

1 Complex Numbers in Quantum Mechanics

Complex numbers and variables can be useful in classical physics. However, they are not essential. To emphasize this, recall that forces, positions, momenta, potentials, electric and magnetic fields are all real quantities, and the equations describing them, Newton's laws, Maxwell's equations, etc. are all differential equations involving strictly real quantities.

Quantum mechanics is different. Factors of $i = \sqrt{-1}$ are everywhere, for example from Heisenberg we have his famous commutation relation,

$$QP - PQ = i\hbar,$$

and from Schrödinger his equally famous equation

$$\frac{\hbar}{i}\partial_t\Psi = H\Psi.$$

The wave function Ψ is complex and so is practically every other quantity needed to formulate quantum mechanics. Granted that classical mechanics fails at distances of nanometer or less, why is it suddenly so necessary that everything be complex?

Comment 1 An interesting question is the distance scale below which a quantum treatment is essential. A nanometer seems like a reasonable choice. But even more intriguing is the question of what happens on much bigger scales, centimeters, meters, and larger. Does quantum mechanics still govern things, or does it somehow break down? In 2011, it is fair to say no definite answer to this question is known.

Returning to the question of why quantum mechanics requires complex quantities, one reason becomes apparent upon trying to describe wavelike phenomena for material particles, electrons, neutrons, etc. To see this point, start with a classical electromagnetic wave. For a plane wave traveling along the z -axis, we can write the following formula for the electric field,

$$\vec{E} = E_0\hat{x}\cos(kz - \omega t). \tag{1}$$

To be more realistic, we should of course superpose nearby frequencies so that the wave has a finite extent. However, the observation we will make applies equally well whether we have a single frequency or a superposition of a narrow range of frequencies. For simplicity, we will use the simple formula Eq.(1). We may break Eq.(1) into pure exponentials, so-called “positive frequency” and “negative frequency” parts;

$$\vec{E} = E_0\hat{x}(e^{-i\omega t+ikx} + e^{i\omega t-ikx}).$$

(By long-established convention $\exp(-i\omega t)$ is taken to be the positive frequency part ($\omega > 0$ here), and $\exp(+i\omega t)$ is the negative frequency part.) The important point is that

in classical physics, since everything is ultimately real, positive and negative frequency parts have to be present in equal measure.

Now turn to material (or massive) particles. They show wave phenomena (interference, diffraction, etc) with wavelengths determined by De Broglie,

$$\lambda = \frac{h}{p} \text{ or } k = \frac{p}{\hbar}.$$

Diffraction phenomena were first demonstrated for electrons by Davisson and Germer (Phys Rev **30**, 705 (1927)), working at the Bell Telephone Lab in NY,NY.

The Davisson-Germer experiment and many other subsequent experiments establish that massive particles with definite momenta have wave-like properties. Given that, one wishes to write down a formula describing traveling waves for massive particles. A first thought would be a formula analogous to Eq.(1). But this cannot work for the following reason. A single material particle traveling along certainly has a **positive energy**. This is true whether we use the full Einstein formula $E = \sqrt{(pc)^2 + (m_0c^2)^2}$ or the non-relativistic kinetic energy, $E_{nr} = p^2/2m$. Furthermore, energy and frequency are related by the Einstein-Planck formula

$$E = \hbar\omega, \tag{2}$$

so given $E > 0$, we must have $\omega > 0$. (The photoelectric effect establishes the Einstein-Planck formula without a doubt.) The conclusion is that a “wave-function” for a single material particle must have only a **positive frequency part**. So Ψ , whatever it represents, must vary as follows

$$\Psi \sim e^{-i\omega t + ikz}.$$

The fact that $\exp(ikz)$ goes with $\exp(-i\omega t)$ is saying the particle is traveling along the +z-axis. If it were traveling in the negative z-direction, we would have $\exp(-ikz)$, but regardless of the direction of travel, the time-dependence must be of the form $\exp(-i\omega t)$ with $\omega = E/\hbar$ positive. So to allow for wavelike phenomena and at the same time allow only positive frequencies as required by positive energy for a free particles, complex functions must enter quantum mechanics.

photons and harmonic oscillators The agreement just given leads to the introduction of a still-mysterious complex “wave-function” for massive particles. But the Einstein-Planck formula Eq.(2) was first written down for photons, and surely applies to them as well. A classical electromagnetic wave whose electric field could be given by Eq.(1) is quantum-mechanically a collection of many photons. Each of them has positive energy, given by Eq.(2), yet the formula for the electric field contains both positive and negative frequencies. This needs some explanation. The first thing to note (and recall from undergrad QM) is that any classical quantity such as a position, a momentum, or an electromagnetic field component, is represented in quantum mechanics as an expected value of an operator. This applies to the electric field of our plane wave. It is composed of photons with energies given by the Einstein-Planck formula Eq.(2), and will show

classical behavior if many photons are present. The numerical value of the electric field will be the expected value of an electric field operator. You will learn how to handle this operator later in the course, but in the end, the electromagnetic field is nothing but a bunch of harmonic oscillators. So we can study the point about positive and negative frequencies by examining a simple harmonic oscillator. Classically, we can write

$$x(t) = Ae^{-i\omega t} + \bar{A}e^{i\omega t}. \quad (3)$$

In quantum mechanics, $x(t)$ will be the expected value of the position operator x , written as $\langle x(t) \rangle$. (We are abusing notation by using the same symbol for the classical position and the quantum operator. This is common practice in physics.) Since the expected value is real, certainly both positive and negative frequencies appear, just as in Eq.(3).

Summary The really important points of the above discussion are: (1) Einstein and Planck have taught us that energy and frequency are proportional, the proportionality constant being Planck's constant. (2) Energy is not like momentum where positive and negative values both occur equally. The energy of any system has a lower bound and allowed energies go up from there. So it is natural when we write a quantity that varies as $\sim \exp(-i\omega t)$, that we will not always have a balancing term the varies as $\sim \exp(i\omega t)$, although this will happen when we are examining an expected value of an operator. Since complex conjugates are bound to occur, there will also be cases where terms $\sim \exp(+i\omega t)$ do not have a balancing $\sim \exp(-i\omega t)$. **Bottom line:** In quantum mechanics, complex numbers are essential.

2 Photon Polarization, Probability and Interference in Quantum Mechanics

The existence of polarized light gives a wonderful example where both classical ideas and one of the most puzzling features of quantum mechanics coexist. Polarization is very familiar in daily life. Almost everyone who wears sunglasses wears polarized sunglasses, which are designed to absorb horizontally polarized light. This can be experienced at the local gas station when you fill your gas tank. If you are wearing polarized sunglasses, you will quickly discover that the light emitted by the display on the gas pump is polarized. For those who do not operate cars, try putting on your polarized sunglasses the next time you are in front of a computer monitor. Tilting your head reveals that the light from the monitor is highly polarized.

Let us start by discussing a classical beam of monochromatic light, traveling along the $+z$ axis. Imagine we have ideal polarizer placed in this beam with its axis along the $+y$ axis. This means that light polarized along \hat{y} is (ideally) 100% transmitted, while light polarized along the \hat{x} axis is totally absorbed. Now if the incident light is polarized at say an angle θ with respect to the y axis, we simply resolve the polarization into a

component along the y axis, and one along the x axis. Only the component along the y axis gets transmitted, and the transmitted intensity satisfies the classical Malus Law:

$$I(\theta) = I(0) \cos^2 \theta, \quad (4)$$

where $I(\theta)$ is the *intensity* of transmitted light, when the incident light is polarized at angle θ to the axis of the polarizer. This is all well-known seemingly pure classical physics.

Now Malus' law puts no restriction on the intensity of the incident light. So imagine that we gradually but steadily reduce the intensity to the point where we are considering the process photon by photon. At this point quantum physics enters the picture, but only in a very general way. We visualize a classical light beam as composed of photons, each satisfying the Einstein-Planck formula relating frequency and energy. For this discussion, we don't need this formula. All we do need is the assertion that photons cannot be broken into pieces, only whole photons are observed beyond the polarizer. Now as long as the photon's polarization vector is either along \hat{y} , or along \hat{x} , there is still no need for new ideas. The photon polarized along \hat{y} gets transmitted, the one polarized along \hat{x} gets absorbed. New ideas do enter when we have a photon incident whose polarization vector is at angle θ with respect to \hat{y} . A polarization vector for an individual photon is a *unit* vector, which we could write for our present discussion as

$$\vec{P} = \cos \theta \hat{y} + \sin \theta \hat{x}. \quad (5)$$

This is the polarization vector of the photon which is incident on the polarizer. Beyond the polarizer, there either a photon or no photon. If there is a photon, its polarization vector is \hat{y} . When this experiment is repeated over and over, always with the same polarization for the incident photon, the frequency of observation of photons transmitted beyond the polarizer is $\cos^2 \theta$. Whether a given photon is transmitted is apparently random. Large numbers of observations are needed to observe the $\cos^2 \theta$ law. This is indeed something new and unfamiliar at the macroscopic level. To interpret this result requires the introduction of probability into quantum mechanics. In our example, probability is invoked as follows. We know that if the photon polarization is along \hat{y} , the photon is transmitted, while if its polarization is along \hat{x} , it is absorbed. We say that the *probability amplitude* of our photon being found with its polarization along \hat{y} is

$$\hat{y} \cdot \vec{P} = \cos \theta, \quad (6)$$

and the *probability* of the photon being found with polarization along \hat{y} is the absolute square of the probability amplitude. Denoting this probability as p_y , we can write

$$p_y = |\hat{y} \cdot \vec{P}|^2 \quad (7)$$

Note that we did not need the absolute value in Eq.(7). However, probability amplitudes are complex in general, so to get a real probability, it is necessary to take the absolute square of the probability amplitude. Eq.(7) allows us to explain the experiment in a statistical sense. No answer can be given to the question of what transpires for an

individual photon. Answers can only be given to questions involving many repeats of the experiment. At this level, quantum probability is like coin-tossing. One cannot say what the outcome will be for a toss of an individual coin. Many tosses are needed to see that the chance of “heads” is 0.50 for an unbiased coin. The idea of “hidden” variables in quantum mechanics can be explained using coin tosses. While we say that the outcome of a coin toss is random, if we were willing to undertake the enormous task of solving the equations of motion taking account of all possible forces and carefully specifying the initial conditions, the outcome of a coin toss is predictable. In the same way, so-called hidden variable theories of quantum behavior postulate the existence of subtle or hidden variables, whose consideration would allow the apparent randomness of outcomes to be understood as not really random after all. Hidden variable theories have had a difficult time and it is safe to say that no real evidence exists that any of the ones proposed so far can accomplish the task of taking the probability out of quantum mechanics. We are left with the situation that the probabilities we have in quantum mechanics are really intrinsic, and there is no deeper level at which there are new variables which if treated would remove randomness. It is well-known that this intrinsic nature of probability in quantum mechanics was bothersome to some very great physicists. Einstein is the most famous example, but Schrödinger also had serious reservations, some of which have been ignored rather than answered.

Given that there is no working alternative, let us go on to state more clearly the situation with respect to probability in quantum mechanics. To warm up, there are some comments that can be made about our polarized photon example. First note that no energy is getting lost. The incident photon either goes through or is absorbed. If it is absorbed, the energy of the photon is given to the absorbing material. That is fairly obvious. The polarization experiment is really a scattering experiment, elastic scattering for the transmitted photon, inelastic for the absorbed photon. Experiments involving scattering often give clear examples of quantum probability. Second, note that we were careful to discuss the polarization of the photon, not its electric field. Classically we think of the polarization is simply the direction of the electric field of the wave. This is fine if we have many photons and in fact an *indefinite* number of photons. However, at the single photon level, there is no “value” of the electric field. As we will see later, the electric field operator is non-diagonal in photon number. Nevertheless, the polarization of an individual photon makes perfect sense. It is a unit vector, perpendicular to the direction of wave propagation. Note that dotting two ordinary vectors does not lead to a quantity whose square could be a probability. However, if they are *unit* vectors, a probability interpretation is possible. The introduction of probability into quantum mechanics was mainly the work of Max Born, who reasoned from the results of scattering experiments.

Probability in Quantum Mechanics

Now we can state the way probability comes into quantum mechanics in a rather general way. Note that this is a *huge* extrapolation from our photon polarization example, but it nevertheless fits all known experience. First the notion of *quantum state*. A quantum state is a vector in a complex vector space, possibly one of an infinite number of dimensions. Call the vector \mathbf{u} . Now suppose there is another state of our system, described by a different vector \mathbf{v} . Generalizing the discussion in the photon case, it is necessary that these both be *unit vectors*, so our space needs to have a notion of length for its vectors. Next, to define probability amplitudes, we need to be able to take the “dot” or inner product of two vectors, denoted as (\mathbf{v}, \mathbf{u}) , so our space needs to be an inner product space. Finally, if the state is \mathbf{u} , the probability that it is found to be in \mathbf{v} is $|(\mathbf{v}, \mathbf{u})|^2$, i.e. the absolute square of the probability amplitude.

- Quantum State $\longleftrightarrow \mathbf{v}$, a *unit* vector in a complex space.
- Probability Amplitude = (\mathbf{v}, \mathbf{u}) , the amplitude to find state \mathbf{v} in a known state \mathbf{u} .
- Probability = $|(\mathbf{v}, \mathbf{u})|^2$, the probability to find state \mathbf{v} in a known state \mathbf{u} .

Interference and Superposition in Quantum Mechanics So far we have used the polarization of photons as a vehicle for understanding how probability comes into quantum mechanics in an essential way. We could use photon polarization for a discussion of interference as well. However, there are more familiar examples. Interference is really a hallmark of *wave* phenomena. Waves arriving at a given point via different paths may add up or cancel each other out. Young’s two slit experiment, the behavior of diffraction gratings, etc are familiar examples from classical physics with light. Many more examples can be given in classical physics, most notably from acoustics. The Davisson-Germer experiment demonstrated for the first time that interference can happen for material particles like electrons. This means that when an electron of reasonably well-defined momentum enters a crystal and then leaves again, there can be maxima and minima observed between electron waves arriving at a given point via different paths. Amplitudes may add constructively or destructively. We have said already that to even describe waves for material particles, we need complex quantities. Since interference involves superposition, adding such quantities must give a quantity of the same type. More formally it is essential that when we add (superpose) objects representing states, we get other objects representing states, and here superposition involves not just addition but addition and multiplication with complex numbers.

Superposition in Quantum Mechanics

A quantum state is a vector in a complex vector space, possibly one of an infinite number of dimensions. Call the vector \mathbf{u} . Now suppose there is another state of our system, described by a different vector \mathbf{v} . The idea of *superposition* states that combining these vectors as in

$$\mathbf{w} = \alpha\mathbf{v} + \beta\mathbf{v} \quad (8)$$

must give another vector \mathbf{w} , which when normalized, is a possible state of the system. (Here α and β are complex numbers.) This is maximally weird when thought of in terms of classical particle physics. Adding twice one state of a particle to 1/10 of another state of the particle to get a third state of the particle? No way in classical physics. But it is absolutely required by experiment in quantum physics!

The next sections will go into some of the features of complex vector spaces in detail.

3 Complex Vectors

Let us start with an ordinary vector in three-dimensional space, as used in elementary physics. We usually expand such a vector in an **orthonormal basis**,

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3, \quad (9)$$

where the \mathbf{e}_i are orthogonal unit vectors, and the v_i are real scalars (not vectors), called the **components** of \mathbf{v} . For the words **orthogonal** and **unit vector** to make sense there must be a scalar or **inner product**. In elementary physics we write this as a “dot product”,

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}. \quad (10)$$

For our work in quantum mechanics, a different notation will be essential. So for the scalar product of \mathbf{e}_i with \mathbf{e}_j , we write $(\mathbf{e}_i, \mathbf{e}_j)$. (Later, we will also use $\langle \mathbf{e}_i, \mathbf{e}_j \rangle$ and $\langle \mathbf{e}_i | \mathbf{e}_j \rangle$ to mean the same thing.) Apart from notation, certain generalizations of ordinary 3-vectors are necessary. First, since our quantum vectors reside generally in the space of **quantum states**, the dimensionality of the space will usually not be three. We will work in n dimensions, where $n = 2, 3, \dots$, including spaces of an infinite number of dimensions. Second, having seen that quantum mechanics must involve complex numbers, the components of vectors must be allowed to be complex numbers. The scalar product is also naturally generalized. The scalar product of a vector \mathbf{v} with a vector \mathbf{u} will be denoted as (\mathbf{v}, \mathbf{u}) . This is a complex number, defined so that

$$(\mathbf{v}, \mathbf{u}) = \overline{(\mathbf{u}, \mathbf{v})}, \quad (11)$$

where a bar over a quantity means complex conjugate. As far as unit vectors are concerned, we still have

$$(\mathbf{e}_i, \mathbf{e}_j) = (\mathbf{e}_j, \mathbf{e}_i) = \delta_{ij}, \quad (12)$$

but for more general vectors, it does matter which vector is on the right or the left in (\mathbf{v}, \mathbf{u}) . The physics convention is that (\mathbf{v}, \mathbf{u}) is **linear** in \mathbf{u} and **antilinear** in \mathbf{v} . Let us illustrate with a simple example. Suppose $\mathbf{u} = \alpha \mathbf{e}_j$, and $\mathbf{v} = \beta \mathbf{e}_k$, where α and β are arbitrary complex numbers. Then by linearity

$$(\mathbf{v}, \mathbf{u}) = (\mathbf{v}, \alpha \mathbf{e}_j) = \alpha (\mathbf{v}, \mathbf{e}_j), \quad (13)$$

and by anti-linearity

$$(\mathbf{v}, \mathbf{e}_j) = (\beta \mathbf{e}_k, \mathbf{e}_j) = \bar{\beta} (\mathbf{e}_k, \mathbf{e}_j), \quad (14)$$

so we would get

$$(\mathbf{v}, \mathbf{u}) = \bar{\beta} \alpha (\mathbf{e}_k, \mathbf{e}_j) = (\bar{\beta} \alpha) \delta_{kj}. \quad (15)$$

Now consider a less trivial example. Suppose our space has n dimensions, so \mathbf{u} and \mathbf{v} can be expanded as

$$\begin{aligned} \mathbf{u} &= \sum_{j=1}^n u_j \mathbf{e}_j \\ \mathbf{v} &= \sum_{k=1}^n v_k \mathbf{e}_k \end{aligned} \quad (16)$$

The scalar product of \mathbf{v} and \mathbf{u} is

$$(\mathbf{v}, \mathbf{u}) = \left(\sum_k v_k \mathbf{e}_k, \sum_j u_j \mathbf{e}_j \right) = \sum_{k,j} \bar{v}_k u_j (\mathbf{e}_k, \mathbf{e}_j) = \sum_j \bar{v}_j u_j \quad (17)$$

It is obvious from this equation that $(\mathbf{v}, \mathbf{v}) \geq 0$ so for any non-zero vector \mathbf{v} , we can define its **norm** by

$$\|\mathbf{v}\| = \sqrt{(\mathbf{v}, \mathbf{v})}. \quad (18)$$

An application of our formulas so far is to prove the Schwarz inequality, which states that

$$|(\mathbf{v}, \mathbf{u})| \leq \|\mathbf{v}\| \cdot \|\mathbf{u}\|, \quad (19)$$

for any two vectors \mathbf{u} and \mathbf{v} . Rearranging the formula Eq.(19), we can write it as

$$|(\mathbf{e}_v, \mathbf{e}_u)| \leq 1, \quad (20)$$

where

$$\mathbf{e}_v = \frac{\mathbf{v}}{\|\mathbf{v}\|} \quad (21)$$

and

$$\mathbf{e}_u = \frac{\mathbf{u}}{\|\mathbf{u}\|}$$

are unit vectors.

Exercise 2 To prove the Schwarz inequality in the form of Eq.(20), consider the vector $\mathbf{w} = \alpha \mathbf{e}_v + \frac{1}{\alpha} \mathbf{e}_u$, where α is a complex number. Eq.(20) may be proven by demanding that $(\mathbf{w}, \mathbf{w}) \geq 0$, and choosing the magnitude and phase of α appropriately.

Linear Independence and Bases The notion of linear independence is a simple but useful one. Suppose our space has n dimensions, and that we have n vectors $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$. This set is linearly independent if any vector in the space can be written as

$$\mathbf{u} = \sum_{j=1}^n u_j \mathbf{b}_j. \quad (22)$$

Said another way, there are n linearly independent vectors in an n dimensional space. Any such set of n linearly independent vectors is called a **basis**. As a simple example of linear independence, think of ordinary 3 dimensional space. Pick any vector in the space for \mathbf{b}_1 , then pick a second vector not parallel to \mathbf{b}_1 for \mathbf{b}_2 . All possible vectors composed of \mathbf{b}_1 and \mathbf{b}_2 compose a plane. Pick a third vector \mathbf{b}_3 which does not lie totally in that plane. Then $\mathbf{b}_1, \mathbf{b}_2$, and \mathbf{b}_3 are a set of linearly independent 3-vectors and form a basis for 3 dimensional space.

Orthogonalization Once we have a basis, we can find the components of a vector. From Eq.(22) we have

$$(\mathbf{b}_k, \mathbf{u}) = \sum_{j=1}^n u_j (\mathbf{b}_k, \mathbf{b}_j), \quad (23)$$

so we can solve for the u_j by solving n equations in n unknowns. Although perfectly possible, this is certainly not a convenient procedure. It is much more convenient to have an orthonormal basis. Orthogonalization (or the Gram-Schmidt procedure) is a method for constructing an orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ from an arbitrary basis $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$. Start by defining

$$\mathbf{e}_1 = \frac{\mathbf{b}_1}{\|\mathbf{b}_1\|}.$$

Next set $\mathbf{h}_2 = \mathbf{b}_2 - \mathbf{e}_1(\mathbf{e}_1, \mathbf{b}_2)$. This removes the part of \mathbf{b}_2 along \mathbf{e}_1 . Then set

$$\mathbf{e}_2 = \frac{\mathbf{h}_2}{\|\mathbf{h}_2\|}.$$

The process continues by setting $\mathbf{h}_3 = \mathbf{b}_3 - \mathbf{e}_1(\mathbf{e}_1, \mathbf{b}_3) - \mathbf{e}_2(\mathbf{e}_2, \mathbf{b}_3)$, which removes the part of \mathbf{b}_3 along \mathbf{e}_1 and \mathbf{e}_2 . Then set

$$\mathbf{e}_3 = \frac{\mathbf{h}_3}{\|\mathbf{h}_3\|},$$

and so on.

Hilbert Space A Hilbert space H is (roughly) the limit as $n \rightarrow \infty$ of our discussion up to now. The example closest to our previous discussion is called l^2 . In this space, a vector is just a infinite list of components. The scalar product of vectors \mathbf{u} and \mathbf{v} is naturally written as the limit of Eq.(17),

$$(\mathbf{u}, \mathbf{v}) = \sum_{j=1}^{\infty} \bar{u}_j v_j, \quad (24)$$

where u_j and v_j are elements of the space at all both

$$(\mathbf{v}, \mathbf{v}) = \sum_{j=1}^{\infty} \bar{v}_j v_j, \quad (25)$$

and

$$(\mathbf{u}, \mathbf{u}) = \sum_{j=1}^{\infty} \bar{u}_j u_j, \quad (26)$$

must be finite. We may define the orthonormal basis of this space by the following equations

$$\begin{aligned} \mathbf{e}_1 &= \{1, 0, 0, \dots\}, \\ \mathbf{e}_2 &= \{0, 1, 0, \dots\}, \\ \mathbf{e}_3 &= \{0, 0, 1, \dots\}, \\ &\dots\dots\dots, \end{aligned} \quad (27)$$

which clearly defines an infinite set of linearly independent vectors.

In order to be mathematically well-defined a Hilbert space H needs to obey certain extra conditions. We will assume that our Hilbert spaces are all “nice” and do obey these requirements. For the record they are called *completeness* and *separability*. The following is a rough description of these. Once we have a scalar product, we can define the norm of a vector, see Eq.(18). The **distance** between two vectors is then defined as the norm of the difference,

$$D(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| \quad (28)$$

Now suppose we have a sequence of vectors \mathbf{u}_n such that

$$D(\mathbf{u}_n, \mathbf{u}_m) \rightarrow 0,$$

as $m, n \rightarrow \infty$. This says the members of the sequence are getting closer and closer together. Now are they getting close to a vector that is the limit of the sequence? If so the space is *complete*. This is very analogous to the relationship of rational numbers to irrational numbers. The requirement of separability is as follows. There is a countable set of vectors in H such that for any vector \mathbf{u} in H , we can find a member of the set \mathbf{g} , such that $D(\mathbf{u}, \mathbf{g}) < \epsilon$ for any small ϵ . This requirement holds for all the spaces used in ordinary quantum mechanics, and we need not consider it explicitly any further.

Another example We have mentioned the Hilbert space l^2 . A second example is L^2 , the space of square integrable functions on an interval, perhaps infinite in length. A simple but important case is square integrable functions on the interval $(0, 2\pi)$. The scalar product is

$$(\mathbf{u}, \mathbf{v}) = \int_0^{2\pi} \bar{u}(x)v(x) dx. \quad (29)$$

An orthonormal basis for $L^2(0, 2\pi)$ is the set of functions

$$\frac{1}{\sqrt{2\pi}} \exp(ikx) \quad (\pm k = 0, 1, 2, \dots). \quad (30)$$

We may establish a correspondence between the vectors in the abstract sense and their realization as functions in $L^2(0, 2\pi)$,

$$\begin{aligned} \mathbf{v} &\leftrightarrow v(x) \\ \mathbf{e}_k &\leftrightarrow \frac{1}{\sqrt{2\pi}} \exp(ikx) \end{aligned} \quad (31)$$

This space is important and we will work with it more later on.

Comment 2 An eternal question in any advanced part of physics is how much math is needed? In the case of Hilbert space, a great deal has been worked out, much of it by John von Neumann in a series of papers written in the 1920's and his famous book, "The Mathematical Foundations of Quantum Mechanics". However, few physicists have the time or the inclination to delve into the technicalities of Hilbert space or other mathematical subjects that deeply. A reasonable attitude is to make sure the physics equations written down are understood and make sense. Physics arguments can often be used to settle fine points. A quote from Feynman on higher mathematics is "**Don't know how to prove what's true. Just know what's true!**" We will try to follow this suggestion.

4 Operators and Adjoins

An operator is a rule generating a vector in our space by acting on an original vector. We will only need to consider **linear** operators, so if T is an operator, we have

$$T(\alpha \mathbf{u}_1 + \beta \mathbf{u}_2) = \alpha T\mathbf{u}_1 + \beta T\mathbf{u}_2. \quad (32)$$

Generally speaking, in quantum mechanics, operators either represent physical quantities like momentum, energy, angular momentum, or represent transformations like rotations, translations, Lorentz transformations, or simple changes of coordinate system in our complex space.

An important concept for a linear operator T is its **adjoint**, denoted as T^\dagger . The defining property of the adjoint of an operator is

$$(\mathbf{u}, T\mathbf{v}) = (T^\dagger \mathbf{u}, \mathbf{v}) \quad (33)$$

To get used to the concept of adjoint, let us look at an example using a orthonormal basis. Our space can be finite- or infinite-dimensional. We will assume all infinite sums converge. Let

$$\mathbf{u} = \sum_j u_j \mathbf{e}_j \quad (34)$$

Acting on \mathbf{u} with T , we have

$$T\mathbf{u} = \sum_j u_j T\mathbf{e}_j. \quad (35)$$

Now $T\mathbf{e}_j$ can be expanded in our space, so we can write

$$T\mathbf{e}_j = \sum_k \mathbf{e}_k t_{kj}, \quad (36)$$

The peculiar way terms are written in Eq.(36) becomes clearer when we take matrix elements of T . We have

$$(\mathbf{e}_k, T\mathbf{e}_j) = t_{kj}, \quad (37)$$

where t_{kj} are a set of numbers, complex in general.

NOTE: Physics notation often uses the same letter to denote several different things. We will do this eventually, but for now, we are trying to distinguish the operator T from its matrix elements in an orthonormal basis, $(\mathbf{e}_j, T\mathbf{e}_k)$, so we use t_{jk} to denote them rather than T_{jk} .

Turning to the adjoint of T , by analogy with Eq.(36) we may write

$$T^\dagger \mathbf{e}_j = \sum_k \mathbf{e}_k \tilde{t}_{kj}, \quad (38)$$

where the \tilde{t}_{kj} are another arbitrary set of complex numbers. Taking matrix elements, we have

$$(\mathbf{e}_k, T^\dagger \mathbf{e}_j) = \tilde{t}_{kj}, \quad (39)$$

so the \tilde{t}_{kj} are the matrix elements of T^\dagger . Computing $T^\dagger \mathbf{u}$ we have

$$T^\dagger \mathbf{u} = \sum_j u_j T^\dagger \mathbf{e}_j = \sum_{\mathbf{k}, j} \mathbf{e}_{\mathbf{k}} \tilde{t}_{kj} u_j. \quad (40)$$

Substituting this into $(T^\dagger \mathbf{u}, \mathbf{v})$ we have

$$(T^\dagger \mathbf{u}, \mathbf{v}) = \left(\sum_{\mathbf{k}, j} u_j \mathbf{e}_{\mathbf{k}} \tilde{t}_{kj}, \sum_{\mathbf{m}} v_{\mathbf{m}} \mathbf{e}_{\mathbf{m}} \right) = \sum_{\mathbf{k}, j} \bar{u}_j (\tilde{t}_{kj})^* v_{\mathbf{k}}. \quad (41)$$

Now consider $(\mathbf{u}, T\mathbf{v})$. An easy calculation gives

$$(\mathbf{u}, T\mathbf{v}) = \sum_{\mathbf{k}, j} \bar{u}_j t_{jk} v_{\mathbf{k}}. \quad (42)$$

By definition, $(\mathbf{u}, T\mathbf{v}) = (T^\dagger \mathbf{u}, \mathbf{v})$. Further, \mathbf{u} and \mathbf{v} are arbitrary vectors in our space, so from Eqs.(40) and (42) we must have

$$(\tilde{t}_{kj})^* = t_{jk} \quad (43)$$

A more illuminating form for this equation follows when we use Eqs.(37) and (39). Using those equations, Eq.(43) becomes

$$(\mathbf{e}_{\mathbf{k}}, T^\dagger \mathbf{e}_j)^* = (\mathbf{e}_j, T\mathbf{e}_{\mathbf{k}}). \quad (44)$$

At this point, let us revert to standard physics notation, and write

$$\begin{aligned} (\mathbf{e}_j, T\mathbf{e}_{\mathbf{k}}) &= T_{jk} \\ (\mathbf{e}_j, T^\dagger \mathbf{e}_{\mathbf{k}}) &= T_{jk}^\dagger, \end{aligned} \quad (45)$$

and $T_{jk} = (T_{kj}^\dagger)^*$. Now if we write T out as a matrix array T_{jk} (this called “defining an operator by its matrix elements”), then the matrix for T^\dagger or T_{jk}^\dagger is just the adjoint in the matrix sense of the matrix for T . (Recall that the matrix adjoint operation takes the complex conjugate of every element and replaces the resulting matrix by its transpose.)

Self-Adjoint and Unitary Operators The two most important classes of operators used in quantum mechanics are self-adjoint (or Hermitian) operators, and unitary operators. A self-adjoint operator is one for which the adjoint is the same as the operator,

$$T^\dagger = T. \quad (46)$$

Self-adjoint operators correspond to physical quantities like energy, momentum, etc. Let us consider the *eigenvalue* of a self-adjoint operator. Suppose T is self-adjoint, and there is a vector \mathbf{u} in our space such that

$$T\mathbf{u} = \lambda\mathbf{u}, \quad (47)$$

where λ is a number called the *eigenvalue*. Consider the following matrix element,

$$(\mathbf{u}, T\mathbf{u}) = \lambda(\mathbf{u}, \mathbf{u}). \quad (48)$$

Now take complex conjugates of both sides,

$$(\mathbf{u}, T\mathbf{u})^* = \lambda^*(\mathbf{u}, \mathbf{u}) \quad (49)$$

Now by definition,

$$(\mathbf{u}, T\mathbf{u})^* = (T\mathbf{u}, \mathbf{u}). \quad (50)$$

Now use the fact that T is self-adjoint to get

$$(T\mathbf{u}, \mathbf{u}) = (\mathbf{u}, T\mathbf{u}) \quad (51)$$

But now from Eqs.(48) and (49) we have

$$\lambda(\mathbf{u}, \mathbf{u}) = \lambda^*(\mathbf{u}, \mathbf{u}), \quad (52)$$

which implies $\lambda = \lambda^*$, i.e. the eigenvalue is *real*. So self-adjoint operators have real eigenvalues. (**NOTE:** The argument just given is fine for λ in the point or discrete spectrum of T . It needs extension for the case where T has a continuous spectrum. Remarks will be made on this in a later section.)

Another characteristic property of self-adjoint operators has to do with multiple eigenvalues. Suppose a self-adjoint operator T has two distinct eigenvalues: $T\mathbf{u} = \lambda\mathbf{u}$, and $T\mathbf{v} = \mu\mathbf{v}$. We have

$$(\mathbf{v}, T\mathbf{u}) = \lambda(\mathbf{v}, \mathbf{u}) = (T\mathbf{v}, \mathbf{u}) = \mu(\mathbf{v}, \mathbf{u}). \quad (53)$$

But this says

$$(\lambda - \mu)(\mathbf{v}, \mathbf{u}) = 0, \quad (54)$$

so if $\lambda \neq \mu$, we must have $(\mathbf{v}, \mathbf{u}) = 0$, so eigenvectors with different eigenvalues must be orthogonal.

Now let us consider unitary operators. The defining property of a unitary operator U is

$$(\mathbf{v}, \mathbf{u}) = (U\mathbf{v}, U\mathbf{u}), \quad (55)$$

i.e. U preserves the inner product of vectors. Now like any operator, U has a adjoint, so we can write

$$(U\mathbf{v}, U\mathbf{u}) = (U^\dagger U\mathbf{v}, \mathbf{u}). \quad (56)$$

But using Eq.(55) we have

$$(\mathbf{v}, \mathbf{u}) = (U^\dagger U\mathbf{v}, \mathbf{u}). \quad (57)$$

Now \mathbf{u} and \mathbf{v} were arbitrary vectors, so we must have

$$U^\dagger U = I \quad (58)$$

i.e. for a unitary operator, the adjoint of the operator is its inverse. Considering the eigenvalues of a unitary operator, suppose

$$U\mathbf{u} = \lambda\mathbf{u}, \quad (59)$$

where λ is a number. Now using the properties of unitary operators, we have

$$(U\mathbf{u}, U\mathbf{u}) = (\lambda\mathbf{u}, \lambda\mathbf{u}) = \lambda^*\lambda(\mathbf{u}, \mathbf{u}) = (\mathbf{u}, \mathbf{u}). \quad (60)$$

which says $\lambda^*\lambda = 1$, i.e. the eigenvalues of a unitary operator are of the form $\exp(i\theta)$. Again, this argument needs more work when dealing with the continuous spectrum.

Summary Self-adjoint operators have real eigenvalues. For a unitary operator, the adjoint is the inverse, and eigenvalues are complex numbers of unit modulus. Both of these results are true in general, although the discussion above has only covered the case of eigenvalues in the discrete spectrum of the operator. For a self-adjoint operator, it is expected that the eigenvectors will span the space, and any vector can be expanded in these eigenfunctions. A simple example of this is Fourier series on the interval $0, 2\pi$. The general expectation that an eigenfunction expansion exists is borne out in problems of physical interest.

Example Let us consider an example, chosen with an eye to physics applications. We work in a three dimensional complex space, and define an operator J by

$$\begin{aligned} J\mathbf{e}_1 &= i\mathbf{e}_2 \\ J\mathbf{e}_2 &= -i\mathbf{e}_1 \\ J\mathbf{e}_3 &= 0 \end{aligned} \quad (61)$$

Exercise 3 Show that J is self-adjoint using the definitions of the previous section.

We want to find eigenvalues and eigenvectors for J . Since we are in three dimension, we expect three eigenvalues and corresponding eigenvectors. We will denote our normalized eigenvectors as \mathbf{p}_λ , where λ is the eigenvalue. From Eq.(61), we have that \mathbf{e}_3 is an eigenvector with eigenvalue 0, so we write

$$\mathbf{e}_3 = \mathbf{p}_0 \quad (62)$$

To find the other eigenvalues and eigenvectors, we apply J a second time. Using Eq.(61) we have

$$\begin{aligned} J(J\mathbf{e}_1) &= iJ\mathbf{e}_2 = \mathbf{e}_1 \\ J(J\mathbf{e}_2) &= -iJ\mathbf{e}_1 = \mathbf{e}_2 \end{aligned} \quad (63)$$

Eq.(63) shows that for any vector in the subspace spanned by \mathbf{e}_1 and \mathbf{e}_2 , the operator JJ acts like the identity. Since the remaining eigenvectors reside in this space, we must have $\lambda^2 = 1$ for any eigenvalue. Let us try $\lambda = -1$. Expand \mathbf{e}_{-1} as

$$\mathbf{p}_{-1} = \alpha\mathbf{e}_1 + \beta\mathbf{e}_2, \quad (64)$$

where α and β are complex coefficients. Demanding that \mathbf{p}_{-1} be an eigenvector of J with eigenvalue -1 , we get

$$J\mathbf{p}_{-1} = \alpha J\mathbf{e}_1 + \beta J\mathbf{e}_2 = i\alpha\mathbf{e}_2 - i\beta\mathbf{e}_1 = -(\alpha\mathbf{e}_1 + \beta\mathbf{e}_2). \quad (65)$$

Matching coefficients of \mathbf{e}_1 and \mathbf{e}_2 , we find that $\beta = -i\alpha$. Choosing α real and positive, we have

$$\mathbf{p}_{-1} = \frac{1}{\sqrt{2}}(\mathbf{e}_1 - i\mathbf{e}_2). \quad (66)$$

There must be one more eigenvector. Demanding $\lambda = -1$ gave \mathbf{p}_{-1} uniquely, up to an overall constant. There can be no other eigenvector with eigenvalue -1 , so the remaining eigenvector must have eigenvalue $+1$. We can easily find it by demanding that it be orthogonal to \mathbf{p}_0 and \mathbf{p}_{-1} . This gives

$$\mathbf{p}_1 = \frac{1}{\sqrt{2}}(\mathbf{e}_1 + i\mathbf{e}_2). \quad (67)$$

The vectors $\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_{-1}$ are a complete set of eigenvectors of J , unique up to phase factors.

In this same space, we can define a unitary operator U by

$$\begin{aligned} U\mathbf{e}_1 &= \cos\theta\mathbf{e}_1 + \sin\theta\mathbf{e}_2 \\ U\mathbf{e}_2 &= \cos\theta\mathbf{e}_2 - \sin\theta\mathbf{e}_1 \\ U\mathbf{e}_3 &= \mathbf{e}_3 \end{aligned} \quad (68)$$

Exercise 4 Show that U is unitary using the definitions of the previous section.

Now let us investigate the action of U on the eigenvectors of J . We have

$$U\mathbf{p}_1 = \frac{1}{\sqrt{2}}(\cos\theta\mathbf{e}_1 + \sin\theta\mathbf{e}_2 + i(\cos\theta\mathbf{e}_2 - \sin\theta\mathbf{e}_1)) = \exp(-i\theta)\mathbf{p}_1. \quad (69)$$

Likewise, we have

$$U\mathbf{p}_{-1} = \frac{1}{\sqrt{2}}(\cos\theta\mathbf{e}_1 + \sin\theta\mathbf{e}_2 - i(\cos\theta\mathbf{e}_2 - \sin\theta\mathbf{e}_1)) = \exp(i\theta)\mathbf{p}_{-1}, \quad (70)$$

and

$$U\mathbf{p}_0 = \mathbf{p}_0. \quad (71)$$

We may summarize these equations by writing

$$U\mathbf{p}_\lambda = \exp(-i\lambda\theta)\mathbf{p}_\lambda. \quad (72)$$

But the λ 's are just the eigenvalues of J , so Eq.(72) is equivalent to

$$U\mathbf{p}_\lambda = \exp(-iJ\theta)\mathbf{p}_\lambda. \quad (73)$$

The exponential of a matrix is a well-defined quantity. Just expand the exponential using the standard power series, and apply it to \mathbf{p}_λ to show that Eq.(73) is correct. Now Eq.(73) applies to any \mathbf{p}_λ , and by linearity, applies to any linear combination of the \mathbf{p}_λ . This means it is true for an arbitrary vector in our space, so we can write the pure operator equation,

$$U = \exp(-iJ\theta). \quad (74)$$

This is an important result, the generalization of which we will see many times. First note that from Eq.(68), U is a rotation around the 3-axis of an ordinary three dimensional coordinate system. **NOTE:** We are taking what is called an *active* viewpoint, i.e. we rotate *vectors* not *coordinate systems*. The second point is from Eq.(74), we see that U is *generated* by the self-adjoint operator J . This structure occurs for all symmetry operations which depend continuously on certain parameters. The general structure is

- Symmetry operation \longleftrightarrow Unitary operator, U .
- Generator \longleftrightarrow Self-adjoint operator, O .
- General form: $U = \exp(-i\xi O)$, where ξ is the *parameter* of the symmetry operation, (angle, translation distance, etc)

As we will learn later, there is a close connection between the parameter of a symmetry operation and a corresponding physical quantity. Relevant for this example is the connection between angle of rotation, and angular momentum. Briefly stated, the operator J we have been discussing is a factor of \hbar times an angular momentum. Since we have established that J is generating a rotation around the 3 axis, we can write

$$J = \hbar\Sigma_3. \quad (75)$$

There will of course be similar objects which generate rotations around 1 and 2 axes. Call them Σ_1, Σ_2 . Returning to our operator U , we can now write it as

$$U = \exp\left(\frac{-i}{\hbar}\Sigma_3\theta\right). \quad (76)$$

Further, it follows from our previous results that

$$\Sigma_3\mathbf{p}_\lambda = \hbar\lambda\mathbf{p}_\lambda, \quad (77)$$

so the \mathbf{p}_λ are eigenstates of Σ_3 with eigenvalues $\hbar\lambda$.

Helicity of the Photon Suppose we have a single photon, traveling in the 3 direction. Like any other particle, a photon will have an *orbital* angular momentum. But we can certainly say that no matter what the orbital angular momentum is, it certainly has *zero* as its component parallel to its momentum. (Think of $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ in classical physics.) So any component of angular momentum we find along the 3 axis must be coming from

photon spin. Now \mathbf{p}_1 and \mathbf{p}_{-1} are possible photon polarization vectors. (Electromagnetic waves are transverse, so \mathbf{p}_0 is not a possible polarization vector.) We showed above that

$$\begin{aligned}\Sigma_3 \mathbf{p}_1 &= \hbar \mathbf{p}_1 \\ \Sigma_3 \mathbf{p}_{-1} &= -\hbar \mathbf{p}_{-1}.\end{aligned}\tag{78}$$

Since there is no component of *orbital* angular momentum along the 3 axis here, we interpret Σ_3 as the 3 component of the *spin* operator, and we have that \mathbf{p}_1 has \hbar and \mathbf{p}_{-1} has $-\hbar$ for their eigenvalues of spin. The component of spin parallel to the momentum of a particle is called its *helicity*. So \mathbf{p}_1 has helicity one, and \mathbf{p}_{-1} has helicity minus one. Transversality says there is no helicity zero component for a photon.

Projection Operators Projection operators are very useful. They are the natural generalization of the idea of projection in ordinary three dimensional space. Suppose we have an ordinary three dimensional vector \mathbf{v} , which we expand as

$$\mathbf{v} = v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 + v_3 \mathbf{e}_3$$

Suppose we want to resolve \mathbf{v} into a “parallel” part (along the 3 axis) and a “perpendicular” part (perpendicular to the 3 axis). This is of course very simple. We write

$$\mathbf{v} = \mathbf{v}_\perp + \mathbf{v}_\parallel,$$

where $\mathbf{v}_\perp = v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2$, and $\mathbf{v}_\parallel = v_3 \mathbf{e}_3$. Now let us generalize. The analog of the “perp” part in a more general setting is a *linear manifold*. Call it \mathcal{M} . The analog of the “parallel” part is “what is left over.” If our entire space is denoted as \mathcal{R} , then the left over part is $\mathcal{R} - \mathcal{M}$. It is a theorem of Hilbert space (totally trivial in three dimensional space) that we can break up any vector \mathbf{v} uniquely as follows:

$$\mathbf{v} = \mathbf{u} + \mathbf{w},$$

where \mathbf{u} is in \mathcal{M} and \mathbf{w} is in $\mathcal{R} - \mathcal{M}$. Now the *projection operator* $P_{\mathcal{M}}$ is defined by

$$P_{\mathcal{M}} \mathbf{v} = \mathbf{u}.$$

A projection operator is *linear*,

$$P_{\mathcal{M}}(\mathbf{v}_1 + \mathbf{v}_2) = P_{\mathcal{M}} \mathbf{v}_1 + P_{\mathcal{M}} \mathbf{v}_2,$$

and it satisfies

$$(P_{\mathcal{M}} \mathbf{v}, \mathbf{u}) = (\mathbf{v}, P_{\mathcal{M}} \mathbf{u}),$$

and

$$P_{\mathcal{M}} P_{\mathcal{M}} \mathbf{v} = P_{\mathcal{M}} \mathbf{v}.$$

A projection operator in general, usually denoted as E , is one that satisfies:

$$(E \mathbf{v}, \mathbf{u}) = (\mathbf{v}, E \mathbf{u}), \text{ and } EE = E$$

It is another theorem of Hilbert space, that such an E must be a projection in some \mathcal{M} , i.e. there will exist an \mathcal{M} such that

$$E = P_{\mathcal{M}}$$

Direct Product Spaces and Entanglement Our Hilbert space vectors often are composed of separate parts, each of which is a vector in a Hilbert space. The vector we are considering will then be in a space which is the “direct product” of two (or more) spaces. The notion of direct product can occur for spaces of any dimension, but is easiest to explain for spaces of finite dimension. Let us consider two spaces with dimensions N_1 and N_2 , with corresponding orthonormal bases. We may expand a vector in space 1 in the usual way,

$$\mathbf{v}_1 = \sum_{j=1}^{N_1} \mathbf{e}_{1j} v_{1j},$$

and likewise for a vector \mathbf{v}_2 in space 2. Now let us define a vector \mathbf{v} in the direct product space which is in fact the direct product of \mathbf{v}_1 and \mathbf{v}_2 . We write

$$\mathbf{v} = \mathbf{v}_1 \otimes \mathbf{v}_2 = \sum_{j,k=1}^{N_1, N_2} \mathbf{e}_{1j} \otimes \mathbf{e}_{2k} v_{1j} v_{2k}$$

The basis vectors in the product space are clearly the $\mathbf{e}_{1j} \otimes \mathbf{e}_{2k}$. They have two indices, and the scalar product is defined by

$$(\mathbf{e}_{1j} \otimes \mathbf{e}_{2k}, \mathbf{e}_{1j'} \otimes \mathbf{e}_{2k'}) \equiv \delta_{j,j'} \delta_{k,k'}$$

Using this definition we have

$$(\mathbf{v}, \mathbf{v}) = \sum_{j,k=1}^{N_1, N_2} (v_{1j}^* v_{1j}) (v_{2k}^* v_{2k})$$

At this point we may define the important concept of *entanglement*. Not every vector in the direct product space is of the form $\mathbf{v} = \mathbf{v}_1 \otimes \mathbf{v}_2$. This is clearly a special case. When the vector \mathbf{v} is not of the form of a direct product of a vector in space 1 times a vector in space 2, we say there is *entanglement*. It is very easy to construct an example. Consider the following vector

$$\mathbf{v} = (\mathbf{e}_{11} \otimes \mathbf{e}_{21}) v_{11} + (\mathbf{e}_{12} \otimes \mathbf{e}_{22}) v_{22}.$$

Here \mathbf{v} involves only the first two unit vectors in spaces 1 and 2, but it is not of the form $\mathbf{v} = \mathbf{v}_1 \otimes \mathbf{v}_2$, so we say the vector is entangled or there is entanglement present. We will delay further exploration of entanglement for now. Here we simply note that the concept is crucial in both quantum information theory and quantum measurement theory. Getting somewhat ahead of the story, suppose we could construct an entangled state of two photons, and the photons are then allowed to separate by a macroscopic distance. The presence of entanglement means that an experiment on system 1, far from system 2, can nevertheless reveal information about system 2 without directly accessing system 2. There are many examples of such experiments involving entangled photons. Another example where entanglement is a crucial concept is in quantum measurement. Here there is a quantum system S which interacts with a measuring device M . The description of

the whole system will involve a vector \mathbf{v} . If the measurement device is to be at all useful, the vector \mathbf{v} must show entanglement between quantum system and measuring device. This is so that extracting information from the system S yields information about the quantum system. If the vector \mathbf{v} were of the form, $\mathbf{v}_S \otimes \mathbf{v}_M$, extracting information from the measuring device by measuring an operator which only involves the measuring device degrees of freedom yields no information about the quantum system. This will become clear later.

The Spectrum of an Operator and Expansion Theorems The situation in the finite dimensional case is simple. A self-adjoint operator O has real eigenvalues λ , and the corresponding eigenvectors form a complete set. There may be and often is, *degeneracy*, i.e. vectors in a certain linear manifold may all have the same eigenvalue of O . We allow for this case by denoting normalized eigenvectors as $\mathbf{e}_{\rho,\nu}$, where ρ is an index labeling *different* eigenvalues, and for a given eigenvalue, $\nu = 1, 2, \dots, \nu_\rho$, where ν_ρ is the degeneracy of the eigenvalue. (The index ν is not needed unless the eigenvalue is degenerate.) We can illustrate this by a simple example. Suppose O written out as a matrix is

$$O = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & \lambda_2 \end{pmatrix}$$

Here we have only 2 eigenvalues; λ_1 and λ_2 , where $\lambda_1 < \lambda_2$. We have 4 orthogonal eigenvectors;

$$\mathbf{e}_1, \mathbf{e}_{2,1}, \mathbf{e}_{2,2}, \mathbf{e}_{2,3}.$$

We may expand any vector \mathbf{u} in our four dimensional space in terms of the eigenvectors of O ;

$$\mathbf{u} = u_1 \mathbf{e}_1 + u_{2,1} \mathbf{e}_{2,1} + u_{2,2} \mathbf{e}_{2,2} + u_{2,3} \mathbf{e}_{2,3},$$

where $u_1 = (\mathbf{e}_1, \mathbf{u})$, etc.

There are various ways of writing out the operator O itself. Here we will show how to do it using projection operators. This generalizes to the case of infinite dimensions, and the case where O has a continuous spectrum. Take any vector in our four dimensional space and resolve it into parts,

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2,$$

where $O\mathbf{v}_1 = \lambda_1 \mathbf{v}_1$, and $O\mathbf{v}_2 = \lambda_2 \mathbf{v}_2$. We define a projection operator which depends on a parameter λ ;

$$E(\lambda)\mathbf{v} = \begin{cases} 0, & \lambda < \lambda_1 \\ \mathbf{v}_1, & \lambda_1 \leq \lambda < \lambda_2 \\ \mathbf{v}, & \lambda_2 \leq \lambda \end{cases}$$

This states that, acting on any vector \mathbf{v} , $E(\lambda)$ projects more and more of the vector as λ increases. Every time an eigenvalue is passed, the part of \mathbf{v} belonging to that eigenvalue is added, until finally as $\lambda \rightarrow \infty$, the whole vector \mathbf{v} is obtained.

To make use of $E(\lambda)$, take two arbitrary vectors \mathbf{u} and \mathbf{v} and form $(\mathbf{u}, O\mathbf{v})$. We will write $(\mathbf{u}, O\mathbf{v})$ as an integral involving $(\mathbf{u}, E(\lambda)\mathbf{v})$ as follows:

$$(\mathbf{u}, O\mathbf{v}) = \int_{-\infty}^{\infty} \lambda d(\mathbf{u}, E(\lambda)\mathbf{v}) \quad (79)$$

To see that this is correct, note that $(\mathbf{u}, E(\lambda)\mathbf{v})$ is constant between eigenvalues and then jumps as an eigenvalue is crossed. This means that $d(\mathbf{u}, E(\lambda)\mathbf{v})$ is zero unless we cross an eigenvalue and gives a Dirac delta function at an eigenvalue. We have

$$d(\mathbf{u}, E(\lambda)\mathbf{v}) = (\delta(\lambda - \lambda_1)(\mathbf{u}, \mathbf{v}_1) + \delta(\lambda - \lambda_2)(\mathbf{u}, \mathbf{v}_2)) d\lambda. \quad (80)$$

Exercise 5 Using what you know about how differentiating a step function produces a Dirac delta function, show that the factor multiplying $\delta(\lambda - \lambda_2)d\lambda$ in Eq.(80) is correct.

Substituting this formula into Eq.(79) gives

$$(\mathbf{u}, O\mathbf{v}) = \lambda_1(\mathbf{u}, \mathbf{v}_1) + \lambda_2(\mathbf{u}, \mathbf{v}_2)$$

which is correct. We can strip off the two vectors and write

$$O = \int_{-\infty}^{\infty} \lambda dE(\lambda).$$

This is a symbolic equation, since it involves integration over an operator. Nevertheless, this is a useful representation, and becomes concrete when we reinstate the two vectors \mathbf{u} and \mathbf{v} . Although we carried through the discussion for our simple four dimensional example, it is clear that the same approach would work for any self-adjoint operator in a finite dimensional space. The basic picture is that as we integrate, each time we pass an eigenvalue, we add in a projection operator for all the eigenvectors of that eigenvalue times the eigenvalue, until all eigenvalues have been included.

Let us now turn to the case of an infinite dimensional space, and a continuous spectrum. First let us specify more carefully what the term “continuous spectrum” means. In a finite dimensional space, when a self-adjoint operator O has an eigenvalue λ , we write

$$O\mathbf{v}_\lambda = \lambda\mathbf{v}_\lambda$$

for an eigenvector \mathbf{v}_λ . It is implicit in this equation that \mathbf{v} can be normalized, i.e. we can choose a normalization factor and require $(\mathbf{v}_\lambda, \mathbf{v}_\lambda) = 1$. This same behavior can and does occur in an infinite dimensional space. A self-adjoint operator may have certain

normalizable eigenvectors, and corresponding real eigenvectors. This is the so-called “discrete spectrum.” Everything here is rather similar to the case of a finite dimensional space. In contrast, for the continuous spectrum there are no normalizable eigenvectors. A very familiar example is the operator X corresponding to an ordinary linear coordinate. The action of X is as follows:

$$(\mathbf{u}, X\mathbf{v}) = \int_{-\infty}^{\infty} dx u^*(x) x v(x)$$

Intuitively, we expect the spectrum of X to be any value from $-\infty$ to $+\infty$, i.e. the spectrum is continuous. However, the operator X does not have normalizable eigenvectors. We can see this as follows. Suppose X had an eigenvalue x_1 and corresponding (normalizable) eigenvector \mathbf{u}_{x_1} . Then by the action of X , we would have

$$X\mathbf{u}_{x_1} = x u_{x_1}(x) = x_1 u_{x_1}(x).$$

But this requires $u_{x_1}(x) = 0$ for any $x \neq x_1$. So clearly normalizable eigenvectors or eigenfunctions are not possible in this case. Nevertheless we “know” that any real value of x should be in the spectrum of X . This is defined in the following way. Let O be any self-adjoint operator, and consider the operator $O(z) \equiv O - zI$, where I is the identity operator and z is any complex number. Then the spectrum of O is those values of z where $O(z)$ has no inverse. This slightly roundabout definition covers both discrete and continuous spectra. It is easy to show with this definition, that the spectrum of X is the real line. Writing out $(\mathbf{u}, O^{-1}(z)\mathbf{v})$, we have

$$(\mathbf{u}, O^{-1}(z)\mathbf{v}) = \int_{-\infty}^{\infty} dx u^*(x) \frac{1}{(x - z)} v(x)$$

This expression is finite as long as z has an imaginary part. However, if z is any real value, the denominator will vanish and the integral will blow up for almost any \mathbf{u} and \mathbf{v} . This is sufficient to show that the spectrum of X is the real axis.

Now let us write the operator X as an integral over its spectrum. First we define the projection operator $E(x_1)$, where x_1 is the eigenvalue. Just as in the finite dimensional case, we want $E(x_1)$ to project more and more of a vector as x_1 increases. We define

$$E(x_1)v(x) = \begin{cases} 0, & x > x_1 \\ v(x), & x \leq x_1 \end{cases}$$

Writing out $(\mathbf{u}, E(x_1)\mathbf{v})$ we have

$$(\mathbf{u}, E(x_1)\mathbf{v}) = \int_{-\infty}^{x_1} u^*(x) v(x) dx.$$

Differentiating with respect to x_1 we have

$$d(\mathbf{u}, E(x_1)\mathbf{v}) = u^*(x_1) v(x_1) dx_1.$$

From this it follows that

$$(\mathbf{u}, X\mathbf{v}) = \int_{-\infty}^{\infty} x_1 d(\mathbf{u}, E(x_1)\mathbf{v}),$$

or symbolically,

$$X = \int_{-\infty}^{\infty} x_1 dE(x_1).$$

This is admittedly a rather simple example. More challenging cases where the operator in question has both a discrete and a continuous spectrum will come up later on.

5 Dirac Notation

Up to now we have been denoting vectors in Hilbert space by boldface letters, \mathbf{u}, \mathbf{v} , scalar products by (\mathbf{u}, \mathbf{v}) , and the action of operators by formulas like $\mathbf{u} = T\mathbf{v}$. All of this is standard in books on Hilbert space, most of which follow von Neumann's pioneering work. (For more on von Neumann see http://en.wikipedia.org/wiki/John_von_Neumann.) In his book "Principles of Quantum Mechanics", P.A.M. Dirac introduced a different way of denoting these things, which has come to be known as "Dirac notation," and is almost universally used by physicists. It is important to emphasize that there is no change in physics implied, Dirac notation is just that, a notation. But notations can be very useful and allow certain concepts to be symbolized in an elegant and suggestive way.

As a first step in developing Dirac's notation, let us slightly change our notation for the scalar product,

$$(\mathbf{u}, \mathbf{v}) \longrightarrow \langle \mathbf{u} | \mathbf{v} \rangle .$$

This is truly just a change in symbols, (is traded for <, a comma becomes |, and) becomes >. The vectors \mathbf{u} and \mathbf{v} , are both exactly the same sort of object. Using an orthonormal basis, each can be thought of as represented by a list,

$$\mathbf{u} \longleftrightarrow \{u_1, u_2, \dots\},$$

$$\mathbf{v} \longleftrightarrow \{v_1, v_2, \dots\}.$$

The scalar product, whether denoted by (\mathbf{u}, \mathbf{v}) or $\langle \mathbf{u} | \mathbf{v} \rangle$ has two slots. In the right slot, we put the list that represents the vector \mathbf{v} , and in the left slot we put the *complex conjugate* of the list representing \mathbf{u} , so

$$(\mathbf{u}, \mathbf{v}) = \langle \mathbf{u} | = \sum_{n=1}^{\infty} \bar{u}_n v_n.$$

So far so good. The next move is to pull the scalar product apart. Quantum states are unit vectors in Hilbert space, but that is generally too much information. Reducing the amount of information to manageable or experimentally accessible size almost always

involves scalar products, with one slot left alone, and the other complex conjugated. Dirac pulls the scalar product apart as follows,

$$\langle \mathbf{u} | \leftarrow \langle \mathbf{u} | \mathbf{v} \rangle \rightarrow | \mathbf{v} \rangle .$$

The difference between $| \mathbf{v} \rangle$ and $\langle \mathbf{u} |$ is in the operation of complex conjugation,

$$| \mathbf{v} \rangle \longleftrightarrow \{v_1, v_2, \dots\},$$

but now

$$\langle \mathbf{u} | \longleftrightarrow \{\bar{u}_1, \bar{u}_2, \dots\},$$

so both are still lists, $| \mathbf{v} \rangle$ being a list of components, while $\langle \mathbf{u} |$ is a list of complex conjugate components. Dirac also introduced names for these objects. A scalar product $\langle \mathbf{u} | \mathbf{v} \rangle$ is called a “bracket.” Striking out the “c”, the object on the left in the scalar product, namely $\langle \mathbf{u} |$ is called “bra” (wince), and the object on the right, namely $| \mathbf{v} \rangle$, is called a “ket” (ugh). The writing in Dirac’s book and physics papers is normally quite elegant, so he may be forgiven for introducing these clunky names.

Now as we have said, $\langle \mathbf{u} |$ and $| \mathbf{v} \rangle$ are simply lists. However, for visualization purposes, it has become customary to represent kets as column vectors and bras as row vectors, so

$$\begin{aligned} | \mathbf{v} \rangle &\longrightarrow \begin{pmatrix} v_1 \\ v_2 \\ \dots \end{pmatrix} \\ \langle \mathbf{u} | &\longrightarrow (\bar{u}_1 \quad \bar{u}_2 \dots \quad) , \end{aligned}$$

so now the scalar product is

$$\langle \mathbf{u} | \mathbf{v} \rangle = (\bar{u}_1 \quad \bar{u}_2 \dots \quad) \cdot \begin{pmatrix} v_1 \\ v_2 \\ \dots \end{pmatrix} = \sum_{n=1}^{\infty} \bar{u}_n v_n .$$

As a further simple move, since every vector in Hilbert space is now either represented as a bra or a ket, we no longer need boldface letters so we can write $|u \rangle, |v \rangle$, or $|\Psi \rangle, |\Phi \rangle$, or likewise, $\langle u |, \langle v |$, or $\langle \Psi |, \langle \Phi |$. We may also simplify the notation for unit vectors in an orthonormal basis, $| \mathbf{e}_j \rangle$ simply becoming $|j \rangle$, and $\langle \mathbf{e}_j |$ becoming the bra, $\langle j |$.

Let us write out the expansion of a vector in an orthonormal basis in old notation and Dirac notation. We have

$$\mathbf{v} = \sum_{n=1}^{\infty} \mathbf{e}_n v_n ,$$

which becomes in Dirac notation,

$$|v \rangle = \sum_{n=1}^{\infty} |n \rangle v_n .$$

The components are the same either way,

$$v_n = (\mathbf{e}_n, \mathbf{v}) = v_n = \langle n | v \rangle .$$

Putting the last form in the Dirac expression, we have

$$|v \rangle = \sum_{n=1}^{\infty} |n \rangle \langle n | v \rangle .$$

We see $|v \rangle$ on both right and left sides of the equation, which allows us to conclude that the sum multiplying $|v \rangle$ on the right side is just the identity,

$$I \equiv \sum_{n=1}^{\infty} |n \rangle \langle n| .$$

In this representation, I is just an infinite matrix, with rows and columns indexed by the orthonormal basis we are using, with ones on the diagonal, and zeroes everywhere else. Inserting the identity in various expressions is quite useful. For example consider a scalar product, $\langle u | v \rangle$. Inserting the identity, we have

$$\langle u | v \rangle = \sum_{n=1}^{\infty} \langle u | n \rangle \langle n | v \rangle .$$

Let us study operators other than I in Dirac's notation. Suppose we have an operator T , and we apply it to a vector \mathbf{v} ,

$$\mathbf{u} = T\mathbf{v} .$$

How do we write this in Dirac's scheme? We can do it in two different ways;

$$|u \rangle = T|v \rangle ,$$

or

$$|u \rangle = |Tv \rangle .$$

It is difficult to draw a clear distinction between these two formulas. The first says, roughly, "write down $|v \rangle$, then apply T to it, and equate the result to $|u \rangle$." The second says, "somehow figure out $|Tv \rangle$, and equate the result to $|u \rangle$." We will regard them as equivalent for most situations. The one case where the second form is preferred is when dealing the differential operators and their adjoints. That said, let us examine the first form. There is surely nothing wrong with putting the identity in front of the right side,

$$|u \rangle = T|v \rangle = \sum_{n=1}^{\infty} |n \rangle \langle n | T | v \rangle .$$

There is also nothing wrong with putting the identity right after the T operator, so we have

$$|u\rangle = T|v\rangle = \sum_{n=1}^{\infty} |n\rangle \langle n| T \sum_{m=1}^{\infty} |m\rangle \langle m| v\rangle = \sum_{n,m=1}^{\infty} |n\rangle \langle n| T |m\rangle \langle m| v\rangle.$$

Comparing the first and last terms on the right side, we can write an expression for T ,

$$T = \sum_{n,m=1}^{\infty} |n\rangle \langle n| T |m\rangle \langle m|.$$

This is known as expanding an operator in terms of its matrix elements, and represents the operator T as an infinite matrix, with matrix elements $\langle n|T|m\rangle = (\mathbf{e}_n, T\mathbf{e}_m)$.

Comment 3 We now have several equations with infinite sums in them. In quantum physics these are very common, representing the fact that most systems have momenta and energies that range to infinite values, and coordinate distances that can become very small. The question arises as to how much effort should be expended on carefully analyzing questions of convergence of such sums. While asserting that it is never necessary to consider such question is too strong, most physicists believe that there is an effective separation of scales. That is, if we are interested in understanding behavior of a system at an energy scale of a say a few eV we do not expect this to be heavily influenced by phenomena at scales of keV , MeV , or GeV . It should be possible to smooth out or cutoff behavior at these high scales without affecting behavior at the eV scale. Likewise, at short distance, the Coulomb potential is about as singular as it gets, so handling very singular potentials is not necessary. So generally, we can avoid convergence issues if the physics is well-understood.

What we have done so far is merely establish a convenient notation. However we describe them, we have vectors with an infinite number of components, and they are acted upon by infinite matrices. The indices on the vectors and matrices are assumed to come from an unspecified orthonormal basis. The next move in developing Dirac's notation is to generalize to *continuous* indices. This was controversial when Dirac first wrote this down, but is widely accepted and used now. To get started, let us assume our Hilbert space comes not from expanding in an orthonormal basis, but from L^2 functions or wave-functions defined over an interval. We will start with functions of one variable, x , which ranges over $(-\infty, +\infty)$. In abstract notation we have vectors like Ψ and Φ , but the scalar product is now an integral,

$$(\Psi, \Phi) = \int_{-\infty}^{+\infty} \bar{\Psi}(x) \Phi(x) dx.$$

In Dirac's notation, we will have a bra, $\langle \Psi |$ and a ket, $|\Phi \rangle$, and the scalar product will be written

$$\langle \Psi | \Phi \rangle = \int_{-\infty}^{+\infty} dx \bar{\Psi}(x) \Phi(x).$$

Now compare this formula to the one for $\langle u | v \rangle$ with identity sandwiched in which we wrote out above. Dirac's next move was to interpret the integral over x as analogous to the sum on n in the case of an orthonormal basis. Doing this we write,

$$\langle \Psi | \Phi \rangle = \int_{-\infty}^{+\infty} dx \langle \Psi | x \rangle \langle x | \Phi \rangle,$$

and comparing left and right sides, the identity must be

$$I = \int_{-\infty}^{+\infty} dx |x \rangle \langle x|.$$

There are some new things here. First, the kets $|x \rangle$ and bras $\langle x|$ are certainly not vectors in our Hilbert space, like $|n \rangle$ and $\langle n|$. Since they refer to an orthonormal basis, we must have $\langle n | m \rangle = \delta_{nm}$, but the normalization of $|x \rangle$ and $\langle x|$ must involve Dirac δ functions,

$$\langle x | x' \rangle = \delta(x - x').$$

The other new thing is that we now interpret the wave function $\Psi(x)$ as an overlap or scalar product, $\langle x | \Psi \rangle$. If we were using an orthonormal basis, we would write the component of $|\Psi \rangle$ along $|n \rangle$ as $\langle n | \Psi \rangle$. In a qualitative fashion we can interpret $\langle x | \Psi \rangle$ in a similar way, as the "component" of $|\Psi \rangle$ "along" $|x \rangle$, where the quotes are to remind us that objects like $|x \rangle$ are not really vectors in our Hilbert space.

When Dirac first wrote all this down, it was treated with suspicion by many. Von Neumann, in particular has harsh words for Dirac's methods in his book. However, much time has passed. Dirac δ functions are very familiar, and Dirac's methods are no longer regarded as on a shaky footing.

NOTE: The methods work without trouble in *Cartesian* coordinates which range $(-\infty, +\infty)$. On intervals or in non-Cartesian coordinates care is needed!

Let us discuss operators in Dirac notation. The place to begin is with the position operator, X . In our original notation we can write

$$(\Psi, X\Phi) = \int_{-\infty}^{+\infty} dx x \bar{\Psi}(x) \Phi(x).$$

Going over to Dirac notation, this is

$$\langle \Psi | X \Phi \rangle = \int_{-\infty}^{+\infty} dx \langle \Psi | x \rangle x \langle x | \Phi \rangle.$$

Since x is just a real number, it can be placed anywhere in the right side integral. The reason for placing it in the middle will become more clear in the next step. We have

learned about sandwiching a factor of the identity in previous examples. Doing this on the left side of the equation we obtain,

$$\langle \Psi | X \Phi \rangle = \int_{-\infty}^{\infty} dx \langle \Psi | x \rangle \langle x | X \Phi \rangle = \int_{-\infty}^{\infty} dx \langle \Psi | x \rangle x \langle x | \Phi \rangle .$$

Note the operation of “pulling out” that we have used here. Returning to the case of a self-adjoint operator with a discrete spectrum for the moment, if $O|\lambda\rangle = \lambda|\lambda\rangle$, we certainly have $\langle \lambda | O = \langle \lambda | \lambda$, so we could write

$$\langle \lambda | O \Psi \rangle = \langle \lambda | O | \Psi \rangle = \lambda \langle \lambda | \Psi \rangle .$$

The operator O “sees” the eigenbra $\langle \lambda |$ to its left and evaluates as the eigenvalue λ . The same operation has been used in pulling X out of the matrix element and evaluating it as x .

We can introduce another important operator by Fourier transform. Let us start by writing the Fourier transform formulae for $\Psi(x)$,

$$\Psi(x) = \int_{-\infty}^{+\infty} dk \frac{1}{\sqrt{2\pi}} \exp(ikx) \tilde{\Psi}(k),$$

$$\tilde{\Psi}(k) = \int_{-\infty}^{+\infty} dx \frac{1}{\sqrt{2\pi}} \exp(-ikx) \Psi(x).$$

Now just as we have written $\Psi(x)$ in Dirac notation as $\langle x | \Psi \rangle$, it is natural to write $\tilde{\Psi}(k)$ as $\langle k | \Psi \rangle$. (We can now use Ψ without the tilde, since $\langle k | \Psi \rangle$ or $\langle x | \Psi \rangle$ tells us whether we are in x or k space. Now rewrite the equations above using our definitions so far. We have

$$\langle x | \Psi \rangle = \int_{-\infty}^{+\infty} dk \frac{1}{\sqrt{2\pi}} \exp(ikx) \langle k | \Psi \rangle .$$

$$\langle k | \Psi \rangle = \int_{-\infty}^{+\infty} dx \frac{1}{\sqrt{2\pi}} \exp(-ikx) \langle x | \Psi \rangle .$$

Comparing these two equations, both have the ket $|\Psi\rangle$ at the far right on both sides. The integrals must then be supplying a factor of the identity operator. We make the identifications

$$\langle x | k \rangle = \frac{1}{\sqrt{2\pi}} \exp(ikx),$$

$$\langle k | x \rangle = \frac{1}{\sqrt{2\pi}} \exp(-ikx),$$

and then our equations neatly become

$$\langle x | \Psi \rangle = \int_{-\infty}^{+\infty} dk \langle x | k \rangle \langle k | \Psi \rangle .$$

$$\langle k|\Psi \rangle = \int_{-\infty}^{+\infty} dx \langle k|x \rangle \langle x|\Psi \rangle,$$

As a consistency check, we should have

$$\langle k|k' \rangle = \int dx \langle k|x \rangle \langle x|k' \rangle = \delta(k - k').$$

But this is correct, since

$$\int dx \frac{1}{2\pi} \exp(i(k - k')x)$$

is a standard representation for $\delta(k - k')$.

The Commutator $KX - XK$ on $(-\infty, +\infty)$

As practice in using Dirac methods, let us compute the commutator $KX - XK$. We will take the matrix element of this between a bra $\langle k|$ and a ket $|x \rangle$. We will use the method of sandwiching the identity, and consider the terms separately. The first one is easy,

$$\langle k|KX|x \rangle = kx \langle k|x \rangle.$$

For the second one we have

$$\begin{aligned} \langle k|XK|x \rangle &= \int dx' dk' \langle k|X|x' \rangle \langle x'|k' \rangle \langle k'|K|x \rangle \\ &= \int dx' dk' \langle k|x' \rangle x' \langle x'|k' \rangle k' \langle k'|x \rangle. \end{aligned}$$

Now

$$\langle k|x' \rangle = \frac{1}{\sqrt{2\pi}} \exp(-ikx'),$$

so we can bring down the factor of x' by acting with $i\partial_k$. Focussing just on the terms involving the integral over x' , we can write

$$\int dx' \langle k|x' \rangle x' \langle x'|k' \rangle = i\partial_k \int dx' dk' \langle k|x' \rangle \langle x'|k' \rangle = i\partial_k \delta(k - k').$$

Putting this back in the main integral, we have

$$\begin{aligned} \langle k|XK|x \rangle &= i\partial_k \int dk' \delta(k - k') k' \langle k'|x \rangle \\ &= i\partial_k k \langle k|x \rangle = (i + kx) \langle k|x \rangle. \end{aligned}$$

Comparing with $\langle k|KX|x \rangle$, we have

$$\langle k|KX - XK|x \rangle = -i \langle k|x \rangle$$

Using this and again sandwiching the identity, we can write

$$\begin{aligned} \langle \Psi | (KX - XK) | \Phi \rangle &= \int dx dk \langle \Psi | k \rangle \langle k | (KX - XK) | x \rangle \langle x | \Phi \rangle \\ &= \int dx dk \langle \Psi | k \rangle \langle k | x \rangle (-i) \langle x | \Phi \rangle = -i \langle \Psi | \Phi \rangle . \end{aligned}$$

But since $|\Phi\rangle$ and $\langle\Psi|$ were arbitrary, the result must hold whatever they are, so we must have

$$KX - XK = -i,$$

that is, whenever $KX - XK$ acts on a vector actually in the Hilbert space, the result is $-i$ times that vector. The result has no direct physics content so far, it is merely a statement about the commutator of two operators in the Hilbert space $L^2(-\infty, +\infty)$.

Generalization to more than one dimension What we have just done generalizes easily out of one dimension. Let us sketch the treatment for three spacial dimensions. Writing the scalar product, we have

$$\langle \Psi | \Phi \rangle = \int_{-\infty}^{+\infty} d^3x \langle \Psi | \vec{x} \rangle \langle \vec{x} | \Phi \rangle ,$$

where we now have bras $\langle \vec{x} |$ and kets $|\vec{x}'\rangle$ which depend on a 3-vector x . The normalization is as expected,

$$\langle \vec{x} | \vec{x}' \rangle = \delta^3(\vec{x} - \vec{x}').$$

Fourier analysis is also the natural generalization of what we had previously. We write

$$\begin{aligned} \langle \vec{x} | \Psi \rangle &= \int_{-\infty}^{+\infty} d^3k \langle \vec{x} | \vec{k} \rangle \langle \vec{k} | \Psi \rangle . \\ \langle \vec{k} | \Psi \rangle &= \int_{-\infty}^{+\infty} d^3x \langle \vec{k} | \vec{x} \rangle \langle \vec{x} | \Psi \rangle , \end{aligned}$$

where

$$\langle \vec{x} | \vec{k} \rangle = \frac{1}{\sqrt{(2\pi)^3}} \exp(i\vec{k} \cdot \vec{x}).$$

Generalizing the discussion in one dimension, we may introduce operators \vec{X} , and \vec{K} , with commutation rule

$$K_j X_l - X_l K_j = i\delta_{jl}.$$

$\hbar\vec{k}$ as linear momentum How do we identify $\hbar\vec{k}$ as the linear momentum of a particle? We certainly **don't** want to say that if we multiply our commutator by \hbar , we get Heisenberg's commutator! That would be totally circular reasoning. A powerful route to identifying $\hbar\vec{k}$ as the linear momentum operator was taken by De Broglie (*ca* 1924). A rough version of this line of argument is as follows. Imagine that all we have is Planck and Einstein's work on black body radiation and the photoelectric effect. From this much we know energy and frequency are related as we discussed earlier.

$$\hbar\omega \longleftrightarrow E$$

De Broglie then notes that waves either classical or quantum will be of the form

$$\exp(-i\omega t + i\vec{k} \cdot \vec{x}).$$

We will certainly have a superposition of many frequencies, and we may have a term which is the complex conjugate of the above. Neither of these will affect De Broglie's argument. De Broglie observes that we always see the combination $\omega t - \vec{k} \cdot \vec{x}$ in the exponent. He then invokes the use of ideas from Einstein's special relativity. Namely ct, \vec{x} are the components of a 4-vector. Now we certainly know that special relativity applies to electromagnetic waves, so we assume the same holds for massive particles, very reasonable. For them, it is true that $E, \vec{p}c$ forms a 4-vector. Next, we use the frequency-energy connection to write

$$\omega t - \vec{k} \cdot \vec{x} = \frac{1}{\hbar c} \left(E(ct) - \hbar\vec{k}c \cdot \vec{x} \right).$$

This will be a scalar product of two 4-vectors and therefore relativistically invariant only if we identify $\hbar\vec{k}$ with the linear 3-momentum \vec{p} . This says that we should indeed think of $\hbar\vec{k}$ as the linear momentum. This is a perfectly wonderful and correct argument. It is also completely consistent with the then known fact (*ca* 1923) that the Compton effect demonstrated that $\hbar\vec{k}$ was the linear momentum for *photons*. So identifying $\hbar\vec{k}$ as linear momentum *in general* makes perfect sense. From this, since $\hbar\vec{k}$ is the eigenvalue of $\hbar\vec{K}$, we have that $\hbar\vec{K}$ is the linear momentum *operator*.

$$\vec{P} = \hbar\vec{K},$$

and of course we have De Broglie's famous equation, $p = h/\lambda$ as a by-product.

Linear Momentum and Translations Consider ordinary 3 dimensional space, and imagine we have a quantum state $|\Psi\rangle$ with wave function $\langle \vec{x}|\Psi\rangle$. Since this is a physical state, we must have $\langle \Psi|\Psi\rangle = 1$, or

$$1 = \int \langle \Psi|\vec{x}\rangle \langle \vec{x}|\Psi\rangle d^3\vec{x}. \quad (81)$$

For visualization purposes, take $\langle \Psi | \vec{x} \rangle \langle \vec{x} | \Psi \rangle$ to be a “lump” centered at $\vec{x} = \vec{b}$, although the argument to be given does not depend on the particular form of $\langle \Psi | \vec{x} \rangle \langle \vec{x} | \Psi \rangle$. Regardless of whether we have a lump or not, we can ask that

$$\int \langle \Psi | \vec{x} \rangle \vec{x} \langle \vec{x} | \Psi \rangle d^3 \vec{x} = \langle \vec{x} \rangle = \vec{b}. \quad (82)$$

Now suppose we translate this state by an amount \vec{a} . (Again, we are taking the “active” viewpoint where we translate the *state*, not the *coordinate system*.) Denote the translated state by $|\Psi'\rangle$. The state $|\Psi'\rangle$ is physically equivalent in every way to $|\Psi\rangle$. In particular, the process of translation must preserve probability amplitudes, i.e.

$$\langle \Psi_2 | \Psi_1 \rangle = \langle \Psi'_2 | \Psi'_1 \rangle, \quad (83)$$

where $|\Psi_i\rangle$ for $i = 1, 2$ are two original states, and $|\Psi'_i\rangle$ are their translated counterparts. Given that the $|\Psi_i\rangle$ are completely arbitrary states, they must be related by a unitary transformation, $U(\vec{a})$, so $|\Psi'_i\rangle = U(\vec{a})|\Psi_i\rangle$. Using unitarity, we have the desired result,

$$\langle \Psi'_2 | \Psi'_1 \rangle = \langle U\Psi_2 | U\Psi_1 \rangle = \langle U^\dagger U\Psi_2 | U\Psi_1 \rangle = \langle \Psi_2 | \Psi_1 \rangle. \quad (84)$$

Our goal is to determine $U(\vec{a})$. Consider $\langle \vec{x} | \Psi' \rangle$. How is this related to $\langle \vec{x} | \Psi \rangle$? Since we are considering a *rigid* translation, the answer is that we must have

$$\langle \vec{x} | \Psi' \rangle = \langle \vec{x} - \vec{a} | \Psi \rangle. \quad (85)$$

This can be seen to be correct by drawing a diagram, or more formally we can evaluate the expected value of \vec{x} in the two cases. We must have $\langle \vec{x} \rangle' = \langle \vec{x} \rangle + \vec{b}$. In words, the lump that used to be centered at \vec{b} is now centered at $\vec{a} + \vec{b}$. Writing this out, we have

$$\begin{aligned} \langle \vec{x} \rangle' &= \int \langle \Psi' | \vec{x} \rangle \vec{x} \langle \vec{x} | \Psi' \rangle d^3 \vec{x} & (86) \\ &= \int \langle \Psi | \vec{x} - \vec{a} \rangle \vec{x} \langle \vec{x} - \vec{a} | \Psi \rangle d^3 \vec{x} \\ &= \int \langle \Psi | \vec{x}' \rangle (\vec{x}' + \vec{a}) \langle \vec{x}' | \Psi \rangle d^3 \vec{x}' \\ &= \vec{b} + \vec{a}. & (87) \end{aligned}$$

To actually determine $U(\vec{a})$, we use Fourier expansions writing

$$\begin{aligned} \langle \vec{x} | \Psi' \rangle &= \int_{-\infty}^{+\infty} d^3 k \langle \vec{x} | \vec{k} \rangle \langle \vec{k} | \Psi' \rangle, & (88) \\ \langle \vec{x} - \vec{a} | \Psi \rangle &= \int_{-\infty}^{+\infty} d^3 k \langle \vec{x} - \vec{a} | \vec{k} \rangle \langle \vec{k} | \Psi \rangle \end{aligned}$$

Now

$$\langle \vec{x} - \vec{a} | \vec{k} \rangle = \frac{1}{\sqrt{(2\pi)^3}} \exp(i\vec{k} \cdot (\vec{x} - \vec{a})) = \langle \vec{x} | \vec{k} \rangle \exp(-i\vec{k} \cdot \vec{a}) \quad (89)$$

Using Eq.(89) in Eqs.(88) we have

$$\int_{-\infty}^{+\infty} d^3k \langle \vec{x} | \vec{k} \rangle \langle \vec{k} | \Psi' \rangle = \int_{-\infty}^{+\infty} d^3k \langle \vec{x} | \vec{k} \rangle \exp(-i\vec{k} \cdot \vec{a}) \langle \vec{k} | \Psi \rangle \quad (90)$$

This implies

$$\langle \vec{k} | \Psi' \rangle = \exp(-i\vec{k} \cdot \vec{a}) \langle \vec{k} | \Psi \rangle. \quad (91)$$

But

$$\exp(-i\vec{k} \cdot \vec{a}) \langle \vec{k} | \Psi \rangle = \langle \vec{k} | \exp(-i\vec{K} \cdot \vec{a}) | \Psi \rangle = \langle \vec{k} | \exp(-\frac{i}{\hbar} \vec{P} \cdot \vec{a}) | \Psi \rangle, \quad (92)$$

and since this is true for all $\langle \vec{k}$, we finally have

$$|\Psi' \rangle = \exp(-\frac{i}{\hbar} \vec{P} \cdot \vec{a}) |\Psi \rangle, \quad (93)$$

or

$$U(\vec{a}) = \exp(-\frac{i}{\hbar} \vec{P} \cdot \vec{a}) \quad (94)$$

Although we carried out the detailed discussion for a particle in 3 dimensions, the result is completely general and can be stated as follows:

The total linear momentum operator \vec{P} is the generator of spatial translations.

Time Evolution and Time Translation Let us start with the same system we had in the previous section. We are in ordinary three dimensional space. We have a quantum state $|\Psi \rangle$, with wave function $\langle \vec{x} | \Psi \rangle$. In order to discuss evolution or translation in time, we need to say more, so we add the information that the system is a free particle of mass m . How are we to let the system evolve in time? We follow Einstein, Planck, and de Broglie and start with the statement that a plane wave should vary as

$$\exp(-i\omega t + i\vec{k} \cdot \vec{x}) \quad (95)$$

with $\hbar\omega = E$, and $\hbar\vec{k} = \vec{p}$. For a free nonrelativistic particle we have $E = \vec{p} \cdot \vec{p}/2m$.

Comment 4 If we were starting with $E = \sqrt{mc^2 + (pc)^2}$, it might seem more correct to write $E = mc^2 + \vec{p} \cdot \vec{p}/2m$. However, as long as we are not creating or destroying particles, the factor mc^2 will be a simple phase present in all wave functions and hence can be transformed away.

Without reference to time, let us Fourier analyze $|\Psi\rangle$, just as we did before. We have

$$\langle \vec{x}|\Psi\rangle = \int d^3k \langle \vec{x}|\vec{k}\rangle \langle \vec{k}|\Psi\rangle \quad (96)$$

Rewriting this with explicit reference to momentum instead of wavevector we have

$$\langle \vec{x}|\Psi\rangle = \int d^3p \langle \vec{x}|\vec{p}\rangle \langle \vec{p}|\Psi\rangle. \quad (97)$$

We want $\langle \vec{x}|\vec{p}\rangle$ to satisfy

$$\int d^3p \langle \vec{x}'|\vec{p}\rangle \langle \vec{p}|\vec{x}\rangle = \delta^3(\vec{x}' - \vec{x}). \quad (98)$$

This will work if we set

$$\langle \vec{x}|\vec{p}\rangle = \frac{1}{(2\pi\hbar)^{3/2}} \exp\left(\frac{i}{\hbar}\vec{p}\cdot\vec{x}\right), \quad (99)$$

which is a simple rescaling of $\langle \vec{x}|\vec{k}\rangle$. So far, no reference to time at all. This wave function will evolve in time if we make the replacement

$$\exp\left(\frac{i}{\hbar}\vec{p}\cdot\vec{x}\right) \longrightarrow \exp\left(-\frac{i}{\hbar}E_p t\right) \exp\left(\frac{i}{\hbar}\vec{p}\cdot\vec{x}\right) \quad (100)$$

with $E_p = \vec{p}\cdot\vec{p}/2m$. Following what happened for the case of spacial translations, we expect this to be induced by a unitary operator acting on $|\Psi\rangle$. We write this as

$$|\Psi(t)\rangle = U(t)|\Psi\rangle. \quad (101)$$

(Why we don't use $|\Psi'\rangle$ instead of $|\Psi(t)\rangle$ will be explained shortly.) So using Eq.(100), Eq.(97) will become

$$\langle \vec{x}|\Psi(t)\rangle = \int d^3p \langle \vec{x}|\vec{p}\rangle \exp\left(-\frac{i}{\hbar}E_p t\right) \langle \vec{p}|\Psi\rangle. \quad (102)$$

From $E_p = \vec{p}\cdot\vec{p}/2m$, we can write

$$\exp\left(-\frac{i}{\hbar}E_p t\right) \langle \vec{p}|\Psi\rangle = \langle \vec{p}|\exp\left(-\frac{i}{\hbar}\frac{\vec{P}\cdot\vec{P}}{2m}t\right)|\Psi\rangle, \quad (103)$$

where \vec{P} is the momentum operator. But we can also momentum expand $|\Psi(t)\rangle$ directly, giving from Eq.(101)

$$\langle \vec{x}|\Psi(t)\rangle = \int d^3p \langle \vec{x}|\vec{p}\rangle \langle \vec{p}|U(t)|\Psi\rangle. \quad (104)$$

From Eqs.(102)-(104) we can finally identify

$$U(t) = \exp\left(-\frac{i}{\hbar}\frac{\vec{P}\cdot\vec{P}}{2m}t\right) = \exp\left(-\frac{i}{\hbar}Ht\right), \quad (105)$$

where

$$H = \frac{\vec{P} \cdot \vec{P}}{2m} \quad (106)$$

is the Hamiltonian of the system.

The wave function $\langle \vec{x} | \Psi(t) \rangle$ is easily shown to satisfy the free Schrödinger equation

$$\frac{\hbar}{i} \partial_t \langle \vec{x} | \Psi(t) \rangle = H \langle \vec{x} | \Psi(t) \rangle = -\frac{\hbar^2}{2m} \vec{\nabla} \cdot \vec{\nabla} \langle \vec{x} | \Psi(t) \rangle. \quad (107)$$

This route to the Schrödinger equation basically uses ideas of de Broglie, Planck, and Einstein. It can be written down without all the apparatus of Hilbert space and probabilities (which actually were introduced into quantum mechanics later on) by simply noting that whatever the meaning of $\langle \vec{x} | \Psi(t) \rangle$, there is no such thing as a perfect plane wave. Waves always come in packets, so an integral over wave numbers or momenta is required. From this point, it seems natural enough to include the case of a non-free particle by simply writing the full Hamiltonian operator

$$H = \frac{\vec{P} \cdot \vec{P}}{2m} + V(\vec{x}), \quad (108)$$

for a particle moving in a potential $V(\vec{x})$. The route to his equation that Schrödinger actually followed made use of the *action* of classical mechanics. This will be discussed after we cover the section on classical mechanics.

A final note. When we introduced the operator $U(t)$, we did not use $|\Psi'\rangle$ to denote the state upon which $U(t)$ acts. This is because we were discussing "evolving" not "translation in time." Evolving means we have a system which is left to (what else?) evolve over a time interval t . Translation in time would mean in effect picking up the system and moving it forward (or backward) in time (easier said than done to be sure.) Translation in time is hardly ever discussed, but for the record if we were to translate a system forward in time by an amount τ , this would be generated by the operator $\exp(i\tau H/\hbar)$, note the + sign in the exponent. This is the only exception to the statement that a symmetry operation is generated by $U = \exp(-i\xi O)$, where ξ is the *parameter* of the symmetry operation, and O is the *generator*.