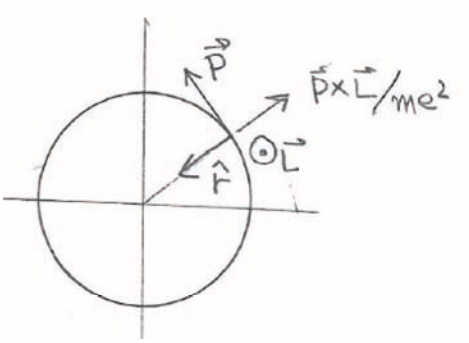


Figure 9: The Runge-Lenz vector vanishes for a circular orbit.

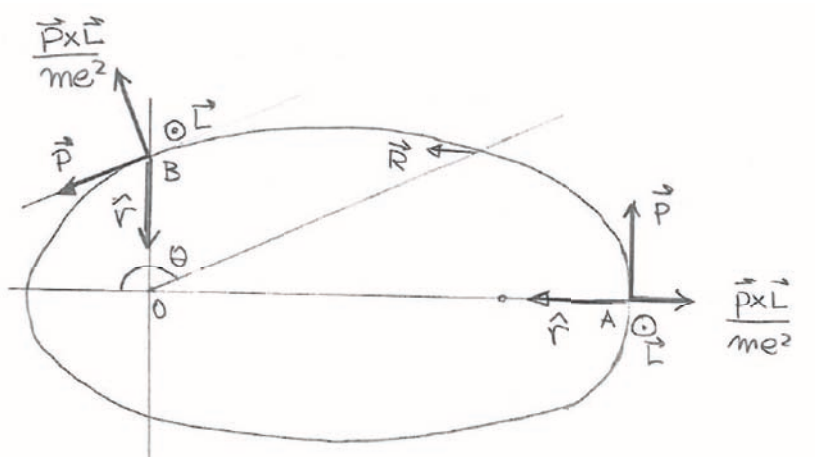


Figure 10: In an elliptic orbit the Runge-Lenz vector is a vector along the major axis of the ellipse and points in the direction from the aphelion to the focus.

which is the statement that the first vector in $\mathbf{R}$, for a circular orbit, is of unit length and being outward directed cancels with the second term. The Runge-Lenz vector indeed vanishes for a circular orbit.

We now argue that for an elliptical orbit the Runge-Lenz vector is not zero. Consider figure 10. At the aphelion (point furthest away from the focal center), denoted as point A we have the first term in $\mathbf{R}$ point outwards and the second term point inwards. Thus, if $\mathbf{R}$ does not vanish it must be a vector along the line joining the focus and the aphelion, a horizontal vector on the figure. Now consider point B right above the focal center of the orbit. At this point $\mathbf{p}$ is no longer perpendicular to the radial vector and therefore $\mathbf{p}\times\mathbf{L}$ is no longer radial. As you can see, it points slightly to the left. It follows that $\mathbf{R}$ points to the left side of the figure. $\mathbf{R}$ is a vector along the major axis of the ellipse and points in the direction from the aphelion to the focus.

To see more quantitatively the role of $\mathbf{R}$ we dot its definition with the radial vector $\mathbf{r}$:

$$\mathbf{r}\cdot\mathbf{R} \;=\; \frac{1}{me^2}\,\mathbf{r}\cdot(\mathbf{p}\times\mathbf{L}) - r \tag{7.97}$$

referring to the figure, with the angle θ as defined there and $R \equiv |\mathbf{R}|$, we get

$$rR\cos\theta \;=\; \frac{1}{me^2}\,\mathbf{L}\cdot(\mathbf{r}\times\mathbf{p}) - r \;=\; \frac{1}{me^2}L^2 - r\,. \tag{7.98}$$

Collecting terms proportional to r:

$$r(1+R\cos\theta) \;=\; \frac{L^2}{me^2} \quad\rightarrow\quad \boxed{\frac{1}{r} \;=\; \frac{me^2}{L^2}(1+R\cos\theta)\,,} \tag{7.99}$$

We identify the magnitude R of the Runge-Lenz vector with the eccentricity of the orbit! Indeed if $R = 0$ the orbit if circular: r does not depend on θ.

This whole analysis has been classical. Quantum mechanically we will need to change some things a bit. The definition of $\mathbf{R}$ only has to be changed to guarantee that $\mathbf{R}$ is a hermitian (vector) operator. As you will verify the hermitization gives

$$\mathbf{R} \;\equiv\; \frac{1}{2me^2}\,(\mathbf{p}\times\mathbf{L} - \mathbf{L}\times\mathbf{p}) - \frac{\mathbf{r}}{r} \tag{7.100}$$

The quantum mechanical conservation of $\mathbf{R}$ is the statement that it commutes with the hydrogen Hamiltonian

$$[\,\mathbf{R}\,, H\,] \;=\; 0\,. \tag{7.101}$$

You will verify this; it is the analog of our classical calculation that showed that the time-derivative of $\mathbf{R}$ is zero. Moreover, the length-squared of the vector is also of interest. You will show that

$$\mathbf{R}^2 \;=\; 1 + \frac{2}{me^4}\,H(\mathbf{L}^2 + \hbar^2)\,. \tag{7.102}$$

ADDITION OF ANGULAR MOMENTUM

B. Zwiebach

December 12, 2013

Contents

1 Adding apples to oranges?

We are going to be adding angular momenta in a variety of ways. We may add the *spin* angular momentum $\mathbf{S}$ of a particle to its *orbital* angular momentum $\mathbf{L}$. Or we may want to add the spin angular momentum $\mathbf{S}^{(1)}$ of a particle to the spin angular momentum $\mathbf{S}^{(2)}$ of another particle. At first sight we may feel like we are trying to add apples to oranges! For a given particle its spin angular momentum has nothing to do with spatial wavefunctions, while its orbital angular momentum does. How could we ever add such things? Adding the spins of two different particles also seems unusual if, for example, the particles are far-away from each other. Vectors that live at different places are seldom added: you don't typically add the electric field at one point to the electric field at another point: the sum has no obvious interpretation. This is even more severe in general relativity: you *cannot* add vectors that 'live' at different points of space-time. To add them you need a procedure to first bring them to a common point. Once they both live at that common point you can add them.

I want to make clear, however, that at a basic algebraic level all angular momenta are apples (granny smith, red-delicious, macintosh, fuji, etc.) and therefore they can be added and it is natural to add them. We are not adding apples to oranges; we are adding apples to apples! The physics

requires it: we will see that energy eigenstates will also be eigenstates of operators in the sum of angular momenta. The mathematics allows it: the sum of angular momenta *is* an angular momentum acting in the appropriate *tensor* product. As we will see below, each angular momentum lives on a different vector space, but the sum finds a home in the tensor product of the vector spaces.

What is an angular momentum? It is a triplet $\hat{J}_i$ of Hermitian linear operators on some complex vector space V satisfying the commutation relations

$$[\hat{J}_i\,,\hat{J}_j\,] \;=\; i\hbar\,\epsilon_{ijk}\,\hat{J}_k\,. \tag{1.1}$$

As we have learned, this is a very powerful statement. When coupled with the requirement that no negative norm-squared states exist, it implies that V can be decomposed into sums of representations of angular momentum, all of which are finite dimensional.

Let us now assume we have two angular momenta:

$$\begin{array}{l} \text{Hermitian operators } J_i^{(1)} \text{ acting on } \;V_1 \text{ and satisfying } \;[\hat{J}_i^{(1)}\,,\hat{J}_j^{(1)}\,] \;=\; i\hbar\,\epsilon_{ijk}\,\hat{J}_k^{(1)} \\ \text{Hermitian operators } J_i^{(2)} \text{ acting on } \;V_2 \text{ and satisfying } \;[\hat{J}_i^{(2)}\,,\hat{J}_j^{(2)}\,] \;=\; i\hbar\,\epsilon_{ijk}\,\hat{J}_k^{(2)} \end{array} \tag{1.2}$$

Our claim is that the 'sum' of angular momenta is an angular momentum in the tensor product:

$$\boxed{\hat{J}_i \;\equiv\; \hat{J}_i^{(1)}\otimes\mathbf{1}+\mathbf{1}\otimes\hat{J}_i^{(2)} \text{ satisfies } [\hat{J}_i\,,\hat{J}_j\,] \;=\; i\hbar\,\epsilon_{ijk}\,\hat{J}_k \text{ acting on } \;V_1\otimes V_2\,.} \tag{1.3}$$

Certainly the sum operator, as defined above, is an operator on $V_1\otimes V_2$. It is in fact a Hermitian operator on $V_1\otimes V_2$. We just need to check that the commutator holds

$$\begin{aligned} [\hat{J}_i\,,\,\hat{J}_j\,] \;&=\; \big[\hat{J}_i^{(1)}\otimes\mathbf{1}+\mathbf{1}\otimes\hat{J}_i^{(2)}\;,\;\hat{J}_j^{(1)}\otimes\mathbf{1}+\mathbf{1}\otimes\hat{J}_j^{(2)}\;\big] \\ &=\; \big[\hat{J}_i^{(1)}\otimes\mathbf{1}\,,\,\hat{J}_j^{(1)}\otimes\mathbf{1}\big]+\big[\,\mathbf{1}\otimes\hat{J}_i^{(2)}\;,\mathbf{1}\otimes\hat{J}_j^{(2)}\;\big]\,, \end{aligned} \tag{1.4}$$

since the mixed terms, which represent commutators of the operators in the different spaces vanish:

$$\big[\hat{J}_i^{(1)}\otimes\mathbf{1}\,,\,\mathbf{1}\otimes\hat{J}_j^{(2)}\,]\;=\;0\,,\quad \big[\mathbf{1}\otimes\hat{J}_i^{(2)}\,,\,\hat{J}_j^{(1)}\otimes\mathbf{1}\,]\;=\;0\,. \tag{1.5}$$

(If this is not 'obvious', think about it and do the one-line computation that will make it obvious!). Writing out the commutators we see that (1.4) becomes

$$[\hat{J}_i\,,\,\hat{J}_j\,]\;=\big[\hat{J}_i^{(1)},\,\hat{J}_j^{(1)}\,\big]\otimes\mathbf{1}+\;\mathbf{1}\otimes\big[\;\hat{J}_i^{(2)},\hat{J}_j^{(2)}\;\big]\,. \tag{1.6}$$

We can now use the independent algebras of angular momentum to find

$$\begin{aligned} [\hat{J}_i\,,\,\hat{J}_j\,] \;&=i\hbar\,\epsilon_{ijk}\,\hat{J}_k^{(1)}\otimes\mathbf{1}+\;i\hbar\,\epsilon_{ijk}\,\mathbf{1}\otimes\hat{J}_k^{(2)} \\ &=i\hbar\,\epsilon_{ijk}\,\big(\hat{J}_k^{(1)}\otimes\mathbf{1}+\mathbf{1}\otimes\hat{J}_k^{(2)}\big) \\ &=i\hbar\,\epsilon_{ijk}\hat{J}_k\,, \end{aligned} \tag{1.7}$$

which is what we set out to prove.

It is important to note that had we added the two angular momenta with some arbitrary coefficients we would not have got an angular momentum. Indeed, suppose we use two non-zero complex constants α and β and write

$$\tilde{J}_i \;\equiv\; \alpha\, \hat{J}_j^{(1)} \otimes \mathbf{1} + \beta\, \mathbf{1} \otimes \hat{J}_j^{(2)}\,. \tag{1.8}$$

If either constant is zero we are not really summing. The commutator calculation above this time yields

$$[\tilde{J}_i\,,\, \tilde{J}_j\,] \;=\; i\hbar\,\epsilon_{ijk}\left(\alpha^2 \hat{J}_k^{(1)} \otimes \mathbf{1} + \beta^2\, \mathbf{1} \otimes \hat{J}_k^{(2)}\right). \tag{1.9}$$

We have an algebra of angular momentum if the operator in parenthesis is $\tilde{J}_k$. This requires $\alpha^2 = \alpha$ and $\beta^2 = \beta$. Since neither α nor β is zero, the only solution is $\alpha = \beta = 1$. This confirms that we are using in (1.3) the *unique way* to add two angular momenta to form a new angular momentum.

By the same arguments that hold for any angular momentum on a vector space, the space $V_1 \otimes V_2$ can be decomposed into sums of representations of the algebra of *total* angular momentum. This property gives us a powerful tool to understand the spectrum of the Hamiltonian in the physical state space $V_1 \otimes V_2$.

2 Adding two spin one-half angular momenta

To set up the notation recall that for a spin one-half particle and spin operators $\mathbf{S}$ we write

$$\mathbf{S}^2|s,m\rangle \;=\; \hbar^2\, s(s+1)|s,m\rangle\,, \quad \hat{S}_z|s,m\rangle \;=\; \hbar m\,|s,m\rangle\,, \quad \text{with } s = \tfrac{1}{2}\,, \text{ and } \; m = \pm\tfrac{1}{2}\,. \tag{2.1}$$

The states that span the vector space are thus

$$|\tfrac{1}{2},\tfrac{1}{2}\rangle\,, \quad |\tfrac{1}{2},-\tfrac{1}{2}\rangle\,, \tag{2.2}$$

states that we used to label as $|+\rangle$ and $|-\rangle$, respectively. The action of $\mathbf{S}^2$ on any of these states gives $\frac{3}{4}\hbar^2$ and the action of $\hat{S}_z/\hbar$ gives $\frac{1}{2}$ on the first and $-\frac{1}{2}$ on the second.

We now consider the case in which our system features two spin one-half particles. For the first particle we have the triplet of spin operators $\mathbf{S}^{(1)}$ acting on the vector space V_1 spanned by

$$|\tfrac{1}{2},\tfrac{1}{2}\rangle_1\,, \quad |\tfrac{1}{2},-\tfrac{1}{2}\rangle_1\,, \tag{2.3}$$

For the second particle we have the triplet spin operators $\mathbf{S}^{(2)}$ acting on the vector space V_2 spanned by

$$|\tfrac{1}{2},\tfrac{1}{2}\rangle_2\,, \quad |\tfrac{1}{2},-\tfrac{1}{2}\rangle_2\,. \tag{2.4}$$

We now form the total spin

$$\hat{S}_i \;\equiv\; \hat{S}_i^{(1)} \otimes \mathbf{1} \,+\, \mathbf{1} \otimes \hat{S}_i^{(2)} \tag{2.5}$$

which we write, for brevity as

$$\hat{S}_i \;=\; \hat{S}_i^{(1)} \,+\, \hat{S}_i^{(2)}\,, \quad \text{for example,} \quad \hat{S}_z \;=\; \hat{S}_z^{(1)} \,+\, \hat{S}_z^{(2)}\,, \tag{2.6}$$

with the understanding that each operator on the right-hand sides acts on the appropriate factor in the tensor product. The state space for the dynamics of the two particles must contain the tensor product $V_1 \otimes V_2$ (more spaces might be needed if the particles have orbital angular momentum or they are moving). As we learned before, $V_1 \otimes V_2$ is a four-dimensional complex vector space spanned by the products of states in (2.3) and (2.4):

$$|\tfrac{1}{2},\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},\tfrac{1}{2}\rangle_2\,, \quad |\tfrac{1}{2},\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},-\tfrac{1}{2}\rangle_2\,, \quad |\tfrac{1}{2},-\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},\tfrac{1}{2}\rangle_2\,, \quad |\tfrac{1}{2},-\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},-\tfrac{1}{2}\rangle_2\,. \tag{2.7}$$

It must be possible to organize these states into finite-dimensional representations of the total spin angular momentum, which is well-defined acting on these states. We have four basis states so the possibilities for multiplets of total spin s are

1. Four singlets ($s = 0$).
2. Two doublets ($s = \frac{1}{2}$).
3. One doublet ($s = \frac{1}{2}$) and two singlets ($s = 0$).
4. One triplet ($s = 1$) and one singlet ($s = 0$).
5. One $s = \frac{3}{2}$ multiplet.

It may be instructive at this point if you pause to make sure no other option exists and then to consider which option is the only likely to be true! Guess it!

The main clue is that the states in the tensor product are eigenstates of $\hat{S}_z$, the total z-component of angular momentum. We see by inspection of (2.7) that the possible values of $\hat{S}_z/\hbar$ and $+1, 0$, and -1. Since we have a state with $m = +1$ and no state with higher m we must have a triplet $s = 1$. Thus the only option is the fourth one listed above: a triplet and a singlet. This is written as

$$(s = \tfrac{1}{2}) \otimes (s = \tfrac{1}{2}) \;=\; (s = 1) \oplus (s = 0) \tag{2.8}$$

Note that in the left-hand side we have the tensor product of the two state spaces, but in the right-hand side the *direct sum* of the representations of total spin angular momentum. This is a fundamental result and is written more briefly as

$$\boxed{\tfrac{1}{2} \otimes \tfrac{1}{2} \;=\; 1 \oplus 0\,.} \tag{2.9}$$

Let us understand this very explicitly by organizing the basis states according to the eigenvalue m of $\hat{S}_z/\hbar$. We readily observe that

$$\begin{aligned} m \;&=\; \;\;1: \qquad |\tfrac{1}{2},\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},\tfrac{1}{2}\rangle_2\,, \\ m \;&=\; \;\;0: \qquad |\tfrac{1}{2},\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},-\tfrac{1}{2}\rangle_2\,, \quad |\tfrac{1}{2},-\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},\tfrac{1}{2}\rangle_2\,, \\ m \;&=\; -1: \qquad |\tfrac{1}{2},-\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},-\tfrac{1}{2}\rangle_2\,. \end{aligned} \tag{2.10}$$

We see that we get two states with $m = 0$. This is as it should be. One linear combination of these two states must correspond to the $m = 0$ state of the triplet and another linear combination must correspond to the singlet $s = m = 0$. Those two states are in fact entangled states. Denoting by $|s, m\rangle$ the eigenstates of $\mathbf{S}^2$ and $\hat{S}_z$ (total spin) we must have a triplet with states

$$\begin{aligned} |1,1\rangle &= |\tfrac{1}{2},\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},\tfrac{1}{2}\rangle_2 \,, \\ |1,0\rangle &= \alpha\, |\tfrac{1}{2},\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},-\tfrac{1}{2}\rangle_2 \;+\; \beta\, |\tfrac{1}{2},-\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},\tfrac{1}{2}\rangle_2 \,, \\ |1,-1\rangle &= |\tfrac{1}{2},-\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},-\tfrac{1}{2}\rangle_2 \,. \end{aligned} \tag{2.11}$$

for some constants α and β, as well as a singlet

$$|0,0\rangle \;=\; \gamma\, |\tfrac{1}{2},\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},-\tfrac{1}{2}\rangle_2 \;+\; \delta\, |\tfrac{1}{2},-\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},\tfrac{1}{2}\rangle_2 \,, \tag{2.12}$$

for some constants γ and δ. We must determine these four constants. Let us begin with the state in the triplet. Recalling the general formula

$$J_\pm |j,m\rangle \;=\; \hbar\sqrt{j(j+1) - m(m\mp 1)}\, |j, m\pm 1\rangle \tag{2.13}$$

we quickly note that

$$\begin{aligned} J_-|1,1\rangle &= \hbar\sqrt{2}\, |1,0\rangle \,, \\ J_-\, |\tfrac{1}{2},\, \tfrac{1}{2}\rangle &= \hbar \sqrt{\tfrac{1}{2}\cdot\tfrac{3}{2} - \tfrac{1}{2}\cdot(-\tfrac{1}{2})}\; |\tfrac{1}{2},-\tfrac{1}{2}\rangle \;=\; \hbar\, |\tfrac{1}{2},-\tfrac{1}{2}\rangle \\ J_+|\tfrac{1}{2},-\tfrac{1}{2}\rangle &= \hbar\sqrt{\tfrac{1}{2}\cdot\tfrac{3}{2} - (-\tfrac{1}{2})\cdot(\tfrac{1}{2})}\; |\tfrac{1}{2},-\tfrac{1}{2}\rangle \;=\; \hbar\, |\tfrac{1}{2},\tfrac{1}{2}\rangle \end{aligned} \tag{2.14}$$

In here J can be spin, of course. We now apply the lowering operator $S_- \;=\; S_-^{(1)} + S_-^{(2)}$ to the top state in the triplet. We have

$$S_-|1,1\rangle \;=\; (S_-^{(1)}|\tfrac{1}{2},\tfrac{1}{2}\rangle_1) \otimes |\tfrac{1}{2},\tfrac{1}{2}\rangle_2 \;+\; |\tfrac{1}{2},\tfrac{1}{2}\rangle_1 \otimes (S_-^{(2)}|\tfrac{1}{2},\tfrac{1}{2}\rangle_2) \,, \tag{2.15}$$

Using the results in (2.14) we find

$$\sqrt{2}\hbar\, |1,0\rangle \;=\; \hbar|\tfrac{1}{2},-\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},\tfrac{1}{2}\rangle_2 \;+\; |\tfrac{1}{2},\tfrac{1}{2}\rangle_1 \otimes \hbar\, |\tfrac{1}{2},-\tfrac{1}{2}\rangle_2 \,. \tag{2.16}$$

Cancelling the common factors of $\hbar$ and switching the order of the terms we find that the $|1,0\rangle$ state takes the form

$$|1,0\rangle \;=\; \tfrac{1}{\sqrt{2}} \Big(|\tfrac{1}{2},\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},-\tfrac{1}{2}\rangle_2 \;+\; |\tfrac{1}{2},-\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},\tfrac{1}{2}\rangle_2 \Big) \,. \tag{2.17}$$

Having found the $m = 0$ state of the $s = 1$ multiplet there are a number of ways of finding the $m = 0$ of the $s = 0$ singlet. One way is orthogonality: the latter state must be orthogonal to the $m = 0$ state above because these are two states with different s, thus different eigenvalues for the Hermitian operator $\mathbf{S}^2$. Since the overall sign or phase is irrelevant, we can simply take for the singlet

$$|0,0\rangle \;=\; \tfrac{1}{\sqrt{2}} \Big(|\tfrac{1}{2},\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},-\tfrac{1}{2}\rangle_2 \;-\; |\tfrac{1}{2},-\tfrac{1}{2}\rangle_1 \otimes |\tfrac{1}{2},\tfrac{1}{2}\rangle_2 \Big) \,. \tag{2.18}$$

You probably remember that we found this state in a previous chapter by requiring that it is annihilated by the sum of spin angular momentum operators. This is exactly the condition for a singlet.

As an interesting calculation and good practice with the operators, let us confirm that $\mathbf{S}^2$ is zero acting on $|0,0\rangle$. For this it is useful to note that

$$\begin{aligned}\mathbf{S}^2 \;=\; (\mathbf{S}^{(1)}+\mathbf{S}^{(2)})^2 \;&=\; (\mathbf{S}^{(1)})^2+(\mathbf{S}^{(2)})^2+2\mathbf{S}^{(1)}\cdot\mathbf{S}^{(2)}\\ &= (\mathbf{S}^{(1)})^2+(\mathbf{S}^{(2)})^2+\,S_+^{(1)}S_-^{(2)}+S_-^{(1)}S_+^{(2)}+2S_z^{(1)}S_z^{(2)}\end{aligned} \tag{2.19}$$

where in the second step we used the general result

$$\hat{\mathbf{J}}^{(1)}\cdot\hat{\mathbf{J}}^{(2)} \;=\; \tfrac{1}{2}\big(\hat{J}_+^{(1)}\hat{J}_-^{(2)}+\hat{J}_-^{(1)}\hat{J}_+^{(2)}\big)\;+\;\hat{J}_z^{(1)}\hat{J}_z^{(2)} \tag{2.20}$$

which is valid for arbitrary angular momenta. Written in explicit tensor notation it reads

$$\hat{\mathbf{J}}^{(1)}\cdot\hat{\mathbf{J}}^{(2)} \;\equiv\; \sum_{i=1}^{3}\hat{J}_i^{(1)}\otimes\hat{J}_i^{(2)} \;=\; \tfrac{1}{2}\big(\hat{J}_+^{(1)}\otimes\hat{J}_-^{(2)}+\hat{J}_-^{(1)}\otimes\hat{J}_+^{(2)}\big)\;+\;\hat{J}_z^{(1)}\otimes\hat{J}_z^{(2)} \tag{2.21}$$

All states in the singlet have $s_1=s_2=\frac{1}{2}$, and therefore $(\mathbf{S}^{(1)})^2=(\mathbf{S}^{(2)})^2=\frac{3}{4}\hbar^2$. We thus have

$$\mathbf{S}^2\,|0,0\rangle \;=\; \frac{3}{2}\hbar^2\,|0,0\rangle+\;\big(S_+^{(1)}S_-^{(2)}+S_-^{(1)}S_+^{(2)}+2S_z^{(1)}S_z^{(2)}\big)|0,0\rangle \tag{2.22}$$

It is simple to see that

$$2S_z^{(1)}S_z^{(2)}\,|0,0\rangle \;=\; 2\,\tfrac{\hbar}{2}\cdot(-\tfrac{\hbar}{2})|0,0\rangle \;=\; -\tfrac{1}{2}\hbar^2\,|0,0\rangle\,, \tag{2.23}$$

because the singlet is a superposition of tensor states where each has one state up and one state down. Similarly recalling that

$$S_\pm|\tfrac{1}{2},\mp\tfrac{1}{2}\rangle \;=\; \hbar^2|\tfrac{1}{2},\pm\tfrac{1}{2}\rangle \tag{2.24}$$

we quickly find that

$$\big(S_+^{(1)}S_-^{(2)}+S_-^{(1)}S_+^{(2)}\big)|0,0\rangle \;=\; -\hbar^2\,|0,0\rangle\,, \tag{2.25}$$

since each of the operators $S_+^{(1)}S_-^{(2)}$ and $S_-^{(1)}S_+^{(2)}$ kills one term in the singlet (call it the first) and acting on the other (the second) it gives $\hbar^2$ times the first. Check it, anyway. Back in (2.26) we get

$$\mathbf{S}^2\,|0,0\rangle \;=\; \tfrac{3}{2}\hbar^2\,|0,0\rangle+\;\big(-\hbar^2-\tfrac{1}{2}\hbar^2\big)|0,0\rangle \;=\; 0 \tag{2.26}$$

as we wanted to show.

Let us summarize our results. The triplet states and singlet states are given by

$$\begin{aligned}|1,1\rangle \;&=\; |\tfrac{1}{2},\tfrac{1}{2}\rangle_1\otimes|\tfrac{1}{2},\tfrac{1}{2}\rangle_2\,,\\ |1,0\rangle \;&=\; \tfrac{1}{\sqrt{2}}\Big(\,|\tfrac{1}{2},\tfrac{1}{2}\rangle_1\otimes|\tfrac{1}{2},-\tfrac{1}{2}\rangle_2\;+\;|\tfrac{1}{2},-\tfrac{1}{2}\rangle_1\otimes|\tfrac{1}{2},\tfrac{1}{2}\rangle_2\,\Big)\\ |1,-1\rangle \;&=\; |\tfrac{1}{2},-\tfrac{1}{2}\rangle_1\otimes|\tfrac{1}{2},-\tfrac{1}{2}\rangle_2\,.\\ |0,0\rangle \;&=\; \tfrac{1}{\sqrt{2}}\Big(\,|\tfrac{1}{2},\tfrac{1}{2}\rangle_1\otimes|\tfrac{1}{2},-\tfrac{1}{2}\rangle_2\;-\;|\tfrac{1}{2},-\tfrac{1}{2}\rangle_1\otimes|\tfrac{1}{2},\tfrac{1}{2}\rangle_2\,\Big)\,.\end{aligned} \tag{2.27}$$

For briefer notation we replace

$$|\tfrac{1}{2},\tfrac{1}{2}\rangle \;\to\; |\uparrow\rangle\,, \quad \text{and} |\tfrac{1}{2},-\tfrac{1}{2}\rangle \;\to\; |\downarrow\rangle \tag{2.28}$$

We then have

$$\begin{aligned} |1,1\rangle &= |\uparrow\rangle_1 \otimes |\uparrow\rangle_2\,, \\ |1,0\rangle &= \tfrac{1}{\sqrt{2}}\Big(|\uparrow\rangle_1 \otimes |\downarrow\rangle_2 + |\downarrow\rangle_1 \otimes |\uparrow\rangle_2\Big) \\ |1,-1\rangle &= |\downarrow\rangle_1 \otimes |\downarrow\rangle_2\,. \\ |0,0\rangle &= \tfrac{1}{\sqrt{2}}\Big(|\uparrow\rangle_1 \otimes |\downarrow\rangle_2 - |\downarrow\rangle_1 \otimes |\uparrow\rangle_2\Big)\,. \end{aligned} \tag{2.29}$$

With the understanding that the first arrow refers to the first particle and the second arrow to the second particle, we can finally write all of this quite briefly

$$\boxed{\begin{aligned} |1,1\rangle &= |\uparrow\uparrow\rangle\,, \\ |1,0\rangle &= \tfrac{1}{\sqrt{2}}\big(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle\big)\,, \\ |1,-1\rangle &= |\downarrow\downarrow\rangle\,. \\ |0,0\rangle &= \tfrac{1}{\sqrt{2}}\big(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle\big)\,. \end{aligned}} \tag{2.30}$$

This decomposition of the tensor product of two spin one-half state spaces is useful for the computation of the hyperfine splitting in the hydrogen atom, where the relevant spins are those of the proton and the electron.

3 Feynman-Hellman lemma

This is a nice and simple result that gives us insight into the way energy levels change under perturbations of the Hamiltonian.

Consider a Hamiltonian $H(\lambda)$ with a parameter λ and some *normalized* energy eigenstate $\psi(\lambda)$ with some energy $E(\lambda)$. We claim that

$$\boxed{\frac{dE(\lambda)}{d\lambda} = \Big\langle \psi(\lambda)\,\Big|\,\frac{dH(\lambda)}{d\lambda}\,\Big|\,\psi(\lambda)\Big\rangle\,.} \tag{3.1}$$

This is the Feynman-Hellman (FH) lemma. Note that by assumption $\psi(\lambda)$ is an eigenstate and is normalized for all values of λ. The proof of (3.1) is straightforward. Note that

$$H(\lambda)|\psi(\lambda)\rangle = E(\lambda)|\psi(\lambda)\rangle\,, \tag{3.2}$$

and therefore

$$E(\lambda) = \langle\psi(\lambda)|H(\lambda)|\psi(\lambda)\rangle\,. \tag{3.3}$$

Now differentiate this relation with respect to λ to get

$$\frac{dE}{d\lambda}(\lambda) \;=\; \Big(\frac{d}{d\lambda}\langle\psi(\lambda)|\Big)H(\lambda)|\psi(\lambda)\rangle + \langle\psi(\lambda)|\frac{dH(\lambda)}{d\lambda}|\psi(\lambda)\rangle + \langle\psi(\lambda)|H(\lambda)\Big(\frac{d}{d\lambda}|\psi(\lambda)\rangle\Big) \tag{3.4}$$

Using the eigenstate equation for the first and last term,

$$\frac{dE}{d\lambda}(\lambda) \;=\; E(\lambda)\Big(\frac{d}{d\lambda}\langle\psi(\lambda)|\Big)|\psi(\lambda)\rangle + \langle\psi(\lambda)|\frac{dH(\lambda)}{d\lambda}|\psi(\lambda)\rangle + E(\lambda)\langle\psi(\lambda)|\Big(\frac{d}{d\lambda}|\psi(\lambda)\rangle\Big) \tag{3.5}$$

We thus have

$$\frac{dE}{d\lambda}(\lambda) \;=\; \langle\psi(\lambda)|\frac{dH(\lambda)}{d\lambda}|\psi(\lambda)\rangle + E(\lambda)\frac{d}{d\lambda}\langle\psi(\lambda)|\psi(\lambda)\rangle\,. \tag{3.6}$$

Since the state is normalized for all λ the last term vanishes and we are left with the result we wanted to prove. Note that the FH lemma is an exact result.

We can use the FH lemma to formulate an approximate result that is quite useful. Consider a Hamiltonian where the λ dependence is as follows

$$H(\lambda) = H(0) + \lambda H_1 \;=\; H(0) + \Delta H\,. \tag{3.7}$$

Here we have identified λH_1 with a perturbation ΔH of the Hamiltonian. Let now $\psi(\lambda)$ be the normalized eigenstate with energy $E(\lambda)$ for all λ. Now by FH

$$\frac{dE(\lambda)}{d\lambda} \;=\; \langle\psi(\lambda)|H_1|\psi(\lambda)\rangle \tag{3.8}$$

and evaluating at $\lambda = 0$ we get

$$\frac{dE(\lambda)}{d\lambda}\Big|_{\lambda=0} \;=\; \langle\psi(0)|H_1|\psi(0)\rangle \tag{3.9}$$

We now note that for small λ

$$E(\lambda) \;=\; E(0) + \lambda\frac{dE(\lambda)}{d\lambda}\Big|_{\lambda=0} + \mathcal{O}(\lambda^2) \tag{3.10}$$

which we write as

$$E(\lambda) \;-\; E(0) \;=\; \langle\psi(0)|\Delta H|\psi(0)\rangle + \mathcal{O}(\Delta H^2) \tag{3.11}$$

The left-hand side is the shift ΔE in energy and is given, finally, to first order in the perturbation ΔH by

$$\boxed{\;\Delta E \;=\; \langle\psi(0)|\Delta H|\psi(0)\rangle\,.\;} \tag{3.12}$$

This is the simplest and most important result in perturbation theory. To compute the energy shift of a state we must just compute the expectation value of the change ΔH of the Hamiltonian in the *unperturbed* state.

The above result can be used without concern when we have a single non-degenerate state of the original Hamiltonian H_0. We will have the situation, however, where we have a set of *degenerate* eigenstates of the Hamiltonian H_0 and we want to figure out what happens to them. One cannot use directly (3.12) since the derivation assumes that the state smoothly evolves into an eigenstate as

λ changes. The problem is that the original states we chose may not remain eigenstates under the perturbation: only certain linear combinations of the original eigenstates become the perturbed energy eigenstates. One must first figure out which are those particular linear combinations and then we can apply (3.12) to any of *those* states.

4 Hyperfine splitting

5 Spin-Orbit coupling

5.1 The physical interaction

5.2 Complete set of observables

5.3 Level splitting

5.4 Computation of $\mathbf{1}\otimes\frac{1}{2}$

6 General aspects of addition of angular momentum

The key formula is

$$j_1 \otimes j_2 \;=\; (j_1 + j_2) \oplus (j_1 + j_2 - 1) \oplus ... \oplus |j_1 - j_2| \,. \tag{6.13}$$

This must be viewed as an equality between vector spaces. Or perhaps, more explicitly, as a way to write the vector space on the left-hand as a direct sum of interesting invariant subspaces. The vector space on the left-hand side is the tensor product of the vector space associated with the multiplet j_1 of $\mathbf{J_1}$ and the vector space associated with the multiplet j_2 of $\mathbf{J_2}$. On the right-hand side we have a collection of multiplets of the angular moment sum $\mathbf{J} = \mathbf{J_1} + \mathbf{J_2}$. Thus each term on the right-hand side is an invariant subspace of the total angular momentum $\mathbf{J}$. Since various invariant subspaces appear, the representation is called reducible.

7 Hydrogen atom and hidden symmetry

Let us review the hydrogen atom spectrum we wish to derive by algebraic methods. We have energy levels with energies E_n, where

$$E_n = -\frac{e^2}{2a_0}\frac{1}{n^2}\,, \quad n = 1, 2, \ldots \tag{7.1}$$

For each value n we have the following ℓ multiplets of angular momentum $\mathbf{L}$:

$$\ell = 0, 1, \ldots, n-1\,. \tag{7.2}$$

A simple computation shows that this gives a total of n^2 states. So at each n have a space $\mathcal{H}_n$ of degenerate energy eigenstates. This space has dimension n^2 and can be written as the direct sum of angular momentum multiplets

$$\mathcal{H}_n \;=\; (\ell = n-1) \oplus (\ell = n-2) \oplus \cdots \oplus (\ell = 0)\,. \tag{7.3}$$

We begin our analysis by recalling the hydrogen Hamiltonian, which takes the form

$$H = \frac{\mathbf{p^2}}{2m} - \frac{e^2}{r}\,. \tag{7.4}$$

There are conserved orbital angular momentum operators $\mathbf{L}$:

$$[H\,,\mathbf{L}\,] \;=\; 0\,, \qquad \mathbf{L}\times\mathbf{L} \;=\; i\hbar\mathbf{L}\,. \tag{7.5}$$

This Runge-Lenz vector R is defined to be:

$$\mathbf{R} \;\equiv\; \frac{1}{2me^2}(\mathbf{p}\times\mathbf{L} - \mathbf{L}\times\mathbf{p}) - \frac{\mathbf{r}}{r}\,. \tag{7.6}$$

As written this operator is Hermitian. It is also conserved, as it too commutes with the Hamiltonian (this was a practice problem!)

$$[H,\mathbf{R}\,] \;=\; 0\,. \tag{7.7}$$

Two alternative rewritings of $\mathbf{R}$ are possible when we use the identity

$$\mathbf{p}\times\mathbf{L} \;=\; -\mathbf{L}\times\mathbf{p} + 2i\hbar\mathbf{p}\,. \tag{7.8}$$

These are

$$\mathbf{R} \;=\; \frac{1}{me^2}\big(\mathbf{p}\times\mathbf{L} - i\hbar\mathbf{p}\big) - \frac{\mathbf{r}}{r} \;=\; \frac{1}{me^2}\big(-\mathbf{L}\times\mathbf{p} + i\hbar\mathbf{p}\big) - \frac{\mathbf{r}}{r}\,. \tag{7.9}$$

If $\mathbf{R}$ and $\mathbf{L}$ are conserved, all of their scalar products are conserved too since the Hamiltonian would commute with them using the derivation property of commutators. We are already familiar with the Casimir $\mathbf{L}^2$ that in fact commutes with all $\hat{L}_i$. Let us consider now $\mathbf{R}^2$, which must be a conserved scalar. Its evaluation is a practice problem (homework) and gives

$$\mathbf{R}^2 \;=\; 1 + \frac{2H}{me^4}\left(\mathbf{L}^2 + \hbar^2\right). \tag{7.10}$$

Notice that H, which appears in the above right-hand side can be moved, if desired, to the right of the parenthesis, as it commutes with $\mathbf{L}$. The right-hand side is indeed a conserved scalar. We now look into $\mathbf{R}\cdot\mathbf{L}$, which must also be conserved. First recall that

$$\mathbf{r}\cdot\mathbf{L} \;=\; 0\,, \quad \mathbf{p}\cdot\mathbf{L} \;=\; 0\,. \tag{7.11}$$

Using these and the first equality in (7.9) we find

$$\mathbf{R}\cdot\mathbf{L} \;=\; \frac{1}{me^2}\,(\mathbf{p}\times\mathbf{L})\cdot\mathbf{L}\,. \tag{7.12}$$

But we now notice that

$$(\mathbf{p}\times\mathbf{L})\cdot\mathbf{L} \;=\; \epsilon_{ijk}p_jL_k\,L_i \;=\; p_j\,\epsilon_{jki}\,L_kL_i \;=\; p_j(\mathbf{L}\times\mathbf{L})_j = \mathbf{p}\cdot\,(i\hbar\mathbf{L}) = 0\,. \tag{7.13}$$

As a result, we have shown that

$$\boxed{\;\mathbf{R}\cdot\mathbf{L} \;=\; 0\,.\;} \tag{7.14}$$

This relation is manifest classically where $\mathbf{R}$ lies along the major axis of the ellipse and $\mathbf{L}$ is perpendicular to the plane of the ellipse.

We now consider the commutation relations. First recall the definition of a vector (operator) under rotations. For such vector $\mathbf{v}$ with components $\hat{v}_i$ we have

$$[\, L_i \,,\, \hat{v}_j \,] \;=\; i\hbar\epsilon_{ijk}\,\hat{v}_k\,. \tag{7.15}$$

This relation can be written using cross products. For this consider the i-th componentd of the following vector

$$\begin{aligned}(\mathbf{L}\times\mathbf{v}+\mathbf{v}\times\mathbf{L})_i \;&=\; \epsilon_{ijk}(L_j\,\hat{v}_k+\hat{v}_j\,L_k)\\ &=\; \epsilon_{ijk}(\,L_j\hat{v}_k-\hat{v}_kL_j\,)\\ &=\; \epsilon_{ijk}\,[\,L_j\,,\,\hat{v}_k\,]\\ &=\; \epsilon_{ijk}\;i\hbar\epsilon_{jkl}\,\hat{v}_l\\ &=\; \epsilon_{ijk}\,\epsilon_{ljk}\;i\hbar\,\hat{v}_l\\ &=2\delta_{il}\;i\hbar\,\hat{v}_l\;=\;2i\hbar\,\hat{v}_i\,,\end{aligned} \tag{7.16}$$

which shows that the condition that $\mathbf{v}$ is a vector under rotations reads

$$\mathbf{L}\times\mathbf{v}+\mathbf{v}\times\mathbf{L}\;=\;2i\hbar\,\mathbf{v}\,. \tag{7.17}$$

The Runge-Lenz vector $\mathbf{R}$ is a vector under rotations. This we know without any computation, since it is built using cross products from $\mathbf{p}$ and $\mathbf{L}$, which are both vectors under rotations. Therefore we must have

$$\mathbf{L}\times\mathbf{R}\;+\;\mathbf{R}\times\mathbf{L}\;=\;2i\hbar\,\mathbf{R}\,. \tag{7.18}$$

We also record the index form of the statement:

$$[\, L_i,\, R_j\,] \;=\; i\hbar\,\epsilon_{ijk}\,R_k\,. \tag{7.19}$$

Exercise. Show that $\mathbf{R}\cdot\mathbf{L}\;=\;\mathbf{L}\cdot\mathbf{R}$.

In order to understand the commutator of two $\mathbf{R}$ operators we need a simple result: The commutator of two conserved operators is a conserved operator. To prove this consider two conserved operator S_1 and S_2:

$$[\, S_1,\, H\,] = [\, S_2,\, H\,] \;=\; 0\,. \tag{7.20}$$

The Jacobi identity applied to the operators S_1, S_2, and H reads:

$$\big[\,[S_1\,,\,S_2\,]\,,H\big] \;+\; \big[\,[H\,,\,S_1\,]\,,S_2\big] \;+\; \big[\,[S_2\,,\,H\,]\,,S_1\big] \;=\; 0\,. \tag{7.21}$$

This gives

$$\big[\,[S_1\,,\,S_2\,]\,,H\big] \;=\; \big[\,[S_1\,,\,H\,]\,,S_2\big] \;-\; \big[\,[S_2\,,\,H\,]\,,S_1\big]\,. \tag{7.22}$$

Each term on the right-hand side vanishes, the first by conservation of S_1 and the second by conservation of S_2. It follows that

$$\big[\,[S_1\,,\,S_2\,]\,,H\big]\;=\;0\,, \tag{7.23}$$

which, as desired, states the conservation of the commutator $[S_1, S_2]$. This result is useful, for it tells us that the commutator $[R_i\,,R_j]$ must be some conserved object. We can focus, equivalently, on the cross product of two $\mathbf{R}$'s which encodes the commutator. We must have

$$\mathbf{R}\times\mathbf{R}\;=\;(\cdots)\text{ ``conserved vector''}\,, \tag{7.24}$$

where the dots represent some conserved scalar. Since $\mathbf{L}$ and $\mathbf{R}$ are conserved vectors, the possible vectors on the right-hand side are $\mathbf{L}, \mathbf{R}$, and $\mathbf{L}\times\mathbf{R}$. To narrow down the options we examine the behavior of various vectors under the parity transformation

$$\mathbf{r}\to-\mathbf{r}\,. \tag{7.25}$$

Under this transformation we must have

$$\mathbf{p}\to-\mathbf{p}\,,\quad \mathbf{L}\to\mathbf{L}\,,\quad \mathbf{R}\to-\mathbf{R}\,. \tag{7.26}$$

The first follows from the preservation of the commutators between $\mathbf{r}$ and $\mathbf{p}$. The second from $\mathbf{L}=\mathbf{r}\times\mathbf{p}$, and the third from the expression for $\mathbf{R}$ in terms of $\mathbf{r}, \mathbf{p}, \mathbf{L}$. Since the left-hand side of (7.24) does not change sign under the parity transformation, nor should the right-hand side. From our choices $\mathbf{L}, \mathbf{R}$, and $\mathbf{L}\times\mathbf{R}$, we can only have $\mathbf{L}$. We must therefore have

$$\mathbf{R}\times\mathbf{R}\;=\;(\cdots)\mathbf{L}\,, \tag{7.27}$$

where, again, the expression in parenthesis must be a conserved scalar. A calculation (that we don't show here) gives

$$\mathbf{R}\times\mathbf{R}\;=\;i\hbar\Big(-\frac{2H}{me^4}\Big)\,\mathbf{L}\,. \tag{7.28}$$

This completes the determination of all commutators relevant to $\mathbf{L}$ and $\mathbf{R}$.

Now we come to the main point. We will derive algebraically the full characterization of the subspaces of degenerate energy eigenstates. For this we will assume we have one such subspace $\mathcal{H}_\nu$ at some energy E_ν, where ν is a parameter to be specified below. We will look at our operators in that subspace. In our operator relations (7.10) and (7.28) we are actually allowed to replace H by the energy E_ν, given that H, which commutes with $\mathbf{L}$, can be brought to the right, to be directly in front of the states. We can therefore set H equal to the energy, that we write in the form

$$H\;=E_\nu\;=\;-\frac{me^4}{2\hbar^2}\frac{1}{\nu^2}\,,\quad \nu\in\mathbb{R}\,. \tag{7.29}$$

Here ν is a unit-free, real constant, to be determined. The rest of the factors, except for a convenient factor of two, provide the right units. Of course, we know that the correct answer for these energies

emerge if ν is a positive integer. This, however, is something we will be able to derive. It follows from the above equation that we can set

$$-\frac{2H}{me^4} \;=\; \frac{1}{\hbar^2\nu^2}\,. \tag{7.30}$$

We can use this expression to simplify our key relations (7.28) and (7.10):

$$\begin{aligned} \mathbf{R}\times\mathbf{R} \;&=\; i\hbar\,\frac{1}{\hbar^2\nu^2}\,\mathbf{L}\,, \\ \mathbf{R}^2 \;&=\; 1\,-\,\frac{1}{\hbar^2\nu^2}\,(\mathbf{L}^2+\hbar^2)\,. \end{aligned} \tag{7.31}$$

A little further rearrangements give:

$$\boxed{\begin{aligned} (\hbar\nu\mathbf{R})\times(\hbar\nu\mathbf{R}) \;&=\; i\hbar\,\mathbf{L}\,, \\ \mathbf{L}^2\,+\,\hbar^2\nu^2\mathbf{R}^2 \;&=\; \hbar^2(\nu^2-1)\,. \end{aligned}} \tag{7.32}$$

These are clear and simple algebraic relations between our operators. The first one shows that $\hbar\nu\mathbf{R}$ has the units of angular momentum and kind of behaves like one, except that the operator to the right is not $\hbar\nu\mathbf{R}$ but rather $\mathbf{L}$.

Our next step is to show that with the help of $\mathbf{L}$ and $\mathbf{R}$ we can construct two independent, commuting algebras of angular momentum. Of course, it is clear that $\mathbf{L}$ is an algebra of angular momentum. But, using suitable linear combinations of $\mathbf{L}$ and $\mathbf{R}$ we will obtain two such algebras. Indeed, define

$$\begin{aligned} \mathbf{J}_1 \;&\equiv\; \tfrac{1}{2}(\mathbf{L}+\hbar\nu\mathbf{R})\,, \\ \mathbf{J}_2 \;&\equiv\; \tfrac{1}{2}(\mathbf{L}-\hbar\nu\mathbf{R})\,. \end{aligned} \tag{7.33}$$

Of course, we can solve for $\mathbf{L}$ and $\hbar\nu\mathbf{R}$ in terms of $\mathbf{J}_1$ and $\mathbf{J}_2$:

$$\begin{aligned} \mathbf{L} \;&=\; \mathbf{J}_1+\mathbf{J}_2 \\ \hbar\nu\mathbf{R} \;&=\; \mathbf{J}_1-\mathbf{J}_2\,. \end{aligned} \tag{7.34}$$

It is important to realize that $\mathbf{L}$ is nothing but the sum of $\mathbf{J}_1$ and $\mathbf{J}_2$. We now claim that the operators $\mathbf{J}_1$ and $\mathbf{J}_2$ commute with each other. This is quickly confirmed by direct computation:

$$\begin{aligned} [J_{1i},J_{2j}] \;&=\; \tfrac{1}{4}\left[L_i+\hbar\nu R_i\,,\,L_j-\hbar\nu R_j\right] \\ &=\; \tfrac{1}{4}\big(\,i\hbar\,\epsilon_{ijk}L_k-\hbar\nu[L_i,R_j]-\hbar\nu[L_j,R_i]-i\hbar\,\epsilon_{ijk}L_k\,\big) \;=\; 0\,, \end{aligned} \tag{7.35}$$

where we noted that the first and last terms on the right-hand side cancel each other out, and the second and third terms also cancel each other out using (7.19). Now we want to show that $\mathbf{J}_1$ and $\mathbf{J}_2$ are indeed angular momenta. We check both operators at the same time using the notation $\mathbf{J}_{1/2}$ with

the subscripts '1' and '2' going with the top and bottom signs, respectively, on the right-hand side:

$$\begin{aligned}\mathbf{J}_{1/2}\times\mathbf{J}_{1/2} &= \tfrac{1}{4}(\mathbf{L}\pm\hbar\nu\mathbf{R})\times(\mathbf{L}\pm\hbar\nu\mathbf{R})\\ &= \tfrac{1}{4}\big(i\hbar\mathbf{L}+i\hbar\mathbf{L}\pm(\mathbf{L}\times\hbar\nu\mathbf{R}+\hbar\nu\mathbf{R}\times\mathbf{L})\big)\\ &= \tfrac{1}{4}(2i\hbar\mathbf{L}\pm 2i\hbar\hbar\nu\mathbf{R})\\ &= i\hbar\tfrac{1}{2}(\mathbf{L}\pm\hbar\nu\mathbf{R})=\mathbf{J}_{1/2}\,.\end{aligned} \tag{7.36}$$

In the first step we used the first equation in (7.32) and in the second step (7.18). In summary, we have confirmed that $\mathbf{J}_1$ and $\mathbf{J}_2$ are indeed two commuting angular momentum operators:

$$\begin{aligned}\mathbf{J}_1\times\mathbf{J}_1 &= i\hbar\,\mathbf{J}_1\,,\\ \mathbf{J}_2\times\mathbf{J}_2 &= i\hbar\,\mathbf{J}_2\,,\\ [\,\mathbf{J}_1\,,\,\mathbf{J}_2\,] &= 0\,.\end{aligned} \tag{7.37}$$

The constraint $\mathbf{R}\cdot\mathbf{L}=0$ gives us crucial information on the angular momenta. Using (7.34) and the the commutativity of $\mathbf{J}_1$ with $\mathbf{J}_2$ we find

$$(\mathbf{J}_1+\mathbf{J}_2)(\mathbf{J}_1-\mathbf{J}_2)=0 \quad\to\quad \mathbf{J}_1^2 \;=\; \mathbf{J}_2^2\,. \tag{7.38}$$

Both angular momenta have the same 'magnitude' on the subspace $\mathcal{H}_\nu$ of degenerate energy eigenstates. Let us look at $\mathbf{J}_1^2$. Again, using $\mathbf{R}\cdot\mathbf{L}=0$ and the second of (7.32), we find

$$\mathbf{J}_1^2 \;=\; \tfrac{1}{4}\left(\mathbf{L}^2+\hbar^2\nu^2\mathbf{R}^2\right) \;=\; \tfrac{1}{4}\,\hbar^2(\nu^2-1)\,. \tag{7.39}$$

Note that the energy parameter ν controls the magnitude of $\mathbf{J}_1^2$. Since both $\mathbf{J}_1$ and $\mathbf{J}_2$ commute with the Hamiltonian, the degenerate subspace $\mathcal{H}_\nu$ must furnish a representation of both of these angular momenta! Any such representation must be specified by some j_1 and some j_2, but the equality $\mathbf{J}_1^2=\mathbf{J}_2^2$ implies $j_1=j_2\equiv j$. We thus have

$$\mathbf{J}_1^2 \;=\; \mathbf{J}_2^2 \;=\; \frac{1}{4}\hbar^2(\nu^2-1) \;=\; \hbar^2 j(j+1)\,. \tag{7.40}$$

Since j is an angular momentum it is quantized: $2j\in\mathbb{Z}$. This means we got the quantization of the energy! Solving for ν in terms of j we find

$$\nu^2=1+4j(j+1)=4j^2+4j+1=(2j+1)^2 \quad\to\quad \nu \;=\; 2j+1\,. \tag{7.41}$$

Note that as anticipated, the energy is determined by the value of j. This shows that in fact each subspace $\mathcal{H}_\nu$ of degenerate energy eigenstates cannot carry more than one value of j. As j runs over all possible values ν takes all positive integer values and thus can be indentified with the principal quantum number n:

$$\begin{aligned} j &= 0,\ \tfrac{1}{2},\ 1,\ \tfrac{3}{2},\ldots\\ n\equiv\nu \;=\; 2j+1 &= 1,\ 2,\ 3,\ 4,\ldots\,.\end{aligned} \tag{7.42}$$

We have therefore recovered exactly the quantization of the energy levels in the hydrogen atom! What is the structure of the degenerate subspace $\mathcal{H}_n$? We already know that each state must be an eigenstate of $\mathbf{J}_1$ with eigenvalue $\hbar^2 j(j+1)$ and *at the same time* an eigenstate of $\mathbf{J}_2$ with the same eigenvalue. Since the two angular momenta are independent, the states must live in a tensor product of the space that carries the representation j of $\mathbf{J}_1$ with the space that carries the representation j of $\mathbf{J}_2$. Therefore $\mathcal{H}_n$, with $n = 2j+1$, must be $j \otimes j$ where the first j is the value of $\mathbf{J}_1$ and the second j the value of $\mathbf{J}_2$. More explicitly, the degenerate subspace has the structure

$$\mathcal{H}_n \;=\; j \otimes j \quad \text{with basis states} \quad |j_1 = j; m_1\rangle \otimes |j_2 = j; m_2\rangle\,, \quad -j \leq\, m_1, m_2 \leq j\,. \tag{7.43}$$

Since m_1 and m_2 each take $2j+1$ values, the dimension of $\mathcal{H}_n$ is $(2j+1)^2 = n^2$. This is indeed, the expected number of states that we have at the energy level. As we are familiar, the tensor product breaks into a sum of representations of the sum of angular momenta. But the sum here is simply the conventional angular momentum $\mathbf{L} = \mathbf{J}_1 + \mathbf{J}_2$. Since we know that

$$j \otimes j = 2j \oplus 2j-1 \oplus \ldots \oplus 0\,, \tag{7.44}$$

the representations on the right-hand side are the ℓ multiplets that arise. Thus the degenerate subspace is a direct sum of the ℓ values

$$(\ell = 2j) \oplus (\ell = 2j-1) \oplus \ldots \oplus 0\,. \tag{7.45}$$

Recalling that $2j+1 = n$, we have obtained

$$\mathcal{H}_n \;=\; (\ell = n-1) \oplus (\ell = n-2) \oplus \ldots \oplus (\ell = 0)\,. \tag{7.46}$$

This is exactly the familiar set of ℓ multiplets at the degenerate subspace labeled by the principal quantum number n. Done!

We should emphasize that the above algebraic analysis characterizes for us the *possible* subspaces $\mathcal{H}_n$ of degenerate energy eigenstates. These subspaces are labeled by the values of j, with $j \geq 0$ and $2j \in \mathbb{Z}$ and $n = 2j+1$. This gives an infinite list $\{0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, \ldots\}$ of possible values of j. The algebraic analysis alone *cannot* tell us which values of j in the above list are used by the hydrogen atom. In physics, however, it is often the case that whatever is possible is in fact compulsory! Thus, it is not all that surprising that all possible values of j actually appear in the hydrogen atom spectrum.

As the simplest non-trivial example of the structure of subspaces of degenerate energies consider $j = \frac{1}{2}$ which gives us $n = 2$. We then have $\mathcal{H}_2 = \frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$, where the right-hand side are the triplet $\ell = 1$ and the singlet $\ell = 0$. The uncoupled basis states are of the form $|\frac{1}{2}, m_1\rangle \otimes |\frac{1}{2}, m_2\rangle$ and thus the four of them can be written briefly as $|\uparrow\uparrow\rangle, |\uparrow\downarrow\rangle, |\downarrow,\uparrow\rangle, |\downarrow\downarrow\rangle$. The singlet and triplet are therefore

$$\ell = 1: \begin{cases} |1,1\rangle \;=\; |\uparrow\uparrow\rangle\,, \\ |1,0\rangle \;=\; \frac{1}{\sqrt{2}}\big(|\uparrow\downarrow\rangle \,+\, |\downarrow\uparrow\rangle\big)\,, \\ |1,-1\rangle \;=\; |\downarrow\downarrow\rangle \end{cases} \quad \ell = 0: \quad |0,0\rangle \;=\; \frac{1}{\sqrt{2}}\big(|\uparrow\downarrow\rangle \,-\, |\downarrow\uparrow\rangle\big)\,. \tag{7.47}$$

These are the four states of the $n = 2$ energy level. Note that we have built them out of spin-one-half states, but there is nowhere in sight such spin-one-half. In the hydrogen atom spectrum we were trying to reproduce the electron and the proton are treated as spinless.

8.05 Quantum Physics II
Fall 2013