

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, in order to show that the kets have certain properties. Particularly important is an expression called the *completeness relation* or the *resolution of identity*. This property is essential for setting up the formalism of quantum mechanics and for obtaining numerical solutions of its equations.

§ 2 *A simple property of probability.* We measure a quantity A that can take one of the values $\{a_1, a_2, \dots, a_n\}$. 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 the 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

$$\sum_{i=1}^n \frac{N_i}{N} = 1 \quad (1)$$

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

$$P(a_i) \equiv \lim_{N \rightarrow \infty} \frac{N_i}{N} \quad (2)$$

Combining Eq. 1 with Eq. 2 gives

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

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

§ 3 *Application to an observable in quantum mechanics.* Assume now that A is a quantum-mechanical observable, its spectrum is $\{a_1, a_2, a_3, \dots\}$, and its continuous spectrum is $\alpha \in D$. If the system is in a normalized state $|\psi\rangle$, the probability that a measurement of A yields a_i is

$$P_\psi(a_i) = |\langle a_i | \psi \rangle|^2 \quad (4)$$

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

$$P_\psi(\alpha) d\alpha = |\langle \alpha | \psi \rangle|^2 d\alpha, \quad \alpha \in D \quad (5)$$

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

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

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

§ 4 Dirac notation: operators and bras. For any complex number z , we have $|z|^2 = z^*z$; for any scalar product, $\langle \lambda | \mu \rangle = \langle \mu | \lambda \rangle^*$. Combining these two equations gives $|\langle a_i | \psi \rangle|^2 = \langle \psi | a_i \rangle \langle a_i | \psi \rangle$. Using this result we rewrite Eq. 6 as

$$\sum_{n=1}^{\infty} \langle \psi | a_n \rangle \langle a_n | \psi \rangle + \int_{\alpha \in D} d\alpha \langle \psi | \alpha \rangle \langle \alpha | \psi \rangle = 1 \quad (7)$$

At this point we follow Dirac and introduce a very innovative and useful notation:

$$\langle \psi | a_n \rangle \langle a_n | \psi \rangle \equiv \langle \psi | (|a_n\rangle \langle a_n|) | \psi \rangle \quad (8)$$

This is an identity in which the left-hand side is written (in the right-hand side) as a product of three distinct symbols: 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. 8. 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 $(\langle a |) |\psi\rangle = \langle a | \psi \rangle$. 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, and vice versa. Thus the bras form a space of their own and there is a one-to-one correspondence between the bra space and the ket space.

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

$$\langle \lambda | \psi \rangle = \langle \psi | \lambda \rangle^* = \langle \psi | \eta \rangle^* + \langle \psi | \mu \rangle^* = \langle \eta | \psi \rangle + \langle \mu | \psi \rangle \quad (9)$$

Therefore we have the rule:

$$\text{if } |\lambda\rangle = |\eta\rangle + |\mu\rangle \text{ then } \langle\lambda| = \langle\eta| + \langle\mu| \quad (10)$$

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

$$\langle\eta|\psi\rangle = \langle\psi|\eta\rangle^* = \langle\psi|\alpha a\rangle^* = (\alpha\langle\psi|a\rangle)^* = \alpha^*\langle a|\psi\rangle \quad (11)$$

and therefore, we have the rule:

$$\text{if } |\eta\rangle = \alpha|a\rangle \text{ then } \langle\eta| = \alpha^*\langle a| \quad (12)$$

The operator

$$\hat{P}(a_n) \equiv |a_n\rangle\langle a_n| \quad (13)$$

is also a new mathematical object. By definition it acts on a ket $|\psi\rangle$ through the rule

$$\hat{P}(a_n)|\psi\rangle = (|a_n\rangle\langle a_n|)|\psi\rangle = |a_n\rangle\langle a_n|\psi\rangle \quad (14)$$

An operator (here, $|a_n\rangle\langle a_n|$) acts on a ket (here, $|\psi\rangle$) to produce another ket (here, $|a_n\rangle\langle a_n|\psi\rangle$). 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 and a ket is a ket. An operator is any object that acts on a ket to generate another ket.

§ 5 *The completeness relation.* Now that we are comfortable with Dirac's notation, let us use it to rewrite Eq. 7 as

$$\langle\psi|\left(\sum_{n=1}^{\infty}|a_n\rangle\langle a_n| + \int_{\alpha\in D} d\alpha|\alpha\rangle\langle\alpha|\right)|\psi\rangle = 1 \quad (15)$$

This is equivalent to setting

$$\sum_{n=1}^{\infty}|a_n\rangle\langle a_n| + \int_{\alpha\in D} d\alpha|\alpha\rangle\langle\alpha| = \hat{I} \quad (16)$$

where \hat{I} is the *unit operator*, defined to give

$$\hat{I}|\psi\rangle = |\psi\rangle \text{ for any ket } |\psi\rangle \quad (17)$$

It is not difficult to show that Eqs. 16 and 17 are equivalent to Eq. 15 (remember that $\langle\psi|\psi\rangle = 1$).

Eq. 16 is called the *completeness relation* for the pure states $|a_n\rangle$, $n = 1, 2, \dots$, 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, and the result must be used with great care. We are going to use it often and I will point out, at the appropriate places, how easy it is to misinterpret and misuse it.

§ 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 X of the particle. The position 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 value of X is x . The completeness relation for these pure states is

$$\hat{I} = \int_{-\infty}^{+\infty} dx |x\rangle\langle x| \quad (18)$$

We say that this equation gives the unit operator in the *coordinate representation*.

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

$$\hat{I} = \int_{-\infty}^{+\infty} dp |p\rangle\langle p| \quad (19)$$

because momentum is an observable with a purely continuous spectrum. Eq. 19 gives the unit operator in the *momentum representation*.

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

$$\hat{I} = \sum_{n=1}^{\infty} |E_n\rangle\langle E_n| + \int_0^{\infty} d\alpha |E_\alpha\rangle\langle E_\alpha| \quad (20)$$

where $|E_n\rangle$ and $|E_\alpha\rangle$ are the pure states of energy.

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, 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

$$|\psi\rangle = \hat{I}|\psi\rangle \quad (21)$$

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

$$\hat{I} = \sum_{n=1}^{\infty} |a_n\rangle \langle a_n| + \int_{\alpha \in D} |\alpha\rangle \langle \alpha| d\alpha \quad (22)$$

Using this in $|\psi\rangle = \hat{I}|\psi\rangle$ (i.e. Eq. 21), we obtain

$$|\psi\rangle = \sum_{n=1}^{\infty} |a_n\rangle \langle a_n | \psi\rangle + \int_{\alpha \in D} |\alpha\rangle \langle \alpha | \psi\rangle d\alpha \quad (23)$$

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 . $\langle \alpha | \psi\rangle$ are also complex numbers and $P_{\psi}(\alpha) = |\langle \alpha | \psi\rangle|^2 d\alpha$ is the probability that A takes values between α and $\alpha + d\alpha$ 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. 23 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 $\{|\alpha\rangle\}_{\alpha \in D}$* .

You will learn later that the pure states of A are also the eigenstates of an operator associated with A , and Eq. 23 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. 23 is central to quantum mechanics. It is used to prove many “theorems”, 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 that in the presentation you have seen so far, the resolution of identity (Eq. 22) and its consequence Eq. 23 have a false generality. If its use is not augmented with common sense and watchful care, it can lead to absurd conclusions or meaningless calculations. After we look at a few examples, I will clarify why I say this.

§ 8 *A hint of how we might use Eq. 23.* 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. 23 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 out what $|\psi\rangle$ is, becomes equivalent to calculating these unknown numbers. One way to do this is to put the expression (Eq. 23) 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 with 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 keen to understand all the implications of Eq. 23. It also tells you that the equation is useless unless we find a way to calculate the pure states of an observable in a form that can be implemented on a computer.

There is another wrinkle that we must take care of. Eq. 23 expresses $|\psi\rangle$ in terms of an *infinite* set of numbers. No computer memory can handle that. There are, however, good reasons to expect that in all practical problems, we can use in Eq. 23 a finite number of terms and still have a good approximation for $|\psi\rangle$. I explain why this is so in §9.

§ 9 *Why we can use a finite number of terms in the energy representation.* Eqs. 22 and 23 have as many implementations as many observables the system has. For example, I can choose A to be the energy of the system. If I denote its pure states by $|E_n\rangle$ and $|E_\alpha\rangle$, then Eq. 23 becomes

$$|\psi\rangle = \sum_{n=1}^{\infty} |E_n\rangle \langle E_n | \psi \rangle + \int_0^{\infty} d\alpha |E_\alpha\rangle \langle E_\alpha | \psi \rangle \quad (24)$$

Remember now that the continuous energy spectrum represents states in which the system breaks up into fragments. These states describe how the fragments move away from each other after the break-up. The probability that a system in state $|\psi\rangle$ broke into such fragments having energy between E_α and $E_\alpha + dE_\alpha$ is $|\langle E_\alpha | \psi \rangle|^2 dE_\alpha$. If the molecule has low internal energy and is stable, the chance that it will break up is exceedingly small; this means that $|\langle E_\alpha | \psi \rangle|^2 \cong 0$. Therefore, I can neglect the integral in Eq. 24 and keep

$$|\psi\rangle \cong \sum_{n=1}^{\infty} |E_n\rangle \langle E_n | \psi \rangle \quad (25)$$

Practice has shown that most often we can use a finite number of terms in this sum and still have a good description of $|\psi\rangle$, given by

$$|\psi\rangle \cong \sum_{n=m}^{m+N} |E_n\rangle \langle E_n | \psi \rangle \quad (26)$$

where m and N are finite.

To understand why this is reasonable, I remind you that

$$P_\psi(E_n) = |\langle E_n | \psi \rangle|^2 \quad (27)$$

is the probability that when I measure the energy of the system in state $|\psi\rangle$, the result is E_n . Imagine now, as an example, that I have exposed a molecule, which was initially in the ground state $|E_0\rangle$, to light containing photons of energy 3 eV. The molecule will either absorb a photon and switch its state to some $|E_n\rangle$, or it will not absorb a photon and remain in the state $|E_0\rangle$. If I want to describe $|\psi\rangle$ well, I must include in Eq. 26 $|E_0\rangle$ and states having the energy around $|E_n\rangle$ and between $|E_n\rangle$ and $|E_0\rangle$. But I don't have to include states $|E_m\rangle$ whose energy is 10 eV, since the probability $|\langle E_m | \psi \rangle|^2$ that they are excited is very small.¹

§ 10 The coordinate representation. When we derived the completeness relation, we emphasized that it is valid for the pure states of any observable. The energy representation is very frequently used because the energy plays such a central role in quantum mechanics. The coordinate representation

¹This is a bit oversimplified. A better understanding of the issues involved is given in the chapter on time-dependent perturbation theory.

plays an equally 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. 18)

$$\hat{I} = \int_{-\infty}^{+\infty} dx |x\rangle\langle x| \quad (28)$$

Here $|x\rangle$ is the pure 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, all 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.

A scalar product $\langle \psi | \phi \rangle$ can be written as

$$\begin{aligned} \langle \psi | \phi \rangle &= \langle \psi | \hat{I} \phi \rangle = \int_{-\infty}^{+\infty} dx \langle \psi | x \rangle \langle x | \phi \rangle = \int_{-\infty}^{+\infty} dx \langle x | \psi \rangle^* \langle x | \phi \rangle \\ &= \int_{-\infty}^{+\infty} \psi(x)^* \phi(x) dx \end{aligned} \quad (29)$$

This is 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. 29 to calculate the scalar product $\langle \psi | \phi \rangle$.

§ 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

$$\sum_{n=1}^{\infty} |e_n\rangle \langle e_n| = \hat{I} \quad (30)$$

Denote the pure states of the energy of a harmonic oscillator by $|E_m\rangle$, $m \geq 0$. Using Eq. 30 I can write

$$|E_m\rangle = \hat{I} |E_m\rangle = \sum_{n=1}^{\infty} |e_n\rangle \langle e_n | E_m \rangle \quad (31)$$

I want to test this equation and explore more deeply its meaning and limitations.

You might have foreseen some trouble ahead already. A particle in a box is 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 state *but give no positional information*. We don't know where the box is located and we don't know 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 present in Eq. 23, we must pay attention to the particle's position. 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). So let us act on Eq. 31 with the bra $\langle x|$ to introduce the position into our equations:

$$\langle x | E_m \rangle = \sum_{n=1}^{\infty} \langle x | e_n \rangle \langle e_n | E_m \rangle \quad (32)$$

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 :

$$\langle x | e_n \rangle = \begin{cases} \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right), & n = 1, 2, 3, \dots & \text{if } x \in [0, L] \\ 0 & & \text{if } x \notin [0, L] \end{cases} \quad (33)$$

This expression was obtained 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. 33 in Eq. 32, we obtain

$$\langle x | E_m \rangle = \begin{cases} \sum_{n=1}^{\infty} \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right) \langle e_n | E_m \rangle & \text{if } x \in [0, L] \\ 0 & \text{if } x \notin [0, L] \end{cases} \quad (34)$$

§ 12 Trouble ahead. As a result of these formal manipulations, 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 expansion in coordinate representation makes it easy to see the limitations of the formal expansion

$\sum_n |e_n\rangle\langle e_n| = \hat{I}$. The wave function $\langle x | E_m \rangle$ 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 wave function $\langle x | e_n \rangle$ differs from zero for $x \in [0, L]$. Eq. 34 can be right only if the region $x \in [x_0 - 3\lambda, x_0 + 3\lambda]$ is inside the region $x \in [0, L]$. To satisfy this condition, we take

$$x_0 = L/2 \quad (35)$$

and

$$L > 3\lambda \quad (36)$$

We placed the oscillator's equilibrium position x_0 in the middle of the box and took the box size L 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' .

The formal expansion

$$|E_m\rangle = \sum_{n=1}^{\infty} |e_n\rangle \langle E_n | E_m \rangle \quad (37)$$

(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.

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 $\langle x | E_0 \rangle$ and $\langle x | e_n \rangle$, $n = 1, 2, 3$. Note that $\langle x | E_0 \rangle$ is zero for $x \in [0, 0.1]$ and $x \in [0.4, 0.5]$, and none of $\langle x | e_n \rangle$, $n = 1, 2, \dots$ is zero in those regions of x . 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 in a way that the terms in the sum cancel each other for $x \in [0, 0.1]$ and $x \in [0.4, 0.5]$. This may not seem likely, but it happens.

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

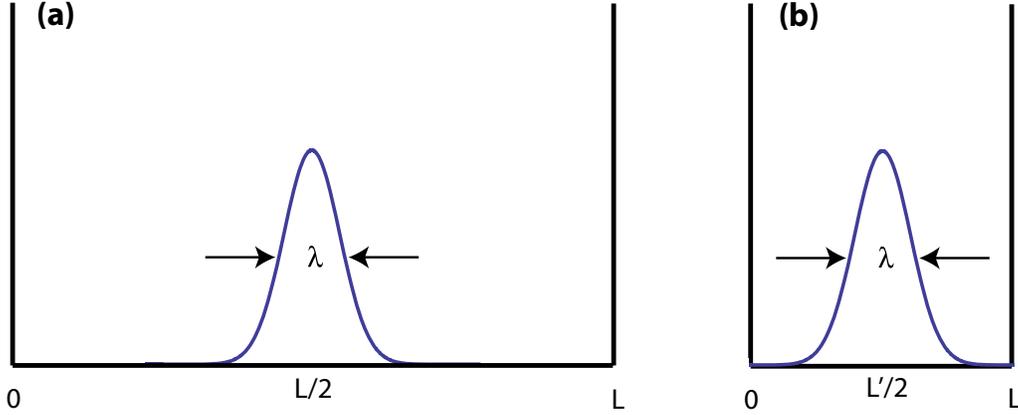


Figure 1: The oscillator ground-state wave function $\langle x | E_0 \rangle$ (a) in a box of width L and (b) in a smaller box, of width L'

the cancellation is harder to achieve than in the case of the smaller box, at the right of Fig. 1. The sum for the box of size L will need more terms to represent $\langle x | E_0 \rangle$ than the sum for the box of length $L' < L$, and the calculation is more time-consuming and more prone to error.

§ 13 *A test of the theory.* Now that we understand what we are up against, let us test Eq. 34. The wave function of the harmonic oscillator is given by (see Metiu, *Quantum Mechanics*, Eqs. 17.8–17.11, pages 272–273)

$$\langle x | E_m \rangle = \sqrt{\frac{1}{\lambda}} \left(\frac{1}{\pi^{1/4} \sqrt{2^m m!}} \right) \exp \left[-\frac{(x - x_0)^2}{2\lambda^2} \right] H_m \left(\frac{x - x_0}{\lambda} \right), \quad m = 0, 1, 2, \dots \quad (38)$$

with

$$\lambda = \sqrt{\frac{\hbar}{\mu\omega}} \quad (39)$$

and

$$x_0 = \frac{L}{2} \quad (40)$$

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

The test of Eq. 34, for $m = 0$, proceeds as follows:

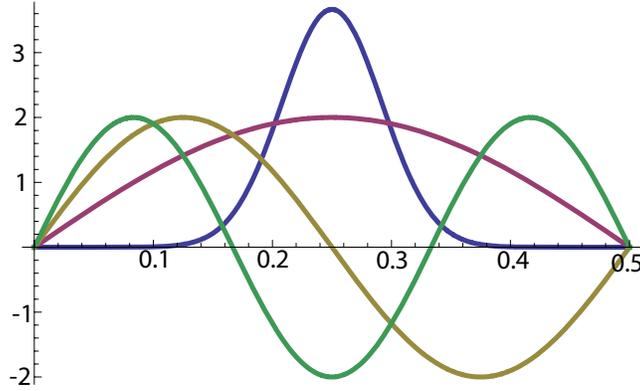


Figure 2: The blue curve is the ground state wave function $\langle x | E_0 \rangle$ of the harmonic oscillator. Purple: $\langle x | e_1 \rangle$; yellow: $\langle x | e_2 \rangle$; green: $\langle x | e_3 \rangle$. I used $L = 0.5 \text{ \AA}$ and $\lambda = 0.042 \text{ \AA}$.

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

$$\langle e_n | E_0 \rangle = \int_{-\infty}^{+\infty} \langle e_n | x \rangle \langle x | E_0 \rangle dx \quad (41)$$

2. Test whether the right-hand side of Eq. 34 is equal to the left-hand side (as given by Eqs. 38–40), by plotting them on the same graph.

Using Eq. 33 for $\langle e_n | x \rangle$ and Eq. 38 for $\langle x | E_0 \rangle$, we evaluate $\langle e_n | E_0 \rangle$ by performing the integral in Eq. 41. `Mathematica` gives (see Cell 4, `WorkBook3_Representation theory.nb`)

$$\begin{aligned} \langle e_n | E_0 \rangle &= -\frac{i}{2} \exp \left[-\frac{n\pi(iL^2 + n\pi\lambda^2)}{2L^2} \right] (-1 + e^{in\pi}) \pi^{1/4} \sqrt{\frac{\lambda}{L}} \\ &\times \left(\operatorname{Erf} \left[\frac{L^2 - 2in\pi\lambda^2}{2\sqrt{2}L\lambda} \right] + \operatorname{Erf} \left[\frac{L^2 + 2in\pi\lambda^2}{2\sqrt{2}L\lambda} \right] \right) \end{aligned} \quad (42)$$

This looks very complicated, but it is not. The Erf function is calculated by `Mathematica` automatically when the argument is a number. The result appears to be a complex number but this is not so: n is an integer, so $\langle e_n | E_0 \rangle$, given by Eq. 42, is a real number. Here are the values of $\langle e_n | E_0 \rangle$ for $L = 0.5$ and $n = 1, 2, \dots, 10$ (see Cell 4, `WorkBook3_Representation theory.nb`):

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

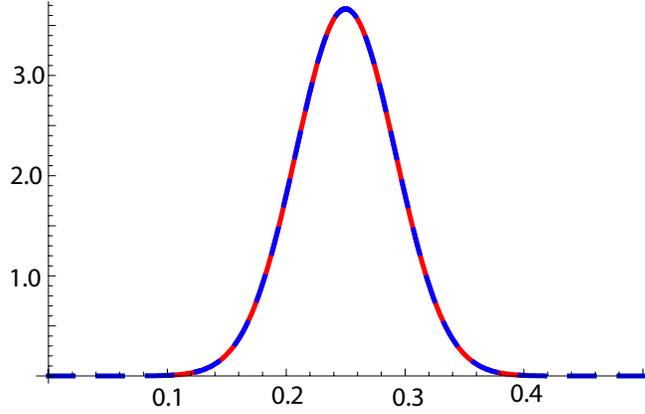


Figure 3: The harmonic oscillator ground state wave function is shown as a dashed blue line. The sum representing it, with 20 terms, is shown as a solid red line. $L = 0.5 \text{ \AA}$, $\lambda = 0.042 \text{ \AA}$. The plot was made in Cell 5 of `WorkBook3_Representation theory.nb`.

That is, $\langle e_1 | E_0 \rangle = 0.7453$, $\langle e_2 | E_0 \rangle = 0$, and so on.

Now that I have the coefficients $\langle e_n | E_0 \rangle$, I can write Eq. 34, for $x \in [0, L]$, as

$$\begin{aligned} \langle x | E_0 \rangle &= \sum_{n=1}^{\infty} \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right) \left(-\frac{i}{2}\right) \exp\left[-\frac{n\pi(iL^2 + n\pi\lambda^2)}{2L^2}\right] (-1 + e^{in\pi}) \\ &\quad \times \pi^{1/4} \sqrt{\frac{\lambda}{L}} \left(\text{Erf}\left[\frac{L^2 - 2in\pi\lambda^2}{2\sqrt{2}L\lambda}\right] + \text{Erf}\left[\frac{L^2 + 2in\pi\lambda^2}{2\sqrt{2}L\lambda}\right] \right) \end{aligned} \quad (43)$$

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

²Those of you who know mathematics will recognize this expression as a 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.

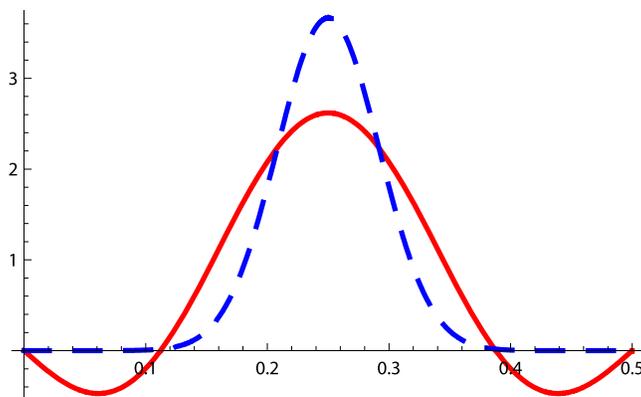


Figure 4: The solid red line is the sum in Eq. 43 with three terms. The dashed blue line is the ground state wave function of the harmonic oscillator. $L = 0.5 \text{ \AA}$, $\lambda = 0.042 \text{ \AA}$. The plot was made in Cell 5 of Workbook3_Representation theory.nb.

In Fig. 3, I compare the sum of the first twenty terms in Eq. 43 with the ground state wave function of the harmonic oscillator. Obviously the representation of $\langle x | E_0 \rangle$ by the sum is excellent. But what happens if we take only the first 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 $L = 0.5 \text{ \AA}$ and $\lambda = 0.042 \text{ \AA}$. 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 $L = 0.4 \text{ \AA}$, to narrow the box. You can see from Fig. 5 that this is the case.

§ 14 *A few general remarks.* We have talked about the space of kets in general, but now we have to emphasize 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 $x = L$.

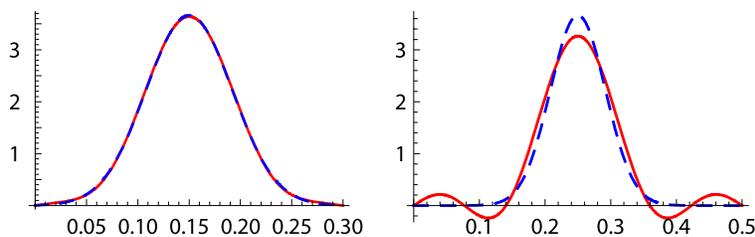


Figure 5: The left panel shows the ground state harmonic oscillator wave function (dashed, blue) and the six-term sum (solid, red), taken with $L = 0.4 \text{ \AA}$. The right panel shows the result when $L = 0.5 \text{ \AA}$ and the same number of terms in the sum. A small box, but not too small, is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

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

$$\sum_{n=1}^{\infty} |e_n\rangle \langle e_n| = \hat{I} \quad (44)$$

and

$$\sum_{n=1}^{\infty} |\varepsilon_n\rangle \langle \varepsilon_n| = \hat{I} \quad (45)$$

(a) I can *formally* write

$$|e_m\rangle = \sum_{n=1}^{\infty} |\varepsilon_n\rangle \langle \varepsilon_n | e_m\rangle \quad (46)$$

and

$$|\varepsilon_m\rangle = \sum_{n=1}^{\infty} |e_n\rangle \langle e_n | \varepsilon_m\rangle \quad (47)$$

Are these equations correct?

(b) Check whether the right-hand side of

$$\langle x | \varepsilon_1\rangle \cong \sum_{n=1}^N \langle x | e_n\rangle \langle e_n | \varepsilon_1\rangle \quad (48)$$

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

$$\langle x | e_1 \rangle \cong \sum_{n=1}^N \langle x | \varepsilon_n \rangle \langle \varepsilon_n | e_1 \rangle \quad (49)$$

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

$$\hat{I} = |\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow| \quad (50)$$

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

$$|\psi\rangle = |\uparrow\rangle\langle\uparrow|\psi\rangle + |\downarrow\rangle\langle\downarrow|\psi\rangle \quad (51)$$

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

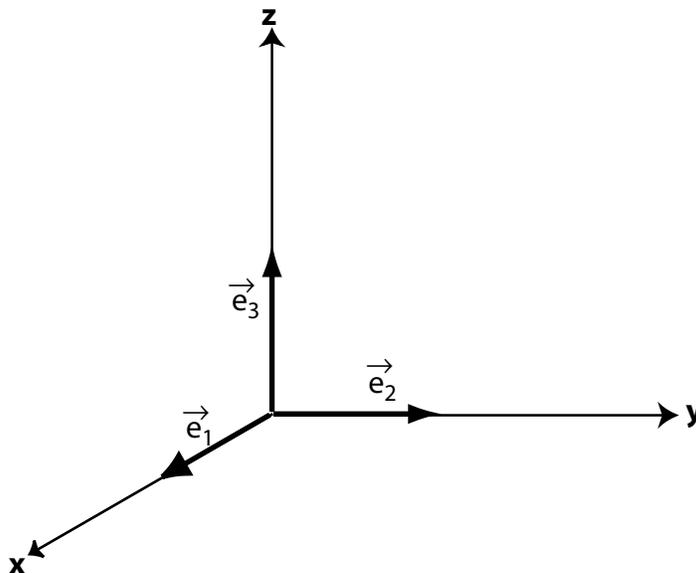
Section 3.3. Generalization: basis set

§ 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

$$|\psi\rangle \cong \sum_{n=1}^N |a_n\rangle \langle a_n | \psi \rangle \quad (52)$$

This follows from the completeness relation

$$\hat{I} \cong \sum_{n=1}^N |a_n\rangle \langle a_n| \quad (53)$$

Figure 6: An orthonormal basis in \mathbb{R}^3

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. 52 depends only on the validity of the relation Eq. 53. 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. 52 becomes

$$\psi(x) \cong \sum_{n=1}^N c_n a_n(x) \quad (54)$$

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. 54? 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 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 calculus. There, you used three vectors $\vec{e}_1, \vec{e}_2, \vec{e}_3$ of unit length, oriented along the axes (see Fig. 6). Any three-dimensional vector \vec{r} can be written as

$$\vec{r} = \alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2 + \alpha_3 \vec{e}_3 \quad (55)$$

We say that the set of vectors $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ generates the space \mathbb{R}^3 or that it is a complete basis set in \mathbb{R}^3 . This basis set is orthonormal because $\vec{e}_i \cdot \vec{e}_j = 0$ 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 \mathbb{R}^3 . Eq. 54 is a generalization of Eq. 55 to the linear space of functions.

§ 17 *Gram-Schmidt orthogonalization.* The basis set $\vec{e}_1, \vec{e}_2, \vec{e}_3$ in \mathbb{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 ℓ^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 $\{|o_1\rangle, |o_2\rangle, \dots, |o_N\rangle\}$ and generate a new set $\{|\nu_1\rangle, |\nu_2\rangle, \dots, |\nu_N\rangle\}$ whose kets satisfy the orthonormality relation

$$\langle \nu_i | \nu_j \rangle = \delta_{ij}, \quad i, j = 1, 2, \dots, N \quad (56)$$

Here is the algorithm that does that.

$$|n_1\rangle = |o_1\rangle \quad (57)$$

$$|n_2\rangle = |o_2\rangle - |n_1\rangle \frac{\langle n_1 | o_2 \rangle}{\langle n_1 | n_1 \rangle} = \left[\hat{I} - \frac{|n_1\rangle \langle n_1|}{\langle n_1 | n_1 \rangle} \right] |o_2\rangle \quad (58)$$

$$|n_3\rangle = |o_3\rangle - |n_1\rangle \frac{\langle n_1 | o_3 \rangle}{\langle n_1 | n_1 \rangle} - |n_2\rangle \frac{\langle n_2 | o_3 \rangle}{\langle n_2 | n_2 \rangle} \quad (59)$$

Or, in general

$$|n_j\rangle = \left[\hat{I} - \sum_{i=1}^{j-1} \frac{|n_i\rangle \langle n_i|}{\langle n_i | n_i \rangle} \right] |o_j\rangle, \quad j = 1, 2, \dots, N \quad (60)$$

You can verify by direct calculation that

$$\langle n_i | n_j \rangle = 0 \text{ for } i \neq j; \quad (61)$$

that is, the set $\{|n_1\rangle, |n_2\rangle, \dots, |n_N\rangle\}$ 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

$$|\nu_i\rangle \equiv \frac{|n_i\rangle}{\sqrt{\langle n_i | n_i \rangle}} \quad (62)$$

satisfy

$$\langle \nu_i | \nu_j \rangle = \delta_{ij} \quad (63)$$

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

§ 18 *An example: constructing an orthonormal basis set in \mathbb{R}^3 .* In Cell 6 of Workbook3, I use a random number generator to create three vectors in \mathbb{R}^3 :

$$|o_1\rangle = \{0.445, -0.781, -0.059\} \quad (64)$$

$$|o_2\rangle = \{0.071, 0.166, -0.412\} \quad (65)$$

$$|o_3\rangle = \{-0.670, -0.202, 0.508\} \quad (66)$$

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

The scalar product in \mathbb{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 the set $\{|n_1\rangle, |n_2\rangle, |n_3\rangle\}$ of vectors that are perpendicular to each other. First,

$$|n_1\rangle = o_1 = \{0.445, -0.781, -0.059\} \quad (67)$$

Second,

$$|n_2\rangle = |o_2\rangle - \frac{\langle o_1 | o_2 \rangle}{\langle o_1 | o_1 \rangle} |o_1\rangle \quad (68)$$

Mathematica easily evaluates the right-hand side, to give

$$|n_2\rangle = \{0.112, 0.095, -0.417\} \quad (69)$$

One intermediate step is

$$\langle n_1 | o_2 \rangle = 0.445 \times 0.071 + (-0.741) \times 0.166 + (-0.059) \times (-0.412) = -0.074$$

The other scalar products are calculated in a similar way. I also remind you that if α is a number then $\alpha|o_1\rangle = \{0.445\alpha, -0.781\alpha, -0.059\alpha\}$. With this reminder, the calculation of Eq. 69 should be clear.

Third, we calculate $|n_3\rangle$ from Eq. 59. `Mathematica` evaluates that to

$$|n_3\rangle = \{-0.251, -0.135, -0.098\} \quad (70)$$

Next we tested that $\langle n_i | n_j \rangle = 0$ if $i \neq j$. Then we used the formula

$$\langle \nu_i \rangle = \frac{|n_i\rangle}{\sqrt{\langle n_i | n_i \rangle}}$$

to construct vectors $|\nu_1\rangle$, $|\nu_2\rangle$, $|\nu_3\rangle$, which are orthonormal. The Gram-Schmidt procedure works just as expected.

The Gram-Schmidt procedure has a simple geometric interpretation. We can always 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. We want the component perpendicular to $|n_1\rangle$, which is

$$|n_2\rangle = |o_2\rangle - \alpha|n_1\rangle; \quad (71)$$

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

$$0 = \langle n_1 | n_2 \rangle = \langle n_1 | (|o_2\rangle - \alpha|n_1\rangle) \rangle = \langle n_1 | o_2 \rangle - \alpha \langle n_1 | n_1 \rangle \quad (72)$$

Rearranging this equation we find

$$\alpha = \frac{\langle n_1 | o_2 \rangle}{\langle n_1 | n_1 \rangle} \quad (73)$$

Inserting this value in Eq. 71 leads to the Gram-Schmidt formula Eq. 58.

Exercise 3 Use a similar geometric argument to derive Eq. 59.

§ 19 *If the original vectors are linearly dependent.* What would happen if one of the “old” vectors depends on the others? For example, what if

$$|o_3\rangle = a|o_1\rangle + b|o_2\rangle? \quad (74)$$

It is intuitively clear that this 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. 74, 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 6 of WorkBook3, I applied the Gram-Schmidt procedure to the set

$$|o_1\rangle = \{0.445, -0.781, -0.059\} \quad (75)$$

$$|o_2\rangle = \{0.071, 0.166, -0.412\} \quad (76)$$

$$|o_3\rangle = a|o_1\rangle + b|o_2\rangle \quad (77)$$

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

$$|n_1\rangle = |o_1\rangle = \{0.445, -0.781, -0.059\} \quad (78)$$

$$|n_2\rangle = \{0.112, 0.095, -0.417\} \quad (79)$$

$$|n_3\rangle = 0 \quad (80)$$

The 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 the new vector $|n_k\rangle$ corresponding to it is 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 the component of $|o_3\rangle$ that is perpendicular to the XOY plane; this component is zero.

The moral: none of the vectors in the starting set $\{|o_1\rangle, |o_2\rangle, \dots, |o_N\rangle\}$ should be linearly dependent on the others. If one of them is, then the Gram-Schmidt procedure will detect it and eliminate it from the orthonormal basis set that it constructs.

§ 20 *An example of constructing an orthonormal basis set in an L^2 space.* Let us construct a basis set in the space of the functions $f(x)$, $x \in [0, \pi]$ that are differentiable and satisfy

$$\int_0^\pi f(x)^2 dx < \infty \quad (81)$$

This space can be used to describe the wave function of a particle in a box whose walls are located at $x = 0$ and $x = \pi$. It can also describe a harmonic oscillator vibrating around $x = \pi/2$ whose maximum amplitude is smaller than $\pi/2$.

As usual, I define the scalar product to be

$$\langle f | \phi \rangle = \int_0^\pi f(x)\phi(x) dx \quad (82)$$

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

I start with the “old” functions

$$\langle x | o_i \rangle = o_i(x) = x^{i-1}, \quad i = 1, \dots, N \quad (83)$$

If I use these functions as a basis set, I can write, for any function $|f\rangle$,

$$|f\rangle = \sum_{i=1}^N \alpha_i |o_i\rangle \quad (84)$$

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

$$\langle x | f \rangle = f(x) = \sum_{i=1}^N \alpha_i \langle x | o_i \rangle = \sum_{i=1}^N \alpha_i o_i(x) = \sum_{i=1}^N \alpha_i x^{i-1} \quad (85)$$

The set $\{o_i(x)\}_{i=1}^N$ is not orthonormal. Let us start with it and construct an orthonormal set. The simple calculations needed for this are performed in

Cell 7 of the `Mathematica` file `WorkBook3.nb`. (I used `Mathematica` because I am lazy; all the calculations can be done by hand.)

Using Eq. 57,

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

which means that

$$\langle x | n_1 \rangle = \langle x | o_1 \rangle = 1 \quad (86)$$

Using Eq. 58:

$$\langle x | n_2 \rangle = \langle x | o_2 \rangle - \langle x | n_1 \rangle \frac{\langle n_1 | o_2 \rangle}{\langle n_1 | n_1 \rangle} \quad (87)$$

with

$$\begin{aligned} \langle n_1 | o_2 \rangle &= \int_0^\pi \langle n_1 | x \rangle \langle x | o_2 \rangle dx \\ &= \int_0^\pi n_1(x) o_2(x) dx = \int_0^\pi 1 \times x dx = \pi^2/2 \end{aligned} \quad (88)$$

$$\begin{aligned} \langle n_1 | n_1 \rangle &= \int_0^\pi \langle n_1 | x \rangle \langle x | n_1 \rangle dx \\ &= \int_0^\pi n_1(x) n_1(x) dx = \int_0^\pi dx = \pi \end{aligned} \quad (89)$$

Using Eqs. 88 and 89 in Eq. 87 gives

$$n_2(x) = o_2(x) - \frac{\pi}{2} = x - \frac{\pi}{2} \quad (90)$$

The third function is (use Eq. 59 with $j = 3$)

$$\begin{aligned} \langle x | n_3 \rangle &= o_3(x) - \langle x | n_1 \rangle \frac{\langle n_1 | o_3 \rangle}{\langle n_1 | n_1 \rangle} - \langle x | n_2 \rangle \frac{\langle n_2 | o_3 \rangle}{\langle n_2 | n_2 \rangle} \\ &= \frac{\pi^2}{6} - \pi x + x^2 \end{aligned} \quad (91)$$

Similarly, we obtain (Cell 7, `WorkBook3`)

$$\langle x | n_4 \rangle \equiv n_4(x) = -\frac{\pi^3}{20} + \frac{3\pi^2}{5}x - \frac{3\pi}{2}x^2 + x^3 \quad (92)$$

You can verify, by performing the integrals, that

$$\langle n_i | n_j \rangle = \int_0^\pi \langle n_i | x \rangle \langle x | n_j \rangle dx = \int_0^\pi n_i(x) n_j(x) dx = 0 \text{ when } i \neq j \quad (93)$$

The new functions are orthogonal. However, they are not normalized. To normalize them we use (see Eq. 62)

$$\nu_i(x) \equiv \frac{n_i(x)}{\sqrt{\langle n_i | n_i \rangle}} \quad (94)$$

The result is (see Cell 7, Workbook3)

$$\nu_1(x) = \frac{1}{\sqrt{\pi}} \quad (95)$$

$$\nu_2(x) = -\sqrt{\frac{3}{\pi}} + \frac{2\sqrt{3}}{\pi^{3/2}} x \quad (96)$$

$$\nu_3(x) = \sqrt{\frac{5}{\pi}} - \frac{6\sqrt{5}}{\pi^{3/2}} x + \frac{6\sqrt{5}}{\pi^{5/2}} x^2 \quad (97)$$

$$\nu_4(x) = -\sqrt{\frac{7}{\pi}} + \frac{12\sqrt{7}}{\pi^{3/2}} x - \frac{30\sqrt{7}}{\pi^{5/2}} x^2 + \frac{20\sqrt{7}}{\pi^{7/2}} x^3 \quad (98)$$

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

$$\langle x | f \rangle = f(x) \equiv x^2 \sin(x) \quad (99)$$

The expansion is

$$|f\rangle \cong \sum_{i=1}^m a_i |\nu_i\rangle \quad (100)$$

Acting with $\langle \nu_j |$ on Eq. 100 gives

$$\langle \nu_j | f \rangle \cong \sum_{i=1}^m a_i \langle \nu_j | \nu_i \rangle = \sum_{i=1}^m a_i \delta_{ji} = a_j \quad (101)$$

If we introduce the expression $\langle \nu_i | f \rangle = a_i$ into Eq. 100, we have

$$|f\rangle \cong \sum_{i=1}^m \langle \nu_i | f \rangle |\nu_i\rangle = \left(\sum_{i=1}^m |\nu_i\rangle \langle \nu_i| \right) |f\rangle \quad (102)$$

which implies that

$$\sum_{i=1}^m |\nu_i\rangle \langle \nu_i| \cong \hat{I} \quad (103)$$

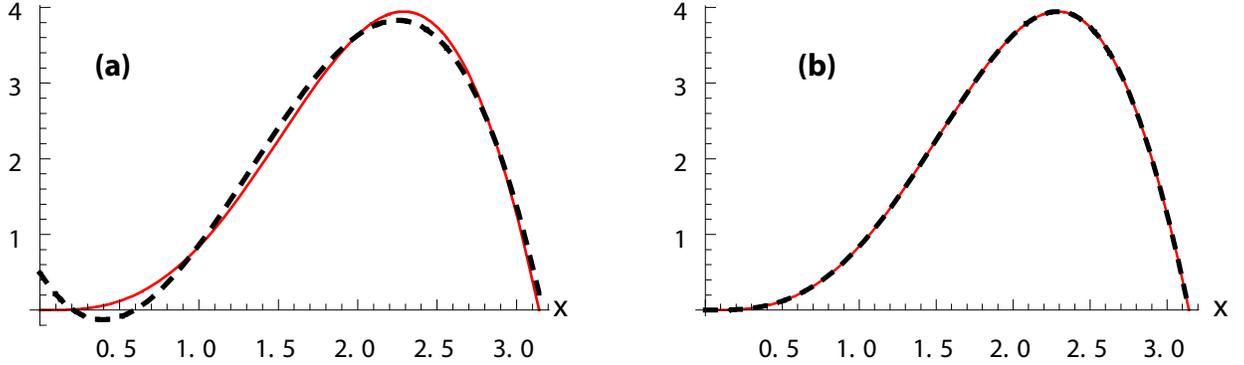


Figure 7: (a) The solid, red curve is $x^2 \sin(x)$ and the dashed, black curve is its representation by a basis set of five orthonormal polynomials. (b) Same as in (a) but with a basis set of ten polynomials. The plots were made in Cell 8.3 of Workbook3.nb.

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. 100) of all the function in the space, is equivalent to assuming that the set satisfies the completeness relation, Eq. 103.

By acting with the bra $\langle x|$ on Eqs. 100–102, we convert them to Schrödinger representation

$$\langle x | f \rangle \equiv x^2 \sin(x) \cong \sum_{i=1}^m a_i \langle x | \nu_i \rangle = \sum_{i=1}^m a_i \nu_i(x) \quad (104)$$

where (see Eq. 101)

$$a_i = \langle \nu_i | f \rangle = \int_0^\pi \langle \nu_i | x \rangle \langle x | f \rangle dx = \int_0^\pi \nu_i(x) x^2 \sin(x) dx \quad (105)$$

I can now test the theory by calculating a_i from Eq. 105 and inserting the result in Eq. 104. If the sum in Eq. 104 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 has the following steps.

1. Assume that the representation given by Eq. 104 is plausible.

2. Calculate the coefficients a_n from Eq. 105, by using the polynomials $\nu_i(x)$ given by Eqs. 95–98.
3. Test whether the sum in Eq. 104 is a good approximation to $f(x) = \langle x | f \rangle = x^2 \sin(x)$.

In Cell 8.3 of Workbook3, we calculate the coefficients a_n by performing the integrals in Eq. 105 For example (see Cell 8.3),

$$\begin{aligned} a_3 &= \int_0^\pi \nu_3(x) x^2 \sin(x) dx \\ &= \int_0^\pi \left(\sqrt{\frac{5}{\pi}} - \frac{6\sqrt{5}}{\pi^{3/2}}x + \frac{6\sqrt{5}}{\pi^{5/2}}x^2 \right) x^2 \sin(x) dx = -1.19836 \end{aligned} \quad (106)$$

The others are

$$a_1 = 3.31157, a_2 = 1.82699, \text{ and } a_4 = -1.2868 \quad (107)$$

The expansion is therefore

$$x^2 \sin(x) \cong \nu_1(x) + 3.31\nu_2(x) - 1.198\nu_3(x) - 1.29\nu_4(x) \quad (108)$$

with $\nu_1(x), \dots, \nu_4(x)$ given by Eqs. 95–98. Using these equations in Eq. 108 gives

$$x^2 \sin(x) \cong 0.49 - 3.31x + 4.92x^2 - 1.24x^3 \quad (109)$$

In Fig. 7(a) I show the function $x^2 \sin(x)$ together with the fit with $m = 5$ in Eq. 104. 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. 109 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

$$f(x) \cong \sum_{k=0}^4 \frac{1}{k!} \left(\frac{\partial^k f}{\partial x^k} \right)_{x=0} x^k \quad (110)$$

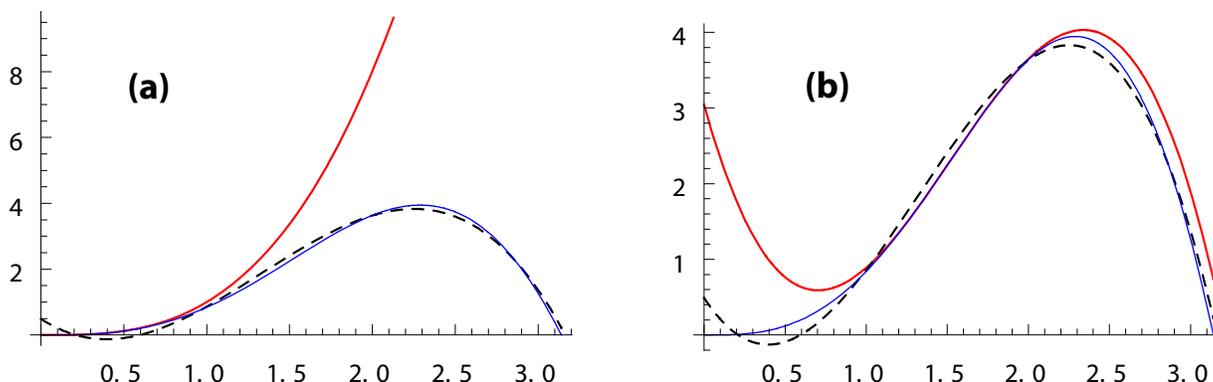


Figure 8: (a) The third-order power-series expansion of $x^2 \sin(x)$ around $x = 0$ (in red), $x^2 \sin(x)$ (in blue), and the polynomial expansion to third order (in black, dashed). (b) Same as in (a) except that the power-series expansion is around $x = \pi/2$.

In the case of $f(x) = x^2 \sin(x)$, this gives (see Cell 8.4 of Workbook3.nb)

$$x^2 \sin(x) \cong x^3 \quad (111)$$

In Fig. 8a, I give plots made in Cell 8.4 of Workbook3 for $x^2 \sin(x)$, its representation Eq. 109, and the Taylor-series approximation Eq. 111. 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 $x = 0$.

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

$$x^2 \sin(x) \cong 3.04 - 7.75x + 7.17x^2 - 1.57x^3 \quad (112)$$

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 In constructing orthonormal basis sets, we follow the example of the space \mathbb{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 $|\phi_i\rangle$, $i = 1, 2, \dots, N$, is to allow us to represent an arbitrary ket $|\psi\rangle$ in the form

$$|\psi\rangle = \sum_{i=1}^N a_i |\phi_i\rangle \quad (113)$$

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 Schrödinger representation this amounts to performing a large number of integrals involving $\langle x | \phi_i \rangle$. Thus, besides completeness (which means that Eq. 113 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 \mathbb{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 \mathbb{R}^3 , because it cannot represent all the vectors of \mathbb{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 \mathbb{R}^2 subspace of \mathbb{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 \mathbb{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

$$\vec{\eta} = \eta_1 \vec{a}_1 + \eta_2 \vec{a}_2 + \eta_3 \vec{a}_3 \quad (114)$$

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

$$\vec{a}_3 = n_1 \vec{a}_1 + n_2 \vec{a}_2 \quad (115)$$

and therefore the third term in Eq. 114 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

$$\sum_{i=1}^N n_i |\phi_i\rangle = 0 \quad (116)$$

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. 116, to obtain

$$\sum_{i=1}^N n_i \langle\phi_j | \phi_i\rangle = 0 \quad (117)$$

We call the number

$$S_{ji} \equiv \langle \phi_j | \phi_i \rangle \quad (118)$$

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. 117 as

$$\sum_{i=1}^N S_{ji} n_i = 0 \quad (119)$$

As j ranges from 1 to N , this constitutes a *homogeneous system of equations* for the unknowns n_i , which is equivalent to Eq. 117. We make use of the following theorem: if the determinant of the matrix (S_{ji}) differs from zero,

$$\det(S_{ij}) \neq 0, \quad (120)$$

then the only solution of Eq. 119 is

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

If

$$\det(S_{ij}) = 0, \quad (121)$$

then Eq. 119 has non-zero solutions.

But the system in Eq. 119 is equivalent to Eq. 117. *We conclude that if $\det(S_{ij}) \neq 0$ then the solution of $\sum_i n_i |\phi_i\rangle = 0$ is $n_1 = \dots = n_N = 0$ 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

$$S_{ij} = \langle \phi_i | \phi_j \rangle = 0 \text{ if } i \neq j \quad (122)$$

The system of equations Eq. 119 reduces to

$$\begin{aligned} S_{11}n_1 &= 0 \\ S_{22}n_2 &= 0 \\ &\vdots \\ S_{NN}n_N &= 0 \end{aligned} \quad (123)$$

Each “self-overlap” $S_{ii} = \langle \phi_i | \phi_i \rangle$ is nonzero and therefore $n_1 = 0$, $n_2 = 0$, \dots , $n_N = 0$. 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 \mathbb{R}^3 . If we regard it as a basis set in \mathbb{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 $\{|a_1\rangle, |a_2\rangle, \dots, |a_N\rangle\}$ to represent kets $|v\rangle$ by expressions

$$|v\rangle = \sum_{i=1}^N c_i |a_i\rangle \quad (124)$$

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

$$\langle a_j | v \rangle = \sum_{i=1}^N c_i \langle a_j | a_i \rangle = \sum_{i=1}^N S_{ji} c_i \quad (125)$$

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. 125. 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. 125 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. 125 is more difficult when the basis set is not orthonormal. If it is orthonormal then

$$S_{ij} = \delta_{ij} \quad (126)$$

and Eq. 125 becomes

$$c_j = \langle a_j | v \rangle \quad (127)$$

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 constructed them as a linear combination of atomic orbitals. For example, one of them is

$$\psi(\vec{r}) = \frac{1}{\sqrt{2}}[1s_A(\vec{r}_A) + 1s_B(\vec{r}_B)] \quad (128)$$

Here $1s_A(\vec{r}_A)$ is the $1s$ orbital of hydrogen atom A and \vec{r}_A is the vector giving the position of the electron with respect to proton A. $1s_B(\vec{r}_B)$ has a similar meaning, for proton B.

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 that $1s_B$ performs a similar service in the neighborhood of B. The sum in Eq. 128 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

$$\psi(\vec{r}) = a_1 1s_A(\vec{r}_A) + a_2 1s_B(\vec{r}_B) + a_3 2p_{xA}(\vec{r}_A) + a_4 2p_{xB}(\vec{r}_B) \quad (129)$$

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

$$\psi(x) = \sum_n a_n \phi_n(x) \quad (130)$$

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. 130 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.