CHAPTER 3

FORMALISM

3.1 HILBERT SPACE

In the last two chapters we have stumbled on a number of interesting properties of simple quantum systems. Some of these are "accidental" features of specific potentials (the even spacing of energy levels for the harmonic oscillator, for example), but others seem to be more general, and it would be nice to prove them once and for all (the uncertainty principle, for instance, and the orthogonality of stationary states). The purpose of this chapter is to recast the theory in a more powerful form, with that in mind. There is not much here that is genuinely *new*; the idea, rather, is to make coherent sense of what we have already discovered in particular cases.

Quantum theory is based on two constructs: *wave functions* and *operators*. The state of a system is represented by its wave function, observables are represented by operators. Mathematically, wave functions satisfy the defining conditions for abstract **vectors**, and operators act on them as **linear transformations**. So the natural language of quantum mechanics is **linear algebra**.[1]

But it is not, I suspect, a form of linear algebra with which you are immediately familiar. In an N-dimensional space it is simplest to represent a vector, $|\alpha\rangle$, by the N-tuple of its components, $\{a_n\}$, with respect to a specified orthonormal basis:

$$|\alpha\rangle \rightarrow \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}. \qquad [3.1]$$

[1] If you have never studied linear algebra, you should read the Appendix before continuing.

The **inner product**, $\langle \alpha | \beta \rangle$, of two vectors (generalizing the dot product in three dimensions) is the complex number,

$$\langle \alpha | \beta \rangle = a_1^* b_1 + a_2^* b_2 + \cdots + a_N^* b_N. \quad [3.2]$$

Linear transformations, T, are represented by **matrices** (with respect to the specified basis), which act on vectors (to produce new vectors) by the ordinary rules of matrix multiplication:

$$|\beta\rangle = T|\alpha\rangle \rightarrow \mathbf{b} = \mathbf{T}\mathbf{a} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1N} \\ t_{21} & t_{22} & \cdots & t_{2N} \\ \vdots & \vdots & & \vdots \\ t_{N1} & t_{N2} & \cdots & t_{NN} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}. \quad [3.3]$$

But the "vectors" we encounter in quantum mechanics are (for the most part) *functions*, and they live in *infinite*-dimensional spaces. For them the N-tuple/matrix notation is awkward, at best, and manipulations that are well-behaved in the finite-dimensional case can be problematic. (The underlying reason is that whereas the *finite* sum in Equation 3.2 always exists, an *infinite* sum—or an integral—may not converge, in which case the inner product does not exist, and any argument involving inner products is immediately suspect.) So even though most of the terminology and notation should be familiar, it pays to approach this subject with caution.

The collection of *all* functions of x constitutes a vector space, but for our purposes it is much too large. To represent a possible physical state, the wave function Ψ must be *normalized*:

$$\int |\Psi|^2 \, dx = 1.$$

The set of all **square-integrable functions**, on a specified interval,[2]

$$f(x) \quad \text{such that} \quad \int_a^b |f(x)|^2 \, dx < \infty, \quad [3.4]$$

constitutes a (much smaller) vector space (see Problem 3.1(a)). Mathematicians call it $L^2(a, b)$; physicists call it **Hilbert space.**[3] In quantum mechanics, then,

Wave functions live in Hilbert space. [3.5]

[2] For us, the limits (a and b) will almost always be $\pm\infty$, but we might as well keep things more general for the moment.

[3] Technically, a Hilbert space is a **complete inner product space**, and the collection of square-integrable functions is only *one example* of a Hilbert space—indeed, every finite-dimensional vector space is trivially a Hilbert space. But since L^2 is the arena of quantum mechanics, it's what physicists generally *mean* when they say "Hilbert space." By the way, the word **complete** here means that any Cauchy sequence of functions in Hilbert space converges to a function that is also in the space: it has no "holes" in it, just as the set of all real numbers has no holes (by contrast, the space of all *polynomials*, for example, like the set of all *rational* numbers, certainly *does* have holes in it). The completeness of a *space* has nothing to do with the completeness (same word, unfortunately) of a *set of functions*, which is the property that any other function can be expressed as a linear combination of them.

We define the **inner product of two functions**, $f(x)$ and $g(x)$, as follows:

$$\langle f|g\rangle \equiv \int_a^b f(x)^* g(x)\,dx. \qquad [3.6]$$

If f and g are both square-integrable (that is, if they are both in Hilbert space), their inner product is guaranteed to exist (the integral in Equation 3.6 converges to a finite number).[4] This follows from the integral **Schwarz inequality**:[5]

$$\left|\int_a^b f(x)^* g(x)\,dx\right| \leq \sqrt{\int_a^b |f(x)|^2\,dx \int_a^b |g(x)|^2\,dx}. \qquad [3.7]$$

You can check for yourself that Equation 3.6 satisfies all the conditions for an inner product (Problem 3.1(b)). Notice in particular that

$$\langle g|f\rangle = \langle f|g\rangle^*. \qquad [3.8]$$

Moreover, the inner product of $f(x)$ with *itself*,

$$\langle f|f\rangle = \int_a^b |f(x)|^2\,dx, \qquad [3.9]$$

is *real* and non-negative; it's *zero* only[6] when $f(x) = 0$.

A function is said to be **normalized** if its inner product with itself is 1; two functions are **orthogonal** if their inner product is 0; and a *set* of functions, $\{f_n\}$, is **orthonormal** if they are normalized and mutually orthogonal:

$$\langle f_m|f_n\rangle = \delta_{mn}. \qquad [3.10]$$

Finally, a set of functions is **complete** if any *other* function (in Hilbert space) can be expressed as a linear combination of them:

$$f(x) = \sum_{n=1}^{\infty} c_n f_n(x). \qquad [3.11]$$

---

[4] In Chapter 2 we were obliged on occasion to work with functions that were *not* normalizable. Such functions lie *outside* Hilbert space, and we are going to have to handle them with special care, as you will see shortly. For the moment, I shall assume that all the functions we encounter *are* in Hilbert space.

[5] For a proof, see F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Unger, New York, 1955), Section 21. In a *finite* dimensional vector space the Schwarz inequality, $|\langle\alpha|\beta\rangle|^2 \leq \langle\alpha|\alpha\rangle\langle\beta|\beta\rangle$, is easy to prove (see Problem A.5). But that proof *assumes* the existence of the inner products, which is precisely what we are trying to *establish* here.

[6] What about a function that is zero everywhere except at a few isolated points? The integral (Equation 3.9) would still vanish, even though the function itself does not. If this bothers you, you should have been a math major. In physics such pathological functions do not occur, but in any case, in Hilbert space two functions that have the same square integral are considered equivalent. Technically, vectors in Hilbert space represent **equivalence classes** of functions.

If the functions $\{f_n(x)\}$ are orthonormal, the coefficients are given by Fourier's trick:

$$c_n = \langle f_n | f \rangle. \qquad [3.12]$$

as you can check for yourself. I anticipated this terminology, of course, back in Chapter 2. (The stationary states for the infinite square well (Equation 2.28) constitute a complete orthonormal set on the interval $(0, a)$; the stationary states for the harmonic oscillator (Equation 2.67 or 2.85) are a complete orthonormal set on the interval $(-\infty, \infty)$.)

Problem 3.1

(a) Show that the set of all square-integrable functions is a vector space (refer to Section A.1 for the definition). *Hint:* The main problem is to show that the sum of two square-integrable functions is itself square-integrable. Use Equation 3.7. Is the set of all *normalized* functions a vector space?

(b) Show that the integral in Equation 3.6 satisfies the conditions for an inner product (Section A.2).

***Problem 3.2**

(a) For what range of ν is the function $f(x) = x^\nu$ in Hilbert space, on the interval $(0, 1)$? Assume ν is real, but not necessarily positive.

(b) For the specific case $\nu = 1/2$, is $f(x)$ in this Hilbert space? What about $xf(x)$? How about $(d/dx)f(x)$?

3.2 OBSERVABLES

3.2.1 Hermitian Operators

The expectation value of an observable $Q(x, p)$ can be expressed very neatly in inner-product notation:[7]

$$\langle Q \rangle = \int \Psi^* \hat{Q} \Psi \, dx = \langle \Psi | \hat{Q} \Psi \rangle. \qquad [3.13]$$

---

[7] Remember that $\hat{Q}$ is the operator constructed from Q by the replacement $p \to \hat{p} \equiv (\hbar/i)d/dx$. These operators are **linear**, in the sense that

$$\hat{Q}[af(x) + bg(x)] = a\hat{Q}f(x) + b\hat{Q}g(x),$$

for any functions f and g and any complex numbers a and b. They constitute *linear transformations* (Section A.3) on the space of all functions. However, they sometimes carry a function *inside* Hilbert

Now, the outcome of a measurement has got to be *real*, and so, *a fortiori*, is the *average* of many measurements:

$$\langle Q \rangle = \langle Q \rangle^*. \qquad [3.14]$$

But the complex conjugate of an inner product reverses the order (Equation 3.8), so

$$\langle \Psi | \hat{Q}\Psi \rangle = \langle \hat{Q}\Psi | \Psi \rangle. \qquad [3.15]$$

and this must hold true for any wave function Ψ. Thus operators representing *observables* have the very special property that

$$\langle f | \hat{Q} f \rangle = \langle \hat{Q} f | f \rangle \quad \text{for all } f(x). \qquad [3.16]$$

We call such operators **hermitian.**

Actually, most books require an ostensibly stronger condition:

$$\langle f | \hat{Q} g \rangle = \langle \hat{Q} f | g \rangle \quad \text{for all } f(x) \text{ and all } g(x). \qquad [3.17]$$

But it turns out, in spite of appearances, that this is perfectly equivalent to my definition (Equation 3.16), as you will prove in Problem 3.3. So use whichever you like. The essential point is that a hermitian operator can be applied either to the first member of an inner product or to the second, with the same result, and hermitian operators naturally arise in quantum mechanics because their expectation values are real:

Observables are represented by hermitian operators. [3.18]

Well, let's *check* this. Is the momentum operator, for example, hermitian?

$$\langle f | \hat{p} g \rangle = \int_{-\infty}^{\infty} f^* \frac{\hbar}{i} \frac{dg}{dx}\, dx = \frac{\hbar}{i} f^* g \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \left(\frac{\hbar}{i} \frac{df}{dx} \right)^* g\, dx = \langle \hat{p} f | g \rangle. \qquad [3.19]$$

I used integration by parts, of course, and threw away the boundary term for the usual reason: If $f(x)$ and $g(x)$ are square integrable, they must go to zero at $\pm\infty$.[8]

---

space into a function *outside* it (see Problem 3.2(b)), and in this case the domain of the operator may have to be restricted.

[8] Actually, this is not quite true. As I mention in Chapter 1, there exist pathological functions that are square-integrable but do *not* go to zero at infinity. However, such functions do not arise in physics, and if you are worried about it we will simply restrict the domain of our operators to exclude them. On *finite* intervals, though, you really *do* have to be more careful with the boundary terms, and an operator that is hermitian on $(-\infty, \infty)$ may *not* be hermitian on $(0, \infty)$ or $(-\pi, \pi)$. If you're wondering about the infinite square well, it's safest to think of those wave functions as residing on the infinite line—they just happen to be *zero* outside $(0, a)$.

Notice how the complex conjugation of i compensates for the minus sign picked up from integration by parts—the operator d/dx (without the i) is *not* hermitian, and it does not represent a possible observable.

---

*Problem 3.3 Show that if $\langle h|\hat{Q}h\rangle = \langle \hat{Q}h|h\rangle$ for all functions h (in Hilbert space), then $\langle f|\hat{Q}g\rangle = \langle \hat{Q}f|g\rangle$ for all f and g (i.e., the two definitions of "hermitian"—Equations 3.16 and 3.17—are equivalent). *Hint:* First let $h = f + g$, and then let $h = f + ig$.

Problem 3.4

(a) Show that the *sum* of two hermitian operators is hermitian.

(b) Suppose $\hat{Q}$ is hermitian, and α is a complex number. Under what condition (on α) is $\alpha\hat{Q}$ hermitian?

(c) When is the *product* of two hermitian operators hermitian?

(d) Show that the position operator ($\hat{x} = x$) and the hamiltonian operator ($\hat{H} = -(\hbar^2/2m)d^2/dx^2 + V(x)$) are hermitian.

Problem 3.5 The **hermitian conjugate** (or **adjoint**) of an operator $\hat{Q}$ is the operator $\hat{Q}^\dagger$ such that

$$\langle f|\hat{Q}g\rangle = \langle \hat{Q}^\dagger f|g\rangle \quad \text{(for all } f \text{ and } g\text{)}. \tag{3.20}$$

(A hermitian operator, then, is equal to its hermitian conjugate: $\hat{Q} = \hat{Q}^\dagger$.)

(a) Find the hermitian conjugates of x, i, and d/dx.

(b) Construct the hermitian conjugate of the harmonic oscillator raising operator, a_+ (Equation 2.47).

(c) Show that $(\hat{Q}\hat{R})^\dagger = \hat{R}^\dagger\hat{Q}^\dagger$.

3.2.2 Determinate States

Ordinarily, when you measure an observable Q on an ensemble of identically prepared systems, all in the same state Ψ, you do *not* get the same result each time—this is the *indeterminacy* of quantum mechanics.[9] *Question:* Would it be possible to prepare a state such that *every* measurement of Q is certain to return the *same* value (call it q)? This would be, if you like, a **determinate state**, for the observable Q. (Actually, we already know one example: Stationary states are determinate states of the Hamiltonian; a measurement of the total energy, on a

---

[9] I'm talking about *competent* measurements, of course—it's always possible to make a *mistake*, and simply get the wrong answer, but that's not the fault of quantum mechanics.

particle in the stationary state Ψ_n, is certain to yield the corresponding "allowed" energy E_n.)

Well, the standard deviation of Q, in a determinate state, would be *zero*, which is to say,

$$\sigma^2 = \langle(\hat{Q} - \langle Q\rangle)^2\rangle = \langle\Psi|(\hat{Q} - q)^2\Psi\rangle = \langle(\hat{Q} - q)\Psi|(\hat{Q} - q)\Psi\rangle = 0. \quad [3.21]$$

(Of course, if every measurement gives q, their average is also q: $\langle Q\rangle = q$. I also used the fact that $\hat{Q}$, and hence also $\hat{Q} - q$, is a *hermitian* operator, to move one factor over to the first term in the inner product.) But the only function whose inner product with itself vanishes is 0, so

$$\hat{Q}\Psi = q\Psi. \quad [3.22]$$

This is the **eigenvalue equation** for the operator $\hat{Q}$; Ψ is an **eigenfunction** of $\hat{Q}$, and q is the corresponding **eigenvalue**. Thus

Determinate states are eigenfunctions of $\hat{Q}$. [3.23]

Measurement of Q on such a state is certain to yield the eigenvalue, q.

Note that the eigen*value* is a *number* (not an operator or a function). You can multiply any eigenfunction by a constant, and it is still an eigenfunction, with the same eigenvalue. Zero does not count as an eigenfunction (we exclude it by definition—otherwise *every* number would be an eigenvalue, since $\hat{Q}\,0 = q\,0 = 0$ for any operator $\hat{Q}$ and all q). But there's nothing wrong with zero as an eigen*value*. The collection of all the eigenvalues of an operator is called its **spectrum**. Sometimes two (or more) linearly independent eigenfunctions share the same eigenvalue; in that case the spectrum is said to be **degenerate**.

For example, determinate states of the total energy are eigenfunctions of the Hamiltonian:

$$\hat{H}\psi = E\psi, \quad [3.24]$$

which is precisely the time-independent Schrödinger equation. In this context we use the letter E for the eigenvalue, and the lower case ψ for the eigenfunction (tack on the factor $\exp(-iEt/\hbar)$ to make it Ψ, if you like; it's still an eigenfunction of H).

Example 3.1 Consider the operator

$$\hat{Q} \equiv i\frac{d}{d\phi}, \quad [3.25]$$

where ϕ is the usual polar coordinate in two dimensions. (This operator might arise in a physical context if we were studying the bead-on-a-ring; see Problem 2.46.) Is $\hat{Q}$ hermitian? Find its eigenfunctions and eigenvalues.

Solution: Here we are working with functions $f(\phi)$ on the *finite* interval $0 \leq \phi \leq 2\pi$, and stipulate that

$$f(\phi + 2\pi) = f(\phi). \qquad [3.26]$$

since ϕ and $\phi + 2\pi$ describe the same physical point. Using integration by parts,

$$\langle f|\hat{Q}\,g\rangle = \int_0^{2\pi} f^* \left(i\frac{dg}{d\phi}\right) d\phi = if^*g\Big|_0^{2\pi} - \int_0^{2\pi} i\left(\frac{df^*}{d\phi}\right) g\,d\phi = \langle \hat{Q}\,f|g\rangle,$$

so $\hat{Q}$ *is* hermitian (this time the boundary term disappears by virtue of Equation 3.26).

The eigenvalue equation,

$$i\frac{d}{d\phi}f(\phi) = qf(\phi), \qquad [3.27]$$

has the general solution

$$f(\phi) = Ae^{-iq\phi}. \qquad [3.28]$$

Equation 3.26 restricts the possible values of the q:

$$e^{-iq2\pi} = 1 \quad \Rightarrow \quad q = 0, \pm 1, \pm 2, \ldots \qquad [3.29]$$

The spectrum of this operator is the set of all integers, and it is nondegenerate.

Problem 3.6 Consider the operator $\hat{Q} = d^2/d\phi^2$, where (as in Example 3.1) ϕ is the azimuthal angle in polar coordinates, and the functions are subject to Equation 3.26. Is $\hat{Q}$ hermitian? Find its eigenfunctions and eigenvalues. What is the spectrum of $\hat{Q}$? Is the spectrum degenerate?

3.3 EIGENFUNCTIONS OF A HERMITIAN OPERATOR

Our attention is thus directed to the *eigenfunctions of hermitian operators* (physically: determinate states of observables). These fall into two categories: If the spectrum is **discrete** (i.e., the eigenvalues are separated from one another) then the eigenfunctions lie in Hilbert space and they constitute physically realizable states. If the spectrum is **continuous** (i.e., the eigenvalues fill out an entire range) then the eigenfunctions are not normalizable, and they do not represent possible wave functions (though *linear combinations* of them—involving necessarily a spread in eigenvalues—may be normalizable). Some operators have a discrete spectrum only (for example, the Hamiltonian for the harmonic oscillator), some have only a continuous spectrum (for example, the free particle Hamiltonian), and some have both a discrete part and a continuous part (for example, the Hamiltonian for a

finite square well). The discrete case is easier to handle, because the relevant inner products are guaranteed to exist—in fact, it is very similar to the finite-dimensional theory (the eigenvectors of a hermitian *matrix*). I'll treat the discrete case first, and then the continuous one.

3.3.1 Discrete Spectra

Mathematically, the normalizable eigenfunctions of a hermitian operator have two important properties:

Theorem 1: Their eigen*values* are *real.*

Proof: Suppose

$$\hat{Q}f = qf,$$

(i.e., $f(x)$ is an eigenfunction of $\hat{Q}$, with eigenvalue q), and[10]

$$\langle f|\hat{Q}f\rangle = \langle \hat{Q}f|f\rangle$$

($\hat{Q}$ is hermitian). Then

$$q\langle f|f\rangle = q^*\langle f|f\rangle$$

(q is a *number*, so it comes outside the integral, and because the first function in the inner product is complex conjugated (Equation 3.6), so too is the q on the right). But $\langle f|f\rangle$ cannot be zero ($f(x) = 0$ is not a legal eigenfunction), so $q = q^*$, and hence q is real. QED

This is comforting: If you measure an observable on a particle in a determinate state, you will at least get a real number.

Theorem 2: Eigenfunctions belonging to distinct eigenvalues are *orthogonal.*

Proof: Suppose

$$\hat{Q}f = qf, \quad \text{and} \quad \hat{Q}g = q'g,$$

and $\hat{Q}$ is hermitian. Then $\langle f|\hat{Q}g\rangle = \langle \hat{Q}f|g\rangle$, so

$$q'\langle f|g\rangle = q^*\langle f|g\rangle$$

(again, the inner products exist because the eigenfunctions are in Hilbert space by assumption). But q is real (from Theorem 1), so if $q' \neq q$ it must be that $\langle f|g\rangle = 0$. QED

---

[10] It is here that we assume the eigenfunctions are in Hilbert space—otherwise the inner product might not exist at all.

That's why the stationary states of the infinite square well, for example, or the harmonic oscillator, are orthogonal—they are eigenfunctions of the Hamiltonian with distinct eigenvalues. But this property is not peculiar to them, or even to the Hamiltonian—the same holds for determinate states of *any* observable.

Unfortunately, Theorem 2 tells us nothing about degenerate states ($q' = q$). However, if two (or more) eigenfunctions share the same eigenvalue, any linear combination of them is itself an eigenfunction, with the same eigenvalue (Problem 3.7(a)), and we can use the **Gram-Schmidt orthogonalization procedure** (Problem A.4) to *construct* orthogonal eigenfunctions within each degenerate subspace. It is almost never necessary to do this explicitly (thank God!), but it can always be done in principle. So *even in the presence of degeneracy* the eigenfunctions can be *chosen* to be orthogonal, and in setting up the formalism of quantum mechanics we shall assume that this has already been done. That licenses the use of Fourier's trick, which depends on the orthonormality of the basis functions.

In a *finite*-dimensional vector space the eigenvectors of a hermitian matrix have a third fundamental property: They span the space (every vector can be expressed as a linear combination of them). Unfortunately, the proof does not generalize to infinite-dimensional spaces. But the property itself is essential to the internal consistency of quantum mechanics, so (following Dirac[11]) we will take it as an *axiom* (or, more precisely, as a restriction on the class of hermitian operators that can represent observables):

> **Axiom:** The eigenfunctions of an observable operator are *complete*: Any function (in Hilbert space) can be expressed as a linear combination of them.[12]

Problem 3.7

(a) Suppose that $f(x)$ and $g(x)$ are two eigenfunctions of an operator $\hat{Q}$, with the same eigenvalue q. Show that any linear combination of f and g is itself an eigenfunction of $\hat{Q}$, with eigenvalue q.

(b) Check that $f(x) = \exp(x)$ and $g(x) = \exp(-x)$ are eigenfunctions of the operator d^2/dx^2, with the same eigenvalue. Construct two linear combinations of f and g that are *orthogonal* eigenfunctions on the interval $(-1, 1)$.

---

[11] P. A. M. Dirac, *The Principles of Quantum Mechanics*, Oxford University Press, New York (1958).

[12] In some specific cases completeness is provable (we know that the stationary states of the infinite square well, for example, are complete, because of Dirichlet's theorem). It is a little awkward to call something an "axiom" that is *provable* in some cases, but I don't know a better way to handle it.

Problem 3.8

(a) Check that the eigenvalues of the hermitian operator in Example 3.1 are real. Show that the eigenfunctions (for distinct eigenvalues) are orthogonal.

(b) Do the same for the operator in Problem 3.6.

3.3.2 Continuous Spectra

If the spectrum of a hermitian operator is *continuous*, the eigenfunctions are not normalizable, and the proofs of Theorems 1 and 2 fail, because the inner products may not exist. Nevertheless, there is a sense in which the three essential properties (reality, orthogonality, and completeness) still hold. I think it's best to approach this subtle case through specific examples.

Example 3.2 Find the eigenfunctions and eigenvalues of the momentum operator.

Solution: Let $f_p(x)$ be the eigenfunction and p the eigenvalue:

$$\frac{\hbar}{i}\frac{d}{dx}f_p(x) = pf_p(x). \qquad [3.30]$$

The general solution is

$$f_p(x) = Ae^{ipx/\hbar}.$$

This is not square-integrable, for *any* (complex) value of p—the momentum operator has *no* eigenfunctions in Hilbert space. And yet, if we restrict ourselves to *real* eigenvalues, we do recover a kind of *ersatz* "orthonormality." Referring to Problems 2.24(a) and 2.26,

$$\int_{-\infty}^{\infty} f_{p'}^*(x)\, f_p(x)\, dx = |A|^2 \int_{-\infty}^{\infty} e^{i(p-p')x/\hbar}\, dx = |A|^2 2\pi\hbar\, \delta(p - p'). \qquad [3.31]$$

If we pick $A = 1/\sqrt{2\pi\hbar}$, so that

$$f_p(x) = \frac{1}{\sqrt{2\pi\hbar}} e^{ipx/\hbar}, \qquad [3.32]$$

then

$$\langle f_{p'} | f_p \rangle = \delta(p - p'), \qquad [3.33]$$

which is strikingly reminiscent of *true* orthonormality (Equation 3.10)—the indices are now continuous variables, and the Kronecker delta has become a Dirac delta, but otherwise it looks just the same. I'll call Equation 3.33 **Dirac orthonormality**.

Most important, the eigenfunctions are *complete*, with the sum (in Equation 3.11) replaced by an integral: Any (square-integrable) function $f(x)$ can be written in the form

$$f(x) = \int_{-\infty}^{\infty} c(p)\, f_p(x)\, dp = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} c(p) e^{ipx/\hbar}\, dp. \qquad [3.34]$$

The expansion coefficient (now a *function*, $c(p)$) is obtained, as always, by Fourier's trick:

$$\langle f_{p'} | f \rangle = \int_{-\infty}^{\infty} c(p) \langle f_{p'} | f_p \rangle\, dp = \int_{-\infty}^{\infty} c(p) \delta(p - p')\, dp = c(p'). \qquad [3.35]$$

Alternatively, you can get them from Plancherel's theorem (Equation 2.102), for the expansion (Equation 3.34) is nothing but a Fourier transform.

The eigenfunctions of momentum (Equation 3.32) are sinusoidal, with wavelength

$$\lambda = \frac{2\pi\hbar}{p}. \qquad [3.36]$$

This is the old de Broglie formula (Equation 1.39), which I promised to prove at the appropriate time. It turns out to be a little more subtle than de Broglie imagined, because we now know that there is actually *no such thing* as a particle with determinate momentum. But we could make a normalizable wave *packet* with a narrow range of momenta, and it is to such an object that the de Broglie relation applies.

What are we to make of Example 3.2? Although none of the eigenfunctions of $\hat{p}$ lives in Hilbert space, a certain family of them (those with real eigenvalues) reside in the nearby "suburbs," with a kind of quasi-normalizability. They do not represent possible physical states, but they are still very useful (as we have already seen, in our study of one-dimensional scattering).[13]

Example 3.3 Find the eigenfunctions and eigenvalues of the position operator.

Solution: Let $g_y(x)$ be the eigenfunction and y the eigenvalue:

$$x\, g_y(x) = y\, g_y(x). \qquad [3.37]$$

---

[13] What about the eigenfunctions with *non*real eigenvalues? These are not merely nonnormalizable—they actually blow up at $\pm\infty$. Functions in what I called the "suburbs" of Hilbert space (the entire metropolitan area is sometimes called a "rigged Hilbert space"; see, for example, Leslie Ballentine's *Quantum Mechanics: A Modern Development*, World Scientific, 1998) have the property that although they have no (finite) inner product with *themselves*, they *do* admit inner products with all members of Hilbert space. This is *not* true for eigenfunctions of $\hat{p}$ with nonreal eigenvalues. In particular, I showed that the momentum operator is hermitian *for functions in Hilbert space*, but the argument depended on dropping the boundary term (in Equation 3.19). That term is still zero if g is an eigenfunction of $\hat{p}$ with a real eigenvalue (as long as f is in Hilbert space), but not if the eigenvalue has an imaginary part. In this sense *any* complex number is an eigenvalue of the operator $\hat{p}$, but only *real* numbers are eigenvalues of the *hermitian* operator $\hat{p}$—the others lie outside the space over which $\hat{p}$ is hermitian.

Here y is a fixed number (for any given eigenfunction), but x is a continuous variable. What function of x has the property that multiplying it by x is the same as multiplying it by the constant y? Obviously it's got to be *zero*, except at the one point $x = y$; in fact, it is nothing but the Dirac delta function:

$$g_y(x) = A\delta(x - y).$$

This time the eigenvalue *has* to be real; the eigenfunctions are not square-integrable, but again they admit *Dirac* orthonormality:

$$\int_{-\infty}^{\infty} g_{y'}^*(x)\, g_y(x)\, dx = |A|^2 \int_{-\infty}^{\infty} \delta(x - y')\delta(x - y)\, dx = |A|^2 \delta(y - y'). \quad [3.38]$$

If we pick $A = 1$, so

$$g_y(x) = \delta(x - y), \quad [3.39]$$

then

$$\langle g_{y'} | g_y \rangle = \delta(y - y'). \quad [3.40]$$

These eigenfunctions are also *complete*:

$$f(x) = \int_{-\infty}^{\infty} c(y)\, g_y(x)\, dy = \int_{-\infty}^{\infty} c(y)\delta(x - y)\, dy, \quad [3.41]$$

with

$$c(y) = f(y) \quad [3.42]$$

(trivial, in this case, but you can get it from Fourier's trick if you insist).

If the spectrum of a hermitian operator is *continuous* (so the eigenvalues are labeled by a continuous variable—p or y, in the examples; z, generically, in what follows), the eigenfunctions are not normalizable, they are not in Hilbert space and they do not represent possible physical states; nevertheless, the eigenfunctions with real eigenvalues are *Dirac* orthonormalizable and complete (with the sum now an integral). Luckily, this is all we really require.

Problem 3.9

(a) Cite a Hamiltonian from Chapter 2 (*other* than the harmonic oscillator) that has only a *discrete* spectrum.

(b) Cite a Hamiltonian from Chapter 2 (*other* than the free particle) that has only a *continuous* spectrum.

APPENDIX

LINEAR ALGEBRA

Linear algebra abstracts and generalizes the arithmetic of ordinary vectors, such as those we encounter in first-year physics. The generalization is in two directions: (1) We allow the scalars to be *complex* numbers, and (2) we do not restrict ourselves to three dimensions.

A.1 VECTORS

A **vector space** consists of a set of **vectors** ($|\alpha\rangle$, $|\beta\rangle$, $|\gamma\rangle$, ...), together with a set of **scalars** (a, b, c, ...),[1] which is **closed**[2] under two operations: vector addition and scalar multiplication.

- **Vector Addition**

  The "sum" of any two vectors is another vector:

$$|\alpha\rangle + |\beta\rangle = |\gamma\rangle. \tag{A.1}$$

  Vector addition is **commutative**:

$$|\alpha\rangle + |\beta\rangle = |\beta\rangle + |\alpha\rangle. \tag{A.2}$$

[1] For our purposes, the scalars will be ordinary complex numbers. Mathematicians can tell you about vector spaces over more exotic fields, but such objects play no role in quantum mechanics. Note that α, β, γ ... are *not* (ordinarily) numbers; they are *names* (labels)—"Charlie," for instance, or "F43A-9GL," or whatever you care to use to identify the vector in question.

[2] That is to say, these operations are always well-defined, and will never carry you outside the vector space.

and **associative**:

$$|\alpha\rangle + (|\beta\rangle + |\gamma\rangle) = (|\alpha\rangle + |\beta\rangle) + |\gamma\rangle. \quad [A.3]$$

There exists a **zero** (or **null**) **vector**,[3] $|0\rangle$, with the property that

$$|\alpha\rangle + |0\rangle = |\alpha\rangle, \quad [A.4]$$

for every vector $|\alpha\rangle$. And for every vector $|\alpha\rangle$ there is an associated **inverse vector** ($|-\alpha\rangle$),[4] such that

$$|\alpha\rangle + |-\alpha\rangle = |0\rangle. \quad [A.5]$$

- **Scalar Multiplication**

  The "product" of any scalar with any vector is another vector:

$$a|\alpha\rangle = |\gamma\rangle. \quad [A.6]$$

  Scalar multiplication is **distributive** with respect to vector addition:

$$a(|\alpha\rangle + |\beta\rangle) = a|\alpha\rangle + a|\beta\rangle, \quad [A.7]$$

  and with respect to scalar addition:

$$(a + b)|\alpha\rangle = a|\alpha\rangle + b|\alpha\rangle. \quad [A.8]$$

  It is also **associative** with respect to the ordinary multiplication of scalars:

$$a(b|\alpha\rangle) = (ab)|\alpha\rangle. \quad [A.9]$$

  Multiplication by the scalars 0 and 1 has the effect you would expect:

$$0|\alpha\rangle = |0\rangle: \quad 1|\alpha\rangle = |\alpha\rangle. \quad [A.10]$$

  Evidently $|-\alpha\rangle = (-1)|\alpha\rangle$ (which we write more simply as $-|\alpha\rangle$).

There's a lot less here than meets the eye—all I have done is to write down in abstract language the familiar rules for manipulating vectors. The virtue of such abstraction is that we will be able to apply our knowledge and intuition about the behavior of ordinary vectors to other systems that happen to share the same formal properties.

---

[3] It is customary, where no confusion can arise, to write the null vector without the adorning bracket: $|0\rangle \rightarrow 0$.

[4] This is funny notation, since α is not a number. I'm simply adopting the name "−Charlie" for the inverse of the vector whose name is "Charlie." More natural terminology will suggest itself in a moment.

A **linear combination** of the vectors $|\alpha\rangle$, $|\beta\rangle$, $|\gamma\rangle$, ..., is an expression of the form

$$a|\alpha\rangle + b|\beta\rangle + c|\gamma\rangle + \cdots. \tag{A.11}$$

A vector $|\lambda\rangle$ is said to be **linearly independent** of the set $|\alpha\rangle$, $|\beta\rangle$, $|\gamma\rangle$, ..., if it cannot be written as a linear combination of them. (For example, in three dimensions the unit vector $\hat{k}$ is linearly independent of $\hat{\imath}$ and $\hat{\jmath}$, but any vector in the xy plane is linearly *dependent* on $\hat{\imath}$ and $\hat{\jmath}$.) By extension, a *set* of vectors is "linearly independent" if each one is linearly independent of all the rest. A collection of vectors is said to **span** the space if *every* vector can be written as a linear combination of the members of this set.[5] A set of *linearly independent* vectors that spans the space is called a **basis**. The number of vectors in any basis is called the **dimension** of the space. For the moment we shall assume that the dimension (n) is *finite*.

With respect to a prescribed basis

$$|e_1\rangle, \ |e_2\rangle, \ldots, |e_n\rangle, \tag{A.12}$$

any given vector

$$|\alpha\rangle = a_1|e_1\rangle + a_2|e_2\rangle + \cdots + a_n|e_n\rangle, \tag{A.13}$$

is uniquely represented by the (ordered) n-tuple of its **components**:

$$|\alpha\rangle \leftrightarrow (a_1, a_2, \ldots, a_n). \tag{A.14}$$

It is often easier to work with the components than with the abstract vectors themselves. To add vectors, you add their corresponding components:

$$|\alpha\rangle + |\beta\rangle \leftrightarrow (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n); \tag{A.15}$$

to multiply by a scalar you multiply each component:

$$c|\alpha\rangle \leftrightarrow (ca_1, ca_2, \ldots, ca_n); \tag{A.16}$$

the null vector is represented by a string of zeroes:

$$|0\rangle \leftrightarrow (0, 0, \ldots, 0); \tag{A.17}$$

and the components of the inverse vector have their signs reversed:

$$|-\alpha\rangle \leftrightarrow (-a_1, -a_2, \ldots, -a_n). \tag{A.18}$$

---

[5] A set of vectors that spans the space is also called **complete**, though I personally reserve that word for the infinite-dimensional case, where subtle questions of convergence may arise.

The only *dis*advantage of working with components is that you have to commit yourself to a particular basis, and the same manipulations will look very different to someone working in a different basis.

Problem A.1 Consider the ordinary vectors in 3 dimensions ($a_x\hat{\imath} + a_y\hat{\jmath} + a_z\hat{k}$), with complex components.

(a) Does the subset of all vectors with $a_z = 0$ constitute a vector space? If so, what is its dimension; if not, why not?

(b) What about the subset of all vectors whose z component is 1? *Hint:* Would the sum of two such vectors be in the subset? How about the null vector?

(c) What about the subset of vectors whose components are all equal?

***Problem A.2** Consider the collection of all polynomials (with complex coefficients) of degree less than N in x.

(a) Does this set constitute a vector space (with the polynomials as "vectors")? If so, suggest a convenient basis, and give the dimension of the space. If not, which of the defining properties does it lack?

(b) What if we require that the polynomials be *even* functions?

(c) What if we require that the leading coefficient (i.e., the number multiplying x^{N-1}) be 1?

(d) What if we require that the polynomials have the value 0 at $x = 1$?

(e) What if we require that the polynomials have the value 1 at $x = 0$?

Problem A.3 Prove that the components of a vector with respect to a given basis are *unique*.

A.2 INNER PRODUCTS

In three dimensions we encounter two kinds of vector products: the dot product and the cross product. The latter does not generalize in any natural way to n-dimensional vector spaces, but the former *does*—in this context it is usually called the **inner product**. The inner product of two vectors ($|\alpha\rangle$ and $|\beta\rangle$) is a complex number, which we write as $\langle\alpha|\beta\rangle$, with the following properties:

$$\langle \beta | \alpha \rangle = \langle \alpha | \beta \rangle^*, \tag{A.19}$$

$$\langle \alpha | \alpha \rangle \geq 0, \quad \text{and} \quad \langle \alpha | \alpha \rangle = 0 \Leftrightarrow |\alpha\rangle = |0\rangle, \tag{A.20}$$

$$\langle \alpha | (b|\beta\rangle + c|\gamma\rangle) = b\langle \alpha | \beta \rangle + c\langle \alpha | \gamma \rangle. \tag{A.21}$$

Apart from the generalization to complex numbers, these axioms simply codify the familiar behavior of dot products. A vector space with an inner product is called an **inner product space.**

Because the inner product of any vector with itself is a nonnegative number (Equation A.20), its square root is *real*—we call this the **norm** of the vector:

$$\|\alpha\| \equiv \sqrt{\langle \alpha | \alpha \rangle}; \tag{A.22}$$

it generalizes the notion of "length." A **unit vector** (one whose norm is 1) is said to be **normalized** (the word should really be "normal," but I guess that sounds too anthropomorphic). Two vectors whose inner product is zero are called **orthogonal** (generalizing the notion of "perpendicular"). A collection of mutually orthogonal normalized vectors,

$$\langle \alpha_i | \alpha_j \rangle = \delta_{ij}, \tag{A.23}$$

is called an **orthonormal set.** It is always possible (see Problem A.4), and almost always convenient, to choose an *orthonormal basis*; in that case the inner product of two vectors can be written very neatly in terms of their components:

$$\langle \alpha | \beta \rangle = a_1^* b_1 + a_2^* b_2 + \cdots + a_n^* b_n, \tag{A.24}$$

the norm (squared) becomes

$$\langle \alpha | \alpha \rangle = |a_1|^2 + |a_2|^2 + \cdots + |a_n|^2, \tag{A.25}$$

and the components themselves are

$$a_i = \langle e_i | \alpha \rangle. \tag{A.26}$$

(These results generalize the familiar formulas $\mathbf{a} \cdot \mathbf{b} = a_x b_x + a_y b_y + a_z b_z$, $|\mathbf{a}|^2 = a_x^2 + a_y^2 + a_z^2$, and $a_x = \hat{\imath} \cdot \mathbf{a}$, $a_y = \hat{\jmath} \cdot \mathbf{a}$, $a_z = \hat{k} \cdot \mathbf{a}$, for the three-dimensional orthonormal basis $\hat{\imath}$, $\hat{\jmath}$, $\hat{k}$.) From now on we shall *always* work in orthonormal bases, unless it is explicitly indicated otherwise.

Another geometrical quantity one might wish to generalize is the *angle* between two vectors. In ordinary vector analysis $\cos\theta = (\mathbf{a} \cdot \mathbf{b})/|\mathbf{a}||\mathbf{b}|$. But because the inner product is in general a complex number, the analogous formula (in an

arbitrary inner product space) does not define a (real) angle θ. Nevertheless, it is still true that the *absolute value* of this quantity is a number no greater than 1,

$$|\langle\alpha|\beta\rangle|^2 \leq \langle\alpha|\alpha\rangle\langle\beta|\beta\rangle. \qquad [A.27]$$

(This important result is known as the **Schwarz inequality**; the proof is given in Problem A.5.) So you can, if you like, define the angle between $|\alpha\rangle$ and $|\beta\rangle$ by the formula

$$\cos\theta = \sqrt{\frac{\langle\alpha|\beta\rangle\langle\beta|\alpha\rangle}{\langle\alpha|\alpha\rangle\langle\beta|\beta\rangle}}. \qquad [A.28]$$

---

*Problem A.4 Suppose you start out with a basis $(|e_1\rangle, |e_2\rangle, \ldots, |e_n\rangle)$ that is *not* orthonormal. The **Gram-Schmidt procedure** is a systematic ritual for generating from it an orthonormal basis $(|e_1'\rangle, |e_2'\rangle, \ldots, |e_n'\rangle)$. It goes like this:

(i) Normalize the first basis vector (divide by its norm):

$$|e_1'\rangle = \frac{|e_1\rangle}{\|e_1\|}.$$

(ii) Find the projection of the second vector along the first, and subtract it off:

$$|e_2\rangle - \langle e_1'|e_2\rangle|e_1'\rangle.$$

This vector is orthogonal to $|e_1'\rangle$; normalize it to get $|e_2'\rangle$.

(iii) Subtract from $|e_3\rangle$ its projections along $|e_1'\rangle$ and $|e_2'\rangle$:

$$|e_3\rangle - \langle e_1'|e_3\rangle|e_1'\rangle - \langle e_2'|e_3\rangle|e_2'\rangle.$$

This is orthogonal to $|e_1'\rangle$ and $|e_2'\rangle$; normalize it to get $|e_3'\rangle$. And so on.

Use the Gram-Schmidt procedure to orthonormalize the 3-space basis $|e_1\rangle = (1+i)\hat{\imath} + (1)\hat{\jmath} + (i)\hat{k}$. $|e_2\rangle = (i)\hat{\imath} + (3)\hat{\jmath} + (1)\hat{k}$. $|e_3\rangle = (0)\hat{\imath} + (28)\hat{\jmath} + (0)\hat{k}$.

Problem A.5 Prove the Schwarz inequality (Equation A.27). *Hint:* Let $|\gamma\rangle = |\beta\rangle - (\langle\alpha|\beta\rangle/\langle\alpha|\alpha\rangle)|\alpha\rangle$, and use $\langle\gamma|\gamma\rangle \geq 0$.

Problem A.6 Find the angle (in the sense of Equation A.28) between the vectors $|\alpha\rangle = (1+i)\hat{\imath} + (1)\hat{\jmath} + (i)\hat{k}$ and $|\beta\rangle = (4-i)\hat{\imath} + (0)\hat{\jmath} + (2-2i)\hat{k}$.

Problem A.7 Prove the **triangle inequality**: $\|(|\alpha\rangle + |\beta\rangle)\| \leq \|\alpha\| + \|\beta\|$.

A.3 MATRICES

Suppose you take every vector (in 3-space) and multiply it by 17, or you rotate every vector by 39° about the z-axis, or you reflect every vector in the xy plane—these are all examples of **linear transformations**. A linear transformation[6] ($\hat{T}$) takes each vector in a vector space and "transforms" it into some other vector ($|\alpha\rangle \rightarrow |\alpha'\rangle = \hat{T}|\alpha\rangle$), subject to the condition that the operation be *linear*:

$$\hat{T}(a|\alpha\rangle + b|\beta\rangle) = a(\hat{T}|\alpha\rangle) + b(\hat{T}|\beta\rangle), \qquad [A.29]$$

for any vectors $|\alpha\rangle$, $|\beta\rangle$ and any scalars a, b.

If you know what a particular linear transformation does to a set of *basis* vectors, you can easily figure out what it does to *any* vector. For suppose that

$$\begin{aligned}
\hat{T}|e_1\rangle &= T_{11}|e_1\rangle + T_{21}|e_2\rangle + \cdots + T_{n1}|e_n\rangle, \\
\hat{T}|e_2\rangle &= T_{12}|e_1\rangle + T_{22}|e_2\rangle + \cdots + T_{n2}|e_n\rangle, \\
&\cdots \\
\hat{T}|e_n\rangle &= T_{1n}|e_1\rangle + T_{2n}|e_2\rangle + \cdots + T_{nn}|e_n\rangle,
\end{aligned}$$

or, more compactly,

$$\hat{T}|e_j\rangle = \sum_{i=1}^{n} T_{ij}|e_i\rangle, \quad (j = 1, 2, \ldots, n). \qquad [A.30]$$

If $|\alpha\rangle$ is an arbitrary vector,

$$|\alpha\rangle = a_1|e_1\rangle + a_2|e_2\rangle + \cdots + a_n|e_n\rangle = \sum_{j=1}^{n} a_j|e_j\rangle, \qquad [A.31]$$

then

$$\hat{T}|\alpha\rangle = \sum_{j=1}^{n} a_j\left(\hat{T}|e_j\rangle\right) = \sum_{j=1}^{n}\sum_{i=1}^{n} a_j T_{ij}|e_i\rangle = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} T_{ij}a_j\right)|e_i\rangle. \qquad [A.32]$$

Evidently $\hat{T}$ takes a vector with components $a_1, a_2, \ldots, a_n$ into a vector with components[7]

$$a'_i = \sum_{j=1}^{n} T_{ij}a_j. \qquad [A.33]$$

---

[6]In this chapter I'll use a hat (^) to denote linear transformations: this is not inconsistent with my convention in the text (putting hats on operators), for (as we shall see) quantum operators *are* linear transformations.

[7]Notice the reversal of indices between Equations A.30 and A.33. This is not a typographical error. Another way of putting it (switching $i \leftrightarrow j$ in Equation A.30) is that if the *components* transform with T_{ij}, the *basis* vectors transform with T_{ji}.

Thus the n^2 **elements** T_{ij} uniquely characterize the linear transformation $\hat{T}$ (with respect to a given basis), just as the n components a_i uniquely characterize the vector $|\alpha\rangle$ (with respect to the same basis):

$$\hat{T} \leftrightarrow (T_{11}, T_{12}, \ldots, T_{nn}). \qquad [A.34]$$

If the basis is orthonormal, it follows from Equation A.30 that

$$T_{ij} = \langle e_i | \hat{T} | e_j \rangle. \qquad [A.35]$$

It is convenient to display these complex numbers in the form of a **matrix**:[8]

$$\mathsf{T} = \begin{pmatrix} T_{11} & T_{12} & \cdots & T_{1n} \\ T_{21} & T_{22} & \cdots & T_{2n} \\ \vdots & \vdots & & \vdots \\ T_{n1} & T_{n2} & \cdots & T_{nn} \end{pmatrix}. \qquad [A.36]$$

The study of linear transformations reduces then to the theory of matrices. The *sum* of two linear transformations $(\hat{S} + \hat{T})$ is defined in the natural way:

$$(\hat{S} + \hat{T})|\alpha\rangle = \hat{S}|\alpha\rangle + \hat{T}|\alpha\rangle; \qquad [A.37]$$

this matches the usual rule for adding matrices (you add their corresponding elements):

$$\mathsf{U} = \mathsf{S} + \mathsf{T} \Leftrightarrow U_{ij} = S_{ij} + T_{ij}. \qquad [A.38]$$

The *product* of two linear transformations $(\hat{S}\hat{T})$ is the net effect of performing them in succession—first $\hat{T}$, then $\hat{S}$:

$$|\alpha'\rangle = \hat{T}|\alpha\rangle; \quad |\alpha''\rangle = \hat{S}|\alpha'\rangle = \hat{S}(\hat{T}|\alpha\rangle) = \hat{S}\hat{T}|\alpha\rangle. \qquad [A.39]$$

What matrix U represents the combined transformation $\hat{U} = \hat{S}\hat{T}$? It's not hard to work it out:

$$a_i'' = \sum_{j=1}^{n} S_{ij} a_j' = \sum_{j=1}^{n} S_{ij} \left(\sum_{k=1}^{n} T_{jk} a_k \right) = \sum_{k=1}^{n} \left(\sum_{j=1}^{n} S_{ij} T_{jk} \right) a_k = \sum_{k=1}^{n} U_{ik} a_k.$$

Evidently

$$\mathsf{U} = \mathsf{ST} \Leftrightarrow U_{ik} = \sum_{j=1}^{n} S_{ij} T_{jk}. \qquad [A.40]$$

[8] I'll use boldface capital letters, sans serif, to denote square matrices.

This is the standard rule for matrix multiplication—to find the ikth element of the product $\mathbf{ST}$, you look at the ith row of $\mathbf{S}$, and the kth column of $\mathbf{T}$, multiply corresponding entries, and add. The same prescription allows you to multiply *rectangular* matrices, as long as the number of columns in the first matches the number of rows in the second. In particular, if we write the n-tuple of components of $|\alpha\rangle$ as an $n \times 1$ **column matrix** (or "column vector"):[9]

$$\mathbf{a} \equiv \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}. \tag{A.41}$$

the transformation rule (Equation A.33) can be expressed as a matrix product:

$$\mathbf{a}' = \mathbf{T}\mathbf{a}. \tag{A.42}$$

Now some matrix terminology:

- The **transpose** of a matrix (which we shall write with a tilde: $\tilde{\mathbf{T}}$) is the same set of elements, but with rows and columns interchanged. In particular, the transpose of a *column* matrix is a **row matrix**:

$$\tilde{\mathbf{a}} = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}. \tag{A.43}$$

  For a *square* matrix taking the transpose amounts to reflecting in the **main diagonal** (upper left to lower right):

$$\tilde{\mathbf{T}} = \begin{pmatrix} T_{11} & T_{21} & \cdots & T_{n1} \\ T_{12} & T_{22} & \cdots & T_{n2} \\ \vdots & \vdots & & \vdots \\ T_{1n} & T_{2n} & \cdots & T_{nn} \end{pmatrix}. \tag{A.44}$$

  A (square) matrix is **symmetric** if it is equal to its transpose; it is **antisymmetric** if this operation reverses the sign:

$$\text{symmetric}: \tilde{\mathbf{T}} = \mathbf{T}; \quad \text{antisymmetric}: \tilde{\mathbf{T}} = -\mathbf{T}. \tag{A.45}$$

- The (complex) **conjugate** of a matrix (which we denote, as usual, with an asterisk, $\mathbf{T}^*$), consists of the complex conjugate of every element:

$$\mathbf{T}^* = \begin{pmatrix} T_{11}^* & T_{12}^* & \cdots & T_{1n}^* \\ T_{21}^* & T_{22}^* & \cdots & T_{2n}^* \\ \vdots & \vdots & & \vdots \\ T_{n1}^* & T_{n2}^* & \cdots & T_{nn}^* \end{pmatrix}; \quad \mathbf{a}^* = \begin{pmatrix} a_1^* \\ a_2^* \\ \vdots \\ a_n^* \end{pmatrix}. \tag{A.46}$$

[9] I'll use boldface lowercase letters, sans serif, for row and column matrices.

A matrix is **real** if all its elements are real, and **imaginary** if they are all imaginary:

$$\text{real}: \mathbf{T}^* = \mathbf{T}; \quad \text{imaginary}: \mathbf{T}^* = -\mathbf{T}. \tag{A.47}$$

- The **hermitian conjugate** (or **adjoint**) of a matrix (indicated by a dagger, $\mathbf{T}^\dagger$) is the transpose conjugate:

$$\mathbf{T}^\dagger \equiv \tilde{\mathbf{T}}^* = \begin{pmatrix} T_{11}^* & T_{21}^* & \cdots & T_{n1}^* \\ T_{12}^* & T_{22}^* & \cdots & T_{n2}^* \\ \vdots & \vdots & & \vdots \\ T_{1n}^* & T_{2n}^* & \cdots & T_{nn}^* \end{pmatrix}: \quad \mathbf{a}^\dagger \equiv \tilde{\mathbf{a}}^* = \begin{pmatrix} a_1^* & a_2^* & \cdots & a_n^* \end{pmatrix}. \tag{A.48}$$

A square matrix is **hermitian** (or **self-adjoint**) if it is equal to its hermitian conjugate; if hermitian conjugation introduces a minus sign, the matrix is **skew hermitian** (or **anti-hermitian**):

$$\text{hermitian}: \mathbf{T}^\dagger = \mathbf{T}; \quad \text{skew hermitian}: \mathbf{T}^\dagger = -\mathbf{T}. \tag{A.49}$$

In this notation the inner product of two vectors (with respect to an orthonormal basis—Equation A.24), can be written very neatly as a matrix product:

$$\langle \alpha | \beta \rangle = \mathbf{a}^\dagger \mathbf{b}. \tag{A.50}$$

Notice that each of the three operations defined in this paragraph, if applied twice, returns you to the original matrix.

Matrix multiplication is not, in general, commutative ($\mathbf{ST} \neq \mathbf{TS}$); the *difference* between the two orderings is called the **commutator**:[10]

$$[\mathbf{S}, \mathbf{T}] \equiv \mathbf{ST} - \mathbf{TS}. \tag{A.51}$$

The transpose of a product is the product of the transposes *in reverse order*:

$$\widetilde{(\mathbf{ST})} = \tilde{\mathbf{T}}\tilde{\mathbf{S}}, \tag{A.52}$$

(see Problem A.11), and the same goes for hermitian conjugates:

$$(\mathbf{ST})^\dagger = \mathbf{T}^\dagger \mathbf{S}^\dagger. \tag{A.53}$$

The **unit matrix** (representing a linear transformation that carries every vector into itself) consists of ones on the main diagonal, and zeroes everywhere else:

$$\mathbf{I} \equiv \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{A.54}$$

[10] The commutator only makes sense for *square* matrices, of course; for rectangular matrices the two orderings wouldn't even be the same size.

In other words,

$$\mathsf{I}_{ij} = \delta_{ij}. \tag{A.55}$$

The **inverse** of a (square) matrix (written T^{-1}) is defined in the obvious way:[11]

$$\mathsf{T}^{-1}\mathsf{T} = \mathsf{T}\mathsf{T}^{-1} = \mathsf{I}. \tag{A.56}$$

A matrix has an inverse if and only if its **determinant**[12] is nonzero; in fact,

$$\mathsf{T}^{-1} = \frac{1}{\det \mathsf{T}}\tilde{\mathsf{C}}, \tag{A.57}$$

where C is the matrix of **cofactors** (the cofactor of element T_{ij} is $(-1)^{i+j}$ times the determinant of the submatrix obtained from T by erasing the ith row and the jth column). A matrix that has no inverse is said to be **singular**. The inverse of a product (assuming it exists) is the product of the inverses *in reverse order*:

$$(\mathsf{ST})^{-1} = \mathsf{T}^{-1}\mathsf{S}^{-1}. \tag{A.58}$$

A matrix is **unitary** if its inverse is equal to its hermitian conjugate:[13]

$$\text{unitary}: \mathsf{U}^{\dagger} = \mathsf{U}^{-1}. \tag{A.59}$$

Assuming the basis is orthonormal, the columns of a unitary matrix constitute an orthonormal set, and so too do its rows (see Problem A.12). Linear transformations represented by unitary matrices preserve inner products, since (Equation A.50)

$$\langle \alpha' | \beta' \rangle = \mathbf{a}'^{\dagger}\mathbf{b}' = (\mathsf{U}\mathbf{a})^{\dagger}(\mathsf{U}\mathbf{b}) = \mathbf{a}^{\dagger}\mathsf{U}^{\dagger}\mathsf{U}\mathbf{b} = \mathbf{a}^{\dagger}\mathbf{b} = \langle \alpha | \beta \rangle. \tag{A.60}$$

∗**Problem A.8** Given the following two matrices:

$$\mathsf{A} = \begin{pmatrix} -1 & 1 & i \\ 2 & 0 & 3 \\ 2i & -2i & 2 \end{pmatrix}, \quad \mathsf{B} = \begin{pmatrix} 2 & 0 & -i \\ 0 & 1 & 0 \\ i & 3 & 2 \end{pmatrix}.$$

---

[11] Note that the left inverse is equal to the right inverse, for if $\mathsf{AT} = \mathsf{I}$ and $\mathsf{TB} = \mathsf{I}$, then (multiplying the second on the left by A and invoking the first) we get $\mathsf{B} = \mathsf{A}$.

[12] I assume you know how to evaluate determinants. If not, see M. Boas, *Mathematical Methods in the Physical Sciences*, 2nd ed. (John Wiley, New York, 1983), Section 3.3.

[13] In a *real* vector space (that is, one in which the scalars are real) the hermitian conjugate is the same as the transpose, and a unitary matrix is **orthogonal**: $\tilde{\mathsf{O}} = \mathsf{O}^{-1}$. For example, rotations in ordinary 3-space are represented by orthogonal matrices.

compute: (a) $\mathbf{A}+\mathbf{B}$, (b) $\mathbf{AB}$, (c) $[\mathbf{A},\mathbf{B}]$, (d) $\tilde{\mathbf{A}}$, (e) $\mathbf{A}^*$, (f) $\mathbf{A}^\dagger$, (g) $\det(\mathbf{B})$, and (h) $\mathbf{B}^{-1}$. Check that $\mathbf{B}\mathbf{B}^{-1}=\mathbf{I}$. Does $\mathbf{A}$ have an inverse?

---

*Problem A.9 Using the square matrices in Problem A.8, and the column matrices

$$\mathbf{a}=\begin{pmatrix} i \\ 2i \\ 2 \end{pmatrix}, \quad \mathbf{b}=\begin{pmatrix} 2 \\ (1-i) \\ 0 \end{pmatrix},$$

find: (a) $\mathbf{Aa}$, (b) $\mathbf{a}^\dagger\mathbf{b}$, (c) $\tilde{\mathbf{a}}\mathbf{B}\mathbf{b}$, (d) $\mathbf{a}\mathbf{b}^\dagger$.

Problem A.10 By explicit construction of the matrices in question, show that any matrix $\mathbf{T}$ can be written

(a) as the sum of a symmetric matrix $\mathbf{S}$ and an antisymmetric matrix $\mathbf{A}$;

(b) as the sum of a real matrix $\mathbf{R}$ and an imaginary matrix $\mathbf{M}$;

(c) as the sum of a hermitian matrix $\mathbf{H}$ and a skew-hermitian matrix $\mathbf{K}$.

---

*Problem A.11 Prove Equations A.52, A.53, and A.58. Show that the product of two unitary matrices is unitary. Under what conditions is the product of two hermitian matrices hermitian? Is the sum of two unitary matrices necessarily unitary? Is the sum of two hermitian matrices hermitian?

Problem A.12 Show that the rows and columns of a unitary matrix constitute orthonormal sets.

Problem A.13 Noting that $\det(\tilde{\mathbf{T}})=\det(\mathbf{T})$, show that the determinant of a hermitian matrix is real, the determinant of a unitary matrix has modulus 1 (hence the name), and the determinant of an orthogonal matrix is either $+1$ or -1.

A.4 CHANGING BASES

The components of a vector depend, of course, on your (arbitrary) choice of basis, and so do the elements of the matrix representing a linear transformation. We might inquire how these numbers change when we switch to a different basis.

The old basis vectors, $|e_i\rangle$ are—like *all* vectors—linear combinations of the new ones, $|f_i\rangle$:

$$\begin{aligned} |e_1\rangle &= S_{11}|f_1\rangle + S_{21}|f_2\rangle + \cdots + S_{n1}|f_n\rangle, \\ |e_2\rangle &= S_{12}|f_1\rangle + S_{22}|f_2\rangle + \cdots + S_{n2}|f_n\rangle, \\ &\cdots \\ |e_n\rangle &= S_{1n}|f_1\rangle + S_{2n}|f_2\rangle + \cdots + S_{nn}|f_n\rangle, \end{aligned}$$

(for some set of complex numbers S_{ij}), or, more compactly,

$$|e_j\rangle = \sum_{i=1}^{n} S_{ij}|f_i\rangle, \quad (j = 1, 2, \ldots, n). \tag{A.61}$$

This is *itself* a linear transformation (compare Equation A.30),[14] and we know immediately how the components transform:

$$a_i^f = \sum_{j=1}^{n} S_{ij} a_j^e. \tag{A.62}$$

(where the superscript indicates the basis). In matrix form

$$\mathbf{a}^f = \mathbf{S}\mathbf{a}^e. \tag{A.63}$$

What about the matrix representing a linear transformation $\hat{T}$—how is *it* modified by a change of basis? Well, in the old basis we had (Equation A.42)

$$\mathbf{a}'^e = \mathbf{T}^e \mathbf{a}^e,$$

and Equation A.63—multiplying both sides by $\mathbf{S}^{-1}$—entails[15] $\mathbf{a}^e = \mathbf{S}^{-1}\mathbf{a}^f$, so

$$\mathbf{a}'^f = \mathbf{S}\mathbf{a}'^e = \mathbf{S}(\mathbf{T}^e \mathbf{a}^e) = \mathbf{S}\mathbf{T}^e \mathbf{S}^{-1}\mathbf{a}^f.$$

Evidently

$$\mathbf{T}^f = \mathbf{S}\mathbf{T}^e\mathbf{S}^{-1}. \tag{A.64}$$

In general, two matrices ($\mathbf{T}_1$ and $\mathbf{T}_2$) are said to be **similar** if $\mathbf{T}_2 = \mathbf{S}\mathbf{T}_1\mathbf{S}^{-1}$ for some (nonsingular) matrix $\mathbf{S}$. What we have just found is that *matrices representing*

---

[14]Notice, however, the radically different perspective: In this case we're talking about one and the same *vector*, referred to two completely different *bases*, whereas before we were thinking of a completely *different* vector, referred to the *same* basis.

[15]Note that $\mathbf{S}^{-1}$ certainly exists—if $\mathbf{S}$ were singular, the $|f_i\rangle$'s would not span the space, so they wouldn't constitute a basis.

the same linear transformation, with respect to different bases, are similar. Incidentally, if the first basis is orthonormal, the second will also be orthonormal if and only if the matrix $\mathbf{S}$ is *unitary* (see Problem A.16). Since we always work in orthonormal bases, we are interested mainly in *unitary* similarity transformations.

While the *elements* of the matrix representing a given linear transformation may look very different in the new basis, two numbers associated with the matrix are unchanged: the determinant and the **trace**. For the determinant of a product is the product of the determinants, and hence

$$\det(\mathbf{T}^f) = \det(\mathbf{S}\mathbf{T}^e\mathbf{S}^{-1}) = \det(\mathbf{S})\det(\mathbf{T}^e)\det(\mathbf{S}^{-1}) = \det\mathbf{T}^e. \qquad [A.65]$$

And the trace, which is the *sum of the diagonal elements*,

$$\mathrm{Tr}(\mathbf{T}) \equiv \sum_{i=1}^{m} T_{ii}, \qquad [A.66]$$

has the property (see Problem A.17) that

$$\mathrm{Tr}(\mathbf{T}_1\mathbf{T}_2) = \mathrm{Tr}(\mathbf{T}_2\mathbf{T}_1), \qquad [A.67]$$

(for any two matrices $\mathbf{T}_1$ and $\mathbf{T}_2$), so

$$\mathrm{Tr}(\mathbf{T}^f) = \mathrm{Tr}(\mathbf{S}\mathbf{T}^e\mathbf{S}^{-1}) = \mathrm{Tr}(\mathbf{T}_e\mathbf{S}^{-1}\mathbf{S}) = \mathrm{Tr}(\mathbf{T}^e). \qquad [A.68]$$

Problem A.14 Using the standard basis $(\hat{\imath}, \hat{\jmath}, \hat{k})$ for vectors in three dimensions:

(a) Construct the matrix representing a rotation through angle θ (counterclockwise, looking down the axis toward the origin) about the z-axis.

(b) Construct the matrix representing a rotation by 120° (counterclockwise, looking down the axis) about an axis through the point (1,1,1).

(c) Construct the matrix representing reflection through the xy-plane.

(d) Check that all these matrices are orthogonal, and calculate their determinants.

Problem A.15 In the usual basis $(\hat{\imath}, \hat{\jmath}, \hat{k})$, construct the matrix $\mathbf{T}_x$ representing a rotation through angle θ about the x-axis, and the matrix $\mathbf{T}_y$ representing a rotation through angle θ about the y-axis. Suppose now we change bases, to $\hat{\imath}' = \hat{\jmath}$, $\hat{\jmath}' = -\hat{\imath}$, $\hat{k}' = \hat{k}$. Construct the matrix $\mathbf{S}$ that effects this change of basis, and check that $\mathbf{S}\mathbf{T}_x\mathbf{S}^{-1}$ and $\mathbf{S}\mathbf{T}_y\mathbf{S}^{-1}$ are what you would expect.

Problem A.16 Show that similarity preserves matrix multiplication (that is, if $\mathbf{A}^e\mathbf{B}^e = \mathbf{C}^e$, then $\mathbf{A}^f\mathbf{B}^f = \mathbf{C}^f$). Similarity does *not*, in general, preserve symmetry,

reality, or hermiticity; show, however, that if $\mathbf{S}$ is *unitary*, and $\mathbf{H}^e$ is hermitian, then $\mathbf{H}^f$ is hermitian. Show that $\mathbf{S}$ carries an orthonormal basis into another orthonormal basis if and only if it is unitary.

∗**Problem A.17** Prove that $\text{Tr}(\mathbf{T}_1\mathbf{T}_2) = \text{Tr}(\mathbf{T}_2\mathbf{T}_1)$. It follows immediately that $\text{Tr}(\mathbf{T}_1\mathbf{T}_2\mathbf{T}_3) = \text{Tr}(\mathbf{T}_2\mathbf{T}_3\mathbf{T}_1)$, but is it the case that $\text{Tr}(\mathbf{T}_1\mathbf{T}_2\mathbf{T}_3) = \text{Tr}(\mathbf{T}_2\mathbf{T}_1\mathbf{T}_3)$, in general? Prove it, or disprove it. *Hint:* The best disproof is always a counterexample—the simpler the better!

A.5 EIGENVECTORS AND EIGENVALUES

Consider the linear transformation in three-space consisting of a rotation, about some specified axis, by an angle θ. Most vectors will change in a rather complicated way (they ride around on a cone about the axis), but vectors that happen to lie *along* the axis have very simple behavior: They don't change at all ($\hat{T}|\alpha\rangle = |\alpha\rangle$). If θ is 180°, then vectors which lie in the "equatorial" plane reverse signs ($\hat{T}|\alpha\rangle = -|\alpha\rangle$). In a complex vector space[16] *every* linear transformation has "special" vectors like these, which are transformed into scalar multiples of themselves:

$$\hat{T}|\alpha\rangle = \lambda|\alpha\rangle; \qquad [A.69]$$

they are called **eigenvectors** of the transformation, and the (complex) number λ is their **eigenvalue**. (The *null* vector doesn't count, even though in a trivial sense it obeys Equation A.69 for *any* $\hat{T}$ and *any* λ; technically, an eigenvector is any *nonzero* vector satisfying Equation A.69.) Notice that any (nonzero) *multiple* of an eigenvector is still an eigenvector, with the same eigenvalue.

With respect to a particular basis, the eigenvector equation assumes the matrix form

$$\mathbf{Ta} = \lambda\mathbf{a}, \qquad [A.70]$$

(for nonzero $\mathbf{a}$), or

$$(\mathbf{T} - \lambda\mathbf{I})\mathbf{a} = \mathbf{0}. \qquad [A.71]$$

(Here $\mathbf{0}$ is the **zero matrix**, whose elements are all zero.) Now, if the matrix $(\mathbf{T} - \lambda\mathbf{I})$ had an *inverse*, we could multiply both sides of Equation A.71 by $(\mathbf{T} - \lambda\mathbf{I})^{-1}$, and

---

[16] This is *not* always true in a *real* vector space (where the scalars are restricted to real values). See Problem A.18.

conclude that $\mathbf{a} = \mathbf{0}$. But by assumption $\mathbf{a}$ is *not* zero, so the matrix $(\mathbf{T} - \lambda\mathbf{I})$ must in fact be singular, which means that its determinant is zero:

$$\det(\mathbf{T} - \lambda\mathbf{I}) = \begin{vmatrix} (T_{11} - \lambda) & T_{12} & \cdots & T_{1n} \\ T_{21} & (T_{22} - \lambda) & \cdots & T_{2n} \\ \vdots & \vdots & & \vdots \\ T_{n1} & T_{n2} & \cdots & (T_{nn} - \lambda) \end{vmatrix} = 0. \qquad [A.72]$$

Expansion of the determinant yields an algebraic equation for λ:

$$C_n\lambda^n + C_{n-1}\lambda^{n-1} + \cdots + C_1\lambda + C_0 = 0. \qquad [A.73]$$

where the coefficients C_i depend on the elements of $\mathbf{T}$ (see Problem A.20). This is called the **characteristic equation** for the matrix; its solutions determine the eigenvalues. Notice that it's an nth-order equation, so (by the **fundamental theorem of algebra**) it has n (complex) roots.[17] However, some of these may be multiple roots, so all we can say for certain is that an $n \times n$ matrix has *at least one* and *at most* n distinct eigenvalues. The collection of all the eigenvalues of a matrix is called its **spectrum**; if two (or more) linearly independent eigenvectors share the same eigenvalue, the spectrum is said to be **degenerate**.

To construct the eigen*vectors* it is generally easiest simply to plug each λ back into Equation A.70 and solve "by hand" for the components of $\mathbf{a}$. I'll show you how it goes by working out an example.

Example A.1 Find the eigenvalues and eigenvectors of the following matrix:

$$\mathbf{M} = \begin{pmatrix} 2 & 0 & -2 \\ -2i & i & 2i \\ 1 & 0 & -1 \end{pmatrix}. \qquad [A.74]$$

Solution: The characteristic equation is

$$\begin{vmatrix} (2 - \lambda) & 0 & -2 \\ -2i & (i - \lambda) & 2i \\ 1 & 0 & (-1 - \lambda) \end{vmatrix} = -\lambda^3 + (1 + i)\lambda^2 - i\lambda = 0, \qquad [A.75]$$

and its roots are 0, 1, and i. Call the components of the first eigenvector (a_1, a_2, a_3); then

$$\begin{pmatrix} 2 & 0 & -2 \\ -2i & i & 2i \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = 0 \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

---

[17] It is here that the case of *real* vector spaces becomes more awkward, because the characteristic equation need not have any (real) solutions at all. See Problem A.18.

which yields three equations:

$$\begin{aligned} 2a_1 - 2a_3 &= 0, \\ -2ia_1 + ia_2 + 2ia_3 &= 0, \\ a_1 - a_3 &= 0. \end{aligned}$$

The first determines a_3 (in terms of a_1): $a_3 = a_1$; the second determines a_2: $a_2 = 0$; and the third is redundant. We may as well pick $a_1 = 1$ (since any multiple of an eigenvector is still an eigenvector):

$$\mathbf{a}^{(1)} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \text{for } \lambda_1 = 0. \qquad [A.76]$$

For the second eigenvector (recycling the same notation for the components) we have

$$\begin{pmatrix} 2 & 0 & -2 \\ -2i & i & 2i \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = 1 \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix},$$

which leads to the equations

$$\begin{aligned} 2a_1 - 2a_3 &= a_1, \\ -2ia_1 + ia_2 + 2ia_3 &= a_2, \\ a_1 - a_3 &= a_3. \end{aligned}$$

with the solution $a_3 = (1/2)a_1$, $a_2 = [(1 - i)/2]a_1$; this time I'll pick $a_1 = 2$, so

$$\mathbf{a}^{(2)} = \begin{pmatrix} 2 \\ 1 - i \\ 1 \end{pmatrix}, \quad \text{for } \lambda_2 = 1. \qquad [A.77]$$

Finally, for the third eigenvector,

$$\begin{pmatrix} 2 & 0 & -2 \\ -2i & i & 2i \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = i \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} ia_1 \\ ia_2 \\ ia_3 \end{pmatrix},$$

which gives the equations

$$\begin{aligned} 2a_1 - 2a_3 &= ia_1, \\ -2ia_1 + ia_2 + 2ia_3 &= ia_2, \\ a_1 - a_3 &= ia_3. \end{aligned}$$

whose solution is $a_3 = a_1 = 0$, with a_2 undetermined. Choosing $a_2 = 1$, we conclude

$$\mathbf{a}^{(3)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \text{for } \lambda_3 = i. \tag{A.78}$$

If the eigenvectors span the space (as they do in the preceding example), we are free to use *them* as a basis:

$$\begin{aligned} \hat{T}|f_1\rangle &= \lambda_1|f_1\rangle, \\ \hat{T}|f_2\rangle &= \lambda_2|f_2\rangle, \\ &\cdots \\ \hat{T}|f_n\rangle &= \lambda_n|f_n\rangle. \end{aligned}$$

In this basis the matrix representing $\hat{T}$ takes on a very simple form, with the eigenvalues strung out along the main diagonal, and all other elements zero:

$$\mathbf{T} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}, \tag{A.79}$$

and the (normalized) eigenvectors are

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \tag{A.80}$$

A matrix that can be brought to **diagonal form** (Equation A.79) by a change of basis is said to be **diagonalizable** (evidently a matrix is diagonalizable if and only if its eigenvectors span the space). The similarity matrix that effects the diagonalization can be constructed by using the normalized eigenvectors (in the old basis) as the columns of $\mathbf{S}^{-1}$:

$$(\mathbf{S}^{-1})_{ij} = (\mathbf{a}^{(j)})_i. \tag{A.81}$$

Example A.2 In Example A.1,

$$\mathbf{S}^{-1} = \begin{pmatrix} 1 & 2 & 0 \\ 0 & (1-i) & 1 \\ 1 & 1 & 0 \end{pmatrix},$$

so (using Equation A.57)

$$\mathbf{S} = \begin{pmatrix} -1 & 0 & 2 \\ 1 & 0 & -1 \\ (i-1) & 1 & (1-i) \end{pmatrix},$$

and you can check for yourself that

$$\mathbf{Sa}^{(1)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{Sa}^{(2)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{Sa}^{(3)} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

and

$$\mathbf{SMS}^{-1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & i \end{pmatrix}.$$

There's an obvious advantage in bringing a matrix to diagonal form: It's much easier to work with. Unfortunately, not every matrix *can* be diagonalized—the eigenvectors have to span the space. If the characteristic equation has n distinct roots, then the matrix is certainly diagonalizable, but it *may* be diagonalizable even if there are multiple roots. (For an example of a matrix that *cannot* be diagonalized, see Problem A.19.) It would be handy to know in advance (before working out all the eigenvectors) whether a given matrix is diagonalizable. A useful sufficient (though not necessary) condition is the following: A matrix is said to be **normal** if it commutes with its hermitian conjugate:

$$\text{normal}: \quad [\mathbf{N}^\dagger, \mathbf{N}] = \mathbf{0}. \qquad [A.82]$$

Every normal matrix is diagonalizable (its eigenvectors span the space). In particular, every hermitian matrix, and every unitary matrix, is diagonalizable.

Suppose we have *two* diagonalizable matrices; in quantum applications the question often arises: Can they be **simultaneously diagonalized** (by the *same* similarity matrix **S**)? That is to say, does there exist a basis in which they are *both* diagonal? The answer is yes *if and only if the two matrices commute* (see Problem A.22).

∗**Problem A.18** The 2×2 matrix representing a rotation of the xy plane is

$$\mathbf{T} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \qquad [A.83]$$

Show that (except for certain special angles—what are they?) this matrix has no real eigenvalues. (This reflects the geometrical fact that no vector in the plane

is carried into itself under such a rotation; contrast rotations in *three* dimensions.) This matrix *does*, however, have *complex* eigenvalues and eigenvectors. Find them. Construct a matrix **S** that diagonalizes **T**. Perform the similarity transformation ($\mathbf{STS}^{-1}$) explicitly, and show that it reduces **T** to diagonal form.

Problem A.19 Find the eigenvalues and eigenvectors of the following matrix:

$$\mathbf{M} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Can this matrix be diagonalized?

Problem A.20 Show that the first, second, and last coefficients in the characteristic equation (Equation A.73) are:

$$C_n = (-1)^n, \quad C_{n-1} = (-1)^{n-1}\mathrm{Tr}(\mathbf{T}), \quad \text{and} \quad C_0 = \det(\mathbf{T}). \qquad [A.84]$$

For a 3×3 matrix with elements T_{ij}, what is C_1?

Problem A.21 It's obvious that the trace of a *diagonal* matrix is the sum of its eigenvalues, and its determinant is their product (just look at Equation A.79). It follows (from Equations A.65 and A.68) that the same holds for any *diagonalizable* matrix. Prove that in fact

$$\det(\mathbf{T}) = \lambda_1\lambda_2\cdots\lambda_n, \quad \mathrm{Tr}(\mathbf{T}) = \lambda_1 + \lambda_2 + \cdots + \lambda_n. \qquad [A.85]$$

for *any* matrix. (The λ's are the n solutions to the characteristic equation—in the case of multiple roots, there may be fewer linearly independent eigen*vectors* than there are solutions, but we still count each λ as many times as it occurs.) *Hint:* Write the characteristic equation in the form

$$(\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda) = 0.$$

and use the result of Problem A.20.

Problem A.22

(a) Show that if two matrices commute in *one* basis, then they commute in any basis. That is:

$$[\mathbf{T}_1^e, \mathbf{T}_2^e] = \mathbf{0} \Rightarrow [\mathbf{T}_1^f, \mathbf{T}_2^f] = \mathbf{0}. \qquad [A.86]$$

Hint: Use Equation A.64.

(b) Show that if two matrices are simultaneously diagonalizable, they commute.[18]

Problem A.23 Consider the matrix

$$\mathbf{M} = \begin{pmatrix} 1 & 1 \\ 1 & i \end{pmatrix}.$$

(a) Is it normal?

(b) Is it diagonalizable?

A.6 HERMITIAN TRANSFORMATIONS

In Equation A.48 I defined the hermitian conjugate (or "adjoint") of a *matrix* as its transpose-conjugate: $\mathbf{T}^\dagger = \tilde{\mathbf{T}}^*$. Now I want to give you a more fundamental definition for the hermitian conjugate of a *linear transformation*: It is that transformation $\hat{T}^\dagger$ which, when applied to the *first* member of an inner product, gives the same result as if $\hat{T}$ itself had been applied to the *second* vector:

$$\langle \hat{T}^\dagger \alpha | \beta \rangle = \langle \alpha | \hat{T} \beta \rangle, \qquad [A.87]$$

(for all vectors $|\alpha\rangle$ and $|\beta\rangle$).[19] I have to warn you that although everybody uses it, this is lousy notation. For α and β are not *vectors* (the *vectors* are $|\alpha\rangle$ and $|\beta\rangle$), they are *names*. In particular, they are endowed with no mathematical properties at all, and the expression "$\hat{T}\beta$" is literally *nonsense*: Linear transformations act on *vectors*, not *labels*. But it's pretty clear what the notation *means*: $\hat{T}\beta$ is the name of the vector $\hat{T}|\beta\rangle$, and $\langle \hat{T}^\dagger \alpha | \beta \rangle$ is the inner product of the vector $\hat{T}^\dagger |\alpha\rangle$ with the vector $|\beta\rangle$. Notice in particular that

$$\langle \alpha | c\beta \rangle = c \langle \alpha | \beta \rangle, \qquad [A.88]$$

whereas

$$\langle c\alpha | \beta \rangle = c^* \langle \alpha | \beta \rangle, \qquad [A.89]$$

for any scalar c.

---

[18] Proving the converse (that if two diagonalizable matrices commute then they are simultaneously diagonalizable) is not so simple. See for example Eugen Merzbacher, *Quantum Mechanics*, 3rd ed., Wiley, New York (1998), Section 10.4.

[19] You may wonder whether such a transformation necessarily exists. Good question! The answer is "yes." See, for instance, P. R. Halmos, *Finite Dimensional Vector Spaces*, 2nd ed., van Nostrand, Princeton (1958), Section 44.

CHAPTER 3

FORMALISM

3.1 HILBERT SPACE

In the last two chapters we have stumbled on a number of interesting properties of simple quantum systems. Some of these are "accidental" features of specific potentials (the even spacing of energy levels for the harmonic oscillator, for example), but others seem to be more general, and it would be nice to prove them once and for all (the uncertainty principle, for instance, and the orthogonality of stationary states). The purpose of this chapter is to recast the theory in a more powerful form, with that in mind. There is not much here that is genuinely *new*; the idea, rather, is to make coherent sense of what we have already discovered in particular cases.

Quantum theory is based on two constructs: *wave functions* and *operators*. The state of a system is represented by its wave function, observables are represented by operators. Mathematically, wave functions satisfy the defining conditions for abstract **vectors**, and operators act on them as **linear transformations**. So the natural language of quantum mechanics is **linear algebra**.[1]

But it is not, I suspect, a form of linear algebra with which you are immediately familiar. In an N-dimensional space it is simplest to represent a vector, $|\alpha\rangle$, by the N-tuple of its components, $\{a_n\}$, with respect to a specified orthonormal basis:

$$|\alpha\rangle \rightarrow \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}. \qquad [3.1]$$

[1] If you have never studied linear algebra, you should read the Appendix before continuing.

The **inner product**, $\langle \alpha | \beta \rangle$, of two vectors (generalizing the dot product in three dimensions) is the complex number,

$$\langle \alpha | \beta \rangle = a_1^* b_1 + a_2^* b_2 + \cdots + a_N^* b_N. \quad [3.2]$$

Linear transformations, T, are represented by **matrices** (with respect to the specified basis), which act on vectors (to produce new vectors) by the ordinary rules of matrix multiplication:

$$|\beta\rangle = T|\alpha\rangle \rightarrow \mathbf{b} = \mathbf{T}\mathbf{a} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1N} \\ t_{21} & t_{22} & \cdots & t_{2N} \\ \vdots & \vdots & & \vdots \\ t_{N1} & t_{N2} & \cdots & t_{NN} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}. \quad [3.3]$$

But the "vectors" we encounter in quantum mechanics are (for the most part) *functions*, and they live in *infinite*-dimensional spaces. For them the N-tuple/matrix notation is awkward, at best, and manipulations that are well-behaved in the finite-dimensional case can be problematic. (The underlying reason is that whereas the *finite* sum in Equation 3.2 always exists, an *infinite* sum—or an integral—may not converge, in which case the inner product does not exist, and any argument involving inner products is immediately suspect.) So even though most of the terminology and notation should be familiar, it pays to approach this subject with caution.

The collection of *all* functions of x constitutes a vector space, but for our purposes it is much too large. To represent a possible physical state, the wave function Ψ must be *normalized*:

$$\int |\Psi|^2 \, dx = 1.$$

The set of all **square-integrable functions**, on a specified interval,[2]

$$f(x) \quad \text{such that} \quad \int_a^b |f(x)|^2 \, dx < \infty, \quad [3.4]$$

constitutes a (much smaller) vector space (see Problem 3.1(a)). Mathematicians call it $L^2(a, b)$; physicists call it **Hilbert space.**[3] In quantum mechanics, then,

Wave functions live in Hilbert space. [3.5]

[2] For us, the limits (a and b) will almost always be $\pm\infty$, but we might as well keep things more general for the moment.

[3] Technically, a Hilbert space is a **complete inner product space**, and the collection of square-integrable functions is only *one example* of a Hilbert space—indeed, every finite-dimensional vector space is trivially a Hilbert space. But since L^2 is the arena of quantum mechanics, it's what physicists generally *mean* when they say "Hilbert space." By the way, the word **complete** here means that any Cauchy sequence of functions in Hilbert space converges to a function that is also in the space: it has no "holes" in it, just as the set of all real numbers has no holes (by contrast, the space of all *polynomials*, for example, like the set of all *rational* numbers, certainly *does* have holes in it). The completeness of a *space* has nothing to do with the completeness (same word, unfortunately) of a *set of functions*, which is the property that any other function can be expressed as a linear combination of them.

We define the **inner product of two functions**, $f(x)$ and $g(x)$, as follows:

$$\langle f|g\rangle \equiv \int_a^b f(x)^* g(x)\,dx. \qquad [3.6]$$

If f and g are both square-integrable (that is, if they are both in Hilbert space), their inner product is guaranteed to exist (the integral in Equation 3.6 converges to a finite number).[4] This follows from the integral **Schwarz inequality**:[5]

$$\left|\int_a^b f(x)^* g(x)\,dx\right| \leq \sqrt{\int_a^b |f(x)|^2\,dx \int_a^b |g(x)|^2\,dx}. \qquad [3.7]$$

You can check for yourself that Equation 3.6 satisfies all the conditions for an inner product (Problem 3.1(b)). Notice in particular that

$$\langle g|f\rangle = \langle f|g\rangle^*. \qquad [3.8]$$

Moreover, the inner product of $f(x)$ with *itself*,

$$\langle f|f\rangle = \int_a^b |f(x)|^2\,dx, \qquad [3.9]$$

is *real* and non-negative; it's *zero* only[6] when $f(x) = 0$.

A function is said to be **normalized** if its inner product with itself is 1; two functions are **orthogonal** if their inner product is 0; and a *set* of functions, $\{f_n\}$, is **orthonormal** if they are normalized and mutually orthogonal:

$$\langle f_m|f_n\rangle = \delta_{mn}. \qquad [3.10]$$

Finally, a set of functions is **complete** if any *other* function (in Hilbert space) can be expressed as a linear combination of them:

$$f(x) = \sum_{n=1}^{\infty} c_n f_n(x). \qquad [3.11]$$

---

[4] In Chapter 2 we were obliged on occasion to work with functions that were *not* normalizable. Such functions lie *outside* Hilbert space, and we are going to have to handle them with special care, as you will see shortly. For the moment, I shall assume that all the functions we encounter *are* in Hilbert space.

[5] For a proof, see F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Unger, New York, 1955), Section 21. In a *finite* dimensional vector space the Schwarz inequality, $|\langle\alpha|\beta\rangle|^2 \leq \langle\alpha|\alpha\rangle\langle\beta|\beta\rangle$, is easy to prove (see Problem A.5). But that proof *assumes* the existence of the inner products, which is precisely what we are trying to *establish* here.

[6] What about a function that is zero everywhere except at a few isolated points? The integral (Equation 3.9) would still vanish, even though the function itself does not. If this bothers you, you should have been a math major. In physics such pathological functions do not occur, but in any case, in Hilbert space two functions that have the same square integral are considered equivalent. Technically, vectors in Hilbert space represent **equivalence classes** of functions.

If the functions $\{f_n(x)\}$ are orthonormal, the coefficients are given by Fourier's trick:

$$c_n = \langle f_n | f \rangle. \qquad [3.12]$$

as you can check for yourself. I anticipated this terminology, of course, back in Chapter 2. (The stationary states for the infinite square well (Equation 2.28) constitute a complete orthonormal set on the interval $(0, a)$; the stationary states for the harmonic oscillator (Equation 2.67 or 2.85) are a complete orthonormal set on the interval $(-\infty, \infty)$.)

Problem 3.1

(a) Show that the set of all square-integrable functions is a vector space (refer to Section A.1 for the definition). *Hint:* The main problem is to show that the sum of two square-integrable functions is itself square-integrable. Use Equation 3.7. Is the set of all *normalized* functions a vector space?

(b) Show that the integral in Equation 3.6 satisfies the conditions for an inner product (Section A.2).

***Problem 3.2**

(a) For what range of ν is the function $f(x) = x^\nu$ in Hilbert space, on the interval $(0, 1)$? Assume ν is real, but not necessarily positive.

(b) For the specific case $\nu = 1/2$, is $f(x)$ in this Hilbert space? What about $xf(x)$? How about $(d/dx)f(x)$?

3.2 OBSERVABLES

3.2.1 Hermitian Operators

The expectation value of an observable $Q(x, p)$ can be expressed very neatly in inner-product notation:[7]

$$\langle Q \rangle = \int \Psi^* \hat{Q} \Psi \, dx = \langle \Psi | \hat{Q} \Psi \rangle. \qquad [3.13]$$

---

[7] Remember that $\hat{Q}$ is the operator constructed from Q by the replacement $p \to \hat{p} \equiv (\hbar/i)d/dx$. These operators are **linear**, in the sense that

$$\hat{Q}[af(x) + bg(x)] = a\hat{Q}f(x) + b\hat{Q}g(x),$$

for any functions f and g and any complex numbers a and b. They constitute *linear transformations* (Section A.3) on the space of all functions. However, they sometimes carry a function *inside* Hilbert

Now, the outcome of a measurement has got to be *real*, and so, *a fortiori*, is the *average* of many measurements:

$$\langle Q \rangle = \langle Q \rangle^*. \qquad [3.14]$$

But the complex conjugate of an inner product reverses the order (Equation 3.8), so

$$\langle \Psi | \hat{Q}\Psi \rangle = \langle \hat{Q}\Psi | \Psi \rangle. \qquad [3.15]$$

and this must hold true for any wave function Ψ. Thus operators representing *observables* have the very special property that

$$\langle f | \hat{Q} f \rangle = \langle \hat{Q} f | f \rangle \quad \text{for all } f(x). \qquad [3.16]$$

We call such operators **hermitian.**

Actually, most books require an ostensibly stronger condition:

$$\langle f | \hat{Q} g \rangle = \langle \hat{Q} f | g \rangle \quad \text{for all } f(x) \text{ and all } g(x). \qquad [3.17]$$

But it turns out, in spite of appearances, that this is perfectly equivalent to my definition (Equation 3.16), as you will prove in Problem 3.3. So use whichever you like. The essential point is that a hermitian operator can be applied either to the first member of an inner product or to the second, with the same result, and hermitian operators naturally arise in quantum mechanics because their expectation values are real:

Observables are represented by hermitian operators. [3.18]

Well, let's *check* this. Is the momentum operator, for example, hermitian?

$$\langle f | \hat{p} g \rangle = \int_{-\infty}^{\infty} f^* \frac{\hbar}{i} \frac{dg}{dx}\, dx = \frac{\hbar}{i} f^* g \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \left(\frac{\hbar}{i} \frac{df}{dx} \right)^* g\, dx = \langle \hat{p} f | g \rangle. \qquad [3.19]$$

I used integration by parts, of course, and threw away the boundary term for the usual reason: If $f(x)$ and $g(x)$ are square integrable, they must go to zero at $\pm\infty$.[8]

---

space into a function *outside* it (see Problem 3.2(b)), and in this case the domain of the operator may have to be restricted.

[8] Actually, this is not quite true. As I mention in Chapter 1, there exist pathological functions that are square-integrable but do *not* go to zero at infinity. However, such functions do not arise in physics, and if you are worried about it we will simply restrict the domain of our operators to exclude them. On *finite* intervals, though, you really *do* have to be more careful with the boundary terms, and an operator that is hermitian on $(-\infty, \infty)$ may *not* be hermitian on $(0, \infty)$ or $(-\pi, \pi)$. If you're wondering about the infinite square well, it's safest to think of those wave functions as residing on the infinite line—they just happen to be *zero* outside $(0, a)$.

Notice how the complex conjugation of i compensates for the minus sign picked up from integration by parts—the operator d/dx (without the i) is *not* hermitian, and it does not represent a possible observable.

---

*Problem 3.3 Show that if $\langle h|\hat{Q}h\rangle = \langle \hat{Q}h|h\rangle$ for all functions h (in Hilbert space), then $\langle f|\hat{Q}g\rangle = \langle \hat{Q}f|g\rangle$ for all f and g (i.e., the two definitions of "hermitian"—Equations 3.16 and 3.17—are equivalent). *Hint:* First let $h = f + g$, and then let $h = f + ig$.

Problem 3.4

(a) Show that the *sum* of two hermitian operators is hermitian.

(b) Suppose $\hat{Q}$ is hermitian, and α is a complex number. Under what condition (on α) is $\alpha\hat{Q}$ hermitian?

(c) When is the *product* of two hermitian operators hermitian?

(d) Show that the position operator ($\hat{x} = x$) and the hamiltonian operator ($\hat{H} = -(\hbar^2/2m)d^2/dx^2 + V(x)$) are hermitian.

Problem 3.5 The **hermitian conjugate** (or **adjoint**) of an operator $\hat{Q}$ is the operator $\hat{Q}^\dagger$ such that

$$\langle f|\hat{Q}g\rangle = \langle \hat{Q}^\dagger f|g\rangle \quad \text{(for all } f \text{ and } g\text{)}. \tag{3.20}$$

(A hermitian operator, then, is equal to its hermitian conjugate: $\hat{Q} = \hat{Q}^\dagger$.)

(a) Find the hermitian conjugates of x, i, and d/dx.

(b) Construct the hermitian conjugate of the harmonic oscillator raising operator, a_+ (Equation 2.47).

(c) Show that $(\hat{Q}\hat{R})^\dagger = \hat{R}^\dagger\hat{Q}^\dagger$.

3.2.2 Determinate States

Ordinarily, when you measure an observable Q on an ensemble of identically prepared systems, all in the same state Ψ, you do *not* get the same result each time—this is the *indeterminacy* of quantum mechanics.[9] *Question:* Would it be possible to prepare a state such that *every* measurement of Q is certain to return the *same* value (call it q)? This would be, if you like, a **determinate state**, for the observable Q. (Actually, we already know one example: Stationary states are determinate states of the Hamiltonian; a measurement of the total energy, on a

---

[9] I'm talking about *competent* measurements, of course—it's always possible to make a *mistake*, and simply get the wrong answer, but that's not the fault of quantum mechanics.

particle in the stationary state Ψ_n, is certain to yield the corresponding "allowed" energy E_n.)

Well, the standard deviation of Q, in a determinate state, would be *zero*, which is to say,

$$\sigma^2 = \langle(\hat{Q} - \langle Q\rangle)^2\rangle = \langle\Psi|(\hat{Q} - q)^2\Psi\rangle = \langle(\hat{Q} - q)\Psi|(\hat{Q} - q)\Psi\rangle = 0. \quad [3.21]$$

(Of course, if every measurement gives q, their average is also q: $\langle Q\rangle = q$. I also used the fact that $\hat{Q}$, and hence also $\hat{Q} - q$, is a *hermitian* operator, to move one factor over to the first term in the inner product.) But the only function whose inner product with itself vanishes is 0, so

$$\hat{Q}\Psi = q\Psi. \quad [3.22]$$

This is the **eigenvalue equation** for the operator $\hat{Q}$; Ψ is an **eigenfunction** of $\hat{Q}$, and q is the corresponding **eigenvalue**. Thus

Determinate states are eigenfunctions of $\hat{Q}$. [3.23]

Measurement of Q on such a state is certain to yield the eigenvalue, q.

Note that the eigen*value* is a *number* (not an operator or a function). You can multiply any eigenfunction by a constant, and it is still an eigenfunction, with the same eigenvalue. Zero does not count as an eigenfunction (we exclude it by definition—otherwise *every* number would be an eigenvalue, since $\hat{Q}\,0 = q\,0 = 0$ for any operator $\hat{Q}$ and all q). But there's nothing wrong with zero as an eigen*value*. The collection of all the eigenvalues of an operator is called its **spectrum**. Sometimes two (or more) linearly independent eigenfunctions share the same eigenvalue; in that case the spectrum is said to be **degenerate**.

For example, determinate states of the total energy are eigenfunctions of the Hamiltonian:

$$\hat{H}\psi = E\psi, \quad [3.24]$$

which is precisely the time-independent Schrödinger equation. In this context we use the letter E for the eigenvalue, and the lower case ψ for the eigenfunction (tack on the factor $\exp(-iEt/\hbar)$ to make it Ψ, if you like; it's still an eigenfunction of H).

Example 3.1 Consider the operator

$$\hat{Q} \equiv i\frac{d}{d\phi}, \quad [3.25]$$

where ϕ is the usual polar coordinate in two dimensions. (This operator might arise in a physical context if we were studying the bead-on-a-ring; see Problem 2.46.) Is $\hat{Q}$ hermitian? Find its eigenfunctions and eigenvalues.

Solution: Here we are working with functions $f(\phi)$ on the *finite* interval $0 \leq \phi \leq 2\pi$, and stipulate that

$$f(\phi + 2\pi) = f(\phi). \qquad [3.26]$$

since ϕ and $\phi + 2\pi$ describe the same physical point. Using integration by parts,

$$\langle f|\hat{Q}\,g\rangle = \int_0^{2\pi} f^* \left(i\frac{dg}{d\phi}\right) d\phi = if^*g\Big|_0^{2\pi} - \int_0^{2\pi} i\left(\frac{df^*}{d\phi}\right) g\,d\phi = \langle \hat{Q}\,f|g\rangle,$$

so $\hat{Q}$ *is* hermitian (this time the boundary term disappears by virtue of Equation 3.26).

The eigenvalue equation,

$$i\frac{d}{d\phi}f(\phi) = qf(\phi), \qquad [3.27]$$

has the general solution

$$f(\phi) = Ae^{-iq\phi}. \qquad [3.28]$$

Equation 3.26 restricts the possible values of the q:

$$e^{-iq2\pi} = 1 \quad \Rightarrow \quad q = 0, \pm 1, \pm 2, \ldots \qquad [3.29]$$

The spectrum of this operator is the set of all integers, and it is nondegenerate.

Problem 3.6 Consider the operator $\hat{Q} = d^2/d\phi^2$, where (as in Example 3.1) ϕ is the azimuthal angle in polar coordinates, and the functions are subject to Equation 3.26. Is $\hat{Q}$ hermitian? Find its eigenfunctions and eigenvalues. What is the spectrum of $\hat{Q}$? Is the spectrum degenerate?

3.3 EIGENFUNCTIONS OF A HERMITIAN OPERATOR

Our attention is thus directed to the *eigenfunctions of hermitian operators* (physically: determinate states of observables). These fall into two categories: If the spectrum is **discrete** (i.e., the eigenvalues are separated from one another) then the eigenfunctions lie in Hilbert space and they constitute physically realizable states. If the spectrum is **continuous** (i.e., the eigenvalues fill out an entire range) then the eigenfunctions are not normalizable, and they do not represent possible wave functions (though *linear combinations* of them—involving necessarily a spread in eigenvalues—may be normalizable). Some operators have a discrete spectrum only (for example, the Hamiltonian for the harmonic oscillator), some have only a continuous spectrum (for example, the free particle Hamiltonian), and some have both a discrete part and a continuous part (for example, the Hamiltonian for a

finite square well). The discrete case is easier to handle, because the relevant inner products are guaranteed to exist—in fact, it is very similar to the finite-dimensional theory (the eigenvectors of a hermitian *matrix*). I'll treat the discrete case first, and then the continuous one.

3.3.1 Discrete Spectra

Mathematically, the normalizable eigenfunctions of a hermitian operator have two important properties:

Theorem 1: Their eigen*values* are *real.*

Proof: Suppose

$$\hat{Q}f = qf,$$

(i.e., $f(x)$ is an eigenfunction of $\hat{Q}$, with eigenvalue q), and[10]

$$\langle f|\hat{Q}f\rangle = \langle \hat{Q}f|f\rangle$$

($\hat{Q}$ is hermitian). Then

$$q\langle f|f\rangle = q^*\langle f|f\rangle$$

(q is a *number*, so it comes outside the integral, and because the first function in the inner product is complex conjugated (Equation 3.6), so too is the q on the right). But $\langle f|f\rangle$ cannot be zero ($f(x) = 0$ is not a legal eigenfunction), so $q = q^*$, and hence q is real. QED

This is comforting: If you measure an observable on a particle in a determinate state, you will at least get a real number.

Theorem 2: Eigenfunctions belonging to distinct eigenvalues are *orthogonal.*

Proof: Suppose

$$\hat{Q}f = qf, \quad \text{and} \quad \hat{Q}g = q'g,$$

and $\hat{Q}$ is hermitian. Then $\langle f|\hat{Q}g\rangle = \langle \hat{Q}f|g\rangle$, so

$$q'\langle f|g\rangle = q^*\langle f|g\rangle$$

(again, the inner products exist because the eigenfunctions are in Hilbert space by assumption). But q is real (from Theorem 1), so if $q' \neq q$ it must be that $\langle f|g\rangle = 0$. QED

---

[10] It is here that we assume the eigenfunctions are in Hilbert space—otherwise the inner product might not exist at all.

That's why the stationary states of the infinite square well, for example, or the harmonic oscillator, are orthogonal—they are eigenfunctions of the Hamiltonian with distinct eigenvalues. But this property is not peculiar to them, or even to the Hamiltonian—the same holds for determinate states of *any* observable.

Unfortunately, Theorem 2 tells us nothing about degenerate states ($q' = q$). However, if two (or more) eigenfunctions share the same eigenvalue, any linear combination of them is itself an eigenfunction, with the same eigenvalue (Problem 3.7(a)), and we can use the **Gram-Schmidt orthogonalization procedure** (Problem A.4) to *construct* orthogonal eigenfunctions within each degenerate subspace. It is almost never necessary to do this explicitly (thank God!), but it can always be done in principle. So *even in the presence of degeneracy* the eigenfunctions can be *chosen* to be orthogonal, and in setting up the formalism of quantum mechanics we shall assume that this has already been done. That licenses the use of Fourier's trick, which depends on the orthonormality of the basis functions.

In a *finite*-dimensional vector space the eigenvectors of a hermitian matrix have a third fundamental property: They span the space (every vector can be expressed as a linear combination of them). Unfortunately, the proof does not generalize to infinite-dimensional spaces. But the property itself is essential to the internal consistency of quantum mechanics, so (following Dirac[11]) we will take it as an *axiom* (or, more precisely, as a restriction on the class of hermitian operators that can represent observables):

> **Axiom:** The eigenfunctions of an observable operator are *complete*: Any function (in Hilbert space) can be expressed as a linear combination of them.[12]

Problem 3.7

(a) Suppose that $f(x)$ and $g(x)$ are two eigenfunctions of an operator $\hat{Q}$, with the same eigenvalue q. Show that any linear combination of f and g is itself an eigenfunction of $\hat{Q}$, with eigenvalue q.

(b) Check that $f(x) = \exp(x)$ and $g(x) = \exp(-x)$ are eigenfunctions of the operator d^2/dx^2, with the same eigenvalue. Construct two linear combinations of f and g that are *orthogonal* eigenfunctions on the interval $(-1, 1)$.

---

[11] P. A. M. Dirac, *The Principles of Quantum Mechanics*, Oxford University Press, New York (1958).

[12] In some specific cases completeness is provable (we know that the stationary states of the infinite square well, for example, are complete, because of Dirichlet's theorem). It is a little awkward to call something an "axiom" that is *provable* in some cases, but I don't know a better way to handle it.

Problem 3.8

(a) Check that the eigenvalues of the hermitian operator in Example 3.1 are real. Show that the eigenfunctions (for distinct eigenvalues) are orthogonal.

(b) Do the same for the operator in Problem 3.6.

3.3.2 Continuous Spectra

If the spectrum of a hermitian operator is *continuous*, the eigenfunctions are not normalizable, and the proofs of Theorems 1 and 2 fail, because the inner products may not exist. Nevertheless, there is a sense in which the three essential properties (reality, orthogonality, and completeness) still hold. I think it's best to approach this subtle case through specific examples.

Example 3.2 Find the eigenfunctions and eigenvalues of the momentum operator.

Solution: Let $f_p(x)$ be the eigenfunction and p the eigenvalue:

$$\frac{\hbar}{i}\frac{d}{dx}f_p(x) = pf_p(x). \qquad [3.30]$$

The general solution is

$$f_p(x) = Ae^{ipx/\hbar}.$$

This is not square-integrable, for *any* (complex) value of p—the momentum operator has *no* eigenfunctions in Hilbert space. And yet, if we restrict ourselves to *real* eigenvalues, we do recover a kind of *ersatz* "orthonormality." Referring to Problems 2.24(a) and 2.26,

$$\int_{-\infty}^{\infty} f_{p'}^*(x)\, f_p(x)\, dx = |A|^2 \int_{-\infty}^{\infty} e^{i(p-p')x/\hbar}\, dx = |A|^2 2\pi\hbar\, \delta(p - p'). \qquad [3.31]$$

If we pick $A = 1/\sqrt{2\pi\hbar}$, so that

$$f_p(x) = \frac{1}{\sqrt{2\pi\hbar}} e^{ipx/\hbar}, \qquad [3.32]$$

then

$$\langle f_{p'} | f_p \rangle = \delta(p - p'), \qquad [3.33]$$

which is strikingly reminiscent of *true* orthonormality (Equation 3.10)—the indices are now continuous variables, and the Kronecker delta has become a Dirac delta, but otherwise it looks just the same. I'll call Equation 3.33 **Dirac orthonormality**.

Most important, the eigenfunctions are *complete*, with the sum (in Equation 3.11) replaced by an integral: Any (square-integrable) function $f(x)$ can be written in the form

$$f(x) = \int_{-\infty}^{\infty} c(p)\, f_p(x)\, dp = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} c(p) e^{ipx/\hbar}\, dp. \qquad [3.34]$$

The expansion coefficient (now a *function*, $c(p)$) is obtained, as always, by Fourier's trick:

$$\langle f_{p'} | f \rangle = \int_{-\infty}^{\infty} c(p) \langle f_{p'} | f_p \rangle\, dp = \int_{-\infty}^{\infty} c(p) \delta(p - p')\, dp = c(p'). \qquad [3.35]$$

Alternatively, you can get them from Plancherel's theorem (Equation 2.102), for the expansion (Equation 3.34) is nothing but a Fourier transform.

The eigenfunctions of momentum (Equation 3.32) are sinusoidal, with wavelength

$$\lambda = \frac{2\pi\hbar}{p}. \qquad [3.36]$$

This is the old de Broglie formula (Equation 1.39), which I promised to prove at the appropriate time. It turns out to be a little more subtle than de Broglie imagined, because we now know that there is actually *no such thing* as a particle with determinate momentum. But we could make a normalizable wave *packet* with a narrow range of momenta, and it is to such an object that the de Broglie relation applies.

What are we to make of Example 3.2? Although none of the eigenfunctions of $\hat{p}$ lives in Hilbert space, a certain family of them (those with real eigenvalues) reside in the nearby "suburbs," with a kind of quasi-normalizability. They do not represent possible physical states, but they are still very useful (as we have already seen, in our study of one-dimensional scattering).[13]

Example 3.3 Find the eigenfunctions and eigenvalues of the position operator.

Solution: Let $g_y(x)$ be the eigenfunction and y the eigenvalue:

$$x\, g_y(x) = y\, g_y(x). \qquad [3.37]$$

---

[13] What about the eigenfunctions with *non*real eigenvalues? These are not merely nonnormalizable—they actually blow up at $\pm\infty$. Functions in what I called the "suburbs" of Hilbert space (the entire metropolitan area is sometimes called a "rigged Hilbert space"; see, for example, Leslie Ballentine's *Quantum Mechanics: A Modern Development*, World Scientific, 1998) have the property that although they have no (finite) inner product with *themselves*, they *do* admit inner products with all members of Hilbert space. This is *not* true for eigenfunctions of $\hat{p}$ with nonreal eigenvalues. In particular, I showed that the momentum operator is hermitian *for functions in Hilbert space*, but the argument depended on dropping the boundary term (in Equation 3.19). That term is still zero if g is an eigenfunction of $\hat{p}$ with a real eigenvalue (as long as f is in Hilbert space), but not if the eigenvalue has an imaginary part. In this sense *any* complex number is an eigenvalue of the operator $\hat{p}$, but only *real* numbers are eigenvalues of the *hermitian* operator $\hat{p}$—the others lie outside the space over which $\hat{p}$ is hermitian.

Here y is a fixed number (for any given eigenfunction), but x is a continuous variable. What function of x has the property that multiplying it by x is the same as multiplying it by the constant y? Obviously it's got to be *zero*, except at the one point $x = y$; in fact, it is nothing but the Dirac delta function:

$$g_y(x) = A\delta(x - y).$$

This time the eigenvalue *has* to be real; the eigenfunctions are not square-integrable, but again they admit *Dirac* orthonormality:

$$\int_{-\infty}^{\infty} g_{y'}^*(x)\, g_y(x)\, dx = |A|^2 \int_{-\infty}^{\infty} \delta(x - y')\delta(x - y)\, dx = |A|^2 \delta(y - y'). \quad [3.38]$$

If we pick $A = 1$, so

$$g_y(x) = \delta(x - y), \quad [3.39]$$

then

$$\langle g_{y'} | g_y \rangle = \delta(y - y'). \quad [3.40]$$

These eigenfunctions are also *complete*:

$$f(x) = \int_{-\infty}^{\infty} c(y)\, g_y(x)\, dy = \int_{-\infty}^{\infty} c(y)\delta(x - y)\, dy, \quad [3.41]$$

with

$$c(y) = f(y) \quad [3.42]$$

(trivial, in this case, but you can get it from Fourier's trick if you insist).

If the spectrum of a hermitian operator is *continuous* (so the eigenvalues are labeled by a continuous variable—p or y, in the examples; z, generically, in what follows), the eigenfunctions are not normalizable, they are not in Hilbert space and they do not represent possible physical states; nevertheless, the eigenfunctions with real eigenvalues are *Dirac* orthonormalizable and complete (with the sum now an integral). Luckily, this is all we really require.

Problem 3.9

(a) Cite a Hamiltonian from Chapter 2 (*other* than the harmonic oscillator) that has only a *discrete* spectrum.

(b) Cite a Hamiltonian from Chapter 2 (*other* than the free particle) that has only a *continuous* spectrum.

APPENDIX

LINEAR ALGEBRA

Linear algebra abstracts and generalizes the arithmetic of ordinary vectors, such as those we encounter in first-year physics. The generalization is in two directions: (1) We allow the scalars to be *complex* numbers, and (2) we do not restrict ourselves to three dimensions.

A.1 VECTORS

A **vector space** consists of a set of **vectors** ($|\alpha\rangle$, $|\beta\rangle$, $|\gamma\rangle$, ...), together with a set of **scalars** (a, b, c, ...),[1] which is **closed**[2] under two operations: vector addition and scalar multiplication.

- **Vector Addition**

  The "sum" of any two vectors is another vector:

$$|\alpha\rangle + |\beta\rangle = |\gamma\rangle. \tag{A.1}$$

  Vector addition is **commutative**:

$$|\alpha\rangle + |\beta\rangle = |\beta\rangle + |\alpha\rangle. \tag{A.2}$$

[1] For our purposes, the scalars will be ordinary complex numbers. Mathematicians can tell you about vector spaces over more exotic fields, but such objects play no role in quantum mechanics. Note that α, β, γ ... are *not* (ordinarily) numbers; they are *names* (labels)—"Charlie," for instance, or "F43A-9GL," or whatever you care to use to identify the vector in question.

[2] That is to say, these operations are always well-defined, and will never carry you outside the vector space.

and **associative**:

$$|\alpha\rangle + (|\beta\rangle + |\gamma\rangle) = (|\alpha\rangle + |\beta\rangle) + |\gamma\rangle. \quad [A.3]$$

There exists a **zero** (or **null**) **vector**,[3] $|0\rangle$, with the property that

$$|\alpha\rangle + |0\rangle = |\alpha\rangle, \quad [A.4]$$

for every vector $|\alpha\rangle$. And for every vector $|\alpha\rangle$ there is an associated **inverse vector** ($|-\alpha\rangle$),[4] such that

$$|\alpha\rangle + |-\alpha\rangle = |0\rangle. \quad [A.5]$$

- **Scalar Multiplication**

  The "product" of any scalar with any vector is another vector:

$$a|\alpha\rangle = |\gamma\rangle. \quad [A.6]$$

  Scalar multiplication is **distributive** with respect to vector addition:

$$a(|\alpha\rangle + |\beta\rangle) = a|\alpha\rangle + a|\beta\rangle, \quad [A.7]$$

  and with respect to scalar addition:

$$(a + b)|\alpha\rangle = a|\alpha\rangle + b|\alpha\rangle. \quad [A.8]$$

  It is also **associative** with respect to the ordinary multiplication of scalars:

$$a(b|\alpha\rangle) = (ab)|\alpha\rangle. \quad [A.9]$$

  Multiplication by the scalars 0 and 1 has the effect you would expect:

$$0|\alpha\rangle = |0\rangle: \quad 1|\alpha\rangle = |\alpha\rangle. \quad [A.10]$$

  Evidently $|-\alpha\rangle = (-1)|\alpha\rangle$ (which we write more simply as $-|\alpha\rangle$).

There's a lot less here than meets the eye—all I have done is to write down in abstract language the familiar rules for manipulating vectors. The virtue of such abstraction is that we will be able to apply our knowledge and intuition about the behavior of ordinary vectors to other systems that happen to share the same formal properties.

---

[3] It is customary, where no confusion can arise, to write the null vector without the adorning bracket: $|0\rangle \rightarrow 0$.

[4] This is funny notation, since α is not a number. I'm simply adopting the name "−Charlie" for the inverse of the vector whose name is "Charlie." More natural terminology will suggest itself in a moment.

A **linear combination** of the vectors $|\alpha\rangle$, $|\beta\rangle$, $|\gamma\rangle$, ..., is an expression of the form

$$a|\alpha\rangle + b|\beta\rangle + c|\gamma\rangle + \cdots. \tag{A.11}$$

A vector $|\lambda\rangle$ is said to be **linearly independent** of the set $|\alpha\rangle$, $|\beta\rangle$, $|\gamma\rangle$, ..., if it cannot be written as a linear combination of them. (For example, in three dimensions the unit vector $\hat{k}$ is linearly independent of $\hat{\imath}$ and $\hat{\jmath}$, but any vector in the xy plane is linearly *dependent* on $\hat{\imath}$ and $\hat{\jmath}$.) By extension, a *set* of vectors is "linearly independent" if each one is linearly independent of all the rest. A collection of vectors is said to **span** the space if *every* vector can be written as a linear combination of the members of this set.[5] A set of *linearly independent* vectors that spans the space is called a **basis**. The number of vectors in any basis is called the **dimension** of the space. For the moment we shall assume that the dimension (n) is *finite*.

With respect to a prescribed basis

$$|e_1\rangle, \ |e_2\rangle, \ldots, |e_n\rangle, \tag{A.12}$$

any given vector

$$|\alpha\rangle = a_1|e_1\rangle + a_2|e_2\rangle + \cdots + a_n|e_n\rangle, \tag{A.13}$$

is uniquely represented by the (ordered) n-tuple of its **components**:

$$|\alpha\rangle \leftrightarrow (a_1, a_2, \ldots, a_n). \tag{A.14}$$

It is often easier to work with the components than with the abstract vectors themselves. To add vectors, you add their corresponding components:

$$|\alpha\rangle + |\beta\rangle \leftrightarrow (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n); \tag{A.15}$$

to multiply by a scalar you multiply each component:

$$c|\alpha\rangle \leftrightarrow (ca_1, ca_2, \ldots, ca_n); \tag{A.16}$$

the null vector is represented by a string of zeroes:

$$|0\rangle \leftrightarrow (0, 0, \ldots, 0); \tag{A.17}$$

and the components of the inverse vector have their signs reversed:

$$|-\alpha\rangle \leftrightarrow (-a_1, -a_2, \ldots, -a_n). \tag{A.18}$$

---

[5] A set of vectors that spans the space is also called **complete**, though I personally reserve that word for the infinite-dimensional case, where subtle questions of convergence may arise.

The only *dis*advantage of working with components is that you have to commit yourself to a particular basis, and the same manipulations will look very different to someone working in a different basis.

Problem A.1 Consider the ordinary vectors in 3 dimensions ($a_x\hat{\imath} + a_y\hat{\jmath} + a_z\hat{k}$), with complex components.

(a) Does the subset of all vectors with $a_z = 0$ constitute a vector space? If so, what is its dimension; if not, why not?

(b) What about the subset of all vectors whose z component is 1? *Hint:* Would the sum of two such vectors be in the subset? How about the null vector?

(c) What about the subset of vectors whose components are all equal?

***Problem A.2** Consider the collection of all polynomials (with complex coefficients) of degree less than N in x.

(a) Does this set constitute a vector space (with the polynomials as "vectors")? If so, suggest a convenient basis, and give the dimension of the space. If not, which of the defining properties does it lack?

(b) What if we require that the polynomials be *even* functions?

(c) What if we require that the leading coefficient (i.e., the number multiplying x^{N-1}) be 1?

(d) What if we require that the polynomials have the value 0 at $x = 1$?

(e) What if we require that the polynomials have the value 1 at $x = 0$?

Problem A.3 Prove that the components of a vector with respect to a given basis are *unique*.

A.2 INNER PRODUCTS

In three dimensions we encounter two kinds of vector products: the dot product and the cross product. The latter does not generalize in any natural way to n-dimensional vector spaces, but the former *does*—in this context it is usually called the **inner product**. The inner product of two vectors ($|\alpha\rangle$ and $|\beta\rangle$) is a complex number, which we write as $\langle\alpha|\beta\rangle$, with the following properties:

$$\langle \beta | \alpha \rangle = \langle \alpha | \beta \rangle^*, \tag{A.19}$$

$$\langle \alpha | \alpha \rangle \geq 0, \quad \text{and} \quad \langle \alpha | \alpha \rangle = 0 \Leftrightarrow |\alpha\rangle = |0\rangle, \tag{A.20}$$

$$\langle \alpha | (b|\beta\rangle + c|\gamma\rangle) = b\langle \alpha | \beta \rangle + c\langle \alpha | \gamma \rangle. \tag{A.21}$$

Apart from the generalization to complex numbers, these axioms simply codify the familiar behavior of dot products. A vector space with an inner product is called an **inner product space.**

Because the inner product of any vector with itself is a nonnegative number (Equation A.20), its square root is *real*—we call this the **norm** of the vector:

$$\|\alpha\| \equiv \sqrt{\langle \alpha | \alpha \rangle}; \tag{A.22}$$

it generalizes the notion of "length." A **unit vector** (one whose norm is 1) is said to be **normalized** (the word should really be "normal," but I guess that sounds too anthropomorphic). Two vectors whose inner product is zero are called **orthogonal** (generalizing the notion of "perpendicular"). A collection of mutually orthogonal normalized vectors,

$$\langle \alpha_i | \alpha_j \rangle = \delta_{ij}, \tag{A.23}$$

is called an **orthonormal set.** It is always possible (see Problem A.4), and almost always convenient, to choose an *orthonormal basis*; in that case the inner product of two vectors can be written very neatly in terms of their components:

$$\langle \alpha | \beta \rangle = a_1^* b_1 + a_2^* b_2 + \cdots + a_n^* b_n, \tag{A.24}$$

the norm (squared) becomes

$$\langle \alpha | \alpha \rangle = |a_1|^2 + |a_2|^2 + \cdots + |a_n|^2, \tag{A.25}$$

and the components themselves are

$$a_i = \langle e_i | \alpha \rangle. \tag{A.26}$$

(These results generalize the familiar formulas $\mathbf{a} \cdot \mathbf{b} = a_x b_x + a_y b_y + a_z b_z$, $|\mathbf{a}|^2 = a_x^2 + a_y^2 + a_z^2$, and $a_x = \hat{\imath} \cdot \mathbf{a}$, $a_y = \hat{\jmath} \cdot \mathbf{a}$, $a_z = \hat{k} \cdot \mathbf{a}$, for the three-dimensional orthonormal basis $\hat{\imath}$, $\hat{\jmath}$, $\hat{k}$.) From now on we shall *always* work in orthonormal bases, unless it is explicitly indicated otherwise.

Another geometrical quantity one might wish to generalize is the *angle* between two vectors. In ordinary vector analysis $\cos\theta = (\mathbf{a} \cdot \mathbf{b})/|\mathbf{a}||\mathbf{b}|$. But because the inner product is in general a complex number, the analogous formula (in an

arbitrary inner product space) does not define a (real) angle θ. Nevertheless, it is still true that the *absolute value* of this quantity is a number no greater than 1,

$$|\langle\alpha|\beta\rangle|^2 \leq \langle\alpha|\alpha\rangle\langle\beta|\beta\rangle. \qquad [A.27]$$

(This important result is known as the **Schwarz inequality**; the proof is given in Problem A.5.) So you can, if you like, define the angle between $|\alpha\rangle$ and $|\beta\rangle$ by the formula

$$\cos\theta = \sqrt{\frac{\langle\alpha|\beta\rangle\langle\beta|\alpha\rangle}{\langle\alpha|\alpha\rangle\langle\beta|\beta\rangle}}. \qquad [A.28]$$

---

*Problem A.4 Suppose you start out with a basis $(|e_1\rangle, |e_2\rangle, \ldots, |e_n\rangle)$ that is *not* orthonormal. The **Gram-Schmidt procedure** is a systematic ritual for generating from it an orthonormal basis $(|e_1'\rangle, |e_2'\rangle, \ldots, |e_n'\rangle)$. It goes like this:

(i) Normalize the first basis vector (divide by its norm):

$$|e_1'\rangle = \frac{|e_1\rangle}{\|e_1\|}.$$

(ii) Find the projection of the second vector along the first, and subtract it off:

$$|e_2\rangle - \langle e_1'|e_2\rangle|e_1'\rangle.$$

This vector is orthogonal to $|e_1'\rangle$; normalize it to get $|e_2'\rangle$.

(iii) Subtract from $|e_3\rangle$ its projections along $|e_1'\rangle$ and $|e_2'\rangle$:

$$|e_3\rangle - \langle e_1'|e_3\rangle|e_1'\rangle - \langle e_2'|e_3\rangle|e_2'\rangle.$$

This is orthogonal to $|e_1'\rangle$ and $|e_2'\rangle$; normalize it to get $|e_3'\rangle$. And so on.

Use the Gram-Schmidt procedure to orthonormalize the 3-space basis $|e_1\rangle = (1+i)\hat{\imath} + (1)\hat{\jmath} + (i)\hat{k}$. $|e_2\rangle = (i)\hat{\imath} + (3)\hat{\jmath} + (1)\hat{k}$. $|e_3\rangle = (0)\hat{\imath} + (28)\hat{\jmath} + (0)\hat{k}$.

Problem A.5 Prove the Schwarz inequality (Equation A.27). *Hint:* Let $|\gamma\rangle = |\beta\rangle - (\langle\alpha|\beta\rangle/\langle\alpha|\alpha\rangle)|\alpha\rangle$, and use $\langle\gamma|\gamma\rangle \geq 0$.

Problem A.6 Find the angle (in the sense of Equation A.28) between the vectors $|\alpha\rangle = (1+i)\hat{\imath} + (1)\hat{\jmath} + (i)\hat{k}$ and $|\beta\rangle = (4-i)\hat{\imath} + (0)\hat{\jmath} + (2-2i)\hat{k}$.

Problem A.7 Prove the **triangle inequality**: $\|(|\alpha\rangle + |\beta\rangle)\| \leq \|\alpha\| + \|\beta\|$.

A.3 MATRICES

Suppose you take every vector (in 3-space) and multiply it by 17, or you rotate every vector by 39° about the z-axis, or you reflect every vector in the xy plane—these are all examples of **linear transformations**. A linear transformation[6] ($\hat{T}$) takes each vector in a vector space and "transforms" it into some other vector ($|\alpha\rangle \rightarrow |\alpha'\rangle = \hat{T}|\alpha\rangle$), subject to the condition that the operation be *linear*:

$$\hat{T}(a|\alpha\rangle + b|\beta\rangle) = a(\hat{T}|\alpha\rangle) + b(\hat{T}|\beta\rangle), \qquad [A.29]$$

for any vectors $|\alpha\rangle$, $|\beta\rangle$ and any scalars a, b.

If you know what a particular linear transformation does to a set of *basis* vectors, you can easily figure out what it does to *any* vector. For suppose that

$$\begin{aligned}
\hat{T}|e_1\rangle &= T_{11}|e_1\rangle + T_{21}|e_2\rangle + \cdots + T_{n1}|e_n\rangle, \\
\hat{T}|e_2\rangle &= T_{12}|e_1\rangle + T_{22}|e_2\rangle + \cdots + T_{n2}|e_n\rangle, \\
&\cdots \\
\hat{T}|e_n\rangle &= T_{1n}|e_1\rangle + T_{2n}|e_2\rangle + \cdots + T_{nn}|e_n\rangle,
\end{aligned}$$

or, more compactly,

$$\hat{T}|e_j\rangle = \sum_{i=1}^{n} T_{ij}|e_i\rangle, \quad (j = 1, 2, \ldots, n). \qquad [A.30]$$

If $|\alpha\rangle$ is an arbitrary vector,

$$|\alpha\rangle = a_1|e_1\rangle + a_2|e_2\rangle + \cdots + a_n|e_n\rangle = \sum_{j=1}^{n} a_j|e_j\rangle, \qquad [A.31]$$

then

$$\hat{T}|\alpha\rangle = \sum_{j=1}^{n} a_j\left(\hat{T}|e_j\rangle\right) = \sum_{j=1}^{n}\sum_{i=1}^{n} a_j T_{ij}|e_i\rangle = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} T_{ij}a_j\right)|e_i\rangle. \qquad [A.32]$$

Evidently $\hat{T}$ takes a vector with components $a_1, a_2, \ldots, a_n$ into a vector with components[7]

$$a'_i = \sum_{j=1}^{n} T_{ij}a_j. \qquad [A.33]$$

---

[6]In this chapter I'll use a hat (^) to denote linear transformations: this is not inconsistent with my convention in the text (putting hats on operators), for (as we shall see) quantum operators *are* linear transformations.

[7]Notice the reversal of indices between Equations A.30 and A.33. This is not a typographical error. Another way of putting it (switching $i \leftrightarrow j$ in Equation A.30) is that if the *components* transform with T_{ij}, the *basis* vectors transform with T_{ji}.

Thus the n^2 **elements** T_{ij} uniquely characterize the linear transformation $\hat{T}$ (with respect to a given basis), just as the n components a_i uniquely characterize the vector $|\alpha\rangle$ (with respect to the same basis):

$$\hat{T} \leftrightarrow (T_{11}, T_{12}, \ldots, T_{nn}). \qquad [A.34]$$

If the basis is orthonormal, it follows from Equation A.30 that

$$T_{ij} = \langle e_i | \hat{T} | e_j \rangle. \qquad [A.35]$$

It is convenient to display these complex numbers in the form of a **matrix**:[8]

$$\mathsf{T} = \begin{pmatrix} T_{11} & T_{12} & \cdots & T_{1n} \\ T_{21} & T_{22} & \cdots & T_{2n} \\ \vdots & \vdots & & \vdots \\ T_{n1} & T_{n2} & \cdots & T_{nn} \end{pmatrix}. \qquad [A.36]$$

The study of linear transformations reduces then to the theory of matrices. The *sum* of two linear transformations $(\hat{S} + \hat{T})$ is defined in the natural way:

$$(\hat{S} + \hat{T})|\alpha\rangle = \hat{S}|\alpha\rangle + \hat{T}|\alpha\rangle; \qquad [A.37]$$

this matches the usual rule for adding matrices (you add their corresponding elements):

$$\mathsf{U} = \mathsf{S} + \mathsf{T} \Leftrightarrow U_{ij} = S_{ij} + T_{ij}. \qquad [A.38]$$

The *product* of two linear transformations $(\hat{S}\hat{T})$ is the net effect of performing them in succession—first $\hat{T}$, then $\hat{S}$:

$$|\alpha'\rangle = \hat{T}|\alpha\rangle; \quad |\alpha''\rangle = \hat{S}|\alpha'\rangle = \hat{S}(\hat{T}|\alpha\rangle) = \hat{S}\hat{T}|\alpha\rangle. \qquad [A.39]$$

What matrix U represents the combined transformation $\hat{U} = \hat{S}\hat{T}$? It's not hard to work it out:

$$a_i'' = \sum_{j=1}^{n} S_{ij} a_j' = \sum_{j=1}^{n} S_{ij} \left(\sum_{k=1}^{n} T_{jk} a_k \right) = \sum_{k=1}^{n} \left(\sum_{j=1}^{n} S_{ij} T_{jk} \right) a_k = \sum_{k=1}^{n} U_{ik} a_k.$$

Evidently

$$\mathsf{U} = \mathsf{ST} \Leftrightarrow U_{ik} = \sum_{j=1}^{n} S_{ij} T_{jk}. \qquad [A.40]$$

[8] I'll use boldface capital letters, sans serif, to denote square matrices.

This is the standard rule for matrix multiplication—to find the ikth element of the product $\mathbf{ST}$, you look at the ith row of $\mathbf{S}$, and the kth column of $\mathbf{T}$, multiply corresponding entries, and add. The same prescription allows you to multiply *rectangular* matrices, as long as the number of columns in the first matches the number of rows in the second. In particular, if we write the n-tuple of components of $|\alpha\rangle$ as an $n \times 1$ **column matrix** (or "column vector"):[9]

$$\mathbf{a} \equiv \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}. \tag{A.41}$$

the transformation rule (Equation A.33) can be expressed as a matrix product:

$$\mathbf{a}' = \mathbf{T}\mathbf{a}. \tag{A.42}$$

Now some matrix terminology:

- The **transpose** of a matrix (which we shall write with a tilde: $\tilde{\mathbf{T}}$) is the same set of elements, but with rows and columns interchanged. In particular, the transpose of a *column* matrix is a **row matrix**:

$$\tilde{\mathbf{a}} = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}. \tag{A.43}$$

  For a *square* matrix taking the transpose amounts to reflecting in the **main diagonal** (upper left to lower right):

$$\tilde{\mathbf{T}} = \begin{pmatrix} T_{11} & T_{21} & \cdots & T_{n1} \\ T_{12} & T_{22} & \cdots & T_{n2} \\ \vdots & \vdots & & \vdots \\ T_{1n} & T_{2n} & \cdots & T_{nn} \end{pmatrix}. \tag{A.44}$$

  A (square) matrix is **symmetric** if it is equal to its transpose; it is **antisymmetric** if this operation reverses the sign:

$$\text{symmetric}: \tilde{\mathbf{T}} = \mathbf{T}; \quad \text{antisymmetric}: \tilde{\mathbf{T}} = -\mathbf{T}. \tag{A.45}$$

- The (complex) **conjugate** of a matrix (which we denote, as usual, with an asterisk, $\mathbf{T}^*$), consists of the complex conjugate of every element:

$$\mathbf{T}^* = \begin{pmatrix} T_{11}^* & T_{12}^* & \cdots & T_{1n}^* \\ T_{21}^* & T_{22}^* & \cdots & T_{2n}^* \\ \vdots & \vdots & & \vdots \\ T_{n1}^* & T_{n2}^* & \cdots & T_{nn}^* \end{pmatrix}; \quad \mathbf{a}^* = \begin{pmatrix} a_1^* \\ a_2^* \\ \vdots \\ a_n^* \end{pmatrix}. \tag{A.46}$$

[9] I'll use boldface lowercase letters, sans serif, for row and column matrices.

A matrix is **real** if all its elements are real, and **imaginary** if they are all imaginary:

$$\text{real}: \mathbf{T}^* = \mathbf{T}; \quad \text{imaginary}: \mathbf{T}^* = -\mathbf{T}. \tag{A.47}$$

- The **hermitian conjugate** (or **adjoint**) of a matrix (indicated by a dagger, $\mathbf{T}^\dagger$) is the transpose conjugate:

$$\mathbf{T}^\dagger \equiv \tilde{\mathbf{T}}^* = \begin{pmatrix} T_{11}^* & T_{21}^* & \cdots & T_{n1}^* \\ T_{12}^* & T_{22}^* & \cdots & T_{n2}^* \\ \vdots & \vdots & & \vdots \\ T_{1n}^* & T_{2n}^* & \cdots & T_{nn}^* \end{pmatrix}: \quad \mathbf{a}^\dagger \equiv \tilde{\mathbf{a}}^* = \begin{pmatrix} a_1^* & a_2^* & \cdots & a_n^* \end{pmatrix}. \tag{A.48}$$

A square matrix is **hermitian** (or **self-adjoint**) if it is equal to its hermitian conjugate; if hermitian conjugation introduces a minus sign, the matrix is **skew hermitian** (or **anti-hermitian**):

$$\text{hermitian}: \mathbf{T}^\dagger = \mathbf{T}; \quad \text{skew hermitian}: \mathbf{T}^\dagger = -\mathbf{T}. \tag{A.49}$$

In this notation the inner product of two vectors (with respect to an orthonormal basis—Equation A.24), can be written very neatly as a matrix product:

$$\langle \alpha | \beta \rangle = \mathbf{a}^\dagger \mathbf{b}. \tag{A.50}$$

Notice that each of the three operations defined in this paragraph, if applied twice, returns you to the original matrix.

Matrix multiplication is not, in general, commutative ($\mathbf{ST} \neq \mathbf{TS}$); the *difference* between the two orderings is called the **commutator**:[10]

$$[\mathbf{S}, \mathbf{T}] \equiv \mathbf{ST} - \mathbf{TS}. \tag{A.51}$$

The transpose of a product is the product of the transposes *in reverse order*:

$$\widetilde{(\mathbf{ST})} = \tilde{\mathbf{T}}\tilde{\mathbf{S}}, \tag{A.52}$$

(see Problem A.11), and the same goes for hermitian conjugates:

$$(\mathbf{ST})^\dagger = \mathbf{T}^\dagger \mathbf{S}^\dagger. \tag{A.53}$$

The **unit matrix** (representing a linear transformation that carries every vector into itself) consists of ones on the main diagonal, and zeroes everywhere else:

$$\mathbf{I} \equiv \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{A.54}$$

[10] The commutator only makes sense for *square* matrices, of course; for rectangular matrices the two orderings wouldn't even be the same size.

In other words,

$$\mathsf{I}_{ij} = \delta_{ij}. \tag{A.55}$$

The **inverse** of a (square) matrix (written T^{-1}) is defined in the obvious way:[11]

$$\mathsf{T}^{-1}\mathsf{T} = \mathsf{T}\mathsf{T}^{-1} = \mathsf{I}. \tag{A.56}$$

A matrix has an inverse if and only if its **determinant**[12] is nonzero; in fact,

$$\mathsf{T}^{-1} = \frac{1}{\det \mathsf{T}}\tilde{\mathsf{C}}, \tag{A.57}$$

where C is the matrix of **cofactors** (the cofactor of element T_{ij} is $(-1)^{i+j}$ times the determinant of the submatrix obtained from T by erasing the ith row and the jth column). A matrix that has no inverse is said to be **singular**. The inverse of a product (assuming it exists) is the product of the inverses *in reverse order*:

$$(\mathsf{ST})^{-1} = \mathsf{T}^{-1}\mathsf{S}^{-1}. \tag{A.58}$$

A matrix is **unitary** if its inverse is equal to its hermitian conjugate:[13]

$$\text{unitary}: \mathsf{U}^{\dagger} = \mathsf{U}^{-1}. \tag{A.59}$$

Assuming the basis is orthonormal, the columns of a unitary matrix constitute an orthonormal set, and so too do its rows (see Problem A.12). Linear transformations represented by unitary matrices preserve inner products, since (Equation A.50)

$$\langle \alpha' | \beta' \rangle = \mathbf{a}'^{\dagger}\mathbf{b}' = (\mathsf{U}\mathbf{a})^{\dagger}(\mathsf{U}\mathbf{b}) = \mathbf{a}^{\dagger}\mathsf{U}^{\dagger}\mathsf{U}\mathbf{b} = \mathbf{a}^{\dagger}\mathbf{b} = \langle \alpha | \beta \rangle. \tag{A.60}$$

∗**Problem A.8** Given the following two matrices:

$$\mathsf{A} = \begin{pmatrix} -1 & 1 & i \\ 2 & 0 & 3 \\ 2i & -2i & 2 \end{pmatrix}, \quad \mathsf{B} = \begin{pmatrix} 2 & 0 & -i \\ 0 & 1 & 0 \\ i & 3 & 2 \end{pmatrix}.$$

---

[11] Note that the left inverse is equal to the right inverse, for if $\mathsf{AT} = \mathsf{I}$ and $\mathsf{TB} = \mathsf{I}$, then (multiplying the second on the left by A and invoking the first) we get $\mathsf{B} = \mathsf{A}$.

[12] I assume you know how to evaluate determinants. If not, see M. Boas, *Mathematical Methods in the Physical Sciences*, 2nd ed. (John Wiley, New York, 1983), Section 3.3.

[13] In a *real* vector space (that is, one in which the scalars are real) the hermitian conjugate is the same as the transpose, and a unitary matrix is **orthogonal**: $\tilde{\mathsf{O}} = \mathsf{O}^{-1}$. For example, rotations in ordinary 3-space are represented by orthogonal matrices.

compute: (a) $\mathbf{A}+\mathbf{B}$, (b) $\mathbf{AB}$, (c) $[\mathbf{A},\mathbf{B}]$, (d) $\tilde{\mathbf{A}}$, (e) $\mathbf{A}^*$, (f) $\mathbf{A}^\dagger$, (g) $\det(\mathbf{B})$, and (h) $\mathbf{B}^{-1}$. Check that $\mathbf{B}\mathbf{B}^{-1}=\mathbf{I}$. Does $\mathbf{A}$ have an inverse?

---

*Problem A.9 Using the square matrices in Problem A.8, and the column matrices

$$\mathbf{a}=\begin{pmatrix} i \\ 2i \\ 2 \end{pmatrix}, \quad \mathbf{b}=\begin{pmatrix} 2 \\ (1-i) \\ 0 \end{pmatrix},$$

find: (a) $\mathbf{Aa}$, (b) $\mathbf{a}^\dagger\mathbf{b}$, (c) $\tilde{\mathbf{a}}\mathbf{B}\mathbf{b}$, (d) $\mathbf{a}\mathbf{b}^\dagger$.

Problem A.10 By explicit construction of the matrices in question, show that any matrix $\mathbf{T}$ can be written

(a) as the sum of a symmetric matrix $\mathbf{S}$ and an antisymmetric matrix $\mathbf{A}$;

(b) as the sum of a real matrix $\mathbf{R}$ and an imaginary matrix $\mathbf{M}$;

(c) as the sum of a hermitian matrix $\mathbf{H}$ and a skew-hermitian matrix $\mathbf{K}$.

---

*Problem A.11 Prove Equations A.52, A.53, and A.58. Show that the product of two unitary matrices is unitary. Under what conditions is the product of two hermitian matrices hermitian? Is the sum of two unitary matrices necessarily unitary? Is the sum of two hermitian matrices hermitian?

Problem A.12 Show that the rows and columns of a unitary matrix constitute orthonormal sets.

Problem A.13 Noting that $\det(\tilde{\mathbf{T}})=\det(\mathbf{T})$, show that the determinant of a hermitian matrix is real, the determinant of a unitary matrix has modulus 1 (hence the name), and the determinant of an orthogonal matrix is either $+1$ or -1.

A.4 CHANGING BASES

The components of a vector depend, of course, on your (arbitrary) choice of basis, and so do the elements of the matrix representing a linear transformation. We might inquire how these numbers change when we switch to a different basis.

The old basis vectors, $|e_i\rangle$ are—like *all* vectors—linear combinations of the new ones, $|f_i\rangle$:

$$\begin{aligned} |e_1\rangle &= S_{11}|f_1\rangle + S_{21}|f_2\rangle + \cdots + S_{n1}|f_n\rangle, \\ |e_2\rangle &= S_{12}|f_1\rangle + S_{22}|f_2\rangle + \cdots + S_{n2}|f_n\rangle, \\ &\cdots \\ |e_n\rangle &= S_{1n}|f_1\rangle + S_{2n}|f_2\rangle + \cdots + S_{nn}|f_n\rangle, \end{aligned}$$

(for some set of complex numbers S_{ij}), or, more compactly,

$$|e_j\rangle = \sum_{i=1}^{n} S_{ij}|f_i\rangle, \quad (j = 1, 2, \ldots, n). \tag{A.61}$$

This is *itself* a linear transformation (compare Equation A.30),[14] and we know immediately how the components transform:

$$a_i^f = \sum_{j=1}^{n} S_{ij} a_j^e. \tag{A.62}$$

(where the superscript indicates the basis). In matrix form

$$\mathbf{a}^f = \mathbf{S}\mathbf{a}^e. \tag{A.63}$$

What about the matrix representing a linear transformation $\hat{T}$—how is *it* modified by a change of basis? Well, in the old basis we had (Equation A.42)

$$\mathbf{a}'^e = \mathbf{T}^e \mathbf{a}^e,$$

and Equation A.63—multiplying both sides by $\mathbf{S}^{-1}$—entails[15] $\mathbf{a}^e = \mathbf{S}^{-1}\mathbf{a}^f$, so

$$\mathbf{a}'^f = \mathbf{S}\mathbf{a}'^e = \mathbf{S}(\mathbf{T}^e \mathbf{a}^e) = \mathbf{S}\mathbf{T}^e \mathbf{S}^{-1}\mathbf{a}^f.$$

Evidently

$$\mathbf{T}^f = \mathbf{S}\mathbf{T}^e\mathbf{S}^{-1}. \tag{A.64}$$

In general, two matrices ($\mathbf{T}_1$ and $\mathbf{T}_2$) are said to be **similar** if $\mathbf{T}_2 = \mathbf{S}\mathbf{T}_1\mathbf{S}^{-1}$ for some (nonsingular) matrix $\mathbf{S}$. What we have just found is that *matrices representing*

---

[14]Notice, however, the radically different perspective: In this case we're talking about one and the same *vector*, referred to two completely different *bases*, whereas before we were thinking of a completely *different* vector, referred to the *same* basis.

[15]Note that $\mathbf{S}^{-1}$ certainly exists—if $\mathbf{S}$ were singular, the $|f_i\rangle$'s would not span the space, so they wouldn't constitute a basis.

the same linear transformation, with respect to different bases, are similar. Incidentally, if the first basis is orthonormal, the second will also be orthonormal if and only if the matrix $\mathbf{S}$ is *unitary* (see Problem A.16). Since we always work in orthonormal bases, we are interested mainly in *unitary* similarity transformations.

While the *elements* of the matrix representing a given linear transformation may look very different in the new basis, two numbers associated with the matrix are unchanged: the determinant and the **trace**. For the determinant of a product is the product of the determinants, and hence

$$\det(\mathbf{T}^f) = \det(\mathbf{S}\mathbf{T}^e\mathbf{S}^{-1}) = \det(\mathbf{S})\det(\mathbf{T}^e)\det(\mathbf{S}^{-1}) = \det\mathbf{T}^e. \qquad [A.65]$$

And the trace, which is the *sum of the diagonal elements*,

$$\mathrm{Tr}(\mathbf{T}) \equiv \sum_{i=1}^{m} T_{ii}, \qquad [A.66]$$

has the property (see Problem A.17) that

$$\mathrm{Tr}(\mathbf{T}_1\mathbf{T}_2) = \mathrm{Tr}(\mathbf{T}_2\mathbf{T}_1), \qquad [A.67]$$

(for any two matrices $\mathbf{T}_1$ and $\mathbf{T}_2$), so

$$\mathrm{Tr}(\mathbf{T}^f) = \mathrm{Tr}(\mathbf{S}\mathbf{T}^e\mathbf{S}^{-1}) = \mathrm{Tr}(\mathbf{T}_e\mathbf{S}^{-1}\mathbf{S}) = \mathrm{Tr}(\mathbf{T}^e). \qquad [A.68]$$

Problem A.14 Using the standard basis $(\hat{\imath}, \hat{\jmath}, \hat{k})$ for vectors in three dimensions:

(a) Construct the matrix representing a rotation through angle θ (counterclockwise, looking down the axis toward the origin) about the z-axis.

(b) Construct the matrix representing a rotation by 120° (counterclockwise, looking down the axis) about an axis through the point (1,1,1).

(c) Construct the matrix representing reflection through the xy-plane.

(d) Check that all these matrices are orthogonal, and calculate their determinants.

Problem A.15 In the usual basis $(\hat{\imath}, \hat{\jmath}, \hat{k})$, construct the matrix $\mathbf{T}_x$ representing a rotation through angle θ about the x-axis, and the matrix $\mathbf{T}_y$ representing a rotation through angle θ about the y-axis. Suppose now we change bases, to $\hat{\imath}' = \hat{\jmath}$, $\hat{\jmath}' = -\hat{\imath}$, $\hat{k}' = \hat{k}$. Construct the matrix $\mathbf{S}$ that effects this change of basis, and check that $\mathbf{S}\mathbf{T}_x\mathbf{S}^{-1}$ and $\mathbf{S}\mathbf{T}_y\mathbf{S}^{-1}$ are what you would expect.

Problem A.16 Show that similarity preserves matrix multiplication (that is, if $\mathbf{A}^e\mathbf{B}^e = \mathbf{C}^e$, then $\mathbf{A}^f\mathbf{B}^f = \mathbf{C}^f$). Similarity does *not*, in general, preserve symmetry,

reality, or hermiticity; show, however, that if $\mathbf{S}$ is *unitary*, and $\mathbf{H}^e$ is hermitian, then $\mathbf{H}^f$ is hermitian. Show that $\mathbf{S}$ carries an orthonormal basis into another orthonormal basis if and only if it is unitary.

∗**Problem A.17** Prove that $\text{Tr}(\mathbf{T}_1\mathbf{T}_2) = \text{Tr}(\mathbf{T}_2\mathbf{T}_1)$. It follows immediately that $\text{Tr}(\mathbf{T}_1\mathbf{T}_2\mathbf{T}_3) = \text{Tr}(\mathbf{T}_2\mathbf{T}_3\mathbf{T}_1)$, but is it the case that $\text{Tr}(\mathbf{T}_1\mathbf{T}_2\mathbf{T}_3) = \text{Tr}(\mathbf{T}_2\mathbf{T}_1\mathbf{T}_3)$, in general? Prove it, or disprove it. *Hint:* The best disproof is always a counterexample—the simpler the better!

A.5 EIGENVECTORS AND EIGENVALUES

Consider the linear transformation in three-space consisting of a rotation, about some specified axis, by an angle θ. Most vectors will change in a rather complicated way (they ride around on a cone about the axis), but vectors that happen to lie *along* the axis have very simple behavior: They don't change at all ($\hat{T}|\alpha\rangle = |\alpha\rangle$). If θ is 180°, then vectors which lie in the "equatorial" plane reverse signs ($\hat{T}|\alpha\rangle = -|\alpha\rangle$). In a complex vector space[16] *every* linear transformation has "special" vectors like these, which are transformed into scalar multiples of themselves:

$$\hat{T}|\alpha\rangle = \lambda|\alpha\rangle; \qquad [A.69]$$

they are called **eigenvectors** of the transformation, and the (complex) number λ is their **eigenvalue**. (The *null* vector doesn't count, even though in a trivial sense it obeys Equation A.69 for *any* $\hat{T}$ and *any* λ; technically, an eigenvector is any *nonzero* vector satisfying Equation A.69.) Notice that any (nonzero) *multiple* of an eigenvector is still an eigenvector, with the same eigenvalue.

With respect to a particular basis, the eigenvector equation assumes the matrix form

$$\mathbf{Ta} = \lambda\mathbf{a}, \qquad [A.70]$$

(for nonzero $\mathbf{a}$), or

$$(\mathbf{T} - \lambda\mathbf{I})\mathbf{a} = \mathbf{0}. \qquad [A.71]$$

(Here $\mathbf{0}$ is the **zero matrix**, whose elements are all zero.) Now, if the matrix $(\mathbf{T} - \lambda\mathbf{I})$ had an *inverse*, we could multiply both sides of Equation A.71 by $(\mathbf{T} - \lambda\mathbf{I})^{-1}$, and

---

[16] This is *not* always true in a *real* vector space (where the scalars are restricted to real values). See Problem A.18.

conclude that $\mathbf{a} = \mathbf{0}$. But by assumption $\mathbf{a}$ is *not* zero, so the matrix $(\mathbf{T} - \lambda\mathbf{I})$ must in fact be singular, which means that its determinant is zero:

$$\det(\mathbf{T} - \lambda\mathbf{I}) = \begin{vmatrix} (T_{11} - \lambda) & T_{12} & \cdots & T_{1n} \\ T_{21} & (T_{22} - \lambda) & \cdots & T_{2n} \\ \vdots & \vdots & & \vdots \\ T_{n1} & T_{n2} & \cdots & (T_{nn} - \lambda) \end{vmatrix} = 0. \qquad [A.72]$$

Expansion of the determinant yields an algebraic equation for λ:

$$C_n\lambda^n + C_{n-1}\lambda^{n-1} + \cdots + C_1\lambda + C_0 = 0. \qquad [A.73]$$

where the coefficients C_i depend on the elements of $\mathbf{T}$ (see Problem A.20). This is called the **characteristic equation** for the matrix; its solutions determine the eigenvalues. Notice that it's an nth-order equation, so (by the **fundamental theorem of algebra**) it has n (complex) roots.[17] However, some of these may be multiple roots, so all we can say for certain is that an $n \times n$ matrix has *at least one* and *at most* n distinct eigenvalues. The collection of all the eigenvalues of a matrix is called its **spectrum**; if two (or more) linearly independent eigenvectors share the same eigenvalue, the spectrum is said to be **degenerate**.

To construct the eigen*vectors* it is generally easiest simply to plug each λ back into Equation A.70 and solve "by hand" for the components of $\mathbf{a}$. I'll show you how it goes by working out an example.

Example A.1 Find the eigenvalues and eigenvectors of the following matrix:

$$\mathbf{M} = \begin{pmatrix} 2 & 0 & -2 \\ -2i & i & 2i \\ 1 & 0 & -1 \end{pmatrix}. \qquad [A.74]$$

Solution: The characteristic equation is

$$\begin{vmatrix} (2 - \lambda) & 0 & -2 \\ -2i & (i - \lambda) & 2i \\ 1 & 0 & (-1 - \lambda) \end{vmatrix} = -\lambda^3 + (1 + i)\lambda^2 - i\lambda = 0, \qquad [A.75]$$

and its roots are 0, 1, and i. Call the components of the first eigenvector (a_1, a_2, a_3); then

$$\begin{pmatrix} 2 & 0 & -2 \\ -2i & i & 2i \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = 0 \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

---

[17] It is here that the case of *real* vector spaces becomes more awkward, because the characteristic equation need not have any (real) solutions at all. See Problem A.18.

which yields three equations:

$$\begin{aligned} 2a_1 - 2a_3 &= 0, \\ -2ia_1 + ia_2 + 2ia_3 &= 0, \\ a_1 - a_3 &= 0. \end{aligned}$$

The first determines a_3 (in terms of a_1): $a_3 = a_1$; the second determines a_2: $a_2 = 0$; and the third is redundant. We may as well pick $a_1 = 1$ (since any multiple of an eigenvector is still an eigenvector):

$$\mathbf{a}^{(1)} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \text{for } \lambda_1 = 0. \qquad [A.76]$$

For the second eigenvector (recycling the same notation for the components) we have

$$\begin{pmatrix} 2 & 0 & -2 \\ -2i & i & 2i \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = 1 \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix},$$

which leads to the equations

$$\begin{aligned} 2a_1 - 2a_3 &= a_1, \\ -2ia_1 + ia_2 + 2ia_3 &= a_2, \\ a_1 - a_3 &= a_3. \end{aligned}$$

with the solution $a_3 = (1/2)a_1$, $a_2 = [(1 - i)/2]a_1$; this time I'll pick $a_1 = 2$, so

$$\mathbf{a}^{(2)} = \begin{pmatrix} 2 \\ 1 - i \\ 1 \end{pmatrix}, \quad \text{for } \lambda_2 = 1. \qquad [A.77]$$

Finally, for the third eigenvector,

$$\begin{pmatrix} 2 & 0 & -2 \\ -2i & i & 2i \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = i \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} ia_1 \\ ia_2 \\ ia_3 \end{pmatrix},$$

which gives the equations

$$\begin{aligned} 2a_1 - 2a_3 &= ia_1, \\ -2ia_1 + ia_2 + 2ia_3 &= ia_2, \\ a_1 - a_3 &= ia_3. \end{aligned}$$

whose solution is $a_3 = a_1 = 0$, with a_2 undetermined. Choosing $a_2 = 1$, we conclude

$$\mathbf{a}^{(3)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \text{for } \lambda_3 = i. \tag{A.78}$$

If the eigenvectors span the space (as they do in the preceding example), we are free to use *them* as a basis:

$$\begin{aligned} \hat{T}|f_1\rangle &= \lambda_1|f_1\rangle, \\ \hat{T}|f_2\rangle &= \lambda_2|f_2\rangle, \\ &\cdots \\ \hat{T}|f_n\rangle &= \lambda_n|f_n\rangle. \end{aligned}$$

In this basis the matrix representing $\hat{T}$ takes on a very simple form, with the eigenvalues strung out along the main diagonal, and all other elements zero:

$$\mathbf{T} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}, \tag{A.79}$$

and the (normalized) eigenvectors are

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \tag{A.80}$$

A matrix that can be brought to **diagonal form** (Equation A.79) by a change of basis is said to be **diagonalizable** (evidently a matrix is diagonalizable if and only if its eigenvectors span the space). The similarity matrix that effects the diagonalization can be constructed by using the normalized eigenvectors (in the old basis) as the columns of $\mathbf{S}^{-1}$:

$$(\mathbf{S}^{-1})_{ij} = (\mathbf{a}^{(j)})_i. \tag{A.81}$$

Example A.2 In Example A.1,

$$\mathbf{S}^{-1} = \begin{pmatrix} 1 & 2 & 0 \\ 0 & (1-i) & 1 \\ 1 & 1 & 0 \end{pmatrix},$$

so (using Equation A.57)

$$\mathbf{S} = \begin{pmatrix} -1 & 0 & 2 \\ 1 & 0 & -1 \\ (i-1) & 1 & (1-i) \end{pmatrix},$$

and you can check for yourself that

$$\mathbf{Sa}^{(1)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{Sa}^{(2)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{Sa}^{(3)} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

and

$$\mathbf{SMS}^{-1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & i \end{pmatrix}.$$

There's an obvious advantage in bringing a matrix to diagonal form: It's much easier to work with. Unfortunately, not every matrix *can* be diagonalized—the eigenvectors have to span the space. If the characteristic equation has n distinct roots, then the matrix is certainly diagonalizable, but it *may* be diagonalizable even if there are multiple roots. (For an example of a matrix that *cannot* be diagonalized, see Problem A.19.) It would be handy to know in advance (before working out all the eigenvectors) whether a given matrix is diagonalizable. A useful sufficient (though not necessary) condition is the following: A matrix is said to be **normal** if it commutes with its hermitian conjugate:

$$\text{normal}: \quad [\mathbf{N}^\dagger, \mathbf{N}] = \mathbf{0}. \qquad [A.82]$$

Every normal matrix is diagonalizable (its eigenvectors span the space). In particular, every hermitian matrix, and every unitary matrix, is diagonalizable.

Suppose we have *two* diagonalizable matrices; in quantum applications the question often arises: Can they be **simultaneously diagonalized** (by the *same* similarity matrix **S**)? That is to say, does there exist a basis in which they are *both* diagonal? The answer is yes *if and only if the two matrices commute* (see Problem A.22).

∗**Problem A.18** The 2×2 matrix representing a rotation of the xy plane is

$$\mathbf{T} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \qquad [A.83]$$

Show that (except for certain special angles—what are they?) this matrix has no real eigenvalues. (This reflects the geometrical fact that no vector in the plane

is carried into itself under such a rotation; contrast rotations in *three* dimensions.) This matrix *does*, however, have *complex* eigenvalues and eigenvectors. Find them. Construct a matrix **S** that diagonalizes **T**. Perform the similarity transformation ($\mathbf{STS}^{-1}$) explicitly, and show that it reduces **T** to diagonal form.

Problem A.19 Find the eigenvalues and eigenvectors of the following matrix:

$$\mathbf{M} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Can this matrix be diagonalized?

Problem A.20 Show that the first, second, and last coefficients in the characteristic equation (Equation A.73) are:

$$C_n = (-1)^n, \quad C_{n-1} = (-1)^{n-1}\mathrm{Tr}(\mathbf{T}), \quad \text{and} \quad C_0 = \det(\mathbf{T}). \qquad [A.84]$$

For a 3×3 matrix with elements T_{ij}, what is C_1?

Problem A.21 It's obvious that the trace of a *diagonal* matrix is the sum of its eigenvalues, and its determinant is their product (just look at Equation A.79). It follows (from Equations A.65 and A.68) that the same holds for any *diagonalizable* matrix. Prove that in fact

$$\det(\mathbf{T}) = \lambda_1\lambda_2\cdots\lambda_n, \quad \mathrm{Tr}(\mathbf{T}) = \lambda_1 + \lambda_2 + \cdots + \lambda_n. \qquad [A.85]$$

for *any* matrix. (The λ's are the n solutions to the characteristic equation—in the case of multiple roots, there may be fewer linearly independent eigen*vectors* than there are solutions, but we still count each λ as many times as it occurs.) *Hint:* Write the characteristic equation in the form

$$(\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda) = 0.$$

and use the result of Problem A.20.

Problem A.22

(a) Show that if two matrices commute in *one* basis, then they commute in any basis. That is:

$$[\mathbf{T}_1^e, \mathbf{T}_2^e] = \mathbf{0} \Rightarrow [\mathbf{T}_1^f, \mathbf{T}_2^f] = \mathbf{0}. \qquad [A.86]$$

Hint: Use Equation A.64.

(b) Show that if two matrices are simultaneously diagonalizable, they commute.[18]

Problem A.23 Consider the matrix

$$\mathbf{M} = \begin{pmatrix} 1 & 1 \\ 1 & i \end{pmatrix}.$$

(a) Is it normal?

(b) Is it diagonalizable?

A.6 HERMITIAN TRANSFORMATIONS

In Equation A.48 I defined the hermitian conjugate (or "adjoint") of a *matrix* as its transpose-conjugate: $\mathbf{T}^\dagger = \tilde{\mathbf{T}}^*$. Now I want to give you a more fundamental definition for the hermitian conjugate of a *linear transformation*: It is that transformation $\hat{T}^\dagger$ which, when applied to the *first* member of an inner product, gives the same result as if $\hat{T}$ itself had been applied to the *second* vector:

$$\langle \hat{T}^\dagger \alpha | \beta \rangle = \langle \alpha | \hat{T} \beta \rangle, \qquad [A.87]$$

(for all vectors $|\alpha\rangle$ and $|\beta\rangle$).[19] I have to warn you that although everybody uses it, this is lousy notation. For α and β are not *vectors* (the *vectors* are $|\alpha\rangle$ and $|\beta\rangle$), they are *names*. In particular, they are endowed with no mathematical properties at all, and the expression "$\hat{T}\beta$" is literally *nonsense*: Linear transformations act on *vectors*, not *labels*. But it's pretty clear what the notation *means*: $\hat{T}\beta$ is the name of the vector $\hat{T}|\beta\rangle$, and $\langle \hat{T}^\dagger \alpha | \beta \rangle$ is the inner product of the vector $\hat{T}^\dagger |\alpha\rangle$ with the vector $|\beta\rangle$. Notice in particular that

$$\langle \alpha | c\beta \rangle = c \langle \alpha | \beta \rangle, \qquad [A.88]$$

whereas

$$\langle c\alpha | \beta \rangle = c^* \langle \alpha | \beta \rangle, \qquad [A.89]$$

for any scalar c.

---

[18] Proving the converse (that if two diagonalizable matrices commute then they are simultaneously diagonalizable) is not so simple. See for example Eugen Merzbacher, *Quantum Mechanics*, 3rd ed., Wiley, New York (1998), Section 10.4.

[19] You may wonder whether such a transformation necessarily exists. Good question! The answer is "yes." See, for instance, P. R. Halmos, *Finite Dimensional Vector Spaces*, 2nd ed., van Nostrand, Princeton (1958), Section 44.

If you're working in an orthonormal basis (as we always shall), the hermitian conjugate of a linear transformation is represented by the hermitian conjugate of the corresponding matrix; for (using Equations A.50 and A.53),

$$\langle \alpha | \hat{T}\beta \rangle = \mathbf{a}^\dagger \mathbf{T}\mathbf{b} = (\mathbf{T}^\dagger \mathbf{a})^\dagger \mathbf{b} = \langle \hat{T}^\dagger \alpha | \beta \rangle. \qquad [A.90]$$

So the terminology is consistent, and we can speak interchangeably in the language of transformations or of matrices.

In quantum mechanics, a fundamental role is played by **hermitian transformations** ($\hat{T}^\dagger = \hat{T}$). The eigenvectors and eigenvalues of a hermitian transformation have three crucial properties:

1. **The eigenvalues of a hermitian transformation are real.**

   Proof: Let λ be an eigenvalue of $\hat{T}$: $\hat{T}|\alpha\rangle = \lambda|\alpha\rangle$, with $|\alpha\rangle \neq |0\rangle$. Then

   $$\langle \alpha | \hat{T}\alpha \rangle = \langle \alpha | \lambda\alpha \rangle = \lambda \langle \alpha | \alpha \rangle.$$

   Meanwhile, if $\hat{T}$ is hermitian, then

   $$\langle \alpha | \hat{T}\alpha \rangle = \langle \hat{T}\alpha | \alpha \rangle = \langle \lambda\alpha | \alpha \rangle = \lambda^* \langle \alpha | \alpha \rangle.$$

   But $\langle \alpha | \alpha \rangle \neq 0$ (Equation A.20), so $\lambda = \lambda^*$, and hence λ is real. QED

2. **The eigenvectors of a hermitian transformation belonging to distinct eigenvalues are orthogonal.**

   Proof: Suppose $\hat{T}|\alpha\rangle = \lambda|\alpha\rangle$ and $\hat{T}|\beta\rangle = \mu|\beta\rangle$, with $\lambda \neq \mu$. Then

   $$\langle \alpha | \hat{T}\beta \rangle = \langle \alpha | \mu\beta \rangle = \mu \langle \alpha | \beta \rangle.$$

   and if $\hat{T}$ is hermitian,

   $$\langle \alpha | \hat{T}\beta \rangle = \langle \hat{T}\alpha | \beta \rangle = \langle \lambda\alpha | \beta \rangle = \lambda^* \langle \alpha | \beta \rangle.$$

   But $\lambda = \lambda^*$ (from **1**), and $\lambda \neq \mu$, by assumption, so $\langle \alpha | \beta \rangle = 0$. QED

3. **The eigenvectors of a hermitian transformation span the space.**
   As we have seen, this is equivalent to the statement that any hermitian matrix can be diagonalized (see Equation A.82). This rather technical fact is, in a sense, the mathematical support on which much of quantum mechanics leans. It turns out to be a thinner reed than one might have hoped, because the proof does not carry over to infinite-dimensional vector spaces.

Problem A.24 A hermitian linear transformation must satisfy $\langle\alpha|\hat{T}\beta\rangle = \langle\hat{T}\alpha|\beta\rangle$ for all vectors $|\alpha\rangle$ and $|\beta\rangle$. Prove that it is (surprisingly) sufficient that $\langle\gamma|\hat{T}\gamma\rangle = \langle\hat{T}\gamma|\gamma\rangle$ for all vectors $|\gamma\rangle$. *Hint:* First let $|\gamma\rangle = |\alpha\rangle + |\beta\rangle$, and then let $|\gamma\rangle = |\alpha\rangle + i|\beta\rangle$.

***Problem A.25** Let

$$\mathbf{T} = \begin{pmatrix} 1 & 1-i \\ 1+i & 0 \end{pmatrix}.$$

(a) Verify that **T** is hermitian.

(b) Find its eigenvalues (note that they are real).

(c) Find and normalize the eigenvectors (note that they are orthogonal).

(d) Construct the unitary diagonalizing matrix **S**, and check explicitly that it diagonalizes **T**.

(e) Check that det(**T**) and Tr(**T**) are the same for **T** as they are for its diagonalized form.

****Problem A.26** Consider the following hermitian matrix:

$$\mathbf{T} = \begin{pmatrix} 2 & i & 1 \\ -i & 2 & i \\ 1 & -i & 2 \end{pmatrix}.$$

(a) Calculate det(**T**) and Tr(**T**).

(b) Find the eigenvalues of **T**. Check that their sum and product are consistent with (a), in the sense of Equation A.85. Write down the diagonalized version of **T**.

(c) Find the eigenvectors of **T**. Within the degenerate sector, construct two linearly independent eigenvectors (it is this step that is always possible for a *hermitian* matrix, but not for an *arbitrary* matrix—contrast Problem A.19). Orthogonalize them, and check that both are orthogonal to the third. Normalize all three eigenvectors.

(d) Construct the unitary matrix **S** that diagonalizes **T**, and show explicitly that the similarity transformation using **S** reduces **T** to the appropriate diagonal form.

Problem A.27 A *unitary transformation* is one for which $\hat{U}^\dagger\hat{U} = 1$.

(a) Show that unitary transformations preserve inner products, in the sense that $\langle\hat{U}\alpha|\hat{U}\beta\rangle = \langle\alpha|\beta\rangle$, for all vectors $|\alpha\rangle$, $|\beta\rangle$.

(b) Show that the eigenvalues of a unitary transformation have modulus 1.

(c) Show that the eigenvectors of a unitary transformation belonging to distinct eigenvalues are orthogonal.

*****Problem A.28** *Functions* of matrices are defined by their Taylor series expansions; for example,

$$e^{\mathbf{M}} \equiv \mathbf{I} + \mathbf{M} + \frac{1}{2}\mathbf{M}^2 + \frac{1}{3!}\mathbf{M}^3 + \cdots. \tag{A.91}$$

(a) Find $\exp(\mathbf{M})$, if

$$\text{(i) } \mathbf{M} = \begin{pmatrix} 0 & 1 & 3 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{pmatrix}; \quad \text{(ii) } \mathbf{M} = \begin{pmatrix} 0 & \theta \\ -\theta & 0 \end{pmatrix}.$$

(b) Show that if $\mathbf{M}$ is diagonalizable, then

$$\det\left(e^{\mathbf{M}}\right) = e^{\mathrm{Tr}(\mathbf{M})}. \tag{A.92}$$

Comment: This is actually *true* even if $\mathbf{M}$ is *not* diagonalizable, but it's harder to prove in the general case.

(c) Show that if the matrices $\mathbf{M}$ and $\mathbf{N}$ *commute*, then

$$e^{\mathbf{M}+\mathbf{N}} = e^{\mathbf{M}} e^{\mathbf{N}}. \tag{A.93}$$

Prove (with the simplest counterexample you can think up) that Equation A.93 is *not* true, in general, for *non*-commuting matrices.

(d) If $\mathbf{H}$ is hermitian, show that $e^{i\mathbf{H}}$ is unitary.

If you're working in an orthonormal basis (as we always shall), the hermitian conjugate of a linear transformation is represented by the hermitian conjugate of the corresponding matrix; for (using Equations A.50 and A.53),

$$\langle \alpha | \hat{T}\beta \rangle = \mathbf{a}^\dagger \mathbf{T}\mathbf{b} = (\mathbf{T}^\dagger \mathbf{a})^\dagger \mathbf{b} = \langle \hat{T}^\dagger \alpha | \beta \rangle. \qquad [A.90]$$

So the terminology is consistent, and we can speak interchangeably in the language of transformations or of matrices.

In quantum mechanics, a fundamental role is played by **hermitian transformations** ($\hat{T}^\dagger = \hat{T}$). The eigenvectors and eigenvalues of a hermitian transformation have three crucial properties:

1. **The eigenvalues of a hermitian transformation are real.**

   Proof: Let λ be an eigenvalue of $\hat{T}$: $\hat{T}|\alpha\rangle = \lambda|\alpha\rangle$, with $|\alpha\rangle \neq |0\rangle$. Then

   $$\langle \alpha | \hat{T}\alpha \rangle = \langle \alpha | \lambda\alpha \rangle = \lambda \langle \alpha | \alpha \rangle.$$

   Meanwhile, if $\hat{T}$ is hermitian, then

   $$\langle \alpha | \hat{T}\alpha \rangle = \langle \hat{T}\alpha | \alpha \rangle = \langle \lambda\alpha | \alpha \rangle = \lambda^* \langle \alpha | \alpha \rangle.$$

   But $\langle \alpha | \alpha \rangle \neq 0$ (Equation A.20), so $\lambda = \lambda^*$, and hence λ is real. QED

2. **The eigenvectors of a hermitian transformation belonging to distinct eigenvalues are orthogonal.**

   Proof: Suppose $\hat{T}|\alpha\rangle = \lambda|\alpha\rangle$ and $\hat{T}|\beta\rangle = \mu|\beta\rangle$, with $\lambda \neq \mu$. Then

   $$\langle \alpha | \hat{T}\beta \rangle = \langle \alpha | \mu\beta \rangle = \mu \langle \alpha | \beta \rangle.$$

   and if $\hat{T}$ is hermitian,

   $$\langle \alpha | \hat{T}\beta \rangle = \langle \hat{T}\alpha | \beta \rangle = \langle \lambda\alpha | \beta \rangle = \lambda^* \langle \alpha | \beta \rangle.$$

   But $\lambda = \lambda^*$ (from **1**), and $\lambda \neq \mu$, by assumption, so $\langle \alpha | \beta \rangle = 0$. QED

3. **The eigenvectors of a hermitian transformation span the space.**
   As we have seen, this is equivalent to the statement that any hermitian matrix can be diagonalized (see Equation A.82). This rather technical fact is, in a sense, the mathematical support on which much of quantum mechanics leans. It turns out to be a thinner reed than one might have hoped, because the proof does not carry over to infinite-dimensional vector spaces.

Problem A.24 A hermitian linear transformation must satisfy $\langle\alpha|\hat{T}\beta\rangle = \langle\hat{T}\alpha|\beta\rangle$ for all vectors $|\alpha\rangle$ and $|\beta\rangle$. Prove that it is (surprisingly) sufficient that $\langle\gamma|\hat{T}\gamma\rangle = \langle\hat{T}\gamma|\gamma\rangle$ for all vectors $|\gamma\rangle$. *Hint:* First let $|\gamma\rangle = |\alpha\rangle + |\beta\rangle$, and then let $|\gamma\rangle = |\alpha\rangle + i|\beta\rangle$.

***Problem A.25** Let

$$\mathbf{T} = \begin{pmatrix} 1 & 1-i \\ 1+i & 0 \end{pmatrix}.$$

(a) Verify that **T** is hermitian.

(b) Find its eigenvalues (note that they are real).

(c) Find and normalize the eigenvectors (note that they are orthogonal).

(d) Construct the unitary diagonalizing matrix **S**, and check explicitly that it diagonalizes **T**.

(e) Check that det(**T**) and Tr(**T**) are the same for **T** as they are for its diagonalized form.

****Problem A.26** Consider the following hermitian matrix:

$$\mathbf{T} = \begin{pmatrix} 2 & i & 1 \\ -i & 2 & i \\ 1 & -i & 2 \end{pmatrix}.$$

(a) Calculate det(**T**) and Tr(**T**).

(b) Find the eigenvalues of **T**. Check that their sum and product are consistent with (a), in the sense of Equation A.85. Write down the diagonalized version of **T**.

(c) Find the eigenvectors of **T**. Within the degenerate sector, construct two linearly independent eigenvectors (it is this step that is always possible for a *hermitian* matrix, but not for an *arbitrary* matrix—contrast Problem A.19). Orthogonalize them, and check that both are orthogonal to the third. Normalize all three eigenvectors.

(d) Construct the unitary matrix **S** that diagonalizes **T**, and show explicitly that the similarity transformation using **S** reduces **T** to the appropriate diagonal form.

Problem A.27 A *unitary transformation* is one for which $\hat{U}^\dagger\hat{U} = 1$.

(a) Show that unitary transformations preserve inner products, in the sense that $\langle\hat{U}\alpha|\hat{U}\beta\rangle = \langle\alpha|\beta\rangle$, for all vectors $|\alpha\rangle$, $|\beta\rangle$.

(b) Show that the eigenvalues of a unitary transformation have modulus 1.

(c) Show that the eigenvectors of a unitary transformation belonging to distinct eigenvalues are orthogonal.

*****Problem A.28** *Functions* of matrices are defined by their Taylor series expansions; for example,

$$e^{\mathbf{M}} \equiv \mathbf{I} + \mathbf{M} + \frac{1}{2}\mathbf{M}^2 + \frac{1}{3!}\mathbf{M}^3 + \cdots. \tag{A.91}$$

(a) Find $\exp(\mathbf{M})$, if

$$\text{(i) } \mathbf{M} = \begin{pmatrix} 0 & 1 & 3 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{pmatrix}; \quad \text{(ii) } \mathbf{M} = \begin{pmatrix} 0 & \theta \\ -\theta & 0 \end{pmatrix}.$$

(b) Show that if $\mathbf{M}$ is diagonalizable, then

$$\det\left(e^{\mathbf{M}}\right) = e^{\mathrm{Tr}(\mathbf{M})}. \tag{A.92}$$

Comment: This is actually *true* even if $\mathbf{M}$ is *not* diagonalizable, but it's harder to prove in the general case.

(c) Show that if the matrices $\mathbf{M}$ and $\mathbf{N}$ *commute*, then

$$e^{\mathbf{M}+\mathbf{N}} = e^{\mathbf{M}} e^{\mathbf{N}}. \tag{A.93}$$

Prove (with the simplest counterexample you can think up) that Equation A.93 is *not* true, in general, for *non*-commuting matrices.

(d) If $\mathbf{H}$ is hermitian, show that $e^{i\mathbf{H}}$ is unitary.
