Chapter 3. The Completeness Relation and Various Ket Representations

Section 3.1. The Completeness Relation

§ 1 *Outline.* In this chapter we use the definition of the probability that a measurement of an observable yields a particular result, to derive an expression called the *completeness relation* or the *resolution of identity.* This property is essential for obtaining numerical solutions of the Schrödinger equation and for analyzing the physical properties of a system.

§ 2 *A simple property of probability.* We measure a quantity A that can take one of the values $\{a_1, a_2, \ldots, a_n\}$. If the measurement consists of flipping a coin, the values that A can take are {head, tail}; if we roll a die, they are $\{1, 2, 3, 4, 5, 6\}$. I denote the probability that a measurement of A gives

the result a_i by $P(a_i)$. To define this probability, I imagine performing the measurement N times, where N is a very large number. I find that a_1 is the result of the measurement N_1 times, a_2 is the result N_2 times, etc.

Obviously I have

$$\sum_{i=1}^{n} \frac{N_i}{N} = 1 \tag{1}$$

The probability that an observation gives a_i is (by definition)

$$P(a_i) \equiv \lim_{N\to\infty} \frac{N_i}{N} \tag{2}$$

Combining Eq. 1 with Eq. 2 gives

$$\sum_{i=1}^{n} P(a_i) = 1 \tag{3}$$

It is essential that *this sum includes all possible values* $\{a_1, a_2, \ldots, a_n\}$ of A. Eq. 3 holds when n is finite or infinite.

§ 3 *Application to quantum mechanics.* Assume now that A is a quantum mechanical observable, its spectrum is $\{a_1, a_2, a_3, \ldots\}$, and its continuous spectrum is $\alpha \in D$. If the state of the system is represented by the *normalized* ket $|\psi\rangle$, the probability that a measurement of A yields the value a_i is (see previous chapter)

$$P_\psi(a_i) = |\langle a_i \,|\, \psi\rangle|^2 \tag{4}$$

The probability that the result of the measurement is a value between α and $\alpha + \,\mathrm{d}\alpha$ is

$$P_\psi(\alpha)\,\mathrm{d}\alpha = |\langle \alpha \,|\, \psi\rangle|^2\,\mathrm{d}\alpha,\ \alpha \in D \tag{5}$$

By definition, the spectrum contains all values that A can take and so Eq. 3 applies for these probabilities:

$$\sum_{i=1}^{\infty} P_\psi(a_i) + \int_{\alpha\in D} \mathrm{d}\alpha\, P_\psi(\alpha) = \sum_{i=1}^{\infty} |\langle a_i \,|\, \psi\rangle|^2 + \int_{\alpha\in D} \mathrm{d}\alpha\, |\langle \alpha \,|\, \psi\rangle|^2 = 1 \tag{6}$$

We have generalized Eq. 3 to include the continuous spectrum by using an integral instead of a sum.

§ 4 *Dirac notation: bras and operators.* Before we dive into the intricacies of Dirac notation, I remind you a few elementary facts about complex numbers and the scalar product. For any complex number z, we have $|z|^2 \equiv z^* z$ where $|z|$ is the absolute value of the complex number and z^* is the complex conjugate of z. I also remind you that for any scalar product, of any two kets, we have $\langle \lambda \,|\, \mu \rangle = \langle \mu \,|\, \lambda \rangle^*$. Combining this knowledge we can rewrite the expression for probability as

$$|\langle a_i \,|\, \psi \rangle|^2 = \langle a_i | \psi \rangle^* \langle a_i | \psi \rangle = \langle \psi \,|\, a_i \rangle \langle a_i \,|\, \psi \rangle \tag{7}$$

. Using this result we can go on and rewrite Eq. 6 as

$$\sum_{n=1}^{\infty} \langle \psi \,|\, a_n \rangle \, \langle a_n \,|\, \psi \rangle + \int_{\alpha \in D} \mathrm{d}\alpha \, \langle \psi \,|\, \alpha \rangle \, \langle \alpha \,|\, \psi \rangle = 1 \tag{8}$$

As you may have noticed by now, mathematics is all about rewriting expressions by using known rules. This may be fun but it is not particularly exciting unless you are Dirac. He examined this expression and took it in a truly innovative direction. He decided that it might be profitable (and it is) to think that the expression

$$\langle \psi \,|\, a_n \rangle \, \langle a_n \,|\, \psi \rangle \equiv \langle \psi \,|\, (|a_n\rangle\langle a_n|) \,|\, \psi \rangle \tag{9}$$

consists of three distinct entities: the *bra* $\langle \psi |$, the *operator* $|a_n\rangle\langle a_n|$, and the ket $|\psi\rangle$.

We already know what a ket is, but the bra and the operator are new objects, defined by Eq. 9. Let us take a look at what they do.

A bra $\langle a|$ must be understood as a symbol that acts on a ket; the rule for this action is

$$(\langle a|)|\psi\rangle = \langle a\,|\,\psi\rangle \tag{10}$$

Thus, a bra $\langle a|$ acting on a ket $|\psi\rangle$ gives the complex number $\langle a\,|\,\psi\rangle$. For every ket $|\eta\rangle$, in a given ket space, there is a corresponding bra $\langle\eta|$ defined by the rule given above. In addition, for every bra there is a ket. Bras and kets are in a one-to-one correspondence. I emphasize that a bra $\langle\eta|$ acting on a ket $|\phi\rangle$ must correspond to a ket that belongs to the same space as $|\phi\rangle$. It makes no sense to act with a bra corresponding to a state of the hydrogen atom on a ket that represents the state of a harmonic operator. On the other hand, one can act with a bra describing a state of a harmonic oscillator on a ket describing the state of a particle in a box, as long as the potential energy of the oscillator is contained in the box (details about this are given later in this section). This happens because the states of the two systems belong to the same linear space. I suspect that this seems rather vague now but it will become very clear (and trivial) as you learn more about quantum mechanics.

The manner in which bras are defined allows us to derive rules of computation for them. If $|\lambda\rangle = |\eta\rangle + |\mu\rangle$ then we can use the properties of the scalar product to write

$$\langle\lambda\,|\,\psi\rangle = \langle\psi\,|\,\lambda\rangle^* = \langle\psi\,|\,\eta\rangle^* + \langle\psi\,|\,\mu\rangle^* = \langle\eta\,|\,\psi\rangle + \langle\mu\,|\,\psi\rangle \tag{11}$$

Therefore we have the rule:

$$\text{if } |\lambda\rangle = |\eta\rangle + |\mu\rangle \text{ then } \langle\lambda| = \langle\eta| + \langle\mu| \tag{12}$$

Similarly, if $|\eta\rangle = \alpha|a\rangle$, where α is a complex number, then

$$\langle\eta\,|\,\psi\rangle = \langle\psi\,|\,\eta\rangle^* = \langle\psi\,|\,\alpha a\rangle^* = (\alpha\langle\psi\,|\,a\rangle)^* = \alpha^*\langle a\,|\,\psi\rangle \tag{13}$$

and therefore, we have the rule:

$$\text{if } |\eta\rangle = \alpha|a\rangle \text{ then } \langle\eta| = \alpha^*\langle a| \tag{14}$$

The operator

$$\hat{P}(a_n) \equiv |a_n\rangle\langle a_n| \tag{15}$$

is also a new mathematical object (meaning that we have not mentioned it so far). By definition it acts on a ket $|\psi\rangle$ through the rule

$$\hat{P}(a_n)|\psi\rangle = (|a_n\rangle\langle a_n|)|\psi\rangle = |a_n\rangle\,\langle a_n\,|\,\psi\rangle \tag{16}$$

An operator (in our case, $|a_n\rangle\langle a_n|$) acts on a ket (in our case, $|\psi\rangle$) to produce another ket (in our case, $|a_n\rangle\,\langle a_n\,|\,\psi\rangle$). The product $|a_n\rangle\,\langle a_n\,|\,\psi\rangle$ is a ket because $\langle a_n\,|\,\psi\rangle$ is a complex number, $|a_n\rangle$ is a ket, and the product of a complex number with a ket is a ket. In general, an operator is any object that acts on a ket to generate another ket.

Keep in mind that the operator $|\,a_n\rangle\langle a_n\,|$ acts only on kets $|\psi\rangle$ that belong to the same linear space as $|a_n\rangle$. If $|a_n\rangle$ describes states of the hydrogen atom, it makes no sense to act with $|a_n\rangle\langle a_n|$ on a state $|\psi\rangle$ of a harmonic oscillator.

§ 5 *The completeness relation.* Now that we are comfortable with Dirac's notation, let us use it to rewrite Eq. 8 as

$$\langle\psi\,|\,(\sum_{n=1}^{\infty}|a_n\rangle\langle a_n| + \int_{\alpha\in D}\mathrm{d}\alpha\,|\alpha\rangle\langle\alpha|)\,|\,\psi\rangle = 1 \tag{17}$$

It will prove very convenient to introduce the notation

$$\sum_{n=1}^{\infty}|a_n\rangle\langle a_n| + \int_{\alpha\in D}\mathrm{d}\alpha\,|\alpha\rangle\langle\alpha| = \hat{I} \tag{18}$$

Obviously $\hat{I}$ is an operator because it is a sum of the operators $|a_n\rangle\langle a_n|$ and the integral over the operators $|\alpha\rangle\langle\alpha|$. $\hat{I}$ is however a peculiar operator as you can see by using it in Eq. 17, which becomes

$$\langle\psi|\hat{I}|\psi\rangle = 1 \tag{19}$$

Eq. 17 originates from the statement that the probability that the measurement gives *some* result must be equal to 1. When we defined the probability in quantum mechanics, I emphasized that the state $|\psi\rangle$ *must* be normalized, that is,

$$\langle\psi|\psi\rangle = 1 \tag{20}$$

When you compare Eq. 19 with Eq. 20, you find that $\hat{I}$ acts on a ket but has no effect on it: in other words

$$\hat{I}|\psi\rangle = |\psi\rangle \quad \text{for any ket } |\psi\rangle \tag{21}$$

Because of this, $\hat{I}$ is called the unit operator and plays in the set of operators the same role as the number 1 does in the multiplication of numbers ($1 \times 3 = 3$). You must keep in mind that $\hat{I}$ is the unit operator in the space to which the $|a_n\rangle$ belong. A unit operator in the space of the states of a harmonic oscillator is not a unit operator in the space formed by the states of hydrogen atom.

The main result here is Eq. 18, which is called the *completeness relation* for the pure states $|a_n\rangle$, $n = 1, 2, \ldots$, and $|\alpha\rangle$, $\alpha \in D$. This is a very complicated way of writing the operator $\hat{I}$, whose effect is to do nothing! The "derivation" given above is not rigorous[1]. On the other hand, the result

[1]This derivation does not show that the infinite sum is convergent and that the integral

rests on solid physical grounds: it is a consequence of the fact that the probability that we get some result when we make a measurement is equal to 1. To appreciate Dirac's ingenuity, consider the fact that one has to study intensely hundreds of pages of dense mathematics before one gets a rigorous proof of the completeness relation. However, what follows will instill respect for the fussiness of the mathematicians as we discover that the very general expression of the completeness relation is misleading and its mechanical use (i.e. without insight into its limitations) can lead to serious damage to the reputation of the user.

§ 6 *Completeness is valid, in principle, for the states of any observable.* The arguments made in deriving the completeness relation put no constraints on the observable A; this relationship is valid for the pure states of any observable.

Let us take as observable the position of a particle. The position is an observable that has a continuous spectrum, which contains all real numbers between $-\infty$ and $+\infty$. Let $|x\rangle$ denote the pure state in which we know for certain that the particle is located at x. The completeness relation for these pure states is

$$\hat{I} = \int_{-\infty}^{+\infty} \mathrm{d}x \, |x\rangle\langle x| \tag{22}$$

We say that this equation expresses the unit operator $\hat{I}$ in the *coordinate representation.*

---

exists, which is particularly irritating to mathematicians. The definition of the operators is also a bit cavalier, and this will lead to interesting complications that are discussed in the remainder of this chapter.

There is nothing special about position and we can play the same game with the pure states $|p\rangle$ of momentum. We have

$$\hat{I} = \int_{-\infty}^{+\infty} \mathrm{d}p \, |p\rangle\langle p| \tag{23}$$

because momentum is an observable with a purely continuous spectrum. Eq. 23 gives the unit operator in the *momentum representation*.

A peculiar thing about the momentum and the position bras is that they can act on any state, except the states of spin. However one must keep in mind the dimensionality of the system. One can use $\hat{I}$ defined by Eq. 22 to act on the state of a harmonic oscillator or a particle in a box. However, one must use

$$\hat{I} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \, |x\rangle|y\rangle|z\rangle \, \langle x|\langle y|\langle z| \tag{24}$$

to act on the states of the electron in the hydrogen atom, because that system is three-dimensional. We will use these formulae for $\hat{I}$ very often and their meaning and usage will become more and more clear as we advance.

The same remarks apply to the representation of $\hat{I}$ that uses momentum states.

Finally, the unit operator in the *energy representation* is

$$\hat{I} = \sum_{n=1}^{\infty} |E_n\rangle\langle E_n| + \int_0^{\infty} \mathrm{d}\alpha \, |E_\alpha\rangle\langle E_\alpha| \tag{25}$$

where $|E_n\rangle$ and $|E_\alpha\rangle$ are the pure states of energy. This representation constructs a unit operator acting only on the space to which the states $|E_n\rangle$ belong or a space that is mathematically equivalent (as we proceed this statement will gain more substance).

Section 3.2. Representation theory

§ 7 *Various representations of a state* $|\psi\rangle$. In many experiments in quantum mechanics, we force the system (a molecule, a solid, etc.) to interact with an external agent (light, an electron beam, etc.). When this interaction stops, the system is left in a state $|\psi\rangle$. Quantum theory is then used to calculate the properties of the system in this state: the probability that the system has the energy E_n, or the average position of the particles in the system, or the evolution of the state in time, etc.

Most of these calculations start by representing $|\psi\rangle$ in a convenient form. One class of representations is generated by starting with the identity

$$|\psi\rangle = \hat{I}|\psi\rangle \tag{26}$$

If A is an observable with the discrete spectrum $\{a_n\}_{n=1}^{\infty}$ and the continuous spectrum $\alpha \in D$, we can write (we use the completeness relation)

$$\hat{I} = \sum_{n=1}^{\infty} |a_n\rangle \langle a_n| + \int_{\alpha\in D} |\alpha\rangle \langle\alpha| \, \mathrm{d}\alpha \tag{27}$$

Using this in $|\psi\rangle = \hat{I}|\psi\rangle$ (i.e. Eq. 26), we obtain

$$|\psi\rangle = \sum_{n=1}^{\infty} |a_n\rangle \langle a_n \,|\, \psi\rangle + \int_{\alpha\in D} |\alpha\rangle \langle\alpha \,|\, \psi\rangle \, \mathrm{d}\alpha \tag{28}$$

Here $|a_n\rangle$ and $|\alpha\rangle$ are the *pure states of the observable* A. The coefficients $\langle a_n \,|\, \psi\rangle$ are *complex numbers*, and $P_\psi(a_n) = |\langle a_n \,|\, \psi\rangle|^2$ is the probability that a measurement of A, when the system is in the state $|\psi\rangle$, yields the value a_n. The scalar products $\langle\alpha \,|\, \psi\rangle$ are also complex numbers, and $P_\psi(\alpha)\,\mathrm{d}\alpha = |\langle\alpha \,|\, \psi\rangle|^2 \,\mathrm{d}\alpha$ is the probability that A takes values between α and $\alpha + \mathrm{d}\alpha$ when the system is in state $|\psi\rangle$. Because of this connection, $\langle a_n \,|\, \psi\rangle$ (or $\langle\alpha \,|\, \psi\rangle$) is

called the *probability amplitude* of $|a_n\rangle$ (or $|\alpha\rangle$) in state $|\psi\rangle$. Eq. 28 is called the *expression of* $|\psi\rangle$ *in the A representation*. Mathematicians will call it the *representation of* $|\psi\rangle$ *as a linear combination of* $\{|a_n\rangle\}_{n=1}^{\infty}$ and $\{|\alpha\rangle\}_{\alpha\in D}$.

The equation is not as universal as it might appear to the innocent. The bras $\langle a_n|$ must be such that the expression $\langle a_n \,|\, \psi\rangle$ is sensible and can be evaluated. For example, it is not possible to expand a state $|\psi\rangle$ of harmonic oscillator in terms of the states $|a_n\rangle$ of a hydrogen atom.

You will learn later that the pure states of A are also the eigenstates of an operator associated with A, and Eq. 28 is said to give $|\psi\rangle$ as a linear combination of the eigenstates of A.

Usually an equation has so many names because it is important, and it has been examined from several points of view. Eq. 28 is central to quantum mechanics. It is used to prove many important relations, to represent the state of the system in ways that illuminate the system's physical properties, and to set up numerical calculations of $|\psi\rangle$. In all these applications it is assumed that we know how to determine the pure states $|a_n\rangle$ and $|\alpha\rangle$ and how to perform calculations with them. You will learn later how to do that.

I warn you again that in the presentation made so far, the resolution of identity (Eq. 27) and its consequence Eq. 28 have a false generality. If their use is not augmented with common sense and watchful care, it can lead to absurd conclusions or meaningless calculations. Later in this section I will use examples that illustrate the traps one can fall in and show that numerical analysis is a combination of art and analysis and that common sense plays a role in any numerical recipe.

§ 8 *A hint of how we might use Eq. 28.* In most experiments in quantum

mechanics, we expose a system to external agents that will change its properties. When the external action stops, the system is left in a state denoted by $|\psi\rangle$. To analyze the results of the experiment, we use the time-dependent Schrödinger equation, which includes the effect of the external agents, to calculate $|\psi\rangle$. In all but the simplest cases, such a calculation is performed on a computer. Computers cannot operate with abstract symbols like kets; they only crunch numbers. To use a computer, we must find a numerical expression for $|\psi\rangle$. This is what Eq. 28 does for us: it expresses the unknown ket $|\psi\rangle$ in terms of the *known* kets $|a_n\rangle$ and $|\alpha\rangle$ and the *unknown numbers* $\langle a_n \,|\, \psi\rangle$ and $\langle \alpha \,|\, \psi\rangle$.

Finding a usable expression for $|\psi\rangle$ means calculating the numbers $\langle a_n \,|\, \psi\rangle$ and $\langle \alpha \,|\, \psi\rangle$. One way to do this is to put the expression (Eq. 28) for $|\psi\rangle$ into the Schrödinger equation for $|\psi\rangle$. This gives equations, for the numbers $\langle a_n \,|\, \psi\rangle$ and $\langle \alpha \,|\, \psi\rangle$, that can be solved by using a computer.

There are some interesting details in the implementation of this idea, which you will learn later. This outline is only telling you why we are so interested in various ways of representing unknown kets. It also highlights the key role played by the pure states of an observable and alerts you to the fact that if we are to make progress we will have to learn how to calculate these pure states.

Note that due to these manipulations the abstract ket $|\psi\rangle$ is completely defined once we know the sequence $\langle a_n \,|\, \psi\rangle$ with $n = 1, 2, \ldots$ and the functions $\langle \alpha_n \,|\, \psi\rangle$. In addition if the state $|\psi\rangle$ describes a bound system (e.g. an atom or a molecule), it is likely that the coefficients $\langle \alpha_n \,|\, \psi\rangle$ are small and their contribution in the representation of $|\psi\rangle$ is negligible. This hap-

pens because the states $|\alpha\rangle$ correspond to the continuous spectrum and such states extend over very large regions of space. Therefore in Eq. 28 we would have a very localized state in the left-hand side and spatially extended states in the right-hand side. It appears that to reconcile the two terms in this equality the coefficients $\langle \alpha_n \,|\, \psi\rangle$ should be zero. This is a fine argument but it is not always correct. You will see later examples where we do represent localized states in terms of delocalized ones. However, in a large number of cases the contribution from the continuous states can be neglected and the representation of ψ reduces to

$$|\psi\rangle \simeq \sum_{n=1}^{\infty} |a_n\rangle \langle a_n \,|\, \psi\rangle \tag{29}$$

It is easy to show that if $|\psi\rangle$ is a bound state, the sequence

$$\{\langle a_1 \,|\, \psi\rangle,\ \langle a_2 \,|\, \psi\rangle, \ldots\} \equiv \{\langle a_n \,|\, \psi\rangle\}_{n=1}^{\infty}$$

belongs to the space $\boldsymbol{\ell}^2$ described in Chapter 1. The resolution of identity allows us to make a one-to-one connection between the elements of the abstract space of kets and the elements of the space $\boldsymbol{\ell}^2$. In Heisenberg versions of quantum mechanics, the state of a system was represented by elements of $\boldsymbol{\ell}^2$ while in the Dirac formulation it was represented by a ket. We see now that these two representations are equivalent. The element of $\boldsymbol{\ell}^2$ representing a given ket $|\psi\rangle$ depends on the states $|a_n\rangle$ chosen to represent $\hat{I}$. For any such choice there is a specific $\boldsymbol{\ell}^2$ space representing the ket.

I realize that this chameleon-like behavior of quantum theory is likely to cause anxiety. However, this is not that unusual in physics. Classical mechanics has also been formulated in a variety of ways, and we have Newton's equation, Hamilton's equations, Lagrange's equations, Hamilton's variational

principle, Hamilton-Jacobi equations, and whole set of canonical transformation that can make the same theory appear very different. The same situation appears in electrodynamics where various gauges can be used. The only difference is that in quantum mechanics, there were three versions of the theory and for a short while people did not know why they all gave correct results.

§ 9 *The energy representation.* We can use in Eq. 28 the pure states of any observable. However, not all pure states are equal and among them the pure states $|E_n\rangle$ of the energy are very popular. The main reason is that their time evolution is very easy to calculate.

If we use these states in Eq. 28 we obtain for $|\psi\rangle$ the expression

$$|\psi\rangle = \sum_{n=1}^{\infty} |E_n\rangle\langle E_n \,|\, \psi\rangle + \int_0^{\infty} \mathrm{d}\alpha \, |E_\alpha\rangle\langle E_\alpha \,|\, \psi\rangle \tag{30}$$

I included the integral because the energy spectrum has continuous values, which represent states in which the energy is so high that the system breaks up into fragments. The numbers $\langle E_n \,|\, \psi\rangle$ have a physical meaning. The probability that a system in state $|\psi\rangle$ has the energy E_n is $P_\psi(E_n) = |\langle E_n|\psi\rangle|^2$.

§ 10 *The coordinate representation.* The coordinate representation also plays a fundamental role because *it connects the abstract Dirac theory to the one proposed by Schrödinger.*

The completeness relation for the pure states of the position is (see Eq. 22)

$$\hat{I} = \int_{-\infty}^{+\infty} \mathrm{d}x \, |x\rangle\langle x| \tag{31}$$

Here $|x\rangle$ is a state in which we know with certainty that the particle is located at x. There is no sum over a discrete spectrum because position has only a continuous spectrum.

This equation has many important uses, most of them based on the fact that

$$\psi(x) \equiv \langle x \,|\, \psi \rangle$$

is the Schrödinger *wave function* of a system in state $|\psi\rangle$. The Schrödinger wave function satisfies the differential Schrödinger equation, which we can solve. Thus we can obtain *explicit expressions* for $\psi(x) \equiv \langle x \,|\, \psi \rangle$ and use them in further calculations.

Let us apply the bra $\langle x|$ to Eq. 30, which gives the expression of $|\psi\rangle$ in energy representation. We obtain

$$\langle x|\psi\rangle = \sum_{n=1}^{\infty} \langle x|E_n\rangle\langle E_n \,|\, \psi\rangle + \int_0^{\infty} \mathrm{d}\alpha \, \langle x|E_\alpha\rangle\langle E_\alpha \,|\, \psi\rangle \tag{32}$$

In the Schrödinger notation the expression $\langle x|E_n\rangle$ is written as $\phi_{E_n}(x)$ and the latter symbol is the wave function of a system that has the energy E_n and the position x. The symbol ϕ has no physical meaning and it is not necessary, but the usage of having a letter such as ϕ or ψ is deeply ingrained. Like most bad habits it is easier to live with it than to try to change it. Since $\langle E_n \,|\, x\rangle = \langle x \,|\, E_n\rangle^*$ we have $\langle E_n \,|\, x\rangle = \phi_{E_n}(x)^*$. With this notation Eq. 32 becomes

$$\psi(x) = \sum_{n=1}^{\infty} \phi_{E_n}(x)\langle E_n \,|\, \psi\rangle + \int_0^{\infty} \mathrm{d}\alpha \, \phi_{E_\alpha}(x)\langle E_\alpha \,|\, \psi\rangle \tag{33}$$

The scalar product $\langle E_n \,|\, \psi\rangle$ can be written as

$$\begin{aligned} \langle E_n \,|\, \psi\rangle &= \langle E_n \,|\, \hat{I}\psi\rangle = \int_{-\infty}^{+\infty} \mathrm{d}x \, \langle E_n \,|\, x\rangle \, \langle x \,|\, \psi\rangle = \int_{-\infty}^{+\infty} \mathrm{d}x \, \langle x \,|\, E_n\rangle^* \, \langle x \,|\, \psi\rangle \\ &= \int_{-\infty}^{+\infty} \phi_{E_n}(x)^* \, \psi(x) \, \mathrm{d}x \end{aligned} \tag{34}$$

In the manipulations made above I used Eq. 22, which is the completeness relation in terms of pure states of position. This procedure for calculating

the scalar product by using the coordinate representation is general. For two arbitrary kets $|\psi\rangle$ and $|\phi\rangle$ we have

$$\langle\phi\,|\,\psi\rangle = \int_{-\infty}^{+\infty} \langle\phi\,|\,x\rangle\langle x\,|\,\psi\rangle\,\mathrm{d}x = \int_{-\infty}^{+\infty} \phi(x)^*\psi(x)\,\mathrm{d}x \tag{35}$$

This is the formula for the scalar product used in Schrödinger's version of quantum theory. If we solve the Schrödinger equation to obtain the wave functions $\phi(x)$ and $\psi(x)$, we can use Eq. 34 to calculate the scalar product $\langle E_n\,|\,\psi\rangle$ which in turn can be used to calculate the probability that the system has the energy E_n. If the system is 30-dimensional (ten particles moving in three-dimensional space) then the equations presented here (valid for one particle moving in one dimension) will contain 30 integrals, one for each coordinate.

§ 11 *An example of using the energy and coordinate representations.* The energy of the particle in a box is an observable with a discrete spectrum and therefore the pure states of energy $|e_n\rangle$ satisfy the completeness relation

$$\sum_{n=1}^{\infty} |e_n\rangle\,\langle e_n| = \hat{I} \tag{36}$$

Denote the pure states of the energy of a harmonic oscillator by $|E_m\rangle$, $m \geq 0$. Using Eq. 36 I can write

$$|E_m\rangle = \hat{I}|E_m\rangle = \sum_{n=1}^{\infty} |e_n\rangle\,\langle e_n\,|\,E_m\rangle \tag{37}$$

I want to use numerical calculations to test this equation and explore more deeply its meaning and limitations. Testing how well this expansion works (if at all) is equivalent to testing the completeness relation Eq. 36.

I have warned you that the completeness relation written in terms of bras and kets appears more general than it is and should be used with care. You will see here an example that explains what I mean.

I will work with a system consisting of particle confined to move within a box of length L, whose walls are located at $x = 0$ and $x = L$. The harmonic oscillator vibrates around a position x_0. The kets $|e_n\rangle$ and $|E_m\rangle$ tell us the energy of the states *but give no position information*. The state $|\, e_n\rangle$ does not tell us where the box is located and how wide it is. The state $|E_n\rangle$ does not tell us the point around which the oscillator vibrates. This missing information is very important: if the box is in London and the oscillator is in New York, we cannot express the wave function of one as a linear combination of the wave functions of the other; the wave function of the particle in a box is zero outside London and that of the oscillator is zero outside New York. To construct the linear combination used in Eq. 28, we must pay attention to the position of the particle in the oscillator and that of the particle in the box. How do we do that?

We know the pure states of the harmonic oscillator and of the particle in a box in the coordinate representation (Schrödinger representation). You can find them in any introductory book on quantum mechanics. So let us act on Eq. 37 with the bra $\langle x|$ to introduce the position into this equation:

$$\langle x \,|\, E_m\rangle = \sum_{n=1}^{\infty} \langle x \,|\, e_n\rangle \, \langle e_n \,|\, E_m\rangle \tag{38}$$

Here x is the location of the particle and $\langle x \,|\, e_n\rangle$ is the wave function of the particle in a box of length L. If you look in an introductory book you will

find that

$$\langle x \,|\, e_n \rangle = \begin{cases} \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right), \ n = 1, 2, 3, \ldots & \text{if } x \in [0, L] \\ 0 & \text{if } x \notin [0, L] \end{cases} \tag{39}$$

This expression is valid for a box whose walls are located at $x = 0$ and $x = L$ (see Metiu, *Quantum Mechanics*, Eq. 8.25, page 103). Moreover, $\langle x \,|\, E_m \rangle$ is the wave function of a harmonic oscillator that has the energy E_m.

If we use Eq. 39 in Eq. 38, we obtain

$$\langle x \,|\, E_m \rangle = \begin{cases} \sum_{n=1}^{\infty} \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right) \langle e_n \,|\, E_m \rangle & \text{if } x \in [0, L] \\ 0 & \text{if } x \notin [0, L] \end{cases} \tag{40}$$

If we assume that we don't know the wave function $\langle x \,|\, E_m \rangle$ then we do not know the complex numbers $\langle e_n \,|\, E_n \rangle$. We can calculate them by forcing Eq. 38 to satisfy the Schrödinger equation for the harmonic oscillator. This will give us equations for the scalar products $\langle e_n \,|\, E_n \rangle$. If we solve those equations, we can use the results for $\langle e_n \,|\, E_n \rangle$ in Eq. 38 and obtain the wave function of the harmonic oscillator. We are going to use this method in a future chapter where numerical procedures for solving the Schrödinger equation are presented. Here we have a more modest goal: we want to test the expansion Eq. 38, which is equivalent to testing the completeness relation. The example will show that the abstract completeness equation $\hat{I} = \sum_{n=1}^{\infty} |e_n\rangle \langle e_n|$ could be very misleading if it is not used with care.

§ 12 *Trouble ahead.* As a result of the formal manipulations used in this chapter, we have written the wave function $\langle x \,|\, E_m \rangle$ of a harmonic oscillator as a sum of wave functions $\langle x \,|\, e_n \rangle$ of the particle in a box. Writing the

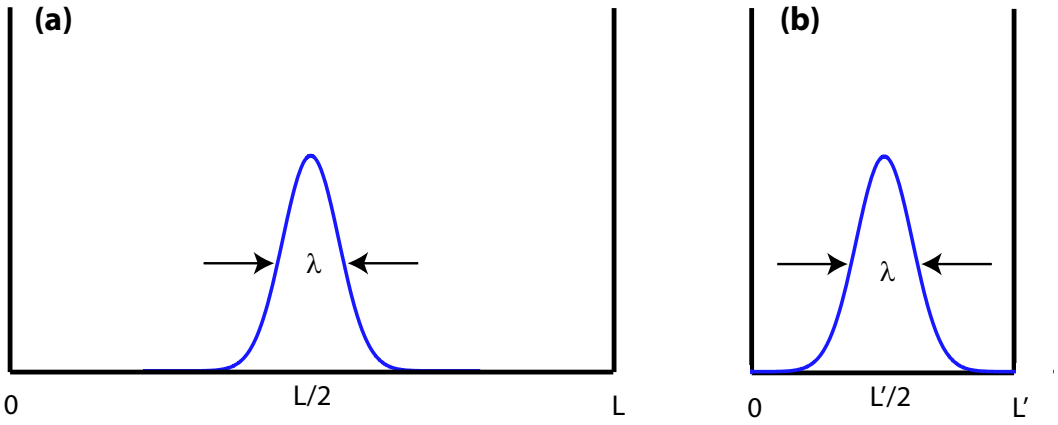

Figure 1: The oscillator ground-state wave function $\langle x \,|\, E_0 \rangle$ (a) in a box of width L and (b) in a smaller box, of width L'

---

expansion in coordinate representation makes it easy to see the limitations of the formal expansion $\sum_n |e_n\rangle \langle e_n| = \hat{I}$. If you examine the wave function $\langle x \,|\, E_m \rangle$ of a harmonic oscillator you will find that it is different from zero in the region $x \in [x_0 - 3\lambda, x_0 + 3\lambda]$ where $\lambda = \sqrt{\hbar/m\omega}$ is a length characterizing the oscillator (m is the mass and ω is the frequency). The particle-in-a-box wave function $\langle x \,|\, e_n \rangle$ differs from zero only if $x \in [0, L]$. Eq. 38 can be correct only if the region $x \in [x_0 - 3\lambda, x_0 + 3\lambda]$ is inside the region $x \in [0, L]$. The amplitude of the oscillator must be contained inside the box. To satisfy this condition, we place the position x_0 (around which the particle oscillates) in the middle of the box. Moreover, we must ensure that the box is large enough to accommodate the wave function of the oscillator. This means that we must have:

$$x_0 = L/2 \tag{41}$$

and

$$L > 2 \times 3\lambda \tag{42}$$

The box-size L must be larger than the largest amplitude the oscillator can have. The graph of $\langle x \,|\, E_0 \rangle$ and that of the box are shown in Fig. 1 for two boxes of lengths L and L'. Both boxes are large enough to contain the wave function of the ground state of the oscillator (the bell-shaped curve). We will see later that if we are interested only in the ground state of the oscillator then the box in Fig. 1(b) is the best choice. However that would not be a good choice if we want to represent the excited states of the oscillator because they would be wider than the box and therefore the representation in terms of particle-in-a-box wave function will *never* give a correct representation of the oscillator wave function. On the other hand, if the box is too large (as in Fig. 1(a)), we obtain an accurate representation only if we take a very large number of terms in the sum.

The formal expansion

$$|E_m\rangle = \sum_{n=1}^{\infty} |e_n\rangle \langle E_n \,|\, E_m\rangle \tag{43}$$

(based on the formal equation $\hat{I} = \sum_{n=1}^{\infty} |e_n\rangle \langle e_n|$) gave no hint that such conditions must be imposed. They become apparent only when we convert this equation to the coordinate representation and give some thought to what we are trying to do. This is why I warned you that a careless use of the resolution of identity $\sum_n |e_n\rangle \langle e_n| = \hat{I}$ can lead to absurd conclusions, not just poor approximations.

As I alluded to earlier, the choice of L, for a given λ, is not without peril. In principle, either box in Fig. 1 is an appropriate choice, but box (b) is

better in practice. To see why I say that, take a look at Fig. 2, which shows graphs of $\langle x \,|\, E_0 \rangle$, which is the function we want to represent, and of $\langle x \,|\, e_n \rangle$, $n = 1, 2, 3$, which are the functions used in the representation (in the sum in the right-hand side of Eq. 38). The wave function of the oscillator (for the lowest energy), namely $\langle x \,|\, E_0 \rangle$, is the dark blue, bell-shaped curve. This function is equal to zero when $x \in [0, 0.1]$ and $x \in [0.4, 0.5]$. The other three curves are the wave functions of the particle in the box $\langle x \,|\, e_n \rangle$, for $n = 1, 2, 3$. None of these functions is zero in those regions of x in which the oscillator wave function is zero. It would seem that we cannot represent $\langle x \,|\, E_0 \rangle$ in the regions [0,0.1] and [0.4,0.5] as a sum of the functions $\langle x \,|\, e_n \rangle$. This is a false impression. The expression $\sum_n \langle x \,|\, e_n \rangle \, \langle e_n \,|\, E_0 \rangle$ can be equal to zero in these ranges of x if some of the coefficients $\langle e_n \,|\, E_0 \rangle$ are negative and some are positive, so that the terms in the sum *cancel each other* when $x \in [0, 0.1]$ or $x \in [0.4, 0.5]$. This may not seem likely, but it happens. We will revisit this situation soon. We will refer to this cancellation as *destructive interference* between the wave functions of the particle in the box.

Now let us go back to the two boxes in Fig. 1. The region of values of x over which the sum must equal zero is larger for the box at the left, and the cancellation by destructive interference is harder to achieve than in the case of the smaller box (at the right of Fig. 1). If we use the box having a larger value of L, we will need more terms to represent $\langle x \,|\, E_0 \rangle$ than when we use the box of length $L' < L$. On the other hand the choice of the larger length is necessary if we want to represent well the excited states of the oscillator, since in these states the oscillator has higher energy and travels further than in the ground state.

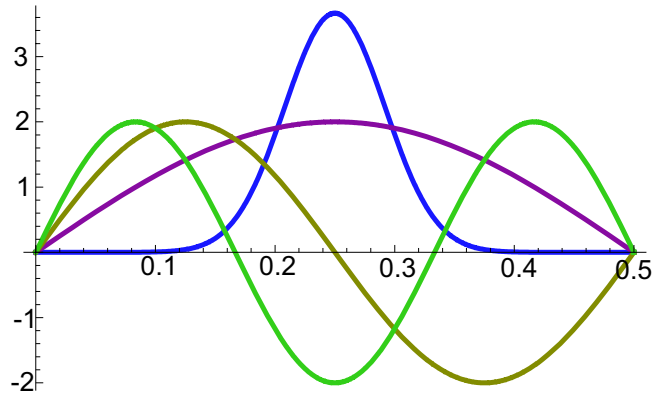

Figure 2: The blue curve is the ground state wave function $\langle x \,|\, E_0 \rangle$ of the harmonic oscillator. Purple: $\langle x \,|\, e_1 \rangle$; yellow: $\langle x \,|\, e_2 \rangle$; green: $\langle x \,|\, e_3 \rangle$. I used $L = 0.5$ Å and $\lambda = 0.042$ Å.

All these statements are based on common sense: they do not involve mathematical sophistication but require careful and practical thinking. They also display why one cannot use the completeness relation mechanically. Numerical analysis is not a field suitable for impractical people who rely solely on logic, especially when the logic is used without the necessary rigor.

§ 13 *A test of the representation theory.* Now that we understand what we are up against, let us test Eq. 38. The wave function of the harmonic oscillator is given by (see Metiu, *Quantum Mechanics*, Eqs. 17.8–17.11, pages 272–273)

$$\langle x \,|\, E_m \rangle = \sqrt{\frac{1}{\lambda}} \left(\frac{1}{\pi^{1/4} \sqrt{2^m m!}} \right) \exp \left[-\frac{(x - x_0)^2}{2\lambda^2} \right] H_m \left(\frac{x - x_0}{\lambda} \right), \; m = 0, 1, 2, \ldots \tag{44}$$

with

$$\lambda = \sqrt{\frac{\hbar}{\mu\omega}} \tag{45}$$

and

$$x_0 = \frac{L}{2} \tag{46}$$

$H_m(x)$ is the Hermite polynomial of order m, and λ is a length.

The test of Eq. 38, for $m = 0$, proceeds as follows:

1. Calculate $\langle e_n \,|\, E_0 \rangle$ by using

$$\langle e_n \,|\, E_0 \rangle = \int_{-\infty}^{+\infty} \langle e_n \,|\, x \rangle \, \langle x \,|\, E_0 \rangle \, \mathrm{d}x \tag{47}$$

2. Test whether the right-hand side of Eq. 38 is equal to the left-hand side (which is given by Eqs. 44–46), by plotting them on the same graph.

Using Eq. 39 for $\langle e_n \,|\, x \rangle$ and Eq. 44 for $\langle x \,|\, E_0 \rangle$, we evaluate $\langle e_n \,|\, E_0 \rangle$ by performing the integral in Eq. 47. `Mathematica` gives (see WorkBook3_Representation theory.nb)

$$\begin{aligned} \langle e_n \,|\, E_0 \rangle &= -\frac{\mathrm{i}}{2} \exp\left[-\frac{n\pi(\mathrm{i}L^2 + n\pi\lambda^2)}{2L^2}\right] \left(-1 + \mathrm{e}^{\mathrm{i}n\pi}\right) \pi^{1/4} \sqrt{\frac{\lambda}{L}} \\ &\quad \times \left(\mathrm{Erf}\left[\frac{L^2 - 2\mathrm{i}n\pi\lambda^2}{2\sqrt{2}\,L\lambda}\right] + \mathrm{Erf}\left[\frac{L^2 + 2\mathrm{i}n\pi\lambda^2}{2\sqrt{2}\,L\lambda}\right] \right) \end{aligned} \tag{48}$$

This looks very complicated, but it poses no serious problems if we use a computer. The Erf function is calculated by `Mathematica` automatically when the argument is a number. This result for $\langle e_n \,|\, E_0 \rangle$ seems to be a complex number but this is not: n is an integer, so $\langle e_n \,|\, E_0 \rangle$, given by Eq. 48, is a real number. Here are the values of $\langle e_n \,|\, E_0 \rangle$ for $L = 0.5$ and $n = 1, 2, \ldots, 10$ (see WorkBook3_Representation theory.nb):

$$0.7453,\ 0,\ -0.5641,\ 0,\ 0.3232,\ 0,\ -0.1401,\ 0,\ 0.04568,\ 0$$

That is, $\langle e_1 \,|\, E_0\rangle = 0.7453$, $\langle e_2 \,|\, E_0\rangle = 0$, and so on. The fact that the coefficients get smaller and smaller is a good sign: it signals that the sum in Eq. 38 seems to be convergent.

Now that I have the coefficients $\langle e_n \,|\, E_0\rangle$, I can write Eq. 38, for $x \in [0, L]$, as

$$\begin{aligned}\langle x \,|\, E_0\rangle &= \sum_{n=1}^{\infty} \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right) \left(-\frac{\mathrm{i}}{2}\right) \exp\left[-\frac{n\pi(\mathrm{i}L^2 + n\pi\lambda^2)}{2L^2}\right] \left(-1 + \mathrm{e}^{\mathrm{i}n\pi}\right) \\ &\quad \times \pi^{1/4} \sqrt{\frac{\lambda}{L}} \left(\mathrm{Erf}\left[\frac{L^2 - 2\mathrm{i}n\pi\lambda^2}{2\sqrt{2}\,L\lambda}\right] + \mathrm{Erf}\left[\frac{L^2 + 2\mathrm{i}n\pi\lambda^2}{2\sqrt{2}\,L\lambda}\right]\right) \qquad (49)\end{aligned}$$

No one would ever write down this equation by accident or by pure contemplation. If it is correct then there is something deep about the completeness relation.[2]

The formal theory developed here tells us that we should take an infinite number of terms in our representation. Of course, no computer can handle that. In practice we can take only a finite number of terms. One of the most important questions in such representations is how many terms do we need and how do we know when we have enough of them? In Fig. 3, I compare the sum of the first twenty terms in Eq. 49 with the ground state wave function of

[2]Those of you who know mathematics will recognize this expression as a truncated Fourier series, discovered by Jean Baptiste Joseph Fourier in 1822. We arrived at this result by using physical arguments about probability in quantum mechanics (to "derive" the completeness relation) and the fact that the energy of a particle in a box is an observable. Before getting too pleased with ourselves, we need to remember that our manipulations do not make a mathematical proof. We did not examine convergence, nor did we establish clearly what kind of functions can be expressed as a sum of sine functions. Nevertheless our "derivation" is a remarkably quick (if "dirty") way of suggesting interesting mathematical connections.

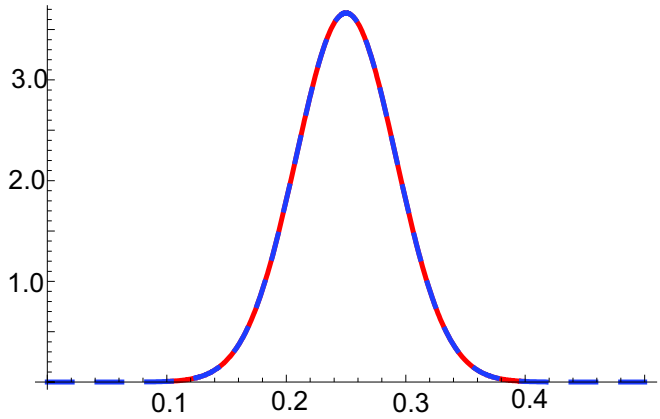

Figure 3: The harmonic oscillator ground state wave function is shown as a dashed blue line. The sum representing it, with 20 terms, is shown as a solid red line. $L = 0.5$ Å, $\lambda = 0.042$ Å. The plot was made in Section I, Cell 5 of WorkBook3_Representation theory.nb.

---

the harmonic oscillator. Obviously the representation of $\langle x \,|\, E_0 \rangle$ by the sum is excellent. But what happens if we take only three terms? The outcome is shown in Fig. 4. It is not a good fit. The peak in the middle is not well developed and the functions in the sum do not cancel each other at the edges of the interval.

When I chose the parameters, I took $L = 0.5$ Å and $\lambda = 0.042$ Å. In Fig. 2 you can see that the box is much wider than the region where the oscillator wave function differs from zero. It seems that I might do better if I take instead $L = 0.4$ Å, to narrow the box. You can see from Fig. 5 that this is the case.

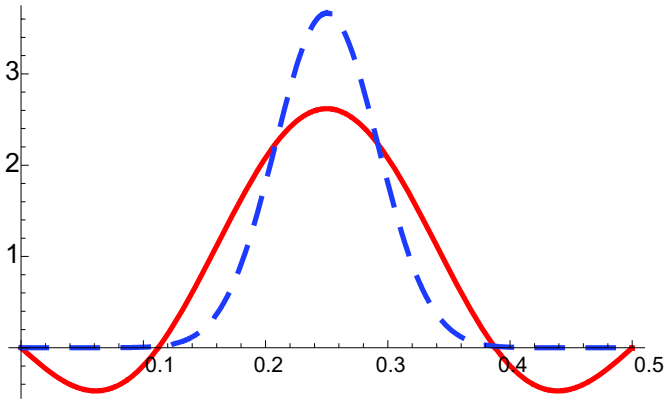

Figure 4: The solid red line is the sum in Eq. 49 with three terms. The dashed blue line is the ground state wave function of the harmonic oscillator. $L = 0.5$ Å, $\lambda = 0.042$ Å. The plot was made in Cell 5 of WorkBook3_Representation theory.nb.

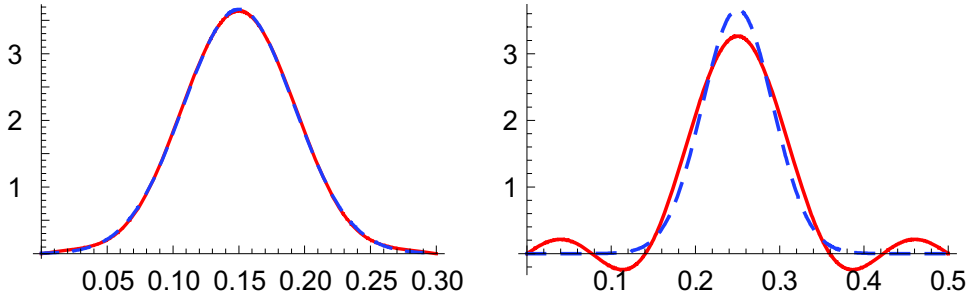

Figure 5: The left panel shows the ground state harmonic oscillator wave function (dashed, blue) and the six-term sum (solid, red), taken with $L = 0.4$ Å. The right panel shows the result when $L = 0.5$ Åand the same number of terms in the sum. A small box, but not too small, is better than a large one.

§ 14 *A few general remarks.* We have talked about the space of kets in general, but now we emphasize again that the pure states of *different* systems generate *different* spaces. The space generated by the pure states of the energy of a particle in a box of length L is different from that generated by the states for a box of length $L' \neq L$. The pure energy states of a particle in a box whose walls are located at $x = -L/2$ and $x = L/2$ generate a different space than do the states of a particle in the same-size box with walls at $x = 0$ and $x = L$.

Exercise 1 $|e_n\rangle$ are the pure states of the energy for a particle in a box of length $L = 0.5$ Å whose walls are located at $x = 0$ and $x = 0.5$. $|\varepsilon_n\rangle$ are the pure states of the energy for a particle in a box of length 0.4 Å with walls located at $x = 0$ and $x = 0.4$. Formally, we have

$$\sum_{n=1}^{\infty} |e_n\rangle \langle e_n| = \hat{I} \tag{50}$$

and

$$\sum_{n=1}^{\infty} |\varepsilon_n\rangle \langle \varepsilon_n| = \hat{I} \tag{51}$$

(a) I can *formally* write

$$|e_m\rangle = \sum_{n=1}^{\infty} |\varepsilon_n\rangle \langle \varepsilon_n \,|\, e_m\rangle \tag{52}$$

and

$$|\varepsilon_m\rangle = \sum_{n=1}^{\infty} |e_n\rangle \langle e_n \,|\, \varepsilon_m\rangle \tag{53}$$

Are these equations correct?

(b) Check whether the right-hand side of

$$\langle x \,|\, \varepsilon_1\rangle \cong \sum_{n=1}^{N} \langle x \,|\, e_n\rangle \langle e_n \,|\, \varepsilon_1\rangle \tag{54}$$

gives a good approximation to the left-hand side. What do you have to say about

$$\langle x \,|\, e_1 \rangle \cong \sum_{n=1}^{N} \langle x \,|\, \varepsilon_n \rangle \, \langle \varepsilon_n \,|\, e_1 \rangle \ ? \tag{55}$$

Exercise 2 Represent the first excited state of a harmonic oscillator as a sum of eigenfunctions of the particle in a box.

The danger of using the completeness relation with blind faith is illustrated dramatically when we consider the spin states of one electron. There are only two of them, denoted by $|\uparrow\rangle$ and $|\downarrow\rangle$. The completeness relation they generate is

$$\hat{I} = |\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow| \tag{56}$$

It would be idiotic to write

$$|E_n\rangle = \hat{I}|E_n\rangle = |\uparrow\rangle\langle\uparrow\ |E_n\rangle + |\downarrow\rangle\langle\downarrow\ |E_n\rangle$$

if $|E_n\rangle$ is a pure energy state of a particle in a box. The kets $|E_n\rangle$ belong to a different space than do $|\uparrow\rangle$ and $|\downarrow\rangle$. However, if I perform an electron-spin resonance (ESR) experiment with an organic radical that has one electron with unpaired spin (e.g. CH_3), then I can write any spin state $|\psi\rangle$ created by the experiment as

$$|\psi\rangle = |\uparrow\rangle\langle\uparrow\ |\psi\rangle + |\downarrow\rangle\langle\downarrow\ |\psi\rangle \tag{57}$$

The state $|\psi\rangle$ of the spin of one electron in a molecule belongs to the space generated by the set $\{|\uparrow\rangle, |\downarrow\rangle\}$.

Section 3.3. Generalization: Basis Set

§ 15 *Summary.* The main result of the previous section is that a ket $|\psi\rangle$ describing the state of a *bound* quantum system can be expanded as

$$|\psi\rangle \cong \sum_{n=1}^{N} |a_n\rangle \langle a_n \,|\, \psi\rangle \tag{58}$$

This follows from the completeness relation

$$\hat{I} \cong \sum_{n=1}^{N} |a_n\rangle \langle a_n| \tag{59}$$

The "proof" of the completeness relation made use of the fact that the kets $|a_n\rangle$ are pure states of an observable. Note, however, that the validity of the representation Eq. 58 depends only on the validity of the relation Eq. 59. We may ask therefore whether it is possible to construct an orthonormal set of kets that satisfy the completeness relation but are not pure states of an observable?

An equivalent question can be posed within the Schrödinger representation, where Eq. 58 becomes

$$\psi(x) \cong \sum_{n=1}^{N} c_n a_n(x) \tag{60}$$

with $\psi(x) \equiv \langle x \,|\, \psi\rangle$, $a_n(x) \equiv \langle x \,|\, a_n\rangle$, and $c_n \equiv \langle a_n \,|\, \psi\rangle$. Can we find an orthonormal set of functions $a_n(x)$ such that the wave function $\psi(x)$ is represented well by Eq. 60? If the answer is affirmative (and it is), can we use this representation in the same way as the one based on expansion in pure states? This would give us more flexibility in solving numerically the equations of quantum mechanics.

In what follows, we show how such a set can be constructed.

§ 16 *A simple example of a basis set.* You encountered the concept of basis set when you studied vector algebra. Much of what you learn about abstract linear spaces is a generalization of results obtained in the linear space of the three-dimensional vectors (which we call R^3). This is why I review the use of a basis set in this simple linear space. When you studied three-dimensional vectors you showed that any three-dimensional vector $\vec{r}$ can be written as

$$\vec{r} = \alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2 + \alpha_3 \vec{e}_3 \tag{61}$$

The three vectors $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ have unit length, and are oriented along the axes of coordinates (see Fig. 6). This means that they are perpendicular to each other. We say that the set of vectors $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ generates the space R^3 or that it is a complete orthonormal basis set in R^3. Because the basis set is orthonormal $\vec{e}_i \cdot \vec{e}_j = 0$ if $i \neq j$ (orthogonality) and $\vec{e}_i \cdot \vec{e}_i = 1$ (unit length). Here the dot product $\vec{e}_i \cdot \vec{e}_j$ is the scalar product in R^3. Eq. 60 is a generalization of Eq. 61 to the linear space of functions.

§ 17 *Gram-Schmidt orthogonalization.* The basis set $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ in R^3, given in §16, is easy to understand. In particular, we have no difficulty in picking the vectors $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ so that they are perpendicular to each other (orthogonal) and have unit length (normalized). The construction of an orthonormal basis set is not as simple in a more general space such as L^2 or $\boldsymbol{\ell}^2$. In what follows I will show you a general scheme, called Gram-Schmidt orthogonalization, that takes an arbitrary set of vectors in a general linear space and converts them into a new set whose vectors are orthonormal.

In terms of kets, the task is to start with a set $\{|o_1\rangle, |o_2\rangle, \ldots, |o_N\rangle\}$ and generate a new set $\{|\nu_1\rangle, |\nu_2\rangle, \ldots, |\nu_N\rangle\}$ whose kets satisfy the orthonormality

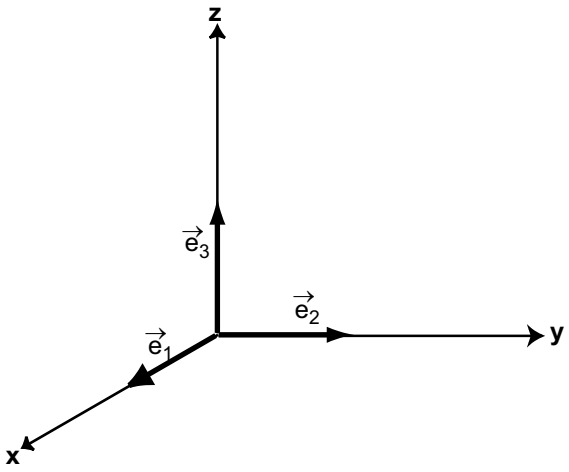

Figure 6: An orthonormal basis in R^3

---

relation

$$\langle \nu_i \,|\, \nu_j \rangle = \delta_{ij}, \quad i,j = 1,2,\ldots,N \tag{62}$$

Here is the algorithm that does that.

$$|n_1\rangle = |o_1\rangle \tag{63}$$

$$|n_2\rangle = |o_2\rangle - |n_1\rangle \frac{\langle n_1 \,|\, o_2 \rangle}{\langle n_1 \,|\, n_1 \rangle} = \left[\hat{I} - \frac{|n_1\rangle \langle n_1|}{\langle n_1 \,|\, n_1 \rangle} \right] |o_2\rangle \tag{64}$$

$$|n_3\rangle = |o_3\rangle - |n_1\rangle \frac{\langle n_1 \,|\, o_3 \rangle}{\langle n_1 \,|\, n_1 \rangle} - |n_2\rangle \frac{\langle n_2 \,|\, o_3 \rangle}{\langle n_2 \,|\, n_2 \rangle} \tag{65}$$

Or, in general

$$|n_j\rangle = \left[\hat{I} - \sum_{i=1}^{j-1} \frac{|n_i\rangle \langle n_i|}{\langle n_i \,|\, n_i \rangle} \right] |o_j\rangle, \;\; j = 1,2,\ldots,N \tag{66}$$

You can verify by direct calculation that

$$\langle n_i \,|\, n_j \rangle = 0 \text{ for } i \neq j; \tag{67}$$

that is, the set $\{|n_1\rangle, |n_2\rangle, \ldots, |n_N\rangle\}$ is orthogonal. For example,

$$\langle n_1 \,|\, n_2\rangle = \langle n_1 \,|\, o_2\rangle - \frac{\langle n_1 \,|\, n_1\rangle \,\langle n_1 \,|\, o_2\rangle}{\langle n_1 \,|\, n_1\rangle} = 0$$

It is also easy to see that the kets

$$|\nu_i\rangle \equiv \frac{|n_i\rangle}{\sqrt{\langle n_i \,|\, n_i\rangle}} \tag{68}$$

satisfy

$$\langle \nu_i \,|\, \nu_j\rangle = \delta_{ij} \tag{69}$$

The set $\{|\nu_1\rangle, |\nu_2\rangle, \ldots, |\nu_N\rangle\}$ is orthonormal!

§ 18 *An example: constructing an orthonormal basis set in* R^3. The calculations are performed with `Mathematica` (they could be done "by hand" but this is tedious) and you should read Section II, Cell 1 of WorkBook3. Here I give a summary of the results. I used a random number generator to create three vectors in R^3:

$$|o_1\rangle = \{0.445, -0.781, -0.059\} \tag{70}$$

$$|o_2\rangle = \{0.071, 0.166, -0.412\} \tag{71}$$

$$|o_3\rangle = \{-0.670, -0.202, 0.508\} \tag{72}$$

I use here the ket notation instead of the customary $\vec{o}_1$, etc.

The scalar product in R^3 is the dot product. I calculated $\langle o_i \,|\, o_j\rangle$ (the dot product of $|o_i\rangle$ with $|o_j\rangle$) and found that these vectors are not orthonormal.

I use the Gram-Schmidt procedure to convert them to a set $\{|n_1\rangle, |n_2\rangle, |n_3\rangle\}$ of vectors that are perpendicular to each other. The first step in the Gram-Schmidt procedure is

$$|n_1\rangle = |o_1\rangle = \{0.445, -0.781, -0.059\} \tag{73}$$

The second vector (supposed to be orthogonal to the first) is

$$|n_2\rangle = |o_2\rangle - \frac{|o_1\rangle}{\langle o_1 \,|\, o_1\rangle}\langle n_1 \,|\, o_2\rangle \tag{74}$$

`Mathematica` easily evaluates the right-hand side, to give

$$|n_2\rangle = \{0.112, 0.095, -0.417\} \tag{75}$$

The third vector, $|n_3\rangle$, is given by Eq. 65. `Mathematica` evaluates that to

$$|n_3\rangle = \{-0.251, -0.135, -0.098\} \tag{76}$$

Next I tested that $\langle n_i \,|\, n_j\rangle = 0$ if $i \neq j$. That turns out to be true, which means that Gram-Schmidt does the job it promised. Then I used the formula $\langle \nu_i\rangle = \frac{|n_i\rangle}{\sqrt{\langle n_i \,|\, n_i\rangle}}$ to create the normalized vectors $|\nu_1\rangle, |\nu_2\rangle, |\nu_3\rangle$, out of $|n_1\rangle, |n_2\rangle, |n_3\rangle$.

The Gram-Schmidt procedure has a simple geometric interpretation. We can think of $|o_2\rangle$ as having two components, one parallel to $|n_1\rangle$ and one perpendicular to it. The component parallel to $|n_1\rangle$ has the form $\alpha|n_1\rangle$, where α is an unknown number. To find the component perpendicular to $|n_1\rangle$, which is

$$|n_2\rangle = |o_2\rangle - \alpha|n_1\rangle, \tag{77}$$

we just remove the part that is along $|n_1\rangle$. The perpendicularity condition allows us to determine α:

$$0 = \langle n_1 \,|\, n_2\rangle = \langle n_1 \,|\, (|o_2\rangle - \alpha|n_1\rangle)\rangle = \langle n_1 \,|\, o_2\rangle - \alpha\langle n_1 \,|\, n_1\rangle \tag{78}$$

Rearranging this equation we find

$$\alpha = \frac{\langle n_1 \,|\, o_2\rangle}{\langle n_1 \,|\, n_1\rangle} \tag{79}$$

Inserting this value in Eq. 77 leads to the Gram-Schmidt formula Eq. 64.

Exercise 3 Use a similar geometric argument to derive Eq. 65.

---

§ 19 *If the original vectors are linearly dependent.* When developing a basis set, we must be careful that none of the basis-set vectors (or kets, in general) depends on the others. To explain what is involved I use the space of three-dimensional vectors as an example. If the basis set is $\{|o_1\rangle, |o_2\rangle, |o_3\rangle\}$ (I use the ket notation instead of the conventional one), we do not want to have

$$|o_3\rangle = a|o_1\rangle + b|o_2\rangle \tag{80}$$

It is intuitively clear that such a relationship causes a problem. If $|o_1\rangle$ is along the x-axis and $|o_2\rangle$ is along the y-axis then $|o_1\rangle$, $|o_2\rangle$, *and* $|o_3\rangle$ lie in the XOY plane. These three vectors provide a basis set only for the vectors lying in the XOY plane. In addition, a basis set for the vectors lying in the XOY plane should have only two vectors; the third is superfluous.

When we have a relationship like Eq. 80, we say that $|o_3\rangle$ is linearly dependent on $|o_1\rangle$ and $|o_2\rangle$. The question is how the Gram-Schmidt procedure will behave when there is such a linear dependence among the set of vectors that we start with?

In Cell 1 of Section II of WorkBook3, I applied the Gram-Schmidt procedure to the set

$$|o_1\rangle \;=\; \{0.445, -0.781, -0.059\} \tag{81}$$

$$|o_2\rangle = \{0.071, 0.166, -0.412\} \quad (82)$$

$$|o_3\rangle = a|o_1\rangle + b|o_2\rangle \quad (83)$$

where a and b are unknown numbers. In this set, $|o_3\rangle$ is linearly dependent on $|o_1\rangle$ and $|o_2\rangle$.

The result of the calculations is

$$|n_1\rangle = |o_1\rangle = \{0.445, -0.781, -0.059\} \quad (84)$$

$$|n_2\rangle = \{0.112, 0.095, -0.417\} \quad (85)$$

$$|n_3\rangle = 0 \quad (86)$$

The Gram-Schmidt procedure "knows" that there is something wrong with $|o_3\rangle$ and makes $|n_3\rangle = 0$. If you examine the algorithm closely you will see that this behavior is general. If a vector $|o_k\rangle$ can be expressed in the form

$$|o_k\rangle = \sum_{j=1}^{k-1} \alpha_j |o_j\rangle$$

then one of the "new" vectors, generated by the Gram-Schmidt procedure, will be equal to zero.

This behavior makes perfect sense. If $|o_1\rangle$ is along OX and $|o_2\rangle$ is along OY, then $|o_3\rangle = a|o_1\rangle + b|o_2\rangle$ is contained in the XOY plane. The Gram-Schmidt procedure calculates two vectors perpendicular to each other and contained in the XOY plane and attempts to combine the "old" vectors to make a vector perpendicular to XOY. However, the only such vector that can be constructed from the starting vectors (which are on in the XOY plane) is the vector equal to zero.

The moral: none of the n vectors in the starting set $\{|o_1\rangle, |o_2\rangle, \ldots, |o_N\rangle\}$ should be linearly dependent on the others. If one of them is, then the Gram-

Schmidt procedure will create an orthonormal basis set of n-1 vectors and produce a vector equal to zero

§ 20 *An example of constructing an orthonormal basis set in* L^2. The Gram-Schmidt procedure is very general and it can be applied to any linear space. For illustration I construct an orthonormal basis set for the space of the functions $f(x)$, $x \in [0, \pi]$ that are differentiable and satisfy

$$\int_0^\pi f(x)^2 \, \mathrm{d}x < \infty \tag{87}$$

This space can be used to describe the wave function of a particle whose position is confined (by the forces acting on it) between $x = 0$ and $x = \pi$. The wave functions of a particle in a box having infinite walls at $x = 0$ and $x = \pi$ are contained in this L^2 space, as are the wave functions of a harmonic oscillator vibrating around $x = \pi/2$ whose maximum amplitude is smaller than $\pi/2$. More important is the fact that basis sets are essential to most methods for numerical solution of the Schrödinger equation. For ambitious calculations a good basis set is essential since it controls both the quality of the results and the amount of computer power required.

I define the scalar product to be

$$\langle f \,|\, \phi \rangle = \int_0^\pi f(x)\phi(x) \, \mathrm{d}x \tag{88}$$

For simplicity, I restrict the space to contain only functions taking real values.

I start with the "old" functions

$$\langle x \,|\, o_i \rangle \equiv o_i(x) = x^{i-1}, \; i = 1, \ldots, N \tag{89}$$

If I use these functions as a basis set, I would write, for any function $|f\rangle$ in the linear space,

$$|f\rangle = \sum_{i=1}^{N} \alpha_i |o_i\rangle \tag{90}$$

where α_i are real numbers. In Schrödinger representation, this means

$$\langle x \,|\, f\rangle \equiv f(x) = \sum_{i=1}^{N} \alpha_i \langle x \,|\, o_i\rangle = \sum_{i=1}^{N} \alpha_i o_i(x) = \sum_{i=1}^{N} \alpha_i x^{i-1} \tag{91}$$

We are attempting to fit the function $f(x)$ with a sum of specially constructed polynomials. There are many methods in numerical analysis for fitting a function with a polynomial. Here I am using the Gram-Schmidt method to illustrate how it works. It will turn out that the polynomials generated this way provide a better representation than the one obtained by performing a Taylor expansion of the function (which also represents the function as a polynomial).

The set $\{o_i(x)\}_{i=1}^{N}$ is not orthonormal. Let us start with it and construct an orthonormal set $\{n_i(x)\}_{i=1}^{N}$. The simple calculations needed for this are performed in Section II, Cell 2, of the `Mathematica` file WorkBook3.nb.

Using Eq. 63,

$$|n_1\rangle = |o_1\rangle$$

which means that

$$\langle x \,|\, n_1\rangle = \langle x \,|\, o_1\rangle = 1 \tag{92}$$

Using Eq. 64:

$$\langle x \,|\, n_2\rangle = \langle x \,|\, o_2\rangle - \langle x \,|\, n_1\rangle \frac{\langle n_1 \,|\, o_2\rangle}{\langle n_1 \,|\, n_1\rangle} \tag{93}$$

with

$$\langle n_1 \,|\, o_2\rangle \;\; = \;\; \int_0^{\pi} \langle n_1 \,|\, x\rangle \, \langle x \,|\, o_2\rangle \, \mathrm{d}x$$

$$= \int_0^\pi n_1(x) o_2(x)\,\mathrm{d}x = \int_0^\pi 1 \times x\,\mathrm{d}x = \pi^2/2 \tag{94}$$

$$\langle n_1 \,|\, n_1 \rangle = \int_0^\pi \langle n_1 \,|\, x \rangle \langle x \,|\, n_1 \rangle\,\mathrm{d}x$$

$$= \int_0^\pi n_1(x) n_1(x)\,\mathrm{d}x = \int_0^\pi \mathrm{d}x = \pi \tag{95}$$

Using Eqs. 94 and 95 in Eq. 93 gives

$$n_2(x) = o_2(x) - \frac{\pi}{2} = x - \frac{\pi}{2} \tag{96}$$

The third function is (use Eq. 65 with $j = 3$)

$$\langle x \,|\, n_3 \rangle = o_3(x) - \langle x \,|\, n_1 \rangle \frac{\langle n_1 \,|\, o_3 \rangle}{\langle n_1 \,|\, n_1 \rangle} - \langle x \,|\, n_2 \rangle \frac{\langle n_2 \,|\, o_3 \rangle}{\langle n_2 \,|\, n_2 \rangle}$$

$$= \frac{\pi^2}{6} - \pi x + x^2 \tag{97}$$

Similarly, we obtain (see WorkBook3)

$$\langle x \,|\, n_4 \rangle \equiv n_4(x) = -\frac{\pi^3}{20} + \frac{3\pi^2}{5} x - \frac{3\pi}{2} x^2 + x^3 \tag{98}$$

We can go on and calculate as many "new" basis set functions (i.e. orthogonal polynomials) as we need by following the recipe.

You can verify, by performing the integrals, that

$$\langle n_i \,|\, n_j \rangle = \int_0^\pi \langle n_i \,|\, x \rangle \langle x \,|\, n_j \rangle\,\mathrm{d}x = \int_0^\pi n_i(x) n_j(x)\,\mathrm{d}x = 0 \text{ when } i \neq j \tag{99}$$

The "new" functions $\{n_i(x)\}_{i=1}^N$ are orthogonal. However, they are not normalized. To normalize them we use (see Eq. 68)

$$\nu_i(x) \equiv \frac{n_i(x)}{\sqrt{\langle n_i \,|\, n_i \rangle}} \tag{100}$$

The result is (see WorkBook3)

$$\nu_1(x) = \frac{1}{\sqrt{\pi}} \tag{101}$$

$$\nu_2(x) = -\sqrt{\frac{3}{\pi}} + \frac{2\sqrt{3}}{\pi^{3/2}}\,x \tag{102}$$

$$\nu_3(x) = \sqrt{\frac{5}{\pi}} - \frac{6\sqrt{5}}{\pi^{3/2}}\,x + \frac{6\sqrt{5}}{\pi^{5/2}}\,x^2 \tag{103}$$

$$\nu_4(x) = -\sqrt{\frac{7}{\pi}} + \frac{12\sqrt{7}}{\pi^{3/2}}\,x - \frac{30\sqrt{7}}{\pi^{5/2}}\,x^2 + \frac{20\sqrt{7}}{\pi^{7/2}}\,x^3 \tag{104}$$

Of course, you can go on and generate more functions. Here I wanted to illustrate the method and give you a chance to see what the orthonormal polynomials look like.

We can now expand any function in our L^2 space in terms of this orthonormal basis set. As an example, let us expand the function

$$\langle x \,|\, f\rangle \equiv f(x) \equiv x^2 \sin(x) \tag{105}$$

The expansion is

$$|f\rangle \cong \sum_{i=1}^{m} a_i \,|\nu_i\rangle \tag{106}$$

Acting with $\langle\nu_j|$ on Eq. 106 gives

$$\langle \nu_j \,|\, f\rangle \cong \sum_{i=1}^{m} a_i \,\langle \nu_j \,|\, \nu_i\rangle = \sum_{i=1}^{m} a_i \,\delta_{ji} = a_j \tag{107}$$

It is in this equation that the orthonormality of the basis set plays a very important role.

Exercise 4 Examine what would happen if the basis set was not orthogonal. Explain why the calculations would be more difficult if that were the case.

If we introduce the expression $\langle \nu_i \,|\, f\rangle = a_i$ into Eq. 106, we have

$$|f\rangle \cong \sum_{i=1}^{m} \langle \nu_i \,|\, f\rangle \,|\nu_i\rangle = \left(\sum_{i=1}^{m} |\nu_i\rangle \,\langle \nu_i|\right) |f\rangle \tag{108}$$

which implies that

$$\sum_{i=1}^{m} |\nu_i\rangle \langle\nu_i| \cong \hat{I} \tag{109}$$

where $\hat{I}$ is the unit operator.

The assumption, that the orthonormal basis set $\{|\nu_i\rangle\}_{i=1}^{m}$ provides a good representation (through Eq. 106) of all the functions in the space, is equivalent to assuming that the set satisfies the completeness relation, Eq. 109.

By acting with the bra $\langle x|$ on Eqs. 106–108, we convert them to Schrödinger representation

$$\langle x \,|\, f\rangle \equiv x^2 \sin(x) \cong \sum_{i=1}^{m} a_i \langle x \,|\, \nu_i\rangle = \sum_{i=1}^{m} a_i \nu_i(x) \tag{110}$$

where (see Eq. 107)

$$a_i = \langle \nu_i \,|\, f\rangle = \int_0^\pi \langle \nu_i \,|\, x\rangle \, \langle x \,|\, f\rangle \, \mathrm{d}x = \int_0^\pi \nu_i(x)\, x^2 \sin(x)\, \mathrm{d}x \tag{111}$$

The second equality is based on our definition of the scalar product in this space.

I can now test the theory by calculating a_i from Eq. 111 and inserting the result in Eq. 110. If the sum in Eq. 110 is nearly equal to $\langle x \,|\, f\rangle = x^2 \sin(x)$, the theory is successfully tested.

§ 21 *A fit of* $x^2 \sin(x)$*,* $x \in [0, \pi]$*, by orthogonal polynomials.* The calculation that provides the fit (and tests its quality) has the following steps.

1. Assume that the representation given by Eq. 110 is plausible.

2. Calculate the coefficients a_n from Eq. 111, by using the polynomials $\nu_i(x)$ given by Eqs. 101–104.

3. Test whether the sum in Eq. 110 is a good approximation to $f(x) = \langle x \,|\, f \rangle = x^2 \sin(x)$.

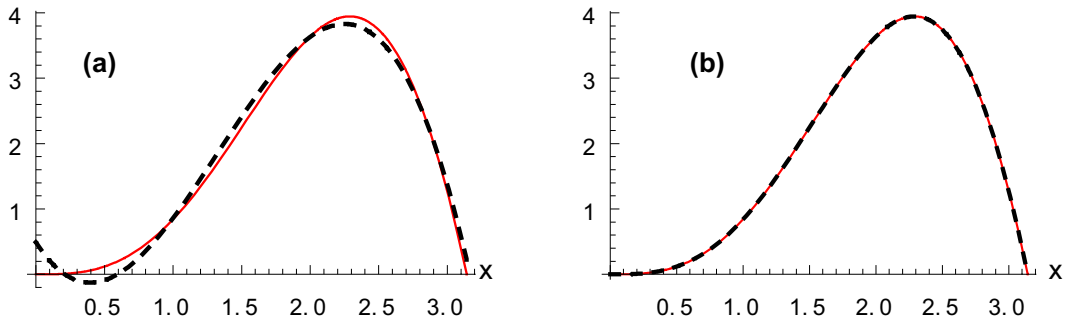

Figure 7: (a) The solid, red curve is $x^2 \sin(x)$ and the dashed, black curve is its representation by a basis set of five orthonormal polynomials. (b) Same as in (a) but with a basis set of ten polynomials. The plots were made in Cell 3 of Section II or WorkBook3.nb.

In Section II, Cell 3, of WorkBook3, I calculated the coefficients a_n by performing the integrals in Eq. 111. For example,

$$\begin{aligned} a_3 &= \int_0^\pi \nu_3(x)\, x^2 \sin(x)\, \mathrm{d}x \\ &= \int_0^\pi \left(\sqrt{\frac{5}{\pi}} - \frac{6\sqrt{5}}{\pi^{3/2}} x + \frac{6\sqrt{5}}{\pi^{5/2}} x^2 \right) x^2 \sin(x)\, \mathrm{d}x = -1.19836 \end{aligned} \quad (112)$$

The others are

$$a_1 = 3.31157, a_2 = 1.82699, \text{ and } a_4 = -1.2868 \quad (113)$$

The expansion is therefore

$$x^2 \sin(x) \cong \nu_1(x) + 3.31\nu_2(x) - 1.198\nu_3(x) - 1.29\nu_4(x) \quad (114)$$

with $\nu_1(x), \ldots, \nu_4(x)$ given by Eqs. 101–104. Using these equations in Eq. 114 gives

$$x^2 \sin(x) \cong 0.49 - 3.31x + 4.92x^2 - 1.24x^3 \tag{115}$$

In Fig. 7(a) I show the function $x^2 \sin(x)$ together with the fit with $m = 5$ in Eq. 110. The fit is respectable and it is much better when $m = 10$ (Fig. 7(b)).

§ 22 *Is this expansion in orthonormal polynomials better than a power-series expansion?* Eq. 115 approximates $x^2 \sin(x)$ with a polynomial. We could have obtained a polynomial approximation for this function by using a Taylor-series expansion. Unless the orthonormal-polynomial expansion is better than a power-series expansion, we have wasted our time developing it. Here we test the performance of two Taylor-series expansions against that obtained by using orthonormal polynomials.

The Taylor-series expansion to fourth order around the point $x = 0$ is

$$f(x) \cong \sum_{k=0}^{4} \frac{1}{k!} \left(\frac{\partial^k f}{\partial x^k} \right)_{x=0} x^k \tag{116}$$

In the case of $f(x) = x^2 \sin(x)$, this gives (see Cell 3 in Section II of WorkBook3.nb)

$$x^2 \sin(x) \cong x^3 \tag{117}$$

In Fig. 8a, I give plots (made in Cell 3 of Section II or WorkBook3) of $x^2 \sin(x)$, of its representation Eq. 115 by a sum of orthonormal polynomials, and of the Taylor-series approximation Eq. 117. The Taylor-series expansion does very well when x is near the expansion point $x = 0$ but is very bad for larger x. The orthonormal-polynomial expansion of the same order does better for all x except those very close to $x = 0$.

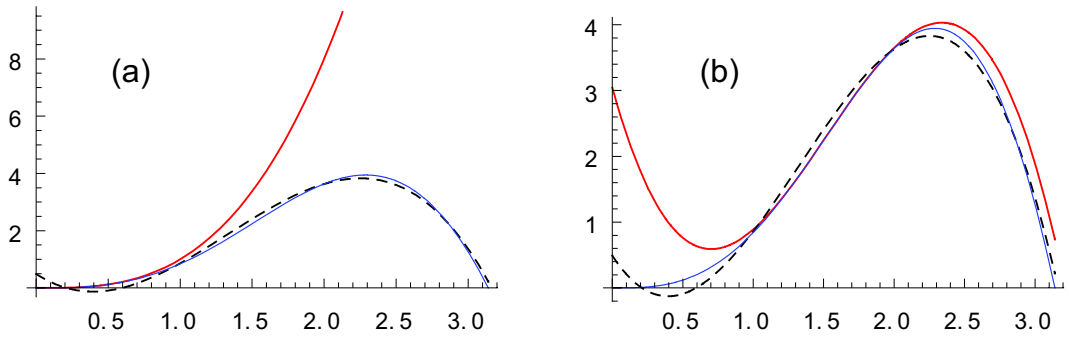

Figure 8: (a) The third-order power-series expansion of $x^2 \sin(x)$ around $x = 0$ (in red), $x^2 \sin(x)$ (in blue), and the polynomial expansion to third order (in black, dashed). (b) Same as in (a) except that the power-series expansion is around $x = \pi/2$.

Perhaps this is an unfair comparison since we took the expansion point to be $x = 0$. A Taylor-series expansion around $x = \pi/2$ might do better. We perform such an expansion in Cell 7.4 and the result is

$$x^2 \sin(x) \cong 3.04 - 7.75x + 7.17x^2 - 1.57x^3 \tag{118}$$

In Fig. 8b, I plot this approximation together with $x^2 \sin(x)$ and the orthonormal-polynomial expansion. The Taylor series does well for x around the expansion point $x = \pi/2$ but is very inaccurate for other values of x.

The Taylor expansion uses only information about the function at the expansion point. The orthonormal polynomial expansion uses information in the whole range of values of x; this gives it an edge.

Section 3.4. Non-orthogonal Basis Sets

§ 23 *Introduction.* In constructing orthonormal basis sets, we follow the example of the space R^3 where the basis set $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$, described in §16, was orthonormal. It turns out that this is not always the best choice when dealing with more general spaces. The main purpose of a basis set $|\phi_i\rangle$, $i = 1, 2, \ldots, N$, is to allow us to represent an arbitrary ket $|\psi\rangle$ in the form

$$|\psi\rangle = \sum_{i=1}^{N} a_i |\phi_i\rangle \tag{119}$$

where a_i are complex numbers. This representation is used in practice to solve the equation satisfied by $|\psi\rangle$. In the process of doing this, we must evaluate a large number of scalar products involving $|\phi_i\rangle$. In Schröding-er representation this amounts to performing a large number of integrals involving $\langle x \,|\, \phi_i\rangle$. Thus, besides completeness (which means that Eq. 119 is a good representation of $|\psi\rangle$), we have the practical requirement that these integrals can be evaluated efficiently. It so happens that many basis sets that satisfy this requirement are not orthonormal. Sometimes we may have to give up orthonormality in favor of easy integrability.

Often the physics of the system suggests to us basis sets that are not orthonormal but permit a convenient physical picture of the wave function. One example is the molecular orbital theory of a diatomic molecule such as H_2. In that case, the molecular orbital is taken to be a linear combination of atomic orbitals centered on the two atoms. The atomic orbitals are the basis set and they are not orthogonal.

It is therefore important to understand the main issues involved in constructing non-orthogonal basis sets. This is the purpose of this section.

§ 24 *Linear independence.* Again we look to R^3 for inspiration. It is clearly possible to write any vector in terms of three arbitrary vectors, $\vec{a}_1$, $\vec{a}_2$, $\vec{a}_3$, even if they are not orthogonal:

$$\vec{v} = \alpha_1\vec{a}_1 + \alpha_2\vec{a}_2 + \alpha_3\vec{a}_3$$

as long as the three vectors $\vec{a}_i$, $i = 1, 2, 3$ *are not co-planar* (i.e. not contained in the same plane). It is this injunction against co-planarity that we explore here.

Let us assume that $\vec{a}_1$, $\vec{a}_2$, and $\vec{a}_3$ *are in the same plane.* There is no loss of generality if we pick the coordinate system so that the OX and OY axes are in this plane. It is immediately clear that the sum

$$\alpha_1\vec{a}_1 + \alpha_2\vec{a}_2 + \alpha_3\vec{a}_3$$

cannot represent a vector that has a component in the OZ direction. This means that the basis set is not complete in R^3, because it cannot represent all the vectors of R^3. It can, however, represent all the vectors contained in the XOY plane (i.e. vectors whose z-component is zero), which is an R^2 subspace of R^3. But this job can be achieved by any two vectors in the set $\{\vec{a}_1, \vec{a}_2, \vec{a}_3\}$, as long as they are not co-linear (i.e. they are not parallel). We say that in the subspace R^2 the basis set $\{\vec{a}_1, \vec{a}_2, \vec{a}_3\}$ is *overcomplete.*

Representing the vectors $\vec{\eta}$ lying in the XOY plane as

$$\vec{\eta} = \eta_1\vec{a}_1 + \eta_2\vec{a}_2 + \eta_3\vec{a}_3 \tag{120}$$

is wasteful and causes all kinds of trouble in applications. The waste comes from the fact that we can write

$$\vec{a}_3 = n_1\vec{a}_1 + n_2\vec{a}_2 \tag{121}$$

and therefore the third term in Eq. 120 is unnecessary.

This suggests that in a good basis set $\{|\phi_i\rangle\}_{i=1}^N$, we should be unable to write any one of the elements as a linear combination of the others. This means that the equality

$$\sum_{i=1}^{N} n_i|\phi_i\rangle = 0 \tag{122}$$

is possible only if all the coefficients n_i are equal to zero. If that is true, we say that the set $\{|\phi_i\rangle\}_{i=1}^N$ is *linearly independent*. If it is not true, we say that the set is *linearly dependent*. Linearly dependent sets are not to be used as basis sets! One way to think of a linearly independent set is that every element gives us information that the others cannot. A linearly dependent set contains one or more vectors that add no new information.

It is easy to derive a test for linear independence. Act with $\langle\phi_j|$ on Eq. 122, to obtain

$$\sum_{i=1}^{N} n_i\langle\phi_j\,|\,\phi_i\rangle = 0 \tag{123}$$

We call the number

$$S_{ji} \equiv \langle\phi_j\,|\,\phi_i\rangle \tag{124}$$

the overlap of $|\phi_j\rangle$ with $|\phi_i\rangle$. If $S_{ji} = 0$, we say that the two kets $|\phi_i\rangle$ and $|\phi_j\rangle$ do not overlap. If $S_{ji} \neq 0$, the kets overlap. Note that orthogonal kets do not overlap.

With this notation, we can write Eq. 123 as

$$\sum_{i=1}^{N} S_{ji}n_i = 0 \tag{125}$$

As j ranges from 1 to N, this constitutes a *homogeneous system of equations* for the unknowns n_i, which is equivalent to Eq. 123. We make use of the

following theorem: if the determinant of the matrix (S_{ji}) differs from zero,

$$\det(S_{ij}) \neq 0, \tag{126}$$

then the only solution of Eq. 125 is

$$n_1 = 0, n_2 = 0, \ldots, n_N = 0$$

On the other hand, if

$$\det(S_{ij}) = 0, \tag{127}$$

then Eq. 125 has non-zero solutions.

The system in Eq. 125 is equivalent to Eq. 123. *We conclude that if* $\det(S_{ij}) \neq 0$ *then the solution of* $\sum_i n_i|\phi_i\rangle = 0$ *is* $n_1 = \cdots = n_N = 0$ *and the kets* $|\phi_i\rangle$ *are linearly independent.*

§ 25 *An orthogonal basis set is always linearly independent.* If a set $\{|\phi_i\rangle\}_{i=1}^N$ is orthogonal then, by the definition of orthogonality, we have

$$S_{ij} = \langle\phi_i \,|\, \phi_j\rangle = 0 \text{ if } i \neq j \tag{128}$$

The system of equations Eq. 125 reduces to

$$\begin{aligned} S_{11}n_1 &= 0 \\ S_{22}n_2 &= 0 \\ &\vdots \\ S_{NN}n_N &= 0 \end{aligned} \tag{129}$$

Each "self-overlap" $S_{ii} = \langle\phi_i \,|\, \phi_i\rangle$ is nonzero and therefore $n_1 = 0$, $n_2 = 0$, $\ldots n_N = 0$. The set is linearly independent.

An example of linear dependence is provided by the set $\vec{a}_1 = \{1, 0, 0\}$, $\vec{a}_2 = \{0, 1, 0\}$, $\vec{a}_3 = \{1, 1, 0\}$. Clearly

$$\vec{a}_3 - \vec{a}_1 - \vec{a}_2 = 0,$$

showing linear dependence.

Also note that all three of these vectors are in the XOY plane: their z-components are zero. Therefore we cannot use this basis to represent a vector whose z-component is not zero. The set is not a complete basis in R^3. If we regard it as a basis set in R^2, it is overcomplete. Any two of the vectors will suffice for representing all vectors in the XOY plane; the third one is superfluous.

§ 26 *The disadvantages of non-orthogonal basis sets.* We use a basis set $\{|a_1\rangle, |a_2\rangle, \ldots, |a_N\rangle\}$ to represent kets $|v\rangle$ by expressions

$$|v\rangle = \sum_{i=1}^{N} c_i |a_i\rangle \tag{130}$$

where the coefficients c_i depend on $|v\rangle$. If we act with $\langle a_j|$ on this expression, we obtain

$$\langle a_j \,|\, v\rangle = \sum_{i=1}^{N} c_i \langle a_j \,|\, a_i\rangle = \sum_{i=1}^{N} S_{ji} c_i \tag{131}$$

As before, $S_{ji} \equiv \langle a_j \,|\, a_i\rangle$ are the elements of the overlap matrix. To calculate the coefficients c_i, we must evaluate $\langle a_j \,|\, v\rangle$, calculate S_{ji}, and then solve the system of equations Eq. 131. It is here where we are hurt if the determinant of the overlap matrix is zero: in such a case, the system does not have a solution. In addition, if the determinant is very small, the solutions of the system in Eq. 131 are likely to be inaccurate when determined by numerical methods (which are the only methods available).

Even if the determinant is robust, obtaining the coefficients c_i from Eq. 131 is more difficult when the basis set is not orthonormal. If it is orthonormal then

$$S_{ij} = \delta_{ij} \tag{132}$$

and Eq. 131 becomes

$$c_j = \langle a_j \,|\, v \rangle \tag{133}$$

so that the values of c_j are easier to calculate.

§ 27 *The advantages of non-orthogonal basis sets.* Given all these troubles, why would we want to use a non-orthogonal basis set? I answer with an example: the molecular orbitals of the hydrogen molecule. We construct them as a linear combination of atomic orbitals. For example, one of them is

$$\psi(\vec{r}) = \frac{1}{\sqrt{2}}[1\mathrm{s}_\mathrm{A}(\vec{r}_\mathrm{A}) + 1\mathrm{s}_\mathrm{B}(\vec{r}_\mathrm{B})] \tag{134}$$

Here $1\mathrm{s}_\mathrm{A}(\vec{r}_\mathrm{A})$ are the 1s orbital of hydrogen atom A and $\vec{r}_\mathrm{A}$ is the vector giving the position of the electron with respect to proton A (see Fig. 9). $1\mathrm{s}_\mathrm{B}(\vec{r}_\mathrm{B})$ has a similar meaning, for proton B.

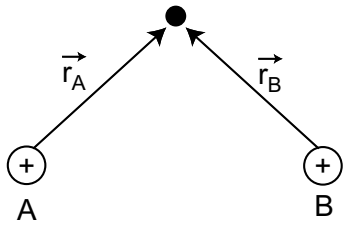

Figure 9: The circles containing + signs represent protons; the closed circle represents an electron.

The motivation for this choice is simple. It is reasonable to assume that $1s_A$ describes well the behavior of the electron when it is located near proton A and away from proton B, and that $1s_B$ performs a similar service in the neighborhood of B. The linear combination should also represent well the state of the electron in the region between nuclei. The sum in Eq. 134 is presumably an acceptable representation of the electron state at all locations.

Here $1s_A(\vec{r}_A)$ and $1s_B(\vec{r}_B)$ were used as a basis set. These functions *are not orthogonal*. When more accuracy is needed, we assume that

$$\psi(\vec{r}) = a_1 1s_A(\vec{r}_A) + a_2 1s_B(\vec{r}_B) + a_3 2p_{xA}(\vec{r}_A) + a_4 2p_{xB}(\vec{r}_B) \tag{135}$$

This basis set is also non-orthogonal. When we construct a basis set, we try to define it so that we need a small number of terms to represent the state we are interested in. This is why the atomic orbitals were used. As a bonus, this basis set has a physical meaning. However, when this representation is used for calculating the energy and the wave function of a molecule, we must perform many integrals involving the atomic orbitals. These integrals are hard to evaluate. Because of this, modern quantum chemistry replaced the atomic-orbital basis set with a set of Gaussian functions centered on the atoms. This basis set is also non-orthogonal and the functions no longer have a physical meaning. However, the integrals are much easier (i.e. faster) to perform.

§ 28 *Summary.* We use basis sets for representing an unknown wave function $\psi(x)$ as a linear combination

$$\psi(x) = \sum_n a_n \phi_n(x) \tag{136}$$

of the known functions $\phi_n(x)$.

Since $\psi(x)$ is unknown, the coefficients a_n are unknown. We determine them by introducing the representation into the Schrödinger equation for $\psi(x)$. Then we determine a_n so that $\psi(x)$ given by Eq. 136 satisfies this equation.

In practical terms, completeness means that the set $\{\phi_n(x)\}_{n=1}^N$ is flexible enough to give a good fit to $\psi(x)$. Since we don't know $\psi(x)$, we have to guess the set $\{\phi_n(x)\}_{n=1}^N$ and this is not always easy. If we do not guess right, $\psi(x)$ does not satisfy the Schrödinger equation with sufficient precision. In that case, we need to add more functions to the basis set or come up with a more sensible set. In many situations, the set is not orthonormal. Then we have a choice: use the Gram-Schmidt procedure to generate an orthonormal set or work with the set as is.