CHAPTER 5

HILBERT SPACE— COMPLETE ORTHONORMAL SETS OF FUNCTIONS

INTRODUCTION

The principal strands of mathematics—algebra, geometry, and analysis—are joined when functions are viewed as vectors in a vector space. This unification is the subject of this chapter; none of the results or methods of mathematics has more relevance to modern physics.

Throughout this chapter we shall be using concepts, terminology, and theorems from Chapters 3 and 4. The present chapter represents the infinite-dimensional generalization of certain of the results obtained there. However, the fact that the vectors are now functions means that there are a number of additional considerations and possibilities in the development of the theory. These result primarily from the attempt to represent a function as a linear combination of some given set of functions—i. e., the problem of series expansions. All the characteristic questions of analysis, such as those of convergence, therefore become relevant.

In particular, the so-called "special functions" of mathematical physics—spherical harmonics, Legendre, Hermite, and Laguerre polynomials, to name a few—are all conveniently treated within the framework of Hilbert space. As we deal with the various special functions, we shall discuss the differential equations of which they are solutions. But the treatment will not be based on the differential equations. The framework of Hilbert space is a much more comprehensive one, and unifies what is otherwise a bewildering maze of special cases and properties.

There will be three stages in the discussion of these functions. In the first stage, where we introduce the basic notions of Hilbert space, the treatment will be quite abstract; the special functions will be mentioned only in passing. Then we shall consider the special functions one by one, somewhat inductively, deemphasizing their common origins and properties. Finally, we shall return to a more abstract level and systematize the properties of these functions in a single comprehensive framework.

Hilbert space is the mathematical setting for quantum physics. Physical observables are represented by operators in Hilbert space, and physical states are

vectors (functions) in Hilbert space. There is no more important property of the functions that describe the possible states of a physical system than that they form a *complete* set. The whole of quantum theory is based on this fact. Thus most of this chapter will in one way or another have to do with the completeness of sets of functions.

At the end of the chapter we shall bring together the key results of the theory of Hilbert space that have been developed throughout the chapter, and emphasize their role in the formulation of quantum mechanics.

5.1 FUNCTION SPACE AND HILBERT SPACE

In discussing finite-dimensional vector spaces, we referred several times to P_n, the vector space of polynomials of degree $\leqq n$. The vectors in that vector space are a class of simple functions, the polynomials.

We are now going to define another vector space whose elements are functions. The elements of this space are the complex-valued functions of a real variable x, defined on the closed interval $[a, b]$,* which are *square integrable*.† We shall show that the set of square integrable functions forms a vector space. This space is called L_2 by mathematicians: we shall call it *function space*. It will be found to be an infinite-dimensional space.

Intuitively, it would seem that function space is far "larger" than the finite-dimensional space P_n, and in a sense this is so. However, suppose that the basis functions for P_n—the set $\{x^m, m = 0, 1, \cdots, n\}$—were extended to contain all possible powers of x by letting $n \to \infty$. Weierstrass's theorem, the central result of this chapter, shows that this infinite set of functions is not as poverty-stricken as it might appear. We shall shortly give a precise account of its latent powers. First we return to the definition of function space.

Addition of the two vectors, f_1 and f_2, in function space is defined according to the natural rule:

$$(f_1 + f_2)(x) \equiv f_1(x) + f_2(x) ,$$

and multiplication by a complex scalar α is defined as

$$(\alpha f)(x) \equiv \alpha f(x) .$$

The only possible difficulty in showing that these operations satisfy the various axioms that define a vector space is establishing closure. That is, are the sums and scalar multiples of square-integrable functions also square-integrable? The

---

* A *closed interval*, written $[a, b]$, is the set of all points $\{x\}$ such that $a \leq x \leq b$. An *open interval*, which we write as (a, b), is the set of points $\{x\}$ such that $a < x < b$. A closed interval is always finite; if either a or b is infinite, the interval is open at that end. Later in the chapter, when we deal with infinite intervals, we shall alert the reader by explicit mention of the shift from finite to infinite intervals.

† A function is *square integrable* on $[a, b]$ if $\int_a^b |f(x)|^2 dx$ exists and is finite.

answer is yes, and so the space is in fact a vector space. We prove closure of the sum.

$$\begin{aligned} |f_1 + f_2|^2 &= |f_1|^2 + |f_2|^2 + f_1^* f_2 + f_1 f_2^* \\ &= |f_1|^2 + |f_2|^2 + 2\mathrm{Re}\,(f_1^* f_2) \\ &\leqq |f_1|^2 + |f_2|^2 + 2|f_1^* f_2| \\ &\leqq |f_1|^2 + |f_2|^2 + 2|f_1||f_2| \,. \end{aligned} \tag{5.1}$$

Also,

$$0 \leqq (|f_1| - |f_2|)^2 = |f_1|^2 + |f_2|^2 - 2|f_1||f_2| \,,$$

so

$$|f_1|^2 + |f_2|^2 \geqq 2|f_1||f_2| \,.$$

We use this last inequality to replace $2|f_1||f_2|$ in Eq. (5.1) with something larger, thereby preserving the inequality. Thus the inequality

$$|f_1 + f_2|^2 \leqq 2|f_1|^2 + 2|f_2|^2 \,, \tag{5.2}$$

holds at every point in $[a, b]$. Integrating over both sides, we see that square integrability of f_1 and f_2 ensures square integrability of their sum.

We now add an inner product to function space.

Definition 5.1. The inner product of two functions, f_1 and f_2, belonging to function space, is defined by

$$(f_1, f_2) \equiv \int_a^b f_1^*(x) f_2(x)\, dx \,. \tag{5.3}$$

Note that square integrability implies that $(f, f) \equiv ||f||^2 = \int_a^b |f|^2\, dx < \infty$. The quantity $||f||$ is called the norm of f. The inner product for any pair of square-integrable functions also exists. Since

$$|f_1^* f_2| = |f_1||f_2| \leqq \tfrac{1}{2}(|f_1|^2 + |f_2|^2) \,,$$

it follows that

$$\int_a^b |f_1^* f_2|\, dx \leqq \tfrac{1}{2}(||f_1||^2 + ||f_2||^2) < \infty \,;$$

but any function whose absolute value is integrable is itself integrable, so (f_1, f_2) exists.

Verification that (f_1, f_2) as defined in Eq. (5.3) is an inner product proceeds very smoothly until we come to the question of positive-definiteness. Clearly,

$$(f, f) \equiv ||f||^2 = \int_a^b |f|^2\, dx \geqq 0 \,.$$

But does $(f, f) = 0$ imply that $f(x) = 0$ for *all* x in $[a, b]$? Not quite, for the function $f(x)$ can be nonzero at any finite number of points, and the integral will not "notice" this. That is, there will be no contribution to the integral even though the integrand is not identically zero throughout $[a, b]$.

The discussion of situations like these is facilitated if a slightly more general notion of integration is used. The Riemann integral suffers from several difficulties; for example, let us consider an extreme case by looking at the bizarre function $f(x)$ defined to be 1 for every rational number in [0, 1] and 0 for every non-rational number. Now, since there are "very few" rationals, only a countable number in fact, we strongly suspect that the integral of this function is zero. However, if we form the upper and lower Riemann integrals by partitioning [0, 1] into small segments Δx_i, and write

$$\overline{\int} f(x)\,dx = \sum_i \Delta x_i \max\,[f(x)], \qquad x_i \leq x \leq x_i + \Delta x_i$$

$$\underline{\int} f(x)\,dx = \sum_i \Delta x_i \min\,[f(x)], \qquad x_i \leq x \leq x_i + \Delta x_i$$

in the usual way, we see that no matter how small the subinterval, Δx_i, the maximum of $f(x)$ on this interval is always 1 and the minimum is always zero. Thus

$$\overline{\int} f(x)\,dx = 1 \qquad \text{and} \qquad \underline{\int} f(x)\,dx = 0\,,$$

so the Riemann integral does not exist.

In the theory of Lebesgue integration, this integral does exist and it equals zero. We say that $f(x) = 0$ except on a set of points of *measure zero*, or $f(x) = 0$ *almost everywhere*. The intuitive content of this phrase is simple. If we have a countable number of points on the real line and are given a small strip of paper of length ϵ, then we can paste a small piece of the strip over each element of the set by dividing it into a countable number of pieces of width $\epsilon/2^n$. Since $\sum_{n=1}^{\infty} \epsilon/2^n = \epsilon$, we use up only our given strip in the process. But since the original strip can be arbitrarily small, the set of points on which $f(x)$ is nonzero is negligible with respect to the set on which it is zero, despite the fact that every real number is arbitrarily close to some rational. Thus, the rationals are a set of measure zero on the real line.

We shall go no further with the notion of the Lebesgue integral, since all the functions which we will actually want to integrate in this book will be Riemann integrable. For example, the Riemann integral always exists for the class of piecewise continuous functions defined on closed intervals. In any case, the Riemann integral, when it exists, is equal to the Lebesgue integral.

To summarize: If f is a function in function space and $(f, f) = 0$, then $f(x)$ need not be identically zero for all x, but it can differ from zero only on a set of measure zero. We say that $(f, f) = 0$ implies $f(x) = 0$ almost everywhere. Any function which equals zero almost everywhere is called the zero function. With this broadened concept of the zero function, Axiom 3 for an inner product is satisfied, so Definition 5.1 does define an inner product.

Thus we have a complex inner product defined in function space, which is itself a complex vector space of complex square-integrable functions of a *real* variable ranging through the closed interval $[a, b]$.

To be useful to the physicist, the inner-product space must also be *complete*. A complete space is one in which there exists no Cauchy sequence of elements of the space which tend toward limits outside the space.* An elementary example of an incomplete space is the set of rational numbers. The sequence of partial sums, $S_N = \sum_{n=0}^{N} 1/n!$ is a sequence of rationals, but it converges to the number e, which is irrational. However, a basic theorem of real analysis states that the set of all real numbers is complete; one cannot get out of the set by taking the limits of Cauchy sequences of real numbers.

Similarly, we want to discover a class of *functions* having the property that there will be no Cauchy sequence of functions in this class whose limits do not belong to the class. Such a class of functions is *complete*. This absolutely fundamental problem in analysis is solved by the Riesz-Fischer theorem, which states that the space of square-integrable functions, i. e., functions with a finite norm, is complete. The proof may be found in any book on functional analysis (e. g., Riesz and Nagy, Rudin). The theorem may be stated as follows.

Riesz-Fischer Theorem. Let the functions $f_1(x), f_2(x), \cdots$ be elements in function space. If

$$\lim_{n,\, m\to\infty} ||f_n - f_m||^2 \equiv \lim_{n,\, m\to\infty} \int_a^b |f_n(x) - f_m(x)|^2\, dx = 0\,,$$

then there exists a square (Lebesgue) integrable function $f(x)$ to which the sequence $f_n(x)$ converges "in the mean"; i. e., there exists an f such that

$$\lim_{n\to\infty} \int_a^b |f(x) - f_n(x)|^2\, dx = 0\,.$$

We shall discuss convergence in the mean in greater detail in the following section. It should be emphasized that this theorem is not true unless the integral used is the Lebesgue integral.

Thus function space, as we have defined it, is in fact complete. Henceforth we shall call this complete inner-product space by its usual name, *Hilbert space*, although function space is not the only Hilbert space.

The notions of orthogonality, normalization, and orthonormal sets of functions are defined exactly as for vectors. Thus the set of functions $\{f_i\}$ is said to be *orthonormal* if

$$(f_i, f_j) \equiv \int_a^b f_i^*(x) f_j(x)\, dx = \delta_{ij}\,.$$

---

* A *Cauchy sequence* is a sequence $\{s_n\}$ which has the property that given an arbitrary number $\epsilon > 0$, there is an index $N(\epsilon)$ such that if m and n are larger than $N(\epsilon)$,

$$||s_n - s_m|| < \epsilon.$$

If the sequence in question is a sequence of numbers, then the norm is interpreted as the absolute value. If, as in our case, the sequence is a sequence of functions, then the norm is given by

$$||f|| \equiv \left[\int_a^b |f|^2\, dx\right]^{1/2}.$$

Example. *The Fourier functions.* Let $f_n(x) = e^{inx}/\sqrt{2\pi}$, where $n = 0, \pm 1, \pm 2, \cdots$ The set $\{f_n\}$ is orthonormal over the interval $[-\pi, \pi]$.

Proof.

$$(f_n, f_m) = \int_{-\pi}^{+\pi} f_n^* f_m \, dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(m-n)x} \, dx$$

$$= \frac{1}{2\pi(m-n)i} e^{i(m-n)x} \Big]_{-\pi}^{\pi} = 0. \qquad \text{if } m \neq n ;$$

$$(f_n, f_n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} 1 \, dx = 1, \qquad \text{if } m = n .$$

Therefore

$$(f_n, f_m) = \delta_{nm} .$$

An often used generalization of orthonormality is the following:

Definition 5.2. A set of functions $\{f_n\}$ is orthonormal with respect to a nonnegative weight function $w(x)$ on $[a, b]$ if

$$(f_n, f_m) \equiv \int_a^b f_n^*(x) f_m(x) w(x) \, dx = \delta_{nm} .$$

We shall work almost exclusively with orthonormal sets of functions, because, as in finite-dimensional vector spaces, they greatly simplify computations.

5.2 COMPLETE ORTHONORMAL SETS OF FUNCTIONS

In the theory of finite-dimensional vector spaces, we found a number of equivalent alternative characterizations of a complete set of basis vectors (Theorem 4.3). The corresponding problem in Hilbert space is that of representing a function as a linear combination of some given set of functions, or, in other words, the problem of series expansions of functions in terms of a given set. The prototype of all such series expansions is the Fourier series. Our treatment, however, will be a general one, embracing many of the functions of mathematical physics; Fourier series will be treated within this framework as a special case.

The first question we must face is that of defining the *completeness of an orthonormal set of functions* in Hilbert space. (Completeness of a *set* of functions is not the same as completeness of a *space*, mentioned briefly in the last section; however, they are intimately connected as we shall see.) We could say that an orthonormal set of functions $\{f_i(x)\}$ is complete if any function $f(x)$ in Hilbert space is expressible as a linear combination of the $f_i(x)$:

$$f(x) = \sum_{i=1}^{\infty} c_i f_i(x) ,$$

the series converging to f at every point x. This would provide a close analog of the idea of completeness of a set of basis vectors in finite-dimensional spaces. But this criterion of convergence is, for many purposes, unnecessarily severe and exclusive. As we shall see, by this criterion there would exist no complete

orthonormal set of functions in Hilbert space. So, instead of demanding strict pointwise convergence, we shall weaken the convergence criterion, and in this way permit the existence of complete sets of functions.

The appropriate weakening of the convergence criterion is suggested by the slight difficulty we encountered with the positive-definiteness of the inner product. There we found that

$$(f, f) = \int_a^b |f|^2 \, dx = 0$$

implied that $f(x)$ vanished not at *every* point x in $[a, b]$, but rather, almost everywhere in $[a, b]$; that is, at all but a set of points of measure zero. Similarly, we shall say that $\sum_i c_i f_i(x)$ converges "in the mean" to $f(x)$ if

$$\lim_{n\to\infty} \int_a^b |f(x) - \sum_i^n c_i f_i(x)|^2 \, dx = 0 \, .$$

This allows the series to differ from $f(x)$ on a set of measure zero.

In order to develop the notion of completeness of an orthonormal set of functions further, we need to define the different kinds of convergence we shall be using:

1. Pointwise convergence;
2. Uniform convergence;
3. Convergence in the mean.

Definition 5.3. A sequence of functions, $h_n(x)$, converges pointwise to $h(x)$ on $[a, b]$ if for every x in $[a, b]$ and every $\epsilon > 0$ there exists an integer $N(x, \epsilon)$ such that for $n > N$,

$$|h(x) - h_n(x)| < \epsilon \, .$$

The $h_n(x)$ may themselves be the partial sums of another sequence; that is,

$$h_n(x) = \sum_{i=1}^n k_i(x) \, .$$

The definition of pointwise convergence, and the other types of convergence to follow, may all be stated equivalently in terms of the infinite series $\sum_{i=1}^{\infty} k_i(x)$, which is the limit of the sequence $h_n(x)$ of partial sums. We say that this series *converges pointwise* to $h(x)$ on $[a, b]$ if for every x in $[a, b]$ and every $\epsilon > 0$, there exists an integer $N(x, \epsilon)$ such that for $n > N$,

$$|h(x) - h_n(x)| = |h(x) - \sum_{i=1}^n k_i(x)| < \epsilon \, .$$

If there is a single N that works in the above definition for all x in $[a, b]$, the convergence is said to be *uniform*. Formally, we have:

Definition 5.4. A sequence of functions, $h_n(x)$, converges *uniformly* to $h(x)$ on $[a, b]$ if for every $\epsilon > 0$ there exists an integer $N(\epsilon)$, *independent of* x,

such that for $n > N$, $|h(x) - h_n(x)| < \epsilon$, for all x in $[a, b]$. Clearly, uniform convergence implies pointwise convergence.

Note that this definition makes explicit mention of the limit $h(x)$ of the sequence of functions, $h_n(x)$. The *Cauchy criterion* for uniform convergence supplies us with a useful alternative to this definition, in which knowledge of the limit function is *not* assumed. Cauchy's criterion is proved in analysis courses; we state it here.

Theorem 5.1. The sequence of functions $h_n(x)$ converges uniformly on $[a, b]$ if for every $\epsilon > 0$ there exists an integer $N(\epsilon)$ such that for all $r > N$, $s > N$, and x in $[a, b]$, $|h_r(x) - h_s(x)| < \epsilon$.
In terms of the partial sums, $h_n(x) \equiv \sum_{i=1}^{n} k_i(x)$, this becomes

$$|h_r(x) - h_s(x)| = \left| \sum_{i=1}^{r} k_i - \sum_{i=1}^{s} k_i \right| = \left| \sum_{i=r+1}^{s} k_i(x) \right| < \epsilon .$$

If we have uniform or pointwise convergence, we may write

$$h(x) = \lim_{n\to\infty} h_n(x) = \sum_{i=1}^{\infty} k_i(x) .$$

The weaker convergence we discussed earlier can now be formalized precisely.

Definition 5.5. A sequence of functions $h_n(x)$, converges *in the mean* to $h(x)$ on $[a, b]$ if

$$\lim_{n\to\infty} \int_a^b |h(x) - h_n(x)|^2 \, dx = 0,$$

that is, if for every ϵ there exists an $N(\epsilon)$ such that for $n > N$,

$$\int_a^b |h(x) - h_n(x)|^2 \, dx < \epsilon .$$

The series $\sum_{i=1}^{\infty} k_i(x)$ converges in the mean to $h(x)$ if

$$\lim_{n\to\infty} \int_a^b \left| h(x) - \sum_{i=1}^{n} k_i(x) \right|^2 dx = 0 .$$

It is easy to see that uniform convergence implies mean convergence. For if convergence is uniform, then for any ϵ there exists an N such that for $n > N$, $|h - h_n| < \epsilon$ for all x in $[a, b]$. Therefore

$$\int_a^b |h - h_n|^2 \, dx < \int_a^b \epsilon^2 \, dx = \epsilon^2 (b - a) ,$$

so the integral can be made arbitrarily small by choice of ϵ. This is just the statement that

$$\lim_{n\to\infty} \int_a^b |h - h_n|^2 \, dx = 0,$$

and so we have mean convergence. Note, however, that pointwise convergence does not imply mean convergence (Problem 19).

It is in terms of *mean* convergence that the completeness of an orthonormal set of functions is defined.

Definition 5.6. Let $g(x)$ be any function in Hilbert space (i. e., any square-integrable function) and let $\{f_i(x)\}$ be an orthonormal set of functions in Hilbert space. If there exist constant coefficients $\{a_i\}$ such that the sequence of partial sums $g_n(x) \equiv \sum_{i=1}^{n} a_i f_i(x)$ converges in the mean to $g(x)$, then the *set of functions* $\{f_i\}$ *is a complete orthonormal set.* Equivalently, if the mean square error can be made arbitrarily small,

$$\lim_{n\to\infty} \int_a^b |g - g_n|^2 \, dx = \lim_{n\to\infty} \int_a^b |g - \sum_{i=1}^{n} a_i f_i|^2 \, dx = 0 \, ,$$

then the set $\{f_i\}$ is a *complete orthonormal set of functions.* It should be noted that the coefficients $\{a_i\}$ are independent of n. Thus as n increases and we include more terms in the partial sum approximating g, the earlier coefficients do not change. We may say that as we extend the sum to infinity, the infinite *series*

$$\sum_{i=1}^{\infty} a_i f_i$$

approximates the arbitrary function g in the mean. We shall write this as

$$g(x) \doteq \sum_{i=1}^{\infty} a_i f_i(x) \, .$$

The dot over the equal sign serves to distinguish mean convergence from pointwise convergence.

Since mean convergence does not necessarily imply pointwise or uniform convergence, it should be clear that the completeness of an orthonormal set of functions $\{f_i\}$, expressed by the relation

$$\lim_{n\to\infty} \int_a^b |f - \sum_{i=1}^{n} a_i f_i|^2 \, dx = 0 \, ,$$

or symbolically,

$$f(x) \doteq \sum_{i=1}^{\infty} a_i f_i(x) \, ,$$

does not imply that

$$f(x) = \sum_{i=1}^{\infty} a_i f_i(x) \, . \tag{5.4}$$

We may only *equate* $f(x)$ and the expansion series if the series converges pointwise or uniformly to $f(x)$.

Let $f(x)$ be an arbitrary function in Hilbert space, and assume for the moment that we have an orthonormal set of functions $\{f_i(x)\}$ such that the series

$\sum_{i=1}^{\infty} c_i f_i(x)$ converges uniformly to $f(x)$:

$$f(x) = \sum_{i=1}^{\infty} c_i f_i(x) .$$

The coefficients c_i are called the *generalized Fourier coefficients*, or *expansion coefficients*. The formula for them is especially simple because $\{f_i\}$ is an orthonormal set:

$$(f_n, f) = \sum_{i=1}^{\infty} c_i (f_n, f_i) = \sum_{i=1}^{\infty} c_i \delta_{ni} = c_n . \tag{5.5}$$

Since the convergence is uniform, mean convergence is also implied. Therefore

$$\lim_{n\to\infty} \int_a^b |f - \sum_{i=1}^{n} c_i f_i|^2 \, dx = 0 ,$$

[where the c_i are given by Eq. (5.5)], and, consequently, the set of functions $\{f_i\}$ is complete.

Now consider the nonnegative quantity

$$M_n = \int_a^b |f(x) - \sum_{i=1}^{n} a_i f_i(x)|^2 \, dx \geq 0 , \tag{5.6}$$

where $\{f_i\}$ is an orthonormal set, and f is any function in Hilbert space. We want to know: What values of the coefficients a_i will minimize M_n? Or, to put it in the language of the physicist: What values of the a_i will give the best least-squares fit to the arbitrary function $f(x)$?

To answer this question, we expand Eq. (5.6) to obtain

$$\begin{aligned} M_n &= \int_a^b (f^*f - \sum_{i=1}^{n} a_i f^* f_i - \sum_{i=1}^{n} a_i^* f f_i^* + \sum_{i,j=1}^{n} a_i^* a_j f_i^* f_j) \, dx \\ &= (f, f) - \sum_{i=1}^{n} a_i c_i^* - \sum_{i=1}^{n} a_i^* c_i + \sum_{i,j=1}^{n} a_i^* a_j \delta_{ij} , \end{aligned}$$

where $c_i \equiv (f_i, f)$. Adding and subtracting the quantity $\sum_{i=1}^{n} |c_i|^2$, we have

$$M_n = (f, f) + \sum_{i=1}^{n} |a_i - c_i|^2 - \sum_{i=1}^{n} |c_i|^2 \geq 0 .$$

It is clear that M_n is minimized by choosing $a_i = c_i$. We then have

$$\begin{aligned} M_n &= \int_a^b |f(x) - \sum_{i=1}^{n} c_i f_i(x)|^2 \, dx \\ &= (f, f) - \sum_{i=1}^{n} |c_i|^2 \geq 0 , \end{aligned} \tag{5.7}$$

which may be rewritten as

$$(f, f) \geq \sum_{i=1}^{n} |c_i|^2 = \sum_{i=1}^{n} |(f_i, f)|^2 . \tag{5.8}$$

Writing $s_n = \sum_{i=1}^{n} |c_i|^2$, we see that the sequence $s_1, s_2, \cdots, s_n, \cdots$ is monotonically increasing and bounded above by (f, f), (Eq. 5.8). Therefore s_n tends toward a limit as n becomes arbitrarily large, and we can write

$$(f, f) \geq \sum_{i=1}^{\infty} |c_i|^2 . \tag{5.9}$$

Hence the infinite series $\sum_{i=1}^{\infty} |c_i|^2$ converges. Equation (5.9) is just Bessel's inequality (compare Theorem 4.2) in an infinite-dimensional space.

The orthonormal set $\{f_i\}$ is complete if and only if there exists a set $\{a_i\}$ such that $\lim_{n\to\infty} M_n = 0$. If the set $\{f_i\}$ is complete, then $a_i = c_i$ and the equal sign holds in Bessel's inequality:

$$(f, f) = \sum_{i=1}^{\infty} |c_i|^2 = \sum_{i=1}^{\infty} |(f_i, f)|^2 , \tag{5.10}$$

for every f. This was also true in the finite-dimensional case (see Theorem 4.3). Equation (5.10) is called the *completeness relation*. As in the finite-dimensional case, it can also be stated in the form of Parseval's equation:

$$(f, g) = \sum_{i=1}^{\infty} (f, f_i)(f_i, g) . \tag{5.11}$$

We leave the proof to the reader (Problem 14).

Another characterization of a complete set of n vectors is that there exists no nonzero vector orthogonal to all n vectors in the set. The corresponding statement holds for a complete orthonormal set of functions. We shall present it as a therorem. It will prove very useful in what follows. To facilitate the statement of the theorem, we first make a definition.

Definition 5.7. A set of orthonormal functions is said to be *closed* if no nonzero function is orthogonal to every function in the set.

Theorem 5.2. A set of orthonormal functions in Hilbert space is complete if and only if it is closed.

Proof. We first prove that completeness of the set implies that the set is closed. Assume that there is a nonzero function $f(x)$ (and let it be normalized), such that

$$(f_i, f) \equiv c_i = \int_a^b f_i^*(x) f(x)\, dx = 0$$

for all i. Then

$$\lim_{n\to\infty} \int_a^b |f - \sum_{i=1}^{n} c_i f_i|^2\, dx = \int_a^b |f|^2\, dx = 1 \neq 0 ,$$

(since f is normalized), so the set $\{f_i\}$ is not complete. Thus completeness of an orthonormal set of functions implies that there are no functions that are orthogonal to every member of the set.

We now prove the converse: if the orthonormal set is closed, it is complete. If it is not complete, then the completeness relation, Eq. (5.10), is not satisfied. Thus there exists some function $f(x)$ such that

$$||f||^2 > \sum_{n=1}^{\infty} |c_n|^2 ,$$

where $c_n = (f_n, f)$. But since the above infinite series is convergent, the sequence $\{g_m(x)\}$, where

$$g_m(x) = \sum_{n=1}^{m} c_n f_n(x)$$

is a Cauchy sequence in Hilbert space, and therefore, because of the *completeness of the space*, the $g_m(x)$ must converge in the mean to a limit *in the space*, call it $g(x)$, such that $c_n = (f_n, g)$. Therefore $(f_n, g) = (f_n, f)$ so $(f_n, f\text{-}g) = 0$. Thus $f(x)\text{-}g(x)$ is orthogonal to $f_n(x)$ for all n. We now show that the norm of $f(x)\text{-}g(x)$ is not equal to zero, so $\{f_n(x)\}$ is not closed, contrary to our assumption. It will then follow by contradiction that the set $\{f_n(x)\}$ is complete and the proof will be finished.

Using the inequality

$$||x - y|| \geq | \, ||x|| - ||y|| \, |$$

(see Problem 4.7b), we have

$$||f - g|| = ||f - g_m - (g - g_m)|| \geq \Big| ||f - g_m|| - ||g - g_m|| \Big|$$

for all m. Now as $m \to \infty$, we know that $||g - g_m|| \to 0$, whereas

$$||f - g_m||^2 = ||f - \sum_{n=1}^{m} c_n f_n||^2 = ||f||^2 - \sum_{n=1}^{m} |c_n|^2 > 0$$

for all m by assumption. Thus $||f - g|| > 0$ and the proof is complete.

This theorem plays a key role in establishing the completeness of the various orthonormal sets of functions which we shall treat in this chapter. Note the crucial use that is made of the completeness of the Hilbert space in proving it.

We conclude this section with some observations on the uniqueness of the representations of functions in series expansions. We first prove that a function in Hilbert space is uniquely determined almost everywhere by its expansion coefficients with respect to a given complete orthonormal set of functions $\{f_i\}$. Suppose that two functions, f and g, have the same expansion coefficients, that is,

$$c_i = (f_i, f) = (f_i, g) .$$

Thus, $(f_i, f - g) = 0$, so by Theorem 5.2, $f - g = 0$, and hence $f = g$.

Now consider the converse problem. Does a given function have a unique set of expansion coefficients? Assume that

$$\lim_{n\to\infty} ||f - \sum_{i=1}^{n} c_i f_i|| = \lim_{n\to\infty} ||f - \sum_{i=1}^{n} d_i f_i|| = 0 ;$$

that is, assume that there are two partial sums with different expansion coefficients that converge in the mean to the same function f. If the expansion coefficients are unique, then $c_i = d_i$. To prove this, we observe that

$$\begin{aligned} \Big|\Big|\sum_{i=1}^{n} c_i f_i - \sum_{i=1}^{n} d_i f_i\Big|\Big| &= \Big|\Big|\sum_i c_i f_i - f + f - \sum_i d_i f_i\Big|\Big| \\ &\leq \Big|\Big|f - \sum_i c_i f_i\Big|\Big| + \Big|\Big|f - \sum_i d_i f_i\Big|\Big| , \end{aligned}$$

where we have used the triangle inequality (see Problem 4.7c). Now given any ϵ, we can choose by assumption an n large enough so that both the last two norms are less than $\epsilon/2$. Therefore, for such an n,

$$\Big|\Big|\sum_i c_i f_i - \sum_i d_i f_i\Big|\Big| = \Big|\Big|\sum_i (c_i - d_i) f_i\Big|\Big| = \Big[\sum_i (c_i - d_i)^2\Big]^{1/2} < \epsilon .$$

But this can only be true if $c_i = d_i$. Therefore the expansion coefficients of a given function are unique. Since the set $\{f_i\}$ is a complete orthonormal set of functions, it follows from our earlier remarks that $c_i = d_i = (f_i, f)$, the Fourier coefficients.

5.3 THE DIRAC δ-FUNCTION

In this section we introduce Dirac's δ-function in an informal way. The δ-function will play an important role in this book as it usually does in physics.

The first thing to understand about the δ-function is that it is not a function at all. A function is a rule that assigns another number to each number in a set of numbers. The δ-function, as used in physics, is instead a shorthand notation for a rather complicated limiting process whose use greatly simplifies calculations. It takes on a meaning only when it appears under an integral sign, in which case it has the following effect:

$$\int_{-\infty}^{\infty} f(x)\delta(x)\, dx = f(0) . \tag{5.12}$$

A special case of this is $f(x) = 1$. Then

$$\int_{-\infty}^{\infty} \delta(x)\, dx = 1 . \tag{5.13}$$

If the singular point is located at an arbitrary point x, then

$$\int_{-\infty}^{\infty} f(x')\delta(x' - x)\, dx' = f(x) . \tag{5.14}$$

Except at the singular point $x = 0$,

$$\delta(x) = 0 . \tag{5.15}$$

Thus $\delta(x)$ behaves as an ordinary function almost everywhere. It vanishes at all points where its argument is not zero, but at that one point it is undefined. Nevertheless its behavior near this point is all that matters.

Now the integral of any real function which vanishes everywhere except at one point must be zero, regardless of the value of the function at the singular point. Thus no *function* which satisfies Eq. (5.15) could possibly satisfy Eqs. (5.12) or (5.13). These equations must be interpreted as a symbolic notation for a process of the following type.

Let $\delta_\alpha(x)$ be a set of functions parametrized by the index α, which have the properties

$$\begin{aligned} &\lim_{\alpha\to 0} \delta_\alpha(x) = 0 \qquad \text{for all } x \neq 0 , \\ &\lim_{\alpha\to 0} \int_{-\infty}^{+\infty} f(x)\delta_\alpha(x)\, dx = f(0) . \end{aligned} \tag{5.16}$$

Our earlier equations result if we denote "$\lim_{\alpha\to 0} \delta_\alpha(x)$" by $\delta(x)$, and interchange the order of the limiting process and the integration, a procedure which is not, in general, valid. The original equations defining the δ-function must be interpreted as standing for the limiting processes of Eqs. (5.16).

Let us look at several sets of functions that have the properties described in Eqs. (5.16).

1. The simplest possible set of functions which has the proper limiting behavior is depicted in Fig. 5.1(a). The function $\delta_c(x)$ is defined (for $c>0$) by

$$\delta_c(x) \equiv \begin{cases} \dfrac{1}{c} & \text{for } |x| \leq \dfrac{c}{2} , \\ 0 & \text{for } |x| > \dfrac{c}{2} . \end{cases} \tag{5.17}$$

Clearly, $\lim_{c\to 0} \delta_c(x) = 0$, at all $x \neq 0$. Also, $\int_{-\infty}^{+\infty} \delta_c(x)\, dx = 1$, independent of c. The function $\delta_c(x)$ (and it *is* a function) is defined for all $c \neq 0$, and the limit

$$\lim_{c\to 0} \int_{-\infty}^{\infty} \delta_c(x)\, dx \tag{5.18}$$

is defined and equals 1. Also,

$$\lim_{c\to 0} \int_{-\infty}^{\infty} f(x)\delta_c(x)\, dx = f(0) , \tag{5.19}$$

which may be shown formally for continuous functions $f(x)$ as follows:

$$\lim_{c\to 0} \int_{-\infty}^{\infty} f(x)\delta_c(x)\, dx = \lim_{c\to 0} \int_{-c/2}^{c/2} f(x)\delta_c(x)\, dx = \lim_{c\to 0} \frac{1}{c} \int_{-c/2}^{c/2} f(x)\, dx ;$$

now, by the mean value theorem of integral calculus,

$$\int_{-c/2}^{c/2} f(x)dx = f(\xi c) \int_{-c/2}^{c/2} dx = cf(\xi c) ,$$

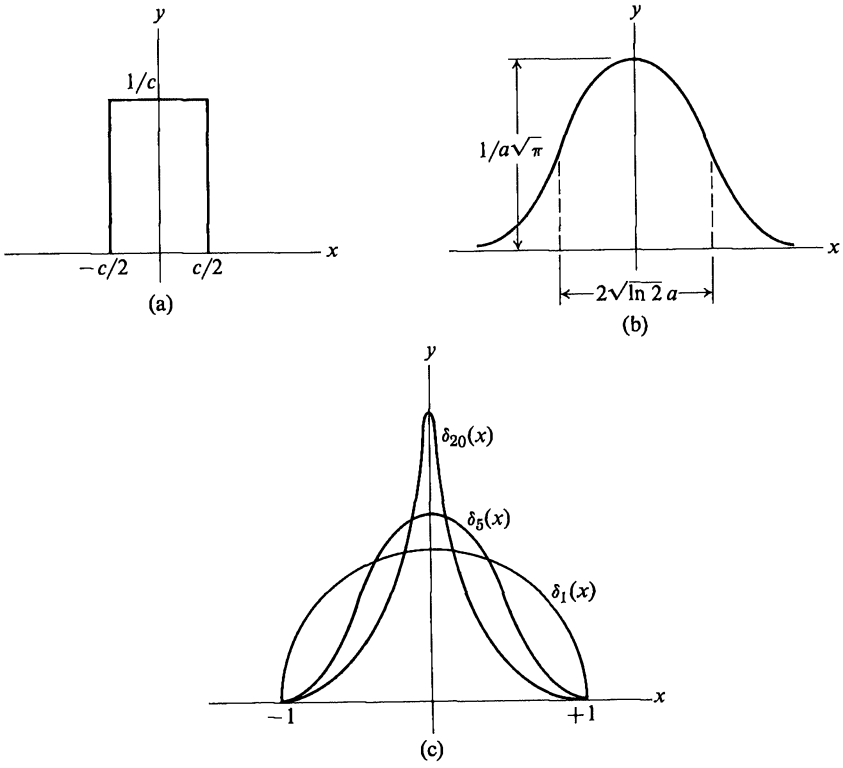

Fig. 5.1. Representations of the δ-function.

where $-1/2 < \xi < 1/2$. Letting $c \to 0$, we obtain

$$\lim_{c\to 0} \int_{-\infty}^{\infty} f(x)\delta_c(x)\, dx = f(0) .$$

2. The sequence of Gaussian distribution functions

$$\delta_a(x) \equiv \frac{1}{a\sqrt{\pi}} e^{-x^2/a^2}$$

provides another representation of the δ-function (see Fig. 5.1b). Note that

$$\lim_{a\to 0} \delta_a(x) = 0 \qquad \text{for all } x \neq 0 ;$$

$$\int_{-\infty}^{\infty} \delta_a(x)\, dx = 1, \qquad \text{independent of } a ;$$

$$\lim_{a\to 0} \int_{-\infty}^{\infty} f(x)\delta_a(x)\, dx = f(0) .$$

The entire contribution to the integral, as $a \to 0$, comes from the neighborhood of $x = 0$. Therefore we may write symbolically,

$$\delta(x) = \lim_{a\to 0} \delta_a(x) = \lim_{a\to 0} \frac{1}{a\sqrt{\pi}} e^{-x^2/a^2}. \tag{5.20}$$

3. Another useful representation for the δ-function is

$$\delta(x) = \lim_{\epsilon \to 0} \delta_\epsilon(x) \equiv \lim_{\epsilon \to 0} \frac{1}{\pi} \frac{\epsilon}{x^2 + \epsilon^2}, \tag{5.21}$$

which the reader can establish as in the above example.

4. The final representation of the δ-function is slightly different from the preceding three. It will play a central role in the proof of Weierstrass's theorem, the basic result of this chapter. It is defined as

$$\delta_n(x) = \begin{cases} c_n(1 - x^2)^n & \text{for } 0 \leq |x| \leq 1, \qquad n = 1, 2, 3, \cdots, \\ 0 & \text{for } |x| > 1, \end{cases} \tag{5.22}$$

where the constant c_n must be determined so that

$$\int_{-1}^{1} \delta_n(x)\, dx = 1 . \tag{5.23}$$

The functions $\delta_n(x)$ form a sequence whose limit is a δ-function (see Fig. 5.1c). We shall show, first informally, then rigorously in the proof of Weierstrass's theorem, that

$$\lim_{n \to \infty} \int_{-1}^{1} f(x)\delta_n(x)\, dx = f(0) . \tag{5.24}$$

Thus

$$\lim_{n \to \infty} \delta_n(x) = \delta(x) . \tag{5.25}$$

This representation of the δ-function differs from the others in that the defining parameter n takes on integral values that increase to infinity, instead of decreasing continuously to zero.

First we determine the normalization constant c_n. We have

$$1/c_n = \int_{-1}^{1} (1 - x^2)^n\, dx = 2\int_{0}^{1} (1 - x^2)^n\, dx . \tag{5.26}$$

Making the change of variable $x = \sin\theta$, we obtain

$$\frac{1}{c_n} = 2\int_0^{\pi/2} \cos^{2n+1}\theta\, d\theta = \frac{2^{n+1} n!}{1 \cdot 3 \cdot 5 \cdots (2n + 1)} . \tag{5.27}$$

From Eq. (5.27), it follows that

$$c_n = (2n + 1)!/2^{2n+1}(n!)^2 . \tag{5.28}$$

It is not clear from this expression for c_n how it behaves as $n \to \infty$. We can estimate its behavior from Eq. (5.26). Picking up again from there, we have

$$1/c_n = 2\int_0^1 (1 - x^2)^n\, dx \geq 2\int_0^{1/\sqrt{n}} (1 - x^2)^n\, dx , \tag{5.29}$$

since $1/\sqrt{n} \leq 1$ for all $n = 1, 2, \cdots$, and since the integrand is positive throughout [0, 1]. Now we shall show that for all n, the integrand $(1 - x^2)^n \geq 1 - nx^2$

for all x in $[0, 1]$. Consider the function $g(x) \equiv (1 - x^2)^n - (1 - nx^2)$. Since $g(0) = 0$, and

$$g'(x) = 2nx[1 - (1 - x^2)^{n-1}] > 0 \qquad \text{for all } 0 < x \leq 1 ,$$

$g(x)$ must be monotonically increasing in the interval $[0, 1]$. Therefore $g(x) \geq 0$, or $(1 - x^2)^n \geq (1 - nx^2)$, for all x in $[0, 1]$. Using this inequality in Eq. (5.29), we have

$$1/c_n \geq 2 \int_0^{1/\sqrt{n}} (1 - nx^2)\, dx = 4/3n^{1/2} > 1/n^{1/2} .$$

Therefore

$$c_n < n^{1/2} . \tag{5.30}$$

This result may also be obtained from Stirling's formula.

As $n \to \infty$, the contribution to the integral $\int_{-1}^{1} \delta_n(x)\, dx$ comes increasingly from the neighborhood surrounding the origin. To see this, note that for $0 < \delta < 1$,

$$\int_{-1}^{-\delta} \delta_n(x)\, dx = \int_{\delta}^{1} \delta_n(x)\, dx , \tag{5.31}$$

since $\delta_n(x)$ is an even function of x. Now

$$\int_{\delta}^{1} \delta_n(x)\, dx < n^{1/2}(1 - \delta^2)^n(1 - \delta) < n^{1/2}(1 - \delta^2)^n , \tag{5.32}$$

since $c_n < n^{1/2}$ [Eq. (5.30)], and $(1 - x^2)^n$ takes on its maximum value at $x = \delta$, and therefore $\int_{\delta}^{1}(1 - x^2)^n\, dx$ is bounded by the area of a rectangle of height $(1 - \delta^2)^n$ and base $(1 - \delta)$, since $0 < \delta < 1$. It is well known that the nth power of any positive number less than 1 will decrease more rapidly with n than any power of n will increase. In particular, the behavior of the term $(1 - \delta^2)^n$ will dominate the term $n^{1/2}$ as $n \to \infty$, and therefore

$$\lim_{n\to\infty} \int_{\delta}^{1} \delta_n(x)\, dx = 0 .$$

Since $\delta_n(x)$ is continuous and never negative, it follows that $\lim_{n\to\infty} \delta_n(x) = 0$ for $0 < x \leqq 1$. Since we have already arranged that $\int_{-1}^{1} \delta_n(x)\, dx = 1$ by our choice of c_n (Eq. 5.28), we obtain the result

$$\lim_{n\to\infty} \int_{-1}^{1} f(x)\delta_n(x)\, dx = f(0) . \tag{5.33}$$

5.4 WEIERSTRASS'S THEOREM: APPROXIMATION BY POLYNOMIALS

Weierstrass's famous approximation theorem proves that one can construct from the set of powers of x a sequence of polynomials which converges *uniformly* to any function that is continuous on the finite closed interval $[a, b]$. From this

result, we can prove that there exists a complete orthonormal set of polynomials on any interval $[a, b]$. Weierstrass's theorem is the starting point for proving the completeness properties of many of the functions of mathematical physics, such as the Legendre polynomials, the trigonometric functions (Fourier series), and the spherical harmonics.

Weierstrass's Theorem. If $f(x)$ is continuous on the closed interval $[a, b]$, there exists a sequence of polynomials $P_n(x)$ such that

$$\lim_{n\to\infty} P_n(x) = f(x)$$

uniformly on $[a, b]$.

Proof. We may assume without loss of generality that $f(x)$ is defined on $[0, 1]$. Suppose that $f(x)$ were defined on $[a, b]$. Then consider the function h defined by

$$h\left(\frac{x-a}{b-a}\right) \equiv f(x) .$$

Clearly, $f(a) = h(0)$, $f(b) = h(1)$, and any x in the interval $[a, b]$ corresponds to a z in $[0, 1]$. Thus if $h(z)$ can be approximated by polynomials in z, then since any polynomial in $z = (x-a)/(b-a)$ is also a polynomial in x, we can go from the polynomial approximating h to a polynomial approximating f. Furthermore, we can assume that $h(z)$ vanishes at $z = 0$ and $z = 1$, for if it does not, we define

$$g(z) = h(z) - h(0) - z[h(1) - h(0)]$$

for z in $[0, 1]$. Clearly, $g(0) = 0$ and $g(1) = 0$. Since $g(z)$ and $h(z)$ differ only by a polynomial, if we can approximate $g(z)$ by a polynomial, we can approximate $h(z)$ by that same polynomial plus the polynomial $h(0) + [h(1) - h(0)]z$.

Therefore we assume our original function $f(x)$ to be defined on $[0, 1]$ and to vanish at $x = 0$ and $x = 1$. We may define $f(x)$ as we choose outside $[0, 1]$, and we define it to be identically zero there.

Now we set

$$P_n(x) = \int_{-1}^{1} f(x + t)\delta_n(t)\, dt, \qquad 0 \le x \le 1 , \tag{5.34}$$

where $\delta_n(t)$ is the sequence of functions, defined in Eq. (5.22), that we claimed represented a δ-function:

$$\delta_n(t) = \begin{cases} c_n(1 - t^2)^n & \text{for } 0 \le |t| \le 1 , \\ 0 & \text{for } |t| > 1 . \end{cases}$$

If that claim were true, then

$$\lim_{n\to\infty} P_n(x) = f(x) ,$$

and the proof would be complete. We mention this just to indicate the motivation behind the rigorous proof we now present. This proof will also demon-

strate that the sequence $\delta_n(t)$ as $n \to \infty$ really does have the properties that characterize a δ-function.

We are assuming that $f(x)$ vanishes outside $[0, 1]$. We may therefore write Eq. (5.34) as

$$P_n(x) = \int_{-x}^{1-x} f(x + t)\delta_n(t)\, dt\,,$$

where $f(x + t) \equiv 0$ whenever $t \leqq -x$ or $t \geqq 1 - x$. By a simple change of variable $(t \to t - x)$, we obtain

$$P_n(x) = \int_0^1 f(t)\delta_n(t - x)\, dt = \int_0^1 f(t)c_n[1 - (t - x)^2]^n\, dt\,.$$

This last integral shows clearly that $P_n(x)$ *is a polynomial* (of degree $2n$) in x. The coefficients of the powers of x are definite integrals over t. Thus $P_n(x)$ is a sequence of polynomials. We shall prove that this sequence converges uniformly to $f(x)$.

In analysis it is shown that a function which is continuous on a finite *closed* interval is uniformly continuous there. For those unfamiliar with this result, its meaning may be abstracted from the following example. The function $q(x) = x^{-1}$ is continuous on the *open* interval $(0, 1)$, but is not uniformly continuous there because

$$|q(x + \delta) - q(x)| = \frac{\delta}{x(x + \delta)}\,,$$

and this difference *cannot* be made arbitrarily small *for all* x by a *single* choice of δ. In fact, by choosing x sufficiently close to zero—yet still in $(0, 1)$—the difference can be made arbitrarily large for any given $\delta > 0$. Thus there does not exist a $\delta > 0$ *independent of* x such that $|q(x + \delta) - q(x)|$ is arbitrarily small, and the continuity is not uniform.

Since $f(x)$ is continuous on the closed interval $[0, 1]$, it is uniformly continuous there. As will be seen, *uniform* continuity is essential to this proof. Therefore we know that given any $\epsilon > 0$, there exists a δ such that

$$|f(x + \delta) - f(x)| < \epsilon$$

for all x in $[0, 1]$.

Now, using Eq. (5.34) for $P_n(x)$, we form the quantity

$$\begin{aligned} |P_n(x) - f(x)| &= \left|\int_{-1}^{1} [f(x + t) - f(x)]\delta_n(t)\, dt\right| \\ &\leq \int_{-1}^{1} |f(x + t) - f(x)|\delta_n(t)\, dt\,, \end{aligned}$$

since $\delta_n(t) \geq 0$ for all t in $[-1, 1]$. We now break up the range of integration into three parts:

$$\int_{-1}^{1} |f(x + t) - f(x)|\delta_n(t)\, dt = \int_{-1}^{-\delta} + \int_{-\delta}^{\delta} + \int_{\delta}^{1}.$$

Since $f(x)$ is continuous on a closed interval, it is bounded there. Let the maximum value of $|f(x)| = M$. Then both $\int_{-1}^{-\delta}$ and $\int_{\delta}^{1}$ are bounded by the quantity $2Mn^{1/2}(1-\delta^2)^n$. We may see this as follows:

$$\int_{\delta}^{1} |f(x+t) - f(x)|\delta_n(t)dt \leq \int_{\delta}^{1} |f(x+t)|\delta_n(t)\,dt + \int_{\delta}^{1} |f(x)|\delta_n(t)\,dt$$

$$\leq 2M\int_{\delta}^{1} \delta_n(t)\,dt < 2Mn^{1/2}(1-\delta^2)^n\,.$$

Here we have used the fact that $|f| < M$ and Eqs. (5.31) and (5.32).

We may estimate the remaining integral, $\int_{-\delta}^{\delta}$, by using the uniform continuity of $f(x)$, and interpreting the limits of integration as a δ which guarantees

$$|f(x+t) - f(x)| < \epsilon/2, \qquad \text{for } |t| < \delta.$$

We find

$$\int_{-\delta}^{\delta} |f(x+t) - f(x)|\delta_n(t)\,dt < \epsilon/2\int_{-\delta}^{\delta} \delta_n(t)\,dt < \epsilon/2\,,$$

since $\int_{-\delta}^{\delta} \delta_n(t)\,dt < 1$.

Collecting these results, we have

$$|P_n(x) - f(x)| < 4Mn^{1/2}(1-\delta^2)^n + \epsilon/2\,.$$

The value of $n^{1/2}(1-\delta^2)^n$ for $0 < \delta < 1$ can be made arbitrarily small for large enough n and, in particular, smaller than $\epsilon/2$. Therefore there exists an N such that for $n > N$,

$$|P_n(x) - f(x)| < \epsilon\,,$$

for any arbitrarily small preassigned ϵ, that is,

$$\lim_{n\to\infty} |P_n(x) - f(x)| = 0.$$

This means that the sequence of polynomials $P_n(x)$ converges uniformly to the continuous function $f(x)$ on $[0, 1]$, and so the proof is complete. In fact, this holds for an arbitrary continuous function on an arbitrary finite closed interval $[a, b]$, as was demonstrated at the outset. QED

Thus Weierstrass's theorem tells us that there exists a set of coefficients

$$a_{nm}; \qquad n = 0, \cdots, \infty, \quad m = 0, \cdots, \mu = 2n\,,$$

such that $\sum_{m=0}^{m=\mu} a_{nm}\,x^m$ tends uniformly to $f(x)$ as $n \to \infty$. It most emphatically does not guarantee a uniformly convergent power series, which would consist of a set of coefficients a_m $(m = 0, \cdots, \infty)$ such that $\sum_{m=0}^{m=n} a_m x^m$ tends uniformly to $f(x)$ as $n \to \infty$. The point is that the Weierstrass coefficients a_{nm} are not independent of n for fixed m. As the approximation improves, by going to polynomials of higher degree, the earlier coefficients change. The theorem would be false if it claimed to produce a uniformly convergent power series. For example,

there is no *power* series that converges uniformly to the continuous function $\sqrt{x}$ in the interval [0, 1].

Weierstrass's theorem for approximation by a sequence of polynomials is in one sense, much stronger than Taylor's theorem for expansion in power series. To expand a function in a Taylor series, its derivatives of all orders must exist: it must be "analytic." In Weierstrass's theorem, only continuity is needed. Furthermore, Weierstrass's theorem demonstrates the existence of polynomial approximations outside the radius of convergence of a Taylor series. But there is, in general, no possibility of rearranging the uniformly convergent sequence of *polynomials* that approximate any continuous function so as to produce a convergent *power* series (Taylor expansion).

In Taylor expansions we need to know the function and its derivatives *locally*, at a point; the radius of convergence of the expansion may be finite or infinite. Weierstrass's theorem applies only to finite intervals, and we need to know the function, but not its derivatives, *globally*, over the entire interval.

As a by-product of this theorem, we have established rigorously that the representation of Dirac's delta function given in Eq. (5.22) has the properties claimed for it [Eq. (5.24)]. Thus Weierstrass, who proved this theorem in 1885 by the above method, anticipated Dirac considerably. In fact, he was not the only one. Heaviside used a closely related symbolic function before Dirac. Yet it was with Dirac's introduction of symbolic functions in his classic book, *The Principles of Quantum Mechanics*, which appeared in 1930, that their use became widespread. Only relatively recently was the theory of δ-functions, and other related symbolic functions, established rigorously by the mathematician Laurent Schwartz in his theory of distributions.

Weierstrass's theorem may also be extended to functions of several variables. By a straightforward generalization of the proof, it can be shown that if a function $f(x_1, x_2, \cdots, x_m)$ is continuous in each variable x_i for x_i in $[a_i, b_i]$ $(i = 1, 2, \cdots, m)$, f may be approximated uniformly by the polynomials

$$P_n(x_1, x_2, \cdots, x_m) = \int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_m}^{b_m} f(t_1, t_2, \cdots, t_m)\delta_n(t_1 - x_1)$$
$$\delta_n(t_2 - x_2)\cdots\delta_n(t_m - x_m)\, dt_1\, dt_2\cdots dt_m . \qquad (5.35)$$

The completeness of the trigonometric functions follows from the special case $m = 2$, as we shall see in Section 5.6.

We close this section with the statement of an important consequence of our basic result. From Section 5.2 it follows that if the set P_n defined in the proof of Weierstrass's theorem approximate any continuous function f uniformly, they also approximate any continuous function *in the mean*; that is, given any ϵ, there exists an n such that $||f - P_n|| < \epsilon$. Now it can be shown that any function in the Hilbert space of square-integrable functions can be approximated arbitrarily closely in the mean by a continuous function. This is discussed further in Section 8.7. It follows that *any* function in this Hilbert space can be approximated *in the mean* by some P_n. This can be seen as follows.

Let ϕ be a function in Hilbert space. Then the difference $\phi - P_n$ can be written as

$$\phi - P_n = (\phi - f) + (f - P_n) ,$$

where f is a continuous function. By the triangle inequality, we have

$$||\phi - P_n|| \leq ||\phi - f|| + ||f - P_n|| .$$

Now f can be chosen so that $||\phi - f||$ is as small as we please; and by Weierstrass's theorem, n can be chosen so that $||f - P_n||$ is as small as we please. Thus P_n approximates ϕ, an arbitrary function in Hilbert space, arbitrarily closely in the mean. This fact will be needed in what follows. A complete proof can be found in most analysis texts, e.g., Riesz and Nagy (Section 46) or Rudin (Theorem 10.38).

5.5 LEGENDRE POLYNOMIALS

We are now in a position to demonstrate that there exists a *complete* orthonormal set of polynomials on the finite closed interval $[a, b]$. The proof of completeness follows from Weierstrass's theorem. After establishing the general result, we shall examine in detail one very important special case—the complete orthonormal set of polynomials on the interval $[-1, 1]$, which, apart from constant normalization factors, is the set of Legendre polynomials. We shall construct the first three functions in this set explicitly, and then provide a formula for calculating all the others.

Weierstrass's theorem tells us that any continuous function f on $[a, b]$ can be approximated uniformly by a sequence of polynomials:

$$P_n(x) = \sum_{m=0}^{2n} a_{nm} x^m .$$

Since the set $\{x^n, n = 0, 1, \cdots\}$ is linearly independent, we may apply the Gram-Schmidt orthogonalization process to construct from it an orthonormal basis $\{Q_n(x)\}$, where $Q_n(x)$ is a polynomial of degree n. We may now express the original functions as finite linear combinations of the orthonormal set:

$$x^m = \sum_{i=0}^{m} c_{mi} Q_i(x) .$$

It follows that the polynomials P_n, which approximate the function f uniformly, may be expressed as

$$P_n(x) = \sum_{m=0}^{2n} a_{nm} \sum_{i=0}^{m} c_{mi} Q_i(x) .$$

We now prove that the orthonormal set $\{Q_n\}$ is complete by showing that it is closed (Theorem 5.2). The orthonormal set $\{Q_n\}$ will be closed if the only function orthogonal to all the Q_n is the zero function. Assume that we have a function in Hilbert space such that $(f, Q_n) = 0$ for all n. It follows immediately from the above equation that $(f, P_n) = 0$ for all n, since the P_n are linear combinations of the Q_n. However, we know from Weierstrass's theorem that f may be approximated in the mean by the P_n. Thus for any ϵ there exists an n such that

$||f - P_n|| < \epsilon$. But since $(f, P_n) = 0$,

$$||f - P_n|| \equiv (f - P_n, f - P_n)^{1/2} = [||f||^2 + ||P_n||^2]^{1/2} .$$

Therefore

$$||f||^2 + ||P_n||^2 < \epsilon^2 .$$

But this implies that $||f||$ is arbitrarily small, and hence f is equal to zero (almost everywhere). Therefore the orthonormal set $\{Q_n\}$ is closed and hence, by Theorem 5.2, it is a complete orthonormal set of polynomials on $[a, b]$.

The completeness of the set Q_i means (according to Definition 5.6) that there exists a set of of constants $\{a_i\}$ such that any function g in the Hilbert space can be approximated in the mean by the sequence of partial sums

$$g_n = \sum_{i=0}^{n} a_i Q_i(x) .$$

But the a_i are independent of n. Thus as we extend the sum to infinity, the approximation improves *without the earlier a_i changing*. Therefore we may say that there exists an infinite *series*

$$\sum_{i=0}^{\infty} a_i Q_i$$

which approximates g in the mean. Symbolically, we write

$$g(x) \doteq \sum_{i=0}^{\infty} a_i Q_i(x) .$$

We have shown that if we have a complete orthonormal set of functions, the coefficients which provide the best approximation in the mean are the Fourier coefficients. Thus the expansion coefficients in the infinite series are given by

$$a_i = (Q_i, g) .$$

Let us now apply these considerations to a concrete case. We shall use the Gram-Schmidt orthogonalization process applied to the basis $\{1, x, x^2, \cdots\}$ to construct the set of orthonormal polynomials on the interval $[-1, 1]$. By the previous result we know that this set will be complete.

The first member of the orthonormal set is $\bar{P}_0(x) = (\frac{1}{2})^{1/2}$, which is evidently normalized for x in $[-1, 1]$. Using the prescription of Section 4.3, we have

$$\begin{aligned} \bar{P}_0(x) &= (\tfrac{1}{2})^{1/2} , \\ \bar{P}_1 &= \frac{x - 1/2^{1/2} \int_{-1}^{1} 1/2^{1/2} x \, dx}{\left\| x - 1/2^{1/2} \int_{-1}^{1} 1/2^{1/2} x \, dx \right\|} = \left(\frac{3}{2}\right)^{1/2} x , \\ \bar{P}_2 &= \frac{x^2 - 1/2^{1/2} \int_{-1}^{1} 1/2^{1/2} x^2 \, dx - (\frac{3}{2})^{1/2} x \int_{-1}^{1} (\frac{3}{2})^{1/2} x^3 \, dx}{\left\| x^2 - 1/2^{1/2} \int_{-1}^{1} 1/2^{1/2} x^2 \, dx - (\frac{3}{2})^{1/2} x \int_{-1}^{1} (\frac{3}{2})^{1/2} x^3 \, dx \right\|} = \left(\frac{5}{2}\right)^{1/2} \left(\frac{3}{2} x^2 - \frac{1}{2}\right) . \end{aligned} \tag{5.36}$$

This is a tedious process; we have carried it out for these three cases only to show that one can construct orthonormal functions directly. One would hope that there is a better way to compute the orthonormal polynomials of higher order, and, indeed, there is.

We claim that a general formula for the complete orthonormal set of polynomials on $[-1, 1]$ is

$$\bar{P}_n(x) = \left(\frac{2n+1}{2}\right)^{1/2} \frac{1}{2^n n!} \frac{d^n}{dx^n}(x^2 - 1)^n . \tag{5.37}$$

The orthogonal, but unnormalized, polynomials

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n}(x^2 - 1)^n \tag{5.38}$$

are known as the *Legendre polynomials*. Equation (5.38) is called *Rodrigues' formula*.

Thus the Legendre polynomials differ from the orthonormal set $\{\bar{P}_n(x)\}$ by constant multiplicative factors. The problem now is to show that the Rodrigues formula gives the *same* set of orthonormal polynomials as the Gram-Schmidt orthogonalization process. Direct computation verifies that they agree through the polynomials of second degree. To show that the Rodrigues formula holds for all orders, we must show that:

1. On any fixed interval there can be only one complete orthonormal set of polynomials in which the nth polynomial is of degree n.
2. The $\bar{P}_n(x)$ given by the Rodrigues formula do in fact form a complete orthonormal set on $[-1, 1]$.

The uniqueness property of the complete orthonormal set of polynomials is embodied in the Gram-Schmidt process, which can generate one and only one complete orthonormal set of polynomials (up to a phase factor) on $[-1, 1]$, or, in fact, on any given interval. Hence we have, in fact, already proved condition (1).

To demonstrate condition (2), we now show that the set of polynomials generated by the Rodrigues formula for the interval $[-1, 1]$ is orthonormal, and that the function $P_n(x)$ given by Rodrigues' formula is a polynomial of degree n. Then completeness follows by our earlier proof. It is immediately clear that $P_n(x)$ is a polynomial of degree n. We prove orthonormality in the following theorem.

Theorem 5.3.

$$(\bar{P}_n, \bar{P}_m) = \int_{-1}^{1} \bar{P}_n(x)\bar{P}_m(x)\, dx = \delta_{nm} . \tag{5.39}$$

Proof. We first prove orthogonality. We denote d^n/dx^n by d^n, and suppose that $n > m$. Dropping constant factors, where we have integrated by parts,

we have

$$\int_{-1}^{1} \bar{P}_n \bar{P}_m \, dx = \int_{-1}^{1} [d^n(x^2 - 1)^n]\,[d^m(x^2 - 1)^m]\, dx$$

$$= [d^{n-1}(x^2 - 1)^n]\,[d^m(x^2 - 1)^m]\Big]_{-1}^{1}$$

$$- \int_{-1}^{1} [d^{n-1}(x^2 - 1)^n]\,[d^{m+1}(x^2 - 1)^m \, dx] ,$$

Since

$$d^{n-1}(x^2 - 1)^n = (\text{a polynomial}) \cdot (x^2 - 1),$$

the first term vanishes upon putting in the limits ± 1, leaving the second term. Therefore, after n partial integrations, we have

$$\int_{-1}^{1} (-1)^n (x^2 - 1)^n d^{m+n}(x^2 - 1)^m \, dx ,$$

since the term which is evaluated at ± 1 always vanishes because it is proportional to some power of $(x^2 - 1)$. Now, since $n > m$, $n + m > 2m$ and so $d^{n+m}(x^2 - 1)^m = 0$. Therefore

$$\int_{-1}^{1} \bar{P}_n \bar{P}_m \, dx = 0 \qquad \text{for } m \neq n .$$

If $m = n$, then as before (but putting in constant factors),

$$\int_{-1}^{1} \bar{P}_n \bar{P}_n \, dx = \frac{(2n + 1)(-1)^n}{2^{2n+1}(n!)^2} \int_{-1}^{1} (x^2 - 1)^n d^{2n}(x^2 - 1)^n \, dx .$$

But $(x^2 - 1)^n$ is a polynomial of degree $2n$, so the $(2n)^{\text{th}}$ derivative is just $(2n)!$. The integral becomes

$$\frac{(-1)^n (2n + 1)!}{2^{2n+1}(n!)^2} \int_{-1}^{1} (x^2 - 1)^n \, dx = 1 ,$$

where we have used Eqs. (5.26) and (5.28).

The functions $\bar{P}_n(x) = [(2n + 1)^{1/2}/2]\, P_n(x)$ are therefore a complete orthonormal set on the interval $[-1, 1]$, and they are the only such set.

We shall now work out several other results pertaining to the Legendre polynomials $P_n(x)$. Our approach will be inductive and straightforward, a continuation of the techniques used to prove orthonormality. There are more elegant ways to obtain most of these results which will be used later to deal with both the Legendre polynomials and other complete sets of functions. But for the moment we proceed in a more pedestrian manner.

First we shall derive from the Rodrigues formula an explicit expression for $P_n(x)$ as a polynomial. We apply the binomial theorem to the factor $(x^2 - 1)^n$ to obtain

$$(x^2 - 1)^n = \sum_{m=0}^{n} \binom{n}{m} x^{2m} (-1)^{n-m} ,$$

where $\binom{n}{m}$ is the binomial coefficient:

$$\binom{n}{m} \equiv \frac{n!}{m!(n-m)!} . \tag{5.40}$$

Then, taking the nth derivative, we obtain

$$P_n(x) = \frac{1}{2^n n!} d^n(x^2-1)^n = \frac{1}{2^n n!} \sum_{m \geq p}^{n} (-1)^{n-m} \binom{n}{m} \frac{(2m)!}{(2m-n)!} x^{2m-n} , \tag{5.41}$$

where $p = n/2$ if n is even and $p = (n+1)/2$ if n is odd.

We next prove that $P_n(1) = 1$ for all n. In itself this is not of staggering importance, although it is interesting. We prove it now to illustrate an important technique. (We shall have an easier way later.) We write

$$P_n(x) = \frac{1}{2^n n!} d^n[(x-1)^n (x+1)^n] ,$$

and evaluate the nth derivative of the product by Leibnitz's rule (the "binomial expansion" for the nth derivative of a product):

$$d^n(uv) = \sum_{m=0}^{n} \binom{n}{m} d^m u d^{n-m} v = u d^n v + n d u d^{n-1} v + \frac{n(n-1)}{2!} d^2 u d^{n-2} v + \cdots + \frac{n!}{k!(n-k)!} d^k u d^{n-k} v + \cdots + v d^n u. \tag{5.42}$$

We now take $u = (x-1)^n$ and $v = (x+1)^n$. It is clear that the only nonzero term in the Leibnitz expansion at $x = 1$ is the term $m = n$, in which the factor $u = (x-1)^n$ is "differentiated down" to the constant $n!$ and hence does not vanish at $x = 1$. Therefore

$$d^n[(x-1)^n (x+1)^n]\Big|_{x=1} = d^n(x-1)^n d^0 (x+1)^n \Big|_{x=1} = n! 2^n ,$$

and consequently,

$$P_n(1) = 1 \qquad \text{for all } n . \tag{5.43}$$

The Legendre polynomials arose originally as the solution to a differential equation, now called Legendre's equation:

$$(1-x^2) P_n''(x) - 2x P_n'(x) + n(n+1) P_n(x) = 0 . \tag{5.44}$$

To prove that the elements of the complete orthonormal set of polynomials on $[-1, 1]$ satisfy this equation, we begin with the identity

$$(x^2-1) \frac{d}{dx} (x^2-1)^n = 2nx(x^2-1)^n .$$

Now differentiate both sides of this identity $n+1$ times. With the help of

Leibnitz's formula (Eq. 5.42), we obtain for the left-hand side

$$d^{n+1}[(x^2-1)d(x^2-1)^n] = (x^2-1)d^{n+2}(x^2-1)^n + 2(n+1)xd^{n+1}(x^2-1)^n + n(n+1)d^n(x^2-1)^n ,$$

and for the right-hand side, we have

$$d^{n+1}[2nx(x^2-1)^n] = 2nx\,d^{n+1}(x^2-1)^n + 2n(n+1)d^n(x^2-1)^n .$$

The difference of these two equations must vanish. Therefore

$$(x^2-1)d^{n+2}(x^2-1)^n + 2x\,d^{n+1}(x^2-1)^n - n(n+1)d^n(x^2-1)^n = 0 .$$

Now, using the Rodrigues formula for $P_n(x)$, we obtain

$$(x^2-1)P_n'' + 2x\,P_n' - n(n+1)P_n = 0 , \tag{5.45}$$

which is equivalent to Legendre's equation (5.44). It may also be written in the form

$$[(1-x^2)P_n'(x)]' + n(n+1)P_n(x) = 0 . \tag{5.46}$$

Thus the $P_n(x)$ are solutions of Legendre's equations; but they are not the only solutions. There exists another linearly independent solution called, collectively, the Legendre *functions* of the second kind, denoted by $Q_n(x)$. However, these functions are not finite at $x = \pm 1$ as are the $P_n(x)$, and so are excluded as possible solutions in many physical applications; they are not polynomials.

There is an alternative proof of the orthogonality of the Legendre polynomials which follows directly from the differential equation (5.46) for $P_n(x)$. We know that P_n and P_m $(n \neq m)$ satisfy the equations

$$\frac{d}{dx}[(1-x^2)P_n'] + n(n+1)P_n = 0 ,$$

$$\frac{d}{dx}[(1-x^2)P_m'] + m(m+1)P_m = 0 .$$

Multiplying these equations by P_m and P_n respectively, subtracting them, and integrating from -1 to $+1$, we obtain

$$\begin{aligned}(n-m)(n+m+1)&\int_{-1}^{1} P_nP_m\,dx \\ &= \int_{-1}^{1}\left[P_n\frac{d}{dx}((1-x^2)P_m') - P_m\frac{d}{dx}((1-x^2)P_n')\right]dx \\ &= \left\{\left[P_n(1-x^2)P_m'\right]_{-1}^{1} - \int_{-1}^{1} dx\,(1-x^2)P_m'P_n'\right. \\ &\quad \left. - \left[P_m(1-x^2)P_n'\right]_{-1}^{1} + \int_{-1}^{1} dx\,(1-x^2)P_n'P_m'\right\} = 0 ;\end{aligned}$$

here the two integrals cancel and the two other terms vanish. Since $n \neq m$, it

follows that

$$(P_n, P_m) \equiv \int_{-1}^{1} P_n P_m \, dx = 0 .$$

This is actually a special case of a proof that holds for a wide variety of functions arising as solutions to the Sturm-Liouville problem, which we shall discuss in Section 5.10. The proof depends on the fact that the differential equations obeyed by orthogonal functions have a certain form.

We have sampled only a few of the properties of the complete orthonormal set of polynomials on $[-1, 1]$. We shall encounter them many times in the course of the book (we met them once before in the multipole expansion (Eqs. 1.100, 1.101, 1.102), and shall discuss their properties in more detail. But our main concern for the moment is the *completeness* of this set of polynomials, and other sets of functions—not all the special properties these sets of functions happen to have.

We now turn to the single most important complete set of functions, the ancestor of them all: the trigonometric functions and Fourier series.

5.6 FOURIER SERIES

The completeness properties of the set of trigonometric functions $\{\sin n\theta, \cos n\theta, n = 0, 1, 2, \cdots, \infty\}$ can be deduced from Weierstrass's theorem for two variables. Equation 5.35 tells us that any function $g(x, y)$ which is continuous in both variables on specified finite closed intervals may be approximated uniformly by the the sequence of functions

$$g_N(x, y) = \sum_{n,m=0}^{N} A_{nm}^{(N)} x^n y^m ;$$

that is, $\lim_{N\to\infty} g_N(x, y) = g(x, y)$, uniformly in x and y. The coefficients $A_{nm}^{(N)}$ are not independent of N for fixed n and m, so this is *not* a power-series expansion.

If we change to polar coordinates and restrict the domain of definition to the unit circle, then $x = \cos\theta$ and $y = \sin\theta$ so that

$$g(\cos\theta, \sin\theta) = f(\theta) = \lim_{N\to\infty} \sum_{n,m=0}^{N} A_{nm}^{(N)} \cos^n\theta \sin^m\theta . \tag{5.47}$$

Clearly, the only functions $f(\theta)$ which can satisfy this last equation are periodic functions with periodicity 2π; this is a consequence of restricting x and y to the unit circle. We shall generalize later to include functions having periodicities different from 2π.

Using Euler's formula,

$$e^{i\theta} = \cos\theta + i \sin\theta ,$$

we obtain expressions for the nth powers of $\sin\theta$ and $\cos\theta$:

$$\cos^n\theta = \left[\frac{1}{2}\left(e^{i\theta} + e^{-i\theta}\right)\right]^n , \qquad \sin^n\theta = \left[\frac{1}{2i}\left(e^{i\theta} - e^{-i\theta}\right)\right]^n .$$

Then Eq. (5.47) can be rewritten in the form

$$f(x) = \lim_{M\to\infty} f_M(x) \equiv \lim_{M\to\infty} \sum_{n=-M}^{M} \frac{c_n^{(M)}}{(2\pi)^{1/2}} e^{inx} , \tag{5.48a}$$

where we have inserted the factor $(2\pi)^{1/2}$ for later convenience, and have replaced the variable θ by x to emphasize the general nature of the result. Clearly, we may write equivalently

$$f(x) = \lim_{M\to\infty} f_M(x) = \lim_{M\to\infty} \left[\frac{a_0^{(M)}}{2} + \sum_{n=1}^{M} (a_n^{(M)} \cos nx + b_n^{(M)} \sin nx) \right]. \tag{5.48b}$$

The complex expansion coefficients in the exponential form of the series, and the real coefficients in the trigonometric form, are related as follows:

$$\begin{aligned} a_0^{(M)} &= \left(\frac{2}{\pi}\right)^{1/2} c_0^{(M)} , & c_0^{(M)} &= \left(\frac{\pi}{2}\right)^{1/2} a_0^{(M)} , \\ a_n^{(M)} &= \frac{1}{(2\pi)^{1/2}} (c_n^{(M)} + c_{-n}^{(M)}) , \quad \text{or} & c_n^{(M)} &= \left(\frac{\pi}{2}\right)^{1/2} (a_n^{(M)} - ib_n^{(M)}) , \\ b_n^{(M)} &= \frac{i}{(2\pi)^{1/2}} (c_n^{(M)} - c_{-n}^{(M)}) , & c_{-n}^{(M)} &= \left(\frac{\pi}{2}\right)^{1/2} (a_n^{(M)} + ib_n^{(M)}) . \end{aligned} \tag{5.49}$$

The superscript M reminds us that the coefficients in the sequence are *not* independent of M; the coefficients of earlier terms in the sequence may change as M increases.

Thus any continuous function $f(x)$, for which $f(x) = f(x + 2\pi)$, can be approximated uniformly by a sequence of trigonometric polynomials:

$$f_M(x) = \sum_{n=-M}^{M} \frac{c_n^{(M)}}{(2\pi)^{1/2}} e^{inx} = \frac{a_0^{(M)}}{2} + \sum_{n=1}^{M} (a_n^{(M)} \cos nx + b_n^{(M)} \sin nx) .$$

Now suppose that $f(x)$ is any continuous function on the closed interval $[a, a + 2\pi]$, and assume further that $f(x)$ does not satisfy the periodicity condition $f(a) = f(a + 2\pi)$ at the endpoints of the interval. It is clear that for any such function f, we can find a continuous function $g(x)$ that both satisfies the periodicity requirement and is such that the quantity

$$\int_a^{a+2\pi} |f(x) - g(x)|^2 \, dx$$

can be made arbitrarily small. (See Courant-Hilbert for proof.) Therefore the trigonometric sequence that converges uniformly to $g(x)$ will converge *in the mean* to $f(x)$ on the interval $[a, a + 2\pi]$. In fact, since any function in Hilbert space can be approximated in the mean to arbitrary precision by some continuous function (see Section 5.4), this same result holds for any function in Hilbert space.

We saw in Section 5.1 that the set of functions

$$\left\{ \frac{1}{(2\pi)^{1/2}} e^{inx}, \; n = 0, \pm 1, \cdots \right\}$$

is orthonormal on the interval $[-\pi, \pi]$. We shall denote this orthonormal set of functions by $\{T_n\}$. We shall now show that this set is a *complete* orthonormal set. We do this, exactly as we did for the orthonormal set of polynomials $\{Q_n\}$ in the last section, by proving that the orthonormal set is closed. Thus, assuming that $(f, T_n) = 0$ for all n, we want to prove $f = 0$. It follows from Eq. (5.48a) that $(f, f_M) = 0$ for all M, because f_M is just a linear combination of the orthonormal funtions T_n. However, we know that any function f may be approximated in the mean by the sequence f_M; that is,

$$||f - f_M|| = [||f||^2 + ||f_M||^2]^{1/2} < \epsilon .$$

Thus $f = 0$ (almost everywhere). Therefore the orthonormal set $\{T_n\}$ is closed, and hence by Theorem 5.2, it is a complete orthonormal set on $[-\pi, \pi]$. Clearly, the set of trigonometric functions which is orthonormal on $[-\pi, \pi]$,

$$\left\{\frac{1}{(2\pi)^{1/2}}, \qquad \frac{1}{\pi^{1/2}} \sin nx, \qquad \frac{1}{\pi^{1/2}} \cos nx, \qquad n = 1, 2, \cdots\right\},$$

is also complete.

It follows from the completeness of the orthonormal set $\{T_n\}$ that we may approximate an arbitrary function f in the mean by an infinite series of the T_n. We write symbolically,

$$f(x) \doteq \sum_{n=-\infty}^{\infty} c_n T_n(x) = \sum_{n=-\infty}^{\infty} \frac{c_n}{(2\pi)^{1/2}} e^{inx} = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx). \quad (5.50)$$

The expansion coefficients which here are constants, independent of M (compare Eq. 5.48a), are given by

$$c_n = (T_n, f) = \frac{1}{(2\pi)^{1/2}} \int_{-\pi}^{\pi} f(x) e^{-inx}\, dx . \quad (5.51)$$

The coefficients a_n and b_n of the trigonometric series may be computed using Eqs. (5.49):

$$a_n = \frac{1}{(2\pi)^{1/2}} (c_n + c_{-n}) = \left(\frac{1}{(2\pi)^{1/2}}\right)^2 \int_{-\pi}^{\pi} f(x) (e^{-inx} + e^{inx})\, dx$$

$$= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx\, dx \qquad \text{for } n = 0, 1, \cdots \quad (5\ 52)$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx\, dx \qquad \text{for } n = 1, 2, \cdots$$

The series (5.50) with the coefficients (5.52) is known as the *Fourier series.*

Since the functions

$$T_n(x) = \frac{1}{(2\pi)^{1/2}} e^{inx}, \qquad n = 0, \pm 1 \cdots$$

form a complete orthonormal set, the completeness relation (Eq. 5.10) must

hold for it. This gives

$$(f,f) = \int_{-\pi}^{\pi} |f|^2\,dx = \sum_{n=-\infty}^{\infty} |c_n|^2 = \pi\left[\frac{a_0^2}{2} + \sum_{n=1}^{\infty} (a_n^2 + b_n^2)\right], \qquad (5.53)$$

where we have used Eqs. (5.49).

Let us now summarize the situation. We have shown that the Fourier series converges in the mean to any function $f(x)$ in Hilbert space on $[-\pi, \pi]$. We also know that there exists *some trigonometric sequence* that converges *uniformly* to any *continuous* function $f(x)$ on $[-\pi, \pi]$ for which $f(-\pi) = f(\pi)$. It is plausible therefore that the result for Fourier series, i. e., the particular trigonometric series with the Fourier coefficients [Eqs. (5.48)], can be strengthened. However, it most definitely is not guaranteed that the particular choice of coefficients which guarantees convergence in the mean, also yields uniform convergence. The converse is easily proved, however; namely, if a trigonometric series converges uniformly, then its coefficients are the Fourier coefficients. This may be demonstrated by multiplying [Eq. (5.50)] by cos mx or sin mx and integrating the uniformly convergent series term by term. But establishing the conditions for uniform convergence is the real problem. We begin our investigation of this question, which is the central question in the theory of Fourier series, by proving a theorem.

Theorem 5.4. The convergence of the Fourier series to $f(x)$ is uniform in the closed interval $[-\pi, \pi]$ if $f(x)$ is continuous and its derivative is piecewise continuous in this interval, and $f(-\pi) = f(\pi)$. If, in addition, $f(x + 2\pi) = f(x)$, the convergence will be uniform everywhere.

An example of a continuous function with a piecewise continuous derivative is given in Fig. 5.2. Such functions are often referred to as *piecewise smooth* functions.

Proof. We shall denote the Fourier coefficients of $f'(x)$ by a_n' and b_n'. Integrating by parts, we have

$$\begin{aligned} a_n' &\equiv \frac{1}{\pi}\int_{-\pi}^{\pi} f'(x)\cos nx\,dx = \frac{n}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx = nb_n\,, \\ b_n' &\equiv \frac{1}{\pi}\int_{-\pi}^{\pi} f'(x)\sin nx\,dx = -\frac{n}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx = -n\,a_n\,, \qquad (5.54) \\ a_0' &= \int_{-\pi}^{\pi} f'(x)\,dx = f(\pi) - f(-\pi) = 0\,. \end{aligned}$$

We have used the fact that $f(\pi) = f(-\pi)$ in several places; the Fourier coefficients of $f(x)$ are denoted by a_n and b_n as before.

Since $f'(x)$ is piecewise continuous on $[-\pi, \pi]$, and hence square integrable, it must satisfy Bessel's inequality, Eqs. (5.9) and (5.53):

$$(f',f') \geq \pi\left[\frac{a_0'^2}{2} + \sum_{n=1}^{\infty} (a_n'^2 + b_n'^2)\right] = \pi \sum_{n=1}^{\infty} n^2(b_n^2 + a_n^2)\,, \qquad (5.55)$$

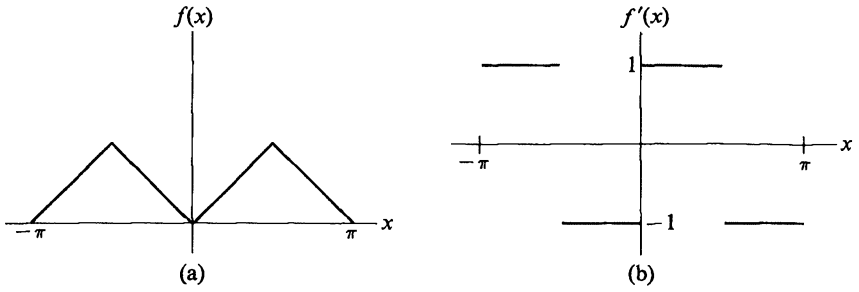

Fig. 5.2. Zig-zag function and its derivative.

where we have used Eqs. (5.54). [In fact, we know that the equality sign holds by the completeness relation, Eq. (5.10)].

The preceding results are all preliminary. We now use the Cauchy criterion to test for the uniform convergence of the Fourier series. Let

$$S_n = \frac{a_0}{2} + \sum_{p=1}^{n} a_p \cos px + \sum_{p=1}^{n} b_p \sin px \,.$$

By Theorem 5.1, if we can show that $|S_n - S_m| < \epsilon$ for all x in$[-\pi, \pi]$, and for all n, m larger than some $N(\epsilon)$, then we will have established uniform convergence. We have

$$\begin{aligned} |S_n - S_m| &= \left| \sum_{p=m+1}^{n} (a_p \cos px + b_p \sin px) \right| \\ &= \left| \sum_{p=m+1}^{n} \frac{1}{p} (pa_p \cos px + pb_p \sin px) \right| \\ &\leq \sqrt{\sum_{p=m+1}^{n} \left|\frac{1}{p}\right|^2 \sum_{p=m+1}^{n} \left| pa_p \cos px + pb_p \sin px \right|^2} \,, \end{aligned}$$

where we have used Schwarz's inequality [Eq. (4.8)]. This last expression may be written as

$$\begin{aligned} |S_n - S_m| &\leq \sqrt{\sum_{p=m+1}^{n} \frac{1}{p^2}} \cdot \sqrt{\sum_{p=m+1}^{n} p^2 |a_p \cos px + b_p \sin px|^2} \\ &\leq \sqrt{\sum_{p=m+1}^{n} \frac{1}{p^2}} \cdot \sqrt{\sum_{p=m+1}^{n} p^2 (a_p^2 + b_p^2)} \,, \end{aligned}$$

since $|a_p \cos px + b_p \sin px|$ can be written as

$$|\sqrt{a_p^2 + b_p^2} \cos (px - \theta)| \leq |\sqrt{a_p^2 + b_p^2}| \,,$$

where $\theta = \tan^{-1} b_p/a_p$. Therefore

$$\left| S_n - S_m \right| \leq \sqrt{\sum_{p=m+1}^{n} \frac{1}{p^2}} \cdot \sqrt{\frac{1}{\pi} \int_{-\pi}^{\pi} |f'(x)|^2 \, dx} \,,$$

where the term $\sqrt{\sum_{p=m+1}^{n} p^2(a_p^2 + b_p^2)}$ has been replaced by the equal or larger term

$$\sqrt{\sum_{p=1}^{\infty} p^2(a_p^2 + b_p^2)} \leq \sqrt{(1/\pi)\int_{-\pi}^{\pi} |f'|^2\, dx}\,, \tag{5.56}$$

from Eq. (5.55). Now the series $\sum_{p=1}^{\infty}(1/p^2)$ converges (it is, in fact, the Riemann zeta function evaluated at 2 and converges to $\pi^2/6$ as we shall show in Section 6.9). If we let

$$M = \sqrt{(1/\pi)\int_{-\pi}^{\pi} |f'(x)|^2\, dx}\,,$$

then we know that M is finite, because $f'(x)$ is piecewise continuous and always finite.

The series $\sum_{p=1}^{\infty}(1/p^2)$, converges, so it must satisfy the Cauchy criterion. Then for any positive number, which we take to be ϵ^2/M^2, there exists an integer N such that $\sum_{p=m+1}^{n}(1/p^2) < \epsilon^2/M^2$ when $n, m > N$. Taking the n and m that appear in $|S_n - S_m|$ larger than this N, we then have

$$|S_n - S_m| < (\epsilon^2/M^2)^{1/2} M = \epsilon\,,$$

and the proof of uniform convergence is complete. We know that the series converges *to* the function $f(x)$ because its convergence in the mean to $f(x)$ has already been established.

Note that the completeness relation could have been used earlier in this proof; there would then be an equals sign where we have an inequality in Eq. (5.56). It *is* valid to use the completeness relation because completeness has been proved already. But in order to demonstrate that prior knowledge of completeness is not required to prove uniform convergence, we have used Bessel's inequality instead. The piecewise continuity of $f'(x)$ implies that M is finite, and this is all that is needed to prove uniform convergence. However, to show that the uniform convergence is *to* $f(x)$ (since Cauchy's criterion makes no mention of the limit function) the convergence in the mean to $f(x)$ (completeness) was used at the very end of the proof. It is possible to prove independently that the uniform convergence is to $f(x)$, and then this result leads to an independent proof of completeness.

This theorem may be extended to deal with the piecewise continuous functions, such as step functions, which may have a finite number of finite discontinuities in the interval $[-\pi, \pi]$.

Theorem 5.5. If $f(x)$ is piecewise continuous in $[-\pi, \pi]$, and has a piecewise continuous derivative there, then its Fourier series converges uniformly to $f(x)$ in every closed subinterval of $[-\pi, \pi]$ in which $f(x)$ is continuous. At points of discontinuity of $f(x)$, its Fourier series converges to the arithmetic mean of the left- and right-hand limits of the function. If $f(x + 2\pi) = f(x)$, then these statements are true everywhere on the real line.

It is interesting that although we know that there exists some *trigonometric* sequence that converges uniformly to every continuous function on a closed interval, continuity of $f(x)$ alone is not sufficient to prove the covergence of its *Fourier* series. There exist continuous functions which are not differentiable, and their *Fourier* series need not converge. The additional condition of continuity of $f'(x)$, however, ensures convergence. Also, the assumption that the intervals be closed is vital to proving *uniform* convergence. The convergence of the Fourier series for piecewise continuous functions—such as step functions—will certainly not be uniform in open intervals (a, b), where one of the endpoints is a point of discontinuity. In fact, it may be shown that just before the Fourier series passes over the discontinuity, it differs from the function by a finite amount. This overshooting effect is called the Gibbs phenomenon (see Problem 4c).

It is primarily the smoothness of the function that determines the size of the Fourier coefficients—the smoother the function, the more rapidly these coefficients decrease and the more rapid is the convergence. The theory of Fourier series is largely the study of the interplay between assumptions about smoothness and conclusions about convergence.

Fejer has constructed a theory of Fourier series based on a special type of series summation (called Cesaro summation) in which one considers the sequence of *arithmetic means* of the partial sums. If a series is convergent in the usual sense, then it is Cesaro summable to the same value. But many divergent series are Cesaro summable; thus Cesaro summability is a natural generalization since it preserves the usual summation as a special case. It was shown by Fejer that the Fourier series of any continuous function $f(x)$ is uniformly Cesaro summable to $f(x)$ (see Rudin or Apostol). No assumptions on smoothness are required. This result is to be contrasted to the fact that there are continuous functions whose Fourier series are divergent at a point.

So far we have restricted our attention to functions defined on the interval $[-\pi, \pi]$. If these functions are periodic of period 2π, the expansions are as good everywhere as within the interval $[-\pi, \pi]$. It is a simple matter to generalize these results to the interval $[-l, l]$, and functions of period $2l$. But the requirement of periodicity cannot be removed if the series is to converge outside the basic interval.

Clearly the set of functions

$$\left\{\frac{e^{in\pi x/l}}{(2l)^{1/2}}\right\}, \qquad n = 0, \pm 1, \pm 2, \cdots$$

is a complete orthonormal set on the interval $[-l, l]$. All our results go through for this set of functions exactly as before. For the complex exponential series, we obtain

$$\begin{aligned} f(x) &= \sum_{n=-\infty}^{\infty} c_n \frac{e^{in\pi x/l}}{(2l)^{1/2}}, \\ c_n &= \frac{1}{(2l)^{1/2}} \int_{-l}^{l} f(x) e^{-in\pi x/l}\, dx . \end{aligned} \tag{5.57}$$

Similarly, for the trigonometric series,

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\frac{n\pi}{l}x + \sum_{n=1}^{\infty} b_n \sin\frac{n\pi}{l}x ,$$

$$a_n = \frac{1}{l}\int_{-l}^{l} f(x)\cos\frac{n\pi}{l}x\,dx \qquad \text{for } n = 0, 1, \cdots, \tag{5.58}$$

$$b_n = \frac{1}{l}\int_{-l}^{l} f(x)\sin\frac{n\pi}{l}x\,dx \qquad \text{for } n = 1, 2, \cdots$$

These formulas hold in the interval $[-l, l]$ under the conditions of the theorems, and will hold outside this interval only if $f(x + 2l) = f(x)$.

For any l (including $l = \pi$), if $f(x)$ is an even function, then $b_n = 0$ for all n and we have a Fourier cosine series; and if $f(x)$ is an odd function, $a_n = 0$ for each n and we have a Fourier sine series.

We close this section with a very simple, but very important observation. It will be recalled that the Legendre polynomials, which we have shown to be a complete orthonormal set, satisfy a second-order differential equation (Eq. 5.44). In this section we have been discussing the more familiar trigonometric functions, and have found that they also are a complete orthonormal set of functions. As we study other such sets of functions we shall always make note of the fact that these functions are the solutions of certain differential equations. Our treatment will not be derived from these differential equations, but rather will culminate in them, deriving them as a by-product of other considerations. Finally, by way of a summary, we shall, in Section 5.11, view all the special polynomials in their capacity as the solutions of second-order differential equations.

The trigonometric functions will, of course, not be included in this summary, since they are not polynomials. For completeness sake, and to help motivate the next section on Fourier integrals, we state the well-known differential equation satisfied by the sine and cosine functions:

$$\frac{d^2}{dx^2}u + \omega^2 u = 0 . \tag{5.59}$$

This second-order differential equation has the two linearly independent solutions

$$u_1 = \sin\omega x, \qquad u_2 = \cos\omega x . \tag{5.60}$$

If $\omega^2 = n^2$, where n is an integer, these solutions give the Fourier functions as n runs from 0 to ∞. It is instructive to prove the orthogonality of these functions from the form of the differential equation as was done for the Legendre polynomials.

5.7. FOURIER INTEGRALS

The restriction of the validity of Fourier series expansions to the basic interval $[-l, l]$ unless the expanded function is periodic is annoying because many functions are not periodic. However, a way around this difficulty is suggested

by the generalization we made in the last section: If we want to expand a nonperiodic function over any specified finite range, all we have to do is expand the function in an interval $[-l, l]$ large enough to contain the specified range. But what if we want an expansion for a nonperiodic function that is valid everywhere, over the whole real line? Throwing all considerations of rigor aside, we shall here derive heuristically a plausible answer to this question. (When a physicist speaks of a "heuristic" derivation, he is usually apologizing in advance for a sloppy, nonrigorous one; actually, the word means "serving to help toward discovery.") We shall take up these matters again, in a more rigorous fashion, in Chapter 9 (corollary to Theorem 9.11).

It is reasonable to try to get around the periodicity requirement by letting $l \to \infty$ in Eqs. (5.57). To carry this out, set $(\pi/l)^{1/2}x = y$ and $n(\pi/l)^{1/2} = k_n$, so $(n\pi/l)\,x = k_n y$ and $\Delta k_n \equiv k_{n+1} - k_n = (\pi/l)^{1/2}$, and $1/(2l)^{1/2} = \Delta k_n/(2\pi)^{1/2}$. Then, replacing c_n by g_{k_n}, Eqs. (5.57) become

$$f(y) = \frac{1}{(2\pi)^{1/2}} \sum_{k_n=-\infty}^{\infty} g_{k_n} e^{ik_n y}\, \Delta k_n ,$$

where

$$g_{k_n} = \frac{1}{(2\pi)^{1/2}} \int_{-\sqrt{\pi l}}^{\sqrt{\pi l}} f(y) e^{-ik_n y}\, dy ,$$

and k_n changes by steps of $\Delta k_n = (\pi/l)^{1/2}$, corresponding to steps of $\Delta n = 1$ in the original sum. Now let $l \to \infty$. As $l \to \infty$, $\Delta k_n \to 0$ and k_n takes on all real values. Therefore the sum over k_n becomes an integral over a continuous variable (which we denote by k), and we have, replacing y by x,

$$\begin{aligned} f(x) &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} g(k) e^{ikx}\, dk , \\ g(k) &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} f(x) e^{-ikx}\, dx . \end{aligned} \tag{5.61}$$

The function $g(k)$ is called the *Fourier transform* of $f(x)$, and vice versa. Two functions which satisfy Eqs. (5.61) are called a Fourier transform pair. This pair may be regarded as the statement and solution of an integral equation of a particular type.

By an easy extension of these results, we get the three-dimensional Fourier integral:

$$\begin{aligned} F(\mathbf{r}) &= \frac{1}{(2\pi)^{3/2}} \int_{-\infty}^{\infty} G(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}}\, d\mathbf{k} , \\ G(\mathbf{k}) &= \frac{1}{(2\pi)^{3/2}} \int_{-\infty}^{\infty} F(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}}\, d\mathbf{r} . \end{aligned} \tag{5.62}$$

Here $\mathbf{r}$ stands for x, y, z, and $d\mathbf{r} = dx\, dy\, dz$; similarly, $\mathbf{k}$ stands for k_x, k_y, k_z, and $d\mathbf{k} = dk_x\, dk_y\, dk_z$. Thus $\mathbf{k}\cdot\mathbf{r} = k_x x + k_y y + k_z z$. The integrals are threefold. We shall use this three-dimensional Fourier transform pair to solve some of the partial differential equations of physics in Chapter 7.

Equation (5.61) may be proved rigorously if one assumes that $f(x)$ is piecewise smooth and that $\int_{-\infty}^{\infty}|f(x)|\,dx$ exists (see Courant-Hilbert), but we shall not go into these issues here (see Section 9.6). Instead, we shall focus on the uses of the Fourier transform in physics.

Example. An interesting result may be obtained if we combine Eqs. (5.61) into the single equation

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(x')e^{-ikx'}\,dx'\right]e^{ikx}\,dk\,,$$

and then interchange the orders of integration to get

$$f(x) = \int_{-\infty}^{\infty} f(x')\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik(x-x')}\,dk\right]dx'\,.$$

The expression in brackets within the integration over x' must be a δ-function if this equation is to hold. Therefore we have an integral representation of the δ-function:

$$\delta(x - x') = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik(x-x')}\,dk\,. \tag{5.63}$$

This equation is to be interpreted as meaning that the expression on either side of it will have the same effect under an integral sign. This "derivation" shows clearly that the δ-functions (as naively conceived and used) originate as a consequence of the illegitimate business of interchanging orders of integration; the original crime must be paid for by the introduction of the extraordinary new mathematical objects called δ-functions. Their symbolic use is so simple, however, that one is apt to get the impression that the crime paid. In a sense it did: violations of the existing rules in mathematics and physics are always required to reach the new, which then turns out to hide the old as a special case.

The δ-function in three dimensions is

$$\delta^3(\mathbf{r} - \mathbf{r}_0) \equiv \delta(x - x_0)\delta(y - y_0)\delta(z - z_0) = \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_0)}\,d\mathbf{k}\,. \tag{5.64}$$

The integral representation of the δ-function can be used to prove the completeness relation (sometimes called Parseval's theorem) for the Fourier integral. Thus

$$\begin{aligned}\int_{-\infty}^{\infty}|f(x)|^2\,dx &= \int_{-\infty}^{\infty}dx\,\frac{1}{(2\pi)^{1/2}}\int_{-\infty}^{\infty}g^*(k)e^{-ikx}\,dk\,\frac{1}{(2\pi)^{1/2}}\int_{-\infty}^{\infty}g(k')e^{ik'x}\,dk'\\ &= \int_{-\infty}^{\infty}dk\,g^*(k)\int_{-\infty}^{\infty}dk'\,g(k')\left[\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{i(k'-k)x}\,dx\right]\\ &= \int_{-\infty}^{\infty}dk\,g^*(k)\int_{-\infty}^{\infty}dk'\,g(k')\delta(k' - k)\\ &= \int_{-\infty}^{\infty}g^*(k)\,g(k)\,dk = \int_{-\infty}^{\infty}|g(k)|^2\,dk\,.\end{aligned}$$

If x is the time t, and k is the frequency υ, then Parseval's theorem has a simple physical interpretation. It reads

$$\int_{-\infty}^{\infty} |f(t)|^2 \, dt = \int_{-\infty}^{\infty} |g(\upsilon)|^2 \, d\upsilon \, .$$

If, for example, $f(t)$ is a radiated electric field, then $|f(t)|^2$ is proportional to the total radiated power, and the integral is a measure of the total radiated energy. On the right-hand side is the integral over all frequencies of the spectrum amplitude squared, $|g(\upsilon)|^2$, which is proportional to the energy radiated per unit frequency interval. Thus Parseval's theorem expresses the conservation of energy.

Before turning to some applications, we state and prove the convolution theorem for Fourier transforms. We shall use it in Chapter 7. Let $f_1(x)$ and $f_2(x)$ be two functions whose Fourier transforms, $g_1(k)$ and $g_2(k)$, are given by Eq. (5.61).

The quantity

$$F(x) \equiv \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} f_1(t) f_2(x - t) \, dt$$

is called the *convolution* of the functions f_1 and f_2. The convolution theorem states that

$$F(x) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} g_1(k) g_2(k) e^{ikx} \, dk \, . \tag{5.65}$$

By taking the Fourier transform of this last equation, and thus solving for the product $g_1(k)\ g_2(k)$, we see that the convolution theorem may also be expressed as follows: The Fourier transforms of the convolution of two functions is the product of the Fourier transforms of these two functions. Denoting the Fourier transform of $F(x)$ by $G(k)$ this may be written

$$G(k) = g_1(k)\ g_2(k) \, .$$

The proof follows immediately from the properties of the Fourier-transform pair. Thus

$$\begin{aligned} F(x) &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} f_1(t) f_2(x - t) \, dt \\ &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} f_1(t) \left[\frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} g_2(k) e^{ik(x-t)} \, dk \right] dt \\ &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \left[\frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} f_1(t) e^{-ikt} \, dt \right] g_2(k) e^{ikx} \, dk \\ &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} g_1(k) g_2(k) e^{ikx} \, dk \, , \end{aligned}$$

which establishes the convolution theorem.

We shall now examine, in some detail, an example of the use of Fourier transforms, and incidentally, the calculation of some specific transforms.

We consider an undamped, one-dimensional harmonic oscillator acted upon by a time-varying but spatially uniform (no x-dependence) external force $F(t)$. This might be, for example, a spring hanging from a board which is jerked upward. We want to compute the energy transferred to the oscillator by such a force. Throughout this example, we shall limit ourselves to force functions $F(t)$ for which the Fourier transform exists.

The equation of motion is

$$\ddot{x} + \omega^2 x = (1/m)F(t) , \tag{5.66}$$

where m is the mass and ω is the natural frequency of the oscillator. We may rewrite this as

$$\dot{z} - i\omega z = (1/m)F(t) , \tag{5.67}$$

where $z = \dot{x} + i\omega x$. The energy of the oscillator at any time t is given by

$$E(t) = m\dot{x}^2/2 + m\omega^2 x^2/2 = (m/2)|z(t)|^2 .$$

Let us now assume that the oscillator is originally at rest ($x = \dot{x} = z = 0$ for $t \leq T_1$). Since the oscillator initially has zero energy, the energy ΔE transferred to the oscillator by $F(t)$ is $E(\infty)$. We now compute this quantity; we shall find that it depends in a simple way on the Fourier transform of $F(t)$.

Multiplying both sides of Eq. (5.67) by $\exp(-i\omega t)$ and integrating from T_1 to T_2 (arbitrary times), we obtain

$$\int_{T_1}^{T_2} (\dot{z} - i\omega z)\, e^{-i\omega t}\, dt = (1/m) \int_{T_1}^{T_2} F(t) e^{-i\omega t}\, dt .$$

Integration by parts of the term on the left gives

$$z(T_2)e^{-i\omega T_2} = (1/m) \int_{T_1}^{T_2} F(t) e^{-i\omega t}\, dt ,$$

where we have used the boundary condition that z vanishes for $t \leq T_1$. Since $F(t)$ vanishes for $t \leq T_1$ we may extend the lower limit of the integration to $-\infty$. The energy transfer is then given by

$$\begin{aligned} \Delta E &= \lim_{T_2 \to \infty} \frac{m}{2} |z(T_2)|^2 \\ &= \lim_{T_2 \to \infty} \frac{m}{2} \left| e^{i\omega T_2} \frac{1}{m} \int_{-\infty}^{T_2} F(t) e^{-i\omega t}\, dt \right|^2 \\ &= \frac{1}{2m} \left| \int_{-\infty}^{\infty} F(t) e^{-i\omega t}\, dt \right|^2 \\ &\equiv \frac{1}{2m} |f(\omega)|^2 , \end{aligned} \tag{5.68}$$

where $f(\omega)$ is the Fourier transform of $F(t)$ evaluated at the natural frequency of the oscillator:

$$f(\omega) \equiv \int_{-\infty}^{\infty} F(t) e^{-i\omega t}\, dt .$$

Thus $F(t)$ is given in terms of $f(\omega)$ by

$$F(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(\omega)e^{i\omega t}\,d\omega\,.$$

We have consolidated the factors of $(2\pi)^{-1/2}$ in one member of this transform pair, which therefore differs slightly in appearance from the symmetric form of Eqs. (5.61). Thus the energy transfer is essentially the absolute value squared of the Fourier component of the force function whose frequency is the natural frequency of the oscillator. The energy transfer can only occur at the resonant frequency ω, because only the ω-component of the Fourier transform of $F(t)$ enters into the formula for ΔE.

To solve the energy transfer problem for a specific force function, we must find its Fourier transform. We develop here some formulas which facilitate the determination of Fourier transforms. Let

$$F_n(t) \equiv \frac{d^{n-1}F(t)}{dt^{n-1}}\,, \qquad F_1(t) \equiv F(t)\,,$$

$$f_n(\omega) = \int_{-\infty}^{\infty} F_n(t)e^{-i\omega t}\,dt\,.$$

An integration by parts gives

$$f_n(\omega) = \left(\frac{1}{i\omega}\right)f_{n+1}(\omega) - \frac{1}{i\omega}\,F_n(t)\,e^{-i\omega t}\Big]_{-\infty}^{\infty}\,.$$

This formula will be useless unless $F_n(-\infty) = F_n(+\infty) = 0$. We therefore consider only such cases. Then, by repeated integration by parts (i.e., iteration of the above formula), we can express $f_1(\omega)$, the Fourier transform needed for this problem, in terms of the Fourier transform of some higher derivative of $F(t)$. This may facilitate the determination of $f_1(\omega)$ if the Fourier transform of a higher derivative is known. We obtain for $f_1(\omega) \equiv f(\omega)$,

$$f(\omega) = \left(\frac{1}{i\omega}\right)^{n-1} f_n(\omega)\,. \tag{5.69}$$

The index n is arbitrary, but in using the formula one naturally chooses the smallest n for which $f_n(\omega)$ is known.

We shall now compute the Fourier transforms of several force functions, and the energy transferred to an oscillator subjected to these forces.

Example 1. *Impulsive force*: $F(t) = P_0\delta(t)$. We obtain

$$\Delta E = \frac{1}{2m}\left|\int_{-\infty}^{\infty} P_0\delta(t)e^{-i\omega t}\,dt\right|^2 = P_0^2/2m\,.$$

This result may also be obtained by simply dropping the term $\omega^2 x$ in the equation of motion (5.66), the so-called impulse approximation.

Example 2. *Gaussian force*: $F(t) = (P_0/\sqrt{\pi}\tau)e^{-t^2/\tau^2}$ (Fig. 5.3). The width of the Gaussian pulse at half-maximum is $2\sqrt{\ln 2}\tau$. The Fourier transform may

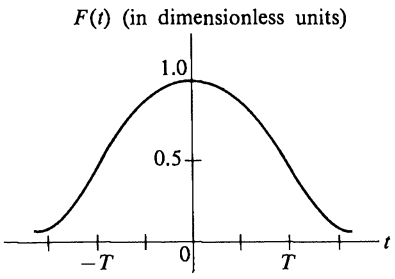

Fig. 5.3.

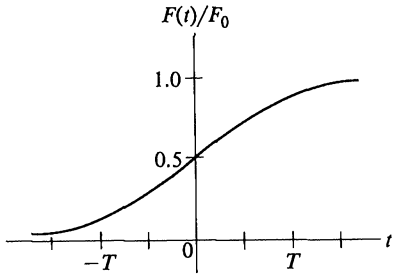

Fig. 5.4.

be computed directly by completing the square in the exponent:

$$f(\omega) = \frac{P_0}{\sqrt{\pi}\tau}\int_{-\infty}^{\infty} e^{-t^2/\tau^2}e^{-i\omega t}\,dt$$

$$= \frac{P_0}{\sqrt{\pi}\tau}\,e^{-\omega^2\tau^2/4}\int_{-\infty}^{\infty} e^{-(t+i\omega\tau^2/2)^2/\tau^2}\,dt$$

$$= \frac{P_0}{\pi^{1/2}}\,e^{-\omega^2\tau^2/4}\int_{-\infty}^{\infty} e^{-x^2}\,dx = P_0 e^{-\omega^2\tau^2/4}\ .$$

By using techniques that will be developed in Section 6.9, we can show that this formal procedure for evaluating the integral is in fact valid. Note that the Fourier transform of a Gaussian is a Gaussian. For the energy transferred to the oscillator, we find that

$$\Delta E = (P_0^2/2m)e^{-\omega^2\tau^2/2}\ .$$

In the "sudden" limit ($\tau << \omega^{-1}$), $\Delta E = (P_0^2/2m)$, as in Example 1. This is to be expected because in the limit $\tau \to 0$, the Gaussian force function becomes a δ-function (see Example 2, Section 5.3). In the adiabatic limit ($\tau >> \omega^{-1}$), there is no energy transferred to the oscillator which also makes sense physically.

Example 3. *Probability integral*:

$$F(t) = (F_0/\sqrt{\pi}\tau)\int_{-\infty}^{t} e^{-t'^2/\tau^2}\,dt'$$

(Fig. 5.4). In this case the Fourier integral does not really exist because the function assumes a nonzero value (F_0) asymptotically as $t \to \infty$. We may still use Eqs. (5.68) and (5.69) to calculate the energy transfer, imagining that the constant force is turned off adiabatically at very large times. This will not affect the net energy transfer, and will bring about convergence of the Fourier integral. We have already determined the Fourier transform of

$$dF(t)/dt \equiv F_2(t) = (F_0/\sqrt{\pi}\tau)e^{-t^2/\tau^2}$$

in Example 2. Using Eq. (5.69) with $n = 2$ and Eq. (5.68), we obtain for the

energy transfer

$$\Delta E = \frac{1}{2m}\left| f(\omega) \right|^2 = \frac{1}{2m}\left|\left(\frac{1}{i\omega}\right) f_2(\omega)\right|^2 = \frac{1}{2m}(F_0/\omega)^2 e^{-\omega^2\tau^2/2} .$$

Again, in the adiabatic limit, $\Delta E \to 0$; in the sudden limit, $\Delta E \to 1/2m\,(F_0/\omega)^2$. In the latter case, the mass, because of its inertia, suddenly finds itself displaced a distance $l_0 = (F_0/m\omega^2)$ from equilibrium and hence acquires the energy $(\frac{1}{2}m\omega^2 l_0^2)$ corresponding to the full displacement. It has no time to follow the shifting equilibrium position in the sudden limit. This result for the sudden limit should agree with the limiting behavior of the energy transfer for the force functions of Problems 7 and 8.

5.8 SPHERICAL HARMONICS AND ASSOCIATED LEGENDRE FUNCTIONS

We have derived the completeness of the Legendre polynomials from Weierstrass's theorem in one variable. From the two-variable generalization of Weierstrass's theorem, we proved the completeness of the trigonometric functions. We now derive the completeness of the spherical harmonics from Weierstrass's for three variables. It tells us that a function F of x, y, z (that is, $\mathbf{r}$) can be approximated uniformly by a sequence of partial sums as follows:

$$F(\mathbf{r}) = \lim_{M\to\infty} F_M(\mathbf{r}) = \lim_{M\to\infty} \sum_{m,n,p=0}^{M} c_{mnp}^{(M)} x_1^m x_2^n x_3^p . \tag{5.70}$$

We may also express $F_M(\mathbf{r})$ in terms of the three variables:

$$z_1 \equiv x_1 + ix_2 = r\sin\theta e^{i\phi} ,$$
$$z_2 \equiv x_1 - ix_2 = r\sin\theta e^{-i\phi} ,$$
$$z_3 \equiv x_3 = r\cos\theta ,$$

which are linear combinations of x_1, x_2, x_3. Thus

$$F_M(\mathbf{r}) = \sum_{\alpha,\beta,\gamma=0}^{M} A_{\alpha\beta\gamma}^{(M)} z_1^\alpha z_2^\beta z_3^\gamma$$
$$= \sum_{l=0}^{3M} r^l \sum_{\substack{\alpha,\beta,\gamma=0 \\ (\alpha+\beta+\gamma=l)}}^{M} A_{\alpha\beta\gamma}^{(M)} e^{l(\alpha-\beta)\phi} \sin^{(\alpha+\beta)}\theta \cos^\gamma\theta . \tag{5.71}$$

In the last expression we have converted the unrestricted sum over all values of α, β, γ to a sum over just those combinations of α, β, γ that sum to the integer l. But then we sum over all possible restrictions (all l), which in effect removes the restriction on α, β, γ and gives the same results as the original unrestricted sum.

We now restrict $F(\mathbf{r})$ to the unit sphere by requiring that $r = 1$; also, we set $(\alpha - \beta) = m$. We shall need the following facts: Since α, β, and γ are

all positive or zero, we have

$$\alpha + \beta = m + 2\beta \geq 0 .$$

Also,

$$\alpha + \beta \geq |\alpha - \beta| = |m| \Longrightarrow \alpha + \beta - |m| = m - |m| + 2\beta \geq 0 .$$

Furthermore, $(\alpha + \beta - |m|)$ is always even because 2β is even and $m - |m| = 0$ if $m \geq 0$ and $m - |m| = -2m$ if $m < 0$. We now rewrite Eq. (5.71) in the form

$$F_M(\theta, \phi) = \sum_{l=0}^{3M} \sum_{\substack{\alpha,\beta,\gamma=0 \\ (\alpha+\beta+\gamma=l)}}^{M} A_{\alpha\beta\gamma}^{(M)} e^{im\phi} \sin^{(\alpha+\beta-|m|)} \theta \cos^{\gamma} \theta \sin^{|m|} \theta .$$

By a trigonometric identity,

$$\sin^{(\alpha+\beta-|m|)}\theta \cos^{\gamma} \theta = (1 - \cos^2 \theta)^{(\alpha+\beta-|m|)/2} \cos^{\gamma} \theta ,$$

which is a polynomial in $\cos\theta$ of maximum degree $\alpha + \beta + \gamma - |m| = l - |m|$, since $(\alpha + \beta - |m|)$ is even. Denoting this polynomial by $f_{lm}(\cos\theta)$, we get

$$F_M(\theta, \phi) = \sum_{l=0}^{3M} \sum_{m} B_{lm}^{(M)} e^{im\phi} \sin^{|m|} \theta f_{lm}(\cos\theta) .$$

The range of the summation over m is still to be specified. Since $m = \alpha - \beta$, $m \leq l$; also, since we sum only over nonnegative α, β, γ, $\alpha + \beta + \gamma - |m| = l - |m| \geq 0$. Therefore, in changing to a sum over m, we must only sum over those m's such that $|m| \leq l$. Thus the last equation becomes

$$F_M(\theta, \phi) = \sum_{l=0}^{3M} \sum_{m=-l}^{m=l} B_{lm}^{(M)} e^{im\phi} \sin^{|m|} \theta f_{lm}(\cos\theta) . \tag{5.72}$$

Therefore the sequence of functions

$$\overline{Y}_{lm}(\theta, \phi) \equiv e^{im\phi} \sin^{|m|} \theta f_{lm}(\cos\theta) ,$$

where $f_{lm}(\cos\theta)$ is a polynomial in $\cos\theta$ of degree $(l - |m|)$, provides a uniform approximation to any continuous function defined on the unit sphere—that is, over the range $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$. If we can now construct from this set of functions an *orthonormal* set of functions, then by reasoning exactly analogous to that applied to the Legendre polynomials and the trigonometric functions, we will know that this set of functions is complete. This is just another instance of the use of our basic result on completeness: whenever we can uniformly approximate an arbitrary function in Hilbert space by a sequence of partial sums which are themselves linear combinations of a set of orthonormal functions, then this set is a complete orthonormal set in Hilbert space.

Thus let us look for functions Y_{lm} satisfying

$$\int_{\Omega} Y_{l'm'}^* Y_{lm} \, d\Omega \equiv \int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta \, d\theta \, Y_{l'm'}^* Y_{lm} = \delta_{ll'}\delta_{mm'} . \tag{5.73}$$

This orthonormality condition uniquely determines the functions Y_{lm} up to a phase factor, exactly as in the case of the normalized Legendre polynomials, $\bar{P}_l$. Thus there can only be one complete set of orthonormal functions defined on the unit sphere. We shall calculate a few of the Y_{lm} explicitly.

If $l = m = 0$, we obtain $Y_{00} = (1/4\pi)^{1/2}$. If $l = 1$, then m may equal -1, 0, or $+1$. Recalling that $f_{lm}(\cos\theta)$ is a polynomial in $\cos\theta$ of degree $l - |m|$, we obtain

$$Y_{10} = a_1 \cos\theta + a_2 ,$$
$$Y_{11} = a_3 e^{i\phi} \sin\theta ,$$
$$Y_{1,-1} = a_4 e^{-i\phi} \sin\theta .$$

The constants a_1, a_2, a_3, a_4 are determined by imposing orthonormality. Thus

$$0 = \delta_{01}\delta_{03} = \int Y_{00}^* Y_{10}\, d\Omega$$
$$= \frac{1}{(4\pi)^{1/2}} \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta (a_1 \cos\theta + a_2) \Longrightarrow a_2 = 0 ,$$
$$1 = \delta_{10}\delta_{10} = \int |Y_{10}|^2\, d\Omega \Longrightarrow a_1 = \left(\frac{3}{4\pi}\right)^{1/2} .$$

Similarly, $a_3 = -a_4 = -(3/8\pi)^{1/2}$. We choose the minus sign to be consistent with the convention to be adopted later, in Eq. (5.75). Therefore the first few members of the complete orthonormal set of functions on the unit sphere are

$$Y_{00} = \left(\frac{1}{4\pi}\right)^{1/2}, \qquad Y_{11} = -\left(\frac{3}{8\pi}\right)^{1/2} e^{i\phi} \sin\theta ,$$
$$Y_{10} = \left(\frac{3}{4\pi}\right)^{1/2} \cos\theta , \qquad Y_{1,-1} = \left(\frac{3}{8\pi}\right)^{1/2} e^{-i\phi} \sin\theta . \tag{5.74}$$

Note that for a given l there are $2l + 1$ functions Y_{lm}. The functions Y_{lm} are called *spherical harmonics*. They are a sort of two-dimensional generalization and combination of the Fourier functions and the Legendre polynomials, since they do for the spherical *surface* what the Fourier functions and the Legendre polynomials do for their respective linear intervals.

We give next a general formula for the Y_{lm}. Since we know that the Y_{lm} are unique, the formula is true if we establish orthonormality—just as in the case of the Legendre polynomials. The formula is

$$Y_{lm}(\theta, \phi) = (-1)^m \left[\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}\right]^{1/2} P_l^m(\cos\theta) e^{im\phi} , \qquad m \geq 0 , \tag{5.75}$$
$$Y_{l,-m} = (-1)^m Y_{lm}^* , \qquad m \geq 0 ,$$

where

$$P_l^m(x) = (1 - x^2)^{m/2} \frac{d^m}{dx^m} P_l(x) , \qquad m \geq 0 , \tag{5.76}$$

are called the *associated Legendre functions*. Thus the functions

$$Y_{l0}(\theta, \phi) = \left(\frac{1}{2\pi}\right)^{1/2}\left(\frac{2l+1}{2}\right)^{1/2} P_l(\cos\theta)$$

are an orthonormal set of functions on the unit sphere.

We shall first show that if u is a solution of Legendre's differential equation, that is,

$$(1 - x^2)u'' - 2xu' + l(l+1)u = 0\,, \tag{5.77}$$

then $v \equiv (1 - x^2)^{m/2}(d^m/dx^m)u$, for integral $m \geq 0$, is a solution of the equation

$$(1 - x^2)v'' - 2xv' + [l(l+1) - m^2/(1 - x^2)]v = 0\,, \tag{5.78}$$

known as Legendre's associated equation. Let $d^m u/dx^m \equiv d^m u \equiv w$. Then

$$v = (1 - x^2)^{m/2}w\,, \qquad v' = (1 - x^2)^{m/2}w' - mx(1 - x^2)^{(m/2)-1}w\,,$$
$$v'' = (1 - x^2)^{m/2}w'' - 2mx(1 - x^2)^{(m/2)-1}w' - m(1 - x^2)^{(m/2)-1}w$$
$$+ mx^2(m - 2)(1 - x^2)^{(m/2)-2}w\,.$$

Substituting into (5.78) and replacing w by $d^m u$, the left-hand side of Eq. (5.78) becomes

$$(1 - x^2)d^{m+2}u - 2(m+1)x\,d^{m+1}u + [l(l+1) - m(m+1)]d^m u\,. \tag{5.79}$$

We shall now show that this expression can be written in the form

$$d^m[(1 - x^2)u'' - 2xu' + l(l+1)u]\,, \tag{5.80}$$

which vanishes, by comparison with Eq. (5.77). This completes the proof that $v = (1 - x^2)^{m/2}(d^m u/dx^m)$ is a solution of Legendre's associated equation. Using Leibnitz's formula [Eq. (5.42)], we obtain

$$d^m[(1 - x^2)u''] = \sum_{n=0}^{m}\binom{m}{n} d^n(1 - x^2)d^{m-n+2}u$$
$$= (1 - x^2)d^{m+2}u - 2xm\,d^{m+1}u - m(m-1)d^m u\,,$$

and also

$$\frac{d^m}{dx^m}[-2xu'] = -2xd^{m+1}u - 2md^m u\,.$$

Combining the last two lines and the term $l(l+1)d^m u$ from Eq. (5.80) establishes the equivalence of Eqs. (5.79) and (5.80) and completes the proof that the functions

$$P_l^m(x) = (1 - x^2)^{m/2}\frac{d^m}{dx^m}P_l(x) = \frac{1}{2^l l!}(1 - x^2)^{m/2}\frac{d^{l+m}}{dx^{l+m}}(x^2 - 1)^l \tag{5.81}$$

are solutions of Legendre's associated equation (5.78). We note that $P_l^0(x) = P_l(x)$.

Using the same procedure that led to the orthonormality relations for the Legendre polynomials, we can show that

$$\int_{-1}^{1} P_l^m(x) P_{l'}^m \, dx = \frac{(l+m)!}{(l-m)!} \frac{2}{2l+1} \delta_{ll'} . \tag{5.82}$$

Once this result has been established, the normalization constant of the Y_{lm} (Eq. 5.75) follows immediately, as the integration over ϕ merely produces a factor of 2π. There is, of course, a free choice of phase factor; ours is a common choice in the physics literature. However, one must be careful, because different authors choose different phase factors for the spherical harmonics. The $P_l^m(x)$ are *not* another orthonormal set of polynomials on $[-1, 1]$, challenging the uniqueness of the Legendre polynomials, for the simple reason that they are not polynomials at all! Equation (5.81) shows this most clearly.

The $P_l^m(x)$ are not the only solutions to Legendre's associated equation. Just as there is a second set of solutions to Legendre's equation (5.77), so too there is a second set of solutions to Eq. (5.78). However, they also are not finite at $x = \pm 1$.

At this point we state an important result; the addition theorem for spherical harmonics. If two vectors $\mathbf{x}(r, \theta, \phi)$ and $\mathbf{x}'(r', \theta', \phi')$ have an angle γ between them, then

$$P_l(\cos\gamma) = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} Y_{lm}^*(\theta', \phi') Y_{lm}(\theta, \phi) , \tag{5.83}$$

where

$$\cos\gamma = \frac{\mathbf{x}\cdot\mathbf{x}'}{|\mathbf{x}|\,|\mathbf{x}'|} = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\phi - \phi') . \tag{5.84}$$

Thus the addition theorem expresses a Legendre polynomial of order l in the angle γ in terms of a sum over products of spherical harmonics of the angles θ, ϕ and θ', ϕ'. We omit the proof; note, however, that for the case $l = 1$, Eq. (5.83) is just a statement of the familiar result of Eq. (5.84).

There is a raft of other properties of the spherical harmonics and associated Legendre functions that are straightforward, but messy, generalizations of the identities proved for Legendre polynomials. We shall not go into these further.

We devote the remainder of this section to a discussion of the spherical harmonics in their capacity as solutions to the angular part of Laplace's equation in spherical coordinates. Since the angular part of the Laplacian is essentially the quantum-mechanical operator that represents the total angular momentum squared, the spherical harmonics play a prominent role in quantum theory. The operator that represents the square of the total angular momentum is

$$L^2 = -\hbar^2\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right) . \tag{5.85}$$

Let us first see why this is the angular-momentum operator. Classically, $\mathbf{L} = \mathbf{r} \times \mathbf{p}$. If we make the usual Schrödinger substitution for $\mathbf{p}$: $\mathbf{p} \rightarrow i\hbar\nabla$, we

find that

$$L_x = -i\hbar\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right),$$
$$L_y = -i\hbar\left(z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z}\right), \tag{5.86}$$
$$L_z = -i\hbar\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right).$$

Now we transform to spherical coordinates. The equations of transformation of coordinates are:

$$x = r\sin\theta\cos\phi, \qquad r = (x^2 + y^2 + z^2)^{1/2},$$
$$y = r\sin\theta\sin\phi, \qquad \phi = \tan^{-1} y/x,$$
$$z = r\cos\theta, \qquad \theta = \tan^{-1}\frac{(x^2 + y^2)^{1/2}}{z}.$$

Therefore

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial}{\partial\theta}\frac{\partial\theta}{\partial x} + \frac{\partial}{\partial\phi}\frac{\partial\phi}{\partial x}$$
$$= \sin\theta\cos\phi\frac{\partial}{\partial r} + \frac{\cos\theta\cos\phi}{r}\frac{\partial}{\partial\theta} - \frac{1}{r}\frac{\sin\phi}{\sin\theta}\frac{\partial}{\partial\phi}.$$

Similarly,

$$\frac{\partial}{\partial y} = \sin\theta\sin\phi\frac{\partial}{\partial r} + \frac{1}{r}\cos\theta\sin\phi\frac{\partial}{\partial\theta} + \frac{1}{r}\frac{\cos\phi}{\sin\theta}\frac{\partial}{\partial\phi},$$
$$\frac{\partial}{\partial z} = \cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial\theta}.$$

Using these relations in Eqs. (5.86), we obtain after some manipulation,

$$L_x = +i\hbar\left(\sin\phi\frac{\partial}{\partial\theta} + \cos\phi\cot\theta\frac{\partial}{\partial\phi}\right),$$
$$L_y = -i\hbar\left(\cos\phi\frac{\partial}{\partial\theta} - \sin\phi\cot\theta\frac{\partial}{\partial\phi}\right), \tag{5.87}$$
$$L_z = -i\hbar\frac{\partial}{\partial\phi}.$$

Now we form L^2:

$$L^2 = L_x^2 + L_y^2 + L_z^2 = -\hbar^2\left[\frac{\partial^2}{\partial\theta^2} + \cot\theta\frac{\partial}{\partial\theta} + (1 + \cot^2\theta)\frac{\partial^2}{\partial\phi^2}\right]$$
$$= -\hbar^2\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right]. \tag{5.88}$$

We observe [see Eq. (1.68)] that the L^2 operator is essentially the angular part

of the Laplacian in spherical coordinates:

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) - \frac{L^2}{\hbar^2 r^2} . \tag{5.89}$$

Thus we may write Laplace's equation in the form

$$\nabla^2\psi = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) - \frac{L^2\psi}{\hbar^2 r^2} = 0 ,$$

or

$$\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) = \frac{1}{\hbar^2}L^2\psi . \tag{5.90}$$

In order to solve this equation, let us assume that $\psi(r, \theta, \phi)$ is a product of two functions, one depending on r alone, and the other on θ and ϕ:

$$\psi(r, \theta, \phi) = R(r)\,Y(\theta, \phi) .$$

If we substitute a solution of this form into Eq. (5.90), and divide the equation by $\psi = RY$, we obtain

$$\frac{1}{R(r)}\frac{\partial}{\partial r}\left(r^2\frac{\partial R(r)}{\partial r}\right) = \frac{1}{\hbar^2 Y(\theta, \phi)}L^2 Y(\theta, \phi) .$$

The left-hand side is a function of r alone, and the right-hand side is a function only of θ and ϕ. This can only hold if both sides equal the same constant, for otherwise we could vary r and yet the right-hand side, depending only on θ and ϕ, would not change. Let us call the constant λ. Then we have the two equations

$$\frac{d}{dr}\left(r^2\frac{dR(r)}{dr}\right) = \lambda R(r) , \tag{5.91}$$

$$L^2 Y(\theta, \phi) = \hbar^2\lambda Y(\theta, \phi) . \tag{5.92}$$

It will be shown below that a solution for the angular equation (5.92) is just the spherical harmonic, Y_{lm} with $\lambda = l(l + 1)$. In other words, the spherical harmonics are eigenvectors of the operator L^2 corresponding to eigenvalues $\hbar^2 l(l + 1)$. For $\lambda = l(l + 1)$, the radial equation (5.91) may be solved by letting $R(r) = U(r)/r$. It then becomes

$$\frac{d^2}{dr^2}U(r) - l(l + 1)\frac{U(r)}{r^2} = 0 , \tag{5.93}$$

which has the solution

$$U(r) = rR(r) = Ar^{l+1} + Br^{-l} , \tag{5.94}$$

where A and B are constants. Together with the spherical harmonics, these radial functions give a solution to Laplace's equation in spherical coordinates. A function which satisfies Laplace's equation is called a harmonic function. This explains the origin of the name "spherical harmonics": these functions

are *harmonic*, since they satisfy the angular part of Laplace's equation, and they are defined on the surface of the unit *sphere*.

The most general solution is a linear combination of solutions for given values of l and m. Thus

$$\psi(r, \theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} [A_{lm} r^l + B_{lm} r^{-l-1}] Y_{lm}(\theta, \phi) . \tag{5.95}$$

It is of interest to separate the variables again, in Eq. (5.92). That is, let $Y_{lm} = \Phi(\phi) P(\theta)$, substitute into the equation, and divide through by $Y = \Phi P$. The result may be written

$$\frac{-\sin\theta}{P} \frac{d}{d\theta}\left(\sin\theta \frac{dP}{d\theta}\right) - l(l+1)\sin^2\theta = \frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} .$$

The θ- and ϕ-dependences have thus been isolated and so the expressions on both sides of the equation must equal the same constant, call it $-m^2$. We then have the two equations

$$\frac{d^2\Phi}{d\phi^2} = -m^2\Phi , \tag{5.96}$$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta \frac{dP}{d\theta}\right) + \left[l(l+1) - \frac{m^2}{\sin^2\theta}\right]P = 0 . \tag{5.97}$$

The azimuthal equation (in ϕ), Eq. (5.96), is immediately solved:

$$\Phi(\phi) = Ae^{\pm im\phi} . \tag{5.98}$$

For $\Phi(\phi)$ to be single-valued over the range $[0, 2\pi]$, m must be an integer. This result was anticipated in choosing the form of the separation constant. The equation in θ (Eq. 5.97) will assume a recognizable form if we simply change variables. Letting $\cos\theta = x$, so $dx = d(\cos\theta)$, and $\sin\theta = (1 - x^2)^{1/2}$, it becomes

$$(1 - x^2)P'' - 2xP' + \left[l(l+1) - \frac{m^2}{1 - x^2}\right]P = 0 , \tag{5.99}$$

which, due to our earlier choice of $\lambda = l(l+1)$, is just Legendre's associated equation (5.78). The solutions are the associated Legendre functions $P(x) = P_l^m(x)$ given in Eq. (5.81). This proves that the spherical harmonics are eigenvectors of L^2 with eigenvalues $\hbar^2 l(l+1)$, as stated above.

It should be noted that we really found *two* solutions to the second-order differential equation (5.91), $R(r) = r^l$ and $R(r) = r^{-l-1}$. This was also true for the differential equation (5.96), where we found the two solutions given by Eq. (5.98). However, for Eq. (5.99), which is also a second order differential equation, we have stated only one solution, $P_l^m(x)$. There is, of course, a second solution, just as there was to the other two equations. The second solution is rarely used in physical problems, however, because it is singular at $x = \pm 1$ ($\theta = 0, \pi$) and must usually be rejected in order to satisfy boundary conditions.

Usually one first encounters all these functions in their capacity as solutions to differential equations that arise in physics. Our approach has been different.

We have emphasized the *completeness* properties of these sets of functions, touching only incidentally on them as solutions to differential equations.

5.9 HERMITE POLYNOMIALS

In the next section we shall develop the properties of the various sets of orthogonal polynomials as special cases of a single comprehensive framework. But in order for us to be able to see the trees, and not just the forest, we shall first work out some of the properties of the Hermite polynomials in an inductive way. We first encountered these polynomials in the discussion of the eigenvalue problem for the quantum-mechanical harmonic oscillator (see Section 3.10).

The Hermite polynomials, $H_n(x)$, are the orthogonal polynomials on the infinite interval $[-\infty, \infty]$ with respect to the nonnegative weight function e^{-x^2}. That is, for $m \neq n$,

$$\int_{-\infty}^{\infty} H_n(x) H_m(x) e^{-x^2} dx = 0 . \tag{5.100}$$

Thus they differ in two important ways from the orthonormal sets of functions dealt with so far. First, they are defined from $-\infty$ to $+\infty$. Therefore Weierstrass' theorem, which deals only with finite intervals, does not apply. Secondly, they are orthogonal with respect to a weight function (see Definition 5.2). The Hermite polynomials are most conveniently *defined* in terms of a generating function $\psi(x, t)$:

$$\psi(x, t) \equiv e^{-t^2+2tx} = e^{x^2} e^{-(t-x)^2} \equiv \sum_{n=0}^{\infty} \frac{H_n(x)}{n!} t^n , \tag{5.101}$$

From this implicit definition of $H_n(x)$, we can get an explicit formula. It is clear that

$$\begin{aligned} H_n(x) &= \left[\frac{\partial^n \psi(x, t)}{\partial t^n}\right]_{t=0} = \left[e^{x^2} \frac{\partial^n}{\partial t^n} e^{-(t-x)^2}\right]_{t=0} \\ &= \left[e^{x^2} \frac{d^n}{dy^n} e^{-y^2}\right]_{y=-x} = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} . \end{aligned} \tag{5.102}$$

This is the Rodrigues formula for the Hermite polynomials. It shows that they are indeed polynomials of degree n.

We next derive some recursion relations for the Hermite polynomials. Since it is true of the generating function that

$$\frac{\partial \psi(x, t)}{\partial x} = 2t \psi(x, t) ,$$

it follows that

$$\sum_n \frac{H_n'(x) t^n}{n!} = \sum_n \frac{2H_n(x) t^{n+1}}{n!} .$$

Equating the coefficients of equal powers of t, we get

$$H_n'(x) = 2nH_{n-1}(x), \qquad n \geq 1 . \tag{5.103}$$

Likewise, from

$$\frac{\partial \psi(x, t)}{\partial t} + 2(t - x)\,\psi(x, t) = 0\,,$$

it follows that

$$H_{n+1}(x) - 2xH_n(x) + 2nH_{n-1}(x) = 0, \qquad n \geq 1\,. \tag{5.104}$$

The reader may show that, by differentiating these recursion relations and combining them properly, the following differential equation for the $H_n(x)$ can be derived:

$$H_n''(x) - 2x\,H_n'(x) + 2nH_n(x) = 0, \qquad n \geq 0\,. \tag{5.105}$$

There are many ways to prove orthogonality of the $H_n(x)$ with respect to the weight function e^{-x^2}. The two methods we have used for the Legendre polynomials—repeated integration by parts, and the proof based directly on the differential equation—also work for the Hermite polynomials. However, we shall use a different method, based on the generating function, in order to illustrate still another technique. Consider the integral

$$I = \int_{-\infty}^{\infty} e^{-t^2+2tx} e^{-s^2+2sx} e^{-x^2}\,dx = \sum_{n,m=0}^{\infty} \frac{t^n s^m}{n!m!} \int_{-\infty}^{\infty} H_n H_m e^{-x^2}\,dx\,.$$

We want to prove that the integral on the right is zero for $m \neq n$. The integral on the left is

$$I = e^{-(t^2+s^2)} \int_{-\infty}^{\infty} e^{-x^2+2(s+t)x}\,dx = e^{-(t^2+s^2)} \pi^{1/2} e^{(s+t)^2} = \pi^{1/2} e^{2st}\,;$$

the integral is done by completing the square. Expanding this expression for I in a Taylor series, we have

$$I = \pi^{1/2} \sum_{n=0}^{\infty} \frac{(2st)^n}{n!} = \sum_{n,m=0}^{\infty} \frac{t^n s^m}{n!\,m!} \int_{-\infty}^{\infty} H_n H_m\, e^{-x^2}\,dx\,.$$

If equal powers of s and t are equated in this identity, we find that

$$\int_{-\infty}^{\infty} H_n H_m e^{-x^2}\,dx = \pi^{1/2} 2^n n! \delta_{nm}\,. \tag{5.106}$$

Thus the functions $u_n(x) = N_n H_n(x) e^{-x^2/2}$ with $N_n = (1/\pi^{1/2} 2^n n!)^{1/2}$ are orthonormal on the interval $(-\infty, \infty)$. The functions $u_n(x)$ are the eigenfunctions of the quantum oscillator. We may say equivalently that the set of Hermite polynomials $H_n(x)$ are orthonormal on the interval $(-\infty, \infty)$ with respect to the weight function $N_n^2 e^{-x^2}$.

Another important set of polynomials are the Laguerre polynomials, $L_n(x)$. They are orthogonal on the interval $[0, \infty]$ with respect to the weight function e^{-x}, that is

$$\int_0^{\infty} L_n(x) L_m(x) e^{-x}\,dx = 0 \qquad \text{for } n \neq m\,. \tag{5.107}$$

Everything that we have done for Hermite polynomials can be done for Laguerre polynomials (Problem 6.36).

5.10 STURM-LIOUVILLE SYSTEMS—ORTHOGONAL POLYNOMIALS

All the special functions we have studied so far are solutions to differential equations. These equations and their solutions look rather different at first, but, in fact, they have a great deal in common. In this section we shall view them as special cases in a single general framework.

All the equations are second-order, linear differential equations of the form

$$Lu = \lambda u \,, \tag{5.108}$$

where

$$L = \alpha(x)\frac{d^2}{dx^2} + \beta(x)\frac{d}{dx} + \gamma(x), \tag{5.109}$$

and λ is a constant; α, β, and γ are real functions of x. The operators L have an important property in common that until now we have not mentioned. Namely, they are *Hermitian* if (1) we use the inner product

$$(f, g) \equiv \int_{-\infty}^{\infty} f^*(x)g(x)w(x)\,dx \,, \tag{5.110}$$

[where the weight function $w(x) \geq 0$], and, (2) the functions u satisfy appropriate boundary conditions. We have not yet committed ourselves to a choice of $w(x)$, and we will find that this freedom is crucial in attaining the generality which we want.

In quantum mechanics, all the operators which correspond to physical observables are Hermitian. It will be recalled that a Hermitian operator H satisfies the equation

$$(Hx, y) = (x, Hy) \,,$$

where x and y are vectors.†

Now suppose that f and g are two vectors (functions) in Hilbert space. For the inner product we defined above, the condition that L be Hermitian is

$$(Lf, g) \equiv \int_{-\infty}^{\infty} (Lf(x))^* g(x)w(x)\,dx = \int_{-\infty}^{\infty} (f(x))^*(Lg(x))w(x)\,dx \equiv (f, Lg) \,. \tag{5.111}$$

Henceforth, we restrict attention to such operators.

To investigate the hermiticity of L, we compute (f, Lg) and (Lf, g) and then form their difference; L is Hermitian if and only if this difference is zero.

† Operators, H, satisfying $(Hx, y) = (x, Hy)$ on an infinite dimensional vector space are called *symmetric* operators by mathematicians. We will refer to them as Hermitian operators. They clearly have real eigenvalves and orthonormal eigenfunctions according to our work in Chapter 4.

We have

$$(f, Lg) = \int_{-\infty}^{\infty} wf^*\alpha g'' \, dx + \int_{-\infty}^{\infty} wf^*\beta g' \, dx + \int_{-\infty}^{\infty} wf^*\gamma g \, dx .$$

Integrating the first two integrals by parts once, we get

$$(f, Lg) = [w\alpha f^* g']_{-\infty}^{\infty} - \int_{-\infty}^{\infty} [(w\alpha)' f^* g' + (w\alpha) f^{*\prime} g'] \, dx$$
$$+ [w\beta f^* g]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} [(w\beta)' f^* g + (w\beta) f^{*\prime} g] \, dx + \int_{-\infty}^{\infty} wf^*\gamma g \, dx .$$

We now subtract the corresponding expression for (Lf, g), which may be obtained from the above by simply interchanging f^* and g, and obtain

$$(f, Lg) - (Lf, g) = [w\alpha(f^* g' - f^{*\prime} g)]_{-\infty}^{\infty}$$
$$- \int_{-\infty}^{\infty} [(w\alpha)' - (w\beta)](f^* g' - g f^{*\prime}) \, dx .$$

For L to be Hermitian, this last expression must be equal to zero for *all* f and g. Therefore the necessary and sufficient conditions for L to be Hermitian are

$$1. \quad [w\alpha(f^* g' - g f^{*\prime})]_{-\infty}^{\infty} = 0 , \tag{5.112}$$

$$2. \quad (w\alpha)' = w\beta . \tag{5.113}$$

Clearly, the weight function $w(x)$ plays an important role in the Hermiticity requirement. In fact, the second equation above actually determines $w(x)$ up to a multiplicative constant, since $\alpha(x)$ and $\beta(x)$ are given functions. The first equation illustrates the importance of the endpoint conditions in determining Hermiticity. For example, if the functions on which L operates vanish sufficiently rapidly at the boundaries, then the first condition above is satisfied. The reader can easily show that there are several less stringent conditions which will fulfill condition one. On the other hand, if $w\alpha$ vanishes sufficiently rapidly at both endpoints, then the functions upon which L operates need not satisfy any special boundary conditions and condition (1) will still be satisfied. The requirement that $w\alpha$ vanish sufficiently rapidly at the endpoints can be met either by having $w\alpha$ identically zero outside some infinite interval $[a, b]$ or having $w\alpha$ fall off sufficiently rapidly at arbitrarily large distances. In what follows, we shall have occasion to utilize both possibilities.

Since $w(x)$ has not been specified, we may consider condition (2) as an equation to be solved for $w(x)$. We rewrite it as

$$(w\alpha)' = \frac{\beta}{\alpha} w\alpha . \tag{5.114}$$

A simple integration gives

$$w\alpha = C \cdot \exp\left[\int \frac{\beta}{\alpha} dx\right] \tag{5.115}$$

in any region in which $(w\alpha)'$ is continuous. The integral is an indefinite integral; C is an arbitrary constant. Note in particular that $C = 0$ $(w\alpha \equiv 0)$ is a solution. For future use, we should point out that it is possible to join a solution $w\alpha \not\equiv 0$ to a solution $w\alpha \equiv 0$ at some point $x = a$, if $w\alpha$ is continuous at this point, that is, $w\alpha = 0$ just to the right *and* the left of $x = a$ on the real line.

In the region where $w\alpha \not\equiv 0$, $\alpha(x)$ must either be nonnegative or nonpositive since we require $w(x) > 0$. In the former case, we must choose C positive, in the latter case, C must be negative. Without loss of generality, we can take $\alpha(x) \geq 0$ and $C > 0$. In summary, for any given $\alpha(x)$ and $\beta(x)$ there exists a weight function $w(x)$ given in any region where $w\alpha \not\equiv 0$ by Eq. (5.115). This guarantees that condition (2) is satisfied.

Any equation of the form we are studying,

$$Lu = \alpha(x)u'' + \beta(x)u' + \gamma(x)u = \lambda u\,, \tag{5.116}$$

can be written in the equivalent form

$$\frac{d}{dx}\left[w\alpha\frac{du}{dx}\right] + (\gamma - \lambda)wu = 0 \tag{5.117}$$

by choosing the weight function $w(x)$ according to Eq. (5.113). An equation of the form (5.117), together with the boundary condition stated in Eq. (5.112), is called a *Sturm-Liouville system*. In Eq. (5.46), Legendre's differential equation appears in Sturm-Liouville form. We used this form of Legendre's equation to prove the orthogonality of eigenfunctions belonging to distinct eigenvalues. We shall now give a general proof of this result, valid for any Sturm-Liouville system.

Let λ_m and λ_n $(\lambda_m \neq \lambda_n)$ be two eigenvalues of Eq. (5.117). Then

$$\frac{d}{dx}\left[(w\alpha)\frac{du_m}{dx}\right] + (\gamma - \lambda_m)wu_m = 0\,,$$

and

$$\frac{d}{dx}\left[(w\alpha)\frac{du_n^*}{dx}\right] + (\gamma - \lambda_n)wu_n^* = 0\,.$$

Multiplying the first of these equations by u_n^*, the second by u_m, and forming their difference, we obtain

$$u_n^*\frac{d}{dx}[(w\alpha)u_m'] - u_m\frac{d}{dx}[(w\alpha)u_n^{*\prime}] = -(\lambda_n - \lambda_m)wu_n^*u_m\,.$$

The left-hand side of this expression is a total derivative. Rewriting this, and integrating over both sides, we find that

$$[w\alpha(u_n^*u_m' - u_mu_n^{*\prime})]_{-\infty}^{\infty} = -(\lambda_n - \lambda_m)\int_{-\infty}^{\infty} wu_n^*u_m\,dx\,.$$

The left-hand side vanishes identically by assumption [Eq. (5.112)]; and since

$\lambda_n \neq \lambda_m$, it follows that

$$(u_m, u_n) = \int_{-\infty}^{\infty} u_m^*(x) u_n(x) w(x)\, dx = 0 .$$

Therefore eigenfunctions belonging to distinct eigenvalues are orthogonal with respect to the weight function $w(x)$. This result merely translates into function language the result on the orthogonality of the eigenvectors of Hermitian operators belonging to distinct eigenvalues (Theorem 4.18).

Let us now restrict our attention to *polynomial* Sturm-Liouville systems. That is, we shall assume that the functions $u_n(x)$ are polynomials of degree n. We denote them by $Q_n(x)$. Then

$$LQ_n = \lambda_n Q_n , \tag{5.118}$$

where

$$L = \alpha(x) \frac{d^2}{dx^2} + \beta(x) \frac{d}{dx} + \gamma(x) , \tag{5.119}$$

as before, and

$$\int_{-\infty}^{\infty} Q_n^*(x) Q_m(x) w(x)\, dx = 0 \tag{5.120}$$

if $m \neq n$, in keeping with the above discussion of orthogonality.

For this case, where Q_n is a polynomial of degree n, the real functions $\alpha(x)$, $\beta(x)$, and $\gamma(x)$ must have the form

$$\alpha(x) = \alpha_0 x^2 + \alpha_1 x + \alpha_2 , \tag{5.121}$$

$$\beta(x) = \beta_0 x + \beta_1 , \tag{5.122}$$

$$\gamma(x) = \gamma_0 . \tag{5.123}$$

That this is the most general form possible is a consequence of the fact that the differential equation is of second order and that we are considering only *polynomial* solutions. We prove this by letting $n = 0, 1$, and 2 in Eq. (5.118). Then

$$\gamma Q_0 = \lambda_0 Q_0 \Longrightarrow \gamma = \lambda_0 = \gamma_0 ,$$

$$\beta Q_1' + \gamma Q_1 = \lambda_1 Q_1 \Longrightarrow \beta = \frac{\lambda_1 - \lambda_0}{Q_1'} Q_1 = \beta_0 x + \beta_1 ,$$

$$\alpha Q_2'' + \beta Q_2' + \gamma Q_2 = \lambda_2 Q_2 \Longrightarrow \alpha = \frac{\lambda_2 - \lambda_0}{Q_2''} Q_2 - \frac{\lambda_1 - \lambda_0}{Q_1' Q_2''} Q_1 Q_2'$$
$$= \alpha_0 x^2 + \alpha_1 x + \alpha_2 .$$

We note that β can never be identically zero because if it were, then it follows from Eq. (5.113) that $w\alpha = \text{const}$. But this makes it impossible to satisfy the other requirement for Hermiticity (Eq. 5.112), since polynomials do not vanish at ∞.

Thus there are no polynomial solutions to the Sturm-Liouville problem (Hermitian L) if $\beta(x) \equiv 0$. If we do not insist on L being Hermitian, there are polynomial solutions of all degrees with eigenvalues equal to $n(n-1)\alpha_0$. The eigenvectors will not be orthogonal. The general expression for $\beta(x)$ above shows that $\lambda_1 = \lambda_0$ if $\beta(x) \equiv 0$. Conversely, if $\beta(x) \not\equiv 0$, then $\lambda_1 \neq \lambda_0$ and $\beta_0 \neq 0$ because Q_1 has a leading term proportional to x.

The restriction to polynomial solutions yields only a few of the possible Sturm-Liouville systems. For example, the well-known Bessel functions do not fall into this category. To ensure the Hermiticity of L, we shall satisfy Eq. (5.112) by choosing $w\alpha$ to vanish as $|x| \to \infty$. In fact, since in Eq. (5.112) $w\alpha$ is multiplied by f^*g' and $gf^{*\prime}$ which are now by assumption polynomials in x, we see that $w\alpha$ must actually tend to zero faster than any inverse power of x as $|x|$ tends to infinity. This also follows from the requirement that $(f, f) < \infty$ for all f in our space.

With these assumptions, all the properties of a polynomial Sturm-Liouville system now are determined by the six parameters α_0, α_1, α_2, β_0, β_1, and γ_0. However, there are four degrees of freedom for inessential changes in L:

1. L can be multiplied by a constant C_1. This leaves Q_n unchanged and multiplies the eigenvalues by C_1.
2. The independent variable x may be shifted by some constant C_2. Thus $Q_n(x)$ is replaced by $Q_n(x + C_2)$, and the eigenvalues are unchanged.
3. The independent variable x may be scaled by a constant C_3: $x \to C_3 x$. Thus $Q_n(x)$ is replaced by $Q_n(C_3 x)$, and the eigenvalues are left unchanged.
4. A constant may be added to L, that is, to γ_0. This leaves Q_n unchanged, and λ_n is replaced by $\lambda_n + C_4$. Throughout this section we shall, for convenience, choose C_4 so that $\gamma_0 \equiv 0$.

With γ_0 fixed (equal to zero), we have only five parameters left which characterize the Sturm-Liouville system (α_0, α_1, α_2, β_0, and β_1); but there are three inessential degrees of freedom remaining. Thus, in fact, the polynomial Sturm-Liouville system is completely characterized by $5 - 3 = 2$ parameters. We now consider the following three cases:

1. $\alpha(x)$ is quadratic;
2. $\alpha(x)$ is linear ($\alpha_0 = 0$), and
3. $\alpha(x)$ is a constant ($\alpha_0 = 0 = \alpha_1$).

Our discussion will completely exhaust the possible *polynomial* solutions to the second-order Sturm-Liouville differential equation. The reader will never be able to discover a new system of Sturm-Liouville polynomials and have them named after himself!

CASE 1. $\alpha(x)$ is quadratic. In this case, Eq. (5.115) is

$$w\alpha = C \cdot \exp\left[\int \frac{\beta_0 x + \beta_1}{\alpha_0 x^2 + \alpha_1 x + \alpha_2}\, dx\right], \tag{5.124}$$

where $\alpha_0 \neq 0$. There are two subcases to distinguish, depending on whether $\alpha(x)$ has real or complex roots. Let us consider the case of complex roots first, and for convenience pick C_1 so that $\alpha_0 = 1$. Then $\alpha(x) = (x - \kappa)(x - \kappa^*)$, where κ and κ^* are the two complex roots of $\alpha(x)$ (the asterisk, as usual, denotes complex conjugation). The integral in Eq. (5.124) is

$$w\alpha = C[\alpha(x)]^{\beta_0/2} \exp\left[\frac{\beta_1 + \beta_0 \operatorname{Re}\kappa}{|\operatorname{Im}\kappa|} \tan^{-1}\left(\frac{x - \operatorname{Re}\kappa}{|\operatorname{Im}\kappa|}\right)\right],$$

where $\operatorname{Re}\kappa$ and $\operatorname{Im}\kappa$ denote, respectively, the real and imaginary parts of κ. Since $\alpha(x)$ is a polynomial in x, $w\alpha$ cannot vanish faster than some inverse power of x as x tends to infinity (the precise power being determined by β_0), so it fails to satisfy our previously derived endpoint requirements, Eq. (5.112). Also, $w\alpha$ is never zero, so we cannot join it onto a solution $w\alpha \equiv 0$ outside some finite interval. Thus our only hope is to consider the case where $\alpha(x)$ has two real roots. For convenience in this case, let us choose the three inessential parameters so that $\alpha_0 = -1$ and the roots of $\alpha(x)$ are at $x = 1$ and $x = -1$. Thus $\alpha(x) = 1 - x^2$. Furthermore, let us adopt as a notational convenience the relations $\beta_1 = q - p$, $\beta_0 = -(p + q + 2)$; this particular convention is made so that $w(x)$ will work out in a nice form; p and q are real but need not be integers. Thus

$$\frac{\beta}{\alpha} = \frac{-(p + q + 2)x + q - p}{1 - x^2} = \frac{q + 1}{1 + x} - \frac{p + 1}{1 - x},$$

and by doing the simple integral in Eq. (5.115) we get

$$w\alpha = C(1 + x)^{q+1}(1 - x)^{p+1} \qquad \text{or} \qquad w(x) = C(1 + x)^q(1 - x)^p.$$

Once again it turns out that, at large distances, $w\alpha$ can tend to zero only as fast as some fixed inverse power of x (depending on the specific values of p and q), so in this form our endpoint requirements cannot be satisfied. However, if $q > -1$ and $p > -1$, $w\alpha$ vanishes at $x = 1$ and $x = -1$, so we can join the above expression for $w(x)$ on $[-1, 1]$ to the solution $w(x)$ identically equal to zero outside $[-1, 1]$. (Note that α_0 and β_0 always have the same sign, a fact we shall use subsequently.) Therefore all the objections we raised in the case of complex roots are overcome, and we have an acceptable weight function for the interval $[-1, 1]$:

$$w(x) = (1 + x)^q(1 - x)^p,$$

where we have set $C = 1$.

The $Q_n(x)$ that correspond to these choices of the essential parameters for the case of $\alpha(x)$ quadratic are called *Jacobi polynomials* of index p, q, and they shall be designated $J_n^{(p,q)}(x)$. (The upper indices distinguish them from Bessel functions.) By making a linear change of the independent variable, they may be recast in another form. Letting $x \to 1 - 2x$ in the differential equation (5.119), we obtain

$$\alpha(x) = x(1 - x) \geq 0 \text{ in } [0, 1], \qquad \beta(x) = -(p + q + 2)x + (p + 1).$$

Thus the left-hand endpoint is now $x = 0$; $\beta_1 = \beta(0) = p + 1$; and

$$\beta_0 = \beta(1) - \beta(0) = -(p + q + 2)$$

is unchanged. It follows that $\alpha w = x^{p+1}(1 - x)^{q+1}$, so $p, q > -1$, as before. The weight function is, therefore, now

$$w(x) = x^p(1 - x)^q .$$

The set of polynomials which are orthogonal to the interval $[-1, 1]$ with respect to the weight function $w(x) = (1 - x)^p(1 + x)^q$, and the polynomials orthogonal on $[0, 1]$ with respect to $w(x) = x^p(1 - x)^q$, differ only by a linear change of variable. Both are referred to as the Jacobi polynomials of index p, q.

We now deal with a number of special cases of the Jacobi polynomials as defined on $[-1, 1]$. For $p = q = m (m > -1)$, $w(x) = (1 - x^2)^m$. The corresponding polynomials are called *ultraspherical* or *Gegenbauer polynomials* of integral index m. We designate them by

$$G_n^m(x) = J_n^{(m, m)}(x) .$$

Note that $p = q$ implies $\beta_1 = 0$, which means that the operator L is unaltered by the change of variable $x \to -x$. If we call P the operator which changes x into $-x$, then the invariance of L under $x \to -x$ may be expressed in operator notation as $PLf(x) = LPf(x)$ for all $f(x)$, so

$$PL = LP ,$$

which according to Theorem (4.22) tells us that any eigenfunction of L can be chosen to be simultaneously an eigenfunction of P. Since $P^2 f(x) = f(x)$ for all $f(x)$, we see that $P^2 = I$, the identity operator. This places severe limitations on the eigenvalues of P. Consider $Pf(x) = \lambda f(x)$. This implies

$$P^2 f(x) = \lambda Pf(x) = \lambda^2 f(x) = f(x) ,$$

since $P^2 = I$. Thus $\lambda^2 = 1$, so $\lambda = \pm 1$. This means that any Gegenbauer polynomial can be characterized by a subscript $+$ or $-$, depending on whether it is an eigenfunction of P with eigenvalue $+1$ or -1 (see Problem 5.15).

P is called the *parity operator*. It plays a fundamental role in many areas of physics. According to the above discussion, one speaks of eigenvectors of positive or negative parity. Since the leading term in $G_n^m(x)$ is proportional to x^n and since $Px^n = (-1)^n x^n$, we see that the Gegenbauer polynomials have positive or negative parity depending on whether n is even or odd. If n is even, $G_n^m(x)$ contains *only* even powers of x, whereas if n is odd, $G_n^m(x)$ contains *only* odd powers of x. The Gegenbauer polynomials occur in the solution of certain relativistically invariant equations in the quantum theory of elementary particles.

If we consider the Jacobi polynomials with $p = q = -\frac{1}{2}$, then $w(x) = (1 - x^2)^{-1/2}$. We designate these polynomials

$$T_n(x) \equiv J_n^{(-1/2, -1/2)}(x) .$$

The orthogonality relation is

$$(T_{n_1}, T_{n_2}) = \int_{-1}^{1} T_{n_1}^*(x)\, T_{n_2}(x)(1 - x^2)^{-1/2}\, dx = 0\,, \qquad (n_1 \neq n_2).$$

If we set $x = \cos\theta$, we get the alternate expression

$$(T_{n_1}, T_{n_2}) = \int_{-\pi}^{\pi} T_{n_1}^*(\cos\theta)\, T_{n_2}^*(\cos\theta)\, d\theta = 0\,.$$

The only set of polynomials in $\cos\theta$ which obeys this relationship on $[-\pi, \pi]$ is

$$T_n(\cos\theta) \propto \cos n\theta\,.$$

Therefore in the x-variable on $[-1, 1]$,

$$T_n(x) \propto \cos(n \cos^{-1} x)\,,$$

so this set of polynomials is determined explicitly up to a constant factor. These polynomials are known as the *Tschebycheff polynomials*. It is probably correct to say that, until now, no two authors have ever spelled their name the same way.

The most familiar of all the polynomials, the Legendre polynomials, are sure to be found somewhere in this cascade of special cases. We see that they result if we take $p = q = m = 0$. Then $w(x) = 1$; thus $J_n^{(0,0)}(x) = P_n(x)$ up to constant factor.

There is a theorem (Problem 20) which tells us that if $\{Q_n\}$ is an orthonormal system of Sturm-Liouville polynomials, then the set $\{Q_n',\ n > 1\}$ is a system of orthogonal polynomials with weight function $w\alpha$. In the case of the Legendre polynomials, where $w = 1$ in $[-1, 1]$ and $\alpha = (1 - x^2)$, this result, applied μ times, tells us immediately that the set of polynomials $(d^\mu/dx^\mu)P_n$ is orthogonal on $[-1, 1]$ with respect to the weight function $\alpha^\mu = (1 - x^2)^\mu$. The uniqueness of sets of orthonormal polynomials on given intervals means that these polynomials are none other than the Gegenbauer polynomials:

$$G_{n-\mu}^{\mu} = \frac{d^\mu}{dx^\mu} P_n\,,$$

which is the only orthonormal set of polynomials with weight function $w = (1 - x^2)^\mu$ on $[-1, 1]$.

It also follows that the set of *functions*

$$P_n^\mu(x) = (1 - x^2)^{\mu/2} \frac{d^\mu}{dx^\mu} P_n(x)$$

is orthogonal with weight function 1 on $[-1, 1]$. These functions are the associated Legendre functions.

CASE 2. $\alpha(x)$ is a linear function. Under these circumstances, $\alpha(x) = x + \alpha_2$, where we have chosen C_1 so that $\alpha_1 = 1$. Also, $\alpha(x)$ has one real zero at $x = -\alpha_2$; adjusting our constant C_2, let us shift the zero to the origin, so finally

$\alpha(x) = x$. Thus

$$w\alpha = C \cdot \exp\left[\int \frac{\beta_0 x + \beta_1}{x}\, dx\right]$$
$$= C \cdot \exp\left(\beta_0 x + \beta_1 \ln x\right) = C x^{\beta_1} e^{\beta_0 x}.$$

If we choose $\beta_0 < 0$, then $w\alpha \to 0$ as $x \to +\infty$, but $w\alpha$ grows exponentially as $x \to -\infty$. Thus to satisfy the boundary conditions, $w(x)$ must be chosen to vanish identically for all $x < a$, if possible. a is still to be determined. If we choose $\beta_1 > 0$, then $w\alpha = 0$ at $x = 0$ and the integral of $w\beta$ over an interval of length ϵ around $x = 0$ goes to zero as ϵ tends to zero, so we may correctly join our identically zero solution (for $x < 0$) to the above nontrivial solution at this point. By adjusting C_3 we may set $\beta_0 = -1$ for convenience. (Note that if we choose $\beta_0 > 0$, this would just reverse the situation, and $w\alpha$ would have to vanish on the positive x-axis. The choice $\beta_0 < 0$ is conventional.) Calling $\beta_1 = s + 1$, $s > -1$, we have

$$w(x) = x^s e^{-x}$$

for positive x and $w(x) = 0$ for negative x. The Q_n that correspond to these conditions are the *associated Laguerre polynomials* of order s. They are designated by $L_n^s(x)$. If $s = 0$, the functions $L_n^0(x) \equiv L(x)$ are called the *Laguerre polynomials*. The presence of only one free parameter (s) reflects the fact that $\alpha(x)$ is *linear* in x: $\alpha_0 = 0$, and since three of the four remaining coefficients in $\alpha(x)$ and $\beta(x)$ may be chosen arbitrarily, there is just a single free parameter left.

CASE 3. $\alpha(x)$ is a constant. Here $\alpha_0 = 0 = \alpha_1$, and by a judicious choice of C_1 we can arrange that $\alpha_2 = 1$. Thus

$$w\alpha = C \cdot \exp\left[\int (\beta_0 x + \beta_1)\, dx\right] = C \cdot \exp\left(\frac{\beta_0}{2} x^2 + \beta_1 x\right)$$
$$= C' \cdot \exp\left[\frac{\beta_0}{2}\left(x + \frac{\beta_1}{\beta_0}\right)^2\right].$$

If we choose $\beta_0 < 0$, then $w\alpha = w$ falls off exponentially as $|x| \to \infty$, so this expression for $w(x)$ is acceptable as it stands; choosing C_2 and C_3 appropriately, we can make $\beta_0 = -2$ and $\beta_1 = 0$, so setting $C' = 1$, we get

$$w(x) = e^{-x^2}.$$

The corresponding $Q_n(x)$ are the Hermite polynomials. We have now run through *all* the possibilities: there are no more Sturm-Liouville polynomials that are solutions of second-order differential equations. Table 5.1 summarizes the results obtained thus far.

We now work out the eigenvalues belonging to each of the three basic cases. We need only collect the terms in x^n in the differential equation. This gives

$$\lambda_n = \alpha_0 n(n-1) + \beta_0 n = n(\alpha_0 n + \beta_0 - \alpha_0). \tag{5.125}$$

It follows immediately that the eigenvalues are spaced linearly for the Hermite

and Laguerre equations (for which $\alpha_0 = 0$), and quadratically for Jacobi's equation, and all the special cases of it.

In particular, for Hermite's equation we have $\lambda_n = -2n$, and for Laguerre's equation we have $\lambda_n = -n$. For Jacobi's equation the eigenvalues are $\lambda_n = -n(n + p + q + 1)$ regardless of whether the equation is the one whose solutions are orthogonal on $[0, 1]$ or the equation whose solutions are orthogonal on $[-1, 1]$. The Gegenbauer special case ($p = q = m$) has eigenvalues $\lambda_n = -n(n + 2m + 1)$. The two further specialized cases, Tschebycheff's ($m = -\frac{1}{2}$) and Legendre's ($m = 0$) equations, have eigenvalues $\lambda_n = -n^2$ and $\lambda_n = -n(n + 1)$, respectively. It should be clear that whether the eigenvalues are positive or negative is merely a matter of the specific sign conventions adopted elsewhere in the treatment.

We conclude this discussion of polynomial Sturm-Liouville systems by proving a general Rodrigues formula which provides explicit expressions for all the polynomials up to a constant factor. The general formula is

$$Q_n(x) = K_n \frac{1}{w} \frac{d^n}{dx^n} (\alpha^n w) , \tag{5.126}$$

where the choice of the constant K_n depends on the physical application. The proof we shall give is a general version of the one in Section 5.5 for the Legendre polynomials. Any set of polynomials is uniquely determined by the requirement that $Q_n(x)$ is a polynomial of nth degree that satisfies the orthogonality relation

$$\int_a^b Q_m^* Q_n w \, dx = 0 , \qquad \text{for } n \neq m . \tag{5.127}$$

Thus we only have to show that the $Q_n(x)$, as given by the generalized Rodrigues formula [Eq. (5.126)], are polynomials of degree n, and satisfy Eq. (5.127) and we shall have proved Eq. (5.126).

We first prove a preliminary result: If $f(x) = \alpha^k w r(x)$, where $r(x)$ is a polynomial of degree l, then $f'(x) = \alpha^{k-1} w s_{l+1}(x)$, where $s_{l+1}(x)$ is a polynomial of degree $l + 1$. To prove this we simply differentiate $f(x)$:

$$\begin{aligned} f'(x) &= (\alpha^{k-1})'(\alpha w) r + \alpha^{k-1}(\alpha w)' r + \alpha^{k-1}(\alpha w) r' \\ &= (k-1)\alpha^{k-2}\alpha'(\alpha w) r + \alpha^{k-1}(\beta w) r + \alpha^{k-1}(\alpha w) r' \\ &= \alpha^{k-1} w [\beta r + (k-1)\alpha' r + \alpha r'] \\ &= \alpha^{k-1} w [r_0(\beta_0 + (2k + l - 2)\alpha_0) x^{l+1} + \cdots] \\ &= \alpha^{k-1} w s_{l+1}(x) . \end{aligned}$$

We have written $r(x) = r_0 x^l + \cdots$. Since $\beta_0 \neq 0$, and since α_0 and β_0 have the same sign if $\alpha_0 \neq 0$, then the expression in brackets is a polynomial of degree $l + 1$ if $k \geq 1$ which is the only situation of interest. From this result applied n times to $\alpha^n w$, ($r = 1$), we infer that

$$\frac{d^n}{dx^n} (\alpha^n w) = \alpha^0 w t(x) = w t_n(x) ,$$

Table 5.1

THE STURM-LIOUVILLE SYSTEMS OF ORTHOGONAL POLYNOMIALS

Name of polynomials	Differential equation	α_0	α_1	α_2	β_0	β_1	$w\alpha$	w	$[a, b]$
(I) α *quadratic*									
Jacobi	$(1-x^2)u'' + [-(p+q+2)x + (q-p)]u' + n(n+p+q+1)u = 0$	-1	0	1	$-(p+q+2)$	$q-p$	$(1-x)^{p+1}(1+x)^{q+1}$	$(1-x)^p(1+x)^q$	$[-1, 1]$
Jacobi	$(x-x)u'' + [-(p+q+2)x + (p+1)]u' + n(n+p+q+1)u = 0$	-1	1	0	$-(p+q+2)$	$p+1$	$x^{p+1}(1-x)^{q+1}$	$x^p(1-x)^q$	$[0, 1]$
Gegenbauer	$(1-x^2)u'' - 2(m+1)xu' + n(n+2m+1)u = 0$	-1	0	1	$-2(m+1)$	0	$(1-x^2)^{m+1}$	$(1-x^2)^m$	$[-1, 1]$
Tschebycheff	$(1-x^2)u'' - xu' + n^2u = 0$	-1	0	1	-1	0	$(1-x^2)^{1/2}$	$(1-x^2)^{-1/2}$	$[-1, 1]$
Legendre	$(1-x^2)u'' - 2xu' + n(n+1)u = 0$	-1	0	1	-2	0	$1-x^2$	1	$[-1, 1]$
(II) α *linear*									
Laguerre	$xu'' - [x-(s+1)]u' + nu = 0$	0	1	0	-1	$s+1$	$x^{s+1}e^{-x}$	x^se^{-x}	$[0, \infty]$
(III) α constant									
Hermite	$u'' - 2xu' + 2nu = 0$	0	0	1	-2	0	e^{-x^2}	e^{-x^2}	$[-\infty, \infty]$

where $t(x)$ is a polynomial of degree n. It follows from Eq. (5.126) that $Q_n(x)$ is indeed a polynomial of degree n.

We can now prove the orthogonality of the $Q_n(x)$ with respect to $w(x)$. We have

$$\int_{-\infty}^{\infty} Q_m^*(x)Q_n(x)w(x)\,dx = \int_{-\infty}^{\infty} Q_m^*(x)\,\frac{d^n}{dx^n}(\alpha^n w)\,dx$$

$$= \sum_{k=0}^{n-1}\left[(-1)^k\left(\frac{d^k}{dx^k}Q_m^*\right)\frac{d^{n-1-k}}{dx^{n-1-k}}(\alpha^n w)\right]_{-\infty}^{\infty}$$

$$+ (-1)^n\int_{-\infty}^{\infty}\left(\frac{d^n}{dx^n}Q_m^*\right)\alpha^n w\,dx\,. \tag{5.128}$$

Here we have integrated by parts n times. If $n > m$, the integral vanishes, so orthogonality will be satisfied if we can show that the terms $d^l/dx^l(\alpha^n w)$, $0 \leq l \leq n-1$, all vanish at the endpoints faster than any inverse power of x. But this follows from the preliminary result which tells us that

$$\frac{d^l}{dx^l}(\alpha^n w) = \alpha^{n-l}wv_l(x)\,,$$

where $v_l(x)$ is a polynomial of degree l. Since by assumption αw vanishes at the endpoints faster than any inverse power of x, so will $\alpha^{n-l}wv_l(x)$ for all $0 \leq l \leq n-1$, and the proof of Eq. (5.126) is complete.

In carrying out this proof we have shown that, in general,

$$\int_{-\infty}^{\infty} R(x)Q_n(x)w(x)\,dx = (-1)^n\int_{-\infty}^{\infty}\frac{d^nR}{dx^n}\alpha^n w\,dx\,. \tag{5.129}$$

This relation can be useful in the evaluation of integrals. In particular, if we take $R(x) \equiv Q_n(x)$, the normalization integral is

$$\int_{-\infty}^{\infty} Q_n^2(x)w(x)\,dx = (-1)^n n!\,q_n\int_{-\infty}^{\infty}\alpha^n w\,dx\,,$$

where q_n is the leading coefficient of $Q_n(x)$, as given in Eq. (5.126).

We can now use this generalized Rodrigues formula to obtain expressions for the polynomials we have considered. The Rodrigues formula will only provide us with expressions up to constant factors, which may then be selected in several ways. In some books they are selected in such a way as to make the coefficient of the highest power in $Q_n(x)$ equal to one for all n. Sometimes they are chosen so as to normalize the polynomials with respect to the corresponding weight function; and sometimes in still other ways. For example, the Legendre polynomials defined in Eq. (5.38) are not normalized on $[-1, 1]$. The normalized Legendre polynomials are given by $(n+\frac{1}{2})^{1/2}P_n(x)$. One must beware of the multiple inconsistent conventions that exist in the literature in defining the various sets of orthogonal polynomials. We shall not put in explicit expressions for the constant factors in the list that follows, except for the Legendre and Hermite polynomials, which we have worked out earlier.

Now, using the generalized Rodrigues formula (Eq. 5.126), we obtain for the Jacobi polynomials,

$$J_n^{(p,q)}(x) = A_n^{(p,q)}(1-x)^{-p}(1+x)^{-q}\frac{d^n}{dx^n}\left[(1-x^2)^n(1-x)^p(1+x)^q\right], \quad (5.130)$$

where $A_n^{(p,q)}$ is a constant factor depending on p, q, and n.

For the Gegenbauer polynomials, $p = q = m$, so this becomes

$$G_n^m(x) = B_n^m(1-x^2)^{-m}\frac{d^n}{dx^n}(1-x^2)^{n+m}, \quad (5.131)$$

where B_n^m is a constant factor depending on m and n.

For the Tschebycheff ($m = -\frac{1}{2}$) and Legendre ($m = 0$) polynomials, we have

$$T_n(x) = C_n(1-x^2)^{1/2}\frac{d^n}{dx^n}(1-x^2)^{n-1/2}, \quad (5.132)$$

where C_n is a constant factor depending on n;

$$P_n(x) = \frac{(-1)^n}{2^n n!}\frac{d^n}{dx^n}(1-x^2)^n. \quad (5.133)$$

For the associated Laguerre polynomials, we obtain

$$L_n^s(x) = D_n^s x^{-s}e^x\frac{d^n}{dx^n}(x^{n+s}e^{-x}), \quad (5.134)$$

where D_n^s is a constant factor which may depend on s and n; conventions vary widely in its choice. The Laguerre polynomials are the case $s = 0$:

$$L_n(x) = D_n e^x\frac{d^n}{dx^n}(n^x e^{-x}); \quad (5.135)$$

here D_n depends only on n.

Finally, the Hermite polynomials are given by

$$H_n(x) = (-1)^n e^{x^2}\frac{d^n}{dx^n}e^{-x^2}. \quad (5.136)$$

We close this section by showing that all the Sturm-Liouville polynomials form complete sets in Hilbert space when the inner product in Hilbert space is given by

$$(f, g) \equiv \int_{-\infty}^{\infty} f^*(x)g(x)w(x)\,dx,$$

with $w(x)$ given for the various polynomials in Table 5.1. In this discussion we shall take into account the Hermite and Laguerre polynomials which are defined on infinite intervals and therefore do not fall within the scope of Weierstrass's Theorem. The proof of completeness is based on Theorem 5.2 which says that if a set of orthonormal functions is closed, then it is complete.

Completeness Theorem. The orthonormal set of Sturm-Liouville polynomials $\{Q_n(x)\}$ is complete in Hilbert space.

Proof. We shall show that the set $\{Q_n(x)\}$ is closed. Suppose there is some function $f(x)$ orthogonal to all the $Q_n(x)$, so that $(f, Q_n) = 0$ for all n. Since x^m can be written as a finite linear combination of the $Q_n(x)$, the first $m + 1$ to be exact, we see immediately that $(x^m, f) = 0$ for all m. Now consider

$$g(k) \equiv (e^{ikx}, f) = \int_{-\infty}^{\infty} f(x)w(x)e^{-ikx}\,dx\,,$$

for all real k. Because of the presence of $w(x)$ in the inner product, e^{-ikx} will belong to Hilbert space for any $w(x)$ appropriate to a Sturm-Liouville polynomial; so will x^m. Now, if we expand e^{-ikx} as

$$e^{-ikx} = \sum_{m=0}^{\infty} \frac{(-ik)^m}{m!} x^m$$

and use $(x^m, f) = 0$, we see that $g(k) = 0$ for all k. But $g(k)$ is precisely the Fourier transform of $f(x)\ w(x)$. Hence by our discussion in Section 5.7, $f(x)\ w(x) = 0$ almost everywhere. Thus we conclude that on any interval where $w(x) \neq 0$, $f(x) = 0$ almost everywhere. We are only interested in $f(x)$ on intervals where $w(x) \neq 0$, since we can define $f(x)$ to be anything we like where $w(x) = 0$. Another way to put this is to say that the Hilbert space is really defined only on the interval where $w(x) \neq 0$, since when $w(x) = 0$ on some finite interval, the inner product gives no structure there at all.

In summary, we conclude that $f(x) = 0$ almost everywhere, so the set $\{Q_n(x)\}$ is closed, and therefore it is a complete orthonormal set. QED

These sets of orthogonal polynomials have many additional polynomials. For example, using the generalized Rodrigues formula, one can derive recursion relations for the various sets of polynomials. Also, the zeros of the polynomials can be characterized in some detail. The interested reader is referred to Lyusternik or Wilf. In Section 6.10 we shall go on to establish the equivalence of the explicit Rodrigues representation of some of these polynomials and their implicit representation in terms of a generating function.

All these sets of functions are solutions to differential equations, as are the common trigonometric functions. All of them are related by various identities, just as the sine and cosine functions are. It is useful to think about them in this light. We study sines and cosines first, but these particular functions are not essentially different from Legendre or Hermite polynomials, or any of the other special functions of physics. They are more familiar to us only because we can apply them to the measurement of distances and because they turn up in macroscopic simple harmonic motion. If we imagine very small intelligent beings who do not measure large distances and instead first investigate quantum effects, it is reasonable to suppose that they discover Hermite polynomials at an early date by investigating the quantum oscillator. After centuries of work,

they spend half their Gross National Product to build a (to us) macroscopic oscillator (to them it will be absolutely gigantic), and, in the course of this project, they discover the trigonometric functions. To them, sines and cosines will seem complicated and troublesome compared to the old familiar Hermite polynomials.

5.11 A MATHEMATICAL FORMULATION OF QUANTUM MECHANICS

In this and in the preceding chapters we have developed many of the mathematical ideas used in the formulation of quantum mechanics, namely, the concepts of vector spaces and, in particular, Hilbert space. We shall now write and interpret the axioms of quantum mechanics in terms of these mathematical tools. What we give here is not a rigorous axiomatization, but rather an introduction to a profound and beautiful subject. For more complete treatments, the reader should see J. M. Jauch's book, *Foundations of Quantum Mechanics*, or J. Von Neumann's classic work, *Mathematical Foundations of Quantum Mechanics*.

The mathematical setting for quantum mechanics is Hilbert space, that is, a complete complex inner-product space. As we have seen, the elements of this space are functions. To give a mathematical formulation of quantum mechanics, we must also make use of *operators* on Hilbert space. In our discussion of the properties of these operators, we will rely heavily on the results of Chapters 3 and 4. We have already noted that the validity of many of the theorems in Chapters 3 and 4 is not restricted to finite-dimensional spaces, and we shall make use of these theorems where they are applicable. We shall also use them to arrive intuitively at certain important generalizations of finite-dimensional results which will be central to our treatment.

We begin by discussing the notion of a *projection* operator (Problem 4.17), which will prove very useful in the mathematical formulation of quantum mechanics. We do this in the context of some familiar finite-dimensional results. Consider an Hermitian operator, A, whose eigenvalues and eigenvectors are $\{\lambda_n\}$ and $\{\phi_n\}$ respectively ($n = 1, 2, \cdots, N$). We assume, for notational convenience, that λ_n is nondegenerate. We now define the operator I by

$$I\psi \equiv \sum_{n=1}^{N} \phi_n(\phi_n, \psi)$$

for any ψ in the finite-dimensional space in question. According to the completeness theorem (Theorem 4.20), this sum is equal to ψ, so that the operator defined in this way is in fact the identity operator, as we had anticipated by our choice of notation. If we define P_n by

$$P_n\psi \equiv \phi_n(\phi_n, \psi) \ , \tag{5.137}$$

then I can be written as

$$I = \sum_{n=1}^{N} P_n \ .$$

The reader can easily verify that $P_n^2 = P_n$, $P_n^\dagger = P_n$, and $P_n P_m = 0$ if $m \neq n$. P_n is a *projection operator*, so called because it projects any vector into the one-dimensional subspace of V spanned by ϕ_n.

What can we say about the operator, B, defined by

$$B \equiv \sum_{n=1}^{N} \lambda_n P_n \, ?$$

Since

$$B\phi = \sum_{n=1}^{N} \lambda_n \phi_n (\phi_n, \phi) \ ,$$

it is immediately clear that B has eigenvalues $\{\lambda_n\}$ with corresponding eigenvectors $\{\phi_n\}$. It must therefore be equal to A (the reader will find it instructive to prove this). In summary, we have

$$I = \sum_{n=1}^{N} P_n \ , \qquad A = \sum_{n=1}^{N} \lambda_n P_n \ . \tag{5.138}$$

Thus we may restate the completeness theorem as follows: For every self-adjoint operator, A, there exists a set of orthogonal projections, $\{P_n\}$, such that Eqs. (5.138) are satisfied. By *orthogonal* projections, we mean projections which satisfy $P_n P_m = 0$ for $m \neq n$. It is in this language that one can generalize the completeness theorem to infinite-dimensional spaces. However, it turns out that only for a special type of infinite-dimensional operator (the completely continuous type, which we will discuss in Chapters 8 and 9) do Eqs. (5.138) generalize directly.

In an infinite-dimensional space, there arises the possibility of "eigenvalues" which are infinitely closely spaced, so that the above formulation needs some alteration. To get the flavor of what these alterations involve, we must say a word about an important generalization of the Riemann integral—the Stieltjes integral. We define

$$\int_a^b f(x)\, dg(x) \equiv \lim \sum_{i=1}^{N} f(\bar{x}_i)[g(x_{i+1}) - g(x_i)] \ , \tag{5.139}$$

where the points $x_1, x_2, \cdots, x_{n+1}$ represent a partition of the interval $[a, b]$, and $\bar{x}_i$ is any point in the ith interval. "Lim" denotes the passage to very small partition intervals. Obviously, if $g(x)$ is a nice differentiable function, we may write

$$\int_a^b f(x)\, dg(x) = \int_a^b f(x) \frac{dg}{dx}\, dx \ ,$$

and we get back to the familiar Riemann integral. However, $g(x)$ need not be differentiable or even continuous. Consider the function

$$g(x) = \begin{cases} 1 & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0 \ . \end{cases}$$

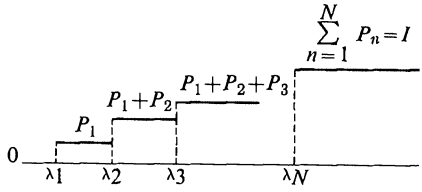

Fig. 5.5. A schematic plot of $E(\lambda)$ for the special case of a finite-dimensional self-adjoint operator whose eigenvalues are $\lambda_1, \lambda_2, \cdots, \lambda_N$.

According to the definition of Eq. (5.139), we have simply

$$\int_{-\infty}^{\infty} f(x)\, dg(x) = f(0) ,$$

since only an arbitrarily small interval around $x = 0$ can contribute to the sum in Eq. (5.139). This result is a δ-function type equation, which is not surprising since in the imprecise δ-function language,

$$\frac{d}{dx} g(x) = \delta(x) .$$

However, in the Stieltjes treatment, we never go beyond the domain of very traditional mathematics.

Now let us define an operator-valued function $E(\lambda)$ as follows:

$$E(\lambda) = \begin{cases} 0 & \text{for } \lambda < \lambda_1 , \\ \sum_{n=1}^{\nu} P_n & \text{for } \lambda_\nu \leqslant \lambda < \lambda_{\nu+1} , \quad \nu = 1, \cdots, N-1 , \\ \sum_{n=1}^{N} P_n & \text{for } \lambda \geq \lambda_N ; \end{cases}$$

the P_n and λ_n are defined above. This operator-valued function is "plotted" schematically in Fig. 5.5. Note that

$$E(-\infty) = 0 \qquad \text{and} \qquad E(+\infty) = I.$$

The reader can easily verify that

$$E(\lambda_1) E(\lambda_2) = E(\bar{\lambda}),$$

where $\bar{\lambda}$ is the smaller of λ_1 and λ_2 (this follows from the fact that the P_n are *orthogonal* projections), and that

$$[E(\lambda)]^2 = E(\lambda).$$

Using the definition of the Stieltjes integral (Eq. 5.139), we see that

$$\int_{-\infty}^{\lambda} dE(\lambda) = \sum_{n=1}^{\nu} P_n$$

if λ lies between λ_ν and $\lambda_{\nu+1}$. The only contributions to this integral come from the points at which $E(\lambda)$ changes discontinuously; at each such point, λ_l (which must be an eigenvalue of A by construction), we pick up the projection operator onto the one-dimensional subspace of V which is spanned by ϕ_l. In this new way of writing things, the first of Eqs. (5.138) becomes

$$I = \int_{-\infty}^{\infty} dE(\lambda) \ ,$$

and the second of Eqs. (5.138) can be written as

$$A = \int_{-\infty}^{\infty} \lambda dE(\lambda) \ .$$

Thus we have succeeded in making the simple finite-dimensional completeness theorem look extremely complicated.

Now let us look at the virtues of this complexity. Consider the familiar problem

$$\frac{d^2\phi_n(x)}{dx^2} = \lambda_n \phi_n(x) \ .$$

We know that with periodic boundary conditions on $[-L, L]$

$$\phi_n = \frac{1}{\sqrt{2L}} e^{i(n\pi/L)x} \ , \qquad \lambda_n = \frac{n^2\pi^2}{L^2} \ , \qquad n = 0, \pm 1, \pm 2, \cdots$$

As we let L become very large, the eigenvalues cluster very close together, so that in Fig. 5.5 the intervals of constancy of $E(\lambda)$ become increasingly small. We may imagine that in this limit $E(\lambda)$ tends to a *continuous* operator-valued function. It is this phenomenon which is given the name of "continuous spectrum," and it is very familiar in quantum mechanics—indeed, it may be said to lie at the heart of quantum mechanics.

Having introduced the function $E(\lambda)$ on a finite-dimensional space, we will generalize, without proof, to an infinite-dimensional Hilbert space. The following theorem is one of the most powerful results obtained in mathematics and is the key to a rigorous formulation of quantum mechanics.

Spectral Theorem. To every self-adjoint operator, A, on a Hilbert space, H, there corresponds an operator-valued function (unique), $E(\lambda)$, such that

1. $E(\lambda_1)E(\lambda_2) = E(\bar{\lambda}) \ ; \quad \bar{\lambda} = \min(\lambda_1, \lambda_2) \ ,$
2. $\lim_{\lambda \to -\infty} E(\lambda) = 0 \ , \quad \lim_{\lambda \to \infty} E(\lambda) = I \ ,$
3. $I = \int_{-\infty}^{\infty} dE(\lambda) \ ,$
4. $A = \int_{-\infty}^{\infty} \lambda dE(\lambda) \ .$

$E(\lambda)$ is called a *resolution of the identity* belonging to A. The set of points at which $E(\lambda)$ is nonconstant is referred to as the *spectrum* of A. For all λ, $E(\lambda)$ commutes with A and with any transformation which commutes with A.

Obviously, the spectral theorem reduces to the familiar completeness theorem in a finite-dimensional space. There are several remarks which should be made at this point. First, note that we have not been precise in saying what we mean by the limit of a sequence of operators (see Part 2 of the above theorem). We shall say more on this point in Chapters 8 and 9, but it is worth remarking that in nonrelativistic quantum mechanics, the spectrum is always bounded below (i.e., we cannot have bound states with arbitrarily large negative energy) so that the limit $\lambda \to -\infty$ is unambiguous. Second, we remark that if A is an *unbounded* operator (i.e., if there are elements of H on which A is *not defined*) then self-adjoint does not mean merely $(Af, g) = (f, Ag)$ whenever Af and Ag are defined. We will not dwell on this point; suffice it to say that all interesting operators which occur in nonrelativistic quantum mechanics are self-adjoint.

With these mathematical preliminaries out of the way, we now proceed to discuss the axioms of quantum mechanics.

Axiom I. Any physical system is completely described by a normalized vector ψ (the *state vector* or *wave function*) in Hilbert space. All possible information about the system can be derived from this state vector by rules which will be given in the following axioms.

Axiom II. To every physical observable there corresponds a self-adjoint operator on Hilbert space.

Examples of observables are position, momentum, energy, angular momentum, spin, etc. This axiom brings with it the powerful machinery associated with the spectral theorem. Notice that Axiom II does not specify how one is to discover the self-adjoint operator which corresponds to a given observable. It merely says that one exists. In practice, the use of classical expressions such as $\mathbf{L} = \mathbf{r} \times \mathbf{p}$, $E = p^2/2m + V(\mathbf{r})$, etc., along with the operator substitutions $\mathbf{p} \to -i\hbar\nabla$ and $\mathbf{r} \to \mathbf{r}$ has been very successful. However, in the case of electron spin, one is forced beyond classical analogy to find the required operator.

Axiom III. The only allowed physical results of measurements of the observable A are the elements of the spectrum of the operator which corresponds to A.

We have already discussed some examples in which the spectrum is discrete. For instance, the operator L^2 of Section 5.8, which corresponds to the

square of the total angular momentum of a particle, satisfies

$$L^2 Y_{lm} = \hbar^2 l(l+1) Y_{lm} .$$

Here there are $2l + 1$ eigenvectors corresponding to each eigenvalue, $\hbar^2 l(l+1)$. Another system whose spectrum is particularly simple is the quantum harmonic oscillator. The classical Hamiltonian of this system is

$$\mathscr{H} = p^2/2m + \tfrac{1}{2}kx^2,$$

and by making the operator substitutions suggested above, we find

$$H = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}kx^2$$

for the *operator* H corresponding to the energy. We saw in Sections 3.10 and 4.4 that this operator has eigenvalues $\lambda_n = \hbar\omega_0(n + \frac{1}{2})$, where $\omega_0 = (k/m)^{1/2}$, and normalized eigenvectors

$$\phi_n(x) = (\sqrt{\pi}\, 2^n n!)^{-1/2} H_n(\alpha x) e^{-\alpha^2 x^2/2} ,$$

where

$$\alpha = (m\omega_0/\hbar)^{1/2} ,$$

and H_n is the nth Hermite polynomial. We have already seen in this chapter that this set of ϕ_n is a complete orthonormal set in Hilbert space.

Now let us look at the situation for a very simple quantum-mechanical system whose spectrum is strictly continuous. The operator corresponding to the momentum of a particle is $p_x = -i\hbar\, d/dx$ (we consider just the x-component). In the space of square-integrable functions (such functions must tend to zero at large distances), p_x is self-adjoint, so according to the spectral theorem, we expect that we should be able to find a resolution of the identity belonging to p_x. If we were very free and easy about this problem, we would simply say that p_x had "eigenfunctions" $\phi_k = (2\pi)^{-1/2} e^{ikx}$ and "eigenvalues" $\hbar k$ for all k. The factor $(2\pi)^{-1/2}$ in ϕ_k is chosen to "normalize e^{ikx} to a δ-function," that is,

$$\int_{-\infty}^{\infty} \phi_k^*(x)\phi_{k'}(x)\, dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i(k'-k)x}\, dx = \delta(k - k') .$$

Also,

$$\int_{-\infty}^{\infty} \phi_k^*(x)\phi_k(x')\, dk = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik(x'-x)} = \delta(x - x') ,$$

which one often thinks of as being analogous to

$$\sum_{n=1}^{N} P_n = I$$

in finite-dimensional spaces. That is, we imagine a P_k defined by

$$P_k f = \phi_k(\phi_k, f) ,$$

where k is a continuous variable $(-\infty < k < \infty)$. Note, however, that P_k is not a conventional operator, since it pushes vectors *out of Hilbert space*. This is due to the fact that ϕ_k is not an element of Hilbert space, and hence is not an eigenvector of p_x in the customary sense. However, this approach does suggest what a correct solution of the problem would be. On the basis of our earlier work, we expect that if we want to obtain a resolution of the identity, $E(k)$, then this resolution should be the sum of all the projection operators P_k, with $k' < k$. Thus we might try

$$E(k)f = \int_{-\infty}^{k} \phi_{k'}(\phi_{k'}, f)\, dk' \ .$$

Written out in detail, this is

$$E(k)f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{k} e^{ik'x} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ik'y} \hat{f}(y)\, dy \right] dk' \ ,$$

or

$$E(k)f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{k} e^{ik'x} \hat{f}(k')\, dk' \ , \tag{5.140}$$

where $\hat{f}(k')$ is the Fourier transform of $f(x)$. We have already indicated that the Fourier transform of every element in the space of square-integrable functions exists and is itself square-integrable (we will prove this rigorously in Chapter 9). Since $\hat{f}(k')$ is square-integrable, the integral in Eq. (5.140) clearly exists. It is a simple exercise to show that the vector $E(k)f(x)$ lies in Hilbert space.

We see by inspection that $E(\infty) = I$, by our main result on the inversion of Fourier transforms, and since $\hat{f}(k')$ is square-integrable, $E(-\infty) = 0$. Now we must show that $E(k_1)E(k_2) = E(k_1)$ (where we assume without loss of generality that $k_1 < k_2$). We reason as follows:

$$\begin{aligned} E(k_1)E(k_2)f &= E(k_1) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{k_2} e^{ik'x} \hat{f}(k')\, dk' \\ &= \frac{1}{2\pi} \int_{-\infty}^{k_1} e^{ikx} F(k)\, dk \ , \end{aligned}$$

where $F(k)$ denotes the Fourier transform of

$$\int_{-\infty}^{k_2} e^{ik'x} \hat{f}(k')\, dk'$$

considered as a function of x. This transform is readily computed:

$$F(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ikx} \int_{-\infty}^{k_2} e^{ik'x} \hat{f}(k') dk'\, dx = \begin{cases} \sqrt{2\pi}\, \hat{f}(k) & \text{if } k_2 > k \ , \\ 0 & \text{if } k_2 < k \ . \end{cases}$$

This result can be most simply obtained by the use of Eq. (5.63). Thus, since $k_1 < k_2$,

$$E(k_1)E(k_2)f = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{k_1} e^{ikx} \hat{f}(k)\, dk = E(k_1)f \ ,$$

as required. Finally, the reader can verify for himself that

$$\int_{-\infty}^{\infty} dE(k) = I ,$$

so we have indeed found a resolution of the identity.

To complete our task, we need only show that the above resolution belongs to p_x, that is, that

$$p_x = \hbar \int_{-\infty}^{\infty} k \, dE(k) .$$

This is easily done. We write

$$\begin{aligned}
\left[\hbar \int_{-\infty}^{\infty} k \, dE(k)\right] f &= \frac{\hbar}{\sqrt{2\pi}} \int_{-\infty}^{\infty} k \, d\left[\int_{-\infty}^{k} e^{ik'x} \hat{f}(k') \, dk'\right] \\
&= \frac{\hbar}{\sqrt{2\pi}} \int_{-\infty}^{\infty} k \frac{d}{dk}\left[\int_{-\infty}^{k} e^{ik'x} \hat{f}(k') \, dk'\right] dk \\
&= \frac{\hbar}{\sqrt{2\pi}} \int_{-\infty}^{\infty} k \, e^{ikx} \hat{f}(k) \, dk \\
&= \frac{-i\hbar}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{d}{dx} e^{ikx} \hat{f}(k) \, dk \\
&= \frac{-i\hbar}{\sqrt{2\pi}} \frac{d}{dx} \int_{-\infty}^{\infty} e^{ikx} \hat{f}(k) \, dk = -i\hbar \frac{df}{dx} = p_x f
\end{aligned}$$

for any f for which df/dx lies in our Hilbert space (under these conditions the above interchange of integration and differentiation is permissible). Thus we have found the resolution of the identity belonging to p_x. Of course, the spectral theorem tells us that such a resolution must exist; we have now explicitly constructed it. As we expect, the spectrum of p_x is purely continuous, consisting of all real numbers. Using the informal ideas above, the reader should be able to show that the resolution of the identity to the position operator X, defined by $Xf(x) \equiv xf(x)$, is the operator-valued function $E(\xi)$, specified by

$$E(\xi)f(x) = \begin{cases} f(x) & \text{if } x \leqslant \xi , \\ 0 & \text{if } x > \xi . \end{cases}$$

For more complicated self-adjoint operators, it is more difficult to find an explicit resolution of I, but it is always possible in principle.

We are now in a position to state the central axiom of quantum mechanics.

Axiom IV. If, on a system in the state ψ, we make a measurement of the observable A, then the probability that the value λ obtained by this measurement will lie between λ_1 and λ_2 ($\lambda_2 > \lambda_1$) is given by

$$P(\lambda_1, \lambda_2) = ||[E(\lambda_2) - E(\lambda_1)]\psi||^2 ,$$

when $E(\lambda)$ is the resolution of the identity belonging to A (by Axiom II plus the Spectral theorem).

Since $||\psi|| = 1$, we see that the probability of finding λ between $-\infty$ and $+\infty$ is 1, as it must be if the theory is to make sense physically. If $E(\lambda)$ is constant between λ_1 and λ_2 (i.e., if there are no points of the spectrum of A between λ_1 and λ_2), then $E(\lambda_2) - E(\lambda_1) = 0$, so a value can never be found in this range. This is in accord with Axiom III; in fact, it is clear that Axiom III is contained in Axiom IV. We have put in Axiom III only because its statement is such a familiar part of elementary quantum mechanics; if we were trying for minimal redundancy, we would omit Axiom III.

A particularly important case is that in which $E(\lambda)$ is constant between λ_1 and λ_2 *except* for a discontinuous jump at $\bar{\lambda}$. We will look at this case in some detail both for its importance to physics and because it enables us to show how a discrete spectrum can emerge from the projection-operator formalism. Intuitively, we should expect from our original motivations that whenever we have a discontinuity in $E(\lambda)$ at, say, $\bar{\lambda}$, on each side of which are intervals of constancy of $E(\lambda)$, we should find an eigenfunction of A. This is easily shown to be the case. Let λ_+ be any point in the interval of constancy to the right of $\bar{\lambda}$, and let λ_- be any point in the interval of constancy to the left of $\bar{\lambda}$. By analogy with our starting point in finite-dimensional spaces, we expect that if ψ is any element of H such that $\phi_{\bar{\lambda}} \equiv [E(\lambda_+) - E(\lambda_-)]\psi \neq 0$, then $\phi_{\bar{\lambda}}$ is an eigenfunction of A. We argue as follows:

$$\begin{aligned} A[E(\lambda_+) - E(\lambda_-)]\psi &= \int_{-\infty}^{\infty} \lambda \, d\{E(\lambda)[E(\lambda_+) - E(\lambda_-)]\psi\} \\ &= \int_{-\infty}^{\infty} \lambda \, d[E(\lambda)E(\lambda_+)\psi] - \int_{-\infty}^{\infty} \lambda \, d[E(\lambda)E(\lambda_-)\psi] \\ &= \int_{-\infty}^{\lambda_+} \lambda \, d[E(\lambda)\psi] - \int_{-\infty}^{\lambda_-} \lambda \, d[E(\lambda)\psi] \\ &= \int_{\lambda_-}^{\lambda_+} \lambda \, d[E(\lambda)\psi] \, . \end{aligned}$$

But by our assumption of intervals of constancy on either side of $\bar{\lambda}$, we have

$$\int_{\lambda_-}^{\lambda_+} \lambda \, d[E(\lambda)\psi] = \bar{\lambda}[E(\lambda_+) - E(\lambda_-)]\psi = \bar{\lambda}\phi_{\bar{\lambda}} \, .$$

Hence

$$A\phi_{\bar{\lambda}} = \bar{\lambda}\phi_{\bar{\lambda}} \, ,$$

so $\bar{\lambda}$ is an eigenvalue of A, and $\phi_{\bar{\lambda}}$ is an eigenvector belonging to that eigenvalue.

Of course, $\phi_{\bar{\lambda}}$ may not be the *only* eigenvector belonging to $\bar{\lambda}$. If one picked a different ψ to start with (note that it is always possible to find a ψ such that $[E(\lambda_+) - E(\lambda_-)]\psi \neq 0$), one might find a vector $\tilde{\phi}_{\bar{\lambda}}$ which belongs to $\bar{\lambda}$ but is not just a multiple of $\phi_{\bar{\lambda}}$. In this case, we say that $\bar{\lambda}$ has multiplicity $\mu > 1$. We leave it as an exercise for the reader to show that if $\Phi_{\bar{\lambda}}$ is the unique normalized eigenvector belonging to $\bar{\lambda}$ (that is, $\bar{\lambda}$ has multiplicity 1), then for any $\psi \in H$,

$$[E(\lambda_+) - E(\lambda_-)]\psi = \Phi_{\bar{\lambda}}(\Phi_{\bar{\lambda}}, \psi) \, , \tag{5.141}$$

where λ_+ and λ_- are as chosen above. Thus, just as we would expect from our original finite-dimensional analogy, the discontinuity in $E(\lambda)$ at an eigenvalue $\bar{\lambda}$ is just the projection operator onto the one-dimensional subspace of H spanned by $\Phi_{\bar{\lambda}}$. If $\bar{\lambda}$ has multiplicity μ greater than one, then the discontinuity in $E(\lambda)$ at $\bar{\lambda}$ is just the projection operator onto the μ-dimensional subspace of H spanned by the μ eigenvectors belonging to $\bar{\lambda}$. Thus, in general, we may write

$$[E(\lambda_+) - E(\lambda_-)]\psi = \sum_{\nu=1}^{\mu} \Phi_{\bar{\lambda}}^{(\nu)} (\Phi_{\bar{\lambda}}^{(\nu)}, \psi) \ , \tag{5.142}$$

where μ is the multiplicity of $\bar{\lambda}$ (see Problem 5.25).

Returning now to our interpretation of Axiom IV, we see that in the case of a discontinuous jump in $E(\lambda)$ between λ_1 and λ_2, the probability of obtaining a measured value of A between λ_1 and λ_2 is, according to Eq. (5.141), just

$$P(\lambda_1, \lambda_2) = |(\Phi_{\bar{\lambda}}, \psi)|^2 \ ,$$

where $\bar{\lambda}$ is the point of discontinuity; we have assumed that $\bar{\lambda}$ has multiplicity 1. Thus $P(\lambda_1, \lambda_2)$ is just the modulus squared of the component of ψ along the "direction" in Hilbert space defined by the eigenvector $\Phi_{\bar{\lambda}}$ belonging to $\bar{\lambda}$. More generally, if $\bar{\lambda}$ has multiplicity μ,

$$P(\lambda_1, \lambda_2) = \sum_{\nu=1}^{\mu} |(\Phi_{\bar{\lambda}}^{(\nu)}, \psi)|^2 \ . \tag{5.143}$$

Hence Eqs. (5.141) and (5.142) lead to very simple results for the probability of finding an observable A in the discrete spectrum.

If we choose an observable with a continuous spectrum, the situation is similar. Consider the x-component of the momentum, p_x. By our previous work, the probability of finding p_x between $\hbar k_1$ and $\hbar k_2$ is just

$$\begin{aligned} P(k_1, k_2) &= ||[E(k_2) - E(k_1)]\psi|| \\ &= \left\| \frac{1}{\sqrt{2\pi}} \int_{k_1}^{k_2} e^{ikx} \hat{\psi}(k)\, dk \right\| \\ &= \int_{k_1}^{k_2} |\hat{\psi}(k)|^2 \, dk \ , \end{aligned}$$

where we have again used Eq. (5.63). In a more conventional language, $P(k_1, k_2)$ is just the continuous superposition of "plane-wave" probability amplitudes.

We now define the *expectation value* of A.

Definition. $\langle A \rangle \equiv \lim \sum_i \bar{\lambda}_i ||[E(\lambda_i + \Delta) - E(\lambda_i)]\psi||^2$ is the *expectation value* of the observable A for the system described by the state vector ψ. Here we have divided the spectrum into intervals of length Δ, and $\bar{\lambda}_i$ is any point in the ith interval. By lim we mean the limit as the interval size tends to zero.

The physical meaning of the expectation value is very simple. It is just the value that would be found by taking the average of many measurements of

A on a large collection of systems all in the state ϕ. Each possible result, $\bar{\lambda}_i$, of a measurement is weighted by the probability of finding a result in the vicinity of $\bar{\lambda}_i$. Note that $\langle A \rangle$ is a number, not an operator.

Theorem. $\langle A \rangle = (\phi, A\phi)$.

Proof. According to the definition (Eq. 5.139) of the Stieltjes integral,

$$\begin{aligned}(\phi, A\phi) &= \int_{-\infty}^{\infty} \lambda \, d[(\phi, E(\lambda)\phi)] \\ &= \lim \sum_i \bar{\lambda}_i (\phi, [E(\lambda_i + \Delta) - E(\lambda_i)]\phi) \ .\end{aligned}$$

But $[E(\lambda_i + \Delta) - E(\lambda_i)]^2 = [E(\lambda_i + \Delta) - E(\lambda_i)]$, so using the self-adjointness of E, we have

$$\begin{aligned}(\phi, A\phi) &= \lim \sum_i \bar{\lambda}_i ([E(\lambda_i + \Delta) - E(\lambda_i)]\phi, [E(\lambda_i + \Delta) - E(\lambda_i)]\phi) \\ &= \lim \sum_i \bar{\lambda}_i ||(E(\lambda_i + \Delta) - E(\lambda_i)]\phi||^2 = \langle A \rangle\end{aligned}$$

by definition.

In terms of the expectation value of A, we define the mean-square deviation $(\Delta A)^2$, which measures the dispersion around the mean value $\langle A \rangle$.

Definition. $(\Delta A)^2$ is the expectation value of $[A - \langle A \rangle]^2$ in the state ϕ in which $\langle A \rangle$ is computed.

Theorem. $(\Delta A)^2 = \langle A^2 \rangle - \langle A \rangle^2$.

Proof.
$$\begin{aligned}(\Delta A)^2 &= \langle [A - \langle A \rangle]^2 \rangle = (\phi, [A - \langle A \rangle]^2 \phi) \\ &= (\phi, [A^2 - 2A\langle A \rangle + \langle A \rangle^2]\phi) \\ &= (\phi, A^2\phi) - 2(\phi, A\phi)\langle A \rangle + \langle A \rangle^2 \\ &= \langle A^2 \rangle - \langle A \rangle^2 \ ,\end{aligned}$$

as required. We have used the previous theorem several times in obtaining this result.

Note that according to our definition,

$$(\Delta A)^2 = (\phi, [A - \langle A \rangle]^2 \phi) = ([A - \langle A \rangle]\phi, [A - \langle A \rangle]\phi) \ ,$$

because A is self-adjoint and $\langle A \rangle$ is real. Thus

$$(\Delta A)^2 = ||[A - \langle A \rangle]\phi||^2 \ ,$$

so if $\Delta A = 0$ for some ϕ, we must have $[A - \langle A \rangle]\phi = 0$; that is, ϕ is an eigenvector of A with eigenvalue $\langle A \rangle$. This has the important implication that unless ϕ is an eigenstate of A, one cannot make a perfectly precise measurement of A in that state. Thus only values of A lying in the discrete spectrum of A can be determined with perfect accuracy. For other values, one must always have some dispersion, ΔA, associated with measurements.

It is instructive to check that if $\psi = \Phi_n$, a normalized eigenstate of A with eigenvalue λ_n, then indeed $\Delta A = 0$. We have

$$\langle A \rangle = (\Phi_n, A\Phi_n) = \lambda_n ,$$
$$\langle A^2 \rangle = (\Phi_n, A^2\Phi_n) = (A\Phi_n, A\Phi_n) = \lambda_n^2 ,$$

so

$$\langle \Delta A \rangle^2 = \langle A^2 \rangle - \langle A \rangle^2 = \lambda_n^2 - \lambda_n^2 = 0 .$$

Furthermore, if $\psi = \Phi_n$, the probability of obtaining a value λ between λ_1 and λ_2 when A is measured is zero if $[\lambda_1, \lambda_2]$ does not contain λ_n, and this probability is 1 if $[\lambda_1, \lambda_2]$ contains λ_n. This follows simply by showing that the probability of finding λ in a small interval around λ_n is 1. According to Axiom IV,

$$\begin{aligned} P(\lambda_n - \epsilon, \lambda_n + \epsilon) &= ||[E(\lambda_n + \epsilon) - E(\lambda_n - \epsilon)]\Phi_n||^2 \\ &= |(\Phi_n, \Phi_n)|^2 = 1 , \end{aligned}$$

where we have made use of Eq. (5.142). Thus the probability of the measured value of A being λ_n is 1. In other words, if a system is in an eigenstate, so that its state vector happens to be an eigenvector of A and not a superposition of many eigenvectors, then the result of a measurement of A is certain to be the eigenvalue corresponding to that eigenvector.

Having considered in detail the case of a single measurement, let us mention briefly an important generalization of Axiom IV which admits the possibility of simultaneous measurement of several observables.

Axiom IV′. Let A, B, and C be observables whose corresponding linear operators commute, that is, $[A, B] = [A, C] = [B, C] = 0$. Then the probability that a simultaneous measurement of A, B, and C in a system whose state vector is ψ will yield a value of A between a_1 and a_2, B between b^1 and b_2, and C between c_1 and c_2 is

$$\begin{aligned} P(a_1, a_2; b_1, b_2; c_1, c_2) = ||&[E_A(a_2) - E_A(a_1)][E_B(b_2) - E_B(b_1)] \\ &\times [E_C(c_2) - E_C(c_1)]\psi||^2 , \end{aligned}$$

where $E_A(a)$, $E_B(b)$, and $E_C(c)$ are the resolutions of the identity belonging to A, B, and C, respectively.

In the case of a single measurement, Axiom IV′ is clearly equivalent to Axiom IV.

Note that since A, B, and C commute, so do $E_A(a)$, $E_B(b)$, and $E_C(c)$, and therefore the ordering of the projection operators on the right-hand side of the above equation is immaterial. Clearly, if their order were significant, the concept of simultaneous measurement would be untenable. The possibility of simultaneous measurement of *noncommuting* observables is denied in quantum mechanics; thus p_x and x, which satisfy $[x, p_x] = i\hbar$, cannot be measured simultaneously. The importance of simultaneous measurement of observables will become apparent when we discuss Axiom VI below.

Everything said so far has been concerned with the properties of a system at a fixed time. We now must say what happens to the system as the time changes.

Axiom V. For every system, there exists an Hermitian operator, H, the Hamiltonian or energy operator, which determines the time development of the state vector, Ψ, of the system through the time-dependent Schrödinger equation:

$$H\Psi(x, t) = i\hbar \frac{\partial \Psi(x, t)}{\partial t} , \tag{5.144}$$

provided that the system is not disturbed (by a measurement, for example). Here $\hbar$ is a constant equal to 1.054×10^{-27} erg-sec.

Axiom I says that the state vector Ψ is normalized. To be certain that our axioms are consistent, we must therefore show that this normalization is preserved in time by the above equation which governs the time dependence of the state vector. Thus we prove the following result:

Theorem.

$$\frac{\partial}{\partial t}(\Psi, \Psi) = 0 .$$

Proof.

$$\frac{\partial}{\partial t}(\Psi, \Psi) = \left(\frac{\partial \Psi}{\partial t}, \Psi\right) + \left(\Psi, \frac{\partial \Psi}{\partial t}\right) .$$

But

$$\frac{\partial \Psi}{\partial t} = \frac{1}{i\hbar} H\Psi ,$$

so,

$$\begin{aligned} \frac{\partial}{\partial t}(\Psi, \Psi) &= \left(\frac{1}{i\hbar} H\Psi, \Psi\right) + \left(\Psi, \frac{1}{i\hbar} H\Psi\right) \\ &= -\frac{1}{i\hbar}(H\Psi, \Psi) + (\Psi, H\Psi)\frac{1}{i\hbar} \\ &= -\frac{i}{i\hbar}(\Psi, H\Psi) + (\Psi, H\Psi)\frac{1}{i\hbar} = 0 , \end{aligned}$$

since H is self-adjoint. Thus Axiom I is consistent with Axiom V.

In general, the solution of Schrödinger's time-dependent equation for a Hamiltonian with complicated time dependence is a difficult job. However, if the Hamiltonian is time-independent, then matters simplify considerably. In this case, it is elementary to verify that the solution to Eq. (5.144), which reduces to $\Psi(x, 0)$ at $t = 0$, is simply

$$\Psi(x, t) = e^{-iHt/\hbar}\Psi(x, 0) . \tag{5.145}$$

Since H is self-adjoint, we have a resolution of the identity belonging to H, and therefore Eq. (5.145) can be written as

$$\Psi(x, t) = \int_{-\infty}^{\infty} e^{-i\epsilon t/\hbar} d[E(\epsilon)\Psi(x, 0)] . \tag{5.146}$$

Again, it is elementary to verify directly that this $\Psi(x, t)$ satisfies Eq. (5.144). Note that if the spectrum of H is purely discrete, then according to Eq. (5.143), Eq. (5.146) takes on the simple form

$$\Psi(x, t) = \sum_n e^{-i\epsilon_n t/\hbar}(\phi_n, \Psi(t = 0))\phi_n(x) , \tag{5.147}$$

where the ϕ_n are the eigenvectors of H and the ϵ_n are its eigenvalues. According to Axiom IV, the probability of finding a value ϵ_n for the energy of the system whose state vector is $\Psi(x, 0)$ is just

$$P(\epsilon_n;\ t = 0) = |(\phi_n, \Psi(t = 0))|^2 .$$

Using Eq. (5.147), we see that at time t, the probability of measuring H and finding ϵ_n is

$$P(\epsilon_n;\ t) = |e^{-i\epsilon_n t/\hbar}(\phi_n, \Psi(t = 0)|^2 = |(\phi_n, \Psi(t = 0))|^2 = P(\epsilon_n;\ t = 0);$$

that is, the probability of finding a particular value of the energy does not change with time.

Let us now see what happens for more general observables. Suppose that A is any observable, with associated resolution of the identity, $E_A(\lambda)$. According to Axiom IV, the probability of finding a value of A between λ_1 and λ_2 is just

$$P(\lambda_1, \lambda_2;\ t) = ||[E_A(\lambda_2) - E_A(\lambda_1)]e^{-iHt/\hbar}\Psi(x, 0)||^2 ,$$

where we have made use of Eq. (5.145) and are continuing to assume that H is independent of time. It is clear that, in general, $P(\lambda_1, \lambda_2;\ t)$ will depend on time, but there are two important situations in which $P(\lambda_1, \lambda_2;\ t)$ will be time-independent.

1. $\Psi(x, 0) = \phi_n(x)$, where $\phi_n(x)$ is an eigenvector of H.

   In this case,

$$P(\lambda_1, \lambda_2;\ t) = ||[E_A(\lambda_2) - E_A(\lambda_1)]e^{-i\epsilon_n t/\hbar}\phi_n(x)||^2 ,$$

   but since $e^{-i\epsilon_n t/\hbar}$ is just a complex number with modulus 1, it cannot affect the norm of a vector, so

$$P(\lambda_1, \lambda_2;\ t) = ||[E_A(\lambda_2) - E_A(\lambda_1)]\phi_n||^2 = P(\lambda_1, \lambda_2;\ t = 0) .$$

2. H commutes with A.

   In this case, H also commutes with $E_A(\lambda)$, so

$$P(\lambda_1, \lambda_2;\ t) = ||e^{-iHt/\hbar}[E_A(\lambda_2) - E_A(\lambda_1)]\Psi(x, 0)||^2 .$$

But according to Theorem 4.13, this is simply

$$P(\lambda_1, \lambda_2; t) = ||[E_A(\lambda_2) - E_A(\lambda_1)]\Psi(x,0)||^2 = P(\lambda_1, \lambda_2; t=0) ,$$

since $\exp[-iHt/\hbar]$ is a unitary operator. This tells us that the results of measuring an observable which commutes with the Hamiltonian are independent of the time at which they are measured.

By way of illustration, let us now investigate the situation for a particular observable, the position of a particle. We have already introduced the operator X defined by $Xf(x) \equiv xf(x)$ and have stated (see Problem 5.26) that its resolution of the identity is given by

$$E_X(\xi)f(x) = \begin{cases} f(x) & \text{if } x \leq \xi , \\ 0 & \text{if } x > \xi . \end{cases}$$

We now ask, what is the probability of obtaining a value of the position between ξ_1 and ξ_2 for a particle in a state Ψ? According to Axiom IV,

$$P(\xi_1, \xi_2; t) = ||[E_X(\xi_2) - E_X(\xi_1)]\Psi(x, t)||^2 .$$

From the definition of $E_X(\xi)$, we see that the vector whose norm is to be evaluated is just the function which is zero outside $[\xi_1, \xi_2]$ and equal to $\Psi(x, t)$ inside $[\xi_1, \xi_2]$. Thus

$$P(\xi_1, \xi_2; t) = \int_{\xi_1}^{\xi_2} |\Psi(x, t)|^2 dx .$$

Note that the probability of finding the particle *someplace* is 1, since Ψ is normalized. If we set $\xi_1 = \xi - \Delta/2$ and $\xi_2 = \xi + \Delta/2$, then for Δ small, we see that the probability of finding the particle in a small interval of length Δ about ξ is just $P = |\Psi(\xi, t)|^2\Delta$; thus $|\Psi(x, t)|^2$ is called the *probability density*, $\rho(x, t)$. This result is often stated at the beginning of quantum mechanics courses. It arose first, historically, in the pioneering work of Born on the probabilistic interpretation of quantum mechanics. Here the result is presented as a special case of Axiom IV. Note that, in general, the probability of finding the particle in a particular region varies with time; however, if $\Psi(x, t)$ is an eigenstate of the Hamiltonian, then $\Psi(x, t)$ is related to $\Psi(x, 0)$ just by a multiplicative phase factor, so $|\Psi(x, t)|^2 = |\Psi(x, 0)|^2$, and the probability of finding the particle in any region is constant in time. This is why eigenstates of the Hamiltonian are often called *stationary states*.

Finally, let us use Axiom V to obtain $P(\xi_1, \xi_2; t)$ in terms of a particular initial state, $\Psi(x, t = 0)$. For simplicity, we assume that $\Psi(x, 0)$ is given by

$$\Psi(x, 0) = (1/\sqrt[4]{\pi a^2})e^{-x^2/2a^2} .$$

Thus the initial probability density is equal to $(1/\pi^{1/2}a)e^{-x^2/a^2}$, so that the probability of finding the particle at a distance greater than a from the origin is very small compared with the probability of finding the particle within a distance a on either side of the origin. Let us assume that we are dealing with a free particle (and confine our attention to the x-dependence of the problem). The

Hamiltonian is just $H = p_x^2/2m$, where m is the particle's mass, so according to Axiom V and Eq. (5.145),

$$\Psi(x, t) = e^{-ip_x^2 t/2m\hbar} \frac{1}{\sqrt[4]{\pi a^2}} e^{-x^2/2a^2} .$$

Using Eq. (5.146) and the spectral resolution which we have already determined for p_x we have immediately

$$\Psi(x, t) = \frac{1}{\sqrt[4]{\pi a^2}} \int_{-\infty}^{\infty} e^{i\hbar k^2 t/2m} \, d[E(k) \, e^{-x^2/2a^2}] .$$

But according to our previous work,

$$E(k) f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{k} e^{ik'x} \hat{f}(k') \, dk' ,$$

where $\hat{f}$ denotes the Fourier transform of f. The Fourier transform of $\Psi(x, 0)$ is easily computed; we find

$$\hat{\Psi}(k, 0) = \sqrt[4]{\frac{a^2}{\pi}} e^{-a^2k^2/2} .$$

Thus

$$\Psi(x, t) = \sqrt[4]{\frac{a^2}{4\pi^3}} \int_{-\infty}^{\infty} e^{-i\hbar k^2 t/2m} \, d\left[\int_{-\infty}^{k} e^{ik'x} \, e^{-a^2k'^2/2} \, dk' \right]$$

$$= \sqrt[4]{\frac{a^2}{4\pi^3}} \int_{-\infty}^{\infty} e^{-i\hbar k^2 t/2m} \, e^{-a^2k^2/2} \, e^{ikx} \, dk .$$

This last integral is readily evaluated. We find that

$$\Psi(x, t) = \frac{1}{\pi^{1/4}} \frac{1}{\sqrt{a + i\dfrac{ht}{ma}}} \exp\left[-\frac{x^2}{2a(a + iht/ma)} \right] ,$$

and therefore the probability density is given by

$$\rho(x, t) = |\Psi(x, t)|^2 = (1/\sqrt{\pi\Delta^2}) \, e^{-x^2/\Delta^2} ,$$

where $\Delta^2 = a^2 + (ht/ma)^2$. Thus the width of the probability distribution increases in time. This phenomenon is known as "the spreading of the wave packet." The initial width was equal to $2a$; the width at time t is $2\sqrt{a^2 + (ht/ma)^2}$. To get a feeling for the magnitudes involved, suppose that we are dealing with electrons and that $a \equiv 1$ mm. This might describe a case in which one creates a beam of electrons collimated by a slit of width approximately equal to one millimeter. Then if we use $m = 9 \times 10^{-28}$ gm and $h = 10^{-27}$ erg-sec, we find that the width has doubled in about 15 msec. This is a very short time on a macroscopic scale, but on a scale appropriate to electrons, it is a very large time. For example, a velocity of 10^8 cm/sec is not unusual for an electron; in 15 msec, such an electron

would travel 1.5×10^6 cm, or 15 km! Thus a beam of electrons initially having a width of 1 mm and moving with a velocity of 10^8 cm/sec would travel 15 km before doubling in width. This assumes, of course, that the electrons are truly free; i.e., we neglect collisions, electromagnetic forces, gravitational forces, etc.

Before we can claim to have a complete system of axioms, we must give some method by which one can gain information about the state vector Ψ that figures so prominently in the previous axioms. The key to this problem, which is clearly of paramount practical importance, is the following axiom.

Axiom VI. If at time $t = 0$ one measures simultaneously the commuting observables A, B, and C, and finds with certainty that the values of these observables lie between a_1 and a_2, b_1 and b_2, and c_1 and c_2 respectively, then immediately after the measurement, the state vector of the system satisfies the relation

$$[E_A(a_2) - E_A(a_1)][E_B(b_2) - E_B(b_1)][E_C(c_2) - E_C(c_1)]\,\Psi = \Psi\,. \quad (5.148)$$

We may say that the measurements project the original wave function into the subspace of Hilbert space associated with the projection operator

$$[E_A(a_2) - E_A(a_1)][E_B(b_2) - E_B(b_1)][E_C(c_2) - E_C(c_1)]\,.$$

Note that because of Axiom IV, Eq. (5.148) implies that if we remeasure A (or B or C) immediately after our first measurement, we will find a value between a_1 and a_2 (or b_1 and b_2, or c_1 and c_2) with certainty.

Let us first consider the measurement of a single observable, say A. If there is only one point, a, in the spectrum of A between a_1 and a_2, then according to Eq. (5.141),

$$\Psi = e^{i\alpha}\Phi_a\,,$$

if a is nondegenerate, when Φ_a is the normalized eigenvector of A belonging to a and $e^{i\alpha}$ is a phase factor which is indeterminate. Clearly, according to our previous axioms, this phase factor can have no observable consequences. In this case, the measurement of A has, so to speak, "forced" the system into an eigenstate of A. Note that this does *not* tell us that the system was in the state Φ_a *before* the measurement, but only that it is in the state Φ_a *after* the measurement. We have "prepared" the system in the state Φ_a. If the eigenvalue a has multiplicity $\mu_a > 1$, then according to Eq. (5.142), after the measurement of A, we can only say that the state vector lies in the subspace of Hilbert space spanned by the μ_a orthonormal eigenvector of A belonging to a, that is,

$$\Psi = \sum_{\nu=1}^{\mu_a} c_\nu \Phi_a^{(\nu)}\,,$$

where

$$\sum_{\nu=1}^{\mu_a} |c_\nu|^2 = 1\,.$$

Thus in this case we have not completely specified the state vector by measuring A.

However, according to Axiom VI, we have additional resources at our disposal; we can choose *another* observable, B, whose self-adjoint operator commutes with A, and measure A and B simultaneously. Suppose we discover that the value of A lies with certainty between a_1 and a_2 and the value of B lies with certainty between b_1 and b_2. Again, we imagine that a is the only point in the spectrum of A between a_1 and a_2 and that a has multiplicity μ_a. Hence

$$[E_A(a_2) - E_A(a_1)]\Psi = \sum_{\nu=1}^{\mu_a} c_\nu \Phi_a^{(\nu)} \ . \tag{5.149}$$

We assume also that there is only one point, b, between b_1 and b_2 in the spectrum of B, and that b has multiplicity μ_b. Using Eqs. (5.142) and (5.149), we find that

$$[E_B(b_2) - E_B(b_1)][E_A(a_2) - E_A(a_1)]\Psi = \sum_{\nu=1}^{\mu_a} \sum_{\nu=1}^{\mu_b} c_\nu (\Phi_b^{(\lambda)}, \Phi_a^{(\nu)}) \Phi_b^{(\lambda)} \ .$$

We know from Theorem 4.22 that every eigenvector of A must be an eigenvector of B *and vice versa*, so $(\Phi_b^{(\lambda)}, \Phi_a^{(\nu)}) = 1$ if $\Phi_b^{(\lambda)} = \Phi_a^{(\nu)}$, and vanishes otherwise. It may happen that there is only one pair, $(\bar{\lambda}, \bar{\nu})$, for which $(\Phi_b^{(\lambda)}, \Phi_a^{(\nu)}) = 1$. In this case, Axiom VI tells us that

$$\Psi = e^{i\beta} \Phi_b^{(\bar{\lambda})} = e^{i\beta} \Phi_a^{(\bar{\nu})} \ ,$$

so that we have uniquely specified the state vector of the system.

If there is more than one pair for which $(\Phi_b^{(\lambda)}, \Phi_a^{(\nu)})$ fails to vanish, then calling these pairs (λ_i, ν_i), we have

$$\Psi = \sum_i c_{\nu_i} \Phi_b^{(\lambda_i)} = \sum_i c_{\nu_i} \Phi_a^{(\nu_i)} \ ;$$

that is, Ψ lies in the subspace of Hilbert space spanned by the mutual eigenvectors of A and B belonging to a and b. In this last case, the state vector is not uniquely determined, so we must find a new observable, C, which commutes with A and B, and repeat the process. In this manner we eventually will specify the state uniquely.

A collection of commuting observables which has the property that simultaneous measurement of the observables specifies uniquely the state vector of a system is called a *complete set* of commuting observables. For example, if we consider an electron in a coulomb field and neglect electron spin, then H, L^2, and L_z (where $H = p^2/2m - e^2/r$ and $\mathbf{L}$ is the orbital angular momentum of the electron) constitute such a set; H and L^2 alone do not constitute a complete set of observables, for after a measurement of the energy and the total angular momentum, there is still a degeneracy corresponding to the possible values of the z-component of the angular momentum. If we include the spin of the electron, then H, L^2, S^2, J^2, and J_z constitute a complete set, where $\mathbf{S}$ is the electron's spin angular momentum, and $\mathbf{J} = \mathbf{L} + \mathbf{S}$ is the total angular momentum

of the electron. This set is not unique; H, L^2, L_z, S^2 and S_z would do just as well. Which set is more useful depends on the practical situation. If we wish to take into account relativistic effects by including magnetic interactions via a term in the Hamiltonian of the form $A\mathbf{L}\cdot\mathbf{S}/r^3$, then the second set mentioned above is no longer a complete commuting set, whereas the first is (because

$$\mathbf{L}\cdot\mathbf{S} = (J^2 - L^2 - S^2)/2) .$$

Thus Axiom VI enables us to perform the all-important task of controlling the initial conditions of an experiment by the process of preparing quantum states. Without this axiom, Axiom IV would have no content, since if one never knew what the state vector of a system was, one could never hope to compute the probability of obtaining a particular result in a measurement.

PROBLEMS

1. Prove that $\int_{-1}^{1} P_n(x)dx = 0$ for $n \neq 0$ in two ways.
2. A metal spherical shell of radius R with its center at $x = y = z = 0$ is cut in half along its intersection with the plane $z = 0$. The two halves are separated by an infinitesimal gap and the upper and lower hemispheres are brought to voltages $+V$ and $-V$ respectively. Show that the potential inside the sphere is

$$\phi(r, \theta) = V \sum_{l=0}^{\infty} (-1)^l \left(\frac{r}{R}\right)^{2l+1} \frac{(2l)!}{(2^l l!)^2} \frac{4l+3}{2l+2} P_{2l+1}(\cos\theta) ,$$

where θ is the polar angle measured relative to the positive z-axis.
Hint: $\nabla^2\phi = 0$ inside and on a spherical shell. ϕ is clearly independent of ϕ. Following the usual procedure of separation of variables, we find that the solution is of the form

$$\phi(r, \theta) = \sum_{l=0}^{\infty} [A_l r^l + B_l r^{-l-1}][C_l P_l(\cos\theta) + D_l Q_l(\cos\theta)] .$$

Immediately set $B_l = 0$ (all l) so $\phi(0, \theta)$ is finite and set $D_l = 0$ (all l) so $\phi(r, 0)$ and $\phi(r, \pi)$ are finite. The problem is now one of mathematics—determining the constants $A_l C_l \equiv \alpha_l$ such that the boundary conditions are satisfied:

$$\phi(R, \theta) = \begin{cases} +V & \text{for} \quad 0 \leq \theta < \pi/2 , \\ -V & \text{for} \quad \pi/2 < \theta \leq \pi . \end{cases}$$

3. A conducting sphere of radius d sits in a charge-free region in an electric field. Its surface is kept at a fixed distribution of electric potential $V = F(\theta)$, where (r, ϕ, θ) are spherical coordinates with the origin at the center of the sphere. Determine the potential at all points in the space inside and outside the sphere; i.e., solve Laplace's equation $\nabla^2 V = 0$, assuming the following boundary conditions:
   i. $\lim_{r\to d} V(r, \theta) = F(\theta), \qquad 0 < \theta < \pi ,$
   ii. $\lim_{r\to\infty} V(r, \theta) = 0$

   (potential vanishes at points infinitely far away from sphere).

Hint: Use the method of separation of variables. Solve the r equation by making the substitution $r = e^t$. When it is solved, set the exponent $-\frac{1}{2} + (\lambda + \frac{1}{4})^{1/2} = n$ so that $\lambda = n(n+1)$ (λ = separation constant). Solve the θ equation by an appropriate change of variable.

Answer:

$$\textit{Inside sphere}: \quad V(r,\theta) = \sum_{n=0}^{\infty} \frac{2n+1}{2} \frac{r^n}{d^n} P_n(\cos\theta) \int_{-1}^{1} F(x) P_n(x)\, dx$$

$$r \leqq d, \quad F(\theta) = f(\cos\theta)\,.$$

$$\textit{Outside sphere}: \quad V(r,\theta) = \sum_{n=0}^{\infty} \frac{2n+1}{2} \frac{d^{n+1}}{r^{n+1}} P_n(\cos\theta) \int_{-1}^{1} F(x) P_n(x)\, dx\,.$$

4. a) Show that the 2π periodic Fourier series representation of the square wave (step function)

$$f(x) = \begin{cases} -1, & -\pi < x < 0\,, \\ +1, & 0 < x < \pi\,, \end{cases} \quad \text{is} \quad f(x) = \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{1}{2n-1} \sin(2n-1)x\,.$$

Compare this result with the expansion of the same function in Legendre polynomials—see Jackson, *Classical Electrodynamics*, pp. 58–59.

b) Derive from the Fourier series obtained in (a) a simple infinite series expression for π; preferably one with rapid convergence. It is instructive to plot several of the approximations to $f(x)$; that is, plot

$$S_m(x) = \frac{4}{\pi} \sum_{n=1}^{m} \frac{1}{2n-1} \sin(2n-1)x$$

for several values of m to see how $S_m(x) \to f(x)$ with increasing m.

c) *Gibbs' phenomenon*: Show that the Fourier series overshoots the function just to the right and left of the origin by about 18%. This is not easy. Guidance may be found in Sommerfeld's *Partial Differential Equations in Physics*, pp. 7–12, or in Morse and Feshbach, p. 747.

5. Prove that if $f(x)$ is real,

$$f(x) = \frac{1}{\pi} \int_0^{\infty} du \int_{-\infty}^{\infty} f(t) \cos u(t-x)\, dt\,.$$

6. Solve the integral equation

$$\int_0^{\infty} g(t) \cos xt\, dt = \begin{cases} e^{-\alpha x} & \text{for } x > 0 \\ e^{\alpha x} & \text{for } x < 0 \end{cases} \qquad (\alpha > 0)\,.$$

Answer:

$$g(t) = \frac{2}{\pi} \frac{\alpha}{\alpha^2 + t^2}\,.$$

7. a) Show that the energy transferred to a spring by a time-dependent force $F(t)$ defined by (Fig. 5.6)

$$F(t) = \begin{cases} 0, & t < -\tau \\ F_0(t/2\tau + 1/2), & |t| \leq \tau \\ F_0, & t > \tau \end{cases} \quad \text{is} \quad \Delta E = \frac{1}{2m} (F_0/\omega)^2 \left(\frac{\sin \omega\tau}{\omega\tau} \right)^2.$$

Hint: Take derivatives of $F(t)$ until you reach a function whose Fourier transform is known.

b) Show that the energy transfers in the adiabatic and sudden limits agree with these limits as determined for the probability integral force function (Example 3, Section 5.7).

c) For certain finite values of $\omega\tau$ there will be no energy transfer to the spring. Determine these values and explain their origin physically.

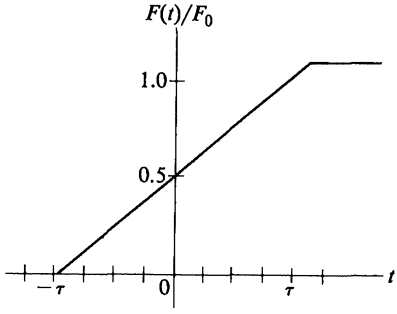

Fig. 5.6.

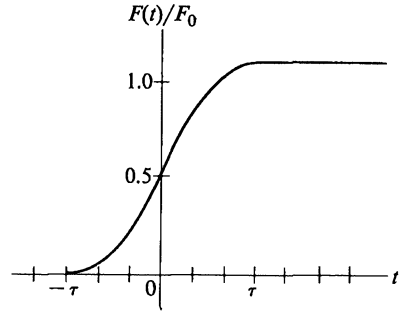

Fig. 5.7.

8. Work Problem 7 for the force function (Fig. 5.7)

$$F(t) = \begin{cases} 0\ , & t < -\tau\ , \\ F_0\left[\frac{1}{2} + \frac{1}{4}\left(\left(\frac{3t}{\tau}\right) - \left(\frac{t}{\tau}\right)^3\right)\right], & |t| \leqslant \tau\ . \\ F_0\ , & t > \tau\ . \end{cases}$$

Answer:

$$\Delta E = \frac{1}{2m}(F_0/\omega)^2 \frac{9}{(\omega_0\tau)^6}(\sin\omega_0\tau - \omega_0\tau\cos\omega_0\tau)^2\ .$$

The adiabatic and sudden limits must and do (prove this) agree with those of Problem 7. For part (c), settle for a determination of the position of the first node of ΔE. You should get

$$\omega\tau \cong 4.49\ .$$

9. Compute the three-dimensional Fourier transforms of the following functions:

a) $H(\mathbf{r}) = \dfrac{e^{-\lambda|\mathbf{r}|}}{|\mathbf{r}|}$

b) $F(\mathbf{r}) = \dfrac{1}{|\mathbf{r}|}$ [To save work consider $F(\mathbf{r})$ as a special case of $H(\mathbf{r})$.]

c) $Y(\mathbf{r}) = e^{-\lambda|\mathbf{r}|}$ [The Fourier transform of this function can be obtained from that of $H(\mathbf{r})$ also.]

10. Maxwell's equations in free space in Gaussian units are:

$$\text{curl}\ \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t}, \qquad \text{div}\ \mathbf{H} = 0\ , \qquad \text{curl}\ \mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t}, \qquad \text{div}\ \mathbf{E} = 0\ .$$

At any time t the electric (and magnetic) fields may be Fourier analyzed:

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{(2\pi)^{3/2}} \int_{-\infty}^{\infty} \mathbf{E}(k, t) e^{i\mathbf{k}\cdot\mathbf{r}}\, d\mathbf{k}, \quad \text{etc.}$$

[Here the Fourier components of $\mathbf{E}(\mathbf{r}, t)$ have been written $\mathbf{E}(\mathbf{k}, t)$; the argument identifies $\mathbf{E}(\mathbf{k}, t)$ as the Fourier component of the field, distinguishing it from the field itself, which has argument $\mathbf{r}$. If you don't like this notation, adopt another.]

a) Prove that the Fourier components of the electric and magnetic fields satisfy the following equations:

i) $\dot{\mathbf{H}}(\mathbf{k}, t) = -ic(\mathbf{k} \times \mathbf{E}(\mathbf{k}, t))$
ii) $\mathbf{k}\cdot\mathbf{E}(\mathbf{k}, t) = 0$
iii) $\dot{\mathbf{E}}(\mathbf{k}, t) = ic(\mathbf{k} \times \mathbf{H}(\mathbf{k}, t))$
iv) $\mathbf{k}\cdot\mathbf{H}(\mathbf{k}, t) = 0$.

These four equations are equivalent to Maxwell's equations. This reformulation is well-suited as a starting point for quantum mechanical considerations of the electromagnetic field.

b) Using the results of part(a), prove that the Fourier components of the electric field satisfy the homogeneous wave equation. That is, prove

$$\left[\frac{\partial^2}{\partial t^2} + c^2k^2\right]\mathbf{E}(\mathbf{k}, t) = 0.$$

11. Show informally that

$$\lim_{g\to\infty} \frac{\sin gx}{\pi x}, \qquad g > 0,$$

is a representation of the Dirac δ-function.

12. Prove that

$$\delta(x) = \begin{cases} \lim_{n\to\infty} \left(\frac{1}{|a|} c_n \left[1 - \left(\frac{x}{a}\right)^2\right]^n\right) & \text{if} \quad 0 \le |x| \le |a|, \\ 0 & \text{otherwise} \end{cases}$$

is also a representation of the δ-function, where c_n is given in Eq. (5.28). Show, therefore, that $\delta(ax) = (1/|a|)\,\delta(x)$.

13. Show that if $r = (x^2 + y^2 + z^2)^{1/2}$, then

$$\nabla^2\left(\frac{1}{r}\right) = -4\pi\delta^3(\mathbf{r})$$

[*Hint*: Use Green's Theorem.]

14. Prove Parseval's equation (5.11) in Hilbert space by applying Bessel's completeness relation (5.10) to the function $f + g$ and then subtracting the corresponding equations for f and g.

15. The parity operator P plays a fundamental role in many areas of physics. Show that the Gegenbauer polynomials are eigenfunctions of P; that is, if the leading power of x in $G_n^m(x)$ is even, then $G_n^m(x)$ contains *only* even powers of x. One way to do this is to assume the contrary. Write $G_n^m(x) = g_n^m(x) + u_n^m(x)$, where $g_n^m(x)$ contains only even powers of x and $u_n^m(x)$ contains only odd powers, and conclude that $(L - \lambda_n)g_n^m = 0$ and $(L - \lambda_n)u_n^m = 0$, where L is the Sturm-Liouville operator which has the Gegenbauer polynomials as eigenfunctions. Thus $g_n^m(x)$ and $u_n^m(x)$

are linearly independent (because they belong to different eigenvalues of P) solutions of the Gegenbauer equation, and they are both polynomial solutions. From this conclude that $u_n^m(x) = 0$. [*Hint*: we know that for the Gegenbauer polynomials $\lambda_n = -n(n + 2m + 1)$.] Thus $G_n^m(x)$ contains only even powers of x and has parity eigenvalue $+1$.

16. Prove that the complete set of orthonormal functions on the unit sphere, Y_{lm}, is unique up to a phase factor. Use induction on l for $m = m'$ fixed.
17. If $[x_i, p_j] = i\hbar\delta_{ij}$, where the indices run from one to three, and if we define $\mathbf{L} = \mathbf{r} \times \mathbf{p}$, then show that

$$[L_i, L_j] = i\hbar \sum_k \epsilon_{ijk} L_k .$$

Defining $L_+ = L_x + iL_y$ and $L_- = L_x - iL_y$, show that $[L_+, L_-] = 2\hbar L_z$. Show that if we define

$$\phi_m = L_+^{l+m} \phi_{-l}, \qquad L_- \phi_{-l} = 0 ,$$

where l is an integer, then

$$L_- \phi_m = \hbar(l + m)(l - m + 1)\phi_{m-1} .$$

18. $H_n(x)$ are the Hermite polynomials. Prove for arbitrary n:

a) $H_{2n}(0) = (-1)^n \dfrac{(2n)!}{n!}$, b) $H'_{2n}(0) = 0$,

c) $H_{2n+1}(0) = 0$, d) $H'_{2n+1}(0) = (-1)^n 2 \dfrac{(2n + 1)!}{n!}$.

19. It should be clear from the definitions that pointwise convergence does not imply uniform convergence. The question comes up, Does pointwise convergence imply convergence in the mean? Show that the sequence of functions

$$f_n(x) = \frac{2n^{1/2}}{(\pi/2)^{1/4}} nxe^{-(nx)^2}$$

converges pointwise for all x to the function zero, that is,

$$\lim_{n\to\infty} f_n(x) = 0 ,$$

but does not converge in the mean to the function zero; i. e.,

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} |0 - f_n|^2 \, dx = 1 \neq 0 .$$

20. Let $\{Q_n\}$ be an orthonormal system of Sturm-Liouville polynomials. Prove that the functions $Q'_1, Q'_2, \cdots$ form an orthogonal system of polynomials with weight function $w\alpha$, where the notation is that of Section 5.10.
21. Show that the set of functions $[x^m, x^{m+1}, \cdots]$ is a complete set in the mean on $[-1, +1]$ for $m \geqslant 1$. (Of course, by Weierstrass's theorem, the above is true for $m = 0$.)
22. Obtain the first four orthonormal polynomials for the complete set $[x, x^2, x^3, \cdots]$ on the interval $[-1, +1]$, and write the first few terms of the expansion of $f(x) = 1$ in terms of this set. What do the first two approximations look like graphically?
23. Consider Bessel's equation $Bf_n = n^2 f_n$ on the interval $[0, \infty]$, where

$$B = x^2 \frac{d^2}{dx^2} + x \frac{d}{dx} + x^2 .$$

a) Using the methods discussed in Section 5.10, find the weight function appropriate to Bessel's equation.

b) Show that if $n \neq m$ and $w(x)$ is the weight function of Part (a), then

$$\int_0^\infty J_{2n}(x)J_{2m}(x)w(x)\,dx = \int_0^\infty J_{2n+1}(x)J_{2m+1}(x)w(x)\,dx = 0\,.$$

You may make use of the fact that for small x,

$$J_n(x) = \left(\frac{x}{2}\right)^n\left[1 - \frac{x^2}{2(2n+2)} + \cdots\right],$$

while for large x,

$$J_n(x) = \sqrt{\frac{2}{\pi x}}\left[\cos\left(x - \frac{n\pi}{2} - \frac{\pi}{4}\right)(1 + O(x^2)) - \frac{4n^2-1}{8}\frac{1}{x}\sin\left(x - \frac{n\pi}{2} - \frac{\pi}{4}\right)(1 + O(x^2))\right].$$

Both of these results can be readily derived from complex variable techniques, which will be discussed in Chapter 6.

c) Using the asymptotic formulas given in Part (b), evaluate

$$\int_0^\infty J_{2n+1}(x)J_{2m}(x)w(x)\,dx\,.$$

d) Is the operator, B, in Bessel's equation an Hermitian operator with respect to the space of functions which satisfy

$$\int_0^\infty |f(x)|^2 w(x)\,dx < \infty\,,$$

where $w(x)$ is given by Part (a)? [It is understood, of course, that B can act only on functions f having the property that Bf lies in the Hilbert space if f does.]

24. We have shown that the Legendre polynomials, $P_l(x)$, are solutions to

$$(1 - x^2)f''(x) - 2xf'(x) + l(l+1)f(x) = 0\,, \qquad l = 0, 1, 2, \cdots,$$

which are orthogonal on $[-1, 1]$. However, these are not the only solutions; they are merely the only polynomial solutions.

a) To reassure yourself on this point, show that the function

$$\tanh^{-1}x = \frac{1}{2}\log\left(\frac{1+x}{1-x}\right)$$

satisfies Legendre's equation with $l = 0$. Note that this function is singular at $x = \pm 1$.

b) Show that if we look for a solution of Legendre's equation of the form

$$Q_l(x) = P_l(x)\tanh^{-1}x + \Pi_l(x)\,,$$

then a solution of this form always exists where $\Pi_l(x)$ is a polynomial of degree $l - 1$. Find the inhomogeneous differential equation that $\Pi_l(x)$ must satisfy, and show that this equation always has a solution. What type of boundary condition has been applied in taking $\Pi_l(x)$ to be a polynomial of order $l - 1$?

c) Since $\Pi_l(x)$ is a polynomial of degree $l - 1$ on $[-1, 1]$, it can be written as a linear combination of Legendre polynomials. Find an expression for the co-

efficients in this linear combination. To be precise, show that

$$\Pi_l(x) = 2 \sum_{n=0,1}^{l-1}{}' \frac{2n+1}{(n-l)(n+l+1)} P_n(x) ,$$

where the sum begins at $n = 0$ if l is odd and at $n = 1$ if l is even. The prime on the summation indicates that the sum runs in steps of two.

d) Show that $(Q_l, P_l) = 0$.
[*Hint*: Don't do any integrals.]

e) From their form it is clear that the $Q_l(x)$ lie in the Hilbert space of square-integrable functions on $[-1, 1]$. Do they form an orthogonal set of functions on $[-1, 1]$? Explain.

25. Let us define $E(\lambda)$, an operator-valued function, as in Section 5.11. Denote by λ_n and ϕ_n $(n = 1, 2, \cdots, N)$ the nondegenerate eigenvalues and corresponding eigenvectors of an Hermitian operator, A, and call P_n the projection operator associated with ϕ_n. Then define

$$E(\lambda) = \begin{cases} 0 \text{ for } \lambda < \lambda_1 , \\ \displaystyle\sum_{n=1}^{\nu} P_n \text{ for } \lambda_\nu \leq \lambda < \lambda_{\nu+1} , \qquad \nu = 1, 2, \cdots, N-1 , \\ \displaystyle\sum_{n=1}^{N} P_n = I \text{ for } \lambda \geq \lambda_N . \end{cases}$$

Show that

i) $E(\lambda_1)E(\lambda_2) = E(\lambda_{\min})$, where $\lambda_{\min} = \min(\lambda_1, \lambda_2)$, ii) $E(\lambda)^2 = E(\lambda)$.

26. Let X be the self-adjoint operator on the Hilbert space of square-integrable functions defined by $Xf(x) \equiv xf(x)$.

i) Is X well defined on all elements of the space?
If not, characterize the elements for which it is defined.

ii) Prove that $E_X(\xi)$, defined by

$$E_X(\xi)f(x) = \begin{cases} f(x) & \text{if } x \leq \xi \\ 0 & \text{if } x > \xi \end{cases} ,$$

is the resolution of the identity belonging to X.

27. Show that if $E_A(\lambda)$ suffers a discontinuous jump at an eigenvalue $\bar{\lambda}$ of A, then

$$[E_A(\lambda_+) - E_A(\lambda_-)]\psi = \sum_{\nu=0}^{\mu} \Phi_{\bar{\lambda}}^{(\nu)}(\Phi_{\bar{\lambda}}^{(\nu)}, \psi) ,$$

where μ is the multiplicity of $\bar{\lambda}$ and $\{\Phi_{\bar{\lambda}}^{(\nu)}: \nu = 1, 2, \cdots, \mu\}$ is a set of μ orthonormal eigenvectors of A belonging to $\bar{\lambda}$, and λ_+ and λ_- are points on either side of $\bar{\lambda}$ in the intervals of constancy of $E(\lambda)$; that is, $E(\lambda)$ is constant in the intervals $[\bar{\lambda}, \lambda_+]$ and $[\lambda_-, \bar{\lambda}]$. A is any self-adjoint operator.

28. Given that $E_A(\lambda)$ is the resolution of the identity belonging to A, show that

$$[E_A(\lambda_2) - E_A(\lambda_1)]^2 = [E_A(\lambda_2) - E_A(\lambda_1)]$$

if $\lambda_2 > \lambda_1$.

29. i) What is the expectation value of the operator X of Problem 5.26 in the state $\Psi(x) = (1/\sqrt[4]{\pi a^2})e^{-x^2/2a^2}$?
    ii) What is the dispersion, ΔX, about this average value?
    iii) What is the expectation value of $p_x = -ih\,d/dx$ in the above state?
    iv) What is the dispersion, Δp_x, about this average value?

    Note that as the dispersion in X gets smaller (i.e., as a gets smaller, which means that the wave function Ψ becomes more sharply peaked about $x = 0$), the dispersion in p_x gets larger. The product of the two dispersions, $\Delta X \Delta p_x$, is a constant, equal to $\hbar/2$. One can show that for any state Ψ we must always have $\Delta X \Delta p_x \geqslant \hbar/2$.

30. Verify that for all z not in the spectrum of A,

$$(A - zI)^{-1} = \int_{-\infty}^{\infty} \frac{1}{\lambda - z}\, dE_A(\lambda) ,$$

    where A is a self-adjoint operator whose associated spectral resolution is $E_A(\lambda)$.

31. Suppose that we wish to find an eigenvector and an eigenvalue of a self-adjoint operator $A = A_0 + \epsilon A_1$. Imagine that we know the resolution of the identity, $E_{A_0}(\lambda)$, belonging to A_0, and assume that for ϵ sufficiently small any eigenvalue of A can be written as

$$\lambda_n = \lambda_n^{(0)} + \epsilon\lambda_n^{(1)} + \epsilon^2\lambda_n^{(2)} + \cdots$$

    and that the corresponding eigenvector can be written as

$$\phi_n = \phi_n^{(0)} + \epsilon\phi_n^{(1)} + \epsilon^2\phi_n^{(2)} + \cdots$$

    where $A_0\phi_n^{(0)} = \lambda_n^{(0)}\phi_n^{(0)}$. Show in the manner of Section 4.11 that we can find an inhomogeneous equation for $\phi_n^{(1)}$, and that the solution to this equation which satisfies $(\phi_n^{(0)}, \phi_n^{(1)}) = 0$ is

$$\phi_n^{(1)} = -\int_{S_0} \frac{1}{\lambda - \lambda_n^{(0)}}\, d[E_{A_0}(\lambda)(A_1 - \lambda_n^{(1)})\phi_n^{(0)}] , \qquad (1)$$

    where $\lambda_n^{(1)} = (\phi_n^{(0)}, A_1\phi_n^{(0)})$ and S_0 denotes the entire spectrum of A_0 except for an arbitrarily small closed interval about the point $\lambda_n^{(0)}$. We assume that the spectra of A and A_0 are nondegenerate. Show that in the case of a finite-dimensional vector space, the results of Section 4.11 are obtained from Eq. (1).

32. i) Let A be a self-adjoint operator on a Hilbert space H. Assume that the spectrum of A consists of a finite set of points, $\{\lambda_i < 0\colon\ i = 1, 2, \cdots, N\}$, plus the positive real line. Show that if the lowest (i.e., most negative) eigenvalue of A, λ_1, is nondegenerate, then $\lambda_1 = \min_{\psi \in H} (\psi, A\psi)$,
    where ψ is constrained by $||\psi|| = 1$.

    [*Hint*: Note that any vector in H can be written as

$$\psi = c_0\phi_0 + \sum_{i=1}^{N} c_i\phi_i ,$$

    where ϕ_i are the eigenvectors of A, and ϕ_0 is a vector orthogonal to all the ϕ_i $(i = 1, 2, \cdots)$. Thus show that

$$(\psi, A\psi) = \sum_{i=0}^{N} |c_i|^2\lambda_i ,$$

where $\{\lambda_i\colon i = 1, 2, \cdots\}$ is the set of eigenvalues of A and λ_0 is a *positive* number. Now proceed as in Section 4.9.]

ii) Write down and prove a result analogous to Eq. (1) for λ_2. How would the result generalize for λ_n?

33. Consider the quantum-mechanical observables L_x, L_y, and L_z, which are represented as matrices as follows:

$$L_x = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad L_y = \begin{bmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{bmatrix}, \quad L_z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

The Hamiltonian for this system has the form $H = H_0 + \alpha L_z$, where H_0 commutes with L_x, L_y, and L_z.

a) Suppose that at $t = 0$ we have prepared the system in an eigenstate of L_z, namely, the state belonging to the eigenvalue $m = +1$ of L_z. If we measure L_z at the later time $t = T$, what is the probability that we will find the value $+1$? the value 0? the value -1?

b) Suppose that instead of measuring L_z at time $t = T$, we measure L_y. What are the possible values of L_x which we can find? What is the probability of finding each of these values?

c) Suppose that at $t = T$ we measure L_y and find with certainty the value -1. After a time τ has elapsed (i.e., at $t = T + \tau$), we remeasure L_y. What is the probability of finding the various allowed values of L_x?

FURTHER READINGS

CHURCHILL, R. V., *Fourier Series and Boundary Value Problems.* New York: McGraw-Hill, 1944. An exceptionally readable book. Good for self-study.

COURANT, R., and D. HILBERT, *Methods of Mathematical Physics*, Vol. I, New York: Wiley-Interscience, 1961.

DAVIS, P. J., *Interpolation and Approximation.* Waltham, Mass.: Blaisdell 1963. An advanced account of the various types of convergence and their roles in the approximation of functions.

DIRAC, P. A. M. *The Principles of Quantum Mechanics*, Fourth edition. New York: Oxford University Press, 1958.

DUNFORD, N., and J. T. SCHWARTZ, *Linear Operators*, Vol. II: Spectral Theory. New York: Wiley-Interscience, 1967. A definitive account of the theory of self-adjoint operators in Hilbert space.

HOCHSTADT, H., *Special Functions of Mathematical Physics.* New York: Holt, Rinehart and Winston, 1961. A brief, very readable treatment of orthogonal polynomials, Legendre, Bessel, and Mathieu functions. On p. 20 there is a highly intuitive probabilistic proof of Weierstrass's theorem in terms of the Bernstein polynomials.

JACKSON, J. D. *Mathematics for Quantum Mechanics.* New York: W. A. Benjamin, 1962.

JAUCH, J.M., *Foundations of Quantum Mechanics.* Reading, Mass: Addison-Wesley, 1968.