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 appalls mathematicians. In chemical physics we can test the correctness of our derivations by comparing them with the results obtained by measurements. Because of this we can afford to cut corners. For example, the definitions of the mathematical objects that we use are often incomplete because we rely on the reader to use the physics context to fill in the missing details. We use often infinite sums and rarely prove that they are convergent. Mathematicians who studied the formalism of quantum mechanics invariably found that the results are confirmed by their rigorous analysis. This happens because we use physics and numerical calculations to convince ourselves that the mathematical manipulations we do are reasonable.

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 corresponding energy eigenfunctions (or eigenstates, or wave functions) $\psi_{E_n}(x)$, $n = 0, 1, 2, \ldots$. 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

$$\int_{-\infty}^{+\infty} \psi_{E_n}(x)^*\, V(x)\, \psi_{E_n}(x)\, \mathrm{d}x \equiv \langle E_n \mid \hat{V} \mid E_n \rangle \tag{1}$$

Here $V(x)$ 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. Since this notation permeates quantum mechanics let us analyze it carefully. The left-hand side contains the integration variable x; the right-hand side does not. As you know from mathematics, an integration variable is a "dummy": we can call it y or z since its name does not matter; the integral depends only on the limits of integration. This is why knowing that we happened to call the integration variable x is superfluous information that is no longer present in the symbol $\langle E_n \,|\, \hat{V} \,|\, E_n \rangle$. The left-hand side of Eq. 1 also contains the symbol ψ which is no longer present in the right-hand side. The symbol ψ_{E_n} tells us that the state of the system is such that its total energy is E_n. The presence of ψ in ψ_{E_n} adds no information. It is a superfluous letter and this is why it does not appear in $\langle E_n \,|\, \hat{V} \,|\, E_n \rangle$.

Another example of this type of abbreviation is the condition that the energy eigenstates must be orthonormal:

$$\int_{-\infty}^{+\infty} \psi_{E_n}(x)^* \, \psi_{E_m}(x) \, \mathrm{d}x \equiv \langle E_n \,|\, E_m \rangle \tag{2}$$

The symbol $\langle E_n \,|\, E_m \rangle$ in Eq. 2 is a convenient shorthand that eliminates all superfluous information from the integral.

Here is a final example, which displays again the generality of the shorthand notation. To calculate the absorption rate of a photon that causes a transition from a state of energy E_n to one of energy E_m, we evaluate

$$\int \psi_{E_n}(x)^* \, \hat{\mu} \, \psi_{E_m}(x) \, \mathrm{d}x \equiv \langle E_n \,|\, \hat{\mu} \,|\, E_m \rangle \tag{3}$$

in which $\hat{\mu}$ is the dipole-moment operator. The shorthand notation $\langle E_n \,|\, \hat{\mu} \,|\, E_m \rangle$ eliminates all superfluous information that is present in the left-hand side of the equation.

It is not important, at this point, that you remember these formulae or that you understand their meaning. What I want to make clear is that any measurable quantity in quantum mechanics can be expressed by such a shorthand notation. We can capture the essence of the theory by using the symbols $|E_n\rangle$ and $\langle E_m|$ and physical quantities such as V and $\hat{\mu}$. Dirac set out to examine how much of the quantum theory can be captured by using these symbols and developing rules for their manipulation.[1]

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, and 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 using 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 and of photons 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 state of spin or of a photon.

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

---

[1] While talking to a friend, who is a computer scientist, I marveled at the fact that Beethoven could compose music while deaf. She told me "this is possible because of a good notation".

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

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

The set of all values that an observable might 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 is used. Consider a gas containing diatomic molecules. The vibrational states of a diatomic molecule are well approximated (when the vibrational energy is small) by the states of a harmonic oscillator. If the temperature of the gas is such that $k_B T$ is much smaller than the energy $\hbar\omega$, the oscillators will be in the ground state. In this case we know that the state of the oscillator is $|E_0\rangle$ where $E_0 = \hbar\omega/2$ (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 $E_1 = (1 + 1/2)\hbar\omega$, and their state will be $|E_1\rangle$.

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

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

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 hydrogen requires three quantum numbers. One of them, $n = 1, 2, \ldots$, tells us (through Eq. 5) the energy of the electron. A second one, $\ell = 0, 1, \ldots, n-1$, tells us that the value of the angular momentum squared is $\hbar^2\ell(\ell+1)$ (see Metiu, Chapter 18). An additional quantum number, often denoted by m, 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 the electron in a hydrogen atom is given by the ket $|n, \ell, m\rangle$, where n, ℓ, and m have specific numerical values. 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 electron energy and it is also a pure state of the square of its angular momentum and of the projection of its angular momentum on the OZ axis. As you can see from this example a system can be in a state that is a pure state of several quantities.

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 either

$-1/2$ or to $1/2$, and tells us how the electron spin is "oriented". A state in which we know the electron spin, its angular momentum and its energy is described by the ket $|n, \ell, m, m_s\rangle$.

I gave this example to show that the ket is a flexible 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 remember 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** *Pure states of observables.* The fact that the energy in some systems can only take discrete values was one of the great surprises introduced by quantum mechanics and has been abundantly described in popular presentations. Let's make sure that we understand what it means.

Consider the energy of a harmonic oscillator, which can only take one of the values $E_n = \hbar\omega(n + 1/2)$, where n could be zero, one, two, etc. Nature does not allow us to prepare a state in which the energy E of an oscillator has a value satisfying $\hbar\omega(n + 1/2) < E < \hbar\omega(1 + n + 1/2)$. Nor can we find any force in nature that can change the energy of the oscillator by an infinitesimal amount. Any natural force[2] acting on the oscillator changes its energy by a multiple of $\hbar\omega$ or not at all.

This statement has immediate experimental consequences. Let us make

---

[2] Supernatural forces are excluded throughout these lectures.

the reasonable assumption that the vibrational motion of a diatomic molecule is well represented by a harmonic oscillator. I have a dilute gas of diatomic molecules and I expose it to infrared radiation of frequency Ω and energy $\hbar\Omega$. My instrument allows me to change the frequency continuously.[3] It won't surprise you to hear that if an atom absorbs an energy $\hbar\Omega$, the energy must be conserved. If the atom was initially in a pure energy state $|E_n\rangle$ and goes into a pure energy state $|E_m\rangle$ when it absorbs the photon, then

$$\hbar\Omega = E_m - E_n = \hbar\omega(m-n), m > n, n = 0, 1, 2, \ldots$$

If this condition is not satisfied, the photons are not absorbed: they pass through the gas unmolested. If the energy is discrete, I can change Ω continuously and observe absorption only for a few values of Ω. This is what is found in experiments.

Is this discreteness confined to energy only? No. The square of angular momentum has discrete values

$$\hbar^2 j(j+1), j = 0, 1, 2, \ldots$$

You might rightly object that angular momentum squared is proportional to the rotational energy and that it is the rotational energy that is discrete. Consider, however, the projection of the angular momentum on an axis. This also takes discrete values. If you believe the quaint notion that the electron in a hydrogen atom moves in an orbit, then the angular momentum is perpendicular to the orbit. The fact that the projection of the angular momentum can only have discrete values would then mean that only orbits

---

[3] Isn't this interesting? I was just telling you that energy takes discrete values. Not always, as you'll see soon.

of a certain "inclination" are allowed. This is very strange. The picture based on orbits is mostly a metaphor[4]; the fact that the projection of the angular momentum takes discrete values follows from the basic rules of quantum mechanics and has been confirmed by all spectroscopic measurements done so far.

If energy takes discrete values, why haven't people noticed this before quantum mechanics? For one thing, classical physics dealt with heavy objects. Their energy spectrum is discrete by the energy difference between the discrete states is so small that our instruments don't have sufficient resolution to detect it.

Is this discreteness a property of all measurable properties of small things? Certainly not. Consider a gas of molecules inside an ultra-high-vacuum chamber. The density of the gas is so low that the molecules present there do not collide with each other; each molecule moves with a constant momentum. There is no natural law that says that we cannot change the momentum infinitesimally. Since the gas is introduced in the chamber through a nozzle, we can change the momentum infinitesimally by making infinitesimal changes of the pressure of the gas going into the nozzle.

The position of a particle can also be changed by an infinitesimal amount. While a harmonic oscillator has discrete energies, there is nothing in theory that prevents us from changing the position of the oscillator by a small amount. This does not mean that we know how to do this in the laboratory

---

[4] This semi-classical interpretation seems to be helpful to people who need to have a picture associated with quantum phenomena. On the other extreme are people who would say that this discretization follows from the commutation relations for the components of the angular momentum vector and that is all you want to know.

— but no law of physics prevents us from doing it.

There are cases in which even energy can take continuous values. A simple example is provided by photo-ionization. We can ionize a molecule by exposing it to photons of sufficiently large energy. Photon absorption causes the ejection of an electron from the molecule and the formation of an ion. If we vary the photon energy continuously, the energy of the fragments varies continuously. In general, we can make infinitesimal changes in energy in any experiments in which a complex system is broken into fragments. The same is true for any experiment in which molecules collide with each other. The energy of the final state in these experiments is the sum of the kinetic energy of the fragments and the energy of their internal motion (vibrational, rotational, electronic). The kinetic energy depends on momentum, and momentum can be changed infinitesimally with no objection from quantum mechanics. The same is true for the total energy of the initial state in these experiments. In the case of photo-ionization we can vary the energy of the photon infinitesimally and in collisions we can change the kinetic energy prior to collision by infinitesimal amount. The upshot: continuous values of energy.

There is no difficulty in using kets to describe such quantities. For example, $|x\rangle$ is a pure state in which a particle is located at position x, and $|p\rangle$ is a pure state in which a particle moves with momentum p. We do not have instruments that can create such states but there is no law of nature preventing us from creating them. They turn out to be of great importance in developing the formalism of quantum mechanics. We can prepare states that are fairly good approximations of these pure states.

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, which are the simplest examples used to illustrate how quantum mechanics works, do not have a continuous spectrum; they are oversimplified models of real systems. The hydrogen atom has a continuous spectrum, but this is rarely mentioned. The analysis of any phenomenon in which two particles collide (e.g. energy transfer, chemical reactions) or a system of particles is broken into fragments (e.g. photo-ionization, photo-dissociation) has to deal with quantities having a continuous spectrum.

§ 4 *The spectrum of an observable.* We can generalize these examples as follows. If A is an observable, then measurements of the value of A can yield the discrete values

$$a_1, a_2, \ldots, a_i, \ldots$$

and/or the continuous values

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

where m and n are real numbers. The set of these possible measurement results is called the *spectrum* of the observable 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). States in which *we are sure* that a measurement of A gives the result a are called *pure states* of A and are represented by the ket $|a\rangle$.

§ 5 *Coherent superposition of pure states.* In many experiments the state of a system is such that the result of a measurement *is not certain*, which means

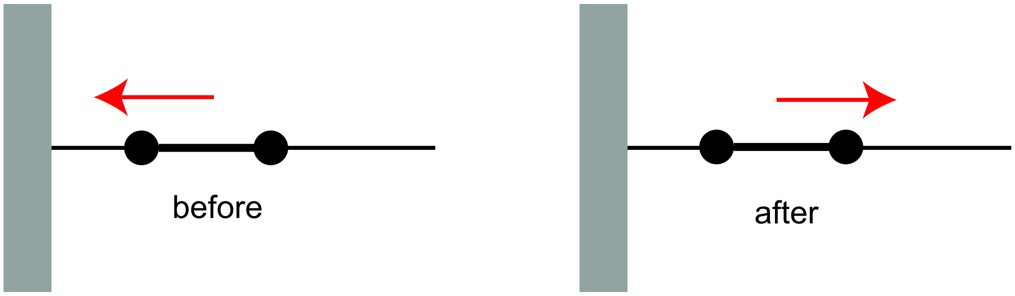

Figure 1: Collision and rebound

that 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 moving with high velocity hits a hard wall (the wall is called 'hard' because it cannot exchange energy with the molecule). To make life simple, I assume that the molecule is prevented from rotating and moves in one dimension, as shown in Fig. 1. Before the collision the vibrational energy of the molecule is E_0 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 of the system. The ket describing this pure state is $|E_0, p\rangle$.

What is the state of the molecule after the collision? If the molecule moves initially very fast, some of its kinetic energy might be converted into vibrational energy. Let us denote the vibrational energy *after the collision* by E_n. Since total energy must be conserved,[5] the energy after the collision

---

[5] You might suspect that energy conservation does not hold in quantum mechanics and therefore conclusions based on it might be wrong. As you will see later, if the energy of a process is conserved in classical mechanics it is also conserved in quantum mechanics.

is

$$\frac{p_n^2}{2m} + E_n = \frac{p^2}{2m} + E_0 \tag{6}$$

The kinetic energy after the collision is $\frac{p_n^2}{2m}$ where p_n is the momentum after the collision. Since E_n is larger than E_0, the kinetic energy after the collision is smaller than before the collision. The fact that the kinetic energy *must be positive* limits the magnitude of the vibrational energy E_n that the molecule can have after the collision. Let us assume that the initial kinetic energy $\frac{p^2}{2m}$ is so large that the molecule could have, after the collision, the vibrational energies E_0, or E_1, or E_2 without violating the restrictions imposed by energy conservation. If the vibrational energy after the collision is E_0 the only effect of the collision is to change the sign of momentum: the molecule has the same kinetic energy as before the collision, but it moves in the opposite direction. The ket describing this state is $|E_0, -p\rangle$. If, after the collision, the molecule has the vibrational energy E_1 then its kinetic energy will be lower than initially and equal to $\frac{p_1^2}{2m} = \frac{p^2}{2m} - (E_1 - E_0)$. The ket describing this state is $|E_1, -p_1\rangle$. Similarly, if the vibrational energy is E_2, the state is described by $|E_2, -p_2\rangle$.

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:

$$|\psi\rangle = c_0|E_0, -p\rangle + c_1|E_1, -p_1\rangle + c_2|E_2, -p_2\rangle \tag{7}$$

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. 7 is a central principle of quantum mechanics. The experiment we have examined has several *possible* final states: $|E_0, -p\rangle$, $|E_1, -p_1\rangle$, and $|E_2, -p_2\rangle$. 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 states that are possible.* In our case this means all the states that do 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 may take one of the values a_1, a_2, …, a_n, then the state of the system is described by the ket

$$|\psi\rangle = c_1|a_1\rangle + c_2|a_2\rangle + \cdots + c_n|a_n\rangle \tag{8}$$

Here $|a_1\rangle$, …, $|a_n\rangle$ are the pure states corresponding to the values a_1, …, a_n, and c_1, …, c_n are complex numbers. It is the business of quantum mechanics to provide equations that allow us to calculate the magnitude of the coefficients c_1, c_2, …, c_n for every specific experiment.

We can consider the statement made by Eq. 8 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 will have the form given by Eq. 8. 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, because it allows me to say something about some physical processes without having to solve equations.

Exercise 1 Consider an experiment in which a rigid diatomic molecule collides with a hard wall and the molecule's kinetic energy is less than that needed for the collision to excite a vibration. We now 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 8 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 properties[6]: (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, and this multiplication 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. Without these properties we could not describe coherent superpositions of pure states.

It is important to 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 does not form a linear space because *adding a hydrogen-atom ket to a harmonic-oscillator ket is devoid by any physical meaning.* Unfortunately the way the ket notation is used does not make this distinction clear, and this can lead to confusion when the general equations, written in terms of kets, are applied to specific situations.

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.

---

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

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.* I give here examples of linear spaces called R^3, L^2, and $\boldsymbol{\ell}^2$, which appear in quantum mechanics and many other areas of physics. R^3 is the space of the three-dimensional vectors that you studied in calculus or classical physics. L^2 is the space to which the Schrödinger wave functions belong. $\boldsymbol{\ell}^2$ is the space used in Heisenberg's version of quantum mechanics.

The space R^3 of ordinary vectors. When you studied mechanics or electricity you worked with three-dimensional vectors. Sometimes you represented each vector by an arrow that showed the direction of the vector and its magnitude. An alternative (equivalent) description was produced by picking a coordinate system and describing the properties of a vector by specifying 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 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 this as

$$\{x_1, x_2, x_3\} + \{y_1, y_2, y_3\} \equiv \{x_1 + y_1, x_2 + y_2, x_3 + y_3\} \tag{9}$$

A mathematician could have invented symbols like $\{x_1, x_2, x_3\}$ and could have defined arbitrary rules for their addition (such as Eq. 9), 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. 9.

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

$$\alpha\{x_1, x_2, x_3\} \equiv \{\alpha x_1, \alpha x_2, \alpha x_3\} \tag{10}$$

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. 9 and 10 for addition and multiplication by a (real) number, form a linear space called R^3. Here 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 R^2? How about the space R^n where n is a finite integer?

The linear space L^2. The space L^2 contains all functions $f(x)$ for which the integral

$$\int_{-\infty}^{+\infty} f(x)^* f(x)\,\mathrm{d}x \tag{11}$$

exists and is finite. Here x is a real variable (meaning that x takes only real-number values) and $f(x)$ is a function that can take complex values. $f(x)^*$ is the complex conjugate[7] of $f(x)$.

The functions $\sin x$ and $\mathrm{e}^{\mathrm{i}kx}$ do not belong to this space, because the integrals

$$\int_{-\infty}^{+\infty} \sin^2 x\,\mathrm{d}x$$

and

$$\int_{-\infty}^{+\infty} \mathrm{e}^{-\mathrm{i}kx}\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x$$

do not exist. On the other hand, $f(x) = \mathrm{e}^{-ax^2}\mathrm{e}^{\mathrm{i}kx}$ does belong to L^2 as long as the real part of the complex number a is positive, since in that case the integral is finite.

It should be clear that the set L^2 is a linear space. Indeed, if two functions $f(x)$ and $\phi(x)$ belong to L^2 then so does their sum. 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.

---

[7] You calculate $f(x)^*$ from $f(x)$ by changing the sign of $\mathrm{i} \equiv \sqrt{-1}$ everywhere it appears in $f(x)$. For example, the complex conjugate of $(x + \mathrm{i}y)\exp(\mathrm{i}\omega t)$ is $(x - \mathrm{i}y)\exp(-\mathrm{i}\omega t)$ if x, y, ω and t are real numbers. If $z = x + \mathrm{i}y$, and x and y are real numbers, then $zz^* = x^2 + y^2$.

Exercise 4 Show that the lowest three energy eigenfunctions of a harmonic oscillator belong to L^2. 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

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y,z)^* f(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z < \infty \tag{12}$$

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

I emphasize again that in physics the linear spaces must be constructed from wave functions that refer to the same physical system. A space that consists of the wave functions of a harmonic oscillator of frequency ω_1 *is a different space* than that of the wave functions of an oscillator having the frequency ω_2. Adding a wave function of an oscillator of one frequency to the wave function of an oscillator of a different frequency is not sensible: the resulting state does not represent any state that can be created by experiment.

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

$$\sum_{n=1}^{\infty} a_n^* a_n < \infty \tag{13}$$

Here a_n^* denotes the complex conjugate of a_n.

If $a \equiv \{a_1, a_2, \ldots\}$ and $b \equiv \{b_1, b_2, \ldots\}$ their sum is

$$a + b \equiv \{a_1 + b_1, a_2 + b_2, \ldots\} \tag{14}$$

The product of $a \equiv \{a_1, a_2, \ldots\}$ with a complex number α is defined by

$$\alpha a = \{\alpha a_1, \alpha a_2, \ldots\} \tag{15}$$

It should be obvious that if $a \equiv \{a_n\}_{n=1}^{\infty}$ and $b \equiv \{b_n\}_{n=1}^{\infty}$ satisfy Eq. 13, then $a + b$ and αa also satisfy it. These sequences form a linear space.

As in the case of other spaces the sequences that form a given linear space must describe the same physical system. This will become clear when we discuss the representation theory which connects wave functions to such sequences.

§ 9 *Heisenberg formulation.* Heisenberg formulated quantum mechanics in terms of elements of $\boldsymbol{\ell}^2$ and Schrödinger used L^2. For a while physicists had two theories, that of Schrödinger and that of Heisenberg, Jordan, and Born. They seemed 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. At about the same time Dirac reformulated the theory by introducing bras, kets, and operators. His formulation is more general and these lecture notes are based on it. When the time is ripe I will show how the Schrödinger formulation and the formulation of Heisenberg, Jordan, and Born follows naturally from Dirac's. It may humble you to know that 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 the kets representing the state of the system with the numerical values produced by experiments. This is done through an operation called *scalar product*.

You have encountered the scalar product when you studied introductory quantum mechanics, but it is likely that you did not learn to look at it as a general operation involving two kets. For example, you learned about the orthonormality condition:

$$\int_{-\infty}^{+\infty} \psi_{E_n}(x)^* \, \psi_{E_m}(x) \, \mathrm{d}x \equiv \langle E_n \,|\, E_m \rangle \tag{16}$$

is 1 when $n = m$ and 0 otherwise. The left-hand side of Eq. 16 is an example of a scalar product of the wave function $\psi_{E_n}(x)$ with the wave function $\psi_{E_m}(x)$. The right-hand side is an example of a scalar product between the kets $|E_n\rangle$ and $|E_m\rangle$. Even though these two expressions (i.e. the integral at the left and the symbol at the right) seem dramatically different, they are both scalar products because they share certain abstract properties.

What are these common properties? Both expressions are “functions” of two “variables”: in one case the variables are $\psi_{E_n}(x)$ and $\psi_{E_m}(x)$, in the other case they are $|E_n\rangle$ and $|E_m\rangle$.

In what follows I will use the notation employed in the right-hand side of Eq. 16. The scalar product of the ket $|a\rangle$ with the ket $|b\rangle$ is denoted by

$$\langle a \,|\, b \rangle \tag{17}$$

This is a function of the two kets. This notation for a function of two “variables” (the two kets) may seem a bit strange. You are used the notation

$f(x, y)$, 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 nearly so in mathematics (some mathematicians use $(a|b)$ instead of $\langle a \,|\, b\rangle$).

The scalar product is a function of two kets *defined* to have 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 non-negative real number:

$$\langle a \,|\, a\rangle \geq 0 \tag{18}$$

3. $\langle a \,|\, a\rangle = 0$ 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.

4. If $|c\rangle = |a\rangle + |b\rangle$ then

$$\langle d \,|\, c\rangle \equiv \langle d \,|\, a + b\rangle = \langle d \,|\, a\rangle + \langle d \,|\, b\rangle \tag{19}$$

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

$$\langle b \,|\, \eta\rangle = \alpha\langle b \,|\, a\rangle \tag{20}$$

6. $\langle a \,|\, b\rangle$ and $\langle b \,|\, a\rangle$ are related through:

$$\langle a \,|\, b\rangle = \langle b \,|\, a\rangle^* \tag{21}$$

I will show shortly several incarnations of this abstract definition of the scalar product. Before doing that I derive some properties of the scalar product that follow from the abstract definition given above.

I emphasize that the scalar product is defined only for kets that belong to the same space. You cannot take the scalar product of a ket representing a harmonic oscillator with a ket representing a hydrogen atom. However, *if one takes the necessary precautions* one can take the scalar product of a ket representing the energy of a particle in a box with a ket representing the energy of a harmonic oscillator. You will see why, in one of the chapters that follows.

§ 11 *Scalar product: other properties.* This six-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. R^3, R^n, L^2, $\boldsymbol{\ell}^2$) 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 six 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

$$\langle \eta \,|\, b\rangle = \alpha^* \langle a \,|\, b\rangle \tag{22}$$

Proof: From defining properties 5 and 6 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 $\eta = \alpha|a\rangle + |b\rangle$ then

$$\langle \eta \,|\, c\rangle = \alpha^*\langle a \,|\, c\rangle + \langle b \,|\, c\rangle \tag{23}$$

Proof: $\langle \eta \,|\, c\rangle \overset{(6)}{=} \langle c \,|\, \eta\rangle^* \overset{(4)}{=} (\langle c \,|\, \alpha a\rangle + \langle c \,|\, b\rangle)^* \overset{(5)}{=} (\alpha\langle c \,|\, a\rangle + \langle c \,|\, b\rangle)^* = \alpha^*\langle c \,|\, a\rangle^* + \langle c \,|\, b\rangle^* \overset{(6)}{=} \alpha^*\langle a \,|\, c\rangle + \langle b \,|\, c\rangle$. A shorthand notation for this property is

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

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

$$\langle \eta \,|\, \eta\rangle = \langle a \,|\, a\rangle + \lambda^*\langle b \,|\, a\rangle + \lambda\langle a \,|\, b\rangle + \lambda^*\lambda\langle b \,|\, b\rangle \tag{25}$$

This follows from (a) and (b). A shorthand notation for this operation is

$$\langle a + \lambda b|a + \lambda b\rangle = \langle a \,|\, a\rangle + \lambda^*\langle b \,|\, a\rangle + \lambda\langle a \,|\, b\rangle + \lambda^*\lambda\langle b \,|\, b\rangle \tag{26}$$

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 *Examples of scalar products*

The scalar product in R^3*.* The simplest and most familiar linear space we encountered is the set 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:

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

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 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 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

$$\int_{-\infty}^{+\infty} f(x)^* f(x) \, \mathrm{d}x < \infty \tag{28}$$

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

$$\langle f \mid \phi \rangle \equiv \int_{-\infty}^{+\infty} f(x)^* \phi(x) \, \mathrm{d}x \tag{29}$$

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 $\boldsymbol{\ell}^2$. The set of all sequences $\{a_1, a_2, \ldots\}$ satisfying

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

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

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

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 R^3 is said to be of unit length if

$$\vec{v}\cdot\vec{v} = 1 \tag{34}$$

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

$$\langle a \,|\, a\rangle = 1 \tag{35}$$

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

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

$$\vec{v}\cdot\vec{u} \equiv \langle\vec{v}\,|\,\vec{u}\rangle = 0 \tag{36}$$

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

$$\langle a \,|\, b\rangle = 0 \tag{37}$$

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 R^3 satisfy the equation

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

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

$$\langle\vec{v}|\vec{u}\rangle = \|\vec{v}\|\,\|\vec{u}\|\cos\theta \tag{39}$$

Because $|\cos\theta|$ is at most 1, Eq. 39 implies that

$$|\langle\vec{v}|\vec{u}\rangle| \leq \|\vec{v}\|\|\vec{u}\| \tag{40}$$

To generalize this result, from R^3 to an arbitrary linear space, we guess that for any two kets in the same linear space we must have

$$|\langle a\,|\,b\rangle| \leq \|a\|\,\|b\| \tag{41}$$

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 looked at the properties required for scalar product of two kets and showed how scalar products are defined in the spaces L^2 and $\boldsymbol{\ell}^2$. Later I will show that there is a one-to-one correspondence between the elements of L^2, $\boldsymbol{\ell}^2$, and the abstract space of the kets.

A lot of work remains to be done. We need to find how the kets represent physical phenomena and we need to show that the theory using the ket notation contains Schrödinger's and Heisenberg's versions of the theory as particular cases. 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 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

$$|\langle a\,|\,b\rangle| \leq \|a\|\,\|b\| \tag{42}$$

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

Let

$$|c\rangle = |a\rangle + \lambda^*|b\rangle \tag{43}$$

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

$$r(\lambda) \equiv \langle c\,|\,c\rangle = \langle a\,|\,a\rangle + \lambda^*\langle a\,|\,b\rangle + \lambda\langle b\,|\,a\rangle + \lambda^*\lambda\langle b\,|\,b\rangle \tag{44}$$

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,

$$r(\lambda) \geq 0. \tag{45}$$

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

$$\lambda = \frac{-\langle b\,|\,a\rangle^*}{\langle b\,|\,b\rangle} = \frac{-\langle a\,|\,b\rangle}{\langle b\,|\,b\rangle} \tag{46}$$

Inserting this value in Eq. 44 gives

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

which implies

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

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 $|b\rangle$.
