Quantum Mechanics

— Incomplete Notes —

DAVID GROSS
Institute for Theoretical Physics
University of Cologne

Thanks to: Lionel J. Dmello

FEBRUARY 1, 2025

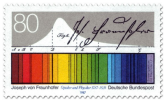
Contents

This symbol indicates that you may skip forward without missing much.

Chapter 0

Introduction

At the beginning of the 20th century, it looked like physics was about to be complete. True, there were a few phenomena that couldn't yet be explained using the laws of mechanics and electrodynamics.



Examples include discrete lines in emission spectra (pictured), quantized angular momentum in a Stern-Gerlach setup, electrons not crashing into the nucleus, and interference patterns behind double slits.

However, there was no reason to believe that more than a little bit of fiddling was required to account for them...

...then a bunch of gifted twenty-somethings took it on them to fix things up, and when they were done, not much was left of the old way of thinking about reality.

0.1 Structure of the theory

In the Hamiltonian formulation of classical physics:

- With every system, one associates a *phase space*. In the simplest case of point particles with n degrees of freedom, it is given by the set of vectors $\boldsymbol{X} = (\boldsymbol{x}, \boldsymbol{p})$, with $\boldsymbol{x} \in \mathbb{R}^n$ describing the positions and $\boldsymbol{p} \in \mathbb{R}^n$ the momenta of the particles.
- The *state* of the system is completely specified if one knows its phase space coordinates $\boldsymbol{X}(t_0)$ at any one time t_0.
- Any *observable property* of the system is a function of $\boldsymbol{X}(t_0)$. This includes the positions and momenta $\boldsymbol{X}(t)$ at any other point in time, which can be found from $\boldsymbol{X}(t_0)$ by solving Hamilton's equations.

Compare this to quantum mechanics:

- With every system, one associates a complex vector space $\mathcal{H}$ with an inner product. (A *Hilbert space*).
- The state of the system is described by a vector $\psi \in \mathcal{H}$ of norm-squared $\langle\psi|\psi\rangle = 1$.
- Any observable property is associated with a linear map $O : \mathcal{H} \to \mathcal{H}$ (called an *operator* in this context). If one measures the property on a system in state ψ, the outcome will follow a probability distribution with expected value equal to $\langle\psi|O\psi\rangle$.

Some Q&A's:

- Q.: *"Complex inner product space"? How in God's name does one come up with that?*
  A.: Yeah, this is not an obvious construction at all! It took a few decades of trial-and-error before the structure became clear.

- Q.: *The quantum description mentions "measurements", the classical one does not.*
  A.: Well spotted! Classical physics tells us how things *are*. QM only tells us how systems respond to measurements.

- Q.: *That seems overly pedantic. Can't we say "a system is in place $\boldsymbol{x}$ if that's where we'd find it if we measure"?*
  As stated, that doesn't work, because QM does *not* tell us where the particle will be when we measure the position. It just makes a prediction about the *average* position found in many repetitions. Though, historically, many scientists felt that there should be a "complete" description of reality that does assign definite positions to particles without referring to acts of measurements or randomness (Einstein: "*Der Alte würfelt nicht*"). Today, it is known that this isn't really possible. In this sense, Nature is fundamentally not classical. (Yes, we'll explain that in more detail).

- Q.: *Enough about philosophy. In practice, how do I know which Hilbert space / state vector / operator to use in order to model a given experiment?*
  A.: There are heuristic *quantization rules* that turn classical into quantum models. ("Heuristic" is a smart-ass term for "rule of thumb"). More systematically, one can often characterize all quantum models that are compatible with a given set of symmetries, e.g. non-relativistic space-time symmetries. We'll have a look at this approach. Finally, physics is an empirical science. So it's perfectly OK to just guess a model and accept it if its predictions agree with experimental outcomes.

- Q.: *Gotcha. Can you show me some details?*
  A.: Sure thing! Read on.

Chapter 1

Wave mechanics

We'll start with a quick and dirty account of the quantum mechanics of a point particle. We will state a number of heuristic "quantization rules". These will be explained and justified in more mathematical detail in later chapters.

Rule 1 The Hilbert space associated with a point particle moving in n dimensions consists of functions $\psi : \mathbb{R}^n \to \mathbb{C}$, with the form

$$\langle \phi | \psi \rangle := \int \phi(\boldsymbol{x})^* \psi(\boldsymbol{x}) \, \mathrm{d}^n \boldsymbol{x}. \tag{1.1}$$

Mathematically, these "wave functions" behave similarly to the classical fields of electrodynamics or continuum mechanics.

The good thing is that you are already familiar with such fields (somewhat, at least). So things will feel very concrete and we'll be able to make quick progress.

The bad thing is that you are already familiar with such fields. So there's the danger that one takes the analogy too seriously and starts thinking of quantum mechanical wave functions as being as real as ripples on the surface of a lake. But that's taking things too far. As we'll discuss in Chapter 2, these functions are really just mathematical tools for computing probabilities. (In particular, I feel that semi-classical notions such as "wave-particle duality" or "the collapse of the wave function", often invoked for pedagogical purposes, result from people taking wave functions too literally, and end up causing more confusion than clarity.)

1.1 Properties of particles in one dimension

We'll now explain what it means when we say that "a particle is in the state $\psi(\boldsymbol{x})$". For simplicity, we restrict to $n = 1$ for the time being.

1.1.1 Position measurements

Rule 2 A position measurement on a particle in state ψ will result in the outcome x with probability density $|\psi(x)|^2$.

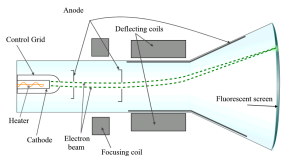
Remark.

- Recall that one goes from probability densities to probabilities by integrating. E.g., the probability of finding the particle in an interval $[a, b]$ is

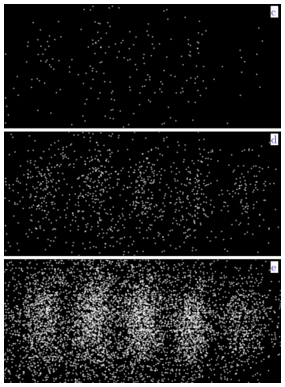
$$\Pr\left[x \in [a,b]\right] = \int_a^b |\psi(x)|^2 \, \mathrm{d}x.$$

- The same situation is sometimes expressed in these words:
  - "The probability of finding the particle in the infinitesimal interval $[x, x + \mathrm{d}x]$ is $|\psi(x)|^2 \, \mathrm{d}x$".
  - "The probability of finding the particle at x is $|\psi(x)|^2$". (It's a little sloppy, but very common, to not strictly distinguish between probabilities and densities).

How are positions of microscopic particles measured, practically speaking?



Until the early 2000's most households had an electron position measurement device in their living rooms: A television set built around a cathode ray tube (top). An electron gun prepares a constant stream of electrons in identical states. They traverse the tube, eventually hitting the fluorescent screen. There, a light flash is created at the impact point, effectively measuring the position of the electron.



A slightly beefed-up version of the setup can be used to reveal interference effects of electron wave functions. The series of pictures on the left is taken from this paper. A low-intensity electron beam was passed through a double slit. Light flashes emanating from a fluorescent screen were recorded with a CCD camera. The series shows the accumulated signals over time. Each point thus represents the outcome of a position measurement on a single electron. As the number of events increases, the density of points becomes proportional to $|\psi(x, y, 0)|^2$ (in coordinates where the electrons move along the z-direction and the screen is located in the $z = 0$-plane).

Picture credit.

The total probability of finding the particle *somewhere* is one, so that any state vector must be *normalized* in the sense that

$$\int_{-\infty}^{\infty} |\psi(x)|^2 \, \mathrm{d}x = 1. \tag{1.2}$$

Using (1.1), the normalization condition can be expressed in terms of the squared norm

$$\|\psi\|^2 = \langle \psi | \psi \rangle = 1.$$

Examples.

- A wave function that gives rise to a uniform distribution on the interval $[a, b]$:

$$\psi(x) = \frac{1}{\sqrt{b-a}} \begin{cases} 1 & x \in [a,b] \\ 0 & \text{else} \end{cases}$$

- *Gaussian wave functions* turn out to be very important, in particular

$$\psi(x) = \pi^{-1/4} e^{-\frac{x^2}{2}}. \tag{1.3}$$

Its probability density

$$|\psi(x)|^2 = \frac{e^{-x^2}}{\sqrt{\pi}}$$

is the normal distribution with mean $\mu = 0$ and standard deviation $\sigma = \frac{1}{\sqrt{2}}$. We'll also soon make contact with the following family of *Gaussian wave packets*, parameterized by $x_0, k_0 \in \mathbb{R}, \sigma \in \mathbb{R}_{>0}$:

$$\psi_{\sigma,x_0,k_0}(x) = \left(\frac{1}{2\pi\sigma^2}\right)^{1/4} e^{-\frac{(x-x_0)^2}{4\sigma^2}} e^{ik_0 x}. \tag{1.4}$$

Because $|e^{ik_0 x}|^2 = 1$, the parameter k_0 does not affect the distribution of position measurements. But the funny complex numbers will turn out to be meaningful in other ways.

Now assume that you perform independent position measurements on N particles, all in the same state ψ. If x_i is the i-th outcome, then, by the law of large numbers, the *mean position* will converge to the *expected value*

$$\frac{1}{N}\sum_{i=1}^{N} x_i \to \int_{-\infty}^{\infty} x|\psi(x)|^2 \,\mathrm{d}x \qquad (N \to \infty). \tag{1.5}$$

In QM, it's common to express the right hand side in different language. Define the *position operator* X to be the linear map that acts on a wave function ψ by multiplying it with its argument:

$$(X\psi)(x) = x\psi(x).$$

Then the expected value can be written as an inner product:

$$\int_{-\infty}^{\infty} x|\psi(x)|^2 \,\mathrm{d}x = \int_{-\infty}^{\infty} \psi(x)^* x\psi(x) \,\mathrm{d}x = \int_{-\infty}^{\infty} \psi(x)^* (X\psi)(x) \,\mathrm{d}x = \langle\psi|X\psi\rangle.$$

Other common notations:

$$\langle\psi|X\psi\rangle = \langle\psi|X|\psi\rangle = \langle X\rangle_\psi = \langle X\rangle.$$

Example. For the Gaussian in (1.3), $\langle\psi|X|\psi\rangle = 0$, because $\psi(x)$ is an even function, but x is an odd function. Similarly, $\langle\psi_{\sigma,x_0,k_0}|X - x_0|\psi_{\sigma,x_0,k_0}\rangle = 0$.

Often, we will compute expected values of *functions* of position.

Examples.

- For

$$f(x) = \begin{cases} 1 & x \in [a,b] \\ 0 & \text{else} \end{cases},$$

the expectation value

$$\int_{-\infty}^{\infty} f(x)|\psi(x)|^2 \,\mathrm{d}x = \int_a^b |\psi(x)|^2 \,\mathrm{d}x = \Pr\left[x \in [a,b]\right]$$

is the just probability of finding the particle in the interval $[a,b]$.

- The expectation value of $f(x) = (x - x_0)^2$ measures the expected squared distance of the position to the value x_0. In particular, the *variance* is the expected value of $f(x) = (x - \langle X\rangle)^2$. It is the most commonly used measure of the *dispersion* (or *spread*) of a distribution.

The expectation value $f(x)$ can also be expressed by an operator, namely $f(X)$. To define that expression, assume that f can be written as a series $f(x) = \sum_k c_k x^k$ and set $f(X) := \sum_k c_k X^k$. From

$$(X^k\psi)(x) = (\underbrace{X\cdots X}_{k\times}\psi)(x) = (\underbrace{X\cdots X}_{(k-1)\times}x\psi)(x) = \cdots = x^k\psi(x)$$

one then easily obtains

$$\big(f(X)\psi\big)(x) = f(x)\psi(x).$$

Therefore, as in (1.5), under repeated measurements,

$$\frac{1}{N}\sum_{i=1}^{N} f(x_i) \to \langle\psi|f(X)|\psi\rangle \qquad (N\to\infty).$$

Example. For the variance, we get the useful formula

$$\mathrm{Var}[X] := \langle (X - \langle X\rangle)^2\rangle = \langle X^2 - 2\langle X\rangle X + \langle X\rangle^2\rangle = \langle X^2\rangle - \langle X\rangle^2.$$

For the Gaussian wave function (1.3), this gives

$$\begin{aligned}
\mathrm{Var}[X] &= \langle\psi|X^2|\psi\rangle - \langle\psi|X|\psi\rangle^2 \\
&= \langle\psi|X^2|\psi\rangle && (\psi(x) \text{ is even and } x \text{ is odd}) \\
&= \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} x^2 e^{-x^2}\,\mathrm{d}x \\
&= \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} x\big(-\frac{1}{2}\partial_x e^{-x^2}\big)\,\mathrm{d}x \\
&= \frac{1}{2}\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-x^2}\,\mathrm{d}x && \text{(partial integration)} \\
&= \frac{1}{2}.
\end{aligned}$$

Similarly, for the wave packet (1.4), the variance of the position distribution is $\mathrm{Var}[X] = \sigma^2$.

1.2 Momentum

Recall the definition of the Fourier transform and its inverse

$$\begin{aligned}
\tilde\psi(\boldsymbol{k}) &= (2\pi)^{-n/2}\int e^{-i\boldsymbol{k}\boldsymbol{x}}\psi(\boldsymbol{x})\,\mathrm{d}^n\boldsymbol{x}, \\
\psi(\boldsymbol{x}) &= (2\pi)^{-n/2}\int e^{i\boldsymbol{k}\boldsymbol{x}}\tilde\psi(\boldsymbol{k})\,\mathrm{d}^n\boldsymbol{k}.
\end{aligned} \tag{1.6}$$

Rule 3 A momentum measurement on a particle in state $\psi(x)$ will result in the outcome p with probability density $|\tilde\psi(k)|^2$, where $k = p/\hbar$.

- Q.: *Why?? What has the Fourier transform of all things to do with momentum?*
  A.: We'll justify this more systematically in Chapter 4. For now, take it as an axiom. It's not completely surprising, because before QM was fully developed, de Broglie conjectured, and Thomson-Reid and Davisson-Germer confirmed experimentally, that electrons with momentum p show interference behavior that is consistent with plane waves with wave vector $k = p/\hbar$. The above box says that the Fourier transform is the right way of making this correspondence precise.

Example. Exercise: The momentum distribution of the wave packet (1.4) is

$$|\tilde{\psi}_{\sigma,x_0,k_0}(k)|^2 = \left(\frac{2\sigma^2}{\hbar^2\pi}\right)^{1/2} e^{-\frac{\sigma^2(k-k_0)^2}{\hbar^2}}.$$

Finding it involves a somewhat lengthy calculation. Next, we'll discuss strategies for avoiding these.

As with position, the expected value of a momentum measurement can be expressed as an inner product involving the wave function and an operator:

$$\begin{aligned}
\int \mathrm{d}k\, \hbar k |\tilde{\psi}(k)|^2 &= \int \mathrm{d}k\, \tilde{\psi}(k)^* \hbar k \tilde{\psi}(k) \qquad (1.7)\\
&= \frac{1}{\sqrt{2\pi}} \int \mathrm{d}k \left(\int \mathrm{d}x\, e^{-ikx}\psi(x)\right)^* \hbar k \tilde{\psi}(k)\\
&= \frac{1}{\sqrt{2\pi}} \int \mathrm{d}k \int \mathrm{d}x\, \psi(x)^* \hbar k e^{ikx} \tilde{\psi}(k)\\
&= \frac{1}{\sqrt{2\pi}} \int \mathrm{d}k \int \mathrm{d}x\, \psi(x)^* (-i\hbar\partial_x) e^{ikx} \tilde{\psi}(k)\\
&= \int \mathrm{d}x\, \psi(x)^* (-i\hbar\partial_x)\psi(x)\\
&= \langle\psi|(-i\hbar\partial_x)|\psi\rangle. \qquad (1.8)
\end{aligned}$$

The *momentum operator* is therefore defined as

$$P := -i\hbar\partial_x. \qquad (1.9)$$

From (1.7), it is manifest that momentum expectation values are real numbers. But that's hard to see from the operator representation (1.8).

Reminder If A is an operator, its *adjoint* is the unique operator $A^\dagger$ such that

$$\langle A^\dagger\phi|\psi\rangle = \langle\phi|A\psi\rangle \qquad \forall\, \phi, \psi.$$

A linear map A is *self-adjoint* or *Hermitian* if $A^\dagger = A$.

If A is Hermitian, then for all ψ,

$$\langle\psi|A\psi\rangle^* = \langle A\psi|\psi\rangle = \langle\psi|A\psi\rangle \qquad \Rightarrow \qquad \langle\psi|A\psi\rangle \in \mathbb{R}.$$

(The converse is also true). Thus, real-valued quantities are represented by Hermitian operators in QM. Let's verify directly that the momentum operator is Hermitian: **Exercise**.

Example. Equation (1.9) allows us to compute expectation values of (functions of) momentum without having to perform Fourier transforms. For example, for $\psi = \psi_{\sigma,x_0,k_0}$ the Gaussian in Eq. (1.4):

$$(P\psi)(x) = -i\hbar \left(\frac{1}{2\pi\sigma^2}\right)^{1/4} \partial_x \left(e^{-\frac{(x-x_0)^2}{4\sigma^2}} e^{ikx}\right) = (x-x_0)\frac{i\hbar}{2\sigma^2}\psi(x) + \hbar k\psi(x)$$

$$\Rightarrow \quad \langle\psi|P|\psi\rangle = \hbar k\langle\psi|\psi\rangle + \frac{i\hbar}{2\sigma^2}\langle\psi|X - x_0|\psi\rangle = \hbar k.$$

The e^{ikx}-factor that did not impact the position distribution is now seen to affect momentum.

Let's compute the variance of the momentum distribution, for simplicity for $\psi := \psi_{\sigma,0,0}$:

$$(P^2\psi)(x) = \frac{\hbar^2}{2\sigma^2}\left(\frac{1}{2\pi\sigma^2}\right)^{1/4} \partial_x \left(xe^{-\frac{x^2}{4\sigma^2}}\right) = \frac{\hbar^2}{2\sigma^2}\left(1 - \frac{x^2}{2\sigma^2}\right)\psi$$

$$\Rightarrow \qquad \mathrm{Var}[P] = \langle\psi|P^2|\psi\rangle = \frac{1}{2}\frac{\hbar^2}{\sigma^2}\left(1 - \frac{\sigma^2}{2\sigma^2}\right) = \frac{1}{4}\frac{\hbar^2}{\sigma^2}.$$

Recall that $\mathrm{Var}[X] = \sigma^2$, so that decreasing σ makes position distributions more concentrated, while making the momentum distribution less concentrated. In fact, their product is constant:

$$\mathrm{Var}[X^2]\,\mathrm{Var}[P^2] = \frac{\hbar^2}{4}. \tag{1.10}$$

For classical point particles, position and momentum are independent parameters. In QM, their distributions are derived from the same object, the wave function $\psi(x)$. This gives rise to relations between the two quantities. The most famous one is the Heisenberg uncertainty relation

$$\mathrm{Var}[X]\,\mathrm{Var}[P] \geq \frac{\hbar^2}{4},$$

which says that no wave function has both a sharp position and a sharp momentum distribution. Proving it is homework. (Eq. (1.10) shows Gaussian wave packets attain Heisenberg's lower bound).

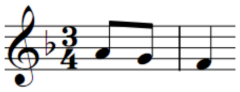
The proof of the uncertainty relation will make use of the fact that X and P do not commute. Indeed, they satisfy (homework!) what is pompously called the *canonical commutation relation*

$$[X, P] = XP - PX = i\hbar\mathbb{1}. \tag{1.11}$$

For now, that's just an algebraic fact. We'll explain where it comes from in Chap. 4.

Remark. Let $\psi(t)$ be the pressure exerted by a sound wave on your eardrums at time t. Creating the pressure takes an amount of mechanical energy proportional to $|\psi(t)|^2$. Likewise, the energy associated with frequency $\frac{\omega}{2\pi}$ is proportional to the square of the Fourier coefficient $|\tilde{\psi}(\omega)|^2$. Can you see where this is going? The distributions of acoustic energy in, respectively, time and frequency space is computed in exactly the same way as position and momentum in quantum mechanics.

Introduce the acoustic time and frequency operators as $(T\psi)(t) = t\psi(t)$ and $(F\psi)(t) = \frac{-i}{2\pi}\partial_t\psi(t)$. Then the same calculation that leads to Heisenberg's uncertainty relation says that the spread in time and frequency of an audio signal must satisfy $\Delta_T\Delta_F \geq \frac{1}{4\pi}$. That's the reason a flute can play much shorter notes than a trombone, while still sounding in tune.



Musical notes encode temporal and frequency information. From the above, this might be problematic! Exercise: Argue why there is no issue for the example shown on the left.

1.3 Quantization of functions of position and momentum

In Hamiltonian mechanics of point particles, any observable quantity is a function $f(\boldsymbol{x}, \boldsymbol{p})$ of position and momentum.

For functions $f(x)$ of position alone, and functions $f(p)$ of momentum alone, we have already constructed operators that represent them in QM: $f(X)$ and $f(P)$ respectively. In particular, this gives operators

$$\frac{P^2}{2m} \quad \text{and} \quad V(X)$$

for the kinetic and potential energy. It's thus natural to conjecture:

Rule 4 A classical phase space function $f(x, p)$ turns into the operator $f(X, P)$.

This turns out to mostly work. There are two related issues. First, multiplication of the classical functions x, p commutes, but the same is not true for the associated operators. This leads to ambiguities. For example, $xp, px, \frac{1}{2}(xp + px)$ all denote the same function, $XP, PX, \frac{1}{2}(XP + PX)$ are completely different. Second, even if $f(x, p)$ is real-valued, $f(X, P)$ need not be Hermitian, and thus need not have real expectation values. Fortunately, neither problem will occur in these notes, so we'll ignore them.

The quantization rule allows us to find the *Hamiltonian operator*, representing energy:

$$H = \frac{P^2}{2m} + V(X), \qquad (H\psi)(x) = \frac{-\hbar^2}{2m}\partial_x^2\psi(x) + V(x)\psi(x). \tag{1.12}$$

For position and momentum, we started describing the possible values encountered in a measurement (any real number), constructed the probability densities ($|\psi(x)|^2$, $|\tilde{\psi}(k)|^2$ respectively), and finally found operators (X, P) that describe expected values. The above quantization rule just gives an operator, but does not by itself describe the possible outcomes and probabilities. There is a standard construction that associates a full measurement process with an operator. It is sometimes called the *projective measurement* or the *von Neumann measurement* of the operator.

Reminder. This construction uses ideas from linear algebra. Recall: Given an operator A, the number λ is an *eigenvalue* if there exists a non-zero *eigenvector* (here also: *eigenfunction*, *eigenstate*) ϕ_λ such that

$$A\phi_\lambda = \lambda\phi_\lambda. \tag{1.13}$$

For a fixed λ, the set of solutions to (1.13) is the *eigenspace* associated with λ. An eigenvalue is *degenerate* if the eigenspace is has dimension larger than 1. If $A = A^\dagger$ is Hermitian, then all eigenvalues are real, eigenvectors to different eigenvalues are orthogonal to each other, and there exists an ortho-normal eigenbasis.

Rule 5 If A is a Hermitian operator describing an observable quantity, then the possible values that can occur in a projective measurement are its eigenvalues.

Next, the basic idea is to derive probabilities from the squared inner product with normalized eigenvectors: $|\langle\psi|\phi_\lambda\rangle|^2$. Unfortunately, the details are quite subtle.

An eigenvalue λ of A is *discrete* if there is a finite distance ϵ such that all other eigenvalues are at least ϵ away. Else, it is *continuous*. We need to associate *probabilities* with discrete eigenvalues, but *probability densities* with continuous ones.

Discrete eigenvalues

If λ is discrete and non-degenerate, one can show that there exist normalized eigenvector

$$A\phi_\lambda = \lambda\phi_\lambda, \qquad \langle\phi_\lambda|\phi_\lambda\rangle = 1. \tag{1.14}$$

It is unique up to a *global phase*: $\phi_\lambda \mapsto e^{i\theta}\phi_\lambda$.

Rule 6d If λ is discrete and non-degenerate, the probability of observing λ when performing the projective measurement of A on a system in state ψ is

$$|\langle\psi|\phi_\lambda\rangle|^2, \tag{1.15}$$

where ϕ_λ solves (1.14).

If λ is degenerate, then its eigenspace has a basis $\{\phi_\lambda^{(i)}\}_i$ that is ortho-normal $\langle\phi_\lambda^{(i)}|\phi_\lambda^{(j)}\rangle = \delta_{ij}$. In this case, Eq. (1.15) generalizes to

$$\sum_i |\langle\psi|\phi_\lambda^{(i)}\rangle|^2.$$

Continuous eigenvalues

If λ is continuous, it is an element of an interval $(a, b) \subset \mathbb{R}$ of eigenvalues. The deep *spectral theorem* of functional analysis then implies that there is a family $\phi_{\lambda'}$ of eigenvectors

$$A\phi_{\lambda'} = \lambda'\phi_{\lambda'}, \qquad \text{normalized in the sense that} \qquad \langle\phi_\lambda|\phi_{\lambda'}\rangle = \delta(\lambda - \lambda'). \tag{1.16}$$

for $\lambda' \in (a, b)$. Here δ is the Dirac delta function.

> ***Remark.*** The "eigenfunctions" ϕ_λ associated with a continuous eigenvalue do not have finite norm, and are often not even realizable as ordinary functions. A careful treatment shows that these are to be understood as *distributions*. See the Appendix for more information. Here, we'll gloss over these details, and just give a few examples.

Rule 6c If λ is continuous and non-degenerate, the probability *density* of observing λ when performing the projective measurement of A on a system in state ψ is

$$|\langle\psi|\phi_\lambda\rangle|^2, \tag{1.17}$$

where ϕ_λ solves (1.16).

If λ is degenerate, we'll need one more index: Then, there exist $\{\phi^{(i)}_{\lambda'}\}$ such that

$$\langle \phi^{(i)}_{\lambda} | \phi^{(j)}_{\lambda'} \rangle = \delta_{ij}\, \delta(\lambda - \lambda')$$

and Eq. (1.17) generalizes to

$$\sum_i |\langle \psi | \phi^{(i)}_{\lambda} \rangle|^2.$$

Example. Let's look at the momentum operator through this lens. The eigenvalue equation $P\phi_\lambda = \lambda\phi_\lambda$ reads

$$-i\hbar\partial_x \phi_\lambda(x) = \lambda\phi_\lambda(x), \qquad \forall x,$$

solved by $\phi_\lambda(x) = ce^{i\frac{\lambda}{\hbar}x}$ for some constant c.

Assume that $\lambda \in \mathbb{R}$. Then ϕ_λ is a plane wave. It follows from Parseval's Theorem that

$$\langle \phi_\lambda | \phi_{\lambda'} \rangle = \frac{|c|^2}{2\pi}\delta(\lambda - \lambda'),$$

so that $c = (2\pi)^{-1/2}$ gives the correct normalization for continuous eigenvectors. The probability density for the outcome λ becomes $|\langle \phi_\lambda | \psi \rangle|^2 = |\tilde{\psi}(\lambda)|^2$. We have recovered the definition of momentum measurement we started with, so all is consistent.

Now consider the case where $\lambda \notin \mathbb{R}$. Then $\phi(x)$ diverges exponential as x tends to either $+\infty$ or $-\infty$, depending on the sign of $\operatorname{Im} z$. But then, even for ψ with finite norm, the integral $|\langle \phi_\lambda | \psi \rangle|$ will usually diverge. We thus reject such solutions.

1.4 Energy

We have seen in (1.12) that the energy operator for point particle is

$$H = \frac{P^2}{2m} + V(X), \qquad (H\psi)(x) = \frac{-\hbar^2}{2m}\partial_x^2\psi(x) + V(x)\psi(x).$$

For reasons explained soon, the eigenvalue equation for H is called the *time-independent Schrödinger equation*. We'll work out its solutions for various potentials V.

1.4.1 Free particle

A particle is *free* if $V = 0$. The eigenvalue equation then reads

$$\frac{-\hbar^2}{2m}\partial_x^2\phi_E(x) = E\phi_E(x), \qquad \text{solved by} \qquad \phi_E(x) = Ae^{i\frac{\sqrt{2mE}}{\hbar}} + Be^{-i\frac{\sqrt{2mE}}{\hbar}}$$

for constants $A, B \in \mathbb{C}$. Energy eigenspaces are thus two-fold degenerate.

A natural basis for the energy eigenspace is given by

$$\phi^+_E(x) = \frac{1}{\sqrt{2\pi}}e^{i\frac{\sqrt{2mE}}{\hbar}x}, \qquad \phi^-_E(x) = \frac{1}{\sqrt{2\pi}}e^{-i\frac{\sqrt{2mE}}{\hbar}x}. \tag{1.18}$$

These are just the momentum eigenfunctions for $p = \pm\sqrt{2mE}$. The relation between energy and momentum is the same as in the classical case (we'll see more exciting results in a minute). Thus, physically, the eigenspace is degenerate because knowing the energy tells you only about the magnitude, but not the direction of momentum.

We are free to choose different bases in the energy eigenspaces. In particular,

$$\begin{aligned}\phi_E^{(e)}(x) &= \frac{1}{\sqrt{2}}(\phi_E^+(x) + \phi_E^-(x)) = \frac{1}{\pi}\cos\Big(\frac{\sqrt{2mE}}{\hbar}x\Big),\\ \phi_E^{(o)}(x) &= \frac{1}{i\sqrt{2}}(\phi_E^+(x) - \phi_E^-(x)) = \frac{1}{\pi}\sin\Big(\frac{\sqrt{2mE}}{\hbar}x\Big)\end{aligned} \tag{1.19}$$

is often a useful choice ("e/o" for even/odd).

1.4.2 The step potential

Consider the "step potential" (Fig. **??**)

$$V(x) = \begin{cases} 0 & x \le 0 \quad (\text{"region I"}) \\ V_{\text{II}} & x > 0 \quad (\text{"region II"}) \end{cases}, \qquad V_{\text{II}} > 0.$$

Of particular interest is the energy range $0 < E < V_{\text{II}}$, which classically is associated with particles that move in from $-\infty$, bounce off the potential step, and return to $-\infty$.

To solve the time-independent Schrödinger equation for this and similar potentials, one first finds the solutions in each region where the potential is constant, and then patches these solutions together. Let's work this out for $0 < E < V_{\text{II}}$.

Piece-wise solution: In region I, the particle is free, so the general solution is

$$\phi_I(x) = Ae^{ikx} + Be^{-ikx}, \qquad k = \sqrt{2mE}/\hbar.$$

In region II, the eigenvalue equation mimics the free one, with E replaced by $V_{\text{II}} - E$:

$$\frac{-\hbar^2}{2m}\partial_x^2\phi_{\text{II}}(x) + V_{\text{II}}\phi_{\text{II}}(x) = E\phi_{\text{II}}(x) \quad \Leftrightarrow \quad \partial_x^2\phi_{\text{II}}(x) - \frac{2m(V_{\text{II}} - E)}{\hbar^2}\phi_{\text{II}}(x) = 0.$$

It is customary to write the general solution as

$$\phi_{\text{II}}(x) = Ce^{\kappa x} + De^{-\kappa x}, \qquad \kappa = \sqrt{2m(V_{\text{II}} - E)}/\hbar. \tag{1.20}$$

The coefficient C has to vanish, for else $\phi_{\text{II}}(x)$ would diverge as $x \to \infty$.

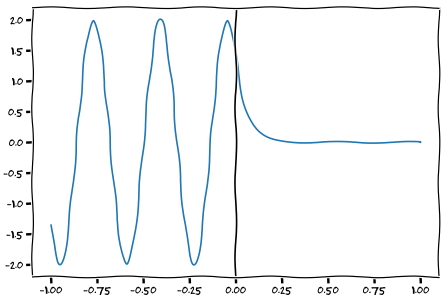
Join at interface: When gluing solutions together we impose the *interface conditions* demanding that $\phi(x)$ and $\phi'(x)$ be continuous (c.f. Eq. (D.11)). In this case:

$$\phi_I(0) = \phi_{\text{II}}(0) \quad \Leftrightarrow \quad A + B = D, \tag{1.21}$$

$$\phi_I'(0) = \phi_{\text{II}}'(0) \quad \Leftrightarrow \quad ikA - ikB = -\kappa D. \tag{1.22}$$

That's a system of linear equations which, after some boring but tedious massaging gives

$$B = -e^{i2\delta}A, \quad D = -2i\sqrt{\tfrac{E}{V_{\text{II}}}}e^{i\delta}A, \qquad \delta = \tan^{-1}\sqrt{\tfrac{E}{V_{\text{II}}-E}}. \tag{1.23}$$



Energy eigenfunction of the potential step Hamiltonian with parameters

$$m = \hbar = 1, V = 300, E = \frac{V}{2}, A = 1.$$

(Compared to (1.23), the wave function has been multiplied by a global phase factor of $ie^{-i\delta}$, which makes its real).

Discussion. We'll interpret these solutions in more detail soon. For now, just a quick analogy. Shine a light wave with incident amplitude A on a perfect mirror. Let B be the amplitude of the reflected wave. Because the energy in the field is proportional to the amplitude squared, energy conservation means that $|A|^2 = |B|^2$. But also, the wave will not be reflected directly at the surface of the mirror: The field will enter the material, with a strength that decreases exponentially from the surface.

Replacing "energy" by "probability", that's exactly what we are seeing here. In particular, Eq. (1.20) says that there is a finite probability of finding the particle in the "classically forbidden" region with $E < V$. We'll come back to this later.[1]

In the limit $V_{\rm II} \to \infty$, Eqs. (1.20, 1.23) say that $\phi_{\rm II}(x) \to 0$: The probability of finding a particle in a region of infinite potential is 0.

1.4.3 The infinite potential well

Next, consider a "infinite potential well" or "particle in a box":

$$V(x) = \begin{cases} \infty & x \in (-\infty, -a/2) \quad (\text{"region I"}) \\ 0 & x \in [-a/2, a/2] \quad (\text{"region II"}) \\ \infty & x \in (a/2, \infty) \quad (\text{"region III"}) \end{cases} .$$

Then $\phi_{\rm I}(x) = \phi_{\rm III}(x) = 0$. Using the basis (1.19) in region II:

$$\phi_{\rm II}(x) = A\sin(kx) + B\cos(kx), \quad \phi_{\rm II}(\pm a/2) = 0, \qquad k = \frac{\sqrt{2mE}}{\hbar}.$$

The boundary conditions $\phi(\pm a/2) = 0$ force $A = 0$ or $B = 0$. Because

$$\begin{aligned} \sin\big(ka/2\big) = 0 \quad &\Leftrightarrow \quad \frac{ka}{2} = \text{a multiple of } \pi \quad &\Leftrightarrow \quad ka = \text{an even multiple of } \pi, \\ \cos\big(ka/2\big) = 0 \quad &\Leftrightarrow \quad \frac{ka}{2} = \frac{\pi}{2} + \text{a multiple of } \pi \quad &\Leftrightarrow \quad ka = \text{an odd multiple of } \pi, \end{aligned}$$

the solutions are labeled by positive natural numbers $n = 1, 2, 3, \ldots$. Specifically

$$\phi_n(x) = \sqrt{\frac{2}{a}} \begin{cases} \sin(k_n x) & n \text{ even} \\ \cos(k_n x) & n \text{ odd} \end{cases}, \qquad k_n = n\frac{\pi}{a}, \qquad E_n = \frac{\hbar^2 k_n^2}{2m} = \frac{\hbar^2\pi^2}{2ma^2}n^2.$$

Discussion: The boundary conditions make it so that the possible energies are a discrete set! This effect is what gives quantum theory its name. Also, the lowest possible energy, $E_1 \propto \frac{1}{a^2}$ increases with decreasing width of the box. This is the uncertainty principle at work: Because $\mathrm{Var}[X] \simeq a^2$, it must be that $\mathrm{Var}[P] \propto \frac{1}{a^2}$. But $\mathrm{Var}[P] = \langle P^2\rangle \propto \langle H\rangle$.

1.5 Dynamics

Recall that in mechanics, the Hamilton function plays two completely different roles.

[1] There's also the somewhat-harder to interpret phase factor $e^{i2\delta}$. Suffice it to say that in two dimensions, its optical analogue gives rise to the Goos-Hänchen effect. I mention it mainly because to a slightly dyslexic speaker of German and English, this has got to be the most hilariously-named result in all of science. The fact that, incongruently, it was demonstrated by a Mr. Goose and a Miss Rooster only adds to its allure. (Also, check out the biography of Hilda Hänchen). (And, no, it didn't escape my attention that the name literally translates to "cockfight". This thing's a goldmine!).

(1) It is an observable quantity, measurable e.g. by a power meter, a thermometer, or, if you know relativity, a weight scale. It is a conserved quantity, and a significant part of the world's economy is concerned with moving our limited supply from one place to another.

(2) It also gives rise to the time evolution, via Hamilton's equations

$$\partial_t \boldsymbol{X} = \boldsymbol{J}\big(\boldsymbol{\nabla} H(\boldsymbol{X})\big), \qquad \boldsymbol{X} = \begin{pmatrix} x \\ p \end{pmatrix}, \qquad \boldsymbol{J} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

(This dual role is an incarnation of Noether's Theorem). It may then not be too surprising that the quantum mechanical rule for time evolution also involves the energy operator.

Rule 7 Let ψ_t be the state of an isolated quantum system at time t. Then

$$i\hbar \partial_t \psi_t = H\psi_t. \tag{1.24}$$

Equation (1.24) is *Schrödinger's equation*.

Depending on context, one should think of ψ_t as a family of functions $\psi_t : \mathbb{R} \to \mathbb{C}$ indexed by the time t, or as a single function $\psi(t, x) = \psi_t(x)$ of two arguments.

Even for $n = 1$, the Schrödinger equation

$$i\hbar \partial_t \psi(t, x) = -\frac{\hbar^2}{2m} \partial_x^2 \psi(t, x) + V(x)\psi(t, x)$$

is a partial differential equation that does not usually allow for an explicit solution. We'll now look a various ways of anyway extracting information from it.

1.5.1 Time evolution operator

Define the *time evolution operator*

$$U(t) = e^{\frac{t}{i\hbar} H} = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{t}{i\hbar} H \right)^k. \tag{1.25}$$

Recalling basic linear algebra,

$$\partial_t U(t) = \left(\frac{1}{i\hbar} H \right) U(t),$$

and one immediately verifies that that given the some initial condition ψ_0, the family

$$\psi_t = U(t)\psi_0$$

solves the Schrödinger equation.

We'll see that this formulation is helpful for analyzing the Schrödinger equation, but it does not usually suggest a way for actually evaluating $\psi_t(x)$.

1.5.2 Reduction to the time-independent Schrödinger equation

Assume that ϕ is an energy eigenfunction, $H\phi = E\phi$. Then $i\hbar \partial_t \phi_t = H\phi_t$ with initial condition $\phi_0 = \phi$ is solved by

$$\phi_t = e^{-i\omega t}\phi, \qquad \omega = \frac{E}{\hbar}.$$

Now assume that H has a discrete eigenbasis $\{\phi_\lambda\}_\lambda$ with eigenvalues E_λ. If

$$\psi_0 = \sum_\lambda c_\lambda \phi_\lambda$$

is the expansion of ψ_0 in terms of that basis, then a solution of (1.24) is given by

$$\psi_t(x) = \sum_\lambda c_\lambda e^{-i\omega_\lambda t}\phi_\lambda, \qquad \omega_\lambda = \frac{E_\lambda}{\hbar}. \tag{1.26}$$

Continuous eigenvalues are treated similarly, with the sum replaced by an integral.

Thus, the Schrödinger equation is easy to solve if we know an eigenbasis of H. It is for this reason that the eigenvalue equation for the Hamiltonian is called "time-independent Schrödinger equation".

Free time evolution

Let's apply this to the free Hamiltonian $H = P^2/(2m)$. Its (continuous) eigenbasis is given by the plane waves $\phi_k = (2\pi)^{-1/2}e^{ikx}$, where $k \in \mathbb{R}$ and $H\phi_k = E_k\phi_k$ with $E_k = \frac{\hbar^2 k^2}{2m}$. The expansion coefficients of ψ_0 in that basis are just the Fourier transform

$$\tilde{\psi_0}(k) = (2\pi)^{-1/2}\int \psi_0(x)e^{-ikx}\,\mathrm{d}x.$$

Hence the above discussion shows that the wave function at time t,

$$\psi_t(x) = \int \tilde{\psi_0}(k)e^{-i\omega_k t}\phi_k(x)\,\mathrm{d}k = (2\pi)^{-1/2}\int \tilde{\psi_0}(k)e^{-i\omega_k t + ikx}\,\mathrm{d}k$$

is the inverse Fourier transform of $\tilde{\psi_0}(k)e^{-i\omega_k t}$.

Remark. The calculation for a Gaussian wave packet is homework. You'll find that the expectation values of position and momentum satisfy the classical relations

$$\langle X\rangle(t) = \langle X\rangle(0) + \frac{\langle P\rangle(0)}{m}t, \qquad \langle P\rangle(t) = \langle P\rangle(0). \tag{1.27}$$

However, asymptotically, $\mathrm{Var}[X] \propto t$. The wave packet becomes "smeared out". ($\mathrm{Var}[P]$ remains constant).

Now here's a puzzle. Put a ping-pong ball on a table. Then no forces act on it in the direction parallel to the table top – these degrees of freedom should thus be "free". Still, if you leave the room and check its position a day later, you'll find it right where you left it (assuming you don't live with my kids) – no broadening of the spatial distribution is observed.

Sometimes the following reason is offered: "The Heisenberg limit $\mathrm{Var}[X] \geq \frac{\hbar^2}{4\,\mathrm{Var}[P]}$ is tiny on account of $\hbar$ being so small, thus macroscopic objects have quasi-definite position and momentum". That's a bogus argument, though, because the Heisenberg limit is only a *lower bound* to the uncertainty. As the wave packet shows, $\mathrm{Var}[X]\,\mathrm{Var}[P]$ may well tend to infinity.

The actual resolution is that Schrödinger's equation only applies to *isolated* quantum system. But macroscopic bodies constantly interact with an enormous number of degrees of freedom, notably electromagnetic radiation.

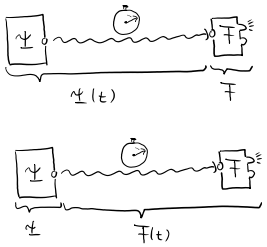
The theory of *decoherence* explains quantitatively how these interactions cause the ball to remain in place. The details are beyond the scope of this lecture, though we might have a superficial look at the theory toward the end of the term.

1.5.3 Heisenberg picture

Consider a quantum system which at time 0 is in the state ψ, and on which an observable F is measured at time t. The expected value is

$$\begin{aligned}\langle F\rangle(t) &= \langle \psi_t | F\psi_t\rangle && (1.28)\\ &= \langle U(t)\psi | FU(t)\psi\rangle \\ &= \langle \psi | \underbrace{U(t)^\dagger FU(t)}_{=:F(t)}\psi\rangle. && (1.29)\end{aligned}$$

Equations (1.28, 1.29) suggest two different ways of modeling the situation.



In Eq. (1.28), the time delay is part of the state preparation. If some procedure produces a system in state ψ, then "that procedure followed by a waiting time" results in the state $\psi_t = U(t)\psi$. This point of view is the *Schrödinger picture*.

In Eq. (1.29), the time delay is part of the measurement process. If some procedure measures F, then "first wait, then measure" is also an observable, described by $F(t) := U(t)^\dagger FU(t)$. This point of view is the *Heisenberg picture*.

Exercises.

- Use $H = H^\dagger$ to show that

$$U(t)^\dagger = U(-t). \tag{1.30}$$

- Prove that for any polynomial function f,

$$[f(X), P] = i\hbar f'(X) \quad \text{and} \quad [X, f(P)] = i\hbar f'(P). \tag{1.31}$$

  So, "taking commutator with P acts on functions of X like a derivative" (and vice-versa).

- Prove that

$$[U(t)^\dagger FU(t), H] = U(t)^\dagger [F, H]U(t). \tag{1.32}$$

The Schrödinger equation is a differential equation for the time evolution of states. To get the analogous *Heisenberg equation of motion*, just differentiate:

$$\begin{aligned}\partial_t F(t) &= \partial_t\, U(-t)FU(t) \\ &= \frac{1}{i\hbar}\big(-HF(t) + F(t)H\big) && \text{product rule, (1.25), (1.30)} \\ &= \frac{1}{i\hbar}[F(t), H]. && (1.33)\end{aligned}$$

Applications (of the Heisenberg picture).

- The constant function $F(t) = F$ solves (1.33) iff $[F, H] = 0$. Thus, an observable is conserved iff it commutes with the Hamiltonian. In particular, energy is preserved.
- For the free Hamiltonian $H = P^2/(2m)$, using (1.31) shows that

$$X(t) = X + \frac{P}{m}t, \qquad P(t) = P \tag{1.34}$$

solves (1.33). Taking expectations, we get the relation (1.27), but now valid for *any* state, not just wave packets. Note that the proof of (1.27) relied on moderately painful Gaussian integrals and a direct generalization to arbitrary states is not feasible. But switching to the Heisenberg picture, we see that the "linear motion of expected position" is actually a property of the X observable, and can be easily proved without any reference to states!

Ehrenfest Theorem

This section is an example for the use of the Heisenberg picture. Nothing much builds on it.

A generalization of (1.34) is known as the *Ehrenfest Theorem*. Compute:

$$\begin{aligned}
\partial_t \langle X\rangle(t) &= \frac{1}{i\hbar}\langle\psi|U(-t)[X,H]U(t)\psi\rangle && \text{(1.32), (1.33)} \\
&= \langle\psi_t|\frac{1}{i\hbar}[X,\frac{1}{2m}P^2+V(X)]\psi_t\rangle \\
&= \frac{\langle P\rangle(t)}{m} && \text{(1.31)} \qquad (1.35) \\
\partial_t \langle P\rangle(t) &= -\langle V'(X)\rangle(t) && \text{(as above).} \qquad (1.36)
\end{aligned}$$

This looks very much like the *classical* equations of motion for a particle with position and momentum given by the expected quantum values

$$x(t) := \langle X\rangle(t), \qquad p(t) := \langle P\rangle(t).$$

The part that spoils it is that the right hand side of (1.36) is the "average force"

$$\langle V'(X)\rangle(t) = \int |\psi_t(x)|^2 V'(x)\,\mathrm{d}x$$

whereas for a classical point particle, we would need the "force at the average" $V'(\langle X\rangle(t))$.

We thus find that position and momentum expectation values behave classically whenever we can exchange taking averages and applying V'. This holds...

- ...approximately, if $|\psi_t(x)|^2$ is concentrated in a region that is much smaller than the length scales on which the force varies.
- ...exactly, if V is at most quadratic in X, so that the force is a linear function.

1.5.4 Numerical diagonalization

Let's face it: In general we'll need computer assistance.

To phrase the problem in terms computers are good at, we replace continuous space by a discrete lattice. Specifically, cover the interval $[0, L]$ by N points of distance $a = N/L$. Then a function $\psi : [0, L] \to \mathbb{C}$ can hopefully be well-approximated by its values $\psi(ar)$, $r \in \{1 \dots N\}$ on the lattice.

It's clear how to discretize the potential energy operator: $(V\psi)(ar) = V(ar)\psi(ar)$.

For the kinetic energy, we have to replace the second derivative by a suitable finite difference quotient. It turns out that

$$(\Delta_a\psi)(x) = \frac{\psi(x-a) - 2\psi(x) + \psi(x+a)}{a^2}$$

is a good choice, because we recover the original version in the continuum limit $a \to 0$, as

$$(\Delta_a \psi)(x) = \frac{\frac{\psi(x+a)-\psi(x)}{a} - \frac{\psi(x)-\psi(x-a)}{a}}{a} \to \frac{\psi'(x+a) - \psi'(x)}{a} \to \psi''(x) \quad (a \to 0).$$

Adopting cyclic boundary conditions so that we'll never reach the "end of the discretized world", we obtain the following matrix approximation to the Hamilton operator:

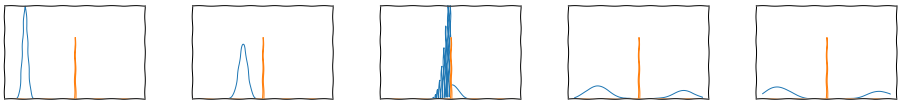
$$H_a = \frac{-\hbar^2}{2ma^2} \begin{pmatrix} -2 & 1 & & & 1 \\ 1 & -2 & 1 & & \\ & & \ddots & & \\ & & 1 & -2 & 1 \\ 1 & & & 1 & -2 \end{pmatrix} + \begin{pmatrix} V(a) & & & & \\ & V(2a) & & & \\ & & \ddots & & \\ & & & & V(Na) \end{pmatrix}.$$

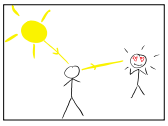
Follow this link to see the pretty results.

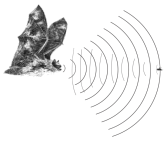
> ***Remark.*** The method presented here has been chosen because it's conceptually close to what one uses in the analytic theory, not because it's the most efficient way to find the time evolution on a classical computer. For the latter, consult the computational physics course!

1.5.5 Scattering & the tunnel effect

Flip through the numerical simulation of a particle tunneling through a potential barrier:







Picture credit.

There's little hope to analytically describe the self-interference effects on display while the particle interacts with the potential (top center). But at the beginning and at the end of the process, the wave function is away from the potential and the time evolution is just free. And indeed, one can often understand quite explicitly how the outgoing state is related to the incoming state.

A situation where particles are initially free, then interact with each other or a potential, only to move out and become free again, is called a *scattering process*. In real life, we learn about objects from their scattering properties all the time (left).

We won't cover scattering theory in general in these notes. Rather, we'll take the one-dimensional tunnel effect as an example to explain how one can read off quantitative properties of a scattering process from suitably chosen eigenfunctions of H.

Tunneling through a rectangular potential barrier

Consider a "potential barrier" of width a and height V_0:

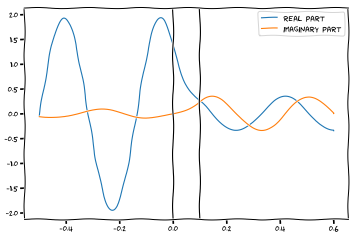
$$V(x) = \begin{cases} 0 & x \in (-\infty, 0) \quad \text{region I} \\ V_0 & x \in [0, a] \quad \text{region II} \\ 0 & x \in (a, \infty) \quad \text{region III} \end{cases}.$$

A classical particle with energy $E < V_0$ cannot transverse the barrier.

As in Sec. 1.4.2, the eigenfunctions of H with energy $E < V_0$ are of the form

$$\phi(x) = \begin{cases} Ae^{ikx} + Be^{-ikx} & x \in (-\infty, 0) \\ Ce^{\kappa x} + De^{-\kappa x} & x \in [0, a] \\ Fe^{ikx} + Ge^{-ikx} & x \in (a, \infty) \end{cases}, \qquad \begin{aligned} k &= \sqrt{2mE}/\hbar, \\ \kappa &= \sqrt{2m(V_0 - E)}/\hbar \end{aligned}.$$

The six coefficients $A, \ldots, G \in \mathbb{C}$ are related by two interface conditions each for $x = 0$ and $x = a$, leaving us with two independent parameters. In other words, energy eigenspaces are two-fold degenerate. As we'll see, the eigenfunctions ϕ_E^+, defined by $A = 1$ and $G = 0$, and ϕ_E^-, defined by $G = 1$ and $A = 0$, are well-suited for describing scattering states that move in from the left or from the right respectively.



We'll focus on ϕ_E^+, given by

$$\phi_E^+(x) = \begin{cases} e^{ikx} + B_k e^{-ikx} & \text{region I} \\ C_k e^{\kappa x} + D_k e^{-\kappa x} & \text{region II} \\ F_k e^{ikx} & \text{region III} \end{cases}.$$

Finding the $B_k, ..., F_k$ requires a lengthy calculation, see below.

We can try to interpret the terms of ϕ_E^+. In region I, it resembles a superposition of a momentum eigenstate with momentum $+\hbar k$ and ampltiude 1, and one with momentum $-\hbar k$ and amplitude B_k. It is then reasonable to suspect that if a particle approaches the barrier with momentum $\hbar k$, the probability to find it in a reflected state after the interaction is $|B_k|^2$. Likewise, in region III, ϕ_E^+ agrees with a momentum eigenstate for $+\hbar k$ and amplitude F_k, and hence we might expect the *tunneling probability* to be $|F_k|^2$.

The interface conditions. Computing the reflection coefficient B_k and the transmission coefficient F_k is not conceptually difficult, but quite lengthy.

The first interface condition at $x = 0$ is the same as (1.21). In matrix form:

$$\begin{pmatrix} A \\ B \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 - i\frac{\kappa}{k} & 1 + i\frac{\kappa}{k} \\ 1 + i\frac{\kappa}{k} & 1 - i\frac{\kappa}{k} \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix}. \tag{1.37}$$

Likewise, the interface conditions are $x = a$ read

$$\begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} e^{-\kappa a} & 0 \\ 0 & e^{\kappa a} \end{pmatrix} \frac{1}{2} \begin{pmatrix} 1 + i\frac{k}{\kappa} & 1 - i\frac{k}{\kappa} \\ 1 - i\frac{k}{\kappa} & 1 + i\frac{k}{\kappa} \end{pmatrix} \begin{pmatrix} e^{ika} & 0 \\ 0 & e^{-ika} \end{pmatrix} \begin{pmatrix} F \\ G \end{pmatrix}. \tag{1.38}$$

You can find this expression either by computing the derivatives directly, or by modifying (1.37) realizing that (i) at $x = a$, the roles of ik and κ are exchanged, and (ii) the diagonal matrices represent the change in the various exponential functions due to the shift $x \mapsto x + a$.

The product of all four matrices expresses A, B in terms of E, F. For $A = 1, F = 0$, an unpleasant calculation leads to

$$F_k = \frac{e^{-ia k}}{\cosh(a\kappa) + \frac{i}{2}\left(\frac{\kappa}{k} - \frac{k}{\kappa}\right)\sinh(a\kappa)}, \qquad B_k = \frac{(k^2 + \kappa^2)}{(k^2 - \kappa^2) + 2ik\kappa\coth(a\kappa)}.$$

For large values of a, the tunnel probability decreases exponentially in the width of the barrier

$$T := |F_k|^2 \propto e^{-2a\kappa}.$$

There's a new phenomenon for energies $E > V_0$, where one finds oscillations in the tunneling probability as a function of k. The details are homework.

Justification of dynamical interpretation

This section justifies why the coefficients that appear in the energy eigenvectors can be interpreted as the dynamical transition & reflection probability. Feel free to skip if you already think that's vaguely plausible.

This way of interpreting the coefficients in the *time-independent* eigenfunction as parameters of the *dynamical* scattering process can indeed be justified. To make this precise, let's track a wave packet $\psi_t(x)$ that at time $t = 0$ is contained in region I. We also assume that the Fourier transform $\tilde{\psi}(k)$ of $\psi_0(x)$ is narrowly concentrated around some $k_0 > 0$, so that the particle is initially approaching the barrier from the left.

Away from the barrier, ϕ_E^+ is a superposition of three momentum eigenstates

$$\phi_E^+(x) = e^{ikx}\theta^-(x) + B_k e^{-ikx}\theta^-(x) + F_k e^{ikx}\theta^+(x) \quad (x \text{ in I or III}), \tag{1.39}$$

where we have used the step functions

$$\theta^-(x) = \begin{cases} 1 & x \leq 0 \\ 0 & x > 0 \end{cases} \quad \text{and} \quad \theta^+(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} .$$

We'll show that (1.39) allows us to express $\psi_t(x)$ in terms of three free wave packets:

$$\alpha_t(x) := \frac{1}{\sqrt{2\pi}} \int \tilde{\psi}(k) e^{+ikx - i\omega_k t}\, \mathrm{d}k,$$

$$\beta_t(x) := \frac{1}{\sqrt{2\pi}} \int \tilde{\psi}(k) B_k e^{-ikx - i\omega_k t}\, \mathrm{d}k \simeq \frac{B_{k_0}}{\sqrt{2\pi}} \int \tilde{\psi}(k) e^{-ikx - i\omega_k t}\, \mathrm{d}k, \tag{1.40}$$

$$\gamma_t(x) := \frac{1}{\sqrt{2\pi}} \int \tilde{\psi}(k) F_k e^{+ikx - i\omega_k t}\, \mathrm{d}k \simeq \frac{F_{k_0}}{\sqrt{2\pi}} \int \tilde{\psi}(k) e^{-ikx - i\omega_k t}\, \mathrm{d}k. \tag{1.41}$$

Their interpretations are (Fig. 1.1):

- $\alpha_t(x)$ is the time evolution of $\psi_0(x)$ under the free Hamiltonian. It will keep on moving to the right as $t \to \infty$.
- $\beta_t(x) \simeq B_{k_0}\alpha_t(-x)$ is a re-scaled mirror image of $\alpha_t(x)$.
- $\gamma_t(x) \simeq F_{k_0}\alpha_t(x)$ is a re-scaled copy of $\alpha_t(x)$.

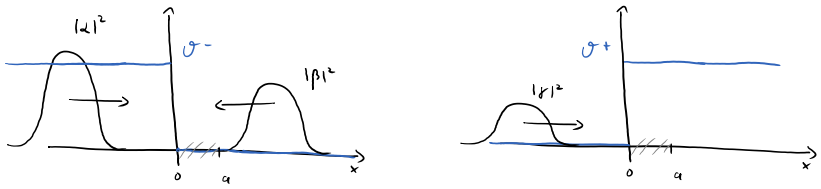


Figure 1.1: For $t \to \pm\infty$, the time evolution of a scattered state $\psi_t(x)$ can be expressed as the superposition of three *free* wave packets (black), multiplied by step functions (blue).

Now consider the following solution to the Schrödinger equation for H:

$$\delta_t(x) = \frac{1}{\sqrt{2\pi}} \int \tilde{\psi}(k) \phi_{E_k}^+(x) e^{-i\omega_k t}\, \mathrm{d}k.$$

(We'll find below that $\delta_t = \psi_t$, but it's better to use different symbols until this has been established). Plugging in (1.39), we see that for all x in region I and III,

$$\delta_t(x) = \alpha_t(x)\theta^-(x) + \beta_t(x)\theta^-(x) + \gamma_t(x)\theta^+(x) \quad (x \text{ in I or III}).$$

Let's analyze the behavior of $\delta_t(x)$.

- For $t \leq 0$: By assumption, the function α_t is contained in region I. Then its mirror image β_t is contained in region III, and thus does not contribute, because the step function θ^- is zero there. The final term does not contribute for similar reasons. It then follows that $\delta_t(x) = \psi_t(x)$. Because both δ_t (by construction) and ψ_t (by assumption) solve the Schrödinger equation and agree at one time, they must be equal: $\psi_t(x) = \delta_t(x)$.
- For $t \gg 0$: α_t and γ_t will have moved into region III under the free time evolution. The mirror image β_t now moves to the left in region I. The step functions kill the first term, leaving us with a reflected and a transmitted copy of $\alpha_t(x)$, as claimed.
- As a bonus, we can explain some of the wild interference fringes seen during the scattering process. They happen for times t where the trailing part of the free incoming wave α_t interferes with the front of its mirror image β_t around $x = 0$.

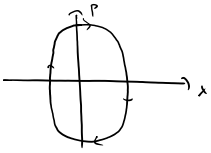
1.6 The harmonic oscillator

Classically, a *harmonic oscillator* is a system with Hamilton function of the form

$$H(x,p) = \frac{1}{2m}p^2 + \frac{1}{2}m\omega^2 x^2.$$

Its trajectories are ellipses in phase space,

$$\begin{pmatrix} x(t) \\ p(t) \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{2}{m\omega^2 E}}\cos(\omega t + \phi) \\ \sqrt{2mE}\sin(\omega t + \phi) \end{pmatrix}.$$



The harmonic oscillator is the most important model system in physics. Some reasons:

- Modes of classical & quantum fields are described by harmonic oscillators.
- Equilibrium positions of classical systems correspond to local minima in the potential. Around a minimum, the potential can be approximated by a parabola. Therefore, small perturbations around an equilibrium position lead to harmonic motion.
- Most physics problems are too difficult to be solved from first principles. You know what kind of systems one can solve efficiently, both with pen and paper and a computer? That's right: coupled harmonic oscillators. This lead to a "filter effect", where systems that cannot be treated this way just can't be treated at all, and physicists stop talking about them. As Peskin is said to have quipped: "*Physics is that subset of human experience which can be reduced to coupled harmonic oscillators.*"

1.6.1 Algebraic solution

The goal is to solve the time-independent Schrödinger equation $H\phi = E\phi$ for

$$H = \frac{1}{2m}P^2 + \frac{1}{2}m\omega^2 X^2.$$

That's a linear, second-order, ordinary differential equation:

$$-\frac{\hbar^2}{2m}\phi''(x) + \frac{1}{2}m\omega^2 x^2\phi(x) - E\phi(x) = 0.$$

It can be solved directly from this point of view, and if your hobbies include studying the intricacies of Hermite polynomials, I invite you to go right ahead. For the rest of us, there is a slightly more abstract, but also vastly more elegant, way of finding the eigenvalues and eigenvectors of H, using not much more than the commutation relation $[X, P] = i\hbar$. We will walk through this *algebraic solution* step by step.

Problem-adapted units

In terms of the dimensionless operators

$$\tilde{X} = \sqrt{\frac{m\omega}{\hbar}}X, \qquad \tilde{P} = \sqrt{\frac{1}{m\hbar\omega}}P,$$

the Hamiltonian and the commutation relations simplify to

$$H = \frac{1}{2}\hbar\omega(\tilde{P}^2 + \tilde{X}^2), \qquad [\tilde{X}, \tilde{P}] = \frac{1}{\hbar}[X, P] = i\mathbb{1}.$$

Ladder operators

Switch to "complex coordinates"

$$a := \frac{1}{\sqrt{2}}(\tilde{X} + i\tilde{P}) \quad \Rightarrow \quad a^\dagger = \frac{1}{\sqrt{2}}(\tilde{X} - i\tilde{P}). \tag{1.42}$$

For reasons we'll see momentarily, these are called *ladder operators*. Then:

$$[a, a^\dagger] = \frac{1}{2}\big(-i[\tilde{X}, \tilde{P}] + i[\tilde{P}, \tilde{X}]\big) = \mathbb{1}$$

and their "absolute value-squared" $N := a^\dagger a$ is essentially equal to the Hamiltonian:

$$N = \frac{1}{2}\left(\tilde{X}^2 + \tilde{P}^2 + i\tilde{X}\tilde{P} - i\tilde{P}\tilde{X}\right) = \frac{1}{2}\left(\tilde{X}^2 + \tilde{P}^2 + i[\tilde{X}, \tilde{P}]\right) = \frac{1}{2}\left(\tilde{X}^2 + \tilde{P}^2 - \mathbb{1}\right).$$

Indeed, re-arranging,

$$H = \hbar\omega\Big(N + \frac{1}{2}\mathbb{1}\Big). \tag{1.43}$$

The operator N is the *occupation number operator*, a name that will become clear soon.

Remark. The complex coordinates $a = \frac{1}{\sqrt{2}}(x + ip)$ are standard in the quantum treatment of the harmonic oscillator, but also make sense classically. See Appendix B.3.

The eigenvalue equation for the occupation number operator

To solve the eigenvalue equation for H it suffices to solve it for N, as

$$N\phi = n\phi \qquad \Leftrightarrow \qquad \hbar\omega\big(N + \frac{1}{2}\mathbb{1}\big)\phi = \hbar\omega\big(n + \frac{1}{2}\big)\phi.$$

We'll now work out the set of eigenvalues of N in a number of mathematical steps.

1. If n is an eigenvalue of N, then $n \geq 0$.

Interpretation: Classically, the harmonic oscillator can attain any non-negative energy. The claim says that the energies of the quantum version are also non-negative... ...but, looking at (1.43), are actually bounded below by $\frac{1}{2}\hbar\omega > 0$. That makes sense, recalling the uncertainty relation: A classical harmonic oscillator attains energy 0 exactly for a particle that is located at $x = 0$ and has momentum $p = 0$, which can't be realized in QM.

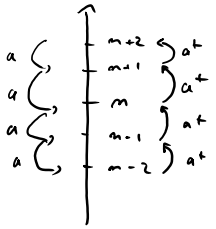
Proof. Let ϕ be an associated eigenvector. Assume that ϕ is normalizable. Then we may as well choose ϕ such that $\langle\phi|\phi\rangle = 1$, in which case

$$n = \langle\phi|N\phi\rangle = \langle\phi|a^\dagger a\phi\rangle = \langle a\phi|a\phi\rangle = \|a\phi\|^2 \geq 0.$$

(It turns out that all eigenvectors of H are normalizable. If you don't want to assume this, a limit argument still gives the same result. We skip the details). □

Let ϕ be a normalized eigenvector of N with eigenvalue n. Then

2. $a^\dagger\phi$ is eigenvector with eigenvalue $n + 1$ and norm $\|a\phi\| = \sqrt{n+1}$.
3. If $n \neq 0$, then $a\phi$ is eigenvector with eigenvalue $n-1$ and norm $\|a\phi\| = \sqrt{n}$.



Interpretation: The result explains the term "ladder operators".

Specifically, $a^\dagger$ is the *creation operator* (because it "adds an excitation" to the oscillator), and a is the *annihilation operator*. (The "dagger" symbol that physicists use to denote the adjoint kinda looks like a $+$, so you can remember which one is the creation operator).

Proof. We only prove the second claim, the first one works analogously. By assumption,

$$\|a\phi\|^2 = \langle\phi|a^\dagger a\phi\rangle = n.$$

In particular, if $n \neq 0$, then $a\phi \neq 0$, too. To see that $a\phi$ is an eigenvector, compute

$$[N, a] = a^\dagger aa - \underbrace{aa^\dagger}_{\mathbb{1}+a^\dagger a}a = -a \quad \Rightarrow \quad N(a\phi) = (aN - a)\phi = (n-1)(a\phi).$$

□

4. The eigenvalues of N are the natural numbers, $n \in \{0, 1, 2, \dots\}$.

Interpretation: The differences between energy levels of a quantum oscillator thus come in multiples of $\hbar\omega$. These are the *quanta* that Max Planck postulated in order to explain the spectrum of the black body radiation, and which give QM its name. It is often said that a harmonic oscillator in an n-eigenstate is "occupied by n excitations", which one talks about like particles. "Huh huh!", I hear you say, "These aren't *real particles*... ...just labels for the eigenfunctions of an oscillator!" Joke's on you, though, because "excitations of an oscillator" is pretty much how particles are *defined* in quantum field theory.

Proof. Let n be an eigenvalue of N.

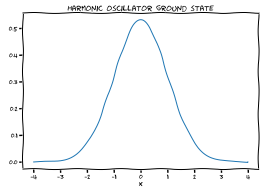
"Downwards": Applying a repeatedly, we get a sequence $n, n-1, n-2, \dots$ of eigenvalues. Because all eigenvalues are non-negative, the sequence has to abort. By the previous claim, it aborts iff it reaches 0. This happens iff $n \in \mathbb{N}$.

"Updwards": Applying $a^\dagger$ repeatedly, every natural number $n' > n$ can be reached.

□

5. The eigenspace for $n = 0$ is non-degenerate. A normalized eigenvector is

$$\phi_0(x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-\frac{1}{2}\frac{m\omega}{\hbar}x^2}. \tag{1.44}$$



That's the only part of the argument which requires us look at the H in detail, rather than rely on commutation relations alone. The lowest-energy eigenstate, or *ground state*, of the oscillator is a Gaussian wave function.

Proof. The eigenvalue equation $N\phi_0 = 0$ can be converted into one for a:

$$N\phi_0 = 0 \quad \Leftrightarrow \quad \langle\phi_0|a^\dagger a\phi_0\rangle = 0 \quad \Leftrightarrow \quad \langle a\phi_0|a\phi_0\rangle = 0 \quad \Leftrightarrow \quad a\phi_0 = 0.$$

The latter expands into the first-order differential equation

$$a\phi_0(x) = \sqrt{\frac{m\omega}{2\hbar}}x\phi_0(x) + \sqrt{\frac{\hbar}{2m\omega}}\phi_0'(x) = 0.$$

Plugging in, one finds that

$$\phi_0(x) = Ae^{-\frac{1}{2}\frac{m\omega}{\hbar}x^2}, \qquad A \in \mathbb{C}$$

is a one-parameter family of solutions. Because the ODE is first-order, these are all. □

6. *All eigenspaces are non-degenerate.*

Proof. Proof by contradiction.

Assume that there are two eigenvectors ϕ_n, ψ_m such that

- Both are n-eigenvectors:

$$N\phi_n = n\phi_n, \qquad N\psi_n = n\psi_n.$$

- They are orthogonal to each other (and hence linearly independent)

$$\langle \phi_n | \psi_n \rangle = 0.$$

Define

$$\phi_{n-1} = a\phi_n, \qquad \psi_{n-1} = a\psi_n.$$

Then both are $(n-1)$-eigenvectors, and

$$\langle \phi_{n-1} | \psi_{n-1} \rangle = \langle a\phi_n | a\psi_n \rangle = \langle \phi_n | a^\dagger a \psi_n \rangle = \langle \phi_n | N\psi_n \rangle = n\langle \phi_n | \psi_n \rangle = 0.$$

Thus $n-1$ is degenerate, and by induction, all eigenvalues $\leq n$ are degenerate.

But 0 is non-degenerate, a contradiction. □

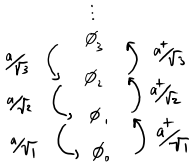
Summary

The eigenvalues of the harmonic oscillator are non-degenerate and given by

$$E_n = \hbar\omega\Big(n + \frac{1}{2}\Big), \qquad n \in \mathbb{N}.$$

The eigenvectors can be be constructed by "laddering" starting from the (Gaussian) ground state ϕ_0. The resulting vector has normalization

$$\|(a^\dagger)^n \phi_0\|^2 = n\|(a^\dagger)^{n-1}\phi_0\|^2 = n(n-1)\|(a^\dagger)^{n-2}\phi_0\|^2 = \cdots = n!\|\phi_0\|^2 = n!$$



so that normalized eigenvectors are

$$\phi_n := \frac{1}{\sqrt{n!}}(a^\dagger)^n \phi_0.$$

1.6.2 Properties of eigenstates

Explicit eigenfunctions

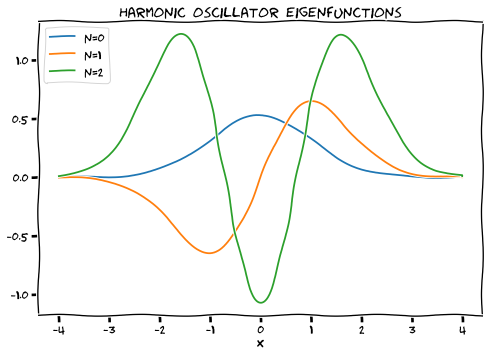
Fortunately, one rarely ever needs to work with the explicit form of the eigenfunctions. But in case you like nice plots, here's a few of them. Setting $m = \omega = \hbar = 1$ for simplicity:

$$\begin{aligned}
\phi_1(x) &= a^\dagger \phi_0(x) = \frac{1}{\sqrt{2}}(X - iP)\phi_0(x) = \frac{1}{\sqrt{2}\pi^{-1/4}}(x - \partial_x)e^{-\frac{1}{2}x^2} \\
&= \frac{1}{\sqrt{2}\pi^{-1/4}} 2xe^{-\frac{1}{2}x^2} \\
\phi_2(x) &= \frac{1}{\sqrt{2}}a^\dagger \phi_1(x) = \frac{1}{\sqrt{2}\pi^{-1/4}}(x - \partial_x)2xe^{-\frac{1}{2}x^2} = \frac{1}{2\pi^{-1/4}}(2x^2 - 2 + 2x^2)e^{-\frac{1}{2}x^2} \\
&= \frac{1}{2\pi^{-1/4}}(4x^2 - 2)e^{-\frac{1}{2}x^2}
\end{aligned}$$

$$\vdots$$

$$\phi_n(x) = \frac{1}{\sqrt{n!2^n}}\frac{1}{\pi^{1/4}}\left(x - \partial_x\right)^n e^{-\frac{1}{2}x^2} = \frac{1}{\sqrt{n!2^n}}H_n(x)e^{-\frac{1}{2}x^2},$$

where $H_n(x)$ is the n-th *Hermite polynomial*.



Expectation values

So how does one compute expectation values if the specific form of the eigenfunctions is rarely used? The trick is to express observables as polynomials in ladder operators. (One can show that this is always possible.) For example, X and P expectations vanish:

$$\begin{aligned}
\langle\phi_n|\tilde{X}\phi_n\rangle &= \frac{1}{\sqrt{2}}\langle\phi_n|(a + a^\dagger)\phi_n\rangle = \frac{1}{\sqrt{2}}\left(\sqrt{n}\langle\phi_n|\phi_{n-1}\rangle + \sqrt{n+1}\langle\phi_n|\phi_{n+1}\rangle\right) = 0, \\
\langle\phi_n|\tilde{P}\phi_n\rangle &= \frac{-i}{\sqrt{2}}\langle\phi_n|(a - a^\dagger)\phi_n\rangle = \cdots = 0.
\end{aligned}$$

Likewise, we can compute the variances from

$$\langle\phi_n|\tilde{X}^2\phi_n\rangle = \frac{1}{2}\langle\phi_n|(a^2 + \underbrace{aa^\dagger}_{a^\dagger a+\mathbb{1}} + a^\dagger a + (a^\dagger)^2)\phi_n\rangle = \frac{1}{2}(n+1+n) = n + \frac{1}{2},$$

$$\langle\phi_n|\tilde{P}^2\phi_n\rangle = ... = n + \frac{1}{2}.$$

In adapted units, the variances are thus equal, never zero, and increase with the energy.

1.6.3 Coherent states

Homework.

Chapter 2

The mathematical framework of QM

The last chapter was phrased in terms of *wave functions* $\psi(x)$. That's a good starting point, because these are very concrete objects. However, the values $\psi(x)$ are not directly physically observable – only probabilities $|\langle\phi_\lambda|\psi\rangle|^2$ or expectation values $\langle\psi|F|\psi\rangle$ are.

This suggests that these linear-algebraic expressions are more fundamental than the explicit functions $\psi(x)$. Also, pragmatically, computing with abstract linear-algebraic objects is often much simpler. Take, e.g. our one-line calculation of the position variance $\langle\phi_n|\tilde{X}^2|\phi_n\rangle$ in a harmonic oscillator eigenstate. One *can* compute this as an integral over explicit Hermite polynomials – but that is unlikely to brighten the mood.

This suggests that one should re-formulate QM, so that abstract Hilbert spaces and linear operators are the fundamental objects. Concrete functions will then appear as the representation of abstract vectors in a concrete basis.

For better or worse, this linear-algebraic language is the one in which QM is most often discussed. In this chapter, we will introduce the relevant notions. (This involves repeating some ideas that already came up in Linear Algebra and Chapter 1).

2.1 Hilbert spaces

A *Hilbert space* $\mathcal{H}$ is a complex vector space with a *sesquilinear* inner product $\langle\cdot|\cdot\rangle$. Sesquilinearity means that for all vectors

$$\alpha, \beta, \gamma \in \mathcal{H}$$

and complex numbers $z \in \mathbb{C}$, we have

$$\langle\alpha|\beta+\gamma\rangle = \langle\alpha|\beta\rangle + \langle\alpha|\gamma\rangle, \tag{2.1}$$
$$\langle\alpha|z\beta\rangle = z\langle\alpha|\beta\rangle, \tag{2.2}$$

as well as

$$\langle\alpha|\beta\rangle = \overline{\langle\beta|\alpha\rangle}. \tag{2.3}$$

From this, it follows that

$$\langle\alpha+\beta|\gamma\rangle = \langle\alpha|\gamma\rangle + \langle\beta|\gamma\rangle,$$
$$\langle z\alpha|\beta\rangle = \bar{z}\langle\alpha|\beta\rangle,$$

i.e. the inner product is anti-linear w.r.t. the first entry and linear w.r.t. the second one.

Remark. Beware that mathematicians usually employ the opposite convention, where the sesquilinear inner product is linear in the first entry!

The *norm* of a vector $\alpha \in \mathcal{H}$ is given by

$$\|\alpha\| := \sqrt{\langle \alpha | \alpha \rangle}.$$

Recall that inner products are required to be *definite*, i.e. to fulfill

$$\|\alpha\| > 0 \quad \forall \alpha \neq 0.$$

There are two examples of Hilbert spaces you should be acquainted with: *column vectors* and *square-integrable functions*. Let's look at both in turn.

The vector space $\mathbb{C}^d$ is formed by d-dimensional complex column vectors

$$\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_d \end{pmatrix}$$

with sesquilinear inner product

$$\langle \alpha | \beta \rangle = \sum_{i=1}^{d} \bar{\alpha}_i \beta_i. \tag{2.4}$$

Such Hilbert spaces appears e.g. in the description of spin degrees of freedom.

The second class are the spaces $L^2(\mathbb{R}^n)$ of *square-integrable complex functions on* $\mathbb{R}^n$. ("wave functions" in physics jargon – we've already encountered the $n = 1$ case in Chapter 1). Given two functions $\alpha, \beta : \mathbb{R}^n \to \mathbb{C}$, define a "continuous analogue" of Eq. (2.4) by

$$\langle \alpha | \beta \rangle = \int \bar{\alpha}(x) \beta(x) \, \mathrm{d}^n x. \tag{2.5}$$

For the non-pedantic physicist, the space of all complex-valued functions, together with the form (2.5) defines a Hilbert space. It is associated with a point particle with n degrees of freedom.

Remark. There are three technical problems that one has to address to define the Hilbert space of functions with mathematical rigor.

The first problem is the integral is not actually defined for all functions. Set, e.g.

$$\psi(x) = \begin{cases} \sin(1/x) & x \neq 0, \\ 0 & x = 0. \end{cases}$$

Then

$$\int |\alpha(x)|^2 \, \mathrm{d}x$$

does not exist. The second problem is that the integral may be defined, but infinite – take e.g. $\alpha(x) = 1$ and compute $\langle \alpha | \alpha \rangle$. To get rid of both problems, we define a function α to be *square-integrable* if

$$\|\alpha\|^2 = \langle \alpha | \alpha \rangle = \int |\alpha(x)|^2 \, \mathrm{d}^n x$$

exists and is finite. If α, β are square-integrable, then the product $\bar{\alpha}\beta$ is integrable, and the *Cauchy-Schwarz inequality* says that

$$|\langle\alpha|\beta\rangle|^2 \leq \|\alpha\|^2 \, \|\beta\|^2 < \infty,$$

so that, by restricting to square-integrable functions, we have rid ourselves of undefined and infinite integrals!

The third problem is that the norm is no longer definite. Indeed, define a function

$$\alpha(x) = \begin{cases} 1 & x = 0 \\ 0 & x \neq 0 \end{cases} .$$

Then $\alpha \neq 0$, but $\|\alpha\|^2 = 0$. Circumventing this problem requires some mathematical gymnastics: We say that two functions are *equivalent* if they differ only on a set of measure zero. This means e.g. that the function α is equivalent to the 0-function, as the two differ only at one point. If we define the Hilbert space $L^2(\mathbb{R}^n)$ to be the *complex vector space of equivalence classes* of square-integrable functions, then one can show that (2.5) becomes a *definite* inner product. Problem solved.

2.2 Linear operators

Recall that a map A between two vector spaces is *linear* if

$$A(\phi + \psi) = A(\phi) + A(\psi) \quad \text{and} \quad A(\lambda\phi) = \lambda A(\phi).$$

In QM, linear maps between Hilbert spaces are traditionally called *operators*.

Examples:

- $\mathcal{H} = \mathbb{C}^d$: In this case, operators can conveniently be specified as matrices, which act on column vectors in the usual way. For example, we will have ample opportunity to work with the *Pauli matrices*:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_x = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

- In $\mathcal{H} = L^2(\mathbb{R})$, we have already made contact with the *position operator*, which acts on a function $\psi : \mathbb{R} \to \mathbb{C}$ by multiplying it with its argument

$$(X\psi)(x) = x\psi(x),$$

and the *momentum operator* maps a function to $-i\hbar$ times its derivative

$$P = -i\hbar\partial_x : \psi \mapsto P\psi = -i\hbar\psi'.$$

2.3 Dirac notation

Physicists often use notational aids to distinguish vector-valued quantities from scalars. In quantum mechanics, the suggestive *Dirac* notation (or "bra-ket" notation) is usually employed. Here, a vector $\alpha \in \mathcal{H}$ is written as $|\alpha\rangle$. This is called a *ket*, for reasons that will be obvious momentarily.

Every vector $\psi \in \mathcal{H}$ defines a linear function

$$\mathcal{H} \to \mathbb{C}, \qquad \phi \mapsto \langle\psi|\phi\rangle, \tag{2.6}$$

the "inner product with ψ". In quantum, we denote this function as $\langle\psi|$ and call it a *bra*. Then we can write

$$\langle\psi|(|\phi\rangle) = \langle\psi|\phi\rangle, \tag{2.7}$$

so a "bra-ket" is a "braket". This passes for humor around here.

Remark. Linear functions from a vector spaces to $\mathbb{C}$ are also called *dual vectors* or *(linear) functionals*. Confusion alert: In the calculus of variation – i.e. the branch of analysis that turns the action principle into the Euler-Lagrange equation – the word "functional" is used instead to refer to a function that takes other functions as arguments.

The genius of this notation is that one doesn't need to expend any thoughts on concepts like "dual vectors" or "linear functionals" – the formalism almost forces one to use these object correctly.

Let's play around with this. Equation (2.7) is the *inner product* between $|\psi\rangle$ and $|\phi\rangle$. One can combine two vectors also to form an *outer product*, namely the linear operator $\mathcal{H} \to \mathcal{H}$ defined as

$$|\beta\rangle \mapsto |\phi\rangle\langle\psi|\left(|\beta\rangle\right) := |\phi\rangle\left(\langle\psi|\beta\rangle\right). \tag{2.8}$$

Definition (2.8) implies that composing bras and kets is associative: One can read the expression

$$|\phi\rangle\langle\psi|\beta\rangle$$

as either

$$\left(|\phi\rangle\langle\psi|\right)\left(|\beta\rangle\right) \qquad \text{"operator acting on vector"}$$

or as

$$|\phi\rangle\left(\langle\psi|\beta\rangle\right) \qquad \text{"vector times inner product"},$$

getting the same result.

Remark. Here's a more systematic interpretation of (2.8). Bras were defined as linear maps $\mathcal{H} \to \mathbb{C}$. Analogously, we can interpret a ket $|\phi\rangle$ as the linear map

$$\mathbb{C} \to \mathcal{H}, \qquad \lambda \mapsto \lambda|\phi\rangle \tag{2.9}$$

(The same way an $n \times 1$-matrix can be read either as a column vector in $\mathbb{C}^n$ or as a linear map $\mathbb{C} \to \mathbb{C}^n$). Then a product of kets and bras is nothing but a concatenation of linear maps. For example, (read from right to left):

$$\underbrace{\langle\gamma|}_{\text{...and on to a number}} \qquad \underbrace{|\beta\rangle}_{\text{...that number to a vector...}} \qquad \underbrace{\langle\alpha|}_{\text{map a vector to a number...}} \qquad : \mathcal{H} \to \mathbb{C}.$$

$\longleftarrow$

The associativity rule found above is thus nothing but the associativity of the composition of linear maps.

2.4 Bases

Let $\mathcal{H}$ be a Hilbert space. A discrete set $\{|e_i\rangle\}_i \subset \mathcal{H}$ is called *ortho-normal* if

$$\langle e_i|e_j\rangle = \delta_{i,j}.$$

If in addition, every element $|\psi\rangle \in \mathcal{H}$ can be expressed as a liner combination

$$|\psi\rangle = \sum_i \psi_i |e_i\rangle \tag{2.10}$$

with suitable *expansion coefficients* $\psi_i \in \mathbb{C}$, then we have an *ortho-normal basis* (ONB).

Remark.

- In physics, unless stated otherwise, "basis" always means "ortho-normal basis".
- Hilbert spaces are often infinite-dimensional. In this case, the "sum" in (2.10) is actually an infinite series, and the equality sign is to be interpreted as the statement

$$\lim_{n\to\infty} \left\| \left(\sum_{i=1}^{n} \psi_i |e_i\rangle \right) - |\psi\rangle \right\| = 0.$$

  Experience has it that ignoring these subtleties tends to not create major problems in physics. If you are bothered by this, it pays to pick up a book on functional analysis.

Every ONB fulfills the *completness relation*

$$\sum_i |e_i\rangle\langle e_i| = \mathbb{1}, \tag{2.11}$$

where $\mathbb{1} : |\psi\rangle \mapsto |\psi\rangle$ is the identity map. To prove it, calculate for an arbitrary $|\psi\rangle = \sum_i \psi_i |e_i\rangle$,

$$\left(\sum_i |e_i\rangle\langle e_i| \right) \left(\sum_j \psi_j |e_j\rangle \right) = \sum_{i,j} \psi_j |e_i\rangle \underbrace{\langle e_i|e_j\rangle}_{\delta_{i,j}} = |\psi\rangle. \tag{2.12}$$

The completeness relation implies the following properties of ONBs:

1. Expansion coefficients are given by inner products

$$|\psi\rangle = \mathbb{1}|\psi\rangle = \Big(\sum_i |e_i\rangle\langle e_i| \Big) |\psi\rangle = \sum_i \underbrace{\langle e_i|\psi\rangle}_{\psi_i} |e_i\rangle.$$

2. Expansion coefficients of bras are the complex conjugate:

$$\langle\psi| = \langle\psi|\mathbb{1} = \sum_i \underbrace{\langle\psi|e_i\rangle}_{\bar{\psi}_i} \langle e_i|.$$

3. Inner products can be computed with respect to an arbitrary ONB:

$$\langle\psi|\phi\rangle = \langle\psi|\mathbb{1}|\phi\rangle = \sum_i \langle\psi|e_i\rangle\langle e_i|\phi\rangle = \sum_i \bar{\psi}_i \phi_i.$$

The special case where $\phi = \psi$ is sometimes called the *Parseval relation*:

$$\|\psi\|^2 = \langle\psi|\psi\rangle = \sum_i |\psi_i|^2.$$

4. Description of operators via *matrix elements*

$$A = \mathbb{1}A\mathbb{1} = \sum_{i,j} |e_i\rangle\langle e_i|A|e_j\rangle\langle e_j| = \sum_{i,j} A_{i,j}|e_i\rangle\langle e_j|,\ A_{i,j} := \langle e_i|A|e_j\rangle. \quad (2.13)$$

so that

$$\langle\phi|A|\psi\rangle = \langle\phi|\mathbb{1}A\mathbb{1}|\psi\rangle = \sum_{ij} \bar{\phi}_j A_{ij}\psi_i.$$

The expression (2.13) also shows that for every basis $\{|e_i\rangle\}_i$ of the Hilbert space, the set $\{|e_i\rangle\langle e_j|\}_{ij}$ is a basis for the vector space of linear operators.

The Dirac notation allows one to save a bit of ink when working with one fixed ONB. Say we have agreed to work with $\{|e_i\rangle\}_i$. Then quantum physicists (and no-one else…) commonly drop the symbol e and just put the index into the ket:

$$|i\rangle := |e_i\rangle.$$

2.4.1 Vector and matrix representations

Assume that $\mathcal{H}$ is finite-dimensional and that some ONB $\{|i\rangle\}_{i=1}^d$ has been fixed. Then the calculations above define a one-one relation between $\mathcal{H}$ and the Hilbert space $\mathbb{C}^d$ of row vectors. Concretely, take the dictionary

$$\text{kets} \leftrightarrow \text{column vectors} \qquad |\psi\rangle \leftrightarrow \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_d \end{pmatrix}$$

$$\text{bras} \leftrightarrow \text{row vectors} \qquad \langle\psi| \leftrightarrow (\bar{\psi}_1, \ldots, \bar{\psi}_d)$$

$$\text{operators} \leftrightarrow \text{matrices} \qquad A \leftrightarrow \begin{pmatrix} A_{1,1} & \ldots & A_{1,d} \\ \vdots & & \vdots \\ A_{d,1} & \ldots & A_{d,d} \end{pmatrix}$$

with

$$\psi_i = \langle i|\psi\rangle, \qquad A_{i,j} = \langle i|A|j\rangle.$$

Under this identification, the composition rules of bras, kets, and operators correspond to the usual rules of matrix-vector multiplication. This representation is particularly useful for computer implementations!

2.5 The adjoint

Recall that in $\mathbb{R}^d$ with Euclidean scalar product

$$(u, v) = \sum_i u_i v_i$$

one can "move a matrix from one entry of the scalar product to the other by taking its transpose"

$$(u, Av) = (A^t u, v), \qquad (A^t)_{i,j} = A_{j,i}.$$

Likewise, if $\mathcal{H}$ is a Hilbert space and A an operator on $\mathcal{H}$, then there is a unique *adjoint operator* $A^\dagger$ such that

$$\langle \phi | A\psi \rangle = \langle A^\dagger \phi | \psi \rangle \qquad \forall\, \psi, \phi \in \mathcal{H}.$$

With respect to a basis, one finds the formula

$$\langle \phi | A\psi \rangle = \sum_{ij} \bar{\phi}_i A_{ij} \psi_j = \sum_j \overline{\left(\sum_i \phi_i \bar{A}_{ij} \right)} \psi_j = \langle A^\dagger \phi | \psi \rangle \quad \Rightarrow \quad (A^\dagger)_{ij} = \bar{A}_{ji}.$$

The basis representation of $A^\dagger$ is therefore the "conjugate transpose" of the one of A. The expression $A^\dagger$ for the adjoint is pronounced "A dagger". An operator A is *self-adjoint* or *Hermitian* if $A = A^\dagger$.

Properties. It is easy to see that taking the adjoint

- ...is anti-linear $(A + zB)^\dagger = A^\dagger + \bar{z} A^\dagger$,
- ...reverses products $(AB)^\dagger = B^\dagger A^\dagger$,
- ...exchanges "bras" and "kets" $\big(|\alpha\rangle\langle\beta|\big)^\dagger = |\beta\rangle\langle\alpha|$.

One can unify the last two properties by slightly generalizing the definition of the adjoint. If $\mathcal{H}$ and $\mathcal{K}$ are two Hilbert spaces, then it is still true that for every operator $A : \mathcal{H} \to \mathcal{K}$, there is a unique adjoint $A^\dagger : \mathcal{K} \to \mathcal{H}$, such that

$$\langle \phi | A\psi \rangle_{\mathcal{K}} = \langle A^\dagger \phi | \psi \rangle_{\mathcal{H}}, \qquad \phi \in \mathcal{K}, \psi \in \mathcal{H}.$$

Recall that $\mathbb{C} = \mathbb{C}^1$ is itself a Hilbert space, and that we have identified a ket $|\beta\rangle \in \mathcal{H}$ with a linear map $\mathbb{C} \to \mathcal{H}$ in (2.9). Then one directly verifies that $|\beta\rangle^\dagger = \langle\beta|$ and that the first two properties still hold. The third property is then a special case of the second one.

Examples.

- The Pauli matrices are self-adjoint, as is evident by taking the conjugate-transpose.
- The momentum operator is self-adjoint:

$$\begin{aligned} \langle \phi | P | \psi \rangle &= \int_{-\infty}^{\infty} \bar{\phi}(x)(-i\hbar)\psi'(x)\,\mathrm{d}x \\ &= -\int_{-\infty}^{\infty} (\bar{\phi})'(x)(-i\hbar)\psi(x)\,\mathrm{d}x \qquad \text{(integration by parts)} \\ &= \overline{\int_{-\infty}^{\infty} \bar{\psi}(x)(-i\hbar)\phi'(x)\,\mathrm{d}x} = \overline{\langle \psi | P | \phi \rangle}, \end{aligned}$$

  where we have used that for square-integrable functions $\lim_{x\to\pm\infty} \psi(x) = 0$, so that no boundary terms appear when integrating by parts.

2.6 Spectral decomposition (discrete case)

Recall our old friend, the *eigenvalue problem*: Given an operator $A : \mathcal{H} \to \mathcal{H}$, find all $\lambda_i, |\psi_i\rangle$ such that

$$A|\psi_i\rangle = \lambda_i |\psi_i\rangle$$

Of course, the λ_i's are the *eigenvalues* and the $|\psi_i\rangle$'s the *eigenvectors* of A.

A *spectral decomposition* (or *eigendecomposition*) of A is a representation of the form

$$A = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|, \qquad \mathbb{1} = \sum_i |\psi_i\rangle\langle\psi_i|. \tag{2.14}$$

The spectral decomposition reveals the eigenvalues and eigenvectors of A:

$$A|\psi_j\rangle = \sum_i \lambda_i |\psi_i\rangle \underbrace{\langle\psi_i|\psi_j\rangle}_{\delta_{ij}} = \lambda_j |\psi_j\rangle.$$

It follows that A has an eigendecomposition if and only if one can find an ONB comprised of eigenvectors. In this case, one refers to it as A's *eigenbasis*, and the λ_i's appearing in the decomposition are exactly the eigenvalues of A.

Not every operator has an eigenbasis, e.g. the "*spin*-$1/2$ *raising operator*"

$$\sigma_+ = \frac{1}{2}(\sigma_x + i\sigma_y) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

does not (why?). There's a theorem in functional analysis that essentially says that A has an eigendecomposition if and only if A commutes with its adjoint, i.e. $AA^\dagger = A^\dagger A$. (Though the case when there is a continuum of eigenvalues needs more attention, see section below).

The most important class of operators for which this holds are, of course, the self-adjoint ones $A = A^\dagger$. What is more, in this case, all eigenvalues are real. Indeed, $A|\psi\rangle = \lambda|\psi\rangle$ implies (taking $|\psi\rangle$ to be normalized without loss of generality)

$$\lambda = \langle\psi|A|\psi\rangle = \overline{\langle\psi|A^\dagger|\psi\rangle} = \overline{\langle\psi|A|\psi\rangle} = \bar{\lambda}.$$

Thus the self-adjoint operators are exactly those of the form

$$A = \sum_i \lambda_i |\phi_i\rangle\langle\phi_i|, \qquad \lambda_i \in \mathbb{R}, \{|\phi_i\rangle\}_i \text{ an ONB}.$$

2.7 Spectral decomposition (continuous case)

When working out eigendecompositions in infinite dimensions, we can run into trouble. Let's see what can go wrong.

First, consider the momentum operator $P = -i\frac{\mathrm{d}}{\mathrm{d}x}$. The eigenvalue equation is trivial to solve:

$$-i\psi' = \lambda\psi \quad \Leftrightarrow \quad \psi(x) = c\, e^{i\lambda x}.$$

Trouble is that these solutions are not square integrable:

$$\|\psi\|^2 = \int_{-\infty}^{\infty} |c|^2 \,\mathrm{d}x = \infty. \qquad \text{☹}$$

For the position operator $(X\psi)(x) = x\psi(x)$, the eigenvalue equation

$$x\psi(x) = \lambda\psi(x) \qquad \forall x$$

is solved by

$$\psi(x) = \begin{cases} c & x = \lambda \\ 0 & \text{else} \end{cases},$$

which has norm $\|\psi\| = 0$. So it seems like there are no eigendecompositions for the two most important operators of QM. ☹☹

To get around the problem, we widen our domain of discourse by allowing for more general objects than just square-integrable functions. Let's first see how this formally solves our problem. Whether we are "allowed to do this", i.e. whether the formal construction will lead to inconsistencies, is something we'll worry about later.

2.7.1 Delta distributions

The *distribution* δ_y is a formal object whose inner product with a smooth function ϕ is *defined* to be

$$\langle \delta_y | \psi \rangle = \int \bar{\delta}_y(x)\psi(x)\,\mathrm{d}x := \psi(y).$$

Then the expression

$$\int_x x|\delta_x\rangle\langle\delta_x|\,\mathrm{d}x$$

provides an eigendecomposition of the position operator X *in the sense* that for any pair of smooth functions ϕ, ψ we get the correct result

$$\langle\phi|\left(\int_x x|\delta_x\rangle\langle\delta_x|\,\mathrm{d}x\right)|\psi\rangle = \int x\langle\phi|\delta_x\rangle\langle\delta_x|\psi\rangle = \int x\bar{\phi}(x)\psi(x) = \langle\phi|X|\psi\rangle. \quad (2.15)$$

Likewise, we have the completeness relation

$$\int_x |\delta_x\rangle\langle\delta_x|\,\mathrm{d}x = \mathbb{1} \quad (2.16)$$

in the same sense, i.e.

$$\langle\phi|\left(\int_x |\delta_x\rangle\langle\delta_x|\,\mathrm{d}x\right)|\psi\rangle = \int \langle\phi|\delta_x\rangle\langle\delta_x|\psi\rangle = \int \bar{\phi}(x)\psi(x) = \langle\phi|\psi\rangle.$$

So *when integrated against smooth functions*, the expressions above behave just like an eigendecomposition should. We can work this that! ☺

2.7.2 Plane waves

We now turn to eigendecomposition of the momentum operator. For $k \in \mathbb{R}$, define the non-normalizable eigenfunction

$$\phi_k(x) = (2\pi)^{-1/2}e^{ikx}.$$

We claim that

$$\int_k |\phi_k\rangle\langle\phi_k|\,\mathrm{d}k = \mathbb{1}, \qquad \int_k \hbar k\,|\phi_k\rangle\langle\phi_k|\,\mathrm{d}k = P,$$

in the sense that for ψ, ϕ vanishing at infinity

$$\langle\phi|\left(\int_k |\phi_k\rangle\langle\phi_k|\,\mathrm{d}k\right)|\psi\rangle = \langle\phi|\psi\rangle, \quad (2.17)$$

$$\langle\phi|\left(\int_k \hbar k\,|\phi_k\rangle\langle\phi_k|\,\mathrm{d}k\right)|\psi\rangle = \langle\phi|P|\psi\rangle. \quad (2.18)$$

To see that this is true, note that the inner product with a function ψ

$$\langle \phi_k | \psi \rangle = (2\pi)^{-1/2} \int e^{-ikx} \psi(x) \, \mathrm{d}x = \tilde{\psi}(k)$$

gives the *Fourier transform* $\tilde{\psi}$ of ψ evaluated at k. Recall that the inverse transform is

$$(2\pi)^{-1/2} \int e^{ikx} \tilde{\psi}(k) \, \mathrm{d}k = \psi(x).$$

The completeness relation Eq. 2.17 thus follows from

$$\langle \delta_x | \left(\int_k |\phi_k\rangle\langle\phi_k| \, \mathrm{d}k \right) |\psi\rangle = (2\pi)^{-1/2} \int_k e^{ikx} \tilde{\psi}(k) \, \mathrm{d}k = \psi(x).$$

Next, for ψ vanishing at infinity, integration by parts give

$$\begin{aligned} \hbar k \langle \phi_k | \psi \rangle &= (2\pi)^{-1/2} \int \hbar k e^{-ikx} \psi(x) \, \mathrm{d}x \\ &= (2\pi)^{-1/2} \int \left(i\hbar \frac{\mathrm{d}}{\mathrm{d}x} e^{-ikx} \right) \psi(x) \, \mathrm{d}x \\ &= (2\pi)^{-1/2} \int \left(-i\hbar \frac{\mathrm{d}}{\mathrm{d}x} \psi(x) \right) e^{-ikx} \, \mathrm{d}x = \langle \phi_k | P | \psi \rangle \end{aligned}$$

which implies Eq. (2.18):

$$\langle \phi | \left(\int_k \hbar k |\phi_k\rangle\langle\phi_k| \, \mathrm{d}k \right) |\psi\rangle = \langle \phi | \left(\int_k |\phi_k\rangle\langle\phi_k| \, \mathrm{d}k \right) P|\psi\rangle = \langle \phi | P | \psi \rangle. \quad ☺☺$$

2.7.3 General eigendecompositions

We can now sketch the way in which a general Hermitian operator A has an eigendecomposition. Consider all solutions to the eigenvalue equation

$$A|\psi_\lambda\rangle = \lambda |\psi_\lambda\rangle,$$

regardless of whether $|\psi_\lambda\rangle$ is square-integrable or not. Assume for simplicity that A is non-degenerate, i.e. that for every $\lambda \in \mathbb{R}$, there is at most one eigenfunction $|\psi_\lambda\rangle$. An eigenvalue $\lambda \in \mathbb{C}$ is called *discrete* if it is separated from all other eigenvalues by a finite distance. Let D be the set of discrete eigenvalues. Eigenvalues that are not discrete are called *continuous*. Collect them in another set C. Choose normalization such that

$$\begin{aligned} \langle \psi_{\lambda'} | \psi_\lambda \rangle &= \delta_{\lambda',\lambda} && \lambda \in D, \\ \langle \psi_{\lambda'} | \psi_\lambda \rangle &= \delta(\lambda' - \lambda) && \lambda \in C. \end{aligned}$$

Then we have the completeness relation and *spectral decomposition*

$$\begin{aligned} \mathbb{1} &= \int_C |\psi_\lambda\rangle\langle\psi_\lambda| \, \mathrm{d}\lambda + \sum_{\lambda \in D} |\psi_\lambda\rangle\langle\psi_\lambda|, \\ A &= \int_C \lambda |\psi_\lambda\rangle\langle\psi_\lambda| \, \mathrm{d}\lambda + \sum_{\lambda \in D} \lambda |\psi_\lambda\rangle\langle\psi_\lambda|. \end{aligned} \tag{2.19}$$

We can unify the treatment of the discrete and the continuous part. Define

$$\rho = \sum_{\lambda \in D} \delta_\lambda + I_C \quad \text{in terms of the } \textit{indicator function } I_C(\lambda') = \begin{cases} 1 & \lambda' \in C \\ 0 & \text{else} \end{cases}.$$

The delta functions allow us to incorporate the sums in (2.19) into the integral:

$$A = \int \lambda |\psi_\lambda\rangle\langle\psi_\lambda| \rho(\lambda)\, \mathrm{d}\lambda. \tag{2.20}$$

The completeness relation generalizes like this: For any subset $S \subset \mathbb{R}$,

$$\int_S |\psi_\lambda\rangle\langle\psi_\lambda| \rho(\lambda)\, \mathrm{d}\lambda = P_S, \tag{2.21}$$

where P_S projects onto the space spanned by $\{|\psi_\lambda\rangle \mid \lambda \in S\}$. This looks somewhat like the formula

$$\int_S \rho(\lambda)\, \mathrm{d}\lambda = \mu(S)$$

for computing the measure of a set S given a density ρ. Therefore, the map $S \mapsto P_S$ is called a *projection-valued measure* and ρ the *density of states* (with respect to $\mathrm{d}\lambda$). The interpretation of ρ is particularly clear when applied to sets S that do not intersect the continuous part $S \cap C = \emptyset$. Then

$$\int_S \rho(\lambda)\, \mathrm{d}\lambda = |S \cap D|$$

equals the number of eigenvalues of A in S.

See Chapter 1 of *Quantum Mechanics* by Ballentine for a more careful, but not too technical exposition. A rigorous version is the *spectral theorem* of functional analysis.

2.8 Functions of operators

Let

$$A = \sum_i \lambda_i |\phi_i\rangle\langle\phi_i|$$

be the eigendecomposition of an operator. Then

$$A^2 = \sum_{ij} \lambda_i \lambda_j |\phi_i\rangle \underbrace{\langle\phi_i|\phi_j\rangle}_{\delta_{ij}} \langle\phi_j| = \sum_i \lambda_i^2 |\phi_i\rangle\langle\phi_i|$$

and likewise

$$A^k = \cdots = \sum_i \lambda_i^k |\phi_i\rangle\langle\phi_i|.$$

Thus, if $p(x) = \sum_k c_k x^k$ is a polynomial, then

$$p(A) = \sum_k c_k A^k = \sum_i p(\lambda_i) |\phi_i\rangle\langle\phi_i|.$$

For an arbitrary function $f : \mathbb{C} \to \mathbb{C}$, one can thus consistently *define* its action on operators with an eigendecomposition as

$$f(A) := \sum_i f(\lambda_i) |\phi_i\rangle\langle\phi_i|.$$

(This convention is sometimes called the *spectral calculus*).

2.9 Unitary operators

Unitary operators are the Hilbert space analogue of orthogonal rotations in Euclidean vector spaces: Invertible linear operators that preserve inner products. Let's work out what that means.

The inner product between $U|\phi\rangle, U|\psi\rangle$ is $\langle\phi|U^\dagger U|\psi\rangle$. Thus U preserves the inner product between any pair of operators if and only if

$$\langle\phi|U^\dagger U|\psi\rangle = \langle\phi|\psi\rangle \quad \forall\, \phi, \psi \in \mathcal{H}.$$

We thus define: An operator is *unitary* if it is invertible and fulfills $U^\dagger U = \mathbb{1}$.

One can work out that these characterizations are equivalent:

1. U is unitary.
2. U has a spectral decomposition of the form
$$U = \sum_i e^{i\phi_i} |\psi_i\rangle\langle\psi_i|, \qquad \phi_i \in \mathbb{R},$$
i.e. all eigenvalues $\lambda_i = e^{i\phi_i}$ have absolute value equal to 1.
3. There is a Hermitian operator A such that $U = e^{iA}$ (in the sense of Sec. 2.8).
4. U is such that $U^\dagger U = \mathbb{1}$ *and* $UU^\dagger = \mathbb{1}$ (in which case U is automatically invertible, so we do not have to list this as an extra requirement).
5. If $\{|e_i\rangle\}_i$ is an ONB, then so is $\{U|e_i\rangle\}_i$.

In quantum mechanics, unitary operators describe symmetries. The most important symmetry is of course time evolution! The Hermitian operator that generates time evolution $U(t)$ in the sense that $U(t) = e^{-it/\hbar H}$ (as in Point 3.) is nothing but $-1/\hbar$ times the Hamiltonian.

2.10 Projections

Recall (see Fig. 2.1) that in $\mathbb{R}^d$ with Euclidean scalar product $(u, v) = \sum_i u_i v_i$, there is a one-one relation between

- Subspaces $V \subset \mathbb{R}^d$, and
- *orthogonal projections* P, i.e. linear maps fulfilling $P = P^t$, $P^2 = P$.

The Hilbert space analogue works like this: An operator P is a *projector* (or *projection*) if

1. $P = P^\dagger$, and
2. $P^2 = P$.

The first property means that P has a spectral decomposition. The second property then implies that the eigenvalues are elements of $\{0, 1\}$. Thus,

$$P = \sum_i |\psi_i\rangle\langle\psi_i|,$$

where the $\{|\psi_i\rangle\}$ form an ONB for the subspace $V \subset \mathcal{H}$ onto which P projects.

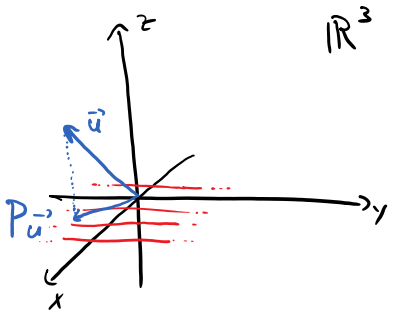


Figure 2.1: Orthogonal projection of $\boldsymbol{u}$ onto the x-y-plane.

Examples.

- For every normalized vector $|\psi\rangle \in \mathcal{H}$, the outer product $P = |\psi\rangle\langle\psi|$ is the projection onto the one-dimensional subspace $V = \{z|\psi\rangle \,|\, z \in \mathbb{C}\}$.
- Define the *parity operator* Π on $\mathcal{H} = L^2(\mathbb{R})$ by

$$\Pi|\delta_x\rangle = |\delta_{-x}\rangle, \quad \text{that is} \quad (\Pi\phi)(x) = \phi(-x).$$

  Then it's easy to see that $P_\pm = \frac{1}{2}(\mathbb{1} + \Pi)$ are projection operators onto the space of even and odd functions respectively (why?).

2.11 The trace

The *trace* of an operator is the sum over its eigenvalues. It can be expressed as

$$\operatorname{tr} A = \sum_i \langle i|A|i\rangle,$$

where the sum is over *any* ONB $\{|i\rangle\}_i$.

Some properties:

- Cyclic invariance:

$$\operatorname{tr} AB = \sum_{ij} \langle i|A|j\rangle\langle j|B|i\rangle = \sum_{ij} \langle j|B|i\rangle\langle i|A|j\rangle = \operatorname{tr} BA.$$

- Trace of outer products are inner products:

$$\operatorname{tr} |\alpha\rangle\langle\beta| = \sum_i \langle i|\alpha\rangle\langle\beta|i\rangle = \sum_i \langle\beta|i\rangle\langle i|\alpha\rangle = \langle\alpha|\beta\rangle.$$

2.12 Commuting operators

Assume that two operators A, B have a *joint eigenbasis* $\{|\psi_i\rangle\}$:

$$A|\psi_i\rangle = a_i|\psi_i\rangle, \qquad B|\psi_i\rangle = b_i|\psi_i\rangle.$$

Then

$$[A, B]|\psi_i\rangle = (AB - BA)|\psi_i\rangle = (a_i b_i - b_i a_i)|\psi_i\rangle = 0 \qquad \forall\, i.$$

so the operators commute.

Less obvious, but still true is that the converse also holds: If a set of normal operators commute, one can construct a joint eigenbasis. (The proof is very simple if one of the operators is non-degenerate, but somewhat lengthy in the general case. We skip it here).

Example. Consider the momentum operator $P = -i\hbar\partial_x$, the free Hamiltonian $H = P^2/2m$, and the parity operator $(\Pi\psi)(x) = \psi(-x)$. Then

1. $[H, P] = 0$ (because any two functions of the same Hermitian operator commute),
2. $[P, \Pi] \neq 0$, because by the chain rule, computing the derivative of the reflected function incurs an extra minus sign: $(P\Pi\psi)(x) = -P\psi(-x)$ while $(\Pi P\psi)(x) = P\psi(-x)$.
3. $[H, \Pi] = 0$, because H depends on the 2nd derivative, and the two signs cancel.

By 1., we can find an energy eigenbasis consisting of states with well-defined momentum. These are the plane waves $\phi_E^{\pm}(x)$ of Eq. (1.18). By 3., there is an energy eigenbasis of states with well-defined parity. These are sine / cosine functions $\phi_E^{(e)}(x), \phi_E^{(o)}(x)$ of (1.19). By 2., we cannot insists on both properties at the same time.

While the eigenspaces of H are two-fold degenerate, the joint eigenbases of H, P and, respectively, of H, Π are unique (up to re-ordering and changing phases).

In the example, the eigenspaces of H and of Π are both degenerate. But the *joint eigenvalue equations*

$$H|E, \pi\rangle = E|E, \pi\rangle, \quad \Pi|E, \pi\rangle = \pi|E, \pi\rangle, \qquad E \geq 0, \pi \in \{\pm 1\}$$

have a unique solution $|E, \pi\rangle$ (up to physically irrelevant phases). In general, a collection of operators with a unique joint eigenbasis is called a *complete set of commuting operators*. The eigenvalues that single out an element of the basis (like E, π in the example) are sometimes called *quantum numbers*.

2.13 Modeling reality, the quantum way

Even though it's common to talk about "the axioms of quantum mechanics", the truth is that "quantum theory" is more of a general framework. Here's the maybe minimal requirement a theory has to fulfill to count as "quantum".

- It takes as input a classical description of a physical setup. "Mount the dilution fridge in the rack, tune the lasers to 400nm, record whether the photo detector clicks."
- The procedure is divided into a "preparation" and a "measurement" part.
- To each preparation procedure, associate a *state*, represented by a vector $|\psi\rangle$ in a Hilbert space. To each measurement procedure, associate an *observable*, represented by a Hermitian operator F.
- The expected value of the measurement outcome is then given by $\langle\psi|F|\psi\rangle$.

TBD: Figure
Comments:

- There's generally a lot of freedom for how to divide up a setup into preparation and measurement. See the discussion of Schrödinger vs. Heisenberg picture in Sec. 1.5.3. (A particularly radical case is commonly used in relativistic quantum field theory, where there's a default preparation procedure essentially saying "don't

do anything at all!", which results in the *quantum vacuum state*. Several books have been written about it!). However, if it's quantum, there has to be a Hilbert space somewhere in the middle!

- So what *is* a "quantum state"? It's natural to define a quantum state as a *classical description of a preparation procedure*, since this is what completely determines the observable behavior. (Or, more pedantically, as an *equivalence class of preparations*, where two procedures are identified if they lead to statistically indistinguishable measurement outcomes). Likewise, an observable *is* an (equivalence class of) classical prescription of a measurement process.

- Q.: *You have defined the more fundamental concept "a quantum state" in terms of the less fundamental "classical description". Shouldn't classical physics just arise as a limit of quantum? Like, in the same way as Galilean space-time is now recognized as a mere approximation to the more fundamental relativistic space-time?*
  Yes, there some tension here! The relation between classical and quantum concepts is a major problem in metaphysics, sometimes discussed as the *measurement problem*.

- Q.: *So "physical reality factors through a Hilbert space". I mean, it's hard to come up with a less intuitive starting point! Can one explain why?*
  A.: Touchy subject.
  Here's what could have been. Special relativity has also fundamentally changed our understanding of reality (e.g. showing that seemingly innocuous concepts like "at the same time" often don't make sense, ...). But: You can start with a simple-to-understand empirical observation – "the speed of light is the same in every reference frame" – and given enough scratch paper and ingenuity, reason your way to SR just from this one physical axiom.
  Trying to emulate this, hundreds of papers have been written in an attempt to find *physically meaningful* axioms from which the Hilbert space structure of QM could be derived. There is currently no consensus on whether this has been achieved to a satisfactory degree, and if so, which axioms to start from.
  For the time being, you may just have to accept it. ¯_(ツ)_/¯

Chapter 3

Two-body quantum mechanics

3.1 Hilbert spaces for composite systems

Consider two quantum mechanical systems with individual Hilbert spaces

- $\mathcal{H}_1 = \{\sum_i \alpha_i |e_i\rangle\}$, with $\langle e_i|e_k\rangle = \delta_{i,k}$.
- $\mathcal{H}_2 = \{\sum_i \beta_i |e_i\rangle\}$, with $\langle f_j|f_l\rangle = \delta_{j,l}$.

Our goal is to assign a Hilbert space $\mathcal{H}_{1,2}$ to the joint system.

If the two particles are distinguishable, then it makes sense to prepare to first one in the state $|e_i\rangle$ and the second one in the state $|f_j\rangle$. There must therefore be an element in the joint Hilbert space $\mathcal{H}_{1,2}$ that describes this state. Write it as

$$|e_i, f_j\rangle \qquad \text{“system 1 in state } |e_i\rangle\text{, system 2 in state } |f_j\rangle\text{”}. \tag{3.1}$$

Being a vector space, $\mathcal{H}_{1,2}$ must also contain any linear combination

$$\sum_{i,j} \psi_{i,j} |e_i, f_j\rangle, \qquad \psi_{i,j} \in \mathbb{C}$$

of such elements. This suggests the following construction:

The *tensor product Hilbert space* $\mathcal{H}_{1,2} = \mathcal{H}_1 \otimes \mathcal{H}_2$ is the vector space

$$\mathcal{H}_{1,2} = \Big\{ \sum_{i,j} \psi_{i,j} |e_i, f_j\rangle \Big\}, \quad \text{with} \quad \langle e_i f_j | e_k f_l\rangle = \delta_{i,k}\delta_{j,l}.$$

3.1.1 Product states, entanglement

Generalizing (3.1), if $|\alpha\rangle = \sum_i \alpha_i |e_i\rangle \in \mathcal{H}_1$ is a general state of the first system, and $|\beta\rangle = \sum_j \beta |f_j\rangle \in \mathcal{H}_j$, one of the second system, associate with it the *product state*

$$|\alpha, \beta\rangle = \sum_{i,j} \alpha_i \beta_j |e_i, f_j\rangle \in \mathcal{H}_{1,2}.$$

It is called so, because the coefficients $\psi_{i,j} = \alpha_i \beta_j$ factorize. Alternative notations:

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

If the basis one is referring to is clear from context, one also writes $|e_i, f_j\rangle = |i, j\rangle$.

It is easy to verify that the tensor product between product vectors factorizes:

$$\langle\alpha\beta|\gamma\delta\rangle = \sum_{ijkl} \alpha_i^* \beta_j^* \gamma_k \delta_l \underbrace{\langle ij|kl\rangle}_{\delta_{ik}\delta_{jl}} = \sum_{ij} \alpha_i^* \beta_j^* \gamma_i \delta_j = \langle\alpha|\gamma\rangle \langle\beta|\delta\rangle. \tag{3.2}$$

A simple parameter counting argument shows that "most" vectors in a tensor product space do not factorize. Such vectors are called *entangled*, and give rise to very exciting phenomena, as we'll see soon.

Remark. Interpreting the expansion coefficients ψ_{ij} as a matrix, it becomes apparent that there are numbers α_i, β_j such that $\psi_{ij} = \alpha_i \beta_j$ if and only if ψ_{ij} has rank 1.

As an example, consider Hilbert spaces $\mathcal{H}_1 = \mathcal{H}_2$ with ONB $\{|\uparrow\rangle, |\downarrow\rangle\}$. Then

$$|\psi\rangle = \frac{1}{2}\big(|\uparrow\uparrow\rangle + |\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle + |\downarrow\downarrow\rangle\big)$$

is not entangled, because the coefficient matrix

$$\psi_{ij} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}_{ij}$$

has rank 1. And indeed, $|\psi\rangle = |\alpha\rangle|\beta\rangle$ for $|\alpha\rangle = |\beta\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle)$. On the other hand, both

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}\big(|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle\big) \quad \text{and} \quad |\Psi^-\rangle = \frac{1}{\sqrt{2}}\big(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle\big) \tag{3.3}$$

are entangled, because

$$\operatorname{rank}\begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{pmatrix} = \operatorname{rank}\begin{pmatrix} 0 & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & 0 \end{pmatrix} = 2.$$

Remark. The tensor product is not the only way to combine local to combine individual Hilbert spaces to a joint one.

Indeed, call two particles *indistinguishable* if one cannot construct measurement devices that are sensitive to one of them, but not to the other. Elementary particles of the same type are indistinguishable: There is no detector that will be triggered only by one specific photon!

For indistinguishable particles, the tensor product Hilbert space is "too large". Indeed, in $\mathcal{H}_1 \otimes \mathcal{H}_2$, the vector $|\alpha\rangle|\beta\rangle$ generally describes a different state than $|\beta\rangle|\alpha\rangle$, even though no measurement device can tell them apart. This redundancy is already present in classical mechanics, but doesn't seem to cause major trouble there. In contrast, in QM, in turns out that one must switch to a "reduced", non-redundant, Hilbert space to get correct predictions. It appears that elementary particles fall in one of two categories, which employ different mechanisms for building a joint Hilbert space from individual ones. These are called *Bosons* and *Fermions* respectively. The details are not part of this class.

3.1.2 Local operators

Let $\mathcal{H}_{1,2} = \mathcal{H}_1 \otimes \mathcal{H}_2$. We can let operators A on $\mathcal{H}_1$, or B on $\mathcal{H}_2$ act on product vectors in the obvious way:

$$A|\alpha, \beta\rangle = (A|\alpha\rangle)|\beta\rangle, \qquad B|\alpha, \beta\rangle = |\alpha\rangle(B|\beta\rangle).$$

This defines A, B on all of $\mathcal{H}_{12}$, because the product vectors $|e_i, f_j\rangle$ form a basis. If not clear from context, the system on which an operator acts is explicitly specified

$$C^{(1)}|\alpha,\beta\rangle = (C|\alpha\rangle)|\beta\rangle, \qquad C^{(2)}|\alpha,\beta\rangle = |\alpha\rangle(C|\beta\rangle).$$

Examples. Consider again $\mathcal{H}_1 = \mathcal{H}_2$ with ONB $\{|\uparrow\rangle, |\downarrow\rangle\}$. With respect to this ONB, let

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

be the operator that exchanges "up" and "down" states. Then $\sigma_x^{(1)}$ acts on the product basis as

$$|\uparrow\uparrow\rangle \mapsto |\downarrow\uparrow\rangle, \quad |\uparrow\downarrow\rangle \mapsto |\downarrow\downarrow\rangle, \quad |\downarrow\uparrow\rangle \mapsto |\uparrow\uparrow\rangle, \quad |\downarrow\downarrow\rangle \mapsto |\uparrow\downarrow\rangle,$$

or, in matrix form w.r.t. that basis,

$$\begin{array}{c} \\ |\uparrow\uparrow\rangle \\ |\uparrow\downarrow\rangle \\ |\downarrow\uparrow\rangle \\ |\downarrow\downarrow\rangle \end{array} \begin{array}{c} \begin{array}{cccc} |\uparrow\uparrow\rangle & |\uparrow\downarrow\rangle & |\downarrow\uparrow\rangle & |\downarrow\downarrow\rangle \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array} .$$

In general, the matrix elements of the product of local operators are the product of the local matrix elements:

$$\langle i,j|A^{(1)}B^{(2)}|k,l\rangle = \langle i|A|k\rangle\,\langle j|B|l\rangle. \tag{3.4}$$

You'll also come across the "tensor product of operators" notation (sometimes called the *Kronecker product*, in particular in computer algebra systems):

$$A \otimes B := A^{(1)}B^{(2)} \quad \Rightarrow \quad A \otimes \mathbb{1} = A^{(1)}, \quad \mathbb{1} \otimes B = B^{(2)}.$$

Local operators acting on different subsystems commute:

$$A^{(1)}B^{(2)}|\alpha\beta\rangle = (A|\alpha\rangle)(B|\beta\rangle) = B^{(2)}A^{(1)}|\alpha\beta\rangle. \tag{3.5}$$

Local observables

Now consider the situation that we have prepared a two-particle system in a product state $|\psi\rangle = |\alpha\rangle|\beta\rangle$, measure the observable A on the first system, and disregard the second one. The expectation value then better be equal to $\langle\alpha|A|\alpha\rangle$, for "preparing a second particle somewhere else and then ignoring it" shouldn't affect the behavior of the first system. And indeed, using (3.2),

$$\langle\psi|A^{(1)}|\psi\rangle = \langle\alpha|\langle\beta|(A|\alpha\rangle)|\beta\rangle = \langle\alpha|A|\alpha\rangle\,\langle\beta|\beta\rangle = \langle\alpha|A|\alpha\rangle. \tag{3.6}$$

More generally,

$$\langle\alpha\beta|\,A^{(1)}B^{(2)})\,|\alpha\beta\rangle = \langle\alpha|A|\alpha\rangle\,\langle\beta|B|\beta\rangle, \tag{3.7}$$

hence the observable $A^{(1)}B^{(2)}$ describes the process of "measuring A on the first system, B on the second system, and multiplying the results".

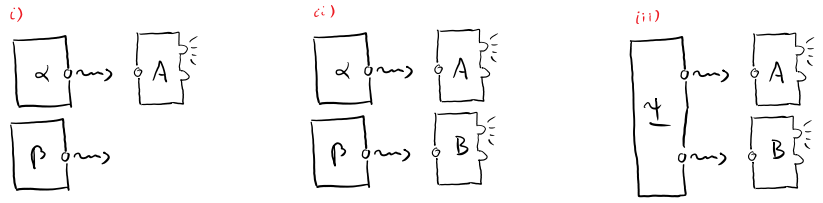


Figure 3.1: Local measurements on two-body quantum states. (i) Assume that particle one has been prepared in state $|\alpha\rangle$ and an observable A has been measured on it. Then the result should be independent of what happened to an unrelated particle. (Verified in Eq. (3.6)). (ii) More generally, if two particles are prepared separately and measured separately, the results should be independent. (Verified in Eq. (3.7)). (iii) Entangled states of two systems can be prepared only in a process involving them both. The results of local measurements on such a state can be correlated. (Verified in Eq. (3.8)).

We can now have a first look at the properties of entangled states. Continuing the example around Eq. (3.3), and using the observable

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

which assigns $+1$ to $|\uparrow\rangle$ and -1 to $|\downarrow\rangle$, compute

$$\begin{aligned}
\langle\Phi^+|\sigma_z^{(1)}|\Phi^+\rangle &= \frac{1}{2}\langle\uparrow\uparrow|\underbrace{\sigma_z^{(1)}|\uparrow\uparrow\rangle}_{+|\uparrow\rangle} + \frac{1}{2}\langle\uparrow\uparrow|\underbrace{\sigma_z^{(1)}|\downarrow\downarrow\rangle}_{-|\downarrow\rangle} = \frac{1}{2} - \frac{1}{2} = 0, \\
\langle\Phi^+|\sigma_z^{(2)}|\Phi^+\rangle &= \ldots = 0, \\
\langle\Phi^+|\sigma_z^{(1)}\sigma_z^{(2)}|\Phi^+\rangle &= \frac{1}{2}\langle\uparrow\uparrow|\underbrace{\sigma_z^{(1)}\sigma_z^{(2)}|\uparrow\uparrow\rangle}_{+|\uparrow\uparrow\rangle} + \frac{1}{2}\langle\uparrow\uparrow|\underbrace{\sigma_z^{(1)}\sigma_z^{(2)}|\downarrow\downarrow\rangle}_{(-1)^2|\downarrow\downarrow\rangle=+|\downarrow\downarrow\rangle} = \frac{1}{2} + \frac{1}{2} = 1.
\end{aligned} \tag{3.8}$$

Interpretation: The first two equations imply that, when measuring σ_z on either particle, the outcomes ± 1 must occur with equal probability, so that the expected value is 0. The final equation says that when measuring σ_z on both particles at the same time, the product of the results is always $+1$. Therefore, the results $+1, +1$ and $-1, -1$ occur with probability $\frac{1}{2}$ each. In other words, the outcomes of the σ_z-measurements are strictly correlated.

Thus, Eq. (3.7) says that product measurements on product states lead to independent outcomes, while product measurements on entangled states can show correlations.

On quanta and socks. In popular science expositions, it is sadly common that the correlations above are presented as something deeply mysterious. "Measuring the first particle immediately tells you about the result a measurement on the second particle would give. Even if they are far apart. Spooky action-at-a-distance!"

This argument is completely wrong. Correlations between two random events are commonplace. Whether or not it will rain on any given day is hard to predict, but if it rains on my place, it will also rain on my neighbor's place. (In the literature, the term Bertlemann's socks has come to be associated with the confused tendency of thinking that correlations are mysterious when they appear in quantum measurements, while unremarkable in everyday objects (like socks)).

However, the correlations that can occur when measuring entangled sates *are actually deeply mysterious*! Seeing this takes a little bit more work. See the section below.

3.2 Bell inequalities

Classical mechanics tells you what is happening. Quantum mechanics only tells you what you will observe when you measure. It does not assign values to unmeasured physical properties.

EINSTEIN ATTACKS QUANTUM THEORY

Scientist and Two Colleagues Find It Is Not 'Complete' Even Though 'Correct.'

SEE FULLER ONE POSSIBLE

Believe a Whole Description of 'the Physical Reality' Can Be Provided Eventually.

The New York Times

Sorry, Einstein. Quantum Study Suggests 'Spooky Action' Is Real.

Figure 3.2: Left panel: 1935 New York Times headline reporting on Einstein-Podolsky-Rosen paper arguing that quantum mechanics was incomplete. I wonder how Podolsky and Rosen felt about the framing. Right panel: 2015 New York Times headline reporting on Einstein being wrong.

From the early days of the theory, some scientists – famously Albert Einstein (Fig. 3.2 – saw this as a sign that quantum mechanics was *incomplete*, and should be supplanted by a more detailed description of Nature that does track the time evolution of all physical properties, measured or not.

In what I feel is one of the most profound findings of modern physics, this program has since been proven to be impossible: The hypothesis

"Physical properties exist independently of measurements" (3.9)

has been *experimentally falsified* as a general property of Nature! On top of the surprising conclusion, this is remarkable because (3.9) feels like a philosophical statement that is too vague to have testable implications. Yet here we are.

In the following derivation, we have to keep in mind that we want to reason about theories different from quantum mechanics. This means that we cannot use any concept that has a meaning only in the context of QM. "Hilbert space", "entanglement", "commutators", even "photon"... ...all these terms are verboten until further notice.[1]

3.2.1 The CHSH scenario

Our challenge now is to come up with a setting in which the vague statement (3.9) leads to quantitative predictions that can be compared to experiments. The most important case is the so-called *CHSH scenario* (Fig. 3.3). While not difficult to understand, it does contain quite a number of elements that seem ill-motivated at this point. Please bear with me for a moment.

[1] Physicists talking about Bell inequalities have a tendency of emphasizing entanglement, or the singlet state and how the fact that it's spin-0 means that angular momentum measurements are anti-correlated, and some such things. These are not wrong and even mildly helpful for the design of experiments that lead to the falsification we are after. All this is also completely secondary to the main point; a case of people sticking to their comfort zone.

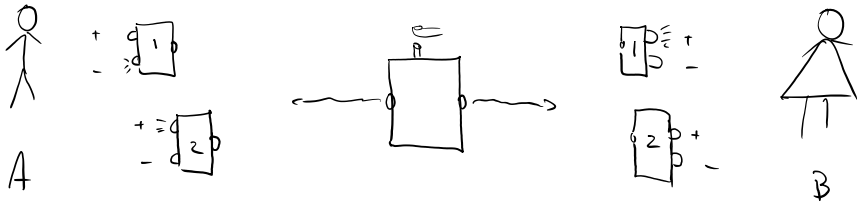


Figure 3.3: The ingredients of the CHSH scenario (for Clauser, Horne, Shimony and Holt). Two experimentalists are located at different ends of a laboratory. Each can perform one of two measurements on systems emanating from a box in the middle. Surprisingly, the analysis of the set of correlations that are compatible with this extremely vaguely defined scenario offers profound insights!

The scenario contains two observers, Alice and Bob, located at different ends of a laboratory. There's a box in the middle. In regular intervals, it emits two systems, one flying to Alice and one to Bob. Each observer has two measurement devices, labeled 1 and 2. The devices work like this: They have an entry port and when one of the systems coming from the central box enters a device, one of two lights will flash. The lights are labeled $+1$ and -1 respectively. Every time a pair of systems leaves the central box, Alice and Bob choose one of their measurement devices at random, put it in the path, and record the observed outcomes.

OK, some Q&A's:

- Q.: *So what's up with the talk of "systems"? What are these? Photons? Spins?*
  A.: Unspecified. For now, these could be puffs of hot air and the measurement devices random number generators. Our analysis does not depend on assumptions about their nature. (Also, what's a photon?)

- Q.: *Are Alice's devices* 1 *and* 2 *different? Is Alice's device 1 different from Bob's device 1?*
  A.: We do not need to make any assumptions about this.

- Q.: *Why are the outcomes labeled* ± 1*?*
  A.: That's not really essential. This particular choice will work well with our analysis, though.

- Q.: *Can Alice rig her boxes together such that she can perform both measurement on the same incoming system?*
  A.: For all we know at this point... maybe?

- Q.: *Look man. You are clearly just avoiding my questions. Why don't you study your system first, and come back once you can give specific answers?!*
  A.: You got it backwards! The fewer assumptions I need to make, the more generally applicable my conclusions will be.[2]

- Q.: *How in the world does one come up with this?*
  A.: Well, it took physics a few decades. Also, literal Einstein missed it.

[2] I once had a long discussion with colleague who refused to conceit this point, despite me applying all the logic, persuasion, and appeals to authority I could muster. Very frustrating.

With the setup established, let's look at the lab book produced by A&B. Here's a possible snapshot:

	Alice		**Bob**	
i	A_1	A_2	B_1	B_2
1	$+$		$-$	
2	$+$			$+$
3	$-$			$-$
4		$+$	$+$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Obviously, in each round i, both Alice and Bob can fill out only the column corresponding to the measurement they chose to make.

We will now argue that Assumption (3.9) puts quantitative constraints on the type of data that can appear in this setting. Later, we will see that there are experiments that violate these constraints—thereby disproving the general validity of (3.9). (Also, QM predicts the violations correctly. That's also interesting, but less relevant).

Concretely, if physical properties exist independently of observations, then there exits a complete table, say

	Alice		**Bob**	
i	A_1	A_2	B_1	B_2
1	$+$	$-$	$-$	$-$
2	$+$	$-$	$+$	$+$
3	$-$	$-$	$+$	$-$
4	$+$	$+$	$+$	$-$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

and in each round, A&B just decide which of the pre-existing values to uncover.

In what may feel like an unmotivated move even by the standards of the present discussion, associate the expression

$$C = A_1B_1 + A_1B_2 + A_2B_1 - A_2B_2$$

which each complete row. There's an elegant geometric construction that leads to this particular formula (the keyword is *Bell polytope*) – but it takes some time to develop, so let's just work with C regardless of where it comes from. In our example:

	Alice		**Bob**		
i	A_1	A_2	B_1	B_2	C
1	$+$	$-$	$-$	$-$	-2
2	$+$	$-$	$+$	$+$	2
3	$-$	$-$	$+$	$-$	2
4	$+$	$+$	$+$	$-$	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Despite being the sum of four terms each valued ± 1, the expression (in fact: its absolute value) is upper-bounded by 2: Factoring out Alice's variables and applying the triangle inequality,

$$|C| = |A_1(B_1 + B_2) + A_2(B_1 - B_2)| \leq |B_1 + B_2| + |B_1 - B_2| = 2.$$

It may seem that we can't extract observable predictions out of this discussion, because the expression C involves all four variables, and by assumption, we only have access to two of them in each round. But there's a nice trick to get around this! Indeed, if $C \leq 2$ in every run, then so is the *average*

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C^{(i)}$$

over N runs. But averages are linear, and therefore $\langle C \rangle$ equals

$$\langle A_1 B_1 + A_1 B_2 + A_2 B_1 - A_2 B_2 \rangle = \langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle.$$

Each of the four terms $\langle A_i B_j \rangle$ *can* be estimated by A&B! If they choose their settings at random, then by the law of large numbers (or, quantitatively, by the *Chernoff bound*), their observed mean will converge to the true expected value in the limit of large N. Thus, Assumption (3.9) implies that the linear combination of these four experimentally accessible numbers is no larger than 2, up to statistical fluctuations that vanish in the large-N limit. Such a test of (3.9) is called a *Bell inequality*.

Following up on pioneering works that led to the 2022 Nobel Prize, it is today fairly routine to perform experiments that are compatible with the CHSH setup and yield a value of $\langle C \rangle \simeq 2.7$.

Thus, Assumption (3.9) must be rejected as a general feature of Nature.

3.2.2 Quantum violations of the CHSH inequality

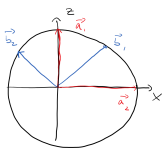
We'll let Alice and Bob perform local measurements on the entangled state

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle).$$

Exercise Let $\boldsymbol{a} = \begin{pmatrix} a_x \\ a_z \end{pmatrix}$ be a vector in the x-z-plane. Assume that it is normalized, $\|\boldsymbol{a}\| = 1$.

- Show that $\hat{a} := a_x \sigma_x + a_z \sigma_z$ is an observable with eigenvalues ± 1.
- Let $\boldsymbol{b}$ be another normalized vector in the x-z-plane. Show that

$$\langle \Phi^+ | \hat{a}^{(1)} \hat{b}^{(2)} | \Phi^+ \rangle = (\boldsymbol{a}, \boldsymbol{b}). \tag{3.10}$$



- Now choose the vectors $\boldsymbol{a}_i, \boldsymbol{b}_i$ according to the diagram to the left (i.e. at angles 0 or ± 45 degrees with respect to the coordinate axes). Show

$$\langle \Phi^+ | \hat{a}_1^{(1)} \hat{b}_1^{(2)} + \hat{a}_1^{(1)} \hat{b}_2^{(2)} + \hat{a}_2^{(1)} \hat{b}_1^{(2)} - \hat{a}_2^{(1)} \hat{b}_2^{(2)} | \Phi^+ \rangle = 2\sqrt{2}.$$

- Interpret the result in the light of the CHSH experiment.

Chapter 4

Symmetries in quantum mechanics

Popcorn time! In this chapter, we discuss *why* some things are the way they are in QM. We won't really talk about *how to compute stuff* (which is what is tested in the written exam). So if there's some detail you don't fully grok – stop worrying and help yourself to some more popcorn!

Picture credit.

4.1 Introduction

Here are some questions you might have:

- *OK, I have gotten used to the idea that mathematical models of reality involve Hilbert spaces and operators. But, when constructing a quantum theory, how do I know* which *vector / operator to assign to a physical setup?*
- *Specifically, what does momentum have to do with i times a spatial derivative??*
- *Also, what's up with all these random*[1] *"$i\hbar$"s floating around QM???*
- Typical textbook quote: "A quantum mechanical angular momentum $\boldsymbol{J}$ is a set of operators J_x, J_y, J_z such that $[J_x, J_y] = i\hbar J_z$...". *Are you trolling me? Who even* comes up *with stuff as bizarre as that????*

It turns out that one can make progress on these questions by working out the conditions a QM model has to fulfill in order to be compatible with this empirical observation:

> The behavior of physical experiments does not depend on when, where, or in which orientation they are being performed.

Outline (spoiler alert!)

The arguments presented here aren't really complicated, but quite lengthy. I give a short outline here, so you are less likely to get lost along the way.

- [Generators]: The operation of "rotating a quantum system" about the z axis by an angle θ is associated with a unitary $U(\theta)$, so that for any state $|\psi\rangle$, the displaced state is $|\psi'\rangle = U(a)|\psi\rangle$. The first-order Taylor approximation is $U(\theta) = \mathbb{1} + \theta l_3$, where $l_3 = U'(0)$ is called the *generator* of z-rotations.

[1] *random*. German adjective. Means "unmotivated, unexpected". Use liberally. Source: Generation Z.

The study of generators is useful for two reasons:

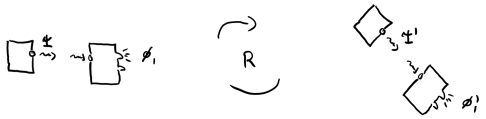
- [Finding Hilbert spaces]: It follows from properties of the rotation group that the generators l_1, l_2, l_3 of rotations about the coordinate axes have to fulfill $[l_1, l_2] = l_3$ (and cyclic permutations). That's useful because one can enumerate all solutions to this equation! The list includes some old friends, (like solutions on $L^2(\mathbb{R}^3)$), but also new possibilities, notably spin-$1/2$ systems.
- [Identifying observables]: It turns out that if g is a generator, then ig is Hermitian. Typically, the observable ig is proportional to a conserved quantity (time to recall Noether's Theorem!). The constant of proportionality is not fixed by theoretical arguments, but has to be experimentally determined. In any case, it's called $\hbar$.

We'll now walk through the arguments in more detail, mainly focusing on rotations.

4.2 Wigner's Theorem

Consider a preparation procedure that produces a quantum system in state $|\psi\rangle$, followed by a measurement of an observable F. Let λ be possible outcome of F, associated with an eigenvector $|\phi_\lambda\rangle$, so that the probability of observing λ is $|\langle\psi|\phi_\lambda\rangle|^2$.

Now compare it to a setup that is identical to the first, except that every part has been displaced by some space-time symmetry. Call the displaced state $|\psi'\rangle$ and the λ-eigenvector of the displaced observable $|\phi'_\lambda\rangle$.



By invariance, the physical behavior of the two setups must be the same, i.e.

$$|\langle\psi|\phi_\lambda\rangle|^2 = |\langle\psi'|\phi'_\lambda\rangle|^2. \tag{4.1}$$

Which types of transformations $|\psi\rangle \mapsto |\psi'\rangle$, $|\phi_\lambda\rangle \mapsto |\phi'_\lambda\rangle$ are compatible with (4.1)? Two possibilities are easy to guess:

- **Unitaries:** Choose a unitary U and set

  $$|\psi'\rangle = U|\psi\rangle, \qquad |\phi'_\lambda\rangle = U|\phi_\lambda\rangle.$$

  Then $\langle\psi'|\phi'_\lambda\rangle = \langle U\psi|U\phi_\lambda\rangle = \langle\psi|\phi_\lambda\rangle$, and hence (4.1) holds.

- **Anti-unitaries:** Choose an ONB $\{|e_i\rangle\}_i$. Define "complex conjugation of coefficients with respect to that basis" as the function

  $$C : |\psi\rangle = \sum_i \psi_i|e_i\rangle \mapsto \sum_i \psi_i^*|e_i\rangle.$$

  Then

  $$|\psi'\rangle = C|\psi\rangle, \quad |\phi'_\lambda\rangle = C|\phi_\lambda\rangle \qquad \Rightarrow \qquad \langle\psi'|\phi'_\lambda\rangle = \langle\psi|\phi_\lambda\rangle^*,$$

  so we found another solution to (4.1). The combination UC of complex conjugation followed by a unitary also does the job. Such maps are called *anti-unitaries*.

A lengthy argument known as *Wigner's Theorem* shows that unitaries and anti-unitaries are the *only* possibilities to satisfy (4.1).

Wigner: For every symmetry S of a quantum system, there is a unitary or anti-unitary $U(S)$ such that state vectors and eigenvectors of observables transform as

$$|\psi'\rangle = U(S)|\psi\rangle, \qquad |\phi'_\lambda\rangle = U(S)|\phi_\lambda\rangle.$$

The most relevant symmetries in physics can be continuously deformed to the trivial transformation. In particular, this is true for rotations and translations – just let the angle / distance go to 0. Such symmetries are always unitarily represented.

Proof. Let L be linear and A anti-linear, both invertible. Then one checks that

$$L(i\mathbb{1})L^{-1} = i\mathbb{1} \quad \text{but} \quad A(i\mathbb{1})A^{-1} = -i\mathbb{1}. \tag{4.2}$$

If a family $U(s)$ of solutions to (4.1) depends *continuously* on a parameter s, then all $U(s)$ are unitary or all are anti-unitary, because a continuous map that takes values in $\{\pm i\mathbb{1}\}$ is constant.

There *are* symmetries not connected to $\mathbb{1}$, and these may be associated with anti-unitaries. We will not consider such symmetries in the notes, though one appears in the...

...homework. Show that in wave mechanics, *time reversal symmetry*, $t \mapsto -t$, is implemented by complex conjugation $\psi(x) \mapsto \psi(x)^*$ in the position basis.

Example. Let's see how wave functions in $L^2(\mathbb{R}^n)$ transform under translations. For $\boldsymbol{a} \in \mathbb{R}^n$, let $T_{\boldsymbol{a}}$ denote the function $\boldsymbol{x} \mapsto \boldsymbol{x} + \boldsymbol{a}$ that translates points by $\boldsymbol{a}$. Its action on wave functions is

$$(U(T_{\boldsymbol{a}})\psi)(\boldsymbol{x}) = \psi(\boldsymbol{x} - \boldsymbol{a}). \tag{4.3}$$

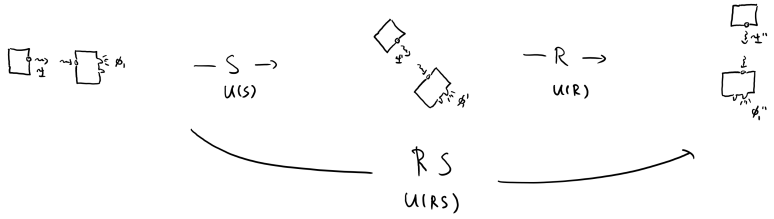
(Mind the minus! The value of the translated function at $\boldsymbol{x}$ equals the value of the original at $\boldsymbol{x} - \boldsymbol{a}$). The operator $U(T_{\boldsymbol{a}})$ is unitary because shifting the integrand does not change integrals.

4.2.1 Projective representations

If R, S are two symmetries, then "R after S", denoted by RS, is another symmetry. Mathematically, the set of symmetries forms a *group*. Therefore, $U(R)U(S)$ must implement the same operation as $U(RS)$ on quantum states. Because two state vectors give rise to the same predictions iff they differ by a phase factor, we get the compatibility condition

$$U(R)U(S) = e^{i\phi(R,S)}U(RS) \tag{4.4}$$

for some phases $\phi(R, S)$.



Let G be a symmetry group. An assignment $S \mapsto U(S)$ of a unitary to each $S \in G$, that satisfies (4.4) is called a *projective unitary representation* of the symmetry group. If the phases all vanish, i.e. if

$$U(R)U(S) = U(RS), \tag{4.5}$$

then U called a *linear unitary representation*, or just a *unitary representation*.

Example. The action of translations on wave functions is indeed a unitary representation:

$$\big(U(T_{\boldsymbol{a}})U(T_{\boldsymbol{b}})\psi\big)(\boldsymbol{x}) = \big(U(T_{\boldsymbol{b}})\psi\big)(\boldsymbol{x}-\boldsymbol{a}) = \psi(\boldsymbol{x}-\boldsymbol{a}-\boldsymbol{b}) = \big(U(T_{\boldsymbol{a}+\boldsymbol{b}})\psi\big)(\boldsymbol{x}).$$

Every continuous symmetry group of a quantum system comes with a unitary projective representation $S \mapsto U(S)$.

It turns out that for the simple cases considered in this chapter, we can ignore the possibility of phase factors $\phi(R,S)$, and focus only on linear representations.

4.3 Rotations: Definition, generators, spin-$1/2$

Motivated by Wigner's Theorem, we'll take a look at representations of rotations.

4.3.1 Rotations

First some basics.

A *rotation* in $\mathbb{R}^n$ is a linear map $\boldsymbol{x} \mapsto R\boldsymbol{x}$ that

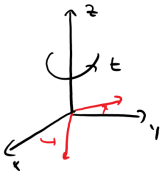
- preserves inner products

$$(R\boldsymbol{x}, R\boldsymbol{y}) = (\boldsymbol{x}, \boldsymbol{y}), \quad \text{or, equivalently,} \quad R^t R = \mathbb{1}, \tag{4.6}$$

- preserves orientation

$$\det R = 1. \tag{4.7}$$

The set of all rotations of $\mathbb{R}^n$ is denoted as $SO(n)$, the "special orthogonal group".



E.g., rotations around the z-axis $\boldsymbol{e}_3$ by an angle θ are represented by

$$R_{\theta \boldsymbol{e}_3} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.8}$$

A theorem of Euler states that for every rotation R of $\mathbb{R}^3$, there is a unit vector $\boldsymbol{\omega}$ and an angle θ such that $R = R_{\theta\boldsymbol{\omega}}$ rotates by θ about the axis $\boldsymbol{\omega}$.

There are two physically important unitary representations of $SO(3)$ that are simple to construct.

Examples. (More details in Chapter 5).

- Let $\mathcal{H} = L^2(\mathbb{R}^3)$. For $R \in SO(3)$, define a linear map $U(R)$ by

$$\big(U(R)\psi\big)(\boldsymbol{x}) = \psi(R^{-1}\boldsymbol{x}). \tag{4.9}$$

(The $^{-1}$ plays the same role as the minus sign in (4.3)). Because $\det R = 1$, the transformation is volume-preserving and hence unitary. It is geometrically obvious that this is a representation (why?), though let's confirm this for those who only trust algebra:

$$\begin{aligned}\big(U(R)U(S)\psi\big)(\boldsymbol{x}) &= \big(U(S)\psi\big)(R^{-1}\boldsymbol{x}) = \psi(S^{-1}R^{-1}\boldsymbol{x}) = \psi((RS)^{-1}\boldsymbol{x}) \\ &= \big(U(RS)\psi\big)(\boldsymbol{x}).\end{aligned}$$

Edit: ...and because the calculation above led to questions, here a more explicit version:

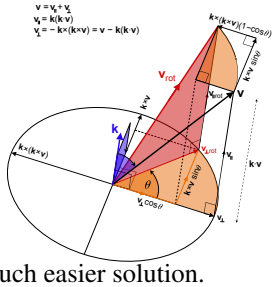
$$\begin{aligned}\big(U(R)\underbrace{U(S)\psi}_{=:\phi}\big)(\boldsymbol{x}) &= \big(U(R)\phi\big)(\boldsymbol{x}) \\ &= \phi(\underbrace{R^{-1}\boldsymbol{x}}_{=:\boldsymbol{y}}) && \text{(by (4.9))} \\ &= (U(S)\psi)(\boldsymbol{y}) && \text{(by def of } \phi) \\ &= \psi(S^{-1}\boldsymbol{y}) && \text{(by (4.9))} \\ &= \psi(S^{-1}R^{-1}\boldsymbol{x}) && \text{(by def of } \boldsymbol{y}) \\ &= \psi((RS)^{-1}\boldsymbol{x}) && \text{(inverse of a composition)} \\ &= (U(RS)\psi)(\boldsymbol{x}) && \text{(by (4.9)).}\end{aligned}$$

The map $R \mapsto U(R)$ is called the *orbital angular momentum representation*. We'll analyze it in detail, in particular for the solution of the hydrogen atom.

- Let $\mathcal{H} = \mathbb{C}^3$. Because $R \in SO(3)$ is just a 3×3-matrix, we can let it act on $\mathcal{H}$ directly. The orthogonality condition (4.6) also implies that R is unitary. For reasons we'll explain later, this realization is called the *spin-1 representation*, $U_1 : R \mapsto R$. It appears e.g. in the quantum theory of photons.

Eventually, we will classify *all* such representations (Chapter 5). The first obstacle we face on the way is that the composition law that appears in (4.4) can be quite complicated.

For example, pick a pair of angles θ_1, θ_2 and axes $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2$. Then the composition $R_{\theta_1\boldsymbol{\omega}_1}R_{\theta_2\boldsymbol{\omega}_2}$ is again a rotation, i.e. of the form $R_{\theta_3\boldsymbol{\omega}_3}$. But how does one find $\theta_3, \boldsymbol{\omega}_3$ as a function of the original rotations? For $SO(3)$, the answer is given by *Rodrigues' rotation formula*[2], which one can work out by analyzing the simple diagram to the right!



What? That doesn't sound like fun? Fortunately, there's a much easier solution.

4.3.2 Generators

The composition law of continuous groups is most easily expressed in terms of their *generators*. (Physicists also talk about *infinitesimal elements* and mathematicians about the *Lie algebra* of the group). The central observation is that

[2] Should rename it *Rodrigues' rotation relation*, to really roll off the tongue.

"*a rotation is the combination of many small ones, so it suffices to understand those*."

More precisely, consider the z-axis rotations $R_{\theta \boldsymbol{e}_3}$ defined in (4.8). Then

$$R_{\theta \boldsymbol{e}_3} = \left(R_{\theta/N \boldsymbol{e}_3}\right)^N \qquad \text{for any } N \in \{1, 2, 3, \dots\}.$$

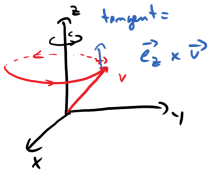
In the limit of $N \to \infty$, the first-order Taylor approximation

$$R_{\theta/N \boldsymbol{e}_3} \simeq \mathbb{1} + \frac{\theta}{N} l_3, \qquad l_3 = \partial_\theta|_0 \, R_{\theta \boldsymbol{e}_3} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{4.10}$$

becomes exact. Therefore, using Euler's characterization of the exponential,

$$R_{\theta \boldsymbol{e}_3} = \lim_{N\to\infty} \left(R_{\theta/N \boldsymbol{e}_3}\right)^N = \lim_{N\to\infty} \left(\mathbb{1} + \frac{\theta}{N} l_3\right)^N = e^{\theta l_3} \qquad \text{(matrix exponential)}.$$

The matrix l_3 is called the *generator* of the *one-parameter group* $R_{\theta \boldsymbol{e}_3}$. Physicist also refer to $\mathbb{1} + \theta l_3$ for "very small" θ as an *infinitesimal rotation*.



Let's generalize this to arbitrary $\boldsymbol{\omega}$. Re-writing (4.10) as

$$\partial_\theta|_0 R_{\theta \boldsymbol{e}_3} \boldsymbol{v} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \boldsymbol{v} = \boldsymbol{e}_z \times \boldsymbol{v}$$

i.e. in terms of the cross product with the rotation axis, suggests that $\partial_\theta|_0 R_{\theta \boldsymbol{\omega}} \boldsymbol{v} = \boldsymbol{\omega} \times \boldsymbol{v}$. This can indeed be verified. Expressing the cross product as a matrix multiplication gives

$$\partial_\theta|_0 R_{\theta \boldsymbol{\omega}} \boldsymbol{v} = \boldsymbol{\omega} \times \boldsymbol{v} = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix} \boldsymbol{v} = \left(\sum_{i=1}^{3} \omega_i l_i\right) \boldsymbol{v},$$

having introduced the following basis for the space of anti-symmetric matrices:

$$l_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad l_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad l_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{4.11}$$

It is often convenient to allow for non-normalized vectors $\boldsymbol{\omega}$. Then, in summary

The exponential $e^{\sum_i \omega_i l_i}$ is the rotation about the axis $\frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|}$ by the angle $|\boldsymbol{\omega}\|$.

Baker-Campbell-Hausdorff

Recall that we aim to express the composition law in a less painful way. In the language introduced above, this means: Given two generators g, h, what is the generator of $e^g e^h$?

The product of two matrix exponentials can be evaluated using the important *Baker-Campbell-Hausdorff*[3] *(BCH) formula*. Let g, h be two arbitrary matrices (not necessarily

[3] Studying in the region, you should learn about the life of Hausdorff. Be warned. There's no happy ending.

generators of rotations). As a warm-up, assume that g and h commute. Then switching to a common eigenbasis shows that

$$e^g e^h = e^{g+h} \qquad (\text{only if } [g,h]=0).$$

In the general case, correction terms appear in the exponent. These are described by the BCH formula (working out the lowest-order correction is homework):

$$e^g e^h = \exp\left(g + h + \frac{1}{2}[g,h] + \frac{1}{12}([g,[g,h]] - [h,[g,h]]) + \dots\right), \tag{4.12}$$

where the dots represent further nested commutators.

Caveat. The exponent in (4.12) is an infinite series. Strictly speaking, the series in the exponent is only guaranteed to converge for group elements g, h that are "sufficiently close to the identity". For now, we can get away with ignoring this subtlety. Having said this, it *is* the basis of one of my favorite quantum effects, and we'll have a brief look at in in Sec. **??**

The precise form of the higher-order terms isn't important. The main lesson is:

The composition law is determined by the commutator relations of the generators.

(If you ever wondered why physicists seem so infatuated with commutators and Poisson brackets and such... ...now you know: *These encode information about groups*!)

For the basis (4.11) a direct computation gives

$$[l_1, l_2] = l_3 \quad \text{and cyclic permutations.} \tag{4.13}$$

By linearity, (4.13) determines arbitrary commutators, and hence, by (4.12) the group law of $SO(3)$. This simple relation thus replaces all the nastiness of Rodrigues' formula!

Exercise. Consider a rotation by a small angle θ first around the x axis and then around the y axis. Can you make sense of the first and second order terms in Eq. (4.12)? Maybe get a ball to play around with and explain what's going on to a partner (or at least to a rubber duck).

4.3.3 Representations of generators

Instead of directly working with representations of groups, the above discussion suggests to work with representations of their generators instead.

To see how this works, let $\mathcal{H}$ be a Hilbert space, and let j_1, j_2, j_3 be operators on $\mathcal{H}$. Define a linear map

$$u\Big(\sum_i \omega_i l_i\Big) = \sum_i \omega_i j_i \tag{4.14}$$

sending generators of rotations to operators on $\mathcal{H}$. Our goal is to find out under which conditions

$$U(e^g) := e^{u(g)} \tag{4.15}$$

defines a representation. Using BCH twice

$$\begin{aligned} U(e^g e^h) &= U(e^{g+h+\frac{1}{2}[g,h]+\dots}) = e^{u(g)+u(h)+\frac{1}{2}u([g,h])+\dots} \\ U(e^g)U(e^h) &= e^{u(g)}e^{u(h)} = e^{u(g)+u(h)+\frac{1}{2}[u(g),u(h)]+\dots} \end{aligned} \tag{4.16}$$

and comparing exponents, we find that

$$U(e^g e^h) = U(e^g)U(e^h) \quad \Leftrightarrow \quad [u(g), u(h)] = u([g, h]). \tag{4.17}$$

The map $U(e^g) = e^{u(g)}$ is a representation iff u preserves commutator relations.

A short calculation using bilinearity of the commutator shows that (4.14) preserves the commutator relation of $SO(3)$ generators iff $[j_1, j_2] = j_3$ (and cyclic permutations). Thus:

There is a one-one correspondence between representations of $SO(3)$ and operators

$$j_1, j_2, j_3 \qquad \text{that satisfy} \qquad [j_1, j_2] = j_3 \quad \text{(and cyclic permutations).} \tag{4.18}$$

Example. Compute the generators for the orbital angular momentum representation (4.9):

$$\begin{aligned}
\partial_\theta|_0 \big(U(R_{\theta \boldsymbol{e}_i})\psi\big)(\boldsymbol{x}) &= \partial_\theta|_0 \psi\big(R^{-1}_{\theta,\boldsymbol{e}_i}\boldsymbol{x}\big) \\
&= \boldsymbol{\nabla}\psi \cdot \big(\partial_\theta|_0 R_{-\theta,\boldsymbol{e}_i}\boldsymbol{x}\big) \\
&= \boldsymbol{\nabla}\psi \cdot \big(\boldsymbol{x} \times \boldsymbol{e}_i\big) \\
&= \boldsymbol{e}_i \cdot \big(\boldsymbol{x} \times \boldsymbol{\nabla}\psi\big) \\
&= \frac{1}{i\hbar}\boldsymbol{e}_i \cdot \big(\boldsymbol{x} \times \boldsymbol{P}\psi\big) \\
&= \frac{1}{i\hbar} \underbrace{\big(\boldsymbol{X} \times \boldsymbol{P}\big)_i}_{=:L_i} \psi.
\end{aligned}$$

Thus, up to constants, $u(l_i)$ is the i-th component of the *orbital angular momentum operator* $\boldsymbol{L} = \boldsymbol{X} \times \boldsymbol{P}$. That is no coincidence (because nothing ever is!), as we will see.

Exercises.

- Compute the commutator $[L_1, L_2]$ between the first two components of angular momentum. Verify that the $\frac{1}{i\hbar}L_i$ satisfy the same commutator relations as the generators of $SO(3)$.
- Show that the inverse of e^g is e^{-g}.
- An operator g is *anti-Hermitian* if $g^\dagger = -g$. Show that g is anti-Hermitian if and only if ig is Hermitian. Show that e^g is unitary if and only if g is anti-Hermitian.

4.3.4 Spin-1/2

We can now find reps of $SO(3)$ by guessing j_1, j_2, j_3 and verifying (4.18). Examples:

- $\dim \mathcal{H} = 1$. All 1×1-matrices commute, so the only solution is $j_1 = j_2 = j_3 = 0$. This generates the *trivial* or *spin*-0 representation $U_0 : R \mapsto e^0 = \mathbb{1}$. While boring, the spin-0 representation is actually quite relevant! It describes spaces of vectors that do not change under rotations. Example: The ground state of the H atom.
- $\dim \mathcal{H} = 3$. We've already found the *defining* or *spin*-1 representation $U_1 : R \mapsto R$.

This leaves a gap at $\dim \mathcal{H} = 2$. It isn't obvious whether one can represent 3D rotations on a 2D space... So, remember how we've constantly annoyed you with Pauli matrices in

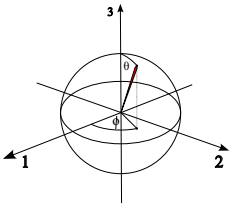
the linear algebra course? Now it's time to cash in! The Paulis anti-commute and satisfy

$$\sigma_1\sigma_2 = i\sigma_3$$

which implies (4.18) for $j_k = \frac{1}{2i}\sigma_k$. They generate the *spin-*$1/2$ *representation*.

The re-scaled Paulis generate the *spin-*$\frac{1}{2}$ *representation* $U_{\frac{1}{2}}(R_{\boldsymbol{\omega}}) = e^{\frac{1}{2i}\sum_k \omega_k\sigma_k}$.

Exercise. So how *can* a three-dimensional rotation act on a two-dimensional space?



The key is to go complex. Up to a (physically irrelevant) global phase, every normalized vector $|\psi\rangle \in \mathbb{C}^2$ can be written as

$$|\psi\rangle = \cos(\theta/2)|0\rangle + e^{i\phi}\sin(\theta/2)|1\rangle, \quad 0 \le \theta \le \pi,\ 0 \le \phi < 2\pi.$$

Interpreting θ, ϕ as polar and azimuthal angles, they specify a point on a sphere. It is called the *Bloch sphere representation* of $|\psi\rangle$. Homework: Show that the Bloch vector of $|\psi'\rangle = U_{\frac{1}{2}}(R)|\psi\rangle$ equals the result of R acting on the Bloch vector of $|\psi\rangle$.

Above, we have denoted the standard basis in $\mathbb{C}^2$ by

$$|0\rangle = |e_0\rangle = \begin{pmatrix}1\\0\end{pmatrix}, \qquad |1\rangle = |e_1\rangle = \begin{pmatrix}0\\1\end{pmatrix}.$$

Yay, we found the lowest-dimensional space that behaves non-trivially under rotations! *That's cute and all*, you might counter, *but the discussion has been purely mathematical. We have yet to work out a physical interpretation which makes testable predictions, so that we can look for signatures of spin-*$\frac{1}{2}$ *in Nature.* OK, then, let's do that.

4.4 Conserved quantities

It turns out that suitable multiples of the generators of symmetries can often be identified with physically relevant observables. The connection goes via the theory of conserved quantities, the basics of which we'll briefly recall.

4.4.1 Classical mechanical energy and momenta

Consider a single free classical point particle in $\mathbb{R}^3$. Its energy $H = \frac{\|\boldsymbol{p}\|^2}{2m}$, momentum $\boldsymbol{p}$, and angular momentum $\boldsymbol{L} = \boldsymbol{x} \times \boldsymbol{p}$ are constants of motion.

If the particle interacts with a second one, its *individual* energy and momenta are no longer conserved. However, Noether's theorem says that if the Hamilton function of the combined system is of the form

$$H = \frac{\|\boldsymbol{p}_1\|^2}{2m_1} + \frac{\|\boldsymbol{p}_2\|^2}{2m_2} + H_I, \qquad H_I = V(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|),$$

i.e. invariant under translations of time and space and under rotations, then conservation holds for the *total* energies and momenta

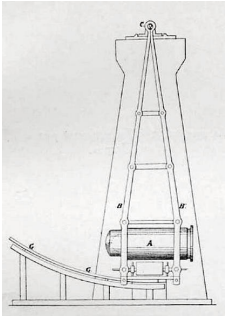
$$H = H_1 + H_2 + H_I, \quad \boldsymbol{p} = \boldsymbol{p}_1 + \boldsymbol{p}_2, \quad \boldsymbol{L} = \boldsymbol{L}_1 + \boldsymbol{L}_2.$$

What is more, in a scattering process, i.e. a process where the two particles do not interact for $t \to \pm\infty$, the interaction term H_I does not contribute to the energy in the limit. Therefore:

In a scattering process where the interaction is invariant under rotations, space and time translations, the sum of the single-body energies and momenta is preserved.

This has practical consequences. For example:

Measurement procedures based on conservation laws



A *ballistic pendulum* (right) uses conservation laws to measure the nozzle velocity of a canon ball (explain how!).

The general idea is to couple a conserved quantity of a difficult-to-access system (a fired canon ball) to a reference system (the pendulum). The method works because conservation laws hold independently of the details of the interaction. In this case, the interaction between the barrel and the ball is mediated by exploding gunpowder – probably very complex and not well-understood.

4.4.2 Generalizations of mechanical energy and momenta

Consider a non-mechanical system, e.g. the electro-magnetic (EM) field or a quantum particle. When does it make sense to call a function H_{gen} of its states a "generalization of mechanical energy"?

Arguably, this is well-justified if one can couple the system to a mechanical reference, and if the sum of H_{gen} and mechanical energy H_{mech} is conserved (in any scattering processes as described above). That's because in this case, any decrease in mechanical energy is accompanied by an equal increase in H_{gen}, so that in a sense, the two functions represent the same kind of stuff. The same logic applies to linear and angular momenta.

A physical quantity represents energy / momenta if it satisfies the corresponding conservation law when interacting with a mechanical reference system.

Example. Consider a mechanical particle that carries a charge and thus interacts with the EM field. Conservation of total energy, linear and angular momentum holds if

$$H_{\text{em}} = \frac{\epsilon_0}{2} \int \boldsymbol{E}^2(\boldsymbol{x}) + c^2 \boldsymbol{B}^2(\boldsymbol{x}) \, \mathrm{d}^3\boldsymbol{x},$$
$$\boldsymbol{p}_{\text{em}} = \epsilon_0 \int \boldsymbol{E}(\boldsymbol{x}) \times \boldsymbol{B}(\boldsymbol{x}) \, \mathrm{d}^3\boldsymbol{x},$$
$$\boldsymbol{L}_{\text{em}} = \epsilon_0 \int \boldsymbol{x} \times (\boldsymbol{E}(\boldsymbol{x}) \times \boldsymbol{B}(\boldsymbol{x})) \, \mathrm{d}^3\boldsymbol{x}.$$

There's no guarantee that such generalized quantities can always be found... ...maybe there *is* a perpetual motion machine. But, while seemingly fundamental concepts had to be abandoned left and right as physics progressed in the past century (determinism, absoluteness of lengths and simultaneity, ...) conservation of energy and momenta has held up!

What is more, *if* one can associate generalized conserved quantities to a new system at all, then their values are essentially unique, and can be determined by physical observations (as we'll see now), or even from theoretical arguments (next section).

Details (on the uniqueness of conserved quantities). Consider a scattering process where the reference particle starts in a state with energy $H^{\text{in}}_{\text{mech}}$ and ends up in a state with energy $H^{\text{out}}_{\text{mech}}$. Then energy conservation gives

$$H^{\text{out}}_{\text{gen}} + H^{\text{out}}_{\text{mech}} = H^{\text{in}}_{\text{mech}} + H^{\text{in}}_{\text{gen}} \quad \Rightarrow \quad H^{\text{out}}_{\text{gen}} - H^{\text{in}}_{\text{gen}} = -(H^{\text{out}}_{\text{mech}} - H^{\text{in}}_{\text{mech}}). \tag{4.19}$$

Hence, the difference in generalized energy between the outgoing and the incoming state of the general system must be equal to minus the energy difference measurable on the mechanical reference. If there is one state of the general system that can be converted into any other by suitable scattering processes, then (4.19) fixes the generalized energy uniquely up to an additive constant (the energy of the state we started from). This remaining ambiguity is unavoidable, because if $H_{\text{mech}} + H_{\text{gen}}$ is conserved, then so is $H_{\text{mech}} + H_{\text{gen}} + H_0$ for any constant H_0.

4.4.3 Conserved quantities and generators

It turns out that if a quantum theory is invariant under a space-time symmetry group, one can often argue that observables satisfy a conservation law if and only if they are proportional to a corresponding generator. Some examples of this correspondence are:

Symmetry	Generator	Observable	Conserved quantity
Rotation about $\boldsymbol{e}_k$	j_k	$J_k = i\hbar\, j_k$	angular momentum
Displacement along $\boldsymbol{e}_k$	p_k	$P_k = i\hbar\, p_k$	linear momentum
Displacement in time	h	$H = -i\hbar\, h$	energy

We'll outline the argument here for the case of spin-$1/2$ systems. (Though some proofs will have to be deferred until more theory is introduced later).

The argument makes use of the transformation properties of classical momenta.

Exercise. Under a rotation $\boldsymbol{x} \mapsto R\boldsymbol{x}$, the momentum and angular momentum of a classical point particle transform as

$$\boldsymbol{p}_{\text{mech}} \mapsto R\boldsymbol{p}_{\text{mech}}, \qquad \boldsymbol{L}_{\text{mech}} \mapsto R\boldsymbol{L}_{\text{mech}}. \tag{4.20}$$

Now proceed as follows:

1. Argue that if a system can couple to a classical point particle, its angular momentum must "inherit" the transformation property (4.20), at least up to additive constants.
2. Show that for spin-$1/2$ particles, there is a unique set of observables that transforms in this way.

Angular momenta transform as vectors under rotation

As a preparation: It is easy to see that *differences* of generalized angular momenta must transform as in (4.20). Indeed, in a scattering process in which angular momenta are conserved, we have that

$$\boldsymbol{L}^{\text{out}}_{\text{gen}} - \boldsymbol{L}^{\text{in}}_{\text{gen}} = \boldsymbol{L}^{\text{in}}_{\text{mech}} - \boldsymbol{L}^{\text{out}}_{\text{mech}}. \tag{4.21}$$

Now consider the same process, but rotated by some R:

$$\begin{aligned} & \boldsymbol{L}_{\text{gen}}^{\text{out}\,\prime} - \boldsymbol{L}_{\text{gen}}^{\text{in}\ \prime} = \boldsymbol{L}_{\text{mech}}^{\text{in}\ \prime} - \boldsymbol{L}_{\text{mech}}^{\text{out}\ \prime} \\ \Rightarrow \quad & \boldsymbol{L}_{\text{gen}}^{\text{out}\,\prime} - \boldsymbol{L}_{\text{gen}}^{\text{in}\ \prime} = R\big(\boldsymbol{L}_{\text{mech}}^{\text{in}} - \boldsymbol{L}_{\text{mech}}^{\text{out}}\big) \qquad \text{(by (4.20))} \\ \Rightarrow \quad & \boldsymbol{L}_{\text{gen}}^{\text{out}\,\prime} - \boldsymbol{L}_{\text{gen}}^{\text{in}\ \prime} = R\big(\boldsymbol{L}_{\text{gen}}^{\text{out}} - \boldsymbol{L}_{\text{gen}}^{\text{in}}\big) \qquad \text{(by (4.21)).} \end{aligned}$$

Arguing that the same holds for each summand is a little bit more complicated, and requires and extra assumption, see box below.

Details. TBD.

Spin-1/2 angular momentum observables

Let's apply the argument to derive the angular momentum observables $\boldsymbol{S} = (S_1, S_2, S_3)$ for spin-$\frac{1}{2}$ systems. We'll write $s_k = \frac{i}{2}\sigma_k$ for the generators of the representation.

Because $\{\sigma_0 = \mathbb{1}, \sigma_1, \sigma_2, \sigma_3\}$ is a basis of the space of 2×2-basis, we can expand

$$S_k = \sum_{l=0}^{3} c_l^{(k)} \sigma_l$$

in terms of yet-to-be-determined coefficients $c_l^{(k)}$. By the previous section, the k-th component of angular momentum is invariant under rotations about the $\boldsymbol{e}_k$-axis. As in Eq. (1.33), and using the commutation relations of the Pauli matrices,

$$\partial_\theta|_0 e^{\theta s_k} S_k e^{-\theta s_k} = \Big[s_k, S_k\Big] = \sum \quad \Leftrightarrow \quad S_k = c_0^{(k)}\sigma_0 + c_k^{(k)}\sigma_k.$$

Because $\sigma_0 = \mathbb{1}$, the first coefficient is just an additive constant, which we may set to 0.

The three coefficients $c_k^{(k)}$ must all be the same. For example, to show that $c^{(1)} = c^{(2)}$, consider the behavior of S_1 under rotations about $\boldsymbol{e}_3$:

$$\partial_\theta|_0 e^{\theta s_3} S_1 e^{-\theta s_3} = c_1^{(1)} \frac{i}{2}[\sigma_3, \sigma_1] = c_1^{(1)}\sigma_2.$$

But the transformation law (4.20) demands

$$\partial_\theta|_0 e^{\theta s_3} S_1 e^{-\theta s_3} = \partial_\theta|_0 \sum_k S_k (R_{\theta \boldsymbol{e}_3})^k{}_1 = S_2 = c_2^{(2)}\sigma_2 \quad \Rightarrow \quad c_2^{(1)} = c_2^{(2)}.$$

The angular momentum observable S_k of a spin-$\frac{1}{2}$ system proportional to σ_k.

The value of the proportionality constant cannot be derived theoretically (beyond that it must be imaginary, for the operators S_k to be Hermitian). Experimentally, it's $i\frac{\hbar}{2}$.

Remark. The argument implies that *any* observable on a spin-$1/2$ system is proportional to an angular momentum (plus an irrelevant multiple of $\mathbb{1}$). In this sense, spin-$1/2$ systems are "bundles of pure angular momentum", with no other physical properties.

Chapter 5

Angular momentum

Under construction. Use Chapter 7 in handwritten notes for now.

As a starting point, we define a quantum mechanical *angular momentum* as a set of hermitian operators L_1, L_2, L_3 satisfying

$$[L_1, L_2] = i\hbar L_3 \quad (+ \text{ cyclic permutations}).$$

Just as momentum is a generator of translations, angular momentum is a generator of rotations and indeed the operators $\frac{1}{i\hbar}L_i$ satisfy the commutation relations of the generators of the rotation group 4.13:

$$[\tfrac{1}{i\hbar}L_1, \tfrac{1}{i\hbar}L_2] = (\tfrac{1}{i\hbar})^2[L_1, L_2] = \tfrac{1}{i\hbar}L_3$$

Note: The quantum mechanical observables differ from the generators by a factor of $i\hbar$. The existance of two normalizations reflects the fact that observables in QM describe measurements as well as generate symmetries. The two roles require different normalizations. We have already seen this in the Schrödinger equation: the generator of time evolution $\frac{1}{i\hbar}$, differs from the observable that corresponds to the energy by the same prefactor.

From now on we will work with observables and the corresponding normalization.

Define the operator

$$L^2 := L_1^2 + L_2^2 + L_3^2.$$

(It's not defined as the square of a "L" operator, just common notation).

The operator L^2 had the special property[1] that it commutes will all the L_i, i.e.,

$$[L^2, L_i] = 0.$$

(the verification of this fact is left to you as homework). Since L^2 commutes with the L_i we can find joint eigenbasis of L^2, L_3. (The fact that we have chosen L_3 is a matter of historical convention. It has no physical relavence. In fact we could replace L_3 with $\sum_i \omega_i L_i$ for any vector $\boldsymbol{\omega} \in \mathbb{R}^3$). Let $|\phi\rangle$ be a common eigenvector. Write the eigenvalue of L_3 as

$$L_3|\phi\rangle = \underbrace{m\hbar}_{\text{(the eigenvalue)}} |\phi\rangle.$$

[1] Casimir Operator. TBD.

Here "m" is called the *magnetic quantum number*.

The operator L^2 is the sum of squares of operators, hence any eigenvalue is non-neqative and has a square roots, can write

$$L^2|\phi\rangle = \lambda^2\hbar^2|\phi\rangle$$

for some $\lambda \in \mathbb{R}$. Even worse, write

$$\ell = \sqrt{\lambda^2 + \frac{1}{4}} - \frac{1}{2}$$

then

$$\ell(\ell+1) = \left(\sqrt{\lambda^2 + \frac{1}{4}} - \frac{1}{2}\right)\left(\sqrt{\lambda^2 + \frac{1}{4}} + \frac{1}{2}\right) = \lambda^2 + \frac{1}{4} - \frac{1}{4} = \lambda^2,$$

so that

$$L^2|\phi\rangle = \ell(\ell+1)\hbar^2|\phi\rangle \qquad L_z|\phi\rangle = m\hbar|\phi\rangle.$$

We can thus choose an ONB of the form

$$\{|k,l,m\rangle =: |klm\rangle\}_{k,l,m}$$

with

ℓ	angular momentum quantum number
m	magnetic quantum number
k	degeneracy

Ex.: Spin-$\frac{1}{2}$. For $L_z = S_z = \frac{\hbar}{2}\sigma_z$,

$$m = \pm\frac{1}{2}, \qquad \lambda^2 = \frac{3}{4} \quad \Rightarrow \quad \ell = \sqrt{\frac{3}{4} + \frac{1}{4}} - \frac{1}{2} = \frac{1}{2}$$

and

$$|\frac{1}{2},\frac{1}{2}\rangle = \begin{pmatrix}1\\0\end{pmatrix} =: |\uparrow\rangle \qquad |\frac{1}{2},-\frac{1}{2}\rangle = \begin{pmatrix}0\\1\end{pmatrix} =: |\downarrow\rangle.$$

5.0.1 Magnetic quantum numbers through laddering

Define (cf. Sec. 1.6.1)

$$L_+ := L_1 + iL_2, \qquad \Rightarrow \qquad L_- := L_+^\dagger = L_1 - iL_2.$$

Ex.: Spin-$\frac{1}{2}$ Ex.:

$$S_+ = \frac{\hbar}{2}(\sigma_x + i\sigma_y) = \frac{\hbar}{2}\begin{pmatrix}0&1\\1&0\end{pmatrix} + i\frac{\hbar}{2}\begin{pmatrix}0&-i\\i&0\end{pmatrix} = \hbar\begin{pmatrix}0&1\\0&0\end{pmatrix},$$
$$S_- = S_+^\dagger = \hbar\begin{pmatrix}0&0\\1&0\end{pmatrix}.$$

Homework:

$$[L_3, L_\pm] = \pm\hbar L_\pm, \qquad L_-L_+ = L^2 - L_3(L_3+\hbar).$$

Ex.: Spin-$\frac{1}{2}$ Ex.:

$$L_z L_+ = \frac{\hbar^2}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix} = \frac{\hbar}{2}L_+$$
$$L_+ L_z = \frac{\hbar^2}{2}\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} = -\frac{\hbar}{2}L_+$$

.

Claim 1: It holds that

$$-\ell \le m \le \ell$$

and

$$\begin{aligned} m = \ell \quad &\Leftrightarrow L_+|k,l,m\rangle = 0\\ m = -\ell \quad &\Leftrightarrow L_-|k,l,m\rangle = 0. \end{aligned}$$

Proof:

$$\begin{aligned} 0 \le \frac{1}{\hbar^2}\|L_+|klm\rangle\|^2 &= \frac{1}{\hbar^2}\langle klm|L_-L_+|klm\rangle\\ &= \frac{1}{\hbar^2}\langle klm|L^2 - L_z(L_z-\hbar)|klm\rangle\\ &= \ell(\ell+1) - m(m+1). \end{aligned}$$

In particular,

$$\|L_\pm|klm\rangle\| = \hbar\sqrt{\ell(\ell+1) - m(m\pm 1)}. \tag{5.1}$$

Ex.: Spin-$\frac{1}{2}$

$$S_+|\uparrow\rangle = \hbar\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}\begin{pmatrix}1\\ 0\end{pmatrix} = \hbar\begin{pmatrix}0\\ 0\end{pmatrix} = 0$$
$$S_+|\downarrow\rangle = \hbar\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}\begin{pmatrix}0\\ 1\end{pmatrix} = \hbar\begin{pmatrix}1\\ 0\end{pmatrix} = \hbar|\uparrow\rangle$$

and indeed,

$$\hbar\sqrt{\tfrac{1}{2}(\tfrac{1}{2}+1) + \tfrac{1}{2}(-\tfrac{1}{2}+1)} = \hbar\sqrt{\frac{3}{4}+\frac{1}{4}} = \hbar.$$

Claim 2:

- If $m < \ell$, then $L_+|klm\rangle$ is L_z eigenvector with eigenvalue $\hbar(m+1)$.
- If $m > -\ell$, then $L_-|klm\rangle$ is L_z eigenvector with eigenvalue $\hbar(m-1)$.

Proof.

$$\begin{aligned} L_z\big(L_+|klm\rangle\big) &= \big(L_+\hbar + L_+L_z\big)|klm\rangle\\ &= \big(L_+\hbar + L_+\hbar m\big)|klm\rangle\\ &= \hbar(m+1)\big(L_+|klm\rangle\big). \end{aligned}$$

TBD: Figure.

Laddering *must* terminate. Hence $m = \pm\ell$ *must* occur. Therefore, we have a $2\ell + 1$ dimensional subspace for each ℓ. Since the dimension of a vector space is a natural number it must be the case that

$$\ell \in \{0, \frac{1}{2}, 1, \frac{3}{2}, 2, ...\}.$$

This in turn implies that the magnetic quantum number is quantized. We call the two possible cases $m = \{-\ell, \cdots, -\frac{1}{2}, \frac{1}{2}, \cdots, \ell\}$ and $m = \{\ell, \cdots, -1, 0, 1, \cdots, \ell\}$, *half-integral* and *integral* respectively. For each choice of ℓ we get a subspace of the full Hilbert space $\mathcal{H}$, i.e.,

$$\mathcal{H}_{(1,\ell)} := \langle \{|1, l, m\rangle\}_{m=-\ell}^{\ell} \rangle \subset \mathcal{H}.$$

In each $\mathcal{H}_{(k,\ell)}$ we can find the common eigenvectors of L^2 and L_3, for example by finding one of the eigenvectors and then using the ladder operators. The collection of all such eigenvectors then forms an ONB of $\mathcal{H}$. In this basis, the previously discussed operators have a *block diagonal* structure. For example

$$\langle k'l'm'|L_3|klm\rangle = \delta_{k'k}\delta_{l'l}\delta_{m'm}\hbar m$$
$$\langle k'l'm'|L_\pm|klm\rangle = \delta_{k'k}\delta_{l'l}\delta_{m',m\pm1}\hbar\sqrt{\ell(\ell+1) - m(m\pm1)}$$

TBD: Block diagonal figure.

5.1 Orbital angular momentum

As an example, let us look at the angular momentum that accompanies circular motion. From classical mechanics, one might be familiar with the angular momentum vector $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$ (for example from planetary motion). The quantum mechanical analog of this quantity is called the *orbital angular momentum* (which one encounters while studying, for example, the structure of atoms).

To define the orbital angular momentum operator, we will first need the vectorial versions of the position and momentum operators on the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^3)$. These are given by:

$$\boldsymbol{R} := \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}, \qquad (\boldsymbol{R}\psi)(x,y,z) = \begin{pmatrix} x\psi(x,y,z) \\ y\psi(x,y,z) \\ z\psi(x,y,z) \end{pmatrix},$$
$$\boldsymbol{P} := -i\hbar \begin{pmatrix} \partial_x \\ \partial_y \\ \partial_z \end{pmatrix}, \qquad (\boldsymbol{P}\psi)(x,y,z) = -i\hbar \begin{pmatrix} \partial_x\psi(x,y,z) \\ \partial_y\psi(x,y,z) \\ \partial_z\psi(x,y,z) \end{pmatrix}.$$

The *orbital angular momentum operator* is defined as

$$\boldsymbol{L} = \begin{pmatrix} L_x \\ L_y \\ L_z \end{pmatrix} := \boldsymbol{R} \times \boldsymbol{P} = -i\hbar \begin{pmatrix} y\partial_z - z\partial_y \\ z\partial_x - x\partial_z \\ x\partial_y - y\partial_x \end{pmatrix}.$$

One can indeed verify that (homework)

$$[L_x, L_y] = i\hbar L_z \quad (+ \text{ cyclic permutations}),$$

and therefore $\boldsymbol{L}$ is a QM angular momentum.

Reduction to functions on a sphere

Switching to *spherical polar coordiantes* we get the following (the details are homework)

$$L_x = -i\hbar(y\partial_z - z\partial_y) = i\hbar\Big(\sin\phi\frac{\partial}{\partial\theta} + \frac{\cos\phi}{\tan\theta}\frac{\partial}{\partial\phi}\Big)$$
$$L_y = -i\hbar(z\partial_x - x\partial_z) = i\hbar\Big(-\cos\phi\frac{\partial}{\partial\theta} + \frac{\sin\phi}{\tan\theta}\frac{\partial}{\partial\phi}\Big)$$
$$L_z = -i\hbar(x\partial_y - y\partial_x) = -i\hbar\frac{\partial}{\partial\phi}$$

The operators are independent of the radial coordinate r! From here we can proceed in one of two ways:

Geometric argument:

TBD.

Analytic argument:

To find the eigenfunctions common eigenfunctions of L^2 and L_z we set up the eigenvalue equations:

$$L^2\Psi_{klm}(r,\theta,\phi) = \hbar^2\ell(\ell+1)\Psi_{klm}(r,\theta,\phi),$$
$$L_z\Psi_{klm}(r,\theta,\phi) = \hbar m\Psi_{klm}(r,\theta,\phi).$$

These are nothing but simultaneous partial differential equations. Since the differential operators are all independent of r, we can use an ansatz of the form

$$\Psi_{klm}(r,\theta,\phi) = f(r)Y_{klm}(\theta,\phi),$$

where $f(r)$ is an arbitrary, normalised function, i.e., $||f|| = 1$. Therefore the above set of differential equations reduces to

$$L^2Y_{klm}(\theta,\phi) = \hbar^2\ell(\ell+1)Y_{klm}(\theta,\phi),$$
$$L_zY_{klm}(\theta,\phi) = \hbar mY_{klm}(\theta,\phi).$$

The functions Y_{klm} live on the unit sphere and have to be square-integrable, i.e., $Y_{klm} \in L^2(S^2) \subset L^2(\mathbb{R}^3)$.

Solving on the sphere: Spherical harmonics

Recall,

$$L_z = -i\hbar\partial_\phi,$$
$$L_\pm = \hbar e^{\pm i\phi}\big(\pm\partial_\theta + i\cot\theta\partial_\phi\big),$$
$$L^2 = -\hbar^2\big(\partial_\theta^2 + \cot\theta\partial_\theta + \csc^2\theta\partial_\phi^2\big).$$

We have to solve the following simultaneous differential equations:

$$L^2Y_{klm}(\theta,\phi) = \hbar^2\ell(\ell+1)Y_{klm}(\theta,\phi), \tag{5.2}$$
$$L_zY_{klm}(\theta,\phi) = \hbar mY_{klm}(\theta,\phi). \tag{5.3}$$

Consider first the differential equation (5.3). Writing out the differential operators, this equation reads

$$-i\hbar\frac{\partial}{\partial\phi}Y_{klm}(\theta,\phi) \stackrel{!}{=} \hbar mY_{klm}(\theta,\phi).$$

This equation is independent of θ and admits a straightforward solution of the form

$$Y_{klm}(\theta,\phi) = e^{i\phi m} F_{klm}(\theta),$$

where $\phi \in [0, 2\pi)$, and (for now) $F_{klm}(\theta)$ is arbitrary. Since we are on a sphere, the function Y_{klm} must be periodic in ϕ, with a period of 2π. From this we get the condition

$$e^{i(\phi+2\pi)m} \stackrel{!}{=} e^{i\phi m} \Rightarrow m \in \mathbb{Z}.$$

Therefore the half-integral values of m and hence ℓ are ruled out.

From here, insted of solving (5.2) directly, we proceed by using the ladder operators. We use the fact that $L_+|k,\ell,\ell\rangle = 0$ to get:

$$\begin{aligned}
0 &\stackrel{!}{=} L_+ Y_{k\ell\ell}(\theta,\phi) \\
\Leftrightarrow 0 &= \hbar e^{i\phi}(\partial_\theta + i\cot\theta\partial_\phi)e^{i\phi\ell}F_{k\ell\ell} \\
\Leftrightarrow 0 &= \hbar e^{i\phi(\ell+1)}(\partial_\theta + i\cot\theta(i\ell))F_{k\ell\ell} \\
\Leftrightarrow 0 &= \Big(\frac{d}{d\theta} - \ell\cot\theta\Big)F_{k\ell\ell}.
\end{aligned}$$

This is a first order ordinary differential equation. This means the solution space of this differential equation is one-dimensional. (TBD: no delocalization?) Indeed the function $F_{\ell\ell}(\theta) = (\sin\theta)^\ell$ solves the above differential equation as

$$\Big(\frac{d}{d\theta} - \ell\cot\theta\Big)(\sin\theta)^\ell = \ell(\sin\theta)^{\ell-1}\cos\theta - \ell\cos\theta(\sin\theta)^{\ell-1} = 0.$$

This solution also exists for every $\ell \in \mathbb{N}_0$. Since we have found *one* solution for each ℓ, and since the solution space is one dimensional, we have infact found *all* possible solutions for the above equation. To find all possible common eigenfunctions of L^2 and L_z, we just repeatedly apply L_- to $Y_{\ell\ell}$ for each ℓ.

More explicitly, we have

$$Y_\ell^\ell(\theta,\phi) = c(\sin\theta)^\ell e^{i\ell\phi}.$$

To find c we normalise

$$1 \stackrel{!}{=} \int \sin\theta d\theta d\phi\, c^2 ||Y_\ell^\ell(\theta,\phi)||^2.$$

This gives us the normalization

$$c = \frac{1}{2^\ell \ell!}\sqrt{\frac{(2\ell+1)!}{4\pi}}$$

The function Y_ℓ^m can now be obtained from this function by applying the ladder operator L_-. The case $\ell = 1$ is homework. For reference, one gets the following functions:

$$Y_\ell^m(\theta,\phi) = (-1)^m\sqrt{\frac{(2\ell+1)}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}}P_\ell^m(\cos\theta)e^{im\phi},$$

where $P_\ell^m(x)$ are the *associated Legendre polynomials*.

In summary: On the Hilbert space $L^2(S^2)$, the following functions on the unit sphere form an ONB

$$Y_\ell^m(\theta,\phi), \quad \text{with} \quad \ell \in \mathbb{N}_0;\ m = -\ell,\cdots,\ell.$$

These are the *spherical harmonics*. The rotations leave the subspaces $\mathcal{H}_\ell = \langle\{Y_\ell^m\}_{m=-\ell}^\ell\rangle$ invarient. The eigenfunctions in $\mathcal{H}_\ell$ with $\ell = 0,1,2,3,\cdots$ are called the $s,p,d,f,\cdots$ orbitals respectively.

5.1.1 Parity

Recall that the *parity operator* Π reflects functions at the origin:

$$(\Pi\psi)(\boldsymbol{x}) = \psi(-\boldsymbol{x}).$$

Rotations and reflections at the origin commute: $R(-\boldsymbol{x}) = -R\boldsymbol{x}$. Hence Π commmutes also with the generators L_i or rotations, and therefore, we could have added

$$\Pi Y_{klm} = \pi Y_{klm}$$

to the joint eigenvalue equations (5.2), (5.3). Because $\Pi^2 = \mathbb{1}$, the parity eigenvalues are $\pi \in \{\pm 1\}$. In fact, no harm was done when we left it out – because the Y_l^m actually all have a definite parity. You can see explicitly this for $Y_l^{m=l} = c(\sin\theta)^l e^{il\phi}$ by evaluating

$$Y_l^{m=l}(\pi-\theta,\phi+\pi) = \underbrace{(\sin(\pi-\theta))}_{\sin(\theta)}{}^l \underbrace{e^{il\pi}}_{(-1)^l} e^{i\phi} = (-1)^l\, Y_l^{m=l}(\theta,\phi).$$

And because Π commutes with L_-, the parity must the be same for the entire basis $\{Y_l^m\}_{m=-l}^l$ of the representation space $\mathcal{H}_{(l)}$.

5.2 Addition of angular momenta

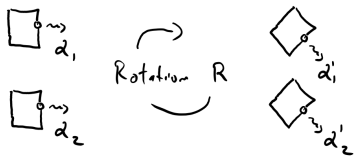
Classically, the *total angular momentum* of an n-body system is just the sum of the individual terms

$$\boldsymbol{L} = \sum_{i=1}^{n} \boldsymbol{L}^{(i)}.$$

Here, we'll work out the quantum theory of combined angular momenta.

For example, take a spin-l_1 and a spin-l_2 particle. Which values can the total angular momentum attain? Concretely, we'd like to know the eigenvalues of L^2 for $\boldsymbol{L} = \boldsymbol{L}^{(1)} + \boldsymbol{L}^{(2)}$.

5.2.1 Total angular momentum and rotations



Recall that we derived the operators representing conserved quantities from the generators of an associated symmetry. In that spirit, the *total angular momentum* of two systems is $i\hbar$ times the generator of rotations acting on each particle individually.

To work out a formula for the generator, consider two Hilbert spaces $\mathcal{H}_i$, and let $\boldsymbol{L}^{(i)}$ be the angular momentum observable of the i-th system. Under a rotation about the $\boldsymbol{e}_k$-axis, state vectors $|\alpha_i\rangle \in \mathcal{H}_i$ transform as

$$|\alpha_i\rangle = e^{\frac{\theta}{i\hbar}L_k^{(i)}}|\alpha_i\rangle.$$

Thus their product in $\mathcal{H}_{1,2}$ transforms as

$$|\alpha_1\rangle|\alpha_2\rangle \mapsto \big(e^{\frac{\theta}{i\hbar}L_k^{(1)}}|\alpha_1\rangle\big)\big(e^{\frac{\theta}{i\hbar}L_k^{(2)}}|\alpha_2\rangle\big). \tag{5.4}$$

Differentiating with respect to θ and using the product rule,

$$\partial_\theta|_0 \left(\left(e^{\frac{\theta}{i\hbar}L_k^{(1)}}|\alpha_1\rangle\right)\left(e^{\frac{\theta}{i\hbar}L_k^{(2)}}|\alpha_2\rangle\right)\right) = \left(\frac{1}{i\hbar}L_k^{(1)} + \frac{1}{i\hbar}L_k^{(2)}\right)|\alpha_1\rangle|\alpha_2\rangle.$$

> If two systems transform under a symmetry, then the generator on the joint system is the sum of the generators on the individual ones.

Multiplying by $i\hbar$, we obtain $\boldsymbol{L} = \boldsymbol{L}^{(1)} + \boldsymbol{L}^{(2)}$ for the total angular momentum.

Because (3.2) is a representation of the rotation group, we know from Chap. 4 that the components of $\boldsymbol{L}$ fulfill the commutator relations of a quantum angular momentum. To verify this directly, use that operators on different systems commute (3.5):

$$\begin{aligned}[L_1, L_2] &= [L_1^{(1)} + L_1^{(2)}, L_2^{(1)} + L_2^{(2)}] = [L_1^{(1)}, L_2^{(1)}] + [L_1^{(2)}, L_2^{(2)}] = i\hbar L_3^{(1)} + i\hbar L_3^{(2)} \\ &= i\hbar L_3 \qquad \text{(and cyclic)}.\end{aligned}$$

5.2.2 Eigenspaces of total angular momentum

[This is the most technical part of this entire course. I don't think I understood what really happened when I first encountered it as a student myself. So, give it a try, but don't beat yourself up if you end up confused on first (or second) reading.]

Assume that $\mathcal{H}_i = \mathcal{H}_{(l_i)}$ carries a spin-l_i-representation of $SO(3)$. In other words,

$$\mathcal{H}_i = \left\{ \sum_{m=-l_i}^{l_i} \psi_m |l, m\rangle \right\}.$$

Then $\boldsymbol{L} = \boldsymbol{L}^{(1)} + \boldsymbol{L}^{(2)}$ generates some representation of the rotation group on

$$\mathcal{H}_{1,2} = \mathcal{H}_{(l_1)} \otimes \mathcal{H}_{(l_2)} = \left\{ \sum_{m_1,m_2} \psi_{m_1,m_2} |l_1, m_1\rangle|l_2, m_2\rangle \right\}.$$

The question is: Which spin representations, i.e. which values of the quantum number l, occur on this joint Hilbert space?

As in the general theory, we need to find a joint eigenbasis

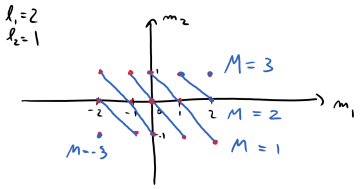
$$L^2|K, L, M\rangle = \hbar^2 L(L+1)|K, L, M\rangle, \qquad L_z|K, L, M\rangle = \hbar M|K, L, M\rangle. \tag{5.5}$$

of the total angular momentum operators on the tensor product Hilbert space. (It will turn out that there are no degeneracies, i.e. the label K will note be required).

Step 1: Find all L_z eigenvalues

Diagonalizing L_z is trivial – just take the product basis:

$$(L_z^{(1)} + L_z^{(2)})|l_1, m_1\rangle|l_2, m_2\rangle = \hbar(\underbrace{m_1 + m_2}_{=M})|l_1, m_1\rangle|l_2, m_2\rangle.$$



To visualize the possible values for M that can occur this way, draw the eigenvalues $-l_i \leq m \leq l_i$ in a plane. Then constant values of M lie on diagonals, as indicated.

Step 2: Find the subrepresentation with highest spin

In particular, the highest L_z-eigenvalue that occurs is $M = l_1+l_2$, with unique eigenvector $|l_1, m_1 = l_1\rangle|l_2, m_2 = l_2\rangle$. It must belong to a spin-$(l_1 + l_2)$-representation. Setting

$$|L, M\rangle := |l_1, m_1 = l_1\rangle|l_2, m_2 = l_2\rangle, \qquad L = M = l_1 + l_2, \tag{5.6}$$

and laddering downwards gives a basis

$$\{|L = l_1 + l_2, M\rangle\}_{M=-(l_1+l_2)}^{l_1+l_2}$$

of $\mathcal{H}_{(l_1+l_2)}$ inside of $\mathcal{H}_{(l_1)} \otimes \mathcal{H}_{(l_2)}$.

Step 3: Find the subrepresentation with second-highest spin

The next-largest eigenvalue is $M = l_1 + l_2 - 1$, which is two-fold degenerate, with basis

$$|l_1, l_1\rangle|l_2, l_2 - 1\rangle, \qquad |l_1, l_1 - 1\rangle|l_2, l_2\rangle.$$

This eigenspaces must contain the vector $|L = l_1+l_2, M = l_1+l_2-1\rangle$ already accounted for. There is therefore a vector $|\psi\rangle$ in this eigenspaces, unique up to a global phase, that is orthogonal to $|L = l_1 + l_2, M = l_1 + l_2 - 1\rangle$. It must belong to a spin-$(l_1 + l_2 - 1)$-representation. Setting

$$|L = l_1 + l_2 - 1, M = l_1 + l_2 - 1\rangle := |\psi\rangle$$

and laddering downwards gives a basis

$$\{|L = l_1 + l_2 - 1, M\rangle\}_{M=-(l_1+l_2-1)}^{l_1+l_2-1}$$

of $\mathcal{H}_{(l_1+l_2-1)} \subset \mathcal{H}_{(l_1+l_2)}^{\perp} \subset \mathcal{H}_{(l_1)} \otimes \mathcal{H}_{(l_2)}$.

Step 4: Iterate

We now iterate this construction until every vector in $\mathcal{H}_{(l_1)} \otimes \mathcal{H}_{(l_2)}$ is accounted for. Assuming (without loss of generality) that $l_1 \geq l_2$, this happens once we have included $L = l_1 - l_2$.

That's because

$$\dim(\mathcal{H}_{(l_1)} \otimes \mathcal{H}_{(l_2)}) = (2l_1 + 1)(2l_2 + 1) = \sum_{l=l_1-l_2}^{l_1+l_2} (2l + 1) = \sum_{l=l_1-l_2}^{l_1+l_2} \dim \mathcal{H}_{(l)}.$$

Proof.

$$\begin{aligned}
&\sum_{l=l_1-l_2}^{l_1+l_2} (2l + 1) = 2l_2 + 1 + 2 \sum_{l=l_1-l_2}^{l_1+l_2} l && \text{(there's } 2l_2 + 1 \text{ summands)}\\
=&2l_2 + 1 + (l_1 + l_2)(l_1 + l_2 + 1) - (l_1 - l_2 - 1)(l_1 - l_2) && \text{(Triangular number)}\\
=&2l_2 + 1 + 4l_1l_2 + 2l_1 = (2l_1 + 1)(2l_2 + 1). && \text{(group terms)}
\end{aligned}$$

We have thus partitioned the tensor product Hilbert space $\mathcal{H}_{(l_1)} \otimes \mathcal{H}_{(l_2)}$ into mutually orthogonal subspaces $\mathcal{H}_{(l)}$. Such a partition is called a *direct sum decomposition*, and stylized like this:

$$\mathcal{H}_{(l_1)} \otimes \mathcal{H}_{(l_2)} \simeq \bigoplus_{L=l_1-l_2}^{l_1+l_2} \mathcal{H}_{(l)}.$$

This particular direct sum is known as the *Clebsch-Gordan decomposition*.

Example. Treat an electron in the p-orbital ($l_1 = 1$) and its spin ($l_2 = \frac{1}{2}$). For brevity, write $|m_1\rangle|m_2\rangle$ for $|l_1, m_1\rangle|l_2, m_2\rangle$. From (5.6), the highest M-value belongs to

$$\Big|\frac{3}{2}, \frac{3}{2}\Big\rangle = \Big|1\Big\rangle\Big|\frac{1}{2}\Big\rangle.$$

Applying the laddering operator $L_- = L_-^{(1)} + L_-^{(2)}$ and using (5.1) gives

$$\Big|\frac{3}{2}, \frac{1}{2}\Big\rangle = \frac{1}{\sqrt{3}}\Big(\sqrt{2}\Big|0\Big\rangle\Big|\frac{1}{2}\Big\rangle + \Big|1\Big\rangle\Big|-\frac{1}{2}\Big\rangle\Big). \tag{5.7}$$

We could ladder downwards again to get the $M = -\frac{1}{2}$ eigenvector, but it's easier to argue by symmetry: Instead of with the highest M-value, we could have started with the lowest one, $M = -(l_1 + l_2)$ and laddered upwards. The analogue of the previous two steps gives

$$\Big|\frac{3}{2}, -\frac{3}{2}\Big\rangle = \Big|-1\Big\rangle\Big|-\frac{1}{2}\Big\rangle,$$
$$\Big|\frac{3}{2}, -\frac{1}{2}\Big\rangle = \frac{1}{\sqrt{3}}\Big(\sqrt{2}\Big|0\Big\rangle\Big|-\frac{1}{2}\Big\rangle + \Big|-1\Big\rangle\Big|\frac{1}{2}\Big\rangle\Big).$$

We have found a basis for the spin-$3/2$ subspace $\mathcal{H}_{(3/2)}$.

From the general theory, the $M = \frac{1}{2}$-subspace must contain a vector orthogonal to $\Big|\frac{3}{2}, \frac{1}{2}\Big\rangle$, unique up to phase. It belongs to a spin-$1/2$ space. Indeed, from (5.7), we can just read it off:

$$\Big|\frac{1}{2}, \frac{1}{2}\Big\rangle = \sqrt{\frac{2}{3}}\Big(-\frac{1}{\sqrt{2}}\Big|0\Big\rangle\Big|\frac{1}{2}\Big\rangle + \Big|1\Big\rangle\Big|-\frac{1}{2}\Big\rangle\Big).$$

(Here, we have followed a general convention, and chosen the phase such that the term with the highest value of m_1 gets a positive coefficient). Laddering down, or arguing by symmetry,

$$\Big|\frac{1}{2}, -\frac{1}{2}\Big\rangle = \sqrt{\frac{2}{3}}\Big(-\frac{1}{\sqrt{2}}\Big|0\Big\rangle\Big|-\frac{1}{2}\Big\rangle + \Big|-1\Big\rangle\Big|\frac{1}{2}\Big\rangle\Big).$$

Let's count dimensions to double-check that we're done:

$$\dim \mathcal{H}_{(1)} \otimes \mathcal{H}_{(1/2)} = 3 \times 2 = 4 + 2 = \dim \mathcal{H}_{(3/2)} + \dim \mathcal{H}_{(1/2)}.$$

The important case $l_1 = l_2 = \frac{1}{2}$ is homework.

Remark. The expansion coefficients $\langle L, M|l_1, m_1\rangle|l_2, m_2\rangle$ which express the common eigenbasis of L^2, L_z in terms of the common eigenbasis of $(L^{(1)})^2, L_z^{(1)}, (L^{(2)})^2, L_z^{(2)}$ are called the *Clebsch-Gordan coefficients*. They are used in atomic physics, and tabulated online. The example above can be summarized as:

M	m_1	m_2	$\langle 3/2, M\|1, m_1\rangle\|1/2, m_2\rangle$
$3/2$	1	$1/2$	1
$1/2$	1	$-1/2$	$\sqrt{2/3}$
$1/2$	-1	$1/2$	$\sqrt{1/3}$
$-1/2$	0	$-1/2$	$\sqrt{2/3}$
$-1/2$	-1	$1/2$	$\sqrt{1/3}$
$-3/2$	-1	$-1/2$	1

M	m_1	m_2	$\langle 1/2, M \vert 1, m_1\rangle \vert 1/2, m_2\rangle$
$1/2$	1	$-1/2$	$\sqrt{2/3}$
$1/2$	0	$1/2$	$-\sqrt{1/3}$
$-1/2$	0	$-1/2$	$-\sqrt{1/3}$
$-1/2$	-1	$1/2$	$\sqrt{2/3}$

Chapter 6

Central potentials

6.1 Recap: Classical theory

Consider a classical particle with Hamilton function

$$H(\boldsymbol{q}, \boldsymbol{p}) = \frac{\|\boldsymbol{p}\|^2}{2m} + V(\|\boldsymbol{q}\|)$$

invariant under rotations. Let $r = \|\boldsymbol{q}\|$ be the distance to the origin, $\hat{\boldsymbol{q}} = \boldsymbol{q}/r$ the *radial vector*, and $p_r = \hat{\boldsymbol{q}} \cdot \boldsymbol{p}$ the *radial momentum*. Decompose $2m$ times the kinetic energy

$$\|\boldsymbol{p}\|^2 = (\hat{\boldsymbol{r}} \cdot \boldsymbol{p})^2 + \|\hat{\boldsymbol{r}} \times \boldsymbol{p}\|^2 = p_r^2 + \frac{\|\boldsymbol{L}\|^2}{r^2} \tag{6.1}$$

in terms of a "radial" and a "rotational" contribution.

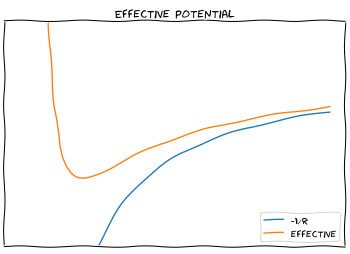
Next, Noether's theorem says that rotational invariance implies that $\boldsymbol{L}$ is conserved. Adjusting coordinates if necessary, we may assume that $\boldsymbol{L} = \ell \boldsymbol{e}_z$ is proportional to $\boldsymbol{e}_z$. Using that $\boldsymbol{L} = \boldsymbol{x} \times \boldsymbol{p}$, it follows that the dynamics takes place in the x–y-plane. Denote polar coordinates in the plane by (r, ϕ). You can trust me that the (r, ϕ, p_r, L_z) are canonical coordinates (or better don't trust me and use the generating function $G_3 = -\boldsymbol{x}(r, \phi) \cdot \boldsymbol{p}$ to verify). We can thus write the Hamiltonian as

$$H(r, \phi, p_r, L_z) = \frac{p_r^2}{2m} + \frac{L_z^2}{2mr^2} + V(r).$$

In particular, the equations of motions for r do not depend on ϕ, and are generated by

$$H(r, p_r) = \frac{p_r^2}{2m} + V_{\text{eff}}(r), \qquad V_{\text{eff}}(r) := \frac{\ell^2}{2mr^2} + V(r), \tag{6.2}$$

which corresponds to a particle moving in one dimension under the *effective potential* V_{eff}.



Gravitational and Coulomb potentials are proportional to $-1/r$. If that were a one-dimensional potential, it would suck everything right into the center.

Fortunately for planet-dwellers (and electrons), we live in three dimensions. If you start off with non-zero angular momentum, the effective potential includes the *centrifugal barrier term* $\ell^2/(2mr^2)$, which keeps us in orbit.

6.2 Stationary Schrödinger equation for central potentials

The Hamilton operator for a point particle in 3D under a central potential $V(r)$ is

$$H = \frac{\|\boldsymbol{P}\|^2}{2m} + V(r) = -\frac{\hbar^2}{2m}\Delta + V(r).$$

Much of the classical analysis of the last section carries over.

6.2.1 Radial angular momentum

In QM, too, we can decompose the kinetic energy into a "radial" and a "rotational" contribution. The most straight-forward way is to convert the Laplacian and the squared-angular momentum operator to spherical coordinates (or to look up the expressions):

$$\begin{aligned}\Delta &= \frac{1}{r}\partial_r^2\, r + \frac{1}{r^2\sin\theta}\partial_\theta\left(\sin\theta\partial_\theta\right) + \frac{1}{r^2\sin^2\theta}\partial_\phi^2,\\ L^2 &= -\hbar^2\left(\frac{1}{\sin\theta}\partial_\theta\left(\sin\theta\partial_\theta\right) + \frac{1}{\sin^2\theta}\partial_\phi^2\right).\end{aligned}$$

The angular components of $-\hbar^2/(2m)\Delta$ and $L^2/(2mr^2)$ are seen to be identical, so that

$$\frac{P^2}{2m} = -\frac{\hbar^2}{2m}\frac{1}{r}\partial_r^2\, r + \frac{L^2}{2mr^2}. \tag{6.3}$$

Radial momentum. The first term looks complicated. Classically, it's p_r^2, where $p_r = \hat{\boldsymbol{q}}\cdot\boldsymbol{p} = \boldsymbol{p}\cdot\hat{\boldsymbol{q}}$ is the radial component of the angular momentum.

Which operator corresponds to p_r? In Sec. 1.3, we mentioned that quantization is ambiguous if operator ordering makes a difference. That happens here, because "putting the radial vector to the right of the momentum gives an additional term from the product rule". Details:

$$(\hat{\boldsymbol{R}}\cdot\boldsymbol{P})\psi = -i\hbar\partial_r\psi,$$

where r is the radial component of spherical coordinates. Ordering operators differently:

$$(\boldsymbol{P}\cdot\hat{\boldsymbol{R}})\psi = -i\hbar\sum_i \partial_i\left(\frac{x_i}{r}\psi\right) = -i\hbar\sum_i\left(\frac{1}{r}\psi - \frac{x_i^2}{r^3}\psi + \frac{x_i}{r}\partial_i\psi\right) = -i\hbar\left(\partial_r + \frac{2}{r}\right)\psi.$$

We might try our luck with the mean of the two possibilities, setting

$$P_r = \frac{1}{2}(\hat{\boldsymbol{R}}\cdot\boldsymbol{P} + \boldsymbol{P}\cdot\hat{\boldsymbol{R}}) = -i\hbar\left(\partial_r + \frac{1}{r}\right). \tag{6.4}$$

Squaring this one gives

$$P_r^2\psi = -\hbar^2\left(\partial_r^2\psi + \partial_r\left(\frac{1}{r}\psi\right) + \frac{1}{r}\partial_r\psi + \frac{1}{r^2}\psi\right) = -\hbar^2\left(\partial_r^2\psi + \frac{2}{r}\partial_r\psi\right), \tag{6.5}$$

which is $2m$ times the operator on the left of (6.3).

$$-\hbar^2\frac{1}{r}\partial_r^2\,(r\psi) = -\hbar^2\frac{1}{r}\partial_r(r\partial_r\psi + \psi) = -\hbar^2\frac{1}{r}(r\partial_r^2\psi + 2\partial_r\psi) = -\hbar^2\left(\partial_r^2\psi + \frac{2}{r}\partial_r\psi\right).$$

This justifies calling (6.4) the *radial momentum operator*.

6.2.2 Reduction to one-dimensional problem

From the last section,

$$H = \frac{P_r^2}{2m} + \frac{L^2}{2mr^2} + V(r).$$

Because the L_k act only on angular coordinates, L^2 and L_z commute with H. We can therefore look for solutions to this system of joint eigenvalue equations

$$H\psi = E\psi, \tag{6.6}$$
$$L^2\psi = \hbar^2 l(l+1)\psi, \tag{6.7}$$
$$L_z\psi = \hbar m\psi. \tag{6.8}$$

Write $\psi = \psi(r, \theta, \phi)$ in spherical coordinates. Then we know from Chap. 5 that, for every value of r, (6.7) and (6.8) together mean that $\psi(r, \theta, \phi)$ is proportional to $Y_l^m(\theta, \phi)$. Thus

$$\psi(r, \theta, \phi) = Y_l^m(\theta, \phi)R(r)$$

where $R(r)$ are the yet-to-be-determined proportionality constants. Then Eq. (6.6) gives

$$\left(-\frac{\hbar^2}{2m}\frac{1}{r}\partial_r^2(rR(r)) + \frac{\hbar^2 l(l+1)}{2mr^2} + V(r)\right) Y_l^m(\theta, \phi)R(r) = E\, Y_l^m(\theta, \phi)R(r).$$

To clean up, set $u(r) := rR(r)$, divide by Y_l^m and multiply by r, to get the *radial equation*

$$H_l u_{l,k}(r) = E_{l,k}\, u_{l,k}(r), \qquad H_l = -\frac{\hbar^2}{2m}\partial_r^2 + \frac{\hbar^2 l(l+1)}{2mr^2} + V(r), \tag{6.9}$$

where we have introduced labels $E_{l,k}, u_{l,k}$ for the k-th eigenvalue and eigenfunction of H_l. As in the classical case, this is formally equivalent to the time-independent Schrödinger equation for a particle moving in 1D under an l-dependent effective potential.

Remark. Taking the Jacobian determinant of spherical coordinates into account,

$$\begin{aligned}\|\psi\|_2^2 &= \iiint |\psi(r, \theta, \phi)|^2\, r^2 \sin\theta \,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi \\ &= \int_0^\infty |R(r)|^2 r^2 \,\mathrm{d}r \iint |Y_l^m(\theta, \phi)|^2 \sin\theta \,\mathrm{d}\theta\,\mathrm{d}\phi = \int_0^\infty |u(r)|^2 \,\mathrm{d}r.\end{aligned} \tag{6.10}$$

In particular, $|u(r)|^2$ is the probability density at distance r from the center.

For most potentials $V(r)$, the radial equations can't be solved explicitly, but there are some general properties, discussed next.

Bound states versus scattering states

Assume that $\lim_{r\to\infty} V(r) = 0$. Then H_l is approximately free asymptotically, and thus eigenfunctions satisfy

$$u(r \to \infty) \propto e^{\pm kr}, \qquad k = \sqrt{\frac{-E\hbar}{2m}}.$$

If $E < 0$, then k is positive. The e^{+kr}-solution diverges and is thus unphysical. The e^{-kr} one gives state whose probability of being found at distance r from the center decreases exponentially. These are *bound states*, and we will focus on this class below.

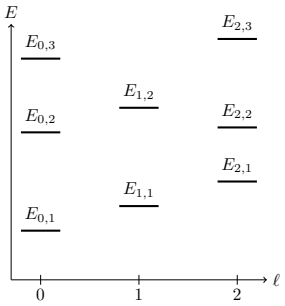
If $E > 0$, then k is imaginary, so $u(r)$ oscillates with constant amplitude as $r \to \infty$, a *spherical wave*. In the Advanced QM course, these solutions are used to study *scattering processes*.

Relations between the radial equations

The Hamiltonians H_l depend on the angular momentum quantum number l. The difference

$$H_{l+1} - H_l = \frac{\hbar^2}{2mr^2}(l+1) \tag{6.11}$$

is strictly positive. The ground state energies $E_{l,1}$ thus increase with l.

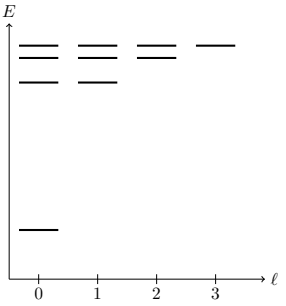


In general, that's the only relation between the eigenvalues of different H_l, so we expect a level diagram to look something like the one sketched.

It *could* happen that two eigenvalues of different Hamiltonians agree. This would then be called an *accidental degeneracy*, because there's really no systematic reason to expect it.

I'm bringing this up because the hydrogen level diagram looks this *this*.

The k-th eigenvalue of H_l is the $(k-1)$st of H_{l+1}. That's not an accident, that's a car crash! Any method for solving the $1/r$-potential better identifies a concrete mechanism that causes this surprising behavior.



6.3 The Coulomb potential

For the important case $V(r) = -\kappa/r$, the radial equation can be solved explicitly. We focus on the bound states, i.e. $E < 0$.

6.3.1 Switching to problem-adapted units

Define problem-adapted units (their names will be justified later)

$$a_0 := \frac{\hbar^2}{m\kappa} \quad \text{Bohr radius,} \qquad \rho := r/a_0 \quad \text{distance in units of } a_0,$$
$$E_I := \frac{m\kappa^2}{2\hbar^2} \quad \text{ionization energy,} \qquad \epsilon := E/E_I \quad \text{energy in units of } E_I.$$

Then a boring substitution...

$$\begin{aligned} H_l &= -\frac{\hbar^2}{2m}\partial_r^2 + \frac{\hbar^2 l(l+1)}{2mr^2} - \frac{\kappa}{r} = -\frac{\hbar^2}{ma_0^2}\frac{1}{2}\partial_r^2 + \frac{\hbar^2}{ma_0^2}\frac{l(l+1)}{2\rho^2} - \frac{\kappa}{a_0}\frac{1}{\rho} \\ &= 2E_I\left(-\frac{1}{2}\partial_\rho^2 + \frac{l(l+1)}{2\rho^2} - \frac{1}{\rho}\right) \end{aligned}$$

...shows that dividing the radial equation by $2E_I$ gives the dimensionless variant

$$h_l\, u_{l,k} = \frac{1}{2}\epsilon_{l,k}\, u_{l,k}, \qquad h_l = -\frac{1}{2}\partial_\rho^2 + \frac{l(l+1)}{2\rho^2} - \frac{1}{\rho}, \tag{6.12}$$

where $u_{l,k}$ is the k-th eigenstate of h_l.

There's various options for solving these eigenvalue equations:

- Treat them as generic differential equations and use standard techniques: split off the asymptotic behavior, make a power series ansatz for the remainder, work out conditions on the coefficients... Stuff like that. Teaching it this way is the grown-up thing to do. Compatible with any textbook and conceptually straight-forward.
  But, boy does it *suck*! It involves zero exciting ideas, is lengthy, and doesn't offer any insights into "why" the accidental degeneracies occur.
- In 1935 (!) Vladimir Fock wrote a paper (in perfect German) relating the Coulomb problem in $\mathbb{R}^3$ to a free particle on the 3-sphere (i.e. the one in $\mathbb{R}^4$). He showed that the Coulomb Hamiltonian is nothing but the spherical Laplace operator in thin disguise, and the hydrogen eigenfunctions are just the (hyper-)spherical harmonics. It's breathtakingly beautiful... ...but I have never found a way to tell the story so that it would work for a 2nd year class. ☹ Oh dear. Back to the power series...
- ...or so it was until, well into my teaching career, I become aware of a trick based on... checks notes... *supersymmetric quantum mechanics*. What's not to love?

6.3.2 Supersymmetric quantum mechanics

Supersymmetry (*SUSY* if you want to sound cutesy) was a big thing in high energy physics from the early 1970s on. It pairs every elementary particle with a *supersymmetric partner*, always one Boson and one Fermion, such that their Hamiltonians share the same eigenvalues. This may sound like random words, but it's actually an essentially unique way to work around a known limitation of relativistic quantum field theory. Taking it seriously made a lot of sense. The hype fizzled out in the 2010s, when the LHC failed to uncover so much as a shred of evidence for supersymmetry being implemented for actual particles.

Oh well.

The math, however, remains sound. And if you ~~dumb it down enough~~ repeatedly pass to simpler models that retain the key ideas, you'll eventually arrive at *supersymmetric quantum mechanics*. It's a somewhat niche field,[1] but I feel its solution of the hydrogen atom is under-appreciated.

Supersymmetric partners

Our development rests on the linear-algebraic triviality that for any operator A, the non-zero eigenvalues of

$$H_- := A^\dagger A \qquad \text{and of} \qquad H_+ := AA^\dagger$$

are the same (their roots being known as the *singular values* of A). Indeed, explicitly,

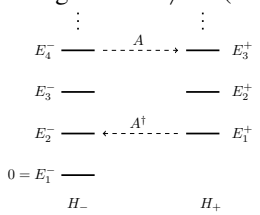
$$\begin{aligned} H_-|\psi_E^-\rangle = E|\psi_E^-\rangle \qquad &\Rightarrow \qquad H_+\, A|\psi_E^-\rangle = AA^\dagger A|\psi_E^-\rangle = E\, A|\psi_E^-\rangle, \\ H_+|\psi_E^+\rangle = E|\psi_E^+\rangle \qquad &\Rightarrow \qquad H_-\, A^\dagger|\psi_E^+\rangle = A^\dagger AA^\dagger|\psi_E^+\rangle = E\, A^\dagger|\psi_E^+\rangle. \end{aligned}$$

[1] In particular, it isn't mentioned in most QM textbooks. This section draws from Chapter 3 of *Supersymmetric Methods* by Junker, and on this article.

In words: If $|\psi_E^-\rangle$ is a H_--eigenvector with eigenvalue E, then $|\psi_E^+\rangle := A|\psi_E^-\rangle$ solves the eigenvalue equation for H_+. Thus if $A|\psi_E^-\rangle$ isn't 0, it is an H_+-eigenvector. But

$$\|A|\psi_E^-\rangle\|^2 = \langle\psi_E^-|A^\dagger A|\psi_E^-\rangle = E\,\|\psi_E^-\|^2, \tag{6.13}$$

so we're good if $E \neq 0$. (In this general setting, one can't say much about $E = 0$).



Below, we'll work with non-degenerate, non-negative Hamiltonians where 0 is an eigenvalue of H_- but not of H_+. The resulting level diagram is pictured.

In analogy to the ladder operators of the harmonic oscillator, A then maps a k-th eigenvector to a $(k-1)$st – however, it also switches from H_- to H_+.

To apply this to QM, we need A's such that $H_\pm$ can be interpreted as Hamiltonians. This is achieved by choosing a function v, called the *supersymmetric potential*, and setting

$$A := \frac{1}{\sqrt{2}}(\partial_x + v) \quad \Rightarrow \quad A^\dagger = \frac{1}{\sqrt{2}}(-\partial_x + v).$$

Then the $H_\pm$ take the form of one-dimensional Hamiltonians, as hoped for:

$$H_\pm = \frac{1}{2}\left(-\partial_x^2 + v^2 \pm [\partial_x, \Phi]\right) = -\frac{1}{2}\partial_x^2 + \frac{1}{2}v^2 \pm \frac{1}{2}v',$$

with a (dimensionless) kinetic energy term $-\frac{1}{2}\partial_x^2$ and potential $V_\pm = \frac{1}{2}(v^2 \pm v')$.

Example. For $v = x$, we recover the ladder operators (1.42) of the harmonic oscillator, and

$$H_\pm = -\frac{1}{2}\partial_x^2 + \frac{1}{2}x^2 \pm \frac{1}{2} = H_{\text{HO}} \pm \frac{1}{2}$$

are shifted versions of the (dimensionless) harmonic oscillator Hamiltonian H_{HO}.

Groundstates

For every state $|\psi\rangle$, we have that

$$\langle\psi|H_-|\psi\rangle = \langle\psi|A^\dagger A|\psi\rangle = \|A|\psi\rangle\|^2 \geq 0,$$

which implies that H_- has non-negative eigenvalues (a similar argument holds for H_+). Also, the eigenvalue 0 occurs for H_- iff there is a state ψ such that

$$0 = \langle\psi|H_-|\psi\rangle = \|A|\psi\rangle\|^2 \quad \Leftrightarrow \quad A\psi = 0 \quad \Leftrightarrow \quad \partial_x\psi(x) = -v(x)\psi(x).$$

Being of first order, this differential equation is trivial to solve:

$$\psi(x) = e^{-\int^x v(x')\,\mathrm{d}x'} \quad \text{(indefinite integral)}. \tag{6.14}$$

It may happen that $\psi(x)$ diverges as $x \to \pm\infty$. Thus, our conclusion is that 0 is an eigenstate of H_- iff the solution of (6.14) can be interpreted as a quantum state.

Example. Continuing the treatment of the harmonic oscillator,

$$v = x \quad \Rightarrow \quad -\int^x v(x')\mathrm{d}x' = -\frac{1}{2}x^2 + C \quad \Rightarrow \quad \psi(x) \propto e^{-\frac{1}{2}x^2},$$

which is normalizable, and indeed the correct ground state wave function.

Application to the Coulomb problem

The effective Hamiltonian h_l of Eq. (6.12) is connected to the supersymmetric potential

$$v_l = \frac{1}{l+1} - \frac{l+1}{\rho}.$$

To see this, compute the two associated partner potentials:

$$\begin{aligned} V_\pm &= \frac{1}{2}v_l^2 \pm \frac{1}{2}v_l' = \frac{1}{2}\left(\frac{1}{(l+1)^2} - \frac{2}{\rho} + \frac{(l+1)^2}{\rho^2} \pm \frac{l+1}{\rho^2}\right) \\ &= \frac{(l+1\pm 1)(l+1)}{2\rho^2} - \frac{1}{\rho} + \frac{1}{2(l+1)^2}. \end{aligned}$$

Now comes the key insight! Comparison with (6.12) shows that

$$H_- = h_l + \frac{1}{2(l+1)^2}, \qquad H_+ = h_{l+1} + \frac{1}{2(l+1)^2}. \tag{6.15}$$

Thus: *Up to a shift of energy levels, "neighboring" radial Hamiltonians h_l, h_{l+1} are supersymmetric partners!* Let's record some simple consequences:

- Zero is an eigenvalue of H_-. Indeed,

$$\int^\rho \frac{l+1}{\rho'} - \frac{1}{l+1}\mathrm{d}\rho' = (l+1) - \frac{\rho}{l+1}\ln\rho + C \quad \Rightarrow \quad u_{l,1}(x) \propto \rho^{l+1}e^{-\frac{\rho}{l+1}},$$

  which is certainly normalizable. Also, we now have the ground state of h_l!

- From (6.15) we then find the ground state energy of h_l:

$$0 = H_- u_{l,1} = h_l\, u_{l,1} + \frac{1}{2(l+1)^2}\, u_{l,1} \qquad \Rightarrow \qquad \epsilon_{l,1} = -\frac{1}{(l+1)^2}.$$

- Because $H_+ = H_- + \frac{l+1}{\rho^2}$ and the added term is strictly positive, 0 is *not* an eigenvalue of H_+. We thus encounter the situation sketched in the level diagram on this page: $E_k^- = E_{k-1}^+$, meaning here that $\epsilon_{l,k} = \epsilon_{l+1,k-1}$.

And that's it. We can sweep the board without any further calculations (c.f. Fig. 6.1)!

- We already know all the ground state energies $\epsilon_{l,1}$ and wave functions $u_{l,1}$.

- But this implies the first excited energies, because $\epsilon_{l,2}$ is the ground state energy $\epsilon_{l+1,1}$ of the partner. Iterating:

$$\epsilon_{l,k} = \epsilon_{l+1,k-1} = \epsilon_{l+2,k-2} = \cdots = \epsilon_{l+k-1,1} = -\frac{1}{(l+k)^2} =: -\frac{1}{n^2}.$$

  Thus, the energy only depends on the sum $n = l + k \in \{1, 2, \dots\}$, called the *principal quantum number* n.

- Each value of n can be realized in n different ways, namely for $l = 0, \dots, n-1$. Recall that the full Hamiltonian H of Eq. (6.6) has $2l+1$ eigenstates associated with each value of l. Thus, the degeneracy of the n-th level of H is

$$\sum_{l=0}^{n-1} 2(l+1) = n + 2\sum_{l=0}^{n-1} l = n + n(n-1) = n^2.$$

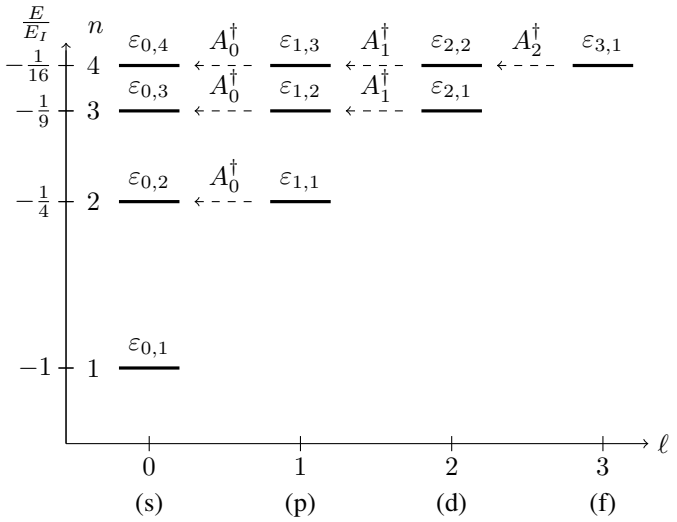


Figure 6.1: Level diagram of the effective Hamiltonians h_l for the Coulomb problem. The "accidental" degeneracies are explained by the fact that neighboring effective Hamiltonians are supersymmetric partners. Energy axis not to scale.

- Eigenfunctions are obtained analogously. Write the generalized annihilation operator for the superpotential v_l as $A_l := \frac{1}{\sqrt{2}}\left(\partial_x + v_l\right)$. Then

$$u_{l,k} \propto A_l^\dagger\, u_{l+1,k-1} \propto A_l^\dagger A_{l+1}^\dagger\, u_{l+2,k-2} \propto \cdots \propto A_l^\dagger A_{l+1}^\dagger \cdots A_{l+k-2}^\dagger\, u_{l+k-1,1}.$$

Normalization. If you're feeling masochistic, you can work out the normalization constants.

For the ground states, *Mathematica* tells me that

$$u_{l,1}(\rho) = N_{l,1}\, \rho^{l+1} e^{-\rho/(l+1)}, \qquad N_{l,1} = \sqrt{\frac{2^{2(l+1)}}{(l+1)^{4+2l}(2l+1)!}}$$

is normalized (w.r.t. (6.10)).

To extend this to all states, we need to preserve normalization under laddering. Arguing as in Eq. (6.13), if $|\psi_E^+\rangle$ is a normalized eigenvector of H_+ with eigenvalue E, then $E^{-1/2}A^\dagger|\psi_E^+\rangle$ is also normalized. The connection with our application is made in Eq. (6.15). Specifically, choosing $|\psi_E^+\rangle = |u_{l+1,k-1}\rangle$, then

$$E = \epsilon_{l,k} + \frac{1}{2(l+1)^2} = \frac{1}{2(l+1)^2} - \frac{1}{2(l+k)^2} = \frac{1}{2(l+1)^2} - \frac{1}{2n^2}.$$

Therefore, the normalized wave functions are

$$u_{l,k} = \prod_{l'=l}^{l+k-2} \left(\left(\frac{1}{2(l+1)^2} - \frac{1}{2n^2} \right)^{-1/2} A_{l'}^\dagger \right) u_{l+k-1,1},$$

where the factors must be order so that l' increases from left to right.

Be sure to multiply with an additional $a_0^{-1/2}$ when switching from ρ to r.

Some complete solutions: TBD.

6.3.3 Summary

The ground states ψ_{nlm} of the $1/r$-potential are thus labeled by the principal quantum number $n = 1, 2, \ldots$, the angular momentum quantum number $l = 0, \ldots, n-1$, and the magnetic quantum number $m = -l, \ldots, l$. Then:

$$\psi_{nlm}(r, \theta, \phi) = N_{nl} \frac{1}{r} u_{l,n-l}(r) Y_l^m(\theta, \phi), \qquad E_n = -\frac{E_I}{n^2}, \tag{6.16}$$

where N_{nl} is a normalization constant.

If you're feeling masochistic, you can be more explicit. For example, recall that the ground state of H_l is

$$u_{l,1}(r) = N_{l+1,l} \left(\frac{r}{a_0} \right)^{l+1} e^{-\frac{r}{a_0(l+1)}}. \tag{6.17}$$

(The normalization constant in (6.16) is the same as the one in (6.17). That's because $|\psi_{nlm}(r, \theta, \phi)|^2$ must be integrated in spherical coordinates, with volume element $r^2 \sin\theta \, \mathrm{d}r \, \mathrm{d}\theta \, \mathrm{d}\phi$. In the radial integral for $|\psi|^2$, the $1/r$ factor in (6.16) exactly kills the r^2 of the volume element, leaving us with $\int_0^\infty |u_{l,r}|^2$). Anyways, Mathematica gives

(whatever...).

There's pretty visualizations online if that floats your boat.

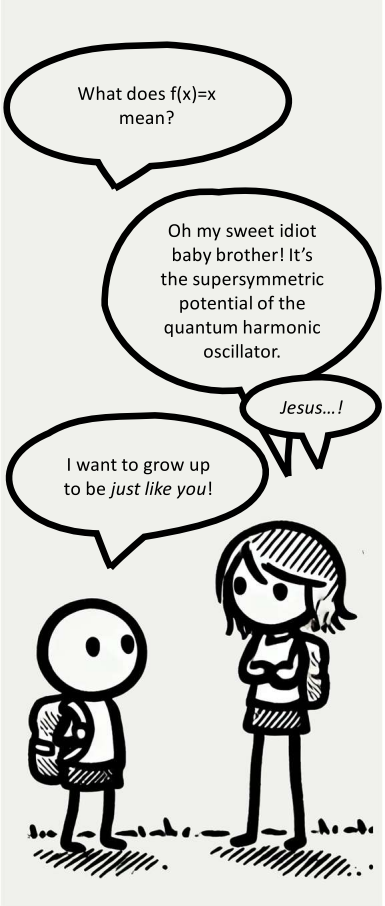
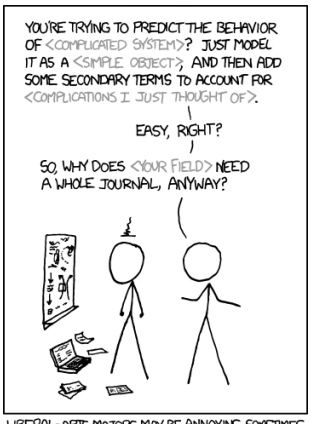


Figure 6.2: I don't know why I did this.

Chapter 7

Perturbation Theory



Obligatory xkcd. This is you (once you understand this chapter).

Most operators cannot be explicitly diagonalized. In perturbation theory, the idea is to start from a Hamiltonian H_0 for which the eigenvalue problem $H_0|\psi_i\rangle = E_i|\psi_i\rangle$ *is* solvable, and to make statements about the "perturbed" operator

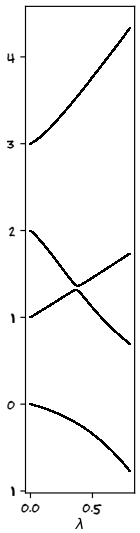
$$H = H_0 + \lambda W.$$

The perturbation W (a "secondary term" as per the comic) is arbitrary, but we only aim for approximations that work for "small" values of λ.

It's not the most conceptually inspiring chapters of QM, but important in practice!

7.1 Stationary perturbation theory

For now, assume that all eigenvalues of H_0 are non-degenerate and discrete. Starting from one solution $H_0|\psi_i\rangle = E_i|\psi_i\rangle$, our goal is to track the evolution of the eigenvalues $E_i(\lambda)$ and eigenvectors $|\psi_i(\lambda)\rangle$ of $H(\lambda) = H_0 + \lambda W$ as λ increases. (Click for numerics).



Here, $E_i(\lambda)$ is the eigenvalue of $H(\lambda)$ that connects to E_i as $\lambda \to 0$. It can happen that the graphs of two eigenvalues cross for larger values of λ – the problem then ceases to be well-defined.

The eigenvector $|\psi_i(\lambda)\rangle$ is defined via

$$H(\lambda)|\psi_i(\lambda)\rangle = E_i(\lambda)|\psi_i(\lambda)\rangle \quad \text{(eigenvalue equation)}, \tag{7.1}$$

$$\langle\psi_i|\psi_i(\lambda)\rangle = 1 \quad \text{("normalization" condition)}. \tag{7.2}$$

Warning: The norm of $|\psi_i(\lambda)\rangle$ is usually larger than one:

$$\langle\psi_i(\lambda)|\psi_i(\lambda)\rangle = \sum_j |\langle\psi_j|\psi_i(\lambda)\rangle|^2 = 1 + \sum_{j\neq 1} |\langle\psi_j|\psi_i(\lambda)\rangle|^2.$$

The basic assumption of perturbation theory is that both the eigenvalue and the eigenvector can be expanded as a power series in λ:

$$E_i(\lambda) = \sum_{n=0}^{\infty} E_i^{(n)}\,\lambda^n \qquad |\psi_i(\lambda)\rangle = \sum_{n=0}^{\infty} |\psi_i^{(n)}\rangle\,\lambda^n.$$

The difficulty lies in finding the scalar and vectorial coefficients $E_i^{(n)}$ and $|\psi_i^{(n)}\rangle$.

The central equation of stationary perturbation theory arises by plugging this ansatz into (7.1):

$$\big(H_0 + \lambda W\big)\left(\sum_{n=0}^{\infty} |\psi_i^{(n)}\rangle\,\lambda^n\right) - \left(\sum_{n=0}^{\infty} E_i^{(n)}\,\lambda^n\right)\left(\sum_{n=0}^{\infty} |\psi_i^{(n)}\rangle\,\lambda^n\right) = 0. \tag{7.3}$$

A power series is identically equal to 0 if and only if the coefficients for each order λ^n vanish. For the rest of the section, we'll analyze these conditions order-by-order.

7.1.1 First-order energy correction

Order-0: Setting the coefficient of λ^0 in (7.3) equal to 0 gives

$$H_0|\psi_i^{(0)}\rangle - E_i^{(0)}|\psi_i^{(0)}\rangle = 0 \quad \Rightarrow \qquad E_i^{(0} = E_i, \quad |\psi_i^{(0)}\rangle = |\psi_i\rangle.$$

Order-1: For the coefficients of λ^1:

$$\begin{aligned} & H_0|\psi_i^{(1)}\rangle + W|\psi_i^{(0)}\rangle - E_i^{(1)}|\psi_i^{(1)}\rangle - E_i^{(0)}|\psi_i^{(0)}\rangle = 0 \\ \Rightarrow \qquad & \big(H_0 - E_i^{(0)}\big)|\psi_i^{(1)}\rangle + \big(W - E_i^{(1)}\big)|\psi_i^{(0)}\rangle = 0. \end{aligned} \tag{7.4}$$

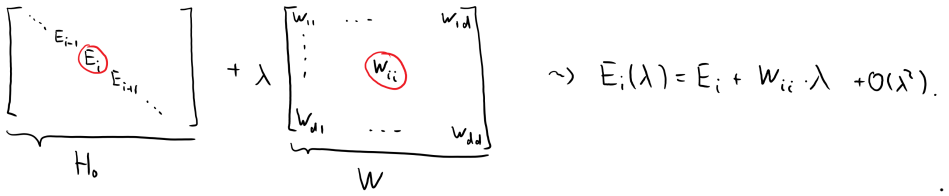
Multiply $\langle\psi_i|$ from the left:

$$\underbrace{\langle\psi_i|H_0 - E_i^{(0)}|\psi_i^{(1)}\rangle}_{\langle\psi_i|E_i - E_i^{(0)}|\psi_i^{(1)}\rangle = 0} + \langle\psi_i|W - E_i^{(1)}|\psi_i^{(0)}\rangle = 0, \tag{7.5}$$

hence $E_i^{(1)} = \langle\psi_i|W|\psi_i\rangle$. An embarrassingly large part of physics is built on the resulting

> First order energy correction formula: $E_i(\lambda) = E_i + \lambda\langle\psi_i|W|\psi_i\rangle + O(\lambda^2)$.

In matrix notation with respect to the H_0-eigenbasis:



Example. The anharmonic oscillator (in units where $\hbar = \omega = m = 1$). Take

$$H_0 = \frac{1}{2}(P^2 + X^2), \qquad W = X^4.$$

To compute the matrix element of W in the ground state, write $X = \frac{1}{\sqrt{2}}(a + a^\dagger)$, so that

$$E_0^{(1)} = \frac{1}{4}\langle 0|(a+a^\dagger)^4|0\rangle = \frac{1}{4}\langle 0|a(aa^\dagger + a^\dagger a)a^\dagger|0\rangle = \frac{1}{4}(\sqrt{1\cdot 2\cdot 2\cdot 1} + \sqrt{1\cdot 1\cdot 1\cdot 1}) = \frac{3}{4}.$$

Thus, the ground state energy increases due to the anharmonic term by

$$E_0(\lambda) = \frac{1}{2} + \frac{3}{4}\lambda + O(\lambda^2).$$

For $n > 0$, there are six non-zero terms.

7.1.2 First-order eigenstate correction

We have obtained the energy correction by projecting Eq. (7.4) onto $|\psi_i\rangle$. We have not yet used the information contained in the projections onto the orthogonal directions. So, let's project onto $|\psi_j\rangle$, for $j \neq i$:

$$\underbrace{\langle\psi_j|H_0 - E_i^{(0)}|\psi_i^{(1)}\rangle}_{(E_j - E_i)\langle\psi_j|\psi_i^{(1)}\rangle} + \underbrace{\langle\psi_j|W - E_i^{(1)}|\psi_i^{(0)}\rangle}_{=\langle\psi_j|W|\psi_i\rangle} = 0 \quad \Rightarrow \quad \langle\psi_j|\psi_i^{(1)}\rangle = \frac{\langle\psi_j|W|\psi_i\rangle}{E_i - E_j}.$$

The contribution of $|\psi_i\rangle$ itself to $|\psi_i(\lambda)\rangle$ is fixed by the normalization condition (7.2):

$$1 = \langle\psi_i|\psi_i(\lambda)\rangle = \langle\psi_i|\sum_{n=0}^{\infty}\lambda^n|\psi_i^{(n)}\rangle = \underbrace{\langle\psi_i|\psi_i^{(0)}\rangle}_{=1} + \sum_{n=1}^{\infty}\lambda^n\langle\psi_i|\psi_i^{(n)}\rangle$$

so that $\langle\psi_i|\psi_i^{(n)}\rangle = 0$ for $n > 0$. We're left with:

$$|\psi_i(\lambda)\rangle = |\psi_i\rangle + \lambda\sum_{j\neq i}\frac{\langle\psi_j|W|\psi_i\rangle}{E_i - E_j}|\psi_j\rangle. \tag{7.6}$$

7.1.3 Higher orders

We'll sketch a recursion relation that gives all $E_i^{(n)}, |\psi_i^{(n)}\rangle$. The λ^n-terms in (7.3) are

$$H_0|\psi_i^{(n)}\rangle + W|\psi_i^{(n-1)}\rangle - \left(\sum_{k=0}^{n}E_i^{(k)}\right)|\psi_i^{(n-k)}\rangle = 0. \tag{7.7}$$

Projecting the above back onto $\langle\psi_i|$, only the term proportional to W and the final summand survive, giving

$$E_i^{(n)} = \langle\psi_i|W|\psi_i^{(n-1)}\rangle. \tag{7.8}$$

Thus: In order to learn the eigenvalue correction of order n, we need to know the eigenvector one of order $n-1$. E.g., with (7.6), the 2nd order coefficient for the energy is

$$E_i^{(2)} = \sum_{j\neq i}\frac{|\langle\psi_j|W|\psi_i\rangle|^2}{E_i - E_j}.$$

Higher orders. We won't be using any other higher-order corrections in this course. If that makes you sad, just project (7.7) onto j-th unperturbed state, re-arrange as above to arrive at

$$|\psi_i^{(n)}\rangle = \sum_{i\neq j} \frac{|\psi_i\rangle\langle\psi_i|}{E_i - E_j} \left(W|\psi_i^{(n-1)}\rangle - \sum_{k=2}^{n} E_i^{(k)} |\psi_i^{(n-k)}\rangle \right).$$

The $E^{(n)}$ term on the right can be eliminated using (7.8), so that only lower-order terms appear.

More is better? Let's talk about convergence.

It can be shown that under the assumptions made, the perturbation series converges to the true function $E_i(\lambda)$ for all $\lambda \in [0, \rho]$, where ρ is some finite convergence radius. Thus, eventually, higher orders will give better results! It is also true that for every n, there is a number ρ_n, such that the order-$n+1$ correction is better than the order-n one if $\lambda \in [0, \rho_n]$.

So it looks like going to higher orders is a good idea – which would be reassuring given that the complexity of the corrections increases quite rapidly.

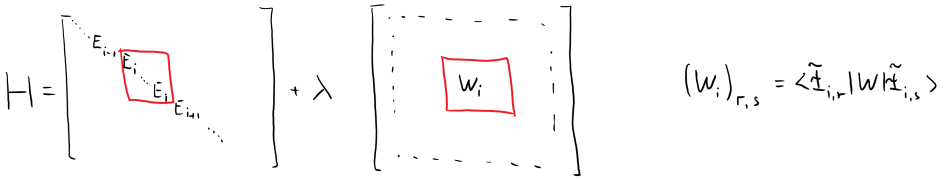
Unfortunately, in physics applications, λ is usually fixed by the problem, and even if $\lambda \in [0, \rho]$, convergence need not be monotonic in the order. The approximation can become worse when increasing n. Adding insult to injury, there's no general way to realize when this happens, and no obvious alternative. Tough.

7.1.4 Perturbations of a degenerate level

The theory changes slightly if the unperturbed eigenvalue is degenerate

$$H_0|\tilde{\psi}_{i,r}\rangle = E_i, \qquad r = 1, \ldots, d$$

for some ONB $|\tilde{\psi}_{i,1}\rangle, \ldots, |\tilde{\psi}_{i,d}\rangle$ of the E_i-eigenspace. In matrix notation:



where W_i is the $d \times d$ submatrix representing the action W on the degenerate space.

Hermiticity of W implies Hermiticity of W_i, so we can find an ONB $|\psi_{i,r}\rangle$ of the E_i-eigenspace that diagonalizes W:

$$W_i\,|\psi_{i,r}\rangle = E_{i,r}^{(1)}\,|\psi_{i,r}\rangle, \qquad r = 1, \ldots, d.$$

The eigenvalues of W_i give the 1st order correction: $E_{i,r}(\lambda) = E_i + \lambda E_{i,r}^{(1)}$.

In particular, if the eigenvalues of W_i are distinct, the degeneracy is lifted for $\lambda > 0$.

Proof. Make the ansatz

$$|\psi_{i,r}^{(1)}\rangle = \sum_{j\neq i} \frac{\langle\psi_j|W|\psi_{i,r}\rangle}{E_i - E_j} |\psi_j\rangle \tag{7.9}$$

for the first-order eigenvector correction. Then setting the λ^1-coefficient in (7.3) to 0 gives

$$\begin{aligned}
&H_0|\psi_{i,r}^{(1)}\rangle + W|\psi_{i,r}\rangle - E_i|\psi_{i,r}^{(1)}\rangle - E_{i,r}^{(1)}|\psi_{i,r}\rangle \\
=&W|\psi_{i,r}\rangle + \sum_{j\neq i} \frac{\langle\psi_j|W|\psi_{i,r}\rangle}{E_i - E_j} \underbrace{(H_0 - E_i)|\psi_j\rangle}_{=(E_j-E_i)|\psi_j\rangle} - E_{i,r}^{(1)}|\psi_{i,r}\rangle \\
=&\underbrace{\Big(\mathbb{1} - \sum_{j\neq i} |\psi_j\rangle\langle\psi_j|\Big)}_{\text{projection onto } E_i\text{-eigenspace}} W|\psi_{i,r}\rangle - E_{i,r}^{(1)}|\psi_{i,r}\rangle = E_{i,r}^{(1)}|\psi_{i,r}\rangle - E_{i,r}^{(1)}|\psi_{i,r}\rangle = 0.
\end{aligned}$$

7.1.5 Application: The linear Stark effect

Work in progress

Add a term $W = cZ$ to the atomic Hamiltonian. This would correspond to an electric field pointing along the z axis with constant field strength.

The ground state does not change to first order

$$\langle nlm|Z|nlm\rangle = 0.$$

For the states with $l = 0$ (in particular the ground state), this is obvious, because these wave functions are rotation-invariant, while Z changes sign under reflections.

All other levels are degenerate. The lowest one is spanned by $|n = 2, l = 0, m = 0\rangle$ and $|n = 2, l = 1, m\rangle$ for $m = -1, 0, 1$. Because L_z commutes with $W \propto Z$, the matrix elements vanish unless $m = m'$.

Thus the only possible non-zero ones are

$$\langle 200|Z|210\rangle = \langle 210|Z|200\rangle^*.$$

We know the wave functions can be represented with real numbers only, so let's call the result w. (Happens to be $-3e|E|a_0$). Thus

$$W = w \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \text{“exchanges } |2,0,0\rangle \text{ and } |2,1,0\rangle\text{”}$$

which is diagonalized by

$$\frac{1}{\sqrt{2}}(|2,0,0\rangle \pm |2,1,0\rangle), |2,1,-1\rangle, |2,1,1\rangle.$$

with eigenvalues $\pm w, 0$.

Thus the $n = 2$ space splits into two perturbed states with energies $\pm w$ and two states that remain unpertrubed to first order.

TBD: Figure.

7.2 Time-dependent perturbation theory

Fig. TBD.

We now study *dynamics* perturbatively. A typical example is the excitation of an atom by laser light. The effects of the light wave on a bound electron can be modeled by a potential term with a sinusoidal time-dependency – so we'll consider this type of interactions in particular.

We consider a system in state $|\phi(t)\rangle$ evolving under a time-dependent Hamiltonian

$$i\hbar\partial_t|\phi(t)\rangle = \big(H_0 + \lambda W(t)\big)|\phi(t)\rangle.$$

At time $t = 0$, the system is assumed to be in an initial eigenstate $|i\rangle$ with energy $H_0|i\rangle = E_i|i\rangle$. Now choose some "final" eigenstate $|f\rangle \neq |i\rangle$ with energy E_f. The goal is to estimate the probability

$$p_{i\to f}(t) = |\langle f|\phi(t)\rangle|^2$$

of finding the system in state $|f\rangle$ when measured at time t.

Series expansion

As is standard in perturbation theory, we assume that one can expand

$$|\phi(t)\rangle = \sum_{n=0}^{\infty} \lambda^n |\phi^{(n)}(t)\rangle$$

as a power series and that low orders are meaningful. For the Schrödinger equation,

$$i\hbar\partial_t\Big(\sum_n \lambda^n|\phi^{(n)}(t)\rangle\Big) = (H_0 + \lambda W(t))\Big(\sum_n \lambda^n|\phi^{(n)}(t)\rangle\Big)$$

the lowest degrees in λ give

$$\begin{aligned} i\hbar\partial_t|\phi^{(0)}\rangle &= H_0|\phi^{(0)}\rangle && \text{0th order,}\\ i\hbar\partial_t|\phi^{(1)}\rangle &= H_0|\phi^{(1)}\rangle + W(t)|\phi^{(0)}\rangle && \text{1st order.}\end{aligned}$$

With initial condition $|\phi(t=0)\rangle = |i\rangle$, the zeroth-order equation is solved by

$$|\phi^{(0)}(t)\rangle = e^{\frac{t}{i\hbar}E_i}|i\rangle.$$

Because $\langle f|i\rangle = 0$, the transition probability $|\langle f|\phi(t)\rangle|^2$ will, to first order, be equal to $\lambda^2|\langle f|\phi^{(1)}(t)\rangle|^2$. To compute the latter, project the 1st order equation onto $\langle f|$ to get

$$\partial_t\langle f|\phi^{(1)}(t)\rangle = \frac{1}{i\hbar}E_0\langle f|\phi^{(1)}(t)\rangle + \frac{1}{i\hbar}e^{\frac{t}{i\hbar}E_i}\langle f|W(t)|i\rangle.$$

Write $\langle f|\phi^{(1)}(t)\rangle = e^{\frac{t}{i\hbar}E_f}b_f(t)$, in terms of an (as of yet unknown) function $b_f(t)$. Plugging in and using the product rule gets us to

$$e^{\frac{t}{i\hbar}E_f}\partial_t b_f(t) = \frac{1}{i\hbar}e^{\frac{t}{i\hbar}E_i}\langle f|W(t)|i\rangle.$$

This equation is solved by the integral

$$b_f(t) = \frac{1}{i\hbar}\int_0^t e^{i\omega_{fi}t'}\langle f|W(t')|i\rangle\,\mathrm{d}t' \tag{7.10}$$

where we have introduced the *Bohr frequency*

$$\omega_{fi} = \frac{E_f - E_i}{\hbar}.$$

We now evaluate the integral for two important special cases of the time dependency.

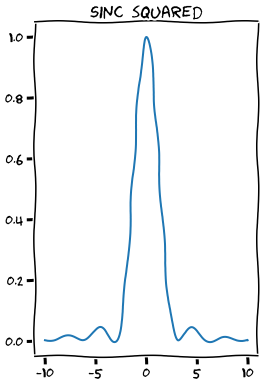
Constant perturbation

First, let $W(t) = W$ be constant. Without doing any calculations, you can see that the integral in (7.10) is oscillatory (and thus small) unless $\omega_{fi} \simeq 0$. Integrating explicitly:

$$b_f(t) = \frac{\langle f|W|i\rangle}{\hbar\omega_{if}}\left(e^{i\omega_{fi}t} - 1\right) \tag{7.11}$$

and using the *half-angle formula* $\sin^2\frac{\theta}{2} = \frac{1}{2}(1-\cos\theta)$,

$$p_{i\to f}(t) = \lambda^2\frac{|\langle f|W|i\rangle|}{\hbar^2}\left(\frac{\sin(\omega_{if}/2)t}{\omega_{fi}/2}\right)^2. \tag{7.12}$$



With $\epsilon := \omega_{fi}/2$, the right hand factor is $\sin^2(\epsilon\tau)/\epsilon^2$, the square of the "sinc function" (pictured). It has a central peak of height t^2, zeroes at $\epsilon = \pm\frac{\pi}{t}$, and shows oscillations of quadratically decreasing amplitude for $\epsilon \to \pm\infty$.

Thus, to first order, only states $|f\rangle$ with energy E_f in the range $E_i \pm \frac{2\pi\hbar}{t}$ pick up significant weight. For such states, the modulus squared is proportional to t^2 and the squared coupling coefficient $|\langle f|\lambda W|i\rangle|^2$.

The result can be interpreted as an approximate version of energy conservation. "*Wait, isn't energy preserved* exactly*?*" There's no contradiction, because we're talking about energy as measured by H_0, but the dynamics is generated by H.

Sinusoidal perturbation

Now assume there is an angular frequency ω such that

$$W(t) = 2\cos(\omega t)W = (e^{i\omega t} + e^{-i\omega t})W.$$

Then

$$b_f(t) = \frac{1}{i\hbar}\int_0^t e^{i(\omega_{fi}+\omega)t'}\langle f|W|i\rangle\,\mathrm{d}t' + \frac{1}{i\hbar}\int_0^t e^{i(\omega_{fi}-\omega)t'}\langle f|W|i\rangle\,\mathrm{d}t'.$$

Each of the integrals reduces to the time-independent case, with ω_{fi} replaced by $\omega_{fi} \pm \omega$. Assume that $\omega_{fi} \gg 0$ and $\omega \simeq \omega_{fi}$. By the above, the first integral will be $\simeq 0$, and the transition probability is thus approximately equal to (7.12), with ω_{if} replaced by $\omega_{if} - \omega$.

A perturbation $\cos(\omega t)W$ causes an efficient transition between H_0-eigenstates $|i\rangle, |f\rangle$ if $|\langle f|W|i\rangle|^2 \gg 0$ and ω is close to the Bohr frequency ω_{fi}.

7.2.1 Coupling to a continuum of states, "Fermi's Golden Rule"

Let's look at the case where E_f is part of the continuous spectrum, and $|f\rangle$ only a generalized eigenstate. Then $|\langle f|\phi(t)\rangle|^2$ is a probability density, and it is more meaningful to work with

$$p_{i\to F} = \langle\phi(t)|P_F|\phi(t)\rangle,$$

for some projection operator P_F. Let $\rho(f)$ be a measure such that

$$P_F = \int_F |f\rangle\langle f|\rho(f)\,\mathrm{d}f.$$

In other words, $\rho(f)$ is the "density of states", in the sense of Sec. 2.7.3. Then

$$\langle\phi(t)|P_F|\phi(t)\rangle = \int_F |\langle f|\phi(t)\rangle|^2\rho(f)\,\mathrm{d}f = \int_F 4|\langle f|V|i\rangle|^2\frac{\sin^2\left((E_i-E_f)\frac{t}{2\hbar}\right)}{(E_i-E_f)^2}\rho(f)\,\mathrm{d}f.$$

It is known (by the Dirichlet integral) that the area under the "sinc-square"–curve is $\tau\pi$. Therefore, the family of functions $f_\tau(\epsilon) := \frac{1}{\pi\tau}\sin^2(\epsilon\tau)/\epsilon^2$ converges to a δ-function centered at 0 as $\tau\to\infty$. Then, optimistically,

$$\langle\phi(t)|P_F|\phi(t)\rangle \simeq t\,\frac{2\pi}{\hbar}\int_F |\langle f|V|i\rangle|^2\,\delta(E_i-E_f)\rho(f)\,\mathrm{d}f =: t\,\Gamma. \tag{7.13}$$

Let's suspend disbelief for a while and take (7.13) at face value. It is called *Fermi's Golden Rule*: The probability $p_{i\to F}(t)$ increases linearly, with slope Γ proportional to the squared coupling and the density of states, integrated over all final states with the right energy.

The "$\simeq$"–step in (7.13) involved quite the leap of faith. The squared-sinc-construction gives a delta function only in the limit of large times, but first-order perturbation theory is valid, at most, at short times. It's unclear whether there's an intermediate regime where both approximations simultaneously hold. But experience has shown that the "golden" rule gives the right answer more often than one could have hoped, hence the moniker.

Appendix A

Complex analysis

These notes are following the books by Jänich (*Funktionentheorie*, in German) and Altland–von Delft. We skip all proofs that don't reduce to simple calculations. None are difficult, though. The Jänich book is particularly clear.

A.1 Holomorphic functions

A function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at z_0 if the limit

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0} = f'(z_0) \tag{A.1}$$

exists.

Now let $f : \mathbb{C} \to \mathbb{C}$. Because one can divide by complex numbers, the expression (A.1) still makes sense if z, z_0 are taken to be complex.

The function f is *complex differentiable* at z_0 if the limit (A.1) exists for $z \in \mathbb{C}$.

Let $U \subset \mathbb{C}$ be an open set.

The function f is *holomorphic* on U if it is complex differentiable for all $z_0 \in U$.

(Often, the domain U is left implicit).
Examples:

- The identity function $f(z) = z$ is obviously holomorphic on $U = \mathbb{C}$.
- The sum, product, and chain rules hold for complex differentiation. In particular, sums, products, and compositions of holomorphic functions are holomorphic
- Thus polynomials are holomorphic on $\mathbb{C}$.
- Assume that for every $z_0 \in U$, there is a power series

$$\sum_{k=0}^{\infty} c_k (z - z_0)^k$$

that converges to $f(z)$ for all z in a vicinity of z_0. Explicitly, its complex derivative can be computed term-by-term

$$f'(z) = \sum_{k=0}^{\infty} c_k \, k z^{k-1}.$$

(As we'll see later, the converse is also true: Every holomorphic function has a locally convergent Taylor series).

- In particular, the exponential and sine/cosine functions are holomorphic on $\mathbb{C}$.
- If f, g are holomorphic on a set U, and if g has no zeros on U, then f/g is holomorphic on U and quotient rule $(f/g)' = (f'g - fg')/g^2$ holds.

Remark on domains. In physics, we're often cavalier about domains. The statement "$1/x$ is a function on $\mathbb{R}$" wouldn't raise eyebrows (even though the function is only defined on $\mathbb{R} \setminus \{0\}$). Often, that sloppiness doesn't cause serious problems. Not so in function theory! Here, we have to be precise for reasons that will be obvious soon.

A.1.1 The Cauchy-Riemann equations

A function $f : \mathbb{C} \to \mathbb{C}$ can equivalently be expressed as a real vector field

$$\begin{pmatrix} u(x,y) \\ v(x,y) \end{pmatrix}, \qquad z = x + iy, \quad f = u + iv. \tag{A.2}$$

In this picture, holomorphy is characterized by the *Cauchy-Riemann equations*:

The function f is holomorphic if and only if

$$\partial_x u = \partial_y v, \qquad \partial_y u = -\partial_x v. \tag{A.3}$$

Interpretation. Let's think about $\mathbb{C}$ as $\mathbb{R}^2$ with basis $1, i$. Multiplication by a complex number then becomes a linear map $\mathbb{R}^2 \to \mathbb{R}^2$. Explicitly, multiplication by 1 acts trivially, and multiplication by i acts on the basis as $1 \mapsto i$ and $i \mapsto -1$. Therefore, in matrix notation:

$$z \mapsto (a + ib)\, z \qquad \text{corresponds to} \qquad a \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + b \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Recall that differentiation can be seen as a method for finding the optimal linear approximation to a function. In particular, if $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^m$, then

$$\boldsymbol{F}(\boldsymbol{x} + \boldsymbol{\delta}) \simeq \boldsymbol{F}(\boldsymbol{x}) + J\boldsymbol{\delta}, \quad \text{where } J \text{ is the } \textit{Jacobi matrix} \text{ with elements } \quad J_{ij} = \partial_j F_i.$$

The Cauchy-Riemann equations say that the Jacobi matrix of the vector field (A.2) is

$$\begin{pmatrix} \partial_x u & \partial_y u \\ \partial_x v & \partial_y v \end{pmatrix} = (\partial_x u) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + (\partial_y u) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{A.4}$$

i.e. that the two-dimensional real linear map given by the l.h.s. of (A.4) corresponds to a one-dimensional complex linear map. This should explain the connection to complex differentiation.

A.2 Complex integration

A.2.1 Cauchy theorem

Vector calculus recap: Line integrals. Let $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^n$ be a vector field.
The *line integral* along a curve $\boldsymbol{\gamma} : [0,1] \to \mathbb{R}^n$ is

$$\int_\gamma \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \int_0^1 \dot{\boldsymbol{\gamma}}(t) \cdot \boldsymbol{F}(\boldsymbol{\gamma}(t))\,\mathrm{d}t.$$

Let f be a holomorphic function on $U \subset \mathbb{C}$, and let $\gamma : [0,1] \to U$ a curve in U. The *complex line integral* is

$$\int_\gamma f \mathrm{d}z := \int_0^1 \dot{\gamma}(t) f(\gamma(t))\,\mathrm{d}t.$$

We can now state the centrally important *Cauchy theorem*. It assumes that U is connected and simply connected (i.e. has no holes).

If U is connected and simply connected, γ a closed curve, and f holomorphic, then

$$\oint_\gamma f\,\mathrm{d}z = 0.$$

Vector calculus recap: Stoke's Theorem. The *fundamental theorem of calculus* links the values of a function on the boundary of a region to the integral of its derivative over the region. One of its generalizations to vector fields is *Stoke's Theorem*. Here, "region" means "an oriented surface Σ in $\mathbb{R}^3$, described by a field $\boldsymbol{n}$ of normal vectors", and the right notion of "derivative" is the *curl* $\boldsymbol{\nabla} \times \boldsymbol{F}$. The boundary of Σ is a curve γ. Then we have

$$\int_\Sigma (\boldsymbol{\nabla} \times \boldsymbol{F}) \cdot \boldsymbol{n}\,\mathrm{d}S = \int_\gamma \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}.$$

TBD: Figure.

Proof. Expand the integrand as

$$\dot{\gamma}\, f = (\mathrm{Re}\,\dot{\gamma} + i\,\mathrm{Im}\,\dot{\gamma})(u + iv) = \begin{pmatrix} \mathrm{Re}\,\dot{\gamma} \\ \mathrm{Im}\,\dot{\gamma} \end{pmatrix} \begin{pmatrix} u \\ -v \end{pmatrix} + i \begin{pmatrix} \mathrm{Re}\,\dot{\gamma} \\ \mathrm{Im}\,\dot{\gamma} \end{pmatrix} \begin{pmatrix} v \\ u \end{pmatrix}.$$

Embed $\mathbb{R}^2$ into the $z = 0$ plane of $\mathbb{R}^3$. Let Σ be the surface enclosed by γ. Then Stoke's Theorem and the Cauchy-Riemann equations give

$$\mathrm{Re} \oint_\gamma f\,\mathrm{d}z = \int_\gamma \begin{pmatrix} u \\ -v \\ 0 \end{pmatrix} \cdot \mathrm{d}\boldsymbol{r} = \int_\Sigma \left(\boldsymbol{\nabla} \times \begin{pmatrix} u \\ -v \\ 0 \end{pmatrix} \right) \cdot \boldsymbol{e}_z \mathrm{d}\boldsymbol{r} = \int_\Sigma (\partial_x v + \partial_y u) \mathrm{d}\boldsymbol{r} = 0,$$

$$\mathrm{Im} \oint_\gamma f\,\mathrm{d}z = \int_\gamma \begin{pmatrix} v \\ u \\ 0 \end{pmatrix} \cdot \mathrm{d}\boldsymbol{r} = \int_\Sigma \left(\boldsymbol{\nabla} \times \begin{pmatrix} v \\ u \\ 0 \end{pmatrix} \right) \cdot \boldsymbol{e}_z \mathrm{d}\boldsymbol{r} = \int_\Sigma (\partial_x u - \partial_y v) \mathrm{d}\boldsymbol{r} = 0.$$

A.2.2 Analyticity

Recap: Radius of convergence of a series. Consider a series

$$f(z) = \sum_{k=0}^{\infty} c_k z^k, \qquad c_k \in \mathbb{C}.$$

Then there exists a number ρ, the *radius of convergence*, such that the series converges for all z with $|z| < r$ and diverges for all z with $|z| > r$. (No simple statement is possible for $|z| = r$).

TBD: Figure.

A function f is *analytic* on an open set $U \subset \mathbb{C}$ if for every point $z_0 \in U$, f can be written as a Taylor series

$$f(z) = \sum_{k=0}^{\infty} c_n (z - z_0)^n, \qquad c_n = \frac{1}{n!} \partial_z^n f(z_0).$$

with a positive radius of convergence.

Remark. Physicists often assume implicitly that every function can be meaningfully Taylor-expanded about every point, or at least about every point at which the function is infinitely often differentiable. (The ansatz made in quantum perturbation theory is a typical example). But that isn't actually true! (Read the WP page on *bump functions* for disturbing counterexamples). Now we have a name for the class of functions for which this does work: *analytic*.

Remarkably one can show that:

A function is holomorphic if and only if it is analytic.

- Some authors therefore do not distinguish between *holomorphic* and *analytic*.
- The statement is remarkable, because the definition of a holomorphic function only talked about *first-order* derivatives. But convergent power series are *infinitely often* differentiable. The implication "first order differentiable $\Rightarrow$ infinitely differentiable" is a remarkable property of complex differentiation, which, of course, absolutely does not hold in the real-valued theory.

A.2.3 Singularities and Laurent series

Let f be holomorphic on U. An isolated point z_0 in the complement of U is called an *isolated singularity*.

There are three types of isolated singularities: TBD. Figure.

...

An isolated singularity at a point z_0 is called a *pole* if there exists an n such that $(z - z_0)^n f(z)$ has a removable singularity at z_0.

If no such n exists, the singularity is *essential*. The *order* of a pole is the smallest number n with that property. A pole of order 1 is called *simple*.

If f has a pole of order n at z_0, then, removing the singularity of $(z-z_0)^n f(z)$ at z_0, we obtain a function that is now holomorphic in a vicinity of z_0. By the last section, this function has a convergent Taylor series

$$(z-z_0)^n f(z) = \sum_{k=0}^{\infty} c_k (z-z_0)^k \quad \text{so that} \quad f(z) = \sum_{k=-n}^{\infty} c_{k+n} (z-z_0)^k.$$

Such a series that includes negative-power terms is called a *Laurent series*.

If f has a pole of order n at z_0, it has a *Laurent series* in a vicinity of z_0:

$$f(z) = \sum_{k=-n}^{\infty} a_k (z-z_0)^k, \qquad z \neq 0.$$

A function that is holomorphic except for finitely many poles is *meromorphic*.

A.2.4 Residues

Let $\gamma(t) = re^{i2\pi t}$ be a positively oriented circular closed curve. Compute the line integral

$$\int_\gamma z^k \,\mathrm{d}z = \int_0^1 \underbrace{r^k e^{2\pi i k t}}_{z^k \text{ at } z=\gamma(t)} \underbrace{2\pi i r e^{2\pi i t}}_{\dot\gamma(t)} \,\mathrm{d}t = 2\pi i r^{k+1} \int_0^1 e^{i2\pi(k+1)t} \,\mathrm{d}t = 2\pi i \delta_{k,-1}.$$

Assume f has a pole at z_0. Let r be sufficiently small that the Laurent series around z_0 converges. Let $\gamma(t) = re^{i2\pi t} + z_0$. Exchanging the order of summation and integration,

$$\frac{1}{2\pi i} \int_\gamma f(z) \,\mathrm{d}z = a_{-1},$$

i.e. the integral around the singularity gives the -1st term of the Laurent series.

If f is meromorphic on U, the *residue* at a point $z_k \in U$ is

$$\mathrm{Res}(f, z_k) := \frac{1}{2\pi i} \int_\gamma f(z) \,\mathrm{d}z,$$

where γ is any positively oriented loop that encloses z_k, but no other singularity.

For physics, the most relevant result in complex analysis is the *residue theorem*:

If f is meromorphic and γ a positively oriented loop, then

$$\oint_\gamma f \,\mathrm{d}z = 2\pi i \sum_k \mathrm{Res}(f, z_k),$$

where the sum is over all poles enclosed by γ.

The proof is essentially graphical: TBD.

A.3 Applications of the residue theorem

A.3.1 Lorentz distribution

Consider the real integral

$$\int_{-\infty}^{\infty} \frac{1}{x^2+a^2}\,\mathrm{d}x = \lim_{R\to\infty}\int_{-R}^{R} \frac{1}{x^2+a^2}\,\mathrm{d}x.$$

Backstory. Bored? Here's the backstory. (Which we just mention, but won't explain. Sorry). Assume an atom decays from an excited state to its ground state, emitting a photon. You might expect the energy of the photon to equal the difference of the atomic energies. But that isn't exactly true! The reason is that "atom in an excited state, no photons present" isn't an eigenstate of the joint atom–field system, and therefore does not have a definite energy. Instead, it turns out that the distribution of energies is well-approximated by a *Lorentz (*or *Cauchy) distribution*

$$p(E) = N\frac{1}{(E-E_0)^2+(\hbar\Gamma/2)^2}.$$

Here, E_0 is the *location* of the distribution (close, but not equal to the difference of the atomic energies), $\hbar\Gamma/2$ controls the width of the distribution, related to the *lifetime* $\tau = 1/\Gamma$ of the excited state, and N is a normalization constant.

We'll compute the normalization constant. You expected more? Sorry. But the mundane stuff also has to be done. It is the reciprocal of the above integral for $a = \hbar\Gamma/2$.

Nothing prevents us from re-interpreting the integrand as a function on $\mathbb{C}$. It factors as

$$f(z) = \frac{1}{z^2+a^2} = \frac{1}{(z+ia)(z-ia)},$$

thus it has two first-order poles, at $z = \pm ia$. Now consider the arc $\gamma_2^{(R)}(t) = Re^{i\pi t}$. Its length is πR. The absolute value of the integrand on the arc is

$$\left|\frac{1}{Re^{i\pi t}+ia}\right|\left|\frac{1}{Re^{i\pi t}-ia}\right| \le \left(\frac{1}{R-|a|}\right)^2. \tag{A.5}$$

It follows that the contribution of the integral over the arc vanishes as $R\to\infty$:

$$\lim_{R\to\infty}\left|\int_{\gamma_2^{(R)}} \frac{1}{z^2+a^2}\,\mathrm{d}z\right| \le \lim_{R\to\infty} \pi R\left(\frac{1}{R-|a|}\right)^2 = 0.$$

Thus, we can *complete the contour* by adding the γ_2 integral. The contour will then enclose the simple pole at $z_0 = ia$. The -1st coefficient of the Laurent series about z_0 is just the 0th term of the Taylor series of

$$(z-ia)f(z) = \frac{1}{z+ia}$$

which in turn is just the value, $1/(2ia)$, of that function at z_0 (after having removed the singularity). Therefore:

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{1}{x^2+a^2}\,\mathrm{d}x &= \lim_{R\to\infty}\left(\int_{-R}^{R}\frac{1}{z^2+a^2}\,\mathrm{d}z + \int_{\gamma_2^{(R)}}\frac{1}{z^2+a^2}\,\mathrm{d}z\right)\\ &= 2\pi i\,\mathrm{Res}(f, ia) = \frac{2\pi i}{2ia} = \frac{\pi}{a}.\end{aligned}$$

The cool aspect of the calculation is that in the residue picture, the only non-zero contribution to the *real* integral came from a *complex* singularity. Magical, no?

A.3.2 Green's functions

Consider the integral

$$G(t) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-i\omega t}}{-\omega^2 - 2i\gamma\omega + \omega_0^2}\,\mathrm{d}\omega, \qquad \gamma, \omega_0 > 0.$$

Backstory. Newton's equation for the position $u(t)$ of a particle subject to a driving force $mf(t)$, viscous damping coefficient $(m\gamma)/2$, and undamped eigenfrequency ω_0 is

$$\left(\partial_t^2 + 2\gamma\partial_t + \omega_0^2\right) u(t) = f(t)$$

The problem is to find $u(t)$ given $f(t)$ and the boundary condition $u(-\infty) = 0$.

It follows from the theory of Green's functions that the solution is given by the convolution

$$u(t) = \int G(t - t')f(t')\,\mathrm{d}t.$$

In this sense, $G(t - t')$ describes the influence at time t of an external impulse at time t'.

The polynomial in the denominator factorizes as

$$-(\omega - \omega_+)(\omega - \omega_-), \qquad \omega_\pm = i\gamma \pm \sqrt{\omega_0^2 - \gamma^2}.$$

The integrand has first-order poles in the positive half-plane. [TBD: Figure.] The strategy is again to close the contour by a half-circle with radius R through the complex numbers, in such a way that the integral over the arc vanishes as $R \to \infty$. The denominator is quadratic in ω and thus goes to 0 fast enough to not cause trouble – compare section above. The enumerator is more tricky: The exponential goes to 0 or to ∞ away from the real axis. To analyze the behavior, we have to treat the cases of positive / negative t separately.

For $t < 0$, the exponential is bounded in the lower half-plane. Closing the integration there leaves the poles outside the contour and thus gives 0.

For $t > 0$, the contour must be closed in the upper half-plane, thus enclosing both poles. The residue theorem then gives

$$G(t) = ie^{-\gamma t}\left(-\frac{1}{\omega_+ - \omega_-}e^{i\sqrt{\omega_0^2-\gamma^2}t} + \frac{1}{\omega_+ - \omega_-}e^{-i\sqrt{\omega_0^2-\gamma^2}t}\right).$$

Since we're done integrating in Fourier space, we can re-use the letter ω, defining it to be $\sqrt{|\omega_0^2 - \gamma^2|}$. Then the above may be simplified (using l'Hôpital for the equality case) to

$$G(t) = \theta(t)\frac{e^{-\gamma t}}{\omega}\begin{cases} \sin(\omega t) & \omega_0 > \gamma \\ e^{-\gamma t}t & \omega_0 = \gamma \\ \sinh(\omega t) & \omega_0 < \gamma \end{cases}, \tag{A.6}$$

where θ is the Heaviside step function.

Discussion:

- The Heaviside function θ assures that the physical system behaves *causally*, in the sense that a past impulse will only affect future behavior. Mathematically, it appeared because the sign of t determined whether the contour would enclose singularities or not.
- You may recognize the functional form of $G(t)$ from classical mechanics. The cases correspond to an underdamped / critical / overdamped oscillator respectively.

A.3.3 Resolvents

Given a linear operator L on a Hilbert space $\mathcal{H}$, the function that maps complex numbers z to the operator

$$R(z; L) := (z\mathbb{1} - L)^{-1},$$

is called the *resolvent* of L. (Warning: Some authors use the sign convention $(L - z\mathbb{1})^{-1}$!).

Assume for simplicity that $L = L^\dagger$ is a Hermitian matrix with eigenvalues $\lambda \in D$ and eigenvectors $|\psi_\lambda\rangle$. Then

$$(z - L)^{-1} = \sum_{\lambda \in D} \frac{1}{z - \lambda} |\psi_\lambda\rangle\langle\psi_\lambda|.$$

Let γ be a positively oriented closed curve in $\mathbb{C}$. The matrix-valued integral

$$\oint_\gamma (z - L)^{-1} \, \mathrm{d}z \tag{A.7}$$

has matrix elements

$$\langle i| \oint_\gamma (z - L)^{-1} \, \mathrm{d}z |j\rangle = \sum_{\lambda \in D} \oint_\gamma \frac{1}{z - \lambda} \, \mathrm{d}z \, \langle i|\psi_\lambda\rangle\langle\psi_\lambda|j\rangle = 2\pi i \sum_{\lambda \in \Sigma \cap D} \langle i|\psi_\lambda\rangle\langle\psi_\lambda|j\rangle,$$

where Σ is the region enclosed by the contour. In other words, (A.7) is $2\pi i$ times the projection onto the eigenspaces whose eigenvalues are enclosed by the contour.

Such operator-valued complex integrals are used e.g. in quantum scattering theory.

A.4 Analytic continuations

TBD.

Appendix B

Technical details

B.1 More on delta distributions

B.1.1 How to think about distributions

Our account of general eigendecompositions and distributions is not mathematically rigorous. It *can* be made precise, but doing so takes a lecture in functional analysis (c.f. the *spectral theorem* and the theory of *distributions*). Given that we won't take the time here to go into more details, how should one deal with distributions that pop up in equations? Some strategies:

1. *Integrate against smooth functions that quickly vanish at infinity*. As in (2.15), even if the intermediate mathematical expression contains δ's, they should have all vanished after one has integrated the expression over smooth functions in order to extract physical quantities. The mathematically rigorous approach is based on this strategy, and it is the one we will have at the back of our heads in this document.

2. *Think of δ is an idealization of "highly concentrated"*. One can in principle replace $\delta_{\boldsymbol{x}}$ by functions $\delta_{\boldsymbol{x}}^{(\epsilon)}$ that are supported on an ϵ-ball around $\boldsymbol{x}$, where ϵ is much smaller than any relevant length scale. The final physical results should then only weakly depend on the actual choice of ϵ, and one should, in fact, be able to take a limit $\epsilon \to 0$. In this sense, the actual distribution is an idealization that allows one to directly obtain the limit, without first introducing an ϵ and eliminating it again in the end.

3. *Shut-up-and-calculate*. The reason δ's are so ubiquitous is that they work well as a computational tool. So in *reality*, people just use them whenever they would have used a Kronecker delta in a discrete analogue, and pretend that all algebraic manipulations that are valid for Kronecker deltas also extend to distributions. This *mostly* works.

B.1.2 Derivatives of delta functions

While the mathematicians look the other way, let's get adventurous and represent the momentum operator in *position* basis.

The *derivative of the delta function* $\delta_y'(x)$ is a formal object whose inner product with smooth functions vanishing at infinity is defined so that formally the rule of *integration by*

parts holds:

$$\langle \delta'_y | \psi \rangle = \int \delta'_x(x)\psi(x)\,\mathrm{d}x := -\int \delta_x(x)\psi'(x)\,\mathrm{d}x = -\phi'(y)$$

and therefore

$$P = i\hbar \int |\delta_x\rangle\langle\delta'_x|\,\mathrm{d}x$$

is valid in the sense that for all smooth ψ, ϕ vanishing at infinity,

$$i\hbar \int \langle\psi|\delta_x\rangle\langle\delta'_x|\phi\rangle\,\mathrm{d}x = -i\hbar \int \bar{\psi}(x)\phi(x')\,\mathrm{d}x = \langle\psi|P|\phi\rangle. \tag{B.1}$$

Other expressions are

$$P = -i\hbar \int |\delta'_x\rangle\langle\delta_x|\,\mathrm{d}x = -i\hbar \int |\delta_x\rangle\partial_x\langle\delta_x|\,\mathrm{d}x = i\hbar \int\int |\delta_y\rangle\, \delta'_y(z)\, \langle\delta_z|\,\mathrm{d}y\,\mathrm{d}z.$$

The first holds because shifting the derivative to the bra means that in (B.1), ψ instead of ϕ gets differentiated, and to remedy that, we need to use integration by parts once more, which causes the change in sign. The second one holds because $\partial_x \delta_x(y) = \partial_x \delta(y-x) = -\delta'_x(y)$, so differentiating the index rather than the argument of the delta function also incurs a sign change. A similar argument verifies the third expression. This last one is interesting, because it is a formal generalization of (2.13) to continuous bases. It expresses P in terms of its "matrix elements"

$$\langle\delta_y|P|\delta_z\rangle = -i\hbar \int \delta_y(x)\delta'_z(x)\,\mathrm{d}x = i\hbar \int \delta'_y(x)\delta_z(x)\,\mathrm{d}x = i\hbar\delta'_y(z).$$

Using these formulas, the kinetic energy operator reads

$$\frac{P^2}{2m} = -\frac{\hbar^2}{2m}\int |\delta_x\rangle\partial_x^2\langle\delta_x|\,\mathrm{d}x = \frac{\hbar^2}{2m}\int |\delta'_x\rangle\langle\delta'_x|\,\mathrm{d}x.$$

B.2 More on Fourier transforms

Let's have a closer look at the n-dimensional Fourier basis $\phi_{\boldsymbol{k}}(\boldsymbol{x}) = (2\pi)^{-n/2}e^{i\boldsymbol{k}\boldsymbol{x}}$, for $\boldsymbol{k} \in \mathbb{R}^n$, and the associated transforms

$$\begin{aligned}\tilde{\psi}(\boldsymbol{k}) &:= \langle\boldsymbol{k}|\psi\rangle = (2\pi)^{-n/2}\int e^{-i\boldsymbol{k}\boldsymbol{x}}\psi(\boldsymbol{x})\,\mathrm{d}^n\boldsymbol{x},\\ \psi(\boldsymbol{x}) &:= \langle\boldsymbol{x}|\psi\rangle = (2\pi)^{-n/2}\int e^{i\boldsymbol{k}\boldsymbol{x}}\tilde{\psi}(\boldsymbol{k})\,\mathrm{d}^n\boldsymbol{k}.\end{aligned} \tag{B.2}$$

B.2.1 Fourier transforms in finite regions

The Fourier basis for functions on $\mathbb{R}^n$ is continuous, which, as discussed above, comes with technical difficulties. Things are much easier for spaces of functions in finite regions.

Concretely, choose some length L and consider the box $B = [-L/2, L/2]^n$ with side length L centered at the origin. Let $L^2(B)$ be the space of functions defined on the region

B with cyclic boundary conditions (i.e. functions take the same values on opposite faces of the box) and with inner products given by integrals over B only:

$$\langle\phi|\psi\rangle = \int_B \bar{\phi}(x)\psi(x)\,\mathrm{d}^n\boldsymbol{x}.$$

A plane wave $e^{i\boldsymbol{k}\boldsymbol{x}}$ complies with the boundary conditions if and only if every component k_i of the wave vector is an integer multiple of $\frac{2\pi}{L}$. Indeed, the discrete set of functions

$$\phi_{\boldsymbol{k}}(\boldsymbol{x}) := \frac{1}{L^{n/2}} e^{i\boldsymbol{k}\boldsymbol{x}}, \qquad k \in \frac{2\pi}{L}\mathbb{Z}^n,$$

forms an ONB for $L^2(B)$ and the formulas for the Fourier transform become

$$\begin{aligned} \tilde{\psi}(\boldsymbol{k}) &= \frac{1}{L^{n/2}} \int_B e^{-i\boldsymbol{k}\boldsymbol{x}}\psi(\boldsymbol{x})\,\mathrm{d}^n\boldsymbol{x}, \\ \psi(\boldsymbol{x}) &= \frac{1}{L^{n/2}} \sum_{\boldsymbol{k}\in\frac{2\pi}{L}\mathbb{Z}^n} \tilde{\psi}(\boldsymbol{k}) e^{i\boldsymbol{k}\boldsymbol{x}}. \end{aligned} \tag{B.3}$$

Comparison with (B.2) shows that, formally, the transition between a finite and an unbounded volume Fourier transform is facilitated by the substitution

$$\frac{1}{\pi^{n/2}} \int_{\mathbb{R}^n} \mathrm{d}^n\boldsymbol{k} \quad \leftrightarrow \quad \frac{1}{L^{n/2}} \sum_{\boldsymbol{k}\in\frac{2\pi}{L}\mathbb{Z}^n} \tag{B.4}$$

Note the asymmetry in (B.3): Fourier transformation takes the compact domain B to the discrete domain $\frac{2\pi}{L}\mathbb{Z}^n$. We can of course reverse the interpretation of the two functions in (B.3). The formula then says that functions $\psi(\boldsymbol{x})$ defined on a lattice $\mathbb{Z}^n\frac{2\pi}{L}$ can be expanded in terms of plane waves $\phi_{\boldsymbol{k}}(\boldsymbol{x}) = \frac{1}{L^{n/2}} e^{-i\boldsymbol{k}\boldsymbol{x}}$ with wave vectors $\boldsymbol{k} \in B$. In this context, B is sometimes called the *Brillouin zone* and $\boldsymbol{k}$ the *crystal momentum* or *quasi-momentum*.

Of course, the universe isn't actually a finite box with cyclic boundary conditions... ...but we may as well pretend it were! Physics is local, so we can assume that all phenomena we are interested in take place in some box that is sufficiently large that the boundary does not affect the predictions we extract from the theory.

B.2.2 Translation symmetry

Fourier transforms are intimately connected to translation symmetry. Let $T_{\boldsymbol{a}}$ be the translation operator that shifts functions along the vector $\boldsymbol{a}$

$$(T_{\boldsymbol{a}}\psi)(\boldsymbol{x}) = \psi(\boldsymbol{x} - \boldsymbol{a}).$$

The Fourier basis diagonalizes translations:

$$\langle\boldsymbol{x}|T_{\boldsymbol{a}}|\phi_{\boldsymbol{k}}\rangle = e^{i\boldsymbol{k}(\boldsymbol{x}-\boldsymbol{a})} = e^{-i\boldsymbol{k}\boldsymbol{a}}\langle\boldsymbol{x}|\phi_{\boldsymbol{k}}\rangle \qquad \Rightarrow \qquad T_{\boldsymbol{a}} = \int e^{-i\boldsymbol{k}\boldsymbol{a}}|\phi_{\boldsymbol{k}}\rangle\langle\phi_{\boldsymbol{k}}|\,\mathrm{d}^n\boldsymbol{k}.$$

It is the unique common eigenbasis of all $T_{\boldsymbol{a}}$ (why?). Therefore, if A is any operator that commutes with translations

$$[T_{\boldsymbol{a}}, A] = 0 \qquad \forall\,\boldsymbol{a}, \tag{B.5}$$

then T must be diagonal in the Fourier basis, too. Explicitly, (B.5) implies that A is fully specified by its "first column"

$$f(\boldsymbol{z}) := \langle \delta_{\boldsymbol{z}} | A | \delta_0 \rangle$$

in the sense that

$$\langle \delta_{\boldsymbol{x}} | A | \delta_{\boldsymbol{y}} \rangle = \langle \delta_{\boldsymbol{x}} | A T_{\boldsymbol{y}} T_{-\boldsymbol{y}} | \delta_{\boldsymbol{y}} \rangle = \langle \delta_{\boldsymbol{x}} | T_{\boldsymbol{y}} A | \delta_0 \rangle = \langle \delta_{\boldsymbol{x}-\boldsymbol{y}} | A | \delta_0 \rangle = f(\boldsymbol{x}-\boldsymbol{y}).$$

It then follows that the eigenvalues of A are proportional to the Fourier transform of f:

$$\begin{aligned}\langle \delta_{\boldsymbol{x}} | A | \phi_{\boldsymbol{k}} \rangle &= (2\pi)^{-n/2} \int \langle \delta_{\boldsymbol{x}} | A | \delta_{\boldsymbol{y}} \rangle e^{i\boldsymbol{k}\boldsymbol{y}} \, \mathrm{d}^n \boldsymbol{y} \\ &= e^{ik\boldsymbol{x}} (2\pi)^{-n/2} \int f(\boldsymbol{x}-\boldsymbol{y}) e^{-i\boldsymbol{k}(\boldsymbol{x}-\boldsymbol{y})} \, \mathrm{d}^n \boldsymbol{y} = (2\pi)^{n/2} \tilde{f}(\boldsymbol{k}) \, \langle \delta_{\boldsymbol{x}} | \phi_k \rangle\end{aligned}$$

so that, summarizing,

$$A = (2\pi)^{n/2} \int \tilde{f}(\boldsymbol{k}) \, |\phi_{\boldsymbol{k}}\rangle\langle\phi_{\boldsymbol{k}}| \, \mathrm{d}^n \boldsymbol{k}. \tag{B.6}$$

B.2.3 Fourier transform for functions depending on space and time

Common notation and sign conventions slightly differ when one coordinate has the interpretation of a time. Write $x = (t, \boldsymbol{x}) \in \mathbb{R}^n$, with t the "temporal" coordinate and $\boldsymbol{x} \in \mathbb{R}^{n-1}$ the "spatial" ones. Wave vectors are denoted by $k = (\omega, \boldsymbol{k})$. To compute inner products, we use the Minkowski form

$$\langle p, x \rangle = \omega t - \boldsymbol{k}\boldsymbol{x},$$

which (at least in the case of $n = 4$) determines the space-time metric in relativity. The commonly used basis of plane waves is

$$\phi_k(x) = (2\pi)^{-n/2} e^{-i\langle p, x\rangle} = (2\pi)^{-n/2} e^{-i\omega t + \boldsymbol{k}\boldsymbol{x}}$$

so that the forward and inverse Fourier transform are, respectively,

$$\begin{aligned}\tilde{\psi}(\omega, \boldsymbol{k}) &= (2\pi)^{-n/2} \int e^{i\omega t - i\boldsymbol{k}\boldsymbol{x}} \psi(t, \boldsymbol{x}) \, \mathrm{d}t \, \mathrm{d}^n \boldsymbol{x}, \\ \psi(t, \boldsymbol{x}) &= (2\pi)^{-n/2} \int e^{-i\omega t + i\boldsymbol{k}\boldsymbol{x}} \tilde{\psi}(\omega, \boldsymbol{k}) \, \mathrm{d}\omega \, \mathrm{d}^n \boldsymbol{k}.\end{aligned} \tag{B.7}$$

This convention extends to the case $n = 1$. That is, if a function ψ depends only on time, then its FT is taken to be $\tilde{\psi}(\omega) = \frac{1}{2\pi} \int e^{i\omega t} \psi(t) \mathrm{d}t$, whereas if the single parameter is interpreted as a spatial coordinate or a generic parameter, then $\tilde{\psi}(k) = \frac{1}{2\pi} \int e^{-ikx} \psi(x) \mathrm{d}x$.

B.2.4 Finite Fourier transform

Occasionally, we'll come across the finite Fourier transform. It is defined for functions $\psi : \mathbb{Z}_N \to \mathbb{C}$, where $\mathbb{Z}_N = \{0, \ldots, N-1\}$ with arithmetic done modulo N. The standard basis on this space is given by delta functions $\delta_x(y) = \delta_{xy}$ so that

$$|\psi\rangle = \sum_{x \in \mathbb{Z}_N} \psi(x) |\delta_x\rangle.$$

The Fourier basis is

$$|\phi_k\rangle = \frac{1}{\sqrt{N}} \sum_{x\in\mathbb{Z}_N} e^{ikx}|\delta_x\rangle, \qquad k \in \frac{2\pi}{N}\mathbb{Z}_N.$$

The Fourier transform and its inverse thus take the form

$$\tilde{\psi}(k) = \langle\phi_k|\psi\rangle = \frac{1}{\sqrt{N}} \sum_{x\in\mathbb{Z}_N} e^{-ikx}\psi(x), \quad \psi(x) = \langle\delta_x|\psi\rangle = \frac{1}{\sqrt{N}} \sum_{k\in\frac{2\pi}{N}\mathbb{Z}_N} e^{ikx}\tilde{\psi}(k).$$

The theory developed above can be easily translated to the finite case.

This appendix is taken from lectures notes on Advanced Quantum Mechanics. Much of the material here, will migrate into the main body of the notes, as the course goes on.

B.3 Complex coordinates for classical harmonic oscillator

Let's retrace the solution of a harmonic oscillator

$$H = \frac{P^2}{2m} + \frac{1}{2}m\omega^2 X^2.$$

in *classical* mechanics. Choose problem-adapted units for length and momentum:

$$\tilde{X} = \sqrt{\frac{m\omega}{\hbar}}X, \qquad \tilde{P} = \sqrt{\frac{1}{m\hbar\omega}}P \quad \Rightarrow \quad H = \frac{1}{2}\hbar\omega(\tilde{P}^2 + \tilde{X}^2).$$

Wait, what's $\hbar$ doing in a classical calculation? Well, it's convenient to work with dimensionless quantities $\tilde{X}, \tilde{P}$. But then $XP/(\tilde{X}\tilde{P})$ is a constant having the dimension of an action. There's no preferred scale of action in classical mechanics – but $\hbar$ does the job and facilitates the later transition to QM.

Next, introduce complex coordinates

$$a := \frac{1}{\sqrt{2}}(\tilde{X} + i\tilde{P}) \quad \Rightarrow \quad a^\dagger = \frac{1}{\sqrt{2}}(\tilde{X} - i\tilde{P}),$$

where we use the "dagger" superscript to denote complex conjugation. These complex coordinates may not have a direct physical interpretation, but they are easy to work with and we can recover the original position and momentum coordinates as

$$X = \sqrt{\frac{\hbar}{2m\omega}}(a + a^\dagger) = \sqrt{\frac{2\hbar}{m\omega}}\,\mathrm{Re}(a), \quad P = -i\sqrt{\frac{m\hbar\omega}{2}}(a - a^\dagger) = \sqrt{2m\hbar\omega}\,\mathrm{Im}(a).$$

The Poisson bracket $\{X, P\} = 1$ implies

$$\{a, a^\dagger\} = \frac{1}{2}\big(-i\{\tilde{X}, \tilde{P}\} + i\{\tilde{P}, \tilde{X}\}\big) = \frac{1}{i\hbar},$$

so the coordinate change $(X, P) \to (a, a^\dagger)$ is canonical up to the factor $1/(i\hbar)$. The Hamilton function reads in complex coordinates

$$H = \frac{1}{2}\hbar\omega(aa^\dagger + a^\dagger a) = \hbar\omega|a|^2. \tag{B.8}$$

and the equations of motion are (using standard properties of Poisson brackets)

$$\partial_t a = \{a, H\} = \hbar\omega\{a, a^\dagger a\} = \hbar\omega(a^\dagger\{a, a\} + \{a, a^\dagger\}a) = -i\omega a,$$

solved by $a(t) = a(0)e^{-i\omega t}$.

Appendix C

Miscellaneous Integrals

C.1 Gaussian and Fresnel integrals

Starting point is the famous formula due to Gauss

$$\int_{-\infty}^{\infty} e^{-x^2}\,\mathrm{d}x = \sqrt{\pi},$$

which can be obtained by evaluating its square in polar coordinates.

From there, we one finds the general form

$$\int_{-\infty}^{\infty} e^{-\alpha x^2+\beta x+\gamma}\,\mathrm{d}x = \sqrt{\frac{\pi}{\alpha}} e^{\frac{\beta^2}{4\alpha}+\gamma} \tag{C.1}$$

which holds for all complex α, β, γ such that the integral converges: either $\mathrm{Re}[\alpha] > 0$; or $\mathrm{Re}[\alpha] = 0$ and $\mathrm{Re}[\beta] = 0$ (though in the latter case, the integral is not absolutely convergent, so it should be handled with care). In the formula, $\sqrt{\pi/\alpha}$ is the *principal square root*, defined to be the unique root with argument in $(-\pi, \pi]$. For real α, β, γ's, the above can be proven by completing the square and using the substitution rule. For complex coefficients, one has to use a suitable contour integration. The special case $\alpha = \mp i$ and $\beta = \gamma = 0$ is the complex asymptotic *Fresnel integral*

$$\int_{-\infty}^{\infty} e^{\pm i x^2}\,\mathrm{d}x = \sqrt{\pi} e^{\pm i\pi/4}. \tag{C.2}$$

The Gaussian integral (C.1) is taken over the entire real real line $x \to \pm\infty$, but in fact, is already close to its asymptotic value if the limits of the integral are large compared to $\sqrt{|\alpha|}$. This is obvious if α has a large real part (because the absolute value of the integrand is decaying with $e^{-\mathrm{Re}\,\alpha x^2}$). Imaginary parts of α also aid convergence, but for a more subtle reason: They cause the integrand to oscillate rapidly for large arguments, so that its contributions to the integral tend to cancel.

To visualize this effect, consider the non-asymptotic real Fresnel integrals

$$C(x) := \int_0^x \cos(t^2)\,\mathrm{d}t, \qquad S(x) := \int_0^x \sin(t^2)\,\mathrm{d}t.$$

Separating real and imaginary parts in (C.2) gives

$$\lim_{x\to\infty} C(x) = \lim_{x\to\infty} S(x) = \sqrt{\frac{\pi}{8}}. \tag{C.3}$$

Their convergence is shown in (Fig. C.1).

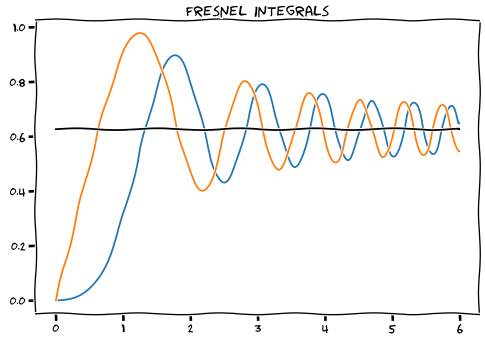


Figure C.1: The Fresnel integrals $C(x)$ (orange) and $S(s)$ (blue). The integral quickly converges towards its asymptotic value $\sqrt{\pi/8}$ (black line), with contributions of larger arguments canceling to the oscillating behavior of the integrand.

C.2 Some Fourier transforms

Rotationally invariant functions

Let $V(\boldsymbol{x}) = V(\|x\|)$ be a rotationally-invariant function in $\mathbb{R}^3$. Its Fourier transform $\tilde{V}(\boldsymbol{k})$ is computed most easily in the coordinate system $(r, \mu = \cos\theta, \phi)$ where (r, θ, ϕ) are spherical coordinates with polar vector parallel to $\boldsymbol{k}$. The volume element is

$$r^2 \sin\theta \, \mathrm{d}r \, \mathrm{d}\theta \, \mathrm{d}\phi = r^2 \, \mathrm{d}r \, \mathrm{d}\mu \, \mathrm{d}\phi \tag{C.4}$$

so that the Fourier transform

$$\begin{aligned}
\tilde{V}(\boldsymbol{k}) &= (2\pi)^{-3/2} \int e^{-i\boldsymbol{k}\boldsymbol{x}} V(r) \, \mathrm{d}^3\boldsymbol{x} \\
&= (2\pi)^{-1/2} \int_0^\infty \mathrm{d}r \, r^2 \, V(r) \int_{-1}^1 \mathrm{d}\mu \, e^{-ikr\mu} \\
&= (2\pi)^{-1/2} \int_0^\infty \mathrm{d}r \, r^2 V(r) \left[\frac{e^{-ikr\mu}}{-ikr} \right]_{\mu=-1}^1 \\
&= \frac{i}{(2\pi)^{1/2} k} \int_0^\infty \mathrm{d}r \, r V(r) \left(e^{-ikr} - e^{ikr} \right) \qquad \text{(C.5)} \\
&= \sqrt{\frac{2}{\pi}} \frac{1}{k} \int_0^\infty r V(r) \sin(kr) \, \mathrm{d}r \qquad \text{(C.6)}
\end{aligned}$$

reduces to a one-dimensional integral.

Appendix D

Function spaces and distributions

In this chapter, we take a more pedantic look at the function spaces that occur in QM. For simplicity of presentation, we'll mainly restrict attention to the one-dimensional case.

D.1 Square-integrable functions

What mathematical properties should a "wave function" $\psi : \mathbb{R} \to \mathbb{C}$ for a particle in one dimension have?

First, according to the Born interpretation, $p(x) := |\psi(x)|^2$ is the probability density describing the distribution of position measurement outcomes. For this interpretation to make sense, $\int |\psi(x)|^2 \, \mathrm{d}x$ must equal 1.

Next, physical predictions depend on ψ only through integrals. Integrals stay the same if the value of the integrand is changed on a set of measure zero. Therefore, two functions that agree *almost everywhere* (i.e. everywhere except on a set of measure zero) define the same physical state and should therefore be identified. For any function $\psi : \mathbb{R} \to \mathbb{C}$, write $[\psi]$ for the set of functions that agree with ψ almost everywhere.

These two conditions suggest that wave functions should belong to the space

$$L^2(\mathbb{R}) = \left\{ [\psi] \,\middle|\, \psi : \mathbb{R} \to \mathbb{C}, \int |\psi(x)|^2 \, \mathrm{d}x < \infty \right\}$$

of equivalence classes of *square-integrable* functions.[1] This is indeed the standard choice.

In practice, the identification of functions agreeing almost everywhere is usually left implicit. That is, $L^2(\mathbb{R})$ is called "the space of square-integrable functions" instead of the more precise "space of equivalence classes of square-integrable functions", and one writes $\psi \in L^2(\mathbb{R})$ as a short-hand for $[\psi] \in L^2(\mathbb{R})$. We will also follow this convention.

The *Cauchy-Schwarz inequality* says that

$$\left| \int \psi(x)^* \phi(x) \, \mathrm{d}x \right| \leq \left| \int \psi(x)^* \psi(x) \, \mathrm{d}x \right|^{1/2} \left| \int \phi(x)^* \phi(x) \, \mathrm{d}x \right|^{1/2}, \qquad \text{(D.1)}$$

so that

$$\langle \psi | \phi \rangle := \int \psi(x)^* \phi(x) \, \mathrm{d}x$$

[1] See any textbook on analysis, e.g. Folland's *Modern analysis*, Chapter 2 for more details on integration theory. Just two comments on terminology: (1) All integrals in the theory of function spaces are to be understood in the sense of Lebesgue. (2) A function f is *integrable* if the integral $\int f$ exists *and* is finite. (So, counter-intuitively, "f is integrable" and "the integral of f exists" are different statements!)

is well-defined as a sesquilinear form $L^2(\mathbb{R}) \to \mathbb{C}$.

Remarks on the use of equivalence classes

Identifying functions that lead to the same physical predictions makes sense. Let's reiterate, though, that consequently elements of $L^2(\mathbb{R})$ aren't strictly speaking functions, but rather *equivalence classes* of functions. In particular, "the value $\psi(x)$" of an element $[\psi] \in L^2(\mathbb{R})$ at a point x is *not* a well-defined concept! This might be surprising, because in practice, we work with point-wise values $\psi(x)$ all the time. We get away with this because either: (1) We use $\psi(x)$ in a context (e.g. under an integral) where it does not matter which representative of the equivalence class has been chosen. Or, (2), there is an (implicit) convention that fixes a representative. For example, it is easy to see that every equivalence class contains at most one continuous function (Fig. **??**). Thus, if we agree to use continuous representatives whenever possible, there is no ambiguity for such classes.

The identification also makes the mathematical theory cleaner. For example, for a function $\psi : \mathbb{R} \to \mathbb{C}$, the integral $\|[\psi]\|^2 := \int |\psi(x)|^2 \,\mathrm{d}x$ vanishes if and only if ψ is supported on a set of measure zero, i.e. iff $[\psi] = [0]$. The implication $\|[\psi]\| = 0 \Rightarrow [\psi] = 0$ is part of the mathematical definition of a *norm*. It is frequently invoked in physics arguments: For example, in the algebraic treatment of the harmonic oscillator, one typically shows that $\|a|0\rangle\| = 0$ and concludes that $a|0\rangle = 0$, i.e. that the attempt to construct a negative-energy eigenstate by laddering leads to the 0 function.

D.1.1 Why go beyond L^2?

The choice of $L^2(\mathbb{R}^n)$ as the space of wave functions was physically well-motivated. But it turns out that for the purpose of doing some calculations, it is "too small", while for others, "too large".

Too small: $L^2(\mathbb{R})$ does not contain the eigenfunctions of some important operators. The eigenfunctions of the momentum operator are plane waves, which have norm ∞, and therefore do not belong to L^2. The eigenfunctions of the position operator are supported only on one single point. As elements of L^2, they are therefore equivalent to the function that is identically 0.

Too large: $L^2(\mathbb{R})$ contains elements for which important operators are undefined. For example, there are classes $[\psi] \in L^2(\mathbb{R})$ that do not contain any continuous representative, in which case the action of the momentum operator is not well-defined. For an example involving the position operator, take the function

$$\psi(x) = \frac{1}{\sqrt{\pi}(x+i)}. \tag{D.2}$$

Then

$$\int |\psi(x)|^2 \,\mathrm{d}x = \frac{1}{\pi} \int \frac{1}{x^2+1} \,\mathrm{d}x = \frac{1}{\pi} \left[\arctan(x)\right]_{-\infty}^{\infty} = 1,$$

so $\psi \in L^2(\mathbb{R})$. But (by comparison with $\int_a^\infty 1/x \,\mathrm{d}x = \infty$), one can easily see that the integral $\langle\psi|X^k|\psi\rangle = \int \frac{x^k}{x^2+1} \,\mathrm{d}x$ is infinite for even $k \in \mathbb{N}$ and undefined for odd $k \in \mathbb{N}$. In particular, $\|X\psi\|^2 = \langle\psi|X^2|\psi\rangle = \infty$, implying that $X\psi \notin L^2(\mathbb{R})$.

Figure D.1: *Rigged Hilbert spaces* are "rigged" in the sense of "fully equipped" (like Imperator Furiosa's *War Rig*, pictured above), not in the sense of "manipulated with the goal to deceive", like a loaded die. (OK, *maayybe* I was just looking for an excuse to include that picture in my lecture notes).

Discussion

Do these issues mean that $L^2(\mathbb{R})$ is not an appropriate mathematical model for the space of wave functions? Arguably not!

For the eigenfunction examples, note that infinitely extended or infinitely concentrated states are unphysical, so we cannot complain that the space $L^2(\mathbb{R})$, designed to model physical wave functions, does not contain them.

Now let's look at the function ψ defined in (D.2). The fact that $X\psi \notin L^2(\mathbb{R})$ does not mean that position measurements aren't well-defined. To the contrary, $p(x) = |\psi(x)|^2 = \frac{1}{\pi(x^2+1)}$ is a perfectly good probability density describing position measurement outcomes. It's just that none of the *moments* $\langle X^k \rangle$ (including the expectation value, $k = 1$) exist and are finite. But nobody ever promised us that all probability distributions can be characterized via moments, so there is no fundamental issue with this. Likewise, any $\psi \in L^2(\mathbb{R})$, even if it exhibits discontinuities, has a Fourier transform $\tilde{\psi}$, and thus a probability density $p(\hbar k) = |\tilde{\psi}(k)|^2$ over momentum measurement outcomes.

However, the discussion does suggest that for the purpose of doing calculations, it would be good to identify a "sandwich of spaces"

$$\Phi \subset L^2(\mathbb{R}) \subset \Phi', \tag{D.3}$$

where Φ is "sufficiently small" that all relevant operators are well-defined on it, and Φ' is "large enough" that it contains a complete set of eigenvectors for all relevant operators.

As we'll see, the spaces Φ and Φ' are usually constructed together. Elements of Φ are called *test functions* and those of Φ' *distributions*. Constellations as in (D.3) are studied as *Gelfand triples* or *rigged Hilbert spaces* (Fig. D.1)).

Which spaces of functions are the best choice for Φ, Φ' depends on the problem one wants to solve. An important set for quantum mechanics is *Schwartz space* (after Laurent Schwartz, not to be confused with Hermann Schwarz of Cauchy-Schwarz-inequality fame) for Φ and the associated space of *tempered distributions* for Φ'. We'll look at this case next, and briefly sketch the general theory in Sec. D.3.

> ***Remark***. The *domain* D of a function f is the set of mathematical objects on which f is defined. One equivalently says that "f is a function on D"... ...except in the theory of Hilbert spaces. We have seen above that P and X are not defined on certain elements of the Hilbert space $L^2(\mathbb{R})$ – their domains $D(P), D(X)$ are strictly smaller. But one still says that "X is the position operator on $L^2(\mathbb{R})$". In general, if T is any linear operator whose domain $D(T)$ is dense in a

Hilbert space $\mathcal{H}$, it is customary in functional analysis to call T a *linear operator on* $\mathcal{H}$. (Of course, physicists don't worry about such details at all).

D.2 Distributions

D.2.1 Schwartz space

The most important set of test functions Φ in QM is *Schwartz space* $\mathcal{S}$, the "smooth functions whose derivatives vanish rapidly":

$$\mathcal{S} = \Big\{ \phi \in C^\infty(\mathbb{R}) \,\Big|\, \forall \alpha, \beta \in \mathbb{N}_0 : \sup_x |x^\alpha \partial_x^\beta \phi(x)| < \infty \Big\}. \tag{D.4}$$

The condition $\phi \in C^\infty(\mathbb{R})$ means that elements of Schwartz space are infinitely differentiable; while the second condition says that ϕ and its derivatives vanish faster than any polynomial function as $|x| \to \infty$. It follows that $\mathcal{S}$ is invariant under P and X. It is also easy to see that any square-integrable function can be arbitrarily-well approximated by Schwartz-class functions, i.e. for every $\psi \in L^2(\mathbb{R})$ and every $\epsilon > 0$, there exists a $\phi \in \mathcal{S}$ such that $\|\psi - \phi\| \leq \epsilon$. (Technically: $\mathcal{S}$ is *dense* in $L^2(\mathbb{R})$ w.r.t. norm topology).

This already solves half of our problems: Because well-behaved functions are dense, there is little loss of generality in assuming that any wave function of physical interest lies in $\mathcal{S}$. One can then apply X and P without any issue.

D.2.2 Tempered distributions

Constructing the space that contains the generalized eigenvectors requires us to to take a little detour: We will first have to study *linear functionals* $\mathcal{S} \to \mathbb{C}$.

A function $u : \mathbb{R} \to \mathbb{C}$ is *locally integrable* if for any compact set $K \subset \mathbb{R}$,

$$\int_K |u(x)| \, \mathrm{d}x < \infty.$$

For example, continuous functions are locally integrable, whereas $1/x$ isn't (e.g. $\int_0^1 |\frac{1}{x}| \, \mathrm{d}x = \infty$). Now, for any locally integrable function u that grows at most polynomially as $|x| \to \infty$, and for any $l \in \mathbb{N}_0$, define a functional $T_{D^l u} : \mathcal{S} \to \mathbb{C}$ by

$$T_{D^l u}(\phi) := \int u(x)(-\partial_x)^l \phi(x) \, \mathrm{d}x. \tag{D.5}$$

(The notation $D^l u$ will be explained below). Then $T_{D^l u}$ is well-defined as a linear functional $\mathcal{S} \to \mathbb{C}$. That's because $\phi \in \mathcal{S}$ implies that $\partial_x^l \phi \in \mathcal{S}$ as well; local integrability of u and continuity of $\partial_x^l \phi$ implies that the integrand is locally integrable; and finally fact that $\partial_x^l \phi$ vanishes faster than any polynomial, together with the matching growth restriction on u, means that the integral remains finite as $|x| \to \infty$. A functional of this form is called a *tempered distribution*, and the space of all tempered distributions is denoted by $\mathcal{S}'$.

In contrast, note that $T_{D^l u}$ is rarely well-defined as a functional on $L^2(\mathbb{R})$. For one, elements $\psi \in L^2(\mathbb{R})$ aren't in general differentiable, and even if they are, they generally vanish too slowly for the integral to converge. So we see that $\mathcal{S}$, on account of being *smaller* than $L^2(\mathbb{R})$, allows for a *larger* set of linear functionals! Recall that we're out to find a set larger than $L^2(\mathbb{R})$, so this seems like a promising direction to explore. Let's look at some examples.

Plane waves: $T_{e^{ikx}}$ defines a linear functional on Schwartz space, but, because e^{ikx} is not normalizable, not on $L^2(\mathbb{R})$.

Delta functional: Let $\theta(x)$ be the step function that is 0 for $x < 0$ and 1 for $x \geq 0$. Then, using integration by parts,

$$T_{D\theta(x)}(\phi) = -\int \theta(x)\partial_x\phi(x)\,\mathrm{d}x = -\int_0^\infty \partial_x\phi(x)\,\mathrm{d}x = \phi(0). \tag{D.6}$$

The operation only makes sense for functions ϕ that are differentiable at 0 – so certainly for elements of $\mathcal{S}$, but not necessarily elements of $L^2(\mathbb{R})$.

"Bra vectors": For every $\psi \in L^2(\mathbb{R})$, the "bra" $\phi \mapsto \langle\psi|\phi\rangle = T_{\psi^*}$ defines a tempered distribution. (Indeed, every square-integrable function is also locally integrable. That's an easy consequence of the Cauchy-Schwarz inequality).

The principal value is important in the theory of partial differential equations, where one often wants to associate a distribution with the function $u(x) = \frac{1}{x}$ in some way. Unfortunately, $\frac{1}{x}$ is not locally integrable, and indeed, $\int \frac{\phi(x)}{x}\,\mathrm{d}x$ does not in general exist. But as we'll see, the *principal value*

$$\mathrm{pv}\left(\frac{1}{x}\right)(\phi) := \lim_{\epsilon\to 0^+}\int_{\mathbb{R}\setminus(-\epsilon,\epsilon)} \frac{\phi(x)}{x}\,\mathrm{d}x \tag{D.7}$$

is finite for all $\phi \in \mathcal{S}$ and, what is more, is given by the tempered distribution $T_{D\log|x|}(\phi)$. To see that this makes sense, we first need to convince ourselves that $\log|x|$, even though it diverges as $x \to 0$, is locally integrable. This follows from the fact that the anti-derivative of $\log|x|$ is $F(x) = x\log|x| - x + C$, which remains finite at the singularity: $\lim_{x\to 0} F(x) = C$. Therefore, $T_{D\log|x|}$ is indeed a tempered distribution. It remains to be shown that it evaluates to the principal value:

$$\begin{aligned}
T_{D\log|x|}(\phi) &= -\int \log|x|\,\phi'(x)\,\mathrm{d}x \\
&= \lim_{\epsilon\to 0^+}\left(-\int_{-\infty}^{-\epsilon}\log(-x)\,\phi'(x)\,\mathrm{d}x - \int_\epsilon^\infty \log x\,\phi'(x)\,\mathrm{d}x\right) \\
&= \lim_{\epsilon\to 0^+}\left(\int_{-\infty}^{-\epsilon}\frac{\phi(x)}{x}\,\mathrm{d}x - \phi(-\epsilon)\log\epsilon + \int_\epsilon^\infty \frac{\phi(x)}{x}\,\mathrm{d}x + \phi(\epsilon)\log\epsilon\right) \\
&= \mathrm{pv}\left(\frac{1}{x}\right)(\phi) + \lim_{\epsilon\to 0^+}\log(\epsilon)(\phi(\epsilon) - \phi(-\epsilon)) \\
&= \mathrm{pv}\left(\frac{1}{x}\right)(\phi) + 2\phi'(0)\underbrace{\lim_{\epsilon\to 0^+}\epsilon\log(\epsilon)}_{=0} = \mathrm{pv}\left(\frac{1}{x}\right)(\phi).
\end{aligned}$$

Powers of $1/r$ in higher dimensions: In contrast to the previous example, $u(\boldsymbol{x}) = \|\boldsymbol{x}\|^{-k}$ *is* locally integrable as a function on $\mathbb{R}^n$, as long as $n > k$. This can be seen by switching to n-dimensional spherical coordinates, where the volume element is proportional to r^{n-1}, which lifts the singularity at 0. The definition (D.5) is easily adapted to higher dimensions, and integrating against such a u thus defines a tempered distribution. Unsurprisingly, the case $k = 1, n = 3$ is important due to its relation to the Coulomb and the gravitational potential.

Regular distributions

Distributions of the form T_u (i.e. those that can be expressed without differentiating the argument before integrating) are called *regular*. For regular distributions, it is common

to use the same symbol for both the distribution $\mathcal{S} \to \mathbb{C}$ and for the function $\mathbb{R} \to \mathbb{C}$ defining it:

$$T(\phi) = \int T(x)\phi(x)\,\mathrm{d}x. \tag{D.8}$$

You might complain that such an overloading of notation is not a nice thing to do. And you'd be right. But things are about to get worse. Such a convention is even used for non-regular distributions!

Consider e.g. the delta distribution $\delta(\phi) = \phi(0)$ discussed above. It is not regular. (Because a hypothetical function giving rise to it would have to be zero everywhere except at $x = 0$ – but an integral over a function supported on only one point is zero). But, in analogy to (D.8), one still writes

$$\delta(\phi) = \int \delta(x)\phi(x)\,\mathrm{dx}.$$

The r.h.s. is *not* an integral and $\delta(x)$ *not* a function – the entire r.h.s. is to be read as an elaborate notation for $\delta(\phi)$. Whether this convention is genius (because it allows practitioners to work with distributions without having to learn the abstract theory) or horrific (because the *one* job of mathematics is to be rigorous and not to pretend that objects exist when in fact they don't) is a question that may be controversially debated.

D.2.3 Operations on distributions

Our goal is still to find generalized eigenvectors for X and P. These will turn out to be distributions. For that to even make sense, we have to define what it means for an operator to act on distributions.

Let A be any operator that maps $\mathcal{S}$ to $\mathcal{S}$. There is a unique operator A^t, the *transpose* of A, such that, for $\phi, \psi \in \mathcal{S}$,

$$\int (A\psi)(x)\phi(x)\,\mathrm{d}x = \int \psi(x)(A^t\phi)(x)\,\mathrm{d}x.$$

(This is the bilinear analogue of the definition of the adjoint for sesquilinear inner products). It directly follows that for regular distributions with $u \in \mathcal{S}$, $T_{Au}(\phi) = T_u(A^t\phi)$. Using the notation in (D.8), this means

$$(AT)(\phi) = T(A^t\phi). \tag{D.9}$$

We take Eq. (D.9) as the general definition for the action of an operator on distributions. In words: Operations on distributions are defined by shifting them onto the argument.

Derivatives of distributions

The most important application is the *differentiation operator* $(D\phi)(x) = \partial_x\phi(x)$. By partial integration, $D^t = -D$ from which we get

$$(DT_u)(\phi) = \int u(x)(-\partial_x^l)\phi(x)\,\mathrm{d}x = T_{Du}(\phi)$$

and, more generally, $D^l T_u = T_{D^l u}$ (which justifies the notation $D^l u$, as promised).

With these conventions in place, we can explain the notion of "derivative in the sense of distribution" that you will likely have come across before. Take for example the step

function θ. Seen as a function $\mathbb{R} \to \mathbb{C}$, it is not differentiable, due to the discontinuity at 0. But the distribution T_θ does have a derivative: $DT_\theta = \delta$, as computed in (D.6). Identifying θ with T_θ, this fact is often expressed as "$\partial_x \theta(x) = \delta(x)$ in the sense of distribution". Note that every locally integrable function is infinitely differentiable in the sense of distribution (by virtue of the elements of the test function space $\mathcal{S}$ having this property).

As an application, let's derive a famous identity that expresses the principal value in terms of a "side limit of a deformed version of $1/x$", namely

$$\lim_{\epsilon \to 0^+} \frac{1}{x \pm i\epsilon} = \mathrm{pv}\left(\frac{1}{x}\right) \mp i\pi\delta. \tag{D.10}$$

First, recall that the principal complex logarithm

$$\mathrm{Log}(x + iy) = \log \sqrt{x^2 + y^2} + i \arctan \frac{x}{y}$$

is an analytic continuation of the logarithm to the complex numbers, except for a branch cut on the negative real axis (Fig. **??**). It follows that

$$\lim_{\epsilon \to 0^+} \mathrm{Log}(x \pm i\epsilon) = \log |x| \mp i\pi\theta(x)$$

which immediately implies (D.10) by differentiating both sides in the sense of distribution.

Generalized eigenvectors

We say that a distribution T is a *generalized eigenvector* of an operator $A : \mathcal{S} \to \mathcal{S}$ if

$$A\,T = \lambda\,T.$$

Plane waves are therefore eigenvectors of the differentiation operator D or the momentum operator $P = -iD$:

$$D\,T_{e^{ikx}} = T_{\partial_x e^{ikx}} = T_{ike^{ikx}} = ik\,T_{e^{ikx}}, \qquad P\,T_{e^{ikx}} = k\,T_{e^{ikx}}.$$

Likewise, if $\delta_a = \partial_x \theta(x - a) : \phi \mapsto \phi(a)$ is the delta distribution at $a \in \mathbb{R}$, then

$$(X\,\delta_a)(\phi) = \delta_a(X\phi) = a\phi(a) = a\,\delta_a(\pi) \quad \Rightarrow \quad X\,\delta_a = a\,\delta_a.$$

So, with all these preparations in the bag, it was pretty easy to identify the generalized eigenvectors!

Interface conditions for piece-wise continuous potentials. In one dimension, the time-independent Schrödinger equation

$$\left(-\frac{\hbar^2}{2m}\partial_x^2 + V(x) - E\right)\psi(x) = 0 \tag{D.11}$$

is an ordinary differential equation. The Picard-Lindelöf theorem says that if V is Lipschitz continuous, then the ODE can be solved.

A staple of introductory QM lectures are potentials that are only piece-wise continuous. In this case, there aren't necessarily solutions to (D.11) in the ordinary sense. Here, we'll work out under which conditions one can stitch piece-wise solutions together to get a generalized eigenvector of the Hamiltonian.

We treat the case where (D.11) has ordinary solutions $\psi_-(x)$ on $(-\infty, 0)$ and $\psi_+(x)$ on $(0, \infty)$. Assume that both solutions and their first derivatives are continuous and bounded around 0, and can thus be extended to 0. We also require that V is bounded around 0. Then

$$\psi(x) := \begin{cases} \psi_-(x) & x \leq 0, \\ \psi_+(x) & x > 0 \end{cases}$$

is a generalized eigenvector if and only if, for all test functions ϕ,

$$\int_{-\infty}^{\infty} \phi(x) \left(-\frac{\hbar^2}{2m} \partial_x^2 + V(x) - E \right) \psi(x) \, \mathrm{d}x = 0.$$

Because ψ is an ordinary solution away from zero, the integral is equal to

$$\lim_{\epsilon \to 0} \int_{-\epsilon}^{\epsilon} \phi(x) \left(\frac{-\hbar^2}{2m} \partial_x^2 + V(x) - E \right) \psi(x) \, \mathrm{d}x = \frac{-\hbar^2}{2m} \lim_{\epsilon \to 0} \int_{-\epsilon}^{\epsilon} \phi(x) \partial_x^2 \psi(x) \, \mathrm{d}x,$$

where we have used that $(V(x) - E)$ does not contribute to the integral in the limit, as ϕ, ψ, and V are bounded around 0. Integrating by parts,

$$\begin{aligned} \lim_{\epsilon \to 0} \int_{-\epsilon}^{\epsilon} \phi(x) \psi''(x) \, \mathrm{d}x =& \phi(0) \big(\psi'_+(0) - \psi'_-(0) \big) - \lim_{\epsilon \to 0} \int_{-\epsilon}^{\epsilon} \phi'(x) \psi'(x) \, \mathrm{d}x \\ =& \phi(0) \big(\psi'_+(0) - \psi'_-(0) \big) - \phi'(0) \big(\psi_+(0) - \psi_-(0) \big), \end{aligned}$$

which vanishes for all ϕ if and only if the *interface conditions*

$$\psi_+(0) = \psi_-(0) \quad \text{and} \quad \psi'_+(0) = \psi'_-(0) \tag{D.12}$$

are met. Notably, these do not imply that ψ is twice differentiable, which would be required for ordinary solutions to (D.11).

Fourier transforms of tempered distributions

Because the Fourier transform exchanges X and P, the characterization (D.4) of $\mathcal{S}$, and hence the space itself, is invariant under Fourier transforms. Applying the general scheme (D.9), the Fourier transform of a tempered distribution T is $(\mathcal{F}T) : \phi \mapsto T(\mathcal{F}^t \phi)$.

To get more explicit formulas, first note that $\mathcal{F}^t = \mathcal{F}$:

$$\int (\mathcal{F}\psi)(k) \phi(k) \, \mathrm{d}k = \frac{1}{\sqrt{2\pi}} \int \int e^{-ikx} \psi(x) \phi(k) \, \mathrm{d}x \, \mathrm{d}k = \int \psi(x) (\mathcal{F}\phi)(x) \, \mathrm{d}x.$$

Thus, using the shorthand "tilde notation" for the Fourier transform,

$$\tilde{T}(\phi) = T(\tilde{\phi}).$$

Delta distribution. For δ, compute

$$\tilde{\delta}(\phi) = \delta(\tilde{\phi}) = \tilde{\phi}(0) = \int \frac{1}{\sqrt{2\pi}} \phi(x) \, \mathrm{d}x = T_{\frac{1}{\sqrt{2\pi}}}(\phi), \tag{D.13}$$

that is, the FT of δ is a regular distribution, arising from the constant function

$$\tilde{\delta}(k) = \frac{1}{\sqrt{2\pi}}. \tag{D.14}$$

One could be tempted to use the following formal calculation to arrive at the same conclusion:

$$\hat{\delta}(k) = \text{“}\frac{1}{\sqrt{2\pi}}\int e^{-ikx}\delta(x)\,\mathrm{d}x\text{”} = \frac{1}{\sqrt{2\pi}}.$$

But, unlike, (D.13), this is *not* a rigorous argument given our development of the theory so far! That's because we have defined $\delta(\phi)$ only for $\phi \in \mathcal{S}$, but e^{-ikx} is most definitely not an element of Schwartz space. The integral is therefore only heuristically defined. One can sometimes make sense of products of distributions – but the issue is subtle and we will not pursue it here.

Constant functions. The constant function $\underline{1}(x) = 1$ does not have a Fourier transform in the ordinary sense. For one, the integral $(2\pi)^{-1}\int \mathrm{d}x$ that would define $\tilde{\underline{1}}(0)$ is infinite. However, because $T_{\underline{1}}$ defines a tempered distribution, it does have a FT. Slightly abusing language once again, we call it the FT of $\underline{1}$ (in the sense of distribution).

We can find it by expressing $\mathcal{F}^{-1}$ in terms of $\mathcal{F}$ and applying it to (D.14). To this end, let Π be the *parity operator*, which mirrors functions about the origin: $(\Pi\phi)(x) = \phi(-x)$. Then it is easy to see that $\mathcal{F}^\dagger = \Pi\mathcal{F}^t$ and hence unitarity of $\mathcal{F}$ implies

$$(\Pi\mathcal{F})\mathcal{F} = \mathcal{F}^\dagger\mathcal{F} = \mathbb{1} \quad \Rightarrow \quad \mathcal{F}^{-1} = \Pi\mathcal{F}. \tag{D.15}$$

Applying this to (D.14) gives

$$\mathcal{F}\left(\underline{1}\right) = \sqrt{2\pi}\delta.$$

The principal value. From an easy contour integration, the FT of $1/(x+i\epsilon)$ is

$$\frac{1}{\sqrt{2\pi}}\int \frac{1}{x+i\epsilon}e^{-ikx}\,\mathrm{d}x = -i\sqrt{2\pi}e^{-\epsilon k}\theta(k) \to -i\sqrt{2\pi}\theta(k) \qquad (\epsilon \to 0^+).$$

Using (D.10), we then find that the FT of the principal value is a regular distribution:

$$\begin{aligned}\big(\mathcal{F}\ \mathrm{pv}(1/x)\big)(k) &= \lim_{\epsilon\to 0^+}\mathcal{F}\left(\frac{1}{x+i\epsilon}\right)(k) + i\pi\mathcal{F}(\delta)(k) \\ &= i\sqrt{\frac{\pi}{2}}\left(-2\theta(k)+1\right) = -i\sqrt{\frac{\pi}{2}}\,\mathrm{sign}(k).\end{aligned} \tag{D.16}$$

Combining this result with (D.15) gives further transforms of common distributions:

$$\left(\mathcal{F}\,\mathrm{sign}\right)(k) = i\sqrt{\frac{2}{\pi}}\,\mathrm{pv}(1/k) \tag{D.17}$$

$$\left(\mathcal{F}\theta\right)(k) = \frac{1}{2}\mathcal{F}\left(\mathrm{sign}+1\right)(k) = i\frac{1}{\sqrt{2\pi}}\left(\mathrm{pv}(1/k) - i\pi\delta\right). \tag{D.18}$$

Coulomb and Yukawa potentials. Up to constants, the Coulomb potential is $u(\boldsymbol{x}) = -\frac{1}{\|\boldsymbol{x}\|}$ in $\mathbb{R}^3$. Just like the constant function treated above, it does not have an ordinary Fourier transform. For example, $\tilde{u}(0)$ would be given by

$$-(2\pi)^{-3/2}\int \frac{1}{\|\boldsymbol{x}\|}\,\mathrm{d}^3\boldsymbol{x} = -2(2\pi)^{-1/2}\int r\,\mathrm{d}r = -\infty. \tag{D.19}$$

But as discussed in Sec. D.2.2, $u(\boldsymbol{x})$ defines a regular distribution whose Fourier transform turns out to be regular again, given by

$$\tilde{u}(\boldsymbol{k}) = -\sqrt{\frac{2}{\pi}}\frac{1}{\|\boldsymbol{k}\|^2}. \tag{D.20}$$

Here's how to find (D.20). Express

$$T_u(\tilde{\phi}) = -\int \frac{\tilde{\phi}(\boldsymbol{k})}{\|\boldsymbol{k}\|}\,\mathrm{d}^3\boldsymbol{k} = -\lim_{s\to 0^+}\int \frac{e^{-s\|\boldsymbol{k}\|}}{\|\boldsymbol{k}\|}\tilde{\phi}(\boldsymbol{k})\,\mathrm{d}^3\boldsymbol{k}. \tag{D.21}$$

as a limit of integrals involving the "regularizing" factor $e^{-s\|\boldsymbol{k}\|}$.

> This is valid, because the integral, interpreted as a function of $s \in [0,\infty)$, is continuous at 0. In fact, it is even differentiable:
>
> $$-\partial_s|_0 \int \frac{e^{-s\|\boldsymbol{k}\|}}{\|\boldsymbol{k}\|}\tilde{\phi}(\boldsymbol{k})\,\mathrm{d}^3\boldsymbol{k} = \int \tilde{\phi}(\boldsymbol{k})\,\mathrm{d}^3\boldsymbol{k}, \tag{D.22}$$
>
> which is finite for $\tilde{\phi} \in \mathcal{S}$. (Note that the same regularization does *not* work for the integral in (D.19), which formally corresponds to the case $\tilde{\phi}(\boldsymbol{k}) = 1$. Of course, the constant function is not an element of Schwartz space, and indeed, this choice would cause (D.22) to diverge).

Plugging in the definition of the FT and exchanging integrals,

$$T_u(\tilde{\phi}) = \lim_{s\to 0}\int \left(-(2\pi)^{-3/2}\int \frac{e^{-s\|\boldsymbol{k}\|}}{\|\boldsymbol{k}\|}e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\phi(\boldsymbol{x})\mathrm{d}^3\boldsymbol{k}\right)\mathrm{d}^3\boldsymbol{x}.$$

The expression in parentheses is the FT of

$$V(\boldsymbol{x}) = -\frac{1}{\|\boldsymbol{x}\|}e^{-s\|\boldsymbol{x}\|}$$

which, up to constants, is known as the *Yukawa potential*. Its Fourier transform follows from the general formula (C.5) for rotationally-invariant functions in terms of $k = \|\boldsymbol{k}\|$:

$$\tilde{V}(\boldsymbol{k}) = \frac{i}{(2\pi)^{1/2}k}\int_0^\infty \left(e^{-sr-ikr} - e^{-sr+ikr}\right)\mathrm{d}r.$$

The one-dimensional integral can immediately be solved as

$$-\left[\frac{e^{r(-s-ik)}}{-s-ik}\right]_0^\infty + \left[\frac{e^{r(-s+ik)}}{-s+ik}\right]_0^\infty = -\frac{1}{-s-ik} + \frac{1}{-s+ik} = -\frac{2ik}{s^2+k^2}.$$

Collecting constants, we get the FT of the Yukawa potential, which gives (D.20) as $s \to 0$:

$$\tilde{V}(\boldsymbol{k}) = -\sqrt{\frac{2}{\pi}}\frac{1}{s^2+k^2}. \tag{D.23}$$

Products and tensor products

If T is a tempered distribution, and v a smooth function that grows at most polynomially as $|x| \to \infty$, then the product vT between v and T is the tempered distribution

$$vT : \phi \mapsto T(v\phi). \tag{D.24}$$

The product between a smooth function and a distribution behaves mostly like the product between functions. In particular, if $T = T_u$ is regular, then $vT_u = T_{uv}$.

However, one cannot extend (D.24) to products between arbitrary distributions, while retaining the basic properties of "multiplication". For example

$$\begin{aligned} & \delta\, x = 0 && \text{by (D.24)} \\ \Rightarrow \quad & (\delta\, x)\,\mathrm{pv}(1/x) = 0\,\mathrm{pv}(1/x) = 0 && \text{by above and (D.24)} \end{aligned}$$

$$\begin{aligned} & x\,\mathrm{pv}(1/x) = 1 && \text{by (D.24)} \\ \Rightarrow \quad & \delta\,(x\,\mathrm{pv}(1/x)) = \delta\, 1 = \delta && \text{by above and (D.24)} \end{aligned}$$

so there is no associative way to assign a meaning to "$\delta\, x\,\mathrm{pv}(1/x)$".

Tensor products between distributions are perfectly well-defined, though. If S, T are distributions, then $S \otimes T$ is the bilinear form that sends $\phi, \psi \in \mathcal{S}$ to $S(\phi)T(\psi)$.

Using the fictional function notation $T(x)$ for distributions T (as in (D.8)), the situation can be summarized as: "$T(x)S(x)$ makes no sense, but $T(x)S(y)$ is unremarkable – just integrate against *two* test functions".

A propose bilinear forms. Take any distribution K on test functions on $\mathbb{R}^2$, not necessarily a tensor product. If ϕ, ψ are test functions on $\mathbb{R}$, then *their* tensor product $(\phi \otimes \psi)(x, y) = \phi(x)\psi(y)$ can be paired with K. This way, K, too, defines a bilinear form:

$$\phi, \psi \mapsto K(\phi \otimes \psi) \quad \text{also written as} \quad K(\phi \otimes \psi) = \int K(x, y)\phi(x)\psi(y)\,\mathrm{d}x\,\mathrm{d}y.$$

One says that K is the *integral kernel* of the bilinear map.

This definition can be extended straight-forwardly to multilinear or sesquilinear functions. The sesquilinear case is frequently used in quantum mechanics – c.f. Eq. (B.1).

D.3 Topological aspects, more pedantry, and generalizations

Our definition of tempered distributions in Eq. (D.5) was *constructive*: We showed how to build distributions concretely given a function u and derivatives ∂_x^l. The mathematical theory is usually formulated *axiomatically*. Distributions are defined indirectly, as linear functionals on test function spaces, subject to some abstract properties. These properties are phrased in the language of point set topology. In this section, we briefly introduce this more abstract point of view.

Topological spaces

Consider a set X. A *topology* on X is a rule that allows us to decide when a sequence $x_k : \mathbb{N} \to X$ converges to an element $x \in X$.

As a first example, assume that X is a vector space equipped with a norm $\|\cdot\|$. This covers an extremely wide range of spaces, from $X = \mathbb{R}$, the real numbers with norm $\|x\| = |x|$ the absolute value, to $X = L^2(\mathbb{R})$ with norm $\|x\| = \langle x|x\rangle^{1/2}$ derived from the inner product. We say that a sequence x_k *converges in norm topology* to x,

$$x_k \to x, \qquad \text{if} \qquad \lim_{k\to\infty} \|x_k - x\| = 0. \tag{D.25}$$

We'll use these concepts to give very general definitions of *continuity* and *completeness*.

Figure D.2: *Sequence continuity* is equivalent to the more familiar "ϵ-δ-definition" of continuity for functions $f : \mathbb{R} \to \mathbb{R}$. It is more general, though, and can also be applied to spaces whose topologies do not derive from a distance measure.

Continuity

A function f between two topological spaces is *continuous* if it maps convergent sequences to convergent sequences (Fig. D.2), i.e. if

$$x_k \to x \qquad \Rightarrow \qquad f(x_k) \to f(x).$$

Example (important!): *If $\psi \in L^2(\mathbb{R})$, then the linear functional $\langle\psi|$ is continuous.*
The proof reduces to the Cauchy-Schwarz inequality. If $\lim_{k\to\infty} \|\phi_k - \phi\| = 0$, then

$$\big|\langle\psi|\phi_k\rangle - \langle\psi|\phi\rangle\big| = \big|\langle\psi|(|\phi_k\rangle - |\phi\rangle)\big| \leq \|\psi\|^{1/2} \big\|\phi_k - \phi\big\|^{1/2} \to 0 \quad (k \to \infty).$$

Completeness and Hilbert spaces

Now for *completeness*. A sequence x_k is *Cauchy* if "its elements eventually become arbitrarily close" in the sense that

$$\forall\, \epsilon > 0,\ \exists\, n \in \mathbb{N} \text{ such that } \forall\, k, l > n, \text{ it holds that } \|x_k - x_l\| \leq \epsilon.$$

A space X is *complete* if every Cauchy sequence converges to an element of X.

As an example of a Cauchy sequence, let x_k be the approximation of $\sqrt{2}$ to k decimal places. This example shows that the rational numbers are not complete: There is no $q \in \mathbb{Q}$ such that $x_k \to q$ (for then q would have to equal $\sqrt{2}$, which, famously, is not rational).

In the mathematical literature, a *Hilbert space* is defined as a

- complex vector space,
- with a sesquilinear inner product $\langle\cdot|\cdot\rangle$,
- that is *complete* with respect to the norm derived from the inner product.

The final condition is often glossed over in physics presentations. It is important, though. For one, it means that series like

$$|\psi(t)\rangle = \sum_{k=0}^{\infty} \frac{1}{k!}(itH)^k|\psi(0)\rangle := \lim_{K\to\infty} \sum_{k=0}^{K} \frac{1}{k!}(itH)^k|\psi(0)\rangle,$$

used ubiquitously, are actually well-defined. Another reason is that the equivalence of "kets" and "bras" requires this property: The set of continuous linear functionals on a Hilbert space $\mathcal{H}$ is denoted by $\mathcal{H}'$. If $|\psi\rangle \in \mathcal{H}$, then we've shown above that $\langle\psi|$ is continuous, i.e. an element of $\mathcal{H}'$. The *Riesz representation theorem* says that the converse is also true: Every continuous linear functional of a Hilbert space is given by some "bra vector".

One can show that $L^2(\mathbb{R})$ is complete, i.e. actually a Hilbert space.

Contrast this with Schwartz space $\mathcal{S}$. It, too is a complex vector space with the same sesquilinear inner product as $L^2(\mathbb{R})$. But it is not complete in norm topology and hence no Hilbert space. The argument works just like the $\sqrt{2}$-example above. Because $\mathcal{S}$ is dense in $L^2(\mathbb{R})$, for every $\psi \in L^2(\mathbb{R})$, there exists a sequence $\phi_k : \mathbb{N} \to \mathcal{S}$ converging to ψ in norm. Thus, if $\psi \notin \mathcal{S}$, the sequence ϕ_k has no limit point in $\mathcal{S}$.

Topology on Schwartz space

Return to Schwartz space $\mathcal{S}$. Because it is a subspace of $L^2(\mathbb{R})$, we can use the norm topology also for $\mathcal{S}$. However, there's a second, important, topology on that space. For $\alpha, \beta \in \mathbb{N}_0$, define the (semi)-norms

$$\|\phi\|_{\alpha,\beta} := \sup_{x\in\mathbb{R}} |X^\alpha \partial_\beta \psi(x)|.$$

A sequence $\phi_k : \mathbb{N} \to \mathcal{S}$ converges with respect to this family of semi-norms,

$$\phi_k \underset{\mathcal{S}}{\to} \phi, \qquad \text{if} \qquad \lim_{k\to\infty} \|\phi - \phi_k\|_{\alpha,\beta} = 0 \quad \forall \alpha, \beta \in \mathbb{N}_0. \tag{D.26}$$

To avoid confusion, we'll write $\phi_k \underset{L^2}{\to} \phi$ if we mean convergence with respect to the Hilbert space norm. It is easy to see that $\phi_k \underset{\mathcal{S}}{\to} \phi$ implies $\phi_k \underset{L^2}{\to} \phi$, but not the other way round. One says that the topology (D.26) is *finer* than norm topology.

There's a non-trivial *regularity theorem* (Reed-Simon, Thm. V.10), which states that the constructive definition (D.5) of tempered distribution characterizes exactly the space of linear functionals on Schwartz space that is continuous in the sense of (D.26).

Generalizations

The topological formulation above is the basis of generalizations. The common recipe is to choose a test function space Φ (often norm-dense in $L^2(\mathbb{R})$), endow it with a finer topology, and then consider the continuous dual Φ'.

The most important choice is to take Φ to be the space of *bump functions* $C_c^\infty(\mathbb{R})$: smooth functions with compact support. "Compact support" means that these functions are identically zero for $|x|$ large enough. (It is not obvious that one can define functions that transition smoothly from being identically zero in some region to being non-zero in other regions, but such functions do exist). In the context of distributions, the space of bump functions is usually denoted by $\mathcal{D}$.

Recall that Schwartz functions vanish faster than any polynomial, and thus integrals against locally integrable functions $u(x)$ that grow at most polynomially are finite. Because bump functions vanish identically for large x, integrals against *any* locally integrable u are well-defined. This suggests, correctly, that the space of distributions $\mathcal{D}'$ is even larger than the space of tempered distributions $\mathcal{S}'$.

The structure of $\mathcal{D}'$ is somewhat more complicated than was the case for $\mathcal{S}'$. We will not discuss it here, but, for completeness, give the topology from which it derives. It is defined in terms of the (semi-)norms

$$\|\phi\|_{K,\alpha} := \sup_{x\in K} |\partial_x^\alpha \phi(x)|$$

indexed by compact subsets $K \subset \mathbb{R}$ and a number $n \in \mathbb{N}_0$, in the same way as (D.26):

$$\phi_k \underset{\mathcal{D}}{\to} \phi \qquad \text{if} \qquad \lim_{k\to\infty} \|\phi - \phi_k\|_{K,\alpha} = 0 \quad \forall \alpha, K.$$

Because elements of $\mathcal{D}$ are smooth, distributions in $\mathcal{D}'$ are arbitrarily often differentiable. However, the space $\mathcal{D}$ is easily seen not be be invariant under Fourier transforms, so the Fourier transform is not defined on $\mathcal{D}'$. This is the reason the space plays a less prominent role in quantum theory.

Terminology

The word "distributions" used without qualification is most likely to refer to $\mathcal{D}'$, but can also mean a general continuous dual space Φ', and may *also* refer to tempered distributions $\mathcal{S}'$, depending on context. Making matters worse, $\mathcal{S}$ is always called "Schwartz space", but the name "Schwartz" is also associated with the general mathematical theory of distributions and in particular also with $\mathcal{D}'$. "Tempered distributions" always means $\mathcal{S}'$, at least.

Lastly, if a science professor answers an inquiry about a questionable derivation by claiming that it is to be understood "in the sense of distribution", they likely mean neither $\mathcal{D}'$ nor Φ' nor $\mathcal{S}'$. Instead, they are probably vaguely aware of the fact that what they are doing isn't quite rigorous, but are optimistic that a smart mathematician could figure it out, and in any case, want to get through their lecture with their dignity intact and have found that "distribution" is a fully general incantation that reliably suppresses follow-up questions.

Needless to say, *I* would never engage in such tactics.