Quantum mechanics for math grads

John Kerl

June 18, 2008

Abstract

An introduction to quantum mechanics is presented, at a level appropriate to a general graduate student in mathematics. The emphasis is on finite-dimensional systems, computation, and visualization.

Contents

1 Introduction

Everything should be as simple as possible, and no simpler.
— Albert Einstein (1879-1955).

The perennial question authors pose in their introductions is: Why write another treatment of an old subject? My reasons for this paper are as follows.

I want to explain basic principles of quantum mechanics to myself in a way that I can understand and remember — for general interest as well as for specific background reference for my research in quantum statistical mechanics. In a recent graduate mathematical physics course (just a few months ago!), given my insufficient level of preparation, I found the level of generality too overwhelming for myself as a newcomer. I became lost in all the technicalities which arise (necessarily!) in the infinite-dimensional case: unbounded operators, domain restrictions, singular measures, abstract spectral theory, etc. Working in the greatest possible generality is the prerogative of experience, but it is the bane of instruction.

Quantum mechanics is not a single theory, but a *recipe* — it asserts some basic postulates and tells you how to write down a PDE describing the behavior of a system. The two most common systems to which that recipe is applied at the introductory level are the infinite well and the harmonic oscillator. Yet, these are not the only choices. In this paper, I choose some finite-dimensional systems which are easy to explicitly compute with and visualize. Infinities notoriously introduce computational challenges; finite-dimensional systems (especially two-level systems) provide the opportunity to experiment very concretely and very computably. These things are worth working out and worth understanding for their own sakes — but, moreover, a solid, computational grounding in the finite-dimensional case makes the more general, abstract treatments of the subject digestible.

This paper (and this talk) are aimed at mathematics graduate students — I am one at present. However, I've attempted to keep the level of presentation accessible to an undergraduate who has seen calculus, differential equations, and linear algebra. It is for this reason that I downplay measure theory in my discussion of probability spaces.

I find [**Griffiths**] accessible, and wholeheartedly recommend it as a prequel to a graduate math-physics course. Some of my added value in this paper is that I put as much focus on numerics as on algebraic solutions. The presentation in [**NC**] is similar in spirit to my own, but with different goals. See [**Griffiths**], [**Mermin**], and [**NC**] for more information on quantum mechanics, finite-dimensional and otherwise.

I acknowledge a large debt of gratitude to Itai Seggev, whose presentation [**Seggev**] of the solar-neutrino problem (which can be viewed as a two-level quantum system) inspired many of the computational results of this paper. As well, the theoretical statements here are due mainly to Janek Wehr's mathematical physics course [**Wehr**].

2 The cartoon version

A friend of mine who is a graduate student in another department — someone non-mathematical but well-educated — recently asked me, "What does the *quantum* mean in phrases such as *quantum mechanics*, *quantum computation*, *quantum information*, etc.?" I believe that we, as mathematical experts, should be able to give concise, accurate answers to good layman questions such as this. This section is my current answer to this question; it comes in three parts.

2.1 First part: granularity

When we look around us, we see things that look continuous, e.g. the surface of a pool of water, the top of a table, etc. But when we look with a sufficiently powerful magnifying glass, things become discrete. Electromagnetic energy is radiated from atoms only with certain energy levels; atoms have most of their mass concentrated in their nuclei with mostly empty space all around; the periodic table is indexed by integers; there is no such thing as half an electron. These little packets of things are called *quanta*, which is the plural of *quantum*. This word comes from the same Latin root as *quantity*.
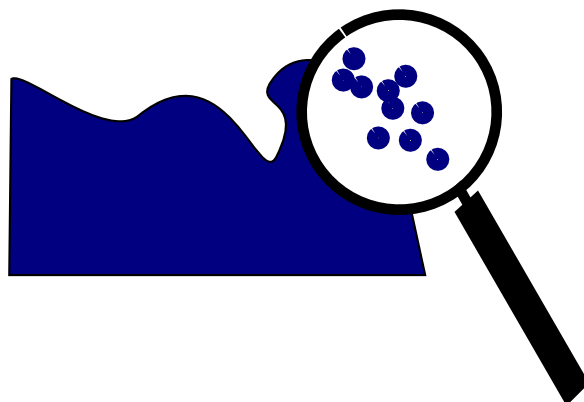
2.2 Second part: indeterminacy

At the same small distance scales where nature reveals its graininess, we also see *indeterminacy*. You can read up (say at Wikipedia) on the double-slit experiment. What this showed is that an electron (this is true for other particles as well) is really not a localized particle — it's sort of a *probability cloud* (or *wave function* as we will discuss below) where the electron, before being observed, has various probabilities of being found in one place or another:
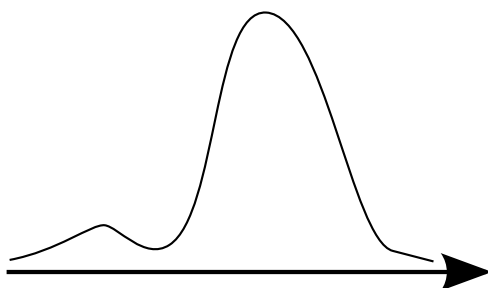
When the particle is observed, it is found to be in one specific place. (This is called the collapse of the wave function, which we will discuss below.)


But if we were to go back in a time machine and repeat our observation (or if we repeat the experiment with many identical electrons), we might find it another place:

(Making a second observation of the *same* particle right after the first observation has a different effect. We'll see why in the next section.)

If we keep repeating that experiment and tabulate all the observed positions and make a histogram of our results, we will start to see the histogram taking the form of the probability distribution which was specified by the wave function:
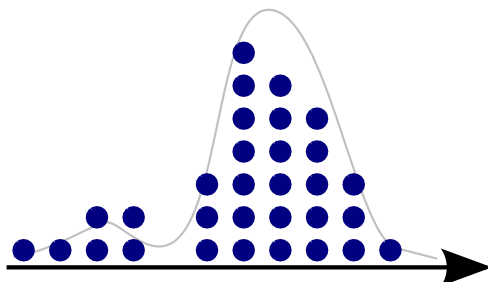
Those s, p orbital figures you saw in high-school chemistry were nothing more than the 95%-confidence-interval boundaries of three-dimensional probability distributions. (Above, I sketched one-dimensional probability distributions.)

This is the connection between probability and statistics:

- Probability predicts the future; it specifies the likelihoods of various events occurring.

- Statistics describes the past; it tabulates the proportions of various events which actually occurred. We use statistical inference to try to discover what probability model underlied the events we observed.

Nature, observed at the macroscopic scale, is deterministic: when we drop a rock, it falls without doubt; planets orbit the sun in very predictable ways. In particular, with the advent of Kepler's laws and Newton's mechanics, the motions of celestial objects could be completely explained without recourse to the supernatural. Many people came to believe that we live in a *clockwork universe*: If one could know exactly the initial conditions of the universe and the laws which govern time evolution of those conditions, one could predict the future completely. This sword is double-edged: on one hand, we would have full knowledge of the universe; on the other hand, there would be no place for free will. This latter is called the *nightmare of determinism.*

Quantum indeterminacy was greeted with some skepticism in the early twentieth century; Albert Einstein famously said, "God does not play dice with the universe." However, indeterminacy was forced upon us by

experiment, and quantum mechanics is a theory which predicts experimental results quite well. This sword is also double-edged: on one hand, nature appears somehow random, but on the other hand, we are relieved of the nightmare of determinism.

2.3   Third part: time evolution

The last piece of quantum theory is that the probability distributions of as-yet-unobserved particles evolve with time:
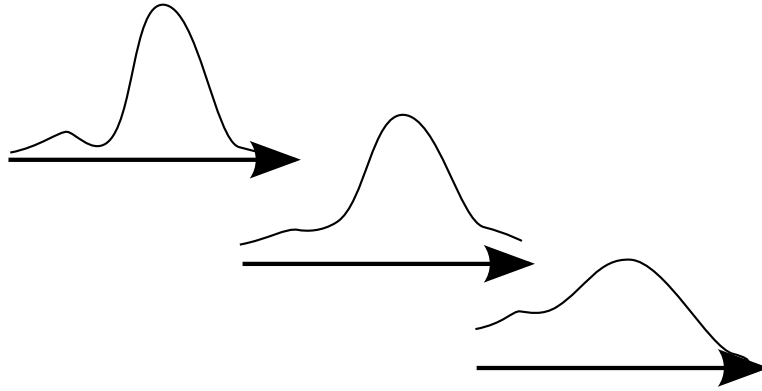
Once a particle is observed, its position is fixed — the probability distribution peaks at the observed location when the wave function collapes — but this distribution continues to evolve with time, typically spreading:
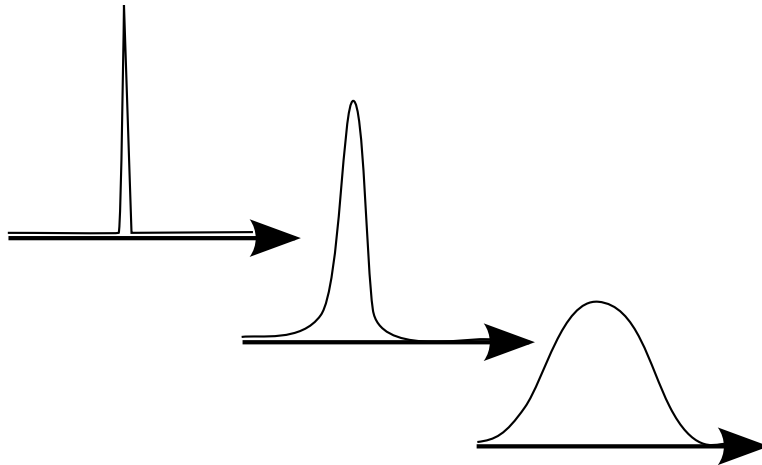
The law governing the time evolution of wave functions is called *Schrödinger's equation.*

That's the qualitative cartoon version. My goal in the remainder of this paper is to show how these notions are made quantitative.

The last question one asks is, *Who cares?* Why study this stuff at all? The best and shortest answer is simple: Without quantum mechanics there would be no Information Age, period.

3   State spaces

In this section we will see how probability spaces (appendix A) and complex vector spaces (appendix B) are combined, in the language of quantum mechanics, into the notion of state spaces.

3.1   The first postulate: wave functions

Consider a single toss of a fair coin. The coin may land heads-up or tails-up on any given toss; we don't know until we carry out the experiment. While the coin is spinning in the air, we say it's neither heads-up nor tails-up, but the fairness of the coin is flying along with it. Only when the coin hits the table do we say that the result of the flip has been determined.

The coin, once it lands, can be in one of two positions, namely, heads or tails, and each occurs half the time — that is, with probability 1/2. We can write this graphically, as in the cartoon section. However, instead of x with all its range of values, there are only two choices on the horizontal axis: H and T.

We can also tabulate these probabilities using an ordered pair of numbers:

$$\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}.$$

(Using the terminology of appendix A, this is a *probability mass function* for the probability space whose two elements are heads and tails.)

Now suppose the coin isn't fair: it lands heads-up with probability 0.64 and tails-up with probability 0.36. We can tabulate the event probabilities as another ordered pair:

$$\begin{pmatrix} 0.64 \\ 0.36 \end{pmatrix}.$$

In general, we have

$$\begin{pmatrix} p \\ 1-p \end{pmatrix}.$$

Physical particles (for example, electrons) likewise can be observed to be in a variety of physical states:

- maybe just one,
- maybe two (e.g. electron spin up or down),
- countably many (e.g. bound-electron energy levels), or
- uncountably infinitely many (e.g. free-electron energy levels).

These situations, however many there may be, are measured with various probabilities. One of the bizarre things about quantum mechanics is that, until the measurement, the particle is like the coin still flipping in mid-air: all we can say about the particle's state is to tabulate the various probabilities of what the state *will* be when it is measured.

That much is perhaps not surprising, but one of the even more bizarre things about quantum mechanics is that the probabilities are *factored* into complex-conjugate pairs. (This fact is not obvious and it was not easily discovered; it is justified only by history and laboratory experiment. See [**Griffiths**] for some history.)

For the coin, we can write down a two-state **wave function**

$$\psi = \begin{pmatrix} a \\ b \end{pmatrix}$$

where a and b are complex numbers. That is,

$$\psi \in \mathbb{C}^2.$$

The probability of the coin landing heads-up (the first state) is

$$P(H) = |a|^2 = a^* a.$$

Likewise, the probability of the coin landing tails-up (the second state) is

$$P(T) = |b|^2 = b^* b.$$

Since it is certain that the coin lands either one way or the other (i.e. with probability 1), we insist on the **normalization**

$$P(H) + P(T) = |a|^2 + |b|^2 = 1.$$

Then we have

$$|a|^2 = p, \quad |b|^2 = 1 - p.$$

For example,

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sqrt{0.5} \\ \sqrt{0.5} \end{pmatrix}$$

is a wave function for the fair coin. Note however that

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sqrt{0.5}i \\ \sqrt{0.5}i \end{pmatrix}$$

also works since

$$(-\sqrt{0.5}i \cdot \sqrt{0.5}i) + (-\sqrt{0.5}i \cdot \sqrt{0.5}i) = 0.5 + 0.5 = 1.$$

Likewise,

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sqrt{0.5}e^{i\alpha} \\ \sqrt{0.5}e^{i\beta} \end{pmatrix}$$

works for real α, β. These $e^{i\alpha}$ and $e^{i\beta}$ are called **phase factors**.

For the unfair coin, we could have wave functions

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.6 \end{pmatrix}, \quad \begin{pmatrix} 0.8i \\ 0.6i \end{pmatrix}, \quad \begin{pmatrix} 0.8e^{i\alpha} \\ 0.6e^{i\beta} \end{pmatrix},$$

etc. In general,

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sqrt{p}e^{i\alpha} \\ \sqrt{1-p}e^{i\beta} \end{pmatrix},$$

is a wave function for the p-weighted coin.

As we will see in section 4, the a and b may change with time. If the complex magnitudes change, then the probabilities change, but if only the complex phases change with time, then the probabilities $|a|^2$ and $|b|^2$ do not change with time.

Remark 3.1. That is to say: *The complex phase factors disappear in the probability computations; they have no observable significance.*

One might ask, why do we need the complex part? It turns out that it is the complex wave function which participates in the differential equation (the Schrödinger equation, section 4.1) which governs the time development of wave functions.

Extensions of this concept:

- If we had not a coin but a six-sided die, then the wave function would be not in $\mathbb{C}^2$ but $\mathbb{C}^6$. The six numbers $|c_1|^2$ through $|c_6|^2$ represent the probability of the die landing in states ϕ_1 through ϕ_6, namely, rolls of 1 through 6 respectively.

- If we were considering countably infinitely many energy levels (e.g. an elecrtron bound to an atom), then the wave function would be infinite-dimensional, and the elements $|c_i|^2$ for $i = 0, 1, 2, 3, \ldots$ would form a discretely probability density function. Here

$$P(\psi = \phi_j) = |c_j|^2,$$

  i.e. the numbers $|c_i|^2$ tabulate the probability of ψ being found in the jth state, ϕ_j, upon measurement. The normalization condition forces ψ to be a **square-summable sequence**, i.e. $\psi \in \ell^2(\mathbb{C})$:

$$\sum_{j=0}^{\infty} P(\psi = \phi_j) = \sum_{j=0}^{\infty} |c_j|^2 = 1.$$

- If we were considering uncountably infinitely many energy levels (e.g. a free electron) then we would need to use the theory of continuous probability distributions. (See [**Kerl**] for more information on this topic.) The normalization criterion forces ψ to be **square integrable**, i.e. $\psi \in L^2(\mathbb{R}^d, \mathbb{C})$:

$$\int_{\mathbb{R}^d} |\psi(\mathbf{x})|^2 d\mathbf{x} = 1.$$

  In fact, this is the historical origin of the study of L^2 spaces.

These examples motivate the following definition.

Definition 3.2. For a finite probability space Ω, with $\#\Omega = N$, the **state space** V is $\mathbb{C}^N$. Each element ψ of V with $\|\psi\| = 1$ is called a **state vector**, and represents a probability distribution (definition A.2) on Ω.

I summarize for emphasis: **Each state vector encodes a probability distribution.**

The **first postulate of quantum mechanics** is (following [**NC**]) simply the following: to any isolated physical system is associated a state space, where the state of the system is described by a unit vector in that space.

The **fourth postulate of quantum mechanics** — a topic which is not pursued here — is that the state of a multi-component system is represented by vectors in the tensor product of the state spaces for each component system.

3.2 Expectations

If we denote the heads-up state as ϕ_0 and the tails-up state as ϕ_1, then we already saw that the wave function

$$\psi = \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sqrt{p}e^{i\alpha} \\ \sqrt{1-p}e^{i\beta} \end{pmatrix}$$

encodes the probability of the coin landing heads-up or tails-up.

We can write down the **mean** (or **expectation** or **expected value**) of various *functions* of the states, much as in section A.3. Suppose you and I are flipping a p-weighted coin with the agreement that if the coin lands heads-up, I win \$2 but if it lands tails-up I lose \$3. Writing 0 for heads and 1 for tails, my winnings are

$$w(0) = 2, \quad w(1) = -3.$$

Then my expected earnings over the long run are

$$\langle\, w(j)\,\rangle = \sum_j w_j P(\psi = \phi_j) = w(0)p + w(1)(1-p) = 2p - 3(1-p) = 5p - 3.$$

(For a fair coin, I'm losing 50 cents per flip on average. As discussed in the appendix, I'd be winning money in the long run as long as this expectation is positive — i.e. if $p > 0.6$.)

Now remember that wave functions factor the probabilities into complex-conjugate pairs. If the result of a coin-flipping experiment depends on the heads-or-tails result, i.e. $f(\psi)$ where ψ is the ordered pair

$$\psi = \begin{pmatrix} \sqrt{p}e^{i\alpha} \\ \sqrt{1-p}e^{i\beta} \end{pmatrix}, = \begin{pmatrix} a \\ b \end{pmatrix},$$

then equation A.1 becomes

$$\langle\, f(\psi)\,\rangle = \sum_j f(\phi_j)P(X = j) = f(\phi_0)|a|^2 + f(\phi_1)|b|^2$$

$$= f(\phi_0)a^*a + f(\phi_1)b^*b$$
$$= a^*f(\phi_0)a + b^*f(\phi_1)b.$$

In general, we put the complex-conjugate part of the probability factor to the left of f and the non-conjugated part to the right:

$$\langle\, f(\psi)\,\rangle = \sum_j c_j^* f(\phi_j)c_j.$$

3.3   The second postulate: observables and measurements

The **second postulate of quantum mechanics**, which again is justified by history and experiment, is as follows:

- An observation corresponds to a Hermitian operator $A : V \to V$ on the state space V.

- Recall from section B.8 that, since A is Hermitian, V has an orthonormal basis $\{\phi_1, \ldots, \phi_N\}$ of eigenvectors of A, with respective eigenvalues $\{a_1, \ldots, a_N\}$. If ψ is the wave function of our system, then it is a linear combination of the eigenbasis:

$$\psi = \sum_{j=1}^N c_j \phi_j$$

  where

$$c_j = \langle\, \phi_j \mid \psi \,\rangle.$$

- Measurement of the quantity to which A corresponds has as its possible outcomes the numbers a_j with probabilities $|c_j|^2$.

If this is to work, we would want the probabilities to be non-negative (which we have by positive-definiteness of the inner product), and we would want them to sum to 1. In fact they do, as shown in the following proposition.

Proposition 3.3. *We have*

$$\sum_j |\langle \phi_j \mid \psi \rangle|^2 = 1.$$

Proof. First, we know we've normalized ψ so that

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

Since the eigenfunctions ϕ_j are an orthonormal spanning set, we have

$$\psi = \sum_j c_j \phi_j$$

with

$$c_j = \langle \phi_j \mid \psi \rangle.$$

Then

$$1 = \langle \psi \mid \psi \rangle = \left\langle \sum_i c_i\,\phi_i \,\Big|\, \sum_j c_j\,\phi_j \right\rangle = \sum_i \sum_j \langle c_i\phi_i \mid c_j\phi_j \rangle$$

$$= \sum_i \sum_j c_i^* c_j \delta_{ij}$$

$$= \sum_j |c_j|^2$$

by the orthonormality of the ϕ_j's. But the c_j's were just shorthand for $\langle \phi_j \mid \psi \rangle$ so we have

$$\sum_j |\langle \phi_j \mid \psi \rangle|^2 = 1$$

which is what we wanted to show. $\qquad\square$

3.4   Expectations, continued

The expected value of A is

$$\langle A \rangle = \sum a_j P(a_j).$$

From the second postulate, those probabilities $P(a_j)$ are

$$P(a_j) = |\langle \phi_j \mid \psi \rangle|^2$$
$$= \langle \psi \mid \phi_j \rangle\langle \phi_j \mid \psi \rangle.$$

Regrouping (since matrix multiplication is associative), we have

$$P(a_j) = \langle \psi \mid \left(\mid \phi_j \rangle\langle \phi_j \mid \right) \mid \psi \rangle.$$

Then

$$\langle A \rangle = \sum \langle \psi \mid \left(a_j \mid \phi_j \rangle\langle \phi_j \mid \right) \mid \psi \rangle.$$

Given the spectral decomposition of A (see section B.8) as

$$A = \sum a_j \mid \phi_j \rangle \langle \phi_j \mid,$$

we have

$$\langle A \rangle = \sum \langle \psi \mid A \mid \psi \rangle.$$

This is the average value or mean of all the observations, averaged over repeated measurements on identical systems.

4 Time evolution

4.1 The Schrödinger equation

As verified by experiment, wave functions evolve according to the **Schrödinger equation**:

$$i\hbar\frac{\partial\psi(t,\mathbf{x})}{\partial t} = \frac{-\hbar^2}{2m}\nabla^2\psi(t,\mathbf{x}) + V(t,\mathbf{x})\psi(t,\mathbf{x}) = H\psi(t,\mathbf{x})$$

i.e.

$$\frac{\partial\psi}{\partial t} = \frac{-iH}{\hbar}\psi$$

where H is the Hermitian operator

$$H = \frac{-\hbar^2}{2m}\nabla^2 + V.$$

The wave function varies in position and time, as in the cartoon version; the **potential function**, conventionally written with the letter V, may also vary in position and time. The potential represents external interactions with the particle. The quantity $\hbar$ is best thought of as being very small.

We can write down a solution

$$\frac{d\psi}{dt} = \frac{-i}{\hbar}\,H\psi \tag{4.1}$$

$$\psi(t) = e^{-iHt/\hbar}\psi(0). \tag{4.2}$$

One checks this solution by differentiating:

$$\frac{d\psi}{dt} = \frac{-iH}{\hbar}\,e^{-iHt/\hbar}\psi(0) = \frac{-iH}{\hbar}\,\psi(t) = \frac{-i}{\hbar}\,H\psi(t).$$

Recall that the Hamiltonian is hermitian. As shown in proposition B.10, the imaginary exponential of a hermitian matrix is **unitary**. The importance of this is that unitary matrices U are **norm-preserving**. That is, if $\|\psi\| = 1$ then $\|U\psi\|$ is still 1; probability functions remain probability functions as they evolve in time.

In fact, [**Wehr**] makes the following grand summary:

> *Quantum mechanics says that the energy operator generates the group of time-evolution operators.*

The group structure is addressed in section 5.2.

4.2 The third postulate: wave-function collapse and subsequent evolution

The **third postulate of quantum mechanics** is that immediately after measurement of an observable A, the state of the system is described *only* by an eigenstate ϕ_k of A (using the notation of section 3.3). This is called the *collapse of the wave function*. The evolution of the single-particle system thereafter is described by the Schrödinger equation, with new initial conditions.

Example 4.1. ▷ Let A be an observable with orthonormal eigenbasis $\{\phi_1,\ldots,\phi_N\}$. We can write our wave function ψ in terms of this basis as

$$\psi = \sum_{j=1}^{N} c_j\phi_j.$$

If $N = 2$ then maybe $c_1 = 0.6$ and $c_2 = 0.8$. Then ψ in this basis is

$$\psi = \begin{pmatrix} 0.6 \\ 0.8 \end{pmatrix}.$$

The probability of the particle being observed in the first or second state is 0.36 or 0.64, respectively. Suppose the former. Then ψ is now

$$\psi = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

After that, as the wave function evolves, probability may leak back into the second slot — or not, as shown in the next section. ◁

4.3 Eigenstates of the Hamiltonian

Suppose a state ψ_k is an eigenstate of the Hamiltonian (not just an observable A, but the same Hamiltonian H which participates in the Schrödiner equation) with eigenvalue E_k. Then

$$H\psi_k = E_k\psi_k.$$

Then equation 4.1 becomes

$$\psi(t) = e^{-iE_k t/\hbar}\psi(0).$$

Notice that the quantity $e^{-iE_k t/\hbar}$ is a complex scalar with magnitude 1. In particular, with a two-dimensional state space,

$$\psi(t) = \begin{pmatrix} a(t) \\ b(t) \end{pmatrix}$$

and so

$$\begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = e^{-iE_k t/\hbar} \begin{pmatrix} a(0) \\ b(0) \end{pmatrix} = \begin{pmatrix} e^{-iE_k t/\hbar}a(0) \\ e^{-iE_k t/\hbar}b(0) \end{pmatrix}.$$

That is, the coefficients $a(t)$ and $b(t)$ evolve in time with only their complex phases being modified; their complex magnitudes do not change in time.

4.4 Inhibition of time evolution

There are two ways to make a system which doesn't evolve in time:

- Construct a system with a zero Hamiltonian, or at least one with a small enough Hamiltonian such that $e^{-iHt/\hbar}$ is sufficiently close to identity for the duration of the experiment, or

- Construct a system in which the initial state is an eigenstate of the Hamiltonian.

5 Numerical results with a two-state system

As described in [**Seggev**], a particular two-state quantum system may be described wherein a neutrino oscillates between one of two flavors. The parameters (which I simplify from Seggev's more physically realistic description) are masses a and b and mixing angle θ. The interpretation of the mixing angle is that if it is zero, no oscillation between neutrino flavors happens; if it is $\pi/4$, then full oscillation occurs.

In this system, the wave function involves t only; there is no notion of position dependence.

The Hamiltonian is (given here without explanation):

$$H = \begin{pmatrix} a^2 \cos^2\theta + b^2 \sin^2\theta & (a^2 - b^2)\cos\theta\sin\theta \\ (a^2 - b^2)\cos\theta\sin\theta & a^2\sin^2\theta + b^2\cos^2\theta \end{pmatrix}.$$

Note that this is real and symmetric, hence Hermitian.

5.1 Algebra

Compute the determinant of $H - \lambda I$. After some algebra, we get eigenvalues

$$\lambda = a^2, b^2$$

with respective eigenvectors

$$\psi = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \begin{pmatrix} \sin\theta \\ -\cos\theta \end{pmatrix}.$$

5.2 Numerics

Time evolution (with $\hbar = 1$):

$$U(t) = e^{-iHt}.$$

Composition:

$$U(s) = e^{-iHs}$$
$$U(s+t) = e^{-iHs}e^{-iHt}$$
$$= e^{-iH(s+t)}.$$

Iterative approximation of $\psi(t)$:

$$\psi(t) = e^{-iHt}\psi(0)$$
$$\psi(t+\Delta t) = e^{-iH(t+\Delta t)}\psi(0)$$
$$= e^{-iH\Delta t}e^{-iHt}\psi(0)$$
$$= e^{-iH\Delta t}\psi(t).$$

Matlab code:

```
a = 0.4; b = 0.5; theta = 0.6 * pi/4;
dt = 0.1; niter = 1000;
```

```
H = [ a^2*cos(theta)^2 + b^2*sin(theta)^2, (a^2-b^2)*cos(theta)*sin(theta);
      (a^2-b^2)*cos(theta)*sin(theta),      a^2*sin(theta)^2 + b^2*cos(theta)^2 ];
Udt = expm(-i * H * dt);
% Other print statements here ...
psit = [1.0; 0.0];
measurement_iter = round(4*niter/5);
for iter = 1:niter
    t     = iter*dt;
    psit  = Udt * psit;

    if (iter == measurement_iter) % Measurement
        u = rand(1)
        if (u < prob(1))
            psit = [1;0];
        else
            psit = [0;1];
        end
    end

    % Other computations / plots ... .
end
```

In black is the probability $|\psi_0|^2$ for the neutrino to be found in state 0, and in blue is the probability $|\psi_1|^2$ for it to be found in state 1. Note that the probabilities always sum to 1 as they vary. After measurement, the wave function collapses onto one of the eigenstates of the observable (not an eigenstate of the Hamiltonian) — i.e. state 0 or state 1 — and then continues to evolve in time.
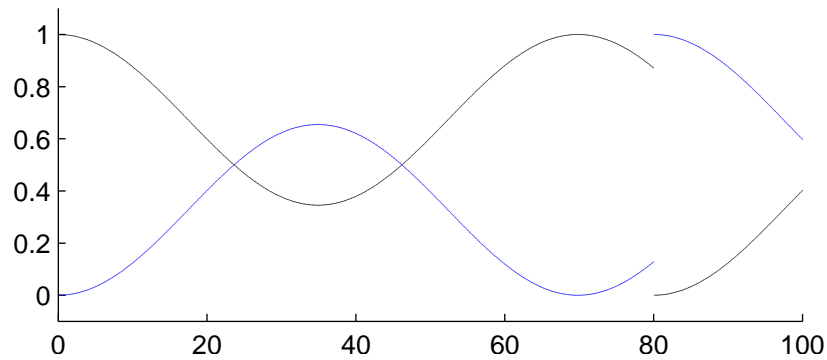
Program output (see code/neutrinos.m):

```
H:
    0.1686   -0.0265
   -0.0265    0.2414

U(dt):
   0.9999 - 0.0169i   0.0001 + 0.0026i
   0.0001 + 0.0026i   0.9997 - 0.0241i

Eigenvalues of H:
```

```
    0.1600
    0.2500


Eigenvectors of H (columns):
   -0.9511    -0.3090
   -0.3090     0.9511


psi(0):
     1
     0


Is H Hermitian?  Here are H* and H:
    0.1686    -0.0265
   -0.0265     0.2414


    0.1686    -0.0265
   -0.0265     0.2414


Is U(dt) unitary?  Here are U(dt)*, U(dt), and U(dt)* U(dt):
   0.9999 + 0.0169i    0.0001 - 0.0026i
   0.0001 - 0.0026i    0.9997 + 0.0241i


   0.9999 - 0.0169i    0.0001 + 0.0026i
   0.0001 + 0.0026i    0.9997 - 0.0241i


   1.0000               -0.0000 + 0.0000i
  -0.0000 - 0.0000i    1.0000
```

6 Numerical results with a discrete 2D system

Here I want to quantify the qualitative picture given in the cartoon version (section 2.1). Namely, let's write down a position-dependent wave function and see how it evolves in time. We can do that numerically (as always in numerics) by sampling the x axis (or x, y axes, etc.) along a discrete mesh — just like a graphing calculator does:
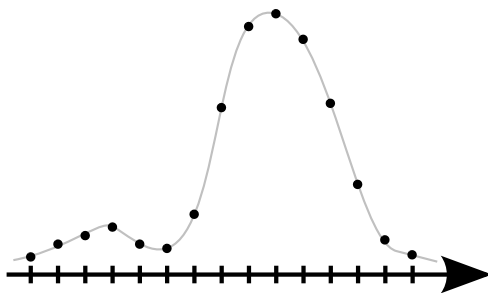
Above, we found a closed-form solution for the Schrödinger equation: given

$$i\hbar \frac{\partial \psi}{\partial t} = \frac{-\hbar^2}{2m} \nabla^2 \psi + V\psi = H\psi,$$

we had

$$\psi(t) = e^{-iHt/\hbar} \psi(0).$$

Here we instead solve the Schrödinger equation directly, using approximation methods.

6.1 The time derivative

Recall from freshman calculus that

$$\frac{\partial \psi}{\partial t} \approx \frac{\psi(t + \Delta t, \mathbf{x}) - \psi(t, \mathbf{x})}{\Delta t}.$$

Given a sequence

$$t_k = t_0 + k\Delta t,$$

we have

$$\frac{\partial \psi}{\partial t} \approx \frac{\psi(t_{k+1}, \mathbf{x}) - \psi(t_k, \mathbf{x})}{\Delta t}.$$

The **Euler method** for solving a single-variable ODE is to use this first-order approximation. Namely, for

$$\frac{dy}{dt} = f(t, y),$$

we use the approximation

$$\frac{y(t_{k+1}) - y(t_k)}{\Delta t} \approx f(t_k, y(t_k))$$

and solve for $y(t_{k+1})$ to obtain

$$y(t_{k+1}) \approx y(t_k) + f(t_k, y(t_k))\Delta t.$$

Here, we start with

$$i\hbar\frac{\partial\psi}{\partial t} = \frac{-\hbar^2}{2m}\nabla^2\psi + V\psi$$
$$\frac{\partial\psi}{\partial t} = i\left(\frac{\hbar}{2m}\nabla^2\psi - \frac{V}{\hbar}\psi\right)$$
$$\frac{\partial\psi}{\partial t} = i\left(c\nabla^2\psi - V\psi\right)$$

where I put

$$c = \frac{\hbar}{2m}$$

and (with abusive notation) set $\tilde{V} = V/\hbar$, then replace $\tilde{V}$ with V. Then the Euler method gives

$$\psi(t_{k+1}, \mathbf{x}) \approx \psi(t_k, \mathbf{x}) + i(c\nabla^2\psi - V\psi)\Delta t.$$

The only question is how to compute the Laplacian on the right-hand side, which I'll do below.

You can read up on the fourth-order **Runge-Kutta method** in any textbook on numerical methods. In short, given

$$\frac{dy}{dt} = f(t, y),$$

the approximation is

$$y(t_{n+1}) = y(t_n) + h/6(k_1 + 2k_2 + 2k_3 + k_4)$$

where

$$k_1 = f(t_n, y_n)$$
$$k_2 = f(t_n + h/2, y_n + k_1 h/2)$$
$$k_3 = f(t_n + h/2, y_n + k_2 h/2)$$
$$k_4 = f(t_n + h, y_n + k_3 h).$$

If this is confusing, just think of it as accomplishing what the Euler method does, but with less error.

6.2   The one-dimensional Laplacian

Also from freshman calculus, the one-dimensional Laplacian is approximately

$$\frac{\partial^2\psi}{\partial x^2} \approx \frac{\psi(t, x + \Delta x) - 2\psi(t, x) + \psi(t, x - \Delta x)}{(\Delta x)^2}.$$

With a mesh of M points $x_0, \ldots, x_{M-1}$, we have

$$\left.\frac{\partial^2\psi}{\partial x^2}\right|_k \approx \frac{\psi(t, x_{k+1}) - 2\psi(t, x_k) + \psi(t, x_{k-1})}{(\Delta x)^2}.$$

Computing the discrete Laplacian at x_k involves sampling ψ there and at two neighbor points:

The question is, what to do for $\psi(t, x_{k-1})$ at the left edge $k = 0$, and for $\psi(t, x_{k+1})$ at the right edge $k = M - 1$.

$$\text{???} \quad x_0 \quad x_1 \quad \cdots \quad\quad\quad\quad\quad\quad\quad\quad x_{M-2} \quad x_{M-1} \quad \text{???}$$

This dilemma is resolved by a choice of boundary conditions:

- Dirichlet-zero, or *clamped*, boundary conditions, taking

$$\psi(t, x_{-1}) = \psi(t, x_M) = 0.$$

- Periodic boundary conditions, taking

$$\psi(t, x_{-1}) = \psi(t, x_{M-1}) \qquad \text{and} \qquad \psi(t, x_M) = \psi(t, x_0).$$

Note that in the 1D case, it is particularly easy to view the discrete Laplacian (and hence the Hamiltonian) as a real symmetric matrix. With Dirichlet-zero boundary conditions we have

$$\Delta\psi = \begin{pmatrix} -2 & 1 & & & & & \\ 1 & -2 & -1 & & & & \\ & 1 & -2 & -1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & 1 & -2 & 1 & \\ & & & & 1 & -2 & 1 \\ & & & & & 1 & -2 \end{pmatrix} \begin{pmatrix} \psi(t, x_0) \\ \psi(t, x_1) \\ \psi(t, x_2) \\ \vdots \\ \psi(t, x_{M-3}) \\ \psi(t, x_{M-2}) \\ \psi(t, x_{M-1}) \end{pmatrix}$$

With periodic boundary conditions, this is

$$\Delta\psi = \begin{pmatrix} -2 & 1 & & & & & 1 \\ 1 & -2 & -1 & & & & \\ & 1 & -2 & -1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & 1 & -2 & 1 & \\ & & & & 1 & -2 & 1 \\ 1 & & & & & 1 & -2 \end{pmatrix} \begin{pmatrix} \psi(t, x_0) \\ \psi(t, x_1) \\ \psi(t, x_2) \\ \vdots \\ \psi(t, x_{M-3}) \\ \psi(t, x_{M-2}) \\ \psi(t, x_{M-1}) \end{pmatrix}$$

The Hamiltonian is written by adding $V(t, x_i)$ along the main diagonal.

Using Matlab, you can pass this matrix to the `eig` routine to find out the eigenvalues and eigenstates of the Hamiltonian.

6.3 The two-dimensional Laplacian

In two dimensions, with an M-point mesh along x and an N-point mesh along y, we have

$$\begin{aligned}
\Delta\psi\Big|_{k,\ell} &= \frac{\partial^2\psi}{\partial x^2}\Big|_{k,\ell} + \frac{\partial^2\psi}{\partial y^2}\Big|_{k,\ell} \\
&\approx \frac{\psi(t, x_{k+1}, y_\ell) - 2\psi(t, x_k, y_\ell) + \psi(t, x_{k-1}, y_\ell)}{(\Delta x)^2} \\
&\quad + \frac{\psi(t, x_k, y_{\ell+1}) - 2\psi(t, x_k, y_\ell) + \psi(t, x_k, y_{\ell-1})}{(\Delta y)^2}.
\end{aligned}$$

Computing the discrete Laplacian at (x_k, y_ℓ) involves sampling ψ there and at four neighbor points:
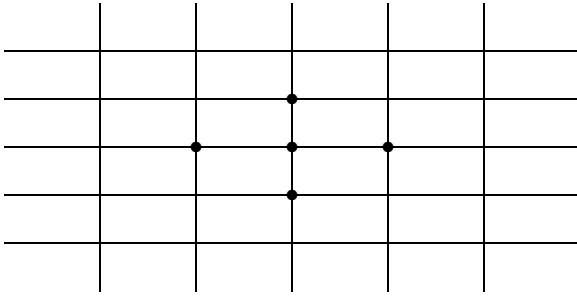
Again, the question of boundary conditions comes up, and some possible solutions are as before: Dirichlet-zero, or periodic (i.e. toric).

6.4 PDE solver

The Matlab code is straightforward, using the `del2` function for the discrete Laplacian with Dirichlet-zero boundary conditions:

```
xlo = -10; xhi =  10; dx  =   0.5;    %% x mesh
ylo = -10; yhi =  10; dy  =   0.5;    %% y mash
[x, y] = meshgrid(xlo:dx:xhi, ylo:dy:yhi);


h   =   0.01; c   =   1.0;
dt  =   0.01; niter = 1000;            %% t mesh

psi0 = exp(-x.^2 + -y.^2);             %% Initial condition
psi0 = psi0 / sqrt(sum(sum(conj(psi0).*psi0))); % Normalize

V = 0.0 * x * y;                       %% Potential
surf(x, y, abs(psi0))                  %% Plot the initial wave function

psie = psi0; psir = psi0; t = 0;

for iter = 1:niter
    % Euler:
    psie = psie + i * dt * (c * del2(psie) - V.*psie);

    % RK4:
    k1 = i * (c * del2(psir)         - V .*  psir);
    k2 = i * (c * del2(psir + k1*h/2) - V .* (psir + k1*h/2));
    k3 = i * (c * del2(psir + k2*h/2) - V .* (psir + k2*h/2));
    k4 = i * (c * del2(psir + k3*h)   - V .* (psir + k3*h));
    psir = psir + dt/6*(k1 + 2*k2 + 2*k3 + k4);

    psie = psie / sqrt(sum(sum(conj(psie).*psie))); % Renormalize for
    psir = psir / sqrt(sum(sum(conj(psir).*psir))); %   round-off error
```
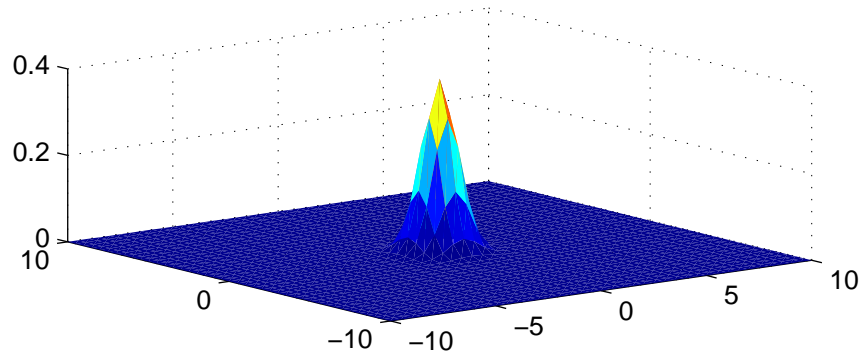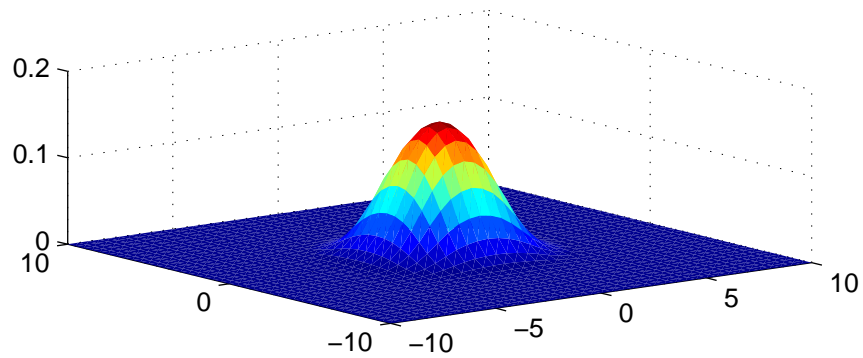
```
    t = t + dt;
end
surf(x, y, abs(psir))                    %% Plot the final wave function
```

Here's the initial wave function:

Here's the final wave function (it looks about the same using Euler or Runge-Kutta):

Much more could be done with this example, e.g. varying the potential. Nonetheless, my main point is that the second-derivative term tends to spread out peaks over time, leading to delocalization of the wave function after measurement.

7   The Heisenberg uncertainty principle

The following is not central to the talk (and in fact, was not presented in the talk). However, I feel that the Heisenberg uncertainty principle is a result that many of us have heard of but which is often not precisely understood. There is a beautiful discussion in [**NC**], which I elaborate on and modify here.

There is some background which will not be proved here. First, take it as given that many observables come in **conjugate pairs** — technically, this means that one is the Fourier transform of the other. These include, among others, position and momentum.

Second, Heisenberg showed that such pairs of observables (call them, say, A and B) satisfy the **canonical commutation relation**

$$[A, B] := AB - BA = i\hbar.$$

Third, as [**NC**] point out, there is a common misconception that measurement of position disturbs a particle, making measurement of its momentum uncertain. Disturbance can occur during measurement, but this is not the content of the uncertainty principle. Rather, it quantifies the products of the *standard deviations* of the measurements of A and B — where the standard deviation of a measurement describes its variability over measurements made in a large number of experiments.

Fourth, recall from probability that the standard deviation of a quantity X is

$$\Delta X := \sqrt{\langle\, (X - \langle\, X\,\rangle)^2\,\rangle} = \sqrt{\langle\, X^2\,\rangle - \langle\, X\,\rangle^2}.$$

If it so happens that X has mean zero, then this reduces to

$$\Delta X = \sqrt{\langle\, X^2\,\rangle}.$$

Fifth, recall from section 3.4 that the expected value of X is

$$\langle\, X\,\rangle = \langle\, \psi \mid X \mid \psi\,\rangle.$$

Proposition 7.1 (Heisenberg uncertainty principle). *If A and B are a conjugate pair of observables, then*

$$\Delta A \Delta B \geq \frac{\hbar}{2}.$$

Proof. Suppose the system is in the state ψ. To simplify the algebra a bit, let

$$C = A - \langle\, A\,\rangle \qquad \text{and} \qquad D = B - \langle\, B\,\rangle.$$

These have mean zero. Since A and A are observables, they are Hermitian; note that C and D are also Hermitian. Also recall that shifting by a constant doesn't affect the standard deviation, so

$$\Delta C = \Delta A \qquad \text{and} \qquad \Delta D = \Delta B.$$

The expected value of the commutator is

$$\langle\, \psi \mid [C, D] \mid \psi\,\rangle = \langle\, \psi \mid CD \mid \psi\,\rangle - \langle\, \psi \mid DC \mid \psi\,\rangle.$$

By the triangle inequality,

$$|\langle\, \psi \mid [C, D] \mid \psi\,\rangle| \leq |\langle\, \psi \mid CD \mid \psi\,\rangle| + |\langle\, \psi \mid DC \mid \psi\,\rangle|.$$

The Cauchy-Schwartz inequality gives us

$$|\langle\,\psi\mid CD\mid\psi\,\rangle| \leq \sqrt{\langle\,\psi\mid C^2\mid\psi\,\rangle\langle\,\psi\mid D^2\mid\psi\,\rangle} = \sqrt{\langle\,C^2\,\rangle\langle\,D^2\,\rangle} = \Delta C \Delta D$$

and likewise

$$|\langle\,\psi\mid DC\mid\psi\,\rangle| \leq \Delta C \Delta D.$$

Then

$$\Delta C \Delta D \geq \frac{|\langle\,\psi\mid [C,D]\mid\psi\,\rangle|}{2}$$

but

$$\langle\,\psi\mid [C,D]\mid\psi\,\rangle = i\hbar\langle\,\psi\mid\psi\,\rangle = i\hbar$$

so

$$\Delta C \Delta D \geq \frac{\hbar}{2}.$$

Since $\Delta A = \Delta C$ and $\Delta B = \Delta D$, we have

$$\Delta A \Delta B \geq \frac{\hbar}{2}.$$

□

A Probability

Probability theory is sketched here at a bare-minimum level to support the concepts in this paper. Footnotes in this section are intended for the reader with a knowledge of probability theory; other readers may ignore them. See [**Griffiths**], [**Kerl**], [**GS**], and [**FG**], respectively, for treatments of probability with increasing levels of rigor.

A.1 Probability spaces

For purposes of this paper, I care only about finite or countably infinite probability spaces.

Definition A.1. A **probability space** is a finite or countably infinite[1] set Ω, called a **sample space**, along with[2] a probability measure which is a function $P : 2^\Omega \to [0, 1]$ which[3] assigns to each $\omega \in \Omega$ a number $P(\omega) \in [0, 1]$ such that
$$\sum_{\omega \in \Omega} P(\omega) = 1.$$
We extend P to a function on finite subsets of Ω by
$$P(\{\omega_1, \ldots, \omega_m\}) = P(\omega_1) + \ldots + P(\omega_m)$$
as long as the ω_i's are distinct. (With the convention that an empty sum is zero, we have $P(\emptyset) = 0$.) Some standard probability terminology is that an $\omega \in \Omega$ is called an **outcome**; a subset of Ω is called an **event**.

Definition A.2. The mapping $\omega \to P(\omega)$ is called the **probability mass function** or **p.m.f.** or **probability distribution** for the probability space (Ω, P).

Example A.3. ▷ Consider the toss of a coin with probability p of landing heads-up. Then
$$\Omega = \{H, T\}$$
and
$$P(H) = p, \quad P(T) = 1 - p.$$
The p.m.f. is

ω	$P(\omega)$
H	p
T	$1 - p$

◁

Example A.4. ▷ If there are n such coins, then $\#\Omega = 2^n$ and P of a particular sequence $(\varepsilon_1, \ldots, \varepsilon_n)$ is
$$\prod_{i=1}^{n} p^{1-\varepsilon_i}(1-p)^{\varepsilon_i}$$
where $\varepsilon = 0$ for heads and $\varepsilon = 1$ for tails. With $n = 2$, the p.m.f. is

ω	$P(\omega)$
(H, H)	p^2
(H, T)	$p(1-p)$
(T, H)	$p(1-p)$
(T, T)	$(1-p)^2$

◁

---

[1]Uncountably infinite probability spaces can be considered; I avoid them here.
[2]The σ-algebra is the full power set $\mathcal{F} = 2^\Omega$ and is implicit in this discussion.
[3]Values of P on singletons suffice to uniquely define the probability measure on all of $\mathcal{F}$.

Example A.5. ▷ Another example is the toss of a fair six-sided die. The sample space is

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

with

$$P(1) = P(2) = \ldots = P(6) = 1/6.$$

Some example events include $\emptyset$, $\{2\}$, and $\{1, 3, 5\}$. We have $P(\emptyset) = 0$, $P(\{2\}) = 1/6$, and

$$P(\{1, 3, 5\}) = P(\{1\}) + P(\{3\}) = P(\{5\}) = 1/6 + 1/6 + 1/6 = 1/2.$$

◁

A.2   Random variables

Definition A.6. A **random variable** X is a function[4] from the probability space (Ω, P) to $\mathbb{R}^d$ or $\mathbb{C}^d$ — in particular, often $\mathbb{R}$ or $\mathbb{C}$.

Example A.7. ▷ For the coin-flip example, we might take

$$X(H) = +2 \qquad \text{and} \qquad X(T) = -3.$$

This represent my winnings in a game where we flip a coin with the rule that you give me \$2 for heads and I give you \$3 for tails. ◁

A.3   Expectations

Definition A.8. The **expectation** or **expected value** or **mean** of a random variable X is

$$E[X] = \langle X \rangle = \sum_{\omega \in \Omega} P(\omega) X(\omega).$$

(For countably infinite Ω, we say the expectation of X is undefined if the sum does not converge.) We usually omit mention of ω's and instead write

$$E[X] = \langle X \rangle = \sum_{x} x P(X = x)$$

where the sum is over all x in the range $X(\Omega)$.

Example A.9. ▷ Continuing example A.7, the range $X(\Omega) = \{2, -3\}$ and my expected winnings over many such games are

$$\langle X \rangle = X(H)P(H) + X(T)P(T) = 2p - 3(1 - p) = 5p - 3.$$

(Then I'll lose money at this in the long run unless the coin has probability of heads $p > 0.6$.) ◁

Proposition A.10. *More generally, for a function $f(X)$ (e.g. $f : \mathbb{C} \to \mathbb{C}$),*

$$\langle f(X) \rangle = \sum_{x} f(x) P(X = x). \tag{A.1}$$

This is called the **Law of the Unconscious Statistician**. See [**Kerl**] for more information.

---

[4]Technically, a measurable function from the probability space$(\Omega, \mathcal{F}, P)$ to a measurable space $(\Omega', \mathcal{F}')$.

B Linear algebra

B.1 Convention for complex inner products

Recall that in complex vector spaces V we must conjugate one argument in order to get positive definiteness of the inner product. (The complex inner product is said to be **sesquilinear** rather than **bilinear**.) That is, for $c \in \mathbb{C}$ and $\xi, \psi \in V$, we have either the **physicists' convention**

$$\langle c\xi, \psi \rangle = c^* \langle \xi, \psi \rangle \qquad \text{and} \qquad \langle \xi, c\psi \rangle = c \langle \xi, \psi \rangle$$

or the **mathematicians' convention**

$$\langle c\xi, \psi \rangle = c \langle \xi, \psi \rangle \qquad \text{and} \qquad \langle \xi, c\psi \rangle = c^* \langle \xi, \psi \rangle.$$

For this paper we use the physicists' convention.

Definition B.1. Recall that the **norm** of a vector is

$$\|\psi\| = \langle \psi, \psi \rangle^{1/2}$$

which is to say

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

What, tangibly, is an inner product? It can take various forms, as long as it satisfies the inner-product axioms (positive-definiteness, conjugate-symmetry, and sesquilinearity). Two important examples are:

- For an L^2 space $\mathcal{H}$ with measure $d\mu$ we write

$$\langle f, g \rangle = \int_{\mathcal{H}} \overline{f}\, g \, d\mu.$$

  In particular, for $\mathcal{H} = [0, L]$ with periodic boundary conditions and Lebesgue measure,

$$\langle f, g \rangle = \frac{1}{L} \int_0^L \overline{f}\, g \, dx.$$

  The $1/L$ scale factor is not required to make an inner product (any non-zero scale factor would do). With the $1/L$, though, the functions $\{e^{i2\pi kx/L}\}_{k \in \mathbb{Z}}$ form an orthonormal set.

- For $\mathbb{C}^N$, with

$$\mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix} \qquad \text{and} \qquad \mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix},$$

  we use

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{j=1}^{N} \overline{u_j} v_j.$$

B.2 Matrix notation

One may think of a **column vector** and a **row vector** in $\mathbb{C}^N$ as an $N \times 1$ matrix and a $1 \times N$ matrix, respectively:

$$\psi = \begin{pmatrix} c_1 \\ \vdots \\ c_N \end{pmatrix} \qquad \text{and} \qquad \psi^* = \begin{pmatrix} c_1 & \cdots & c_N \end{pmatrix}$$

In fact, this point of view makes Dirac notation (below) clear and intuitive.

B.3   Bra-ket notation

Notation due to Paul Dirac, and firmly entrenched in the physics community, is as follows. Let $\{\phi_1, \ldots, \phi_N\}$ be a basis for an N-dimensional complex vector space V. Then vectors ψ are sometimes written

$$| \, \psi \, \rangle.$$

The basis vectors

$$\phi_1, \phi_2, \ldots, \phi_N$$

are sometimes written simply as the **ket**

$$| \, 1 \, \rangle, | \, 2 \, \rangle, \ldots, | \, N \, \rangle.$$

For example, in $\mathbb{C}^2$ the standard basis is

$$| \, 1 \, \rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad | \, 2 \, \rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Of course, if the basis vectors are numbered

$$| \, 0 \, \rangle, | \, 1 \, \rangle, \ldots, | \, N - 1 \, \rangle.$$

then we would instead have (with $N = 2$)

$$| \, 0 \, \rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad | \, 1 \, \rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Since the dot product $\langle \xi, \psi \rangle$ may be written written with respect to an orthonormal basis as the matrix product

$$\xi^* \psi,$$

e.g.

$$\begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix} = \begin{pmatrix} a_1^* & \cdots & a_N^* \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix} = \sum_{j=1}^{N} a_j^* b_j,$$

we write the conjugate transpose row vector, or **bra**, as

$$\langle \, \psi \, | \quad = \quad | \, \psi \, \rangle^*.$$

Then the inner product

$$\langle \xi, \psi \rangle$$

is written by putting the bra and ket together as a **bra-ket**, or bracket:

$$\langle \, \xi \, | \, \psi \, \rangle$$

Now let's put a matrix in the picture. Let $M \in \mathbb{C}^{N,N}$ be

$$M = \begin{pmatrix} m_{11} & \cdots & m_{1N} \\ \vdots & & \vdots \\ m_{N1} & \cdots & m_{NN} \end{pmatrix}.$$

Then $M\psi$ is another vector, with entries

$$(M\psi)_i = \sum_{j=1}^{N} m_{ij} b_j,$$

so

$$\langle\, \xi \mid M\psi \,\rangle = \sum_{i=1}^{N} \overline{a_i}(M\psi)_i = \sum_{i,j=1}^{N} m_{ij}\overline{a_i}b_j.$$

Likewise,

$$(M\xi)_j = \sum_{i=1}^{N} m_{ji} a_i,$$

so

$$\langle\, M\xi \mid \psi \,\rangle = \sum_{j=1}^{N} \overline{(M\xi)_j}b_j = \sum_{i,j=1}^{N} \overline{m_{ji}}\,\overline{a_i}\,b_j.$$

In the particular case $M = M^*$, i.e. $m_{ij} = \overline{m_{ji}}$, then

$$\langle\, \xi \mid M\psi \,\rangle = \langle\, M\xi \mid \psi \,\rangle$$

and we write that complex number as

$$\langle\, \xi \mid M \mid \psi \,\rangle.$$

B.4   Outer products

Definition B.2. If

$$\xi = \sum_{i=1}^{N} a_i \phi_i \qquad \text{and} \qquad \psi = \sum_{j=1}^{N} b_j \phi_j$$

are elements[5] of $\mathbb{C}^N$, written as a linear combination of a basis $\{\phi_1, \ldots, \phi_N\}$, then the **outer product** of ξ and ψ with itself, written

$$\mid \xi \,\rangle\langle\, \psi \mid,$$

is the $N \times N$ matrix with elements

$$a_i b_j^*.$$

Example B.3. ▷ If

$$\psi = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

then

$$\mid \psi \,\rangle\langle\, \psi \mid = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

◁

---

[5]One can do this for $\mathbb{C}^M$ and $\mathbb{C}^N$, although we won't need that for this paper.

The intuition is that an outer product is a *deferred inner product operator*. For example, let $\xi, \phi, \psi \in \mathbb{C}^2$. Then we may write the vector $| \xi \rangle \langle \phi | \psi \rangle$ as the product of three matrices. By associativity of matrix multiplication, we can do the product $\langle \phi | \psi \rangle$ first:

$$| \xi \rangle [\langle \phi | \psi \rangle] = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_N \end{pmatrix} \left[\begin{pmatrix} \phi_1^* & \cdots & \phi_N^* \end{pmatrix} \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_N \end{pmatrix} \right].$$

On the other hand, we could do the product $| \xi \rangle \langle \phi |$ first:

$$[| \xi \rangle \langle \phi |] | \psi \rangle = \left[\begin{pmatrix} \xi_1 \\ \vdots \\ \xi_N \end{pmatrix} \begin{pmatrix} \phi_1^* & \cdots & \phi_N^* \end{pmatrix} \right] \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_N \end{pmatrix}.$$

This is why we conjugate the second operand's coefficients in the definition of the outer product.

B.5 Self-adjoint operators

Let A be an operator, i.e. linear transformation, from a complex vector space V to itself[6]. There are two ways to think of adjoint and self-adjoint operators: the first is coordinate-free and the second is coordinate-dependent.

B.5.1 Coordinate-free definition

The **adjoint** of A is the unique operator A^* such that for all $\xi, \psi \in V$,

$$\langle \xi, A\psi \rangle = \langle A^* \xi, \psi \rangle.$$

(It can be shown that the adjoint exists and is unique; we'll construct it below.) The operator A is called **self-adjoint** if $A = A^*$, i.e. for all $\xi, \psi \in V$,

$$\langle \xi, A\psi \rangle = \langle A\xi, \psi \rangle.$$

Using Dirac's bra-ket notation, we can write either

$$\langle \xi, A\psi \rangle \qquad \text{or} \qquad \langle A\xi, \psi \rangle$$

when A is self-adjoint. Since it doesn't matter, the convention is to put the operator in the middle:

$$\langle \xi | A | \psi \rangle.$$

B.5.2 Coordinate-dependent construction

The rule of thumb for computing adjoints is, as I heard at an AMS/MAA talk recently, "The adjoint of A is whatever you have to do to A to get it onto the other side of the inner product." Let's do that. I spent some scratch paper playing around with summation symbols — but then I realized it's easy as long as we think of row vectors, matrices, and column vectors all as matrices. That is,

$$\langle \xi, A\psi \rangle = \xi^* A\psi.$$

---

[6]In fact, the adjoint is defined for $A : V \to W$ to be an operator $A^* : W \to V$, but we won't need that here.

For example, with $N = 2$, we have

$$\begin{pmatrix} \xi_1^* & \xi_2^* \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}.$$

We know that for matrices B and C, $(BC)^* = C^*B^*$. So here we have

$$\langle \xi, A\psi \rangle = \xi^* A\psi = (A^*\xi)^* \psi = \langle A^*\xi, \psi \rangle.$$

That is, the adjoint of the matrix A is just its **conjugate transpose**. For example,

$$\begin{pmatrix} 1 & 2i \\ 3i & 4 \end{pmatrix}^* = \begin{pmatrix} 1 & -3i \\ -2i & 4 \end{pmatrix}.$$

Then, when you are looking at coordinates, you can tell A is self-adjoint if it is **Hermitian**, i.e. equal to its own conjugate transpose. For example,

$$\begin{pmatrix} 1 & 2i \\ -2i & 3 \end{pmatrix}^* = \begin{pmatrix} 1 & 2i \\ -2i & 3 \end{pmatrix}$$

so this matrix is self-adjoint.

We will see in section B.8 that a self-adjoint operator has an orthonormal eigenbasis.

B.6   Unitary operators

Definition B.4. An operator $A : V \to V$ is said to be **unitary** if it is norm-preserving, i.e. for all $\psi \in V$,

$$\|A\psi\| = \|\psi\|.$$

In terms of definition B.1, this means that for all $\psi \in V$,

$$\langle A\psi, A\psi \rangle = \langle \psi, \psi \rangle.$$

Remark B.5. An alternate characterization of unitarity is

$$A^*A = I,$$

which is to say

$$A^{-1} = A^*.$$

B.7   Matrix exponentials; Hermitian and unitary matrices

Definition B.6. The **exponential** of $A : V \to V$ is defined to be

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

where A^0 is the identity matrix.

It can be shown that for finite-dimensional V, this sum converges for all A.

Remark B.7. Note in particular that if A is diagonal, with entries

$$A = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_N \end{pmatrix},$$

then e^A is also diagonal with

$$e^A = \begin{pmatrix} e^{d_1} & & 0 \\ & \ddots & \\ 0 & & e^{d_N} \end{pmatrix}.$$

In particular,

$$e^0 = I.$$

Proposition B.8. *Exponential commutes with conjugate transpose. That is,*

$$(e^A)^* = e^{A^*}.$$

Proof. Using formal manipulation of symbols, this can be justified by moving the conjugate transpose through the power series term by term. □

Proposition B.9. *For $A : V \to V$,*
$$e^{-A} = (e^A)^{-1}.$$

Proof. This can also be proved using the power-series expansion. □

Proposition B.10. *If A is Hermitian, then e^{iA} is unitary.*

Proof. Suppose A is Hermitian. Again using formal manipulation of symbols, and using the characterization in remark B.5 as well as proposition B.9,

$$(e^{iA})^* = e^{(iA)^*} = e^{-iA^*} = e^{-iA} = (e^{iA})^{-1}.$$

□

B.8   The finite-dimensional spectral theorem

Theorem B.11 (The finite-dimensional spectral theorem)**.** *Let V be a finite-dimensional complex vector space, and let $A : V \to V$ be a Hermitian operator. Then*

(i) *The eigenvalues of A are all real, and*

(ii) *the eigenvectors of A form an orthonormal basis for V.*

Proof. (Proofs follow [**Wik**] and appendix of [**Griffiths**].)

Part (i). Suppose $A\phi = \lambda\phi$. Then

$$\langle \phi, A\phi \rangle = \langle \phi, \lambda\phi \rangle = \lambda \langle \phi, \phi \rangle.$$

33

But since A is Hermitian, we also have

$$\langle\,\phi,\,A\phi\,\rangle=\langle\,A\phi,\,\phi\,\rangle=\langle\,\lambda\phi,\,\phi\,\rangle=\lambda^*\langle\,\phi,\,\phi\,\rangle.$$

Eigenvectors are never zero, and only the zero vector has norm zero, so in

$$\lambda\langle\,\phi,\,\phi\,\rangle=\lambda^*\langle\,\phi,\,\phi\,\rangle$$

we can divide by $\langle\,\phi,\,\phi\,\rangle$ to conclude

$$\lambda=\lambda^*,$$

which is to say that λ is real.

Part (ii). This has two subparts: that the eigenvectors of A span V, and that they are orthogonal. (Normality can be obtained by scaling the eigenvectors to have norm 1.) Both subparts are sketched at once in [**Wik**], but a little abstractly. I'll show this (completing the sketch) but I'll also include Griffith's very concrete proof of orthogonality.

Taking as given that any operator on a complex vector space has at least one eigenvector (this follows from the fundamental theorem of algebra applied to the characteristic polynomial $\det(A-\lambda I)$), let ϕ_1 be an eigenvector of A with eigenvalue λ_1. Let V_1 be the eigenspace of ϕ_1, i.e. the span of $\{\phi_1\}$, and consider the perpendicular space (or orthogonal complement)

$$V_1^\perp=\{\psi\in V:\langle\,\phi_1,\,\psi\,\rangle=0\}.$$

We claim that $V_1^\perp$ is an invariant subspace of A. To prove this claim, we need to show that if $\psi\in V_1^\perp$, then $A\psi\in V_1^\perp$. Let $\psi\in V_1^\perp$, i.e. $\langle\,\phi_1,\,\psi\,\rangle=0$. Then

$$\langle\,\phi_1,\,A\psi\,\rangle=\langle\,A\phi_1,\,\psi\,\rangle=\lambda_1\langle\,\phi_1,\,\psi\,\rangle=0.$$

Now inductively apply this argument to $V_1^\perp$.

Here's Griffith's bit on orthogonality: Suppose we have two eigenvectors ϕ and ψ, with distinct eigenvalues $\lambda\neq\mu$, respectively. Then

$$\langle\,\phi,\,A\psi\,\rangle=\langle\,\phi,\,\mu\psi\,\rangle=\mu\langle\,\phi,\,\psi\,\rangle.$$

But since A is Hermitian, and recalling that λ and μ are both real by part (i), we also have

$$\langle\,\phi,\,A\psi\,\rangle=\langle\,A\phi,\,\psi\,\rangle=\langle\,\lambda\phi,\,\psi\,\rangle=\lambda\langle\,\phi,\,\psi\,\rangle.$$

These two things are equal so we have

$$\lambda\langle\,\phi,\,\psi\,\rangle=\mu\langle\,\phi,\,\psi\,\rangle$$
$$(\lambda-\mu)\langle\,\phi,\,\psi\,\rangle=0.$$

Since $\lambda\neq\mu$, the first factor is non-zero, and so the second factor must be zero. This means that

$$\phi\perp\psi.$$

Note that this applies only for $\lambda\neq\mu$. In the degenerate case, i.e. when an eigenvalue appears with multiplicity, we need to use Gram-Schmidt to obtain an orthonormal basis for that eigenspace. $\square$

References

Quantum mechanics:

[**Griffiths**] Griffiths, D.J. *Introduction to Quantum Mechanics* (2nd ed.). Pearson Prentice Hall, 2005.

[**Mermin**] Mermin, X. *Lecture Notes on Quantum Computation* (nth ed.).
`http://people.ccmr.cornell.edu/~mermin/qcomp/CS483.html`.

[**NC**] Nielsen, M.A. and Chuang, I.L. *Quantum Computation and Quantum Information.* Cambridge, 2001.

[**Seggev**] Seggev, I. *More Quantum Mechanics and Linear Algebra: Neutrinos and Other Subatomic Particles.* Presentation at the AMS/MAA Joint Meetings, San Diego, 2008.

[**Wehr**] Wehr, J. Math 541. Mathematical Physics course at University of Arizona, fall 2007.

[**Wik**] Articles on *Density operator*, *Matrix exponential*, and *Partial trace*. http://en.wikipedia.org.

Probability:

[**FG**] Fristedt, B. and Gray, L. *A Modern Approach to Probability Theory.* Birkhauser, 1997.

[**GS**] Grimmett, G. and Stirzaker, D. *Probability and Random Processes*, 3rd ed. Oxford, 2001.

[**Kerl**] Kerl, J. *Probability notes.* `http://math.arizona.edu/~kerl/doc/prb.pdf`

Hilbert spaces:

[**AG**] Akhiezer, N.I. and Glazman, I.M. *Theory of linear operators in Hilbert Space.* Dover, 1993.

[**BS**] Berezin, F.A. and Shubin M.A.. *The Schrödinger Equation.* Springer, 1991.

[**Lax**] P. Lax. *Functional Analysis.* Wiley Interscience, 2002.

[**RS**] Reed, M. and Simon, B. *Methods of Modern Mathematical Physics* (rev. ed.). Academic Press, 1980.

Index