

Chapter 1. Pure States, Kets and Linear Spaces with Scalar Product

§ 1 *Why bras and kets?* You are likely to have studied a version of quantum mechanics in which the wave function played a central role. This is not always the best choice. Dirac developed a more general theory, which incorporates the wave function naturally along with useful alternatives. The chapters that follow explain the physical assumptions and the mathematical apparatus of this formulation.

We introduce the mathematical concepts with a lack of rigor that would appall a mathematician but which is good enough for a practicing quantum mechanic. In chemical physics we can test the correctness of our results by comparing them with the results obtained by measurements. Mathematicians do not have this option. They must be very careful and thorough in their use of logic while we can afford to cut corners.

In the introductory lectures on Quantum Mechanics, you were taught that if you want to know the properties of a system, you write down the appropriate Schrödinger equation and solve it. This provides a set of allowed energies E_n and the corresponding energy eigenfunctions (or eigenstates) $\psi_{E_n}(x)$, $n = 0, 1, 2, \dots$. Here x is the position of the particle. To make life simple I pretend for now that the universe consists of one particle moving along a line. The generalization to many particles, moving in three dimensions, will turn out to be trivial and will be done later.

The wave function $\psi_{E_n}(x)$ is not a measurable quantity. To make contact with experiments, Schrödinger theory tells us that we must calculate certain integrals. For example, if the system is in the state $\psi_{E_n}(x)$, then the average potential energy is

$$\int_{-\infty}^{+\infty} \psi_{E_n}(x)^* V(x) \psi_{E_n}(x) dx \equiv \langle E_n | \hat{V} | E_n \rangle \quad (1)$$

where $V(x)$ is the potential energy given by classical mechanics. The right-hand side of Eq. 1 is a shorthand notation for the integral on the left. This notation permeates quantum mechanics. For example, the condition that the various energy eigenstates must be orthonormal contains the expression

$$\int_{-\infty}^{+\infty} \psi_{E_n}(x)^* \psi_{E_m}(x) dx \equiv \langle E_n | E_m \rangle \quad (2)$$

The symbol $\langle E_n | E_m \rangle$ in Eq. 2 is a convenient shorthand for the integral.

To calculate the rate of a photon absorption in spectroscopy, for a transition from a state of energy E_n to one of energy E_m , we evaluate

$$\int \psi_{E_n}(x)^* \hat{\mu} \psi_{E_m}(x) dx \equiv \langle E_n | \hat{\mu} | E_m \rangle \quad (3)$$

in which $\hat{\mu}$ is the dipole-moment operator. The shorthand notation $\langle E_n | \hat{\mu} | E_m \rangle$ appears again.

It is not important, at this point, that you remember these formulae or that you understand their meaning. We focus here on the shorthand notations in the right-hand side of these equations. Their widespread use in quantum mechanics suggests that perhaps we capture the essence of the theory by using the symbols $|E_n\rangle$ and $\langle E_m|$ and the physical quantities V and $\hat{\mu}$. The presence of x and of the integral may be of secondary importance. Probably Dirac set out to examine how much of the theory available at the time when he did his work can be captured by giving meaning to these symbols and developing rules for their manipulation.

Dirac confessed that he did physics by playing with equations. In this particular case, the game was fruitful and led him to a new notation, which produced a more general formulation of quantum mechanics. In this theory the left-hand sides of the equations given above are particular “representations” of the symbols in the right-hand side. Other representations, equivalent to the wave function $\psi_{E_n}(x)$, can be generated, and in some applications these other representations are more useful than the wave function. In addition, the new formulation can deal with the quantum mechanics of spin in a straightforward way. The formulation based on wave functions does not, because there is no wave function $\psi_a(x)$ that can represent the spin state.

§ 2 *Kets representing pure states.* We call ‘observable’ any quantity whose magnitude can be measured. Position, momentum, energy, angular momentum, spin are all observables. The wave function $\psi_{E_n}(x)$ is not.

From experiments, and from Schrödinger’s version of quantum theory, we know that some observables can only take *discrete* values. For example, the energy of the harmonic oscillator of frequency ω can only take one of the values

$$E_n = \hbar\omega \left(n + \frac{1}{2} \right), \quad n = 0, 1, \dots \quad (4)$$

The set of all values that an observable is allowed to take is called the *spectrum* of that observable. The energy spectrum of the harmonic oscillator consists of the discrete values given by Eq. 4.

If an experimentalist managed to prepare a system so that *we are certain* that a measurement of energy will give the value E_n , then Dirac would say that the system is in a *pure energy state* represented by the symbol $|E_n\rangle$, which he called a ‘ket’. When I say that the state of the system is $|E_n\rangle$, I am telling you that if you measure the energy of the system the result *is guaranteed* to be E_n .

Let us look at a few examples to see how this notation might be used. The vibrational motion of a diatomic molecule is described reasonably well by a harmonic oscillator. The molecules of gaseous H_2 at low temperature are in the ground vibrational state. This is true because, at low temperature, the kinetic energy of the molecules is small and collisions between them cannot provide the energy $\hbar\omega$ needed for lifting the vibrational energy to E_1 . The ket describing this pure state is $|E_0\rangle$ and a particle in this state is guaranteed to have the vibrational energy $E_0 = \hbar\omega/2$, where ω is the vibrational frequency. However, if we shine infrared light of frequency ω on the gas, some of the molecules will absorb a photon, their energy will increase to $E_1 = 3\hbar\omega/2$ and their state will be $|E_1\rangle$.

The energy of an electron in the hydrogen atom offers a more complex example. We know that the energy takes the discrete values (see H. Metiu, *Physical Chemistry: Quantum Mechanics*, page 300)

$$E_n = - \left(\frac{\mu e^4}{2(4\pi\epsilon_0)^2 \hbar^2} \right) \frac{1}{n^2}, \quad n = 1, 2, \dots \quad (\text{SI units}) \quad (5)$$

where μ , e , ϵ_0 , \hbar are constants whose meaning does not matter here. One would think that the state of a hydrogen atom whose energy is E_n should be denoted by $|E_n\rangle$. This is fine if all you care about is energy. However, a complete description of the state of the electron in H requires three quantum numbers. One of them, $n = 1, 2, \dots$, tells us (through Eq. 5) the energy of the electron. A second one, $\ell = 0, 1, \dots, n - 1$, tells us that the value of the angular momentum squared is $\hbar^2\ell(\ell + 1)$ (see Metiu, Chapter 18). An additional quantum number, $m = -\ell, -\ell + 1, \dots, +\ell$, tells us that the projection of the angular momentum vector on an axis is equal to $\hbar m$. A complete specification of a pure state of a hydrogen atom is given by the ket $|n, \ell, m\rangle$. If I tell you that the atom is in this state, you know that if you measure the energy of the atom you obtain E_n , if you measure the square of the angular momentum you obtain $\hbar^2\ell(\ell + 1)$, and if you measure the projection of the angular momentum on an axis, you obtain $\hbar m$. The ket

$|n, \ell, m\rangle$ conveys all this information. This ket is a pure state of energy and it is also a pure state of the square of the angular momentum and of its projection on the OZ axis.

This description of the hydrogen atom is limited. An experimentalist can make molecular beams of hydrogen atoms moving with a known momentum p . Since we know the value of p , we must add this information to the ket describing the state, which becomes $|p, n, \ell, m\rangle$. Knowing the momentum of the atom is important if we plan to do experiments in which the atom collides with a molecule and reacts. Slow atoms will have a different reaction probability than atoms moving fast. If we plan to analyze electron spin resonance (ESR) experiments on the hydrogen atom, then we need to know the state of the spin of the electron. This is characterized by a new quantum number m_s , which is equal to $-1/2$ or $1/2$ and tells us how the spin is “oriented”. An experimentalist who prepares a beam of H atoms having a given orientation of the spin and a known momentum will describe the state by giving the ket $|p, n, \ell, m, m_s\rangle$. The ket $|p, n, \ell, m, \frac{1}{2}\rangle$ tells me that the hydrogen atoms have “spin up”, and $|p, n, \ell, m, -\frac{1}{2}\rangle$ tells me they have “spin down”. If one examines NMR experiments then the pure state must give information about the nuclear spin as well.

The ket is a shorthand notation telling us what we know about the state of the system. It is important to be clear that pure states and kets are distinct concepts. A system is in a pure state when we know *for certain* the result of a specific measurement. The ket is a more general concept and it can be used to describe states that are not pure, meaning that we do not know for certain the result of a specific measurement. Also note that while energy plays an important role in quantum mechanics, the concept of pure state extends to any other observable, such as momentum, angular momentum, and spin.

§ 3 Continuous spectra. We have discussed examples in which the energy takes only discrete values. However, the hydrogen atom displays another interesting possibility. If we give the electron in the atom an energy larger than a certain amount E_I , the atom falls apart: the electron and the proton are no longer bound and they move away from each other. An atom can acquire such a large amount of energy when it absorbs an X-ray photon of energy $\hbar\omega$. A part of this energy is spent to overcome the ionization energy E_I . The rest is taken by the kinetic energy of the electron and proton flying

away from each other. Energy conservation tells me that

$$\hbar\omega - E_I = \frac{\hbar^2 k^2}{2\mu} \quad (6)$$

where $\hbar^2 k^2/2\mu$ is the kinetic energy after ionization ($\hbar k$ is the relative momentum of the two particles and μ is the reduced mass). If you do not remember or you do not understand this formula do not panic; just trust that this equation is correct.

If I use a synchrotron to perform this experiment, I can change $\hbar\omega$ by an infinitesimal amount. Since E_I is constant, for a given atom, Eq. 6 tells me that if I change $\hbar\omega$ continuously then I change the relative kinetic energy $\hbar^2 k^2/2\mu$ of the electron-proton system continuously. When the energy of the atom exceeds E_I , its energy spectrum is *no longer discrete*. Such experiments tell us that the energy of the electron-proton system is discrete when it is smaller than E_I , but it is *continuous* when it exceeds E_I . The ket describing the pure energy state of the *dissociated* hydrogen atom is $|k\rangle$ where k specifies the relative kinetic energy of the two particles (i.e. electron and proton). The point I am making is that in this symbol k takes continuous values, which means that I can change the energy by an infinitesimal amount.

If the continuous spectrum is a novelty to you, it is because the books on introductory quantum mechanics prefer to ignore it. The particle in a box and the harmonic oscillator do not have a continuous spectrum, because they are oversimplified models of real systems. The hydrogen atom has a continuous spectrum, which is rarely mentioned, like an embarrassing uncle in a proper family. The continuous spectrum cannot, however, be ignored, because it is indispensable for describing molecules that break up into fragments and for describing collisions between two molecules. It is an experimental fact that the energy of two particles flying apart or coming together can be changed by an infinitesimal amount. The energy spectrum in these processes is continuous. Since this energy is the kinetic energy (we neglect the internal energy of the particles, which plays no role here) $p^2/2m$, this means that momentum has a continuous spectrum. In addition, it is an experimental fact that position has a continuous spectrum; there is no law of nature that says that I cannot displace a particle by an arbitrarily small amount.

§ 4 We can generalize these examples as follows. If A is an observable, then

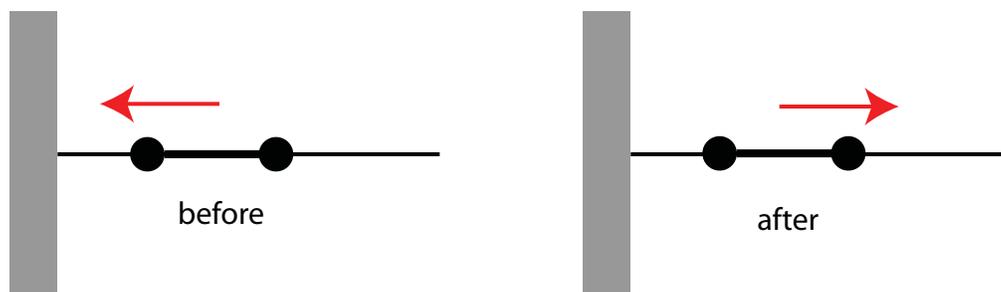


Figure 1: Collision and rebound

measurements of the value of A can yield the discrete values

$$a_1, a_2, \dots, a_i, \dots$$

and the continuous values

$$\alpha \in [m, n]$$

where m and n are real numbers. The set of all values that A can take is called the *spectrum* of A . Some observables have only a discrete spectrum (e.g. the energy of a harmonic oscillator), some have only a continuous spectrum (e.g. position), most have both (e.g. the energy of the hydrogen atom, or that of any system that can dissociate or be ionized).

§ 5 *Coherent superposition of pure states.* In many situations created by experiments, the result of a measurement *is not certain* and therefore the system is *not in a pure state*. We can see how this might happen by examining the following experiment. Imagine that a diatomic molecule hits a hard wall with high velocity. To make life simple, I assume that the molecule is prevented from rotating and moves on a line, as shown in Fig. 1. Before the collision the vibrational energy is $E_0 = \hbar\omega/2$ and the kinetic energy is $p^2/2m$ (p is the momentum of the molecule and m is its mass). I will call this the initial state and I will assume that the experimentalist can control it. The ket describing this pure state is $|E_0, p\rangle$. I would like to know how to write the ket that describes the state *after the collision* (the final state).

The collision can increase the vibrational energy of the molecule by using a fraction of the kinetic energy $p^2/2m$. The vibrational energy is discrete (see Eq. 4) and to change it from E_0 to E_1 , the kinetic energy $p^2/2m$ must

exceed the difference $E_1 - E_0 = \hbar\omega$. If $p^2/2m < \hbar\omega$, the result of the experiment is simple: the vibrational energy after the collision is E_0 and the only effect of the collision is to change the momentum from p to $-p$. I took the momentum axis so that the momentum of a molecule moving toward the wall is positive ($p > 0$). The post-collision state is described by the ket $|E_0, -p\rangle$. This is a pure state of energy, because energy conservation tells me that it is guaranteed that the energy of the state $|E_0, -p\rangle$ is $\frac{p^2}{2m} + E_0$.

Consider now the case when the initial kinetic energy is sufficiently high to excite the molecule to E_1 and to E_2 . Energy conservation says that the kinetic energy after the collision is

$$\frac{p_n^2}{2m} = \frac{p^2}{2m} + E_0 - E_n, \quad n = 0, 1, 2 \quad (7)$$

Here E_n is the vibrational energy after the collision and $p_n^2/2m$ is the kinetic energy of the molecule after the collision. Note that the molecule has the kinetic energy $p_n^2/2m$ if and only if the vibrational energy is E_n .

Because $p_n^2/2m$ must be positive, Eq. 7 limits the possible vibrational energy E_n after the collision. In the case we consider, the final state may have the vibrational energy E_0 and the momentum $-p$, and the ket describing it is $|E_0, -p\rangle$; or the vibrational energy could be E_1 and the momentum $p_1 = -\sqrt{2m((p^2/2m) + E_0 - E_1)}$, in which case the ket is $|E_1, p_1\rangle$. When the incident kinetic energy is large enough to excite the molecule in the vibrational state of energy E_2 , the final state is $|E_2, p_2\rangle$ with $p_2 = -\sqrt{2m((p^2/2m) + E_0 - E_2)}$ and $E_2 = \frac{3}{2}\hbar\omega$.

If the initial kinetic energy $p^2/2m$ is sufficiently high, energy conservation no longer pins down the final state. In the example given above, the state after the collision could be $|E_0, -p\rangle$, or $|E_1, p_1\rangle$, or $|E_2, p_2\rangle$. Which one is it? None of them!

Quantum mechanics tells us that the state $|\psi\rangle$ after the collision is described by a ket of the form:

$$|\psi\rangle = c_0|E_0, -p\rangle + c_1|E_1, p_1\rangle + c_2|E_2, p_2\rangle \quad (8)$$

Here c_0, c_1, c_2 are complex numbers. The expression in the right-hand side is called a *linear combination* or a *linear superposition* or a *coherent superposition* of the pure states $|E_0, -p\rangle$, $|E_1, p_1\rangle$, and $|E_2, p_2\rangle$.

§ 6 Coherent superposition of states. The statement contained in Eq. 8 is a central principle of quantum mechanics. The experiment we have examined

has several possible final states: $|E_0, -p\rangle$, $|E_1, p_1\rangle$, or $|E_2, p_2\rangle$. Since we had no way of knowing that one will be preferred over the others, the ket describing the final state of the system (after the collision) is linear combination (or a coherent superposition) of all the possible states. By possible we mean all states that would not violate energy conservation or any other law of physics.

We generalize this observation as follows: if an experiment is such that the observable A can take one of the values a_1, a_2, \dots, a_n , the state of the system is described by the ket

$$|\psi\rangle = c_1|a_1\rangle + c_2|a_2\rangle + \dots + c_n|a_n\rangle \quad (9)$$

Here $|a_1\rangle, \dots, |a_n\rangle$ are the pure states corresponding to the values a_1, \dots, a_n , and c_1, \dots, c_n are complex numbers.

We can consider the statement above to be a postulate, called the *principle of linear superposition*. However, if we postulate the validity of the time-dependent Schrödinger equation and solve it for the collision discussed above, we will find that the state $|\psi\rangle$ after the collision must have the form given by Eq. 9. If we postulate the Schrödinger equation, then the linear superposition principle is a consequence of the linearity of the equation. If we postulate the principle of linear superposition, then it follows that the Schrödinger equation must be linear. I prefer to consider that the principle of linear superposition is a basic postulate of quantum mechanics.

Exercise 1 Consider an experiment in which the diatomic molecule is very rigid and does not have enough kinetic energy for the collision to increase its vibrational energy. However, now we allow the diatomic to rotate and therefore the collision can change the rotational energy. Analyze this experiment in the manner done in §5 and §6.

Exercise 2 Use the line of reasoning developed in §5 and §6 to describe the collision of a diatomic molecule with a flat, rigid wall. The molecule moves in three dimensions. The collision can excite the rotation, the vibration, and the electrons, and there is enough energy to break the molecule.

§ 7 *Linear spaces.* Equation 9 tells us that to describe quantum phenomena in terms of kets, we must know how to multiply a ket $|a_n\rangle$ by a complex number c_n to obtain another ket $c_n|a_n\rangle$, and how to add two kets $|a_n\rangle$ and $|a_m\rangle$ to obtain a new ket denoted by $|a_n\rangle + |a_m\rangle$ and called the sum of the two kets.

A *linear space* is a set of objects with the following properties¹: (1) We have defined an addition operation which is such that the sum of any two objects in the set gives an object in the same set. (2) We have defined a multiplication between any object of the set and any complex number, which yields another object in the set. Sometimes a linear space is called a *vector space*. Obviously the kets form a linear space since the sum of any two kets is a ket and the product of a complex number with a ket is a ket.

Keep in mind that only kets of the same kind form a linear space. For example, the kets describing the energy states of the hydrogen atom form a linear space, and the kets describing the energy states of a harmonic oscillator form *another* linear space. The set consisting of the pure energy states of the hydrogen atom *and* the pure energy states of a harmonic oscillator do not form a linear space because adding a hydrogen-atom ket to a harmonic-oscillator ket is devoid of any physical meaning.

While we recognize that the kets describing harmonic oscillators form a different linear space than do the kets describing a hydrogen atom, we must also recognize that any two linear spaces have common abstract properties. For the remainder of this chapter, we study mostly these common properties. The linear space of hydrogen-atom kets and the linear space of harmonic-oscillator kets are concrete representations of the abstract concept of linear space just as three monkeys and three apples are concrete realizations of the abstract number 3.

§ 8 *Examples of linear spaces.* There are many examples of linear spaces, some of which are very familiar to you, such as \mathbb{R}^3 , the set of three-dimensional vectors. Others, called in what follows L^2 and ℓ^2 , appear in quantum mechanics and many other areas of physics. Different methods for calculating the physical properties of a system in quantum mechanics will use objects belonging to different linear spaces.

The space \mathbb{R}^3 of ordinary vectors. When you studied mechanics or electricity you worked with ordinary three-dimensional vectors. Sometimes you rep-

¹Mathematicians give a more precise and more general definition; the one given here satisfies our needs.

resented them by arrows that showed the direction of the vector and its magnitude. Most often you picked a coordinate system and showed that you can describe completely any vector by three real numbers (the *components* of the vector). Thus a vector may be represented by the symbol $\{x_1, x_2, x_3\}$ where the symbols inside the braces are real numbers. Note that the symbol $\{x_1, x_2, x_3\}$ is ordered: $\{x_1, x_2, x_3\}$ and $\{x_2, x_1, x_3\}$ are different vectors (if $x_1 \neq x_2$). The two representations of a given vector (arrow or three numbers) are equivalent: if you know the three numbers you can calculate all the properties of the arrow, and if you know the arrow you can calculate the three components of the vector.

Any two vectors in \mathbb{R}^3 can be added according to the following rule: the sum of two vectors \vec{x} and \vec{y} , having the coordinates $\{x_1, x_2, x_3\}$ and $\{y_1, y_2, y_3\}$, respectively, is the vector having the coordinates $\{x_1 + y_1, x_2 + y_2, x_3 + y_3\}$. We can write that

$$\{x_1, x_2, x_3\} + \{y_1, y_2, y_3\} \equiv \{x_1 + y_1, x_2 + y_2, x_3 + y_3\} \quad (10)$$

A mathematician could have invented symbols like $\{x_1, x_2, x_3\}$ and defined arbitrary rules for their addition (such as Eq. 10), but this is not how the vectors came about. This addition rule was derived from experiments in which two forces (two vectors) were made to act on the same point of a body and then the net force was measured. This net force was called the sum of the two forces and happens to be obtained from the individual forces through the addition rule introduced in Eq. 10.

We can also define the multiplication of a vector $\{x_1, x_2, x_3\}$ with a real number α through the rule

$$\alpha\{x_1, x_2, x_3\} \equiv \{\alpha x_1, \alpha x_2, \alpha x_3\} \quad (11)$$

This also has a physical meaning. Imagine that a person pulls on a rope to move a boat. The force of this pull is a vector with the components $\{x_1, x_2, x_3\}$. If he starts pulling twice as hard, in the same direction, the force will be $\{2x_1, 2x_2, 2x_3\}$.

The set of all objects of the form $\{x_1, x_2, x_3\}$, with the definitions Eqs. 10 and 11 for addition and multiplication with a (real) number, form a linear space called \mathbb{R}^3 . Here \mathbb{R} tells us that the coordinates of the vector are real numbers and 3 tells us that three coordinates are needed for defining a vector.

Exercise 3 What would be a good definition for the space \mathbb{R}^2 ? How about the space \mathbb{R}^n where n is a finite integer?

The linear space L^2 . The space L^2 contains all functions $f(x)$ that have the property

$$\int_{-\infty}^{+\infty} f(x)^* f(x) dx < \infty. \quad (12)$$

Here x is a real variable and $f(x)$ is a function that can take complex values. The functions $\sin x$ and e^{ikx} do not belong to this space, because the integrals

$$\int_{-\infty}^{+\infty} \sin^2 x dx$$

and

$$\int_{-\infty}^{+\infty} e^{-ikx} e^{ikx} dx$$

do not satisfy Eq. 12. On the other hand, $f(x) = e^{-ax^2} e^{ikx}$ does belong to L^2 as long as the real part of the complex number a is positive and k is a real number, since in that case the integral is finite.

The set L^2 is a linear space. Indeed, if two functions $f(x)$ and $\phi(x)$ satisfy Eq. 12 then so does their sum and therefore the sum is also an element of L^2 . If $f(x)$ is in L^2 then $\alpha f(x)$ is also in L^2 for any finite, complex number α . This space plays an fundamental role in the Schrödinger version of quantum mechanics. The wave functions of any bound system (atoms, molecules, harmonic oscillator, etc.) belong to L^2 .

Exercise 4 Show that the lowest three energy eigenfunctions of a harmonic oscillator satisfy the condition Eq. 12. Explain why the physics of the system demands that such a condition must be fulfilled.

It is easy to generalize the space L^2 to functions of more than one variable. For example, the complex functions of three real variables satisfying the condition

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y, z)^* f(x, y, z) dx dy dz < \infty \quad (13)$$

form a space denoted by $L^2(\mathbb{R}^3)$. The notation indicates that the space contains functions of three real variables. With this notation the space of one-dimensional functions satisfying Eq. 12 is denoted by $L^2(\mathbb{R})$.

The space ℓ^2 . Another linear space of interest to quantum mechanics consists of the set of all infinite, ordered sequences $\{a_1, a_2, \dots\} \equiv \{a_n\}_{n=1}^{\infty}$ of complex numbers that satisfy

$$\sum_{n=1}^{\infty} a_n^* a_n < \infty \quad (14)$$

Here a_n^* denotes the complex conjugate of a_n .

If $a \equiv \{a_1, a_2, \dots\}$ and $b \equiv \{b_1, b_2, \dots\}$ then we define their sum to be

$$a + b \equiv \{a_1 + b_1, a_2 + b_2, \dots\} \quad (15)$$

and the product of $a \equiv \{a_1, a_2, \dots\}$ with a complex number α to be

$$\alpha a = \{\alpha a_1, \alpha a_2, \dots\} \quad (16)$$

It should be obvious that if $a \equiv \{a_n\}_{n=1}^{\infty}$ and $b \equiv \{b_n\}_{n=1}^{\infty}$ satisfy Eq. 14, then $a + b$ and αa also satisfy it. These types of sequences therefore form a linear space.

§ 9 Heisenberg formulated quantum mechanics in terms of elements of ℓ^2 . For a while physicists had two theories, that of Schrödinger and that of Heisenberg, Jordan, and Born. They appeared completely different but gave the same results when used to calculate measurable quantities (such as the energy of an oscillator or of a hydrogen atom). Schrödinger showed that the two formulations were mathematically equivalent. Later Dirac reformulated the theory, by introducing bras, kets, and operators. His formulation is more general and will be explained in these lecture notes. Dirac, Heisenberg and Jordan were about 25 years old when they did this work. Schrödinger and Born were in their forties.

§ 10 *Scalar product: the definition.* Abstract symbols, such as kets, cannot take us too far in physics. Measurements produce numbers and the theory needs an operation that associates numerical values with the kets representing the state of the system. This is done through an operation called *scalar product*.

The scalar product is a “function” that takes as an argument two kets $|a\rangle$ and $|b\rangle$ and produces a complex number denoted by

$$\langle a | b \rangle \quad (17)$$

This notation may seem a bit strange. You are used to notation like $f(x, y)$, where the arguments x and y are real or complex numbers and where f is a short name for the operations performed on x and y to produce the number denoted by $f(x, y)$. The scalar product is a function that takes as arguments two elements $|x\rangle$ and $|y\rangle$ of the linear space and gives as a result a complex number. The traditional notation for this should have been $s(|x\rangle, |y\rangle)$ where s would be a symbol for the rules telling us how to combine the two kets to generate a complex number. It turns out that $\langle a | b \rangle$ is a more convenient notation and it is universally used in physics and mathematics (some mathematicians use $(a|b)$ instead of $\langle a | b \rangle$).

The scalar product is *defined by* the following properties.

1. If $|a\rangle$ and $|b\rangle$ are arbitrary kets then $\langle a | b \rangle$ is a complex number. The set of complex numbers includes the real numbers (which are complex numbers whose imaginary part is zero). Therefore this requirement does not exclude real values for $\langle a | b \rangle$.
2. For an arbitrary ket $|a\rangle$, $\langle a | a \rangle$ is a real number and

$$\langle a | a \rangle \geq 0 \quad (18)$$

with equality possible only if $|a\rangle = 0$. $|a\rangle = 0$ means that $|a\rangle + |b\rangle = |b\rangle$ for all kets $|b\rangle$. In quantum mechanics the state $|a\rangle = 0$ is physically impossible.

3. If $|c\rangle = |a\rangle + |b\rangle$ and $|d\rangle$ is an arbitrary ket then

$$\langle d | c \rangle \equiv \langle d | a + b \rangle = \langle d | a \rangle + \langle d | b \rangle \quad (19)$$

4. If $|\eta\rangle = \alpha|a\rangle$ where α is an arbitrary complex number then

$$\langle b | \eta \rangle = \alpha \langle b | a \rangle \quad (20)$$

5. Complex conjugation² reverses the order of the kets in the scalar product:

$$\langle a | b \rangle = \langle b | a \rangle^* \quad (21)$$

§ 11 *Scalar product: other properties.* This five-point, abstract definition does not tell us how to calculate a scalar product (you will learn how to do that later). Different linear spaces (e.g. \mathbb{R}^3 , \mathbb{R}^n , L^2 , ℓ^2) use different definitions for the scalar product, dictated by the nature of the physical applications for which the space is defined. However, all these different definitions must satisfy the five conditions given above. Because of this, the scalar products in different linear spaces share a number of general properties, which we derive below. These properties and the manipulations used to derive them are used very frequently in quantum mechanics, and you must become very familiar with them if you want to master the subject.

In what follows $|a\rangle$, $|b\rangle$ and $|d\rangle$ are arbitrary kets and α is an arbitrary complex number.

- (a) If $|\eta\rangle = \alpha|a\rangle$ then

$$\langle \eta | b \rangle = \alpha^* \langle a | b \rangle \quad (22)$$

Proof: From defining properties 4 and 5 it follows that $\langle \eta | b \rangle = \langle b | \eta \rangle^* = (\alpha \langle b | a \rangle)^* = \alpha^* \langle b | a \rangle^* = \alpha^* \langle a | b \rangle$.

- (b) If $|\eta\rangle = \alpha|a\rangle + |b\rangle$ then

$$\langle \eta | c \rangle = \alpha^* \langle a | c \rangle + \langle b | c \rangle \quad (23)$$

Proof: $\langle \eta | c \rangle \stackrel{(5)}{=} \langle c | \eta \rangle^* \stackrel{(b)}{=} (\langle c | \alpha a \rangle + \langle c | b \rangle)^* \stackrel{(4)}{=} (\alpha \langle c | a \rangle + \langle c | b \rangle)^* = \alpha^* \langle c | a \rangle^* + \langle c | b \rangle^* \stackrel{(5)}{=} \alpha^* \langle a | c \rangle + \langle b | c \rangle$. A shorthand notation for this property is

$$\langle (\alpha a + b) | c \rangle = \alpha^* \langle a | c \rangle + \langle b | c \rangle \quad (24)$$

- (c) If $|\eta\rangle = |a\rangle + \lambda|b\rangle$ where λ is a complex number, then

$$\langle \eta | \eta \rangle = \langle a | a \rangle + \lambda^* \langle b | a \rangle + \lambda \langle a | b \rangle + \lambda^* \lambda \langle b | b \rangle \quad (25)$$

²The complex conjugate of a function or a number is obtained by replacing $i = \sqrt{-1}$ with $-i$; for example, $(e^{ix})^* = e^{-ix}$ if x is a real number, or $(e^{iz})^* = e^{-iz^*} = e^{-i(x-iy)}$ if $z = x + iy$ and x and y are real. Note that when $z = x + iy$ with x and y real, $z^*z = x^2 + y^2$ is always a real number.

This follows from (a) and (b). A short hand notation for this property is

$$\langle a + \lambda b | a + \lambda b \rangle = \langle a | a \rangle + \lambda^* \langle b | a \rangle + \lambda \langle a | b \rangle + \lambda^* \lambda \langle b | b \rangle \quad (26)$$

Exercise 5 Show that if $|a\rangle = 0$ then $\langle a | b \rangle = 0$ for every ket $|b\rangle$. (Hint: use $|c\rangle + |0\rangle = |c\rangle$.)

§ 12 Scalar product: examples.

The scalar product in \mathbb{R}^3 . The simplest and most familiar linear space we encountered is the set \mathbb{R}^3 of the three-dimensional vectors having real components. One of the operations involving two vectors $\vec{v} = \{v_1, v_2, v_3\}$ and $\vec{u} = \{u_1, u_2, u_3\}$ is the dot product:

$$\vec{v} \cdot \vec{u} \equiv \sum_{i=1}^3 v_i u_i \quad (27)$$

You can easily verify that the dot product has all the properties required in §10 and it is therefore a scalar product. A person accustomed to the notation of quantum mechanics will be tempted to use the notation $\langle \vec{v} | \vec{u} \rangle$ for $\vec{v} \cdot \vec{u}$.

Exercise 6 Consider the linear space \mathbb{C}^3 consisting of three-dimensional vectors whose components are complex numbers. How would the definition of the scalar product in this space differ from that in \mathbb{R}^3 ? The definition must satisfy all the requirements of §10.

The scalar product in L^2 . Recall that L^2 is the linear space consisting of all functions $f(x)$ satisfying the condition

$$\int_{-\infty}^{+\infty} f(x)^* f(x) dx < \infty \quad (28)$$

The argument x is a real number and $f(x)$ has complex values. You can easily verify that the integral in the right-hand side of

$$\langle f | \phi \rangle \equiv \int_{-\infty}^{+\infty} f(x)^* \phi(x) dx \quad (29)$$

has all the properties required in §10, justifying the notation used in the left-hand side. This scalar product is used in Schrödinger's version of quantum mechanics.

The scalar product in ℓ^2 . The set of all sequences $\{a_1, a_2, \dots\}$ satisfying

$$\sum_{n=1}^{\infty} a_n^* a_n < \infty \tag{30}$$

forms the linear space ℓ^2 . Show that if $A = \{a_1, a_2, \dots\}$ and $B = \{b_1, b_2, \dots\}$ are two sequences in ℓ^2 then

$$\langle A | B \rangle \equiv \sum_{n=1}^{\infty} a_n^* b_n \tag{31}$$

is their scalar product. This scalar product is used in Heisenberg's version of quantum mechanics.

§ 13 Generalizations involving the scalar product. Many concepts and some of the nomenclature used for the vectors of \mathbb{R}^3 are generalized for use in other linear spaces. In vector algebra

$$\sqrt{\langle \vec{v} | \vec{v} \rangle} \equiv \sqrt{\vec{v} \cdot \vec{v}} = \sqrt{v_1^2 + v_2^2 + v_3^2}$$

is the length of the vector \vec{v} . In general,

$$\|a\| \equiv \sqrt{\langle a | a \rangle} \tag{32}$$

is thought of as the 'length' of $|a\rangle$ and it is called the *norm* of $|a\rangle$. Using this nomenclature for the space L^2 , we call

$$\|f\| \equiv \sqrt{\int_{-\infty}^{+\infty} f(x)^* f(x) dx} \tag{33}$$

the length (or the norm) of f .

Exercise 7 Show that for kets $|a\rangle$ and $|b\rangle = \alpha|a\rangle$, where α is a complex number, $\|b\| = |\alpha| \|a\|$ (where $|\alpha| \equiv \sqrt{\alpha^* \alpha}$).

Normalized kets. A vector \vec{v} in \mathbb{R}^3 is said to be of unit length if

$$\vec{v} \cdot \vec{v} = 1 \quad (34)$$

We extend this definition to all linear spaces and say that a ket having the property

$$\langle a | a \rangle = 1 \quad (35)$$

is *normalized*: its ‘length’ is equal to 1.

Orthogonal kets. If two non-zero vectors \vec{v} and \vec{u} in \mathbb{R}^3 satisfy

$$\vec{v} \cdot \vec{u} \equiv \langle \vec{v} | \vec{u} \rangle = 0 \quad (36)$$

then they are perpendicular (orthogonal) to each other. We extend this definition to say that if two kets $|a\rangle$ and $|b\rangle$ in a linear space satisfy

$$\langle a | b \rangle = 0 \quad (37)$$

and neither is equal to $|0\rangle$, they are *orthogonal* to each other.

Schwartz’s inequality. Any two vectors \vec{v} and \vec{u} in \mathbb{R}^3 satisfy the equation

$$\vec{v} \cdot \vec{u} = (\text{length of } \vec{v}) \times (\text{length of } \vec{u}) \times \cos \theta \quad (38)$$

where θ is the angle between the vectors. Also, the length of \vec{v} is equal to $\sqrt{\vec{v} \cdot \vec{v}} = \|\vec{v}\|$. Using this relationship we can rewrite Eq. 38 as

$$\langle \vec{v} | \vec{u} \rangle = \|\vec{v}\| \|\vec{u}\| \cos \theta \quad (39)$$

Because $|\cos \theta| \leq 1$, Eq. 39 implies that

$$|\langle \vec{v} | \vec{u} \rangle| \leq \|\vec{v}\| \|\vec{u}\| \quad (40)$$

To generalize this result, valid in \mathbb{R}^3 , we guess that for any two kets in a linear space we must have

$$|\langle a | b \rangle| \leq \|a\| \|b\| \quad (41)$$

Equality takes place only when $|a\rangle = \alpha|b\rangle$ where α is a complex number. This is called the *Schwarz inequality* and it is very useful in quantum mechanics.

Writing down equations by analogy does not prove that they are correct. A proof of the Schwarz inequality based on the abstract properties of the scalar product is given in Appendix 1.

§ 14 *Summary.* Stimulated by a certain notation in Schrödinger's version of quantum mechanics, we have introduced mathematical objects called kets, which represent pure states and coherent states of a quantum system. We assumed that the kets form a linear space, which means that if we add any two kets we obtain a ket and if we multiply a ket by a complex number we obtain a ket.

We have also postulated that in this linear space we can define a scalar product. You have seen several examples of mathematical objects that have the properties required of the kets. Later I will show that there is a one-to-one correspondence between the elements of L^2 , ℓ^2 , and the abstract space of the kets.

A lot of work remains to be done. We need to connect kets to physical phenomena and we need to show that this notation contains Schrödinger's and Heisenberg's versions of the theory. We also must show that the Dirac version of the theory is more powerful and more convenient than the other versions. Otherwise there is no point in developing this new notation.

Appendix 1: The Schwarz Inequality

We want to prove that if $|a\rangle$ and $|b\rangle$ are nonzero elements in a linear space in which a scalar product $\langle a | b \rangle$ has been defined, then

$$|\langle a | b \rangle| \leq \|a\| \|b\| \quad (42)$$

Moreover, equality holds exactly when $|a\rangle = \alpha|b\rangle$, where α is some complex number.

Let

$$|c\rangle = |a\rangle + \lambda^*|b\rangle \quad (43)$$

where λ is an undetermined complex number. Define $r(\lambda)$ through:

$$r(\lambda) \equiv \langle c | c \rangle = \langle a | a \rangle + \lambda^* \langle a | b \rangle + \lambda \langle b | a \rangle + \lambda^* \lambda \langle b | b \rangle \quad (44)$$

From the properties of the scalar product (§10) we know that $\langle c | c \rangle$ is real and nonnegative, and it is equal to zero only if $|c\rangle = |0\rangle$. Using this and eq. 44 it follows that $r(\lambda)$ is real,

$$r(\lambda) \geq 0. \quad (45)$$

and $r(\lambda) = 0$ only when $|c\rangle \equiv |a\rangle + \lambda^*|b\rangle = 0$.

The inequality in Eq. 45 is valid for any value of λ , including

$$\lambda = \frac{-\langle b|a\rangle^*}{\langle b|b\rangle} = \frac{-\langle a|b\rangle}{\langle b|b\rangle} \quad (46)$$

Inserting this value in Eq. 44 gives

$$\begin{aligned} & \langle a|a\rangle + \left(\frac{-\langle b|a\rangle}{\langle b|b\rangle}\right) \langle a|b\rangle + \left(\frac{-\langle a|b\rangle}{\langle b|b\rangle}\right) \langle b|a\rangle + \left(\frac{-\langle b|a\rangle}{\langle b|b\rangle}\right) \left(\frac{-\langle a|b\rangle}{\langle b|b\rangle}\right) \langle b|b\rangle \\ &= \langle a|a\rangle - \frac{\langle b|a\rangle\langle a|b\rangle}{\langle b|b\rangle} \geq 0 \end{aligned} \quad (47)$$

which implies

$$\langle a|a\rangle \geq \frac{\langle b|a\rangle\langle a|b\rangle}{\langle b|b\rangle} \quad (48)$$

Using the notation $\langle a|a\rangle = \|a\|^2$ and

$$\langle b|a\rangle\langle a|b\rangle = \langle a|b\rangle^* \langle a|b\rangle = |\langle a|b\rangle|^2$$

we can rewrite Eq. 48 as

$$|\langle a|b\rangle| \leq \|a\| \|b\|$$

which is what we set out to prove.

Exercise 8 Show that, when $|a\rangle$ and $|b\rangle$ are nonzero, $|\langle a|b\rangle| = \|a\| \|b\|$ if and only if $|a\rangle$ is a multiple of $|b\rangle$.
