Chapter 2

Function Spaces

Many differential equations of physics are relations involving *linear differential operators*. These operators, like matrices, are linear maps acting on vector spaces. The new feature is that the elements of the vector spaces are functions, and the spaces are infinite dimensional. We can try to survive in these vast regions by relying on our experience in finite dimensions, but sometimes this fails, and more sophistication is required.

2.1 Motivation

In the previous chapter we considered two variational problems:

1) Find the stationary points of

$$F(\mathbf{x}) = \frac{1}{2}\mathbf{x}\cdot\mathbf{A}\mathbf{x} = \frac{1}{2}x_i A_{ij} x_j \tag{2.1}$$

on the surface $\mathbf{x}\cdot\mathbf{x} = 1$. This led to the matrix eigenvalue equation

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}. \tag{2.2}$$

2) Find the stationary points of

$$J[y] = \int_a^b \frac{1}{2}\left\{p(x)(y')^2 + q(x)y^2\right\}\,dx, \tag{2.3}$$

subject to the conditions $y(a) = y(b) = 0$ and

$$K[y] = \int_a^b y^2\,dx = 1. \tag{2.4}$$

This led to the differential equation

$$-(py')' + qy = \lambda y, \quad y(a) = y(b) = 0. \tag{2.5}$$

There will be a solution that satisfies the boundary conditions only for a discrete set of values of λ.

The stationary points of both function and functional are therefore determined by *linear eigenvalue problems*. The only difference is that the finite matrix in the first is replaced in the second by a linear differential operator. The theme of the next few chapters is an exploration of the similarities and differences between finite matrices and linear differential operators. In this chapter we will focus on how the functions on which the derivatives act can be thought of as vectors.

2.1.1 Functions as vectors

Consider $F[a, b]$, the set of all real (or complex) valued functions $f(x)$ on the interval $[a, b]$. This is a *vector space* over the field of the real (or complex) numbers: Given two functions $f_1(x)$ and $f_2(x)$, and two numbers λ_1 and λ_2, we can form the sum $\lambda_1 f_1(x)+\lambda_2 f_2(x)$ and the result is still a function on the same interval. Examination of the axioms listed in appendix A will show that $F[a, b]$ possesses all the other attributes of a vector space as well. We may think of the array of numbers $(f(x))$ for $x \in [a, b]$ as being the components of the vector. Since there is an infinity of independent components — one for each point x — the space of functions is infinite dimensional.

The set of *all* functions is usually too large for us. We will restrict ourselves to subspaces of functions with nice properties, such as being continuous or differentiable. There is some fairly standard notation for these spaces: The space of C^n functions (those which have n continuous derivatives) is called $C^n[a, b]$. For smooth functions (those with derivatives of all orders) we write $C^\infty[a, b]$. For the space of analytic functions (those whose Taylor expansion actually converges to the function) we write $C^\omega[a, b]$. For C^∞ functions defined on the whole real line we write $C^\infty(\mathbb{R})$. For the subset of functions with compact support (those that vanish outside some finite interval) we write $C_0^\infty(\mathbb{R})$. There are no non-zero analytic functions with compact support: $C_0^\omega(\mathbb{R}) = \{0\}$.

2.2 Norms and inner products

We are often interested in "how large" a function is. This leads to the idea of *normed* function spaces. There are many measures of function size. Suppose $R(t)$ is the number of inches per hour of rainfall. If your are a farmer you are probably most concerned with the total amount of rain that falls. A big rain has big $\int |R(t)|\,dt$. If you are the Urbana city engineer worrying about the capacity of the sewer system to cope with a downpour, you are primarily concerned with the maximum value of $R(t)$. For you a big rain has a big "$\sup |R(t)|$."[1]

2.2.1 Norms and convergence

We can seldom write down an exact solution function to a real-world problem. We are usually forced to use numerical methods, or to expand as a power series in some small parameter. The result is a sequence of approximate solutions $f_n(x)$, which we hope will converge to the desired exact solution $f(x)$ as we make the numerical grid smaller, or take more terms in the power series.

Because there is more than one way to measure of the "size" of a function, the convergence of a sequence of functions f_n to a limit function f is not as simple a concept as the convergence of a sequence of numbers x_n to a limit x. Convergence means that the distance between the f_n and the limit function f gets smaller and smaller as n increases, so each different measure of this distance provides a new notion of what it means to converge. We are not going to make much use of formal "ε, δ" analysis, but you must realize that this distinction between different forms of convergence is not merely academic: real-world engineers must be precise about the kind of errors they are prepared to tolerate, or else a bridge they design might collapse. Graduate-level *engineering* courses in mathematical methods therefore devote much time to these issues. While physicists do not normally face the same legal liabilities as engineers, we should at least have it clear in our own minds what we mean when we write that $f_n \to f$.

[1] Here "sup," short for *supremum*, is synonymous with the "least upper bound" of a set of numbers, *i.e.* the smallest number that is exceeded by no number in the set. This concept is more useful than "maximum" because the supremum need not be an element of the set. It is an axiom of the real number system that any bounded set of real numbers has a least upper bound. The "greatest lower bound" is denoted "inf", for *infimum*.

Here are some common forms of convergence:

i) If, for each x in its domain of definition $\mathcal{D}$, the set of numbers $f_n(x)$ converges to $f(x)$, then we say the sequence converges *pointwise*.

ii) If the maximum separation

$$\sup_{x\in\mathcal{D}} |f_n(x) - f(x)| \tag{2.6}$$

goes to zero as $n \to \infty$, then we say that f_n converges to f *uniformly* on $\mathcal{D}$.

iii) If

$$\int_{\mathcal{D}} |f_n(x) - f(x)|\, dx \tag{2.7}$$

goes to zero as $n \to \infty$, then we say that f_n converges *in the mean* to f on $\mathcal{D}$.

Uniform convergence implies pointwise convergence, but not *vice versa*. If $\mathcal{D}$ is a finite interval, then uniform convergence implies convergence in the mean, but convergence in the mean implies neither uniform nor pointwise convergence.

Example: Consider the sequence $f_n = x^n$ ($n = 1, 2, \ldots$) and $\mathcal{D} = [0, 1)$. Here, the round and square bracket notation means that the point $x = 0$ is included in the interval, but the point 1 is excluded.

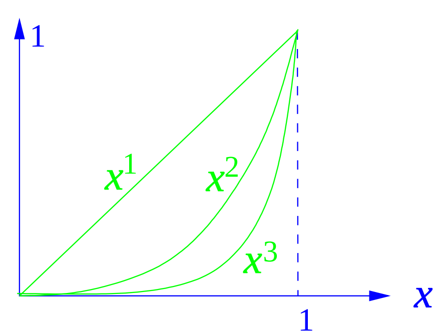

Figure 2.1: $x^n \to 0$ *on* $[0, 1)$, *but not uniformly.*

As n becomes large we have $x^n \to 0$ pointwise in $\mathcal{D}$, but the convergence is *not* uniform because

$$\sup_{x\in\mathcal{D}} |x^n - 0| = 1 \tag{2.8}$$

for all n.

Example: Let $f_n = x^n$ with $\mathcal{D} = [0,1]$. Now the the two square brackets mean that both $x = 0$ and $x = 1$ are to be included in the interval. In this case we have neither uniform nor pointwise convergence of the x^n to zero, but $x^n \to 0$ in the mean.

We can describe uniform convergence by means of a *norm* — a generalization of the usual measure of the length of a vector. A norm, denoted by $\|f\|$, of a vector f (a function, in our case) is a real number that obeys

i) positivity: $\|f\| \geq 0$, and $\|f\| = 0 \Leftrightarrow f = 0$,
ii) the *triangle inequality*: $\|f + g\| \leq \|f\| + \|g\|$,
iii) linear homogeneity: $\|\lambda f\| = |\lambda| \|f\|$.

One example is the "sup" norm, which is defined by

$$\|f\|_\infty = \sup_{x \in \mathcal{D}} |f(x)|. \tag{2.9}$$

This number is guaranteed to be finite if f is continuous and $\mathcal{D}$ is compact. In terms of the sup norm, uniform convergence is the statement that

$$\lim_{n \to \infty} \|f_n - f\|_\infty = 0. \tag{2.10}$$

2.2.2 Norms from integrals

The space $L^p[a, b]$, for any $1 \leq p < \infty$, is defined to be our $F[a, b]$ equipped with

$$\|f\|_p = \left(\int_a^b |f(x)|^p \, dx \right)^{1/p}, \tag{2.11}$$

as the measure of length, and with a restriction to functions for which $\|f\|_p$ is finite.

We say that $f_n \to f$ in L^p if the L^p distance $\|f - f_n\|_p$ tends to zero. We have already seen the L^1 measure of distance in the definition of convergence in the mean. As in that case, convergence in L^p says nothing about pointwise convergence.

We would like to regard $\|f\|_p$ as a norm. It is possible, however, for a function to have $\|f\|_p = 0$ without f being identically zero — a function that vanishes at all but a finite set of points, for example. This pathology violates number i) in our list of requirements for something to be called a norm, but we circumvent the problem by simply declaring such functions to be zero. This means that elements of the L^p spaces are not really functions, but only *equivalence classes* of functions — two functions being regarded as the same is they differ by a function of zero length. Clearly these spaces are not for use when anything significant depends on the value of the function at any precise point. They *are* useful in physics, however, because we can never measure a quantity at an exact position in space or time. We usually measure some sort of local average.

The L^p norms satisfy the triangle inequality for all $1 \le p \le \infty$, although this is not exactly trivial to prove.

An important property for any space to have is that of being *complete*. Roughly speaking, a space is complete if when some sequence of elements of the space look as if they are converging, then they are indeed converging and their limit is an element of the space. To make this concept precise, we need to say what we mean by the phrase "look as if they are converging." This we do by introducing the idea of a *Cauchy sequence*.

Definition: A sequence f_n in a normed vector space is Cauchy *if for any $\varepsilon > 0$ we can find an N such that $n, m > N$ implies that $\|f_m - f_n\| < \varepsilon$.*

This definition can be loosely paraphrased to say that the elements of a Cauchy sequence get arbitrarily close to each other as $n \to \infty$.

A normed vector space is *complete* with respect to its norm if every Cauchy sequence actually converges to some element in the space. Consider. for example, the normed vector space $\mathbb{Q}$ of rational numbers with distance measured in the usual way as $\|q_1 - q_2\| \equiv |q_1 - q_2|$. The sequence

$$\begin{aligned} q_0 &= 1.0, \\ q_1 &= 1.4, \\ q_2 &= 1.41, \\ q_3 &= 1.414, \\ &\vdots \end{aligned}$$

consisting of successive decimal approximations to $\sqrt{2}$, obeys

$$|q_n - q_m| < \frac{1}{10^{\min(n,m)}} \tag{2.12}$$

and so is Cauchy. Pythagoras famously showed that $\sqrt{2}$ is irrational, however, and so this sequence of rational numbers has no limit in $\mathbb{Q}$. Thus $\mathbb{Q}$ is *not* complete. The space $\mathbb{R}$ of real numbers is constructed by filling in the gaps between the rationals, and so *completing* $\mathbb{Q}$. A real number such as $\sqrt{2}$ is defined as a Cauchy sequence of rational numbers (by giving a rule, for example, that determines its infinite decimal expansion), with two rational sequences q_n and q'_n defining the same real number if $q_n - q'_n$ converges to zero.

A complete normed vector space is called a *Banach space*. If we interpret the norms as Lebesgue integrals[2] then the $L^p[a,b]$ are complete, and therefore Banach spaces. The theory of Lebesgue integration is rather complicated, however, and is not really necessary. One way of avoiding it is explained in exercise 2.2.

Exercise 2.1: Show that any convergent sequence is Cauchy.

2.2.3 Hilbert space

The Banach space $L^2[a,b]$ is special in that it is also a *Hilbert space*. This means that its norm is derived from an inner product. If we define the inner product

$$\langle f, g\rangle = \int_a^b f^* g\, dx \tag{2.13}$$

then the $L^2[a,b]$ norm can be written

$$\|f\|_2 = \sqrt{\langle f, f\rangle}. \tag{2.14}$$

When we omit the subscript on a norm, we mean it to be this one. You are probably familiar with this Hilbert space from your quantum mechanics classes.

Being positive definite, the inner product satisfies the *Cauchy-Schwarz-Bunyakovsky inequality*

$$|\langle f, g\rangle| \le \|f\|\|g\|. \tag{2.15}$$

That this is so can be seen by observing that

$$\langle \lambda f + \mu g, \lambda f + \mu g\rangle = (\,\lambda^*,\ \mu^*\,)\begin{pmatrix} \|f\|^2 & \langle f, g\rangle \\ \langle f, g\rangle^* & \|g\|^2 \end{pmatrix}\begin{pmatrix} \lambda \\ \mu \end{pmatrix}, \tag{2.16}$$

must be non-negative for any choice of λ and μ. We therefore select $\lambda = \|g\|$, $\mu = -\langle f, g\rangle^* \|g\|^{-1}$, in which case the non-negativity of (2.16) becomes the statement that

$$\|f\|^2\|g\|^2 - |\langle f, g\rangle|^2 \ge 0. \tag{2.17}$$

[2]The "L" in L^p honours Henri Lebesgue. Banach spaces are named after Stefan Banach, who was one of the founders of functional analysis, a subject largely developed by him and other habitués of the Scottish Café in Lvóv, Poland.

From Cauchy-Schwarz-Bunyakovsky we can establish the triangle inequality:

$$\begin{aligned}
\|f+g\|^2 &= \|f\|^2+\|g\|^2+2\mathrm{Re}\langle f,g\rangle \\
&\leq \|f\|^2+\|g\|^2+2|\langle f,g\rangle|, \\
&\leq \|f\|^2+\|g\|^2+2\|f\|\|g\|, \\
&= (\|f\|+\|g\|)^2,
\end{aligned} \tag{2.18}$$

so

$$\|f+g\| \leq \|f\|+\|g\|. \tag{2.19}$$

A second important consequence of Cauchy-Schwarz-Bunyakovsky is that if $f_n \to f$ in the sense that $\|f_n - f\| \to 0$, then

$$\begin{aligned}
|\langle f_n, g\rangle - \langle f, g\rangle| &= |\langle (f_n - f), g\rangle| \\
&\leq \|f_n - f\|\,\|g\|
\end{aligned} \tag{2.20}$$

tends to zero, and so

$$\langle f_n, g\rangle \to \langle f, g\rangle. \tag{2.21}$$

This means that the inner product $\langle f, g\rangle$ is a *continuous* functional of f and g. Take care to note that this continuity hinges on $\|g\|$ being finite. It is for this reason that we do not permit $\|g\| = \infty$ functions to be elements of our Hilbert space.

Orthonormal sets

Once we are in possession of an inner product, we can introduce the notion of an *orthonormal set*. A set of functions $\{u_n\}$ is orthonormal if

$$\langle u_n, u_m\rangle = \delta_{nm}. \tag{2.22}$$

For example,

$$2\int_0^1 \sin(n\pi x)\sin(m\pi x)\,dx = \delta_{nm}, \quad n, m = 1, 2, \ldots \tag{2.23}$$

so the set of functions $u_n = \sqrt{2}\sin n\pi x$ is orthonormal on $[0, 1]$. This set of functions is also *complete* — in a different sense, however, from our earlier use of this word. A orthonormal set of functions is said to be complete if any function f for which $\|f\|^2$ is finite, and hence f an element of the Hilbert space, has a convergent expansion

$$f(x) = \sum_{n=0}^{\infty} a_n u_n(x).$$

If we assume that such an expansion exists, and that we can freely interchange the order of the sum and integral, we can multiply both sides of this expansion by $u_m^*(x)$, integrate over x, and use the orthonormality of the u_n's to read off the expansion coefficients as $a_n = \langle u_n, f\rangle$. When

$$\|f\|^2 = \int_0^1 |f(x)|^2\, dx \tag{2.24}$$

and $u_n = \sqrt{2}\sin(n\pi x)$, the result is the half-range sine Fourier series.

Example: Expanding unity. Suppose $f(x) = 1$. Since $\int_0^1 |f|^2 dx = 1$ is finite, the function $f(x) = 1$ can be represented as a convergent sum of the $u_n = \sqrt{2}\sin(n\pi x)$.

The inner product of f with the u_n's is

$$\langle u_n, f\rangle = \int_0^1 \sqrt{2}\sin(n\pi x)\, dx = \begin{cases} 0, & n \text{ even},\\ \frac{2\sqrt{2}}{n\pi}, & n \text{ odd}.\end{cases}$$

Thus,

$$1 = \sum_{n=0}^{\infty} \frac{4}{(2n+1)\pi}\sin\Big((2n+1)\pi x\Big), \quad \text{in} \quad L^2[0,1]. \tag{2.25}$$

It is important to understand that the sum converges to the left-hand side in the closed interval $[0,1]$ only in the L^2 sense. The series does not converge pointwise to unity at $x = 0$ or $x = 1$ — every term is zero at these points.

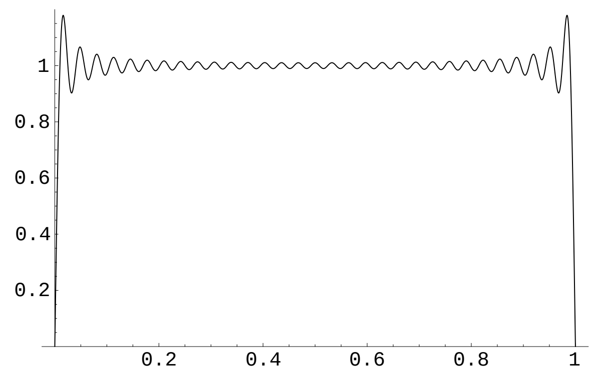

Figure 2.2: *The sum of the first 31 terms in the sine expansion of $f(x) = 1$.*

Figure 2.2 shows the sum of the series up to and including the term with $n = 30$. The $L^2[0,1]$ measure of the distance between $f(x) = 1$ and this sum is

$$\int_0^1 \left| 1 - \sum_{n=0}^{30} \frac{4}{(2n+1)\pi} \sin\Big((2n+1)\pi x\Big) \right|^2 dx = 0.00654. \tag{2.26}$$

We can make this number as small as we desire by taking sufficiently many terms.

It is perhaps surprising that a set of functions that vanish at the endpoints of the interval can be used to expand a function that does not vanish at the ends. This exposes an important technical point: Any finite sum of continuous functions vanishing at the endpoints is also a continuous function vanishing at the endpoints. It is therefore tempting to talk about the "subspace" of such functions. This set is indeed a vector space, and a subset of the Hilbert space, but it is not itself a Hilbert space. As the example shows, a Cauchy sequence of continuous functions vanishing at the endpoints of an interval can converge to a continuous function that does *not* vanish there. The "subspace" is therefore not complete in our original meaning of the term. The set of continuous functions vanishing at the endpoints fits into the whole Hilbert space much as the rational numbers fit into the real numbers: A finite sum of rationals is a rational number, but an infinite sum of rationals is not in general a rational number and we can obtain any real number as the limit of a sequence of rational numbers. The rationals $\mathbb{Q}$ are therefore a *dense* subset of the reals, and, as explained earlier, the reals are obtained by completing the set of rationals by adding to this set its limit points. In the same sense, the set of continuous functions vanishing at the endpoints is a dense subset of the whole Hilbert space and the whole Hilbert space is its completion.

Exercise 2.2: In this technical exercise we will explain in more detail how we "complete" a Hilbert space. The idea is to mirror the construction to the real numbers and *define* the elements of the Hilbert space to be Cauchy sequences of continuous functions. To specify a general element of $L^2[a,b]$ we must therefore exhibit a Cauchy sequence $f_n \in C[a,b]$. The choice is not unique: two Cauchy sequences $f_n^{(1)}(x)$ and $f_n^{(2)}(x)$ will specify the the same element if

$$\lim_{n\to\infty} \|f_n^{(1)} - f_n^{(2)}\| = 0.$$

Such sequences are said to be *equivalent*. For convenience, we will write "$\lim_{n\to\infty} f_n = f$" but bear in mind that, in this exercise, this means that the sequence f_n *defines* the symbol f, and not that f is the limit of the sequence, as this limit need have no prior existence. We have deliberately written "f", and not "$f(x)$", for the "limit function" to warn us that f is assigned no unique numerical value at any x. A continuous function $f(x)$ can still be considered to be an element of $L^2[a,b]$—take a sequence in which every $f_n(x)$ is equal to $f(x)$—but an equivalent sequence of $f_n(x)$ can alter the limiting $f(x)$ on a set of measure zero without changing the resulting element $f \in L^2[a,b]$.

i) If f_n and g_n are Cauchy sequences defining f, g, respectively, it is natural to try to define the inner product $\langle f, g\rangle$ by setting

$$\langle f, g\rangle \equiv \lim_{n\to\infty} \langle f_n, g_n\rangle.$$

Use the Cauchy-Schwarz-Bunyakovsky inequality to show that the numbers $F_n = \langle f_n, g_n\rangle$ form a Cauchy sequence in $\mathbb{C}$. Since $\mathbb{C}$ is complete, deduce that this limit exists. Next show that the limit is unaltered if either f_n or g_n is replaced by an equivalent sequence. Conclude that our tentative inner product is well defined.

ii) The next, and harder, task is to show that the "completed" space is indeed complete. The problem is to show that given a Cauchy sequence $f_k \in L^2[a,b]$, where the f_k are *not* necessarily in $C[a,b]$, has a limit in $L^2[a,b]$. Begin by taking Cauchy sequences $f_{ki} \in C[a,b]$ such that $\lim_{i\to\infty} f_{ki} = f_k$. Use the triangle inequality to show that we can select a subsequence $f_{k,i(k)}$ that is Cauchy and so defines the desired limit.

Later we will show that the elements of $L^2[a,b]$ can be given a concrete meaning as *distributions*.

Best approximation

Let $u_n(x)$ be an orthonormal set of functions. The sum of the first N terms of the Fourier expansion of $f(x)$ in the u_n, is the closest— measuring distance with the L^2 norm — that one can get to f whilst remaining in the space spanned by $u_1, u_2, \ldots, u_N$.

To see this, consider the square of the error-distance:

$$\begin{aligned} \Delta &\stackrel{\text{def}}{=} \|f - \sum_1^N a_n u_n\|^2 = \langle f - \sum_{m=1}^N a_m u_m, f - \sum_{n=1}^N a_n u_n \rangle \\ &= \|f\|^2 - \sum_{n=1}^N a_n \langle f, u_n \rangle - \sum_{m=1}^N a_m^* \langle u_m, f \rangle + \sum_{n,m=1}^N a_m^* a_n \langle u_m, u_n \rangle \\ &= \|f\|^2 - \sum_{n=1}^N a_n \langle f, u_n \rangle - \sum_{m=1}^N a_m^* \langle u_m, f \rangle + \sum_{n=1}^N |a_n|^2, \end{aligned} \tag{2.27}$$

In the last line we have used the orthonormality of the u_n. We can complete the squares, and rewrite Δ as

$$\Delta = \|f\|^2 - \sum_{n=1}^N |\langle u_n, f \rangle|^2 + \sum_{n=1}^N |a_n - \langle u_n, f \rangle|^2. \tag{2.28}$$

We seek to minimize Δ by a suitable choice of coefficients a_n. The smallest we can make it is

$$\Delta_{\text{min}} = \|f\|^2 - \sum_{n=1}^N |\langle u_n, f \rangle|^2, \tag{2.29}$$

and we attain this bound by setting each of the $|a_n - \langle u_n, f \rangle|$ equal to zero. That is, by taking

$$a_n = \langle u_n, f \rangle. \tag{2.30}$$

Thus the Fourier coefficients $\langle u_n, f \rangle$ are the optimal choice for the a_n.

Suppose we have some *non-orthogonal* collection of functions g_n, $n = 1, \ldots N$, and we have found the best approximation $\sum_{n=1}^N a_n g_n(x)$ to $f(x)$. Now suppose we are given a g_{N+1} to add to our collection. We may then seek an improved approximation $\sum_{n=1}^{N+1} a'_n g_n(x)$ by including this new function — but finding this better fit will generally involve tweaking *all* the a_n, not just trying different values of a_{N+1}. The great advantage of approximating by orthogonal functions is that, given another member of an orthonormal family, we can improve the precision of the fit by adjusting only the coefficient of the new term. We do not have to perturb the previously obtained coefficients.

Parseval's theorem

The "best approximation" result from the previous section allows us to give an alternative definition of a "complete orthonormal set," and to obtain the formula $a_n = \langle u_n, f\rangle$ for the expansion coefficients without having to assume that we can integrate the infinite series $\sum a_n u_n$ term-by-term. Recall that we said that a set of points S is a *dense* subset of a space T if any given point $x \in T$ is the limit of a sequence of points in S, *i.e.* there are elements of S lying arbitrarily close to x. For example, the set of rational numbers $\mathbb{Q}$ is a dense subset of $\mathbb{R}$. Using this language, we say that a set of orthonormal functions $\{u_n(x)\}$ is complete if the set of all finite linear combinations of the u_n is a dense subset of the entire Hilbert space. This guarantees that, by taking N sufficently large, our best approximation will approach arbitrarily close to our target function $f(x)$. Since the best approximation containing all the u_n up to u_N is the N-th partial sum of the Fourier series, this shows that the Fourier series actually converges to f.

We have therefore proved that if we are given $u_n(x)$, $n = 1, 2, \ldots$, a complete orthonormal set of functions on $[a, b]$, then any function for which $\|f\|^2$ is finite can be expanded as a convergent Fourier series

$$f(x) = \sum_{n=1}^{\infty} a_n u_n(x), \tag{2.31}$$

where

$$a_n = \langle u_n, f\rangle = \int_a^b u_n^*(x) f(x)\, dx. \tag{2.32}$$

The convergence is guaranteed only in the L^2 sense that

$$\lim_{N\to\infty} \int_a^b \left| f(x) - \sum_{n=1}^{N} a_n u_n(x)\right|^2 dx = 0. \tag{2.33}$$

Equivalently

$$\Delta_N = \|f - \sum_{n=1}^{N} a_n u_n\|^2 \to 0 \tag{2.34}$$

as $N \to \infty$. Now, we showed in the previous section that

$$\begin{aligned}\Delta_N &= \|f\|^2 - \sum_{n=1}^{N} |\langle u_n, f\rangle|^2 \\ &= \|f\|^2 - \sum_{n=1}^{N} |a_n|^2, \end{aligned} \tag{2.35}$$

and so the L^2 convergence is equivalent to the statement that

$$\|f\|^2 = \sum_{n=1}^{\infty} |a_n|^2. \tag{2.36}$$

This last result is called *Parseval's theorem*.
Example: In the expansion (2.25), we have $\|f^2\| = 1$ and

$$|a_n|^2 = \begin{cases} 8/(n^2\pi^2), & n \text{ odd}, \\ 0, & n \text{ even}. \end{cases} \tag{2.37}$$

Parseval therefore tells us tells us that

$$\sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} = 1 + \frac{1}{3^2} + \frac{1}{5^2} + \cdots = \frac{\pi^2}{8}. \tag{2.38}$$

Example: The functions $u_n(x) = \frac{1}{\sqrt{2\pi}} e^{inx}$, $n \in \mathbb{Z}$ form a complete orthonormal set on the interval $[-\pi, \pi]$. Let $f(x) = \frac{1}{\sqrt{2\pi}} e^{i\zeta x}$. Then its Fourier expansion is

$$\frac{1}{\sqrt{2\pi}} e^{i\zeta x} = \sum_{n=-\infty}^{\infty} c_n \frac{1}{\sqrt{2\pi}} e^{inx}, \quad -\pi < x < \pi, \tag{2.39}$$

where

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\zeta x} e^{-inx}\, dx = \frac{\sin(\pi(\zeta - n))}{\pi(\zeta - n)}. \tag{2.40}$$

We also have that

$$\|f\|^2 = \int_{-\pi}^{\pi} \frac{1}{2\pi} dx = 1. \tag{2.41}$$

Now Parseval tells us that

$$\|f\|^2 = \sum_{n=-\infty}^{\infty} \frac{\sin^2(\pi(\zeta - n))}{\pi^2(\zeta - n)^2}, \tag{2.42}$$

the left hand side being unity.

Finally, as $\sin^2(\pi(\zeta - n)) = \sin^2(\pi\zeta)$, we have

$$\operatorname{cosec}^2(\pi\zeta) \equiv \frac{1}{\sin^2(\pi\zeta)} = \sum_{n=-\infty}^{\infty} \frac{1}{\pi^2(\zeta - n)^2}. \tag{2.43}$$

The end result is a quite non-trivial expansion for the square of the cosecant.

2.2.4 Orthogonal polynomials

A useful class of orthonormal functions are the sets of *orthogonal polynomials* associated with an interval $[a, b]$ and a positive weight function $w(x)$ such that $\int_a^b w(x)\,dx$ is finite. We introduce the Hilbert space $L^2_w[a, b]$ with the real inner product

$$\langle u, v\rangle_w = \int_a^b w(x)u(x)v(x)\,dx, \tag{2.44}$$

and apply the *Gram-Schmidt procedure* to the monomial powers $1, x, x^2, x^3, \ldots$ so as to produce an orthonomal set. We begin with

$$P_0(x) \equiv 1/\|1\|_w, \tag{2.45}$$

where $\|1\|_w = \sqrt{\int_a^b w(x)\,dx}$, and define recursively

$$P_{n+1}(x) = \frac{xP_n(x) - \sum_0^n P_i(x)\langle P_i, xP_n\rangle_w}{\|xP_n - \sum_0^n P_i\langle P_i, xP_n\rangle\|_w}. \tag{2.46}$$

Clearly $P_n(x)$ is an n-th order polynomial, and by construction

$$\langle P_n, P_m\rangle_w = \delta_{nm}. \tag{2.47}$$

All such sets of polynomials obey a three-term recurrence relation

$$xP_n(x) = b_n P_{n+1}(x) + a_n P_n(x) + b_{n-1}P_{n-1}(x). \tag{2.48}$$

That there are only three terms, and that the coefficients of P_{n+1} and P_{n-1} are related, is due to the identity

$$\langle P_n, xP_m\rangle_w = \langle xP_n, P_m\rangle_w. \tag{2.49}$$

This means that the matrix (in the P_n basis) representing the operation of multiplication by x is symmetric. Since multiplication by x takes us from P_n only to P_{n+1}, the matrix has just one non-zero entry above the main diagonal, and hence, by symmetry, only one below.

The completeness of a family of polynomials orthogonal on a finite interval is guaranteed by the *Weierstrass approximation theorem* which asserts that for any continuous real function $f(x)$ on $[a, b]$, and for any $\varepsilon > 0$, there exists a polynomial $p(x)$ such that $|f(x) - p(x)| < \varepsilon$ for all $x \in [a, b]$. This means that polynomials are dense in the space of continuous functions equipped with the $\|\ldots\|_\infty$ norm. Because $|f(x) - p(x)| < \varepsilon$ implies that

$$\int_a^b |f(x) - p(x)|^2 w(x)\, dx \leq \varepsilon^2 \int_a^b w(x)\, dx, \tag{2.50}$$

they are also a dense subset of the continuous functions in the sense of $L^2_w[a, b]$ convergence. Because the Hilbert space $L^2_w[a, b]$ is defined to be the completion of the space of continuous functions, the continuous functions are automatically dense in $L^2_w[a, b]$. Now the triangle inequality tells us that a dense subset of a dense set is dense in the larger set, so the polynomials are dense in $L^2_w[a, b]$ itself. The normalized orthogonal polynomials therefore constitute a complete orthonormal set.

For later use, we here summarize the properties of the families of polynomials named after Legendre, Hermite and Tchebychef.

Legendre polynomials

Legendre polynomials have $a = -1$, $b = 1$ and $w = 1$. The standard Legendre polynomials are not normalized by the scalar product, but instead by setting $P_n(1) = 1$. They are given by *Rodriguez'* formula

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n. \tag{2.51}$$

The first few are

$$\begin{aligned}
P_0(x) &= 1, \\
P_1(x) &= x, \\
P_2(x) &= \frac{1}{2}(3x^2 - 1),
\end{aligned}$$

$$\begin{aligned} P_3(x) &= \frac{1}{2}(5x^3 - 3x), \\ P_4(x) &= \frac{1}{8}(35x^4 - 30x^2 + 3). \end{aligned}$$

Their inner product is

$$\int_{-1}^{1} P_n(x)P_m(x)\,dx = \frac{2}{2n+1}\delta_{nm}. \tag{2.52}$$

The three-term recurrence relation is

$$(2n+1)xP_n(x) = (n+1)P_{n+1}(x) + nP_{n-1}(x). \tag{2.53}$$

The P_n form a complete set for expanding functions on $[-1, 1]$.

Hermite polynomials

The Hermite polynomials have $a = -\infty$, $b = +\infty$ and $w(x) = e^{-x^2}$, and are defined by the generating function

$$e^{2tx-t^2} = \sum_{n=0}^{\infty} \frac{1}{n!} H_n(x) t^n. \tag{2.54}$$

If we write

$$e^{2tx-t^2} = e^{x^2-(x-t)^2}, \tag{2.55}$$

we may use Taylor's theorem to find

$$H_n(x) = \frac{d^n}{dt^n} e^{x^2-(x-t)^2}\bigg|_{t=0} = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2}, \tag{2.56}$$

which is a a useful alternative definition. The first few Hermite polynomials are

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= 2x, \\ H_2(x) &= 4x^2 - 2 \\ H_3(x) &= 8x^3 - 12x, \\ H_4(x) &= 16x^4 - 48x^2 + 12, \end{aligned}$$

The normalization is such that

$$\int_{-\infty}^{\infty} H_n(x)H_m(x)e^{-x^2}\,dx = 2^n n!\sqrt{\pi}\delta_{nm}, \tag{2.57}$$

as may be proved by using the generating function. The three-term recurrence relation is

$$2xH_n(x) = H_{n+1}(x) + 2nH_{n-1}(x). \tag{2.58}$$

Exercise 2.3: Evaluate the integral

$$F(s,t) = \int_{-\infty}^{\infty} e^{-x^2}e^{2sx-s^2}e^{2tx-t^2}\,dx$$

and expand the result as a double power series in s and t. By examining the coefficient of $s^n t^m$, show that

$$\int_{-\infty}^{\infty} H_n(x)H_m(x)e^{-x^2}\,dx = 2^n n!\sqrt{\pi}\delta_{nm}.$$

Problem 2.4: Let

$$\varphi_n(x) = \frac{1}{\sqrt{2^n n!\sqrt{\pi}}}H_n(x)e^{-x^2/2}$$

be the *normalized Hermite functions*. They form a complete orthonormal set in $L^2(\mathbb{R})$. Show that

$$\sum_{n=0}^{\infty} t^n\varphi_n(x)\varphi_n(y) = \frac{1}{\sqrt{\pi(1-t^2)}}\exp\left\{\frac{4xyt-(x^2+y^2)(1+t^2)}{2(1-t^2)}\right\}, \quad 0\le t<1.$$

This is *Mehler's formula*. (Hint: Expand of the right hand side as $\sum_{n=0}^{\infty} a_n(x,t)\varphi_n(y)$. To find $a_n(x,t)$, multiply by $e^{2sy-s^2-y^2/2}$ and integrate over y.)

Exercise 2.5: Let $\varphi_n(x)$ be the same functions as in the preceding problem. Define a Fourier-transform operator $F: L^2(\mathbb{R})\to L^2(\mathbb{R})$ by

$$F(f) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{ixs}f(s)\,ds.$$

With this normalization of the Fourier transform, F^4 is the identity map. The possible eigenvalues of F are therefore ± 1, $\pm i$. Starting from (2.56), show that the $\varphi_n(x)$ are eigenfunctions of F, and that

$$F(\varphi_n) = i^n\varphi_n(x).$$

Tchebychef polynomials

Tchebychef polynomials are defined by taking $a = -1$, $b = +1$ and $w(x) = (1-x^2)^{\pm 1/2}$. The *Tchebychef polynomials of the first kind* are

$$T_n(x) = \cos(n\cos^{-1} x). \tag{2.59}$$

The first few are

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= x, \\ T_2(x) &= 2x^2 - 1, \\ T_3(x) &= 4x^3 - 3x. \end{aligned}$$

The *Tchebychef polynomials of the second kind* are

$$U_{n-1}(x) = \frac{\sin(n\cos^{-1} x)}{\sin(\cos^{-1} x)} = \frac{1}{n}T_n'(x). \tag{2.60}$$

and the first few are

$$\begin{aligned} U_{-1}(x) &= 0, \\ U_0(x) &= 1, \\ U_1(x) &= 2x, \\ U_2(x) &= 4x^2 - 1, \\ U_3(x) &= 8x^3 - 4x. \end{aligned}$$

T_n and U_n obey the same recurrence relation

$$\begin{aligned} 2xT_n &= T_{n+1} + T_{n-1}, \\ 2xU_n &= U_{n+1} + U_{n-1}, \end{aligned}$$

which are disguised forms of elementary trigonometric identities. The orthogonality is also a disguised form of the orthogonality of the functions $\cos n\theta$ and $\sin n\theta$. After setting $x = \cos\theta$ we have

$$\int_0^\pi \cos n\theta \cos m\theta \, d\theta = \int_{-1}^1 \frac{1}{\sqrt{1-x^2}} T_n(x) T_m(x)\, dx = h_n \delta_{nm}, \quad n, m, \geq 0, \tag{2.61}$$

where $h_0 = \pi$, $h_n = \pi/2$, $n > 0$, and

$$\int_0^\pi \sin n\theta \sin m\theta \, d\theta = \int_{-1}^1 \sqrt{1-x^2} U_{n-1}(x) U_{m-1}(x) \, dx = \frac{\pi}{2} \delta_{nm}, \quad n, m > 0. \tag{2.62}$$

The set $\{T_n(x)\}$ is therefore orthogonal and complete in $L^2_{(1-x^2)^{-1/2}}[-1,1]$, and the set $\{U_n(x)\}$ is orthogonal and complete in $L^2_{(1-x^2)^{1/2}}[-1,1]$. Any function continuous on the closed interval $[-1,1]$ lies in both of these spaces, and can therefore be expanded in terms of either set.

2.3 Linear operators and distributions

Our theme is the analogy between linear differential operators and matrices. It is therefore useful to understand how we can think of a differential operator as a continuously indexed "matrix."

2.3.1 Linear operators

The action of a matrix on a vector $\mathbf{y} = \mathbf{A}\mathbf{x}$ is given in components by

$$y_i = A_{ij} x_j. \tag{2.63}$$

The function-space analogue of this, $g = Af$, is naturally to be thought of as

$$g(x) = \int_a^b A(x,y) f(y) \, dy, \tag{2.64}$$

where the summation over adjacent indices has been replaced by an integration over the dummy variable y. If $A(x,y)$ is an ordinary function then $A(x,y)$ is called an *integral kernel*. We will study such linear operators in the chapter on integral equations.

The identity operation is

$$f(x) = \int_a^b \delta(x-y) f(y) \, dy, \tag{2.65}$$

and so the Dirac delta function, which is **not** an ordinary function, plays the role of the identity matrix. Once we admit *distributions* such as $\delta(x)$, we can

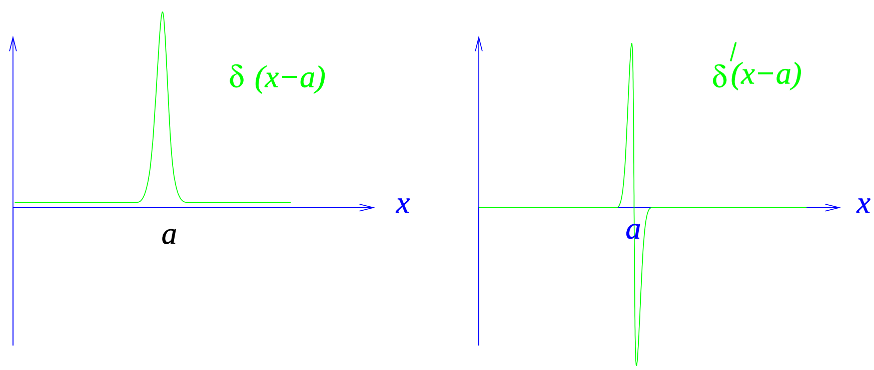

Figure 2.3: *Smooth approximations to* $\delta(x-a)$ *and* $\delta'(x-a)$.

think of differential operators as continuously indexed matrices by using the distribution

$$\delta'(x) = \text{“}\frac{d}{dx}\delta(x)\text{”}. \tag{2.66}$$

The quotes are to warn us that we are not really taking the derivative of the highly singular delta function. The symbol $\delta'(x)$ is properly defined by its behaviour in an integral

$$\begin{aligned}
\int_a^b \delta'(x-y)f(y)\,dy &= \int_a^b \frac{d}{dx}\delta(x-y)f(y)\,dy \\
&= -\int_a^b f(y)\frac{d}{dy}\delta(x-y)\,dy \\
&= \int_a^b f'(y)\delta(x-y)\,dy, \quad \text{(Integration by parts)} \\
&= f'(x).
\end{aligned}$$

The manipulations here are purely formal, and serve only to motivate the defining property

$$\int_a^b \delta'(x-y)f(y)\,dy = f'(x). \tag{2.67}$$

It is, however, sometimes useful to think of a smooth approximation to $\delta'(x-a)$ being the genuine derivative of a smooth approximation to $\delta(x-a)$, as illustrated in figure 2.3.

We can now define higher "derivatives" of $\delta(x)$ by

$$\int_a^b \delta^{(n)}(x) f(x) dx = (-1)^n f^{(n)}(0), \tag{2.68}$$

and use them to represent any linear differential operator as a formal integral kernel.

Example: In chapter one we formally evaluated a functional second derivative and ended up with the distributional kernel (1.186), which we here write as

$$\begin{aligned} k(x,y) &= -\frac{d}{dy}\left(p(y)\frac{d}{dy}\delta(y-x)\right) + q(y)\delta(y-x) \\ &= -p(y)\delta''(y-x) - p'(y)\delta'(y-x) + q(y)\delta(y-x). \end{aligned} \tag{2.69}$$

When k acts on a function u, it gives

$$\begin{aligned} \int k(x,y)u(y)\,dy &= \int \{-p(y)\delta''(y-x) - p'(y)\delta'(y-x) + q(y)\delta(y-x)\}\,u(y)\,dy \\ &= \int \delta(y-x)\,\{-[p(y)u(y)]'' + [p'(y)u(y)]' + q(y)u(y)\}\,dy \\ &= \int \delta(y-x)\,\{-p(y)u''(y) - p'(y)u'(y) + q(y)u(y)\}\,dy \\ &= -\frac{d}{dx}\left(p(x)\frac{du}{dx}\right) + q(x)u(x). \end{aligned} \tag{2.70}$$

The continuous matrix (1.186) therefore does, as indicated in chapter one, represent the Sturm-Liouville operator L defined in (1.182).

Exercise 2.6: Consider the distributional kernel

$$k(x,y) = a_2(y)\delta''(x-y) + a_1(y)\delta'(x-y) + a_0(y)\delta(x-y).$$

Show that

$$\int k(x,y)u(y)\,dy = (a_2(x)u(x))'' + (a_1(x)u(x))' + a_0(x)u(x).$$

Similarly show that

$$k(x,y) = a_2(x)\delta''(x-y) + a_1(x)\delta'(x-y) + a_0(x)\delta(x-y),$$

leads to

$$\int k(x,y)u(y)\,dy = a_2(x)u''(x) + a_1(x)u'(x) + a_0(x)u(x).$$

Exercise 2.7: The distributional kernel (2.69) was originally obtained as a functional second derivative

$$\begin{aligned} k(x_1, x_2) &= \frac{\delta}{\delta y(x_1)}\left(\frac{\delta J[y]}{\delta y(x_2)}\right) \\ &= -\frac{d}{dx_2}\left(p(x_2)\frac{d}{dx_2}\delta(x_2 - x_1)\right) + q(x_2)\delta(x_2 - x_1). \end{aligned}$$

By analogy with conventional partial derivatives, we would expect that

$$\frac{\delta}{\delta y(x_1)}\left(\frac{\delta J[y]}{\delta y(x_2)}\right) = \frac{\delta}{\delta y(x_2)}\left(\frac{\delta J[y]}{\delta y(x_1)}\right),$$

but x_1 and x_2 appear asymmetrically in $k(x_1, x_2)$. Define

$$k^T(x_1, x_2) = k(x_2, x_1),$$

and show that

$$\int k^T(x_1, x_2)u(x_2)\, dx_2 = \int k(x_1, x_2)u(x_2)\, dx_2.$$

Conclude that, superficial appearance notwithstanding, we do have $k(x_1, x_2) = k(x_2, x_1)$.

The example and exercises show that linear differential operators correspond to continuously-infinite matrices having entries only infinitesimally close to their main diagonal.

2.3.2 Distributions and test-functions

It is possible to work most the problems in this book with no deeper understanding of what a delta-function is than that presented in section 2.3.1. At some point however, the more careful reader will wonder about the logical structure of what we are doing, and will soon discover that too free a use of $\delta(x)$ and its derivatives can lead to paradoxes. How do such creatures fit into the function-space picture, and what sort of manipulations with them are valid?

We often think of $\delta(x)$ as being a "limit" of a sequence of functions whose graphs are getting narrower and narrower while their height grows to keep the area under the curve fixed. An example would be the spike function $\delta_\varepsilon(x - a)$ appearing in figure 2.4.

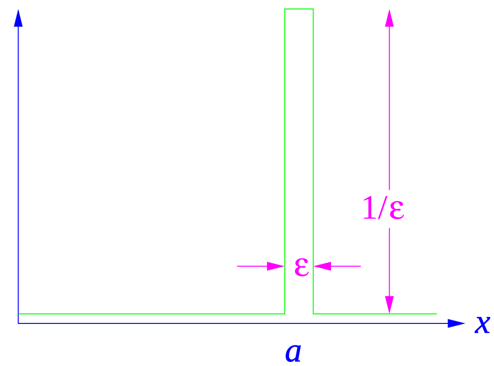

Figure 2.4: *Approximation* $\delta_\varepsilon(x-a)$ *to* $\delta(x-a)$.

The L^2 norm of δ_ε,

$$\|\delta_\varepsilon\|^2 = \int |\delta_\varepsilon(x)|^2\,dx = \frac{1}{\varepsilon}, \tag{2.71}$$

tends to infinity as $\varepsilon \to 0$, so δ_ε cannot be tending to any function in L^2. This delta function has infinite "length," and so is *not* an element of our Hilbert space.

The simple spike is not the only way to construct a delta function. In Fourier theory we meet

$$\delta_\Lambda(x) = \int_{-\Lambda}^{\Lambda} e^{ikx}\,\frac{dk}{2\pi} = \frac{1}{\pi}\frac{\sin\Lambda x}{x}, \tag{2.72}$$

which becomes a delta-function when Λ becomes large. In this case

$$\|\delta_\Lambda\|^2 = \int_{-\infty}^{\infty} \frac{\sin^2\Lambda x}{\pi^2 x^2}\,dx = \Lambda/\pi. \tag{2.73}$$

Again the "limit" has infinite length and cannot be accommodated in Hilbert space. This $\delta_\Lambda(x)$ is even more pathological than δ_ε. It provides a salutary counter-example to the often asserted "fact" that $\delta(x) = 0$ for $x \neq 0$. As Λ becomes large $\delta_\Lambda(0)$ diverges to infinity. At any fixed non-zero x, however, $\delta_\Lambda(x)$ oscillates between $\pm 1/x$ as Λ grows. Consequently the limit $\lim_{\Lambda\to\infty}\delta_\Lambda(x)$ exists nowhere. It therefore makes no sense to assign a numerical value to $\delta(x)$ at any x.

Given its wild behaviour, is not surprising that mathematicians looked askance at Dirac's $\delta(x)$. It was only in 1944, long after its effectiveness in solving physics and engineering problems had become an embarrassment, that Laurent Schwartz was able to tame $\delta(x)$ by creating his *theory of distributions.* Using the language of distributions we can state precisely the conditions under which a manoeuvre involving singular objects such as $\delta'(x)$ is legitimate.

Schwartz' theory is built on a concept from linear algebra. Recall that the *dual space* V^* of a vector space V is the vector space of linear functions from the original vector space V to the field over which it is defined. We consider $\delta(x)$ to be an element of the dual space of a vector space $\mathcal{T}$ of *test functions*. When a test function $\varphi(x)$ is plugged in, the δ-machine returns the number $\varphi(0)$. This operation is a linear map because the action of δ on $\lambda\varphi(x)+\mu\chi(x)$ is to return $\lambda\varphi(0)+\mu\chi(0)$. Test functions are smooth (infinitely differentiable) functions that tend rapidly to zero at infinity. Exactly what class of function we chose for $\mathcal{T}$ depends on the problem at hand. If we are going to make extensive use of Fourier transforms, for example, we mght select the *Schwartz space*, $\mathcal{S}(\mathbb{R})$. This is the space of infinitely differentiable functions $\varphi(x)$ such that the *seminorms*[3]

$$|\varphi|_{m,n} = \sup_{x\in\mathbb{R}} \left\{ |x|^n \left| \frac{d^m\varphi}{dx^m} \right| \right\} \tag{2.74}$$

are finite for all positive integers m and n. The Schwartz space has the advantage that if φ is in $\mathcal{S}(\mathbb{R})$, then so is its Fourier transform. Another popular space of test functions is $\mathcal{D}$ consisting of C^∞ functions of *compact support*—meaning that each function is identically zero outside some finite interval. Only if we want to prove theorems is a precise specification of $\mathcal{T}$ essential. For most physics calculations infinite differentiability and a rapid enough decrease at infinity for us to be able to ignore boundary terms is all that we need.

The "nice" behaviour of the test functions compensates for the "nasty" behaviour of $\delta(x)$ and its relatives. The objects, such as $\delta(x)$, composing the dual space of $\mathcal{T}$ are called *generalized functions*, or *distributions*. Actually, not every linear map $\mathcal{T} \to \mathbb{R}$ is to be included in the dual space because, for technical reasons, we must require the maps to be *continuous*. In other words, if $\varphi_n \to \varphi$, we want our distributions u to obey $u(\varphi_n) \to u(\varphi)$. Making precise what we mean by $\varphi_n \to \varphi$ is part of the task of specifying $\mathcal{T}$. In the

[3]A seminorm $|\cdots|$ has all the properties of a norm except that $|\varphi| = 0$ does not imply that $\varphi = 0$.

Schwartz space, for example, we declare that $\varphi_n \to \varphi$ if $|\varphi_n - \varphi|_{n,m} \to 0$, for all positive m, n. When we restrict a dual space to continuous functionals, we usually denote it by V' rather than V^*. The space of distributions is therefore $\mathcal{T}'$.

When they wish to stress the dual-space aspect of distribution theory, mathematically-minded authors use the notation

$$\delta(\varphi) = \varphi(0), \tag{2.75}$$

or

$$(\delta, \varphi) = \varphi(0), \tag{2.76}$$

in place of the common, but purely formal,

$$\int \delta(x)\varphi(x)\,dx = \varphi(0). \tag{2.77}$$

The expression (δ, φ) here represents the *pairing* of the element φ of the vector space $\mathcal{T}$ with the element δ of its dual space $\mathcal{T}'$. It should not be thought of as an inner product as the distribution and the test function lie in different spaces. The "integral" in the common notation is purely symbolic, of course, but the common notation should not be despised even by those in quest of rigour. It suggests correct results, such as

$$\int \delta(ax - b)\varphi(x)\,dx = \frac{1}{|a|}\varphi(b/a), \tag{2.78}$$

which would look quite unmotivated in the dual-space notation.

The distribution $\delta'(x)$ is now defined by the pairing

$$(\delta', \varphi) = -\varphi'(0), \tag{2.79}$$

where the minus sign comes from imagining an integration by parts that takes the "derivative" off $\delta(x)$ and puts it on to the smooth function $\varphi(x)$:

$$\text{``}\int \delta'(x)\varphi(x)\,dx\text{''} = -\int \delta(x)\varphi'(x)\,dx. \tag{2.80}$$

Similarly $\delta^{(n)}(x)$ is now defined by the pairing

$$(\delta^{(n)}, \varphi) = (-1)^n \varphi^{(n)}(0). \tag{2.81}$$

The "nicer" the class of test function we take, the "nastier" the class of distributions we can handle. For example, the Hilbert space L^2 is its own dual: the *Riesz-Fréchet theorem* (see exercise 2.10) asserts that any continuous linear map $F : L^2 \to \mathbb{R}$ can be written as $F[f] = \langle l, f \rangle$ for some $l \in L^2$. The delta-function map is not continuous when considered as a map from $L^2 \to \mathbb{R}$ however. An arbitrarily small change, $f \to f + \delta f$, in a function (small in the L^2 sense of $\|\delta f\|$ being small) can produce an arbitrarily large change in $f(0)$. Thus L^2 functions are not "nice" enough for their dual space to be able accommodate the delta function. Another way of understanding this is to remember that we regard two L^2 functions as being the same whenever $\|f_1 - f_2\| = 0$. This distance will be zero even if f_1 and f_2 differ from one another on a countable set of points. As we have remarked earlier, this means that elements of L^2 are not really functions at all — they do not have an assigned valued at each point. They are, instead, only *equivalence classes* of functions. Since $f(0)$ is undefined, any attempt to interpret the statement $\int \delta(x) f(x)\, dx = f(0)$ for f an arbitrary element L^2 is necessarily doomed to failure. Continuous functions, however, do have well-defined values at every point. If we take the space of test of functions $\mathcal{T}$ to consist of all continuous functions, but not demand that they be differentiable, then $\mathcal{T}'$ will include the delta function, but not its "derivative" $\delta'(x)$, as this requires us to evaluate $f'(0)$. If we require the test functions to be once-differentiable, then $\mathcal{T}'$ will include $\delta'(x)$ but not $\delta''(x)$, and so on.

When we add suitable spaces $\mathcal{T}$ and $\mathcal{T}'$ to our toolkit, we are constructing what is called a *rigged*[4] Hilbert space. In such a rigged space we have the inclusion

$$\mathcal{T} \subset L^2 \equiv [L^2]' \subset \mathcal{T}'. \tag{2.82}$$

The idea is to take the space $\mathcal{T}'$ big enough to contain objects such as the limit of our sequence of "approximate" delta functions δ_ε, which does not converge to anything in L^2.

Ordinary functions can also be regarded as distributions, and this helps illuminate the different senses in which a sequence u_n can converge. For example, we can consider the functions

$$u_n = \sin n\pi x, \quad 0 < x < 1, \tag{2.83}$$

[4] "Rigged" as in a sailing ship ready for sea, not "rigged" as in a corrupt election.

as being either elements of $L^2[0,1]$ or as distributions. As distributions we evaluate them on a smooth function φ as

$$(u_n, \varphi) = \int_0^1 \varphi(x) u_n(x)\, dx. \tag{2.84}$$

Now

$$\lim_{n\to\infty} (u_n, \varphi) = 0, \tag{2.85}$$

since the high-frequency Fourier coefficients of any smooth function tend to zero. We deduce that *as a distribution* we have $\lim_{n\to\infty} u_n = 0$, the convergence being pointwise on the space of test functions. Considered as elements of $L^2[0,1]$, however, the u_n do not tend to zero. Their norm is $\|u_n\| = 1/2$ and so all the u_n remain at the same fixed distance from 0.

Exercise 2.8: Here we show that the elements of $L^2[a,b]$, which we defined in exercise 2.2 to be the formal limits of of Cauchy sequences of continuous functions, may be thought of as distributions.

i) Let $\varphi(x)$ be a test function and $f_n(x)$ a Cauchy sequence of continuous functions defining $f \in L^2$. Use the Cauchy-Schwarz-Bunyakovsky inequality to show that the sequence of numbers $\langle \varphi, f_n \rangle$ is Cauchy and so deduce that $\lim_{n\to\infty} \langle \varphi, f_n \rangle$ exists.

ii) Let $\varphi(x)$ be a test function and $f_n^{(1)}(x)$ and $f_n^{(2)}(x)$ be a pair of equivalent sequences defining the same element $f \in L^2$. Use Cauchy-Schwarz-Bunyakovsky to show that

$$\lim_{n\to\infty} \langle \varphi, f_n^{(1)} - f_n^{(2)} \rangle = 0.$$

Combine this result with that of the preceding exercise to deduce that we can set

$$(\varphi, f) \equiv \lim_{n\to\infty} \langle \varphi^*, f_n \rangle,$$

and so define $f \equiv \lim_{n\to\infty} f_n$ as a distribution.

The interpretation of elements of L^2 as distributions is simultaneously simpler and more physical than the classical interpretation *via* the Lebesgue integral.

Weak derivatives

By exploiting the infinite differentiability of our test functions, we were able to make mathematical sense of the "derivative" of the highly singular delta function. The same idea of a formal integration by parts can be used to define the "derivative" for any distribution, and also for ordinary functions that would not usually be regarded as being differentiable.

We therefore define the *weak* or *distributional* derivative $v(x)$ of a distribution $u(x)$ by requiring its evaluation on a test function $\varphi \in \mathcal{T}$ to be

$$\int v(x)\varphi(x)\,dx \stackrel{\text{def}}{=} -\int u(x)\varphi'(x)\,dx. \tag{2.86}$$

In the more formal pairing notation we write

$$(v,\varphi) \stackrel{\text{def}}{=} -(u,\varphi'). \tag{2.87}$$

The right hand side of (2.87) is a continuous linear function of φ, and so, therefore, is the left hand side. Thus the weak derivative $u' \equiv v$ is a well-defined distribution for any u.

When $u(x)$ is an ordinary function that is differentiable in the conventional sense, its weak derivative coincides with the usual derivative. When the function is not conventionally differentiable the weak derivative still exists, but does not assign a numerical value to the derivative at each point. It is therefore a distribution and not a function.

The elements of L^2 are not quite functions — having no well-defined value at a point — but are particularly mild-mannered distributions, and their weak derivatives may themselves be elements of L^2. It is in this weak sense that we will, in later chapters, allow differential operators to act on L^2 "functions."

Example: In the weak sense

$$\frac{d}{dx}|x| = \text{sgn}(x), \tag{2.88}$$

$$\frac{d}{dx}\text{sgn}(x) = 2\delta(x). \tag{2.89}$$

The object $|x|$ is an ordinary function, but $\text{sgn}(x)$ has no definite value at $x=0$, whilst $\delta(x)$ has no definite value at any x.

Example: As a more subtle illustration, consider the weak derivative of the function $\ln|x|$. With $\varphi(x)$ a test function, the improper integral

$$I = -\int_{-\infty}^{\infty} \varphi'(x)\,\ln|x|\,dx \equiv -\lim_{\varepsilon,\varepsilon'\to 0}\left(\int_{-\infty}^{-\varepsilon} + \int_{\varepsilon'}^{\infty}\right)\varphi'(x)\,\ln|x|\,dx \tag{2.90}$$

is convergent and defines the pairing $(-\ln|x|, \varphi')$. We wish to integrate by parts and interpret the result as $([\ln|x|]', \varphi)$. The logarithm is differentiable in the conventional sense away from $x = 0$, and

$$[\ln|x|\varphi(x)]' = \frac{1}{x}\varphi(x) + \ln|x|\varphi'(x), \quad x \neq 0. \tag{2.91}$$

From this we find that

$$-(\ln|x|, \varphi') = \lim_{\varepsilon,\varepsilon'\to 0}\left\{\left(\int_{-\infty}^{-\varepsilon} + \int_{\varepsilon'}^{\infty}\right)\frac{1}{x}\varphi(x)\,dx + \Big(\varphi(\varepsilon')\ln|\varepsilon'| - \varphi(-\varepsilon)\ln|\varepsilon|\Big)\right\}. \tag{2.92}$$

So far ε and ε' are unrelated except in that they are both being sent to zero. If, however, we choose to make them equal, $\varepsilon = \varepsilon'$, then the integrated-out part becomes

$$\Big(\varphi(\varepsilon) - \varphi(-\varepsilon)\Big)\ln|\varepsilon| \sim 2\varphi'(0)\varepsilon\ln|\varepsilon|, \tag{2.93}$$

and this tends to zero as ε becomes small. In this case

$$-([\ln|x|], \varphi') = \lim_{\varepsilon\to 0}\left\{\left(\int_{-\infty}^{-\varepsilon} + \int_{\varepsilon}^{\infty}\right)\frac{1}{x}\varphi(x)\,dx\right\}. \tag{2.94}$$

By the definition of the weak derivative, the left hand side of (2.94) is the pairing $([\ln|x|]', \varphi)$. We conclude that

$$\frac{d}{dx}\ln|x| = P\left(\frac{1}{x}\right), \tag{2.95}$$

where $P(1/x)$, the *principal-part* distribution, is defined by the right-hand-side of (2.94). It is evaluated on the test function $\varphi(x)$ by forming $\int \varphi(x)/x\,dx$, but with an infinitesimal interval from $-\varepsilon$ to $+\varepsilon$, omitted from the range of integration. It is essential that this omitted interval lie symmetrically about the dangerous point $x = 0$. Otherwise the integrated-out part will not vanish in the $\varepsilon \to 0$ limit. The resulting *principal-part integral*, written $P\int \varphi(x)/x\,dx$, is then convergent and $P(1/x)$ is a well-defined distribution despite the singularity in the integrand. Principal-part integrals are common in physics. We will next meet them when we study Green functions.

For further reading on distributions and their applications we recommend M. J. Lighthill *Fourier Analysis and Generalised Functions*, or F. G. Friedlander *Introduction to the Theory of Distributions*. Both books are published by Cambridge University Press.

2.4 Further exercises and problems

The first two exercises lead the reader through a proof of the Riesz-Fréchet theorem. Although not an essential part of our story, they demonstrate how "completeness" is used in Hilbert space theory, and provide some practice with "ϵ, δ" arguments for those who desire it.

Exercise 2.9: Show that if a norm $\| \;\; \|$ is derived from an inner product, then it obeys the *parallelogram law*

$$\|f+g\|^2 + \|f-g\|^2 = 2(\|f\|^2 + \|g\|^2).$$

Let N be a complete linear subspace of a Hilbert space H. Let $g \notin N$, and let

$$\inf_{f \in N} \|g - f\| = d.$$

Show that there exists a sequence $f_n \in N$ such that $\lim_{n\to\infty} \|f_n - g\| = d$. Use the parallelogram law to show that the sequence f_n is Cauchy, and hence deduce that there is a unique $f \in N$ such that $\|g - f\| = d$. From this, conclude that $d > 0$. Now show that $\langle (g-f), h\rangle = 0$ for all $h \in N$.

Exercise 2.10: *Riesz-Fréchet theorem.* Let $L[h]$ be a continuous linear functional on a Hilbert space H. Here *continuous* means that

$$\|h_n - h\| \to 0 \Rightarrow L[h_n] \to L[h].$$

Show that the set $N = \{f \in H : L[f] = 0\}$ is a complete linear subspace of H.

Suppose now that there is a $g \in H$ such that $L(g) \neq 0$, and let $l \in H$ be the vector "$g - f$" from the previous problem. Show that

$$L[h] = \langle \alpha l, h\rangle, \quad \text{where } \alpha = L[g]/\langle l, g\rangle = L[g]/\|l\|^2.$$

A continuous linear functional can therefore be expressed as an inner product.

Next we have some problems on orthogonal polynomials and three-term recurrence relations. They provide an excuse for reviewing linear algebra, and also serve to introduce the theory behind some practical numerical methods.

Exercise 2.11: Let $\{P_n(x)\}$ be a family of polynomials orthonormal on $[a, b]$ with respect to a a positive weight function $w(x)$, and with $\deg\,[P_n(x)] = n$. Let us also scale $w(x)$ so that $\int_a^b w(x)\,dx = 1$, and $P_0(x) = 1$.

a) Suppose that the $P_n(x)$ obey the three-term recurrence relation

$$xP_n(x) = b_n P_{n+1}(x) + a_n P_n(x) + b_{n-1} P_{n-1}(x); \quad P_{-1}(x) = 0,\ P_0(x) = 1.$$

Define

$$p_n(x) = P_n(x)(b_{n-1} b_{n-2} \cdots b_0),$$

and show that

$$xp_n(x) = p_{n+1}(x) + a_n p_n(x) + b_{n-1}^2 p_{n-1}(x); \quad p_{-1}(x) = 0,\ p_0(x) = 1.$$

Conclude that the $p_n(x)$ are *monic* — *i.e.* the coefficient of their leading power of x is unity.

b) Show also that the functions

$$q_n(x) = \int_a^b \frac{p_n(x) - p_n(\xi)}{x - \xi} w(\xi)\, d\xi$$

are degree $n-1$ monic polynomials that obey the same recurrence relation as the $p_n(x)$, but with initial conditions $q_0(x) = 0$, $q_1(x) \equiv \int_a^b w\, dx = 1$.

Warning: while the $q_n(x)$ polynomials defined in part b) turn out to be very useful, they are *not* mutually orthogonal with respect to $\langle\ ,\ \rangle_w$.

Exercise 2.12: *Gaussian quadrature*. Orthogonal polynomials have application to numerical integration. Let the polynomials $\{P_n(x)\}$ be orthonormal on $[a, b]$ with respect to the positive weight function $w(x)$, and let x_ν, $\nu = 1, \ldots, N$ be the zeros of $P_N(x)$. You will show that if we define the weights

$$w_\nu = \int_a^b \frac{P_N(x)}{P_N'(x_\nu)(x - x_\nu)} w(x)\, dx$$

then the approximate integration scheme

$$\int_a^b f(x) w(x)\, dx \approx w_1 f(x_1) + w_2 f(x_2) + \cdots w_N f(x_N),$$

known as *Gauss' quadrature rule*, is *exact* for $f(x)$ any polynomial of degree less than or equal to $2N - 1$.

a) Let $\pi(x) = (x - \xi_1)(x - \xi_2) \cdots (x - \xi_N)$ be a polynomial of degree N. Given a function $F(x)$, show that

$$F_L(x) \stackrel{\text{def}}{=} \sum_{\nu=1}^{N} F(\xi_\nu) \frac{\pi(x)}{\pi'(\xi_\nu)(x - \xi_\nu)}$$

is a polynomial of degree $N - 1$ that coincides with $F(x)$ at $x = \xi_\nu$, $\nu = 1, \ldots, N$. (This is *Lagrange's interpolation formula*.)

b) Show that if $F(x)$ is polynomial of degree $N-1$ or less then $F_L(x) = F(x)$.
c) Let $f(x)$ be a polynomial of degree $2N-1$ or less. Cite the polynomial division algorithm to show that there exist polynomials $Q(x)$ and $R(x)$, each of degree $N-1$ or less, such that

$$f(x) = P_N(x)Q(x) + R(x).$$

d) Show that $f(x_\nu) = R(x_\nu)$, and that

$$\int_a^b f(x)w(x)\,dx = \int_a^b R(x)w(x)\,dx.$$

e) Combine parts a), b) and d) to establish Gauss' result.
f) Show that if we normalize $w(x)$ so that $\int w\,dx = 1$ then the weights w_ν can be expressed as $w_\nu = q_N(x_\nu)/p'_N(x_\nu)$, where $p_n(x)$, $q_n(x)$ are the monic polynomials defined in the preceding problem.

The ultimate large-N exactness of Gaussian quadrature can be expressed as

$$w(x) = \lim_{N\to\infty}\left\{\sum_\nu \delta(x-x_\nu)w_\nu\right\}.$$

Of course, a sum of Dirac delta-functions can never become a continuous function in any ordinary sense. The equality holds only after both sides are integrated against a smooth test function, *i.e.*, when it is considered as a statement about distributions.

Exercise 2.13: The completeness of a set of polynomials $\{P_n(x)\}$, orthonormal with respect to the positive weight function $w(x)$, is equivalent to the statement that

$$\sum_{n=0}^{\infty} P_n(x)P_n(y) = \frac{1}{w(x)}\delta(x-y).$$

It is useful to have a formula for the partial sums of this infinite series.

Suppose that the polynomials $P_n(x)$ obey the three-term recurrence relation

$$xP_n(x) = b_nP_{n+1}(x) + a_nP_n(x) + b_{n-1}P_{n-1}(x); \quad P_{-1}(x) = 0,\ P_0(x) = 1.$$

Use this recurrence relation, together with its initial conditions, to obtain the *Christoffel-Darboux* formula

$$\sum_{n=0}^{N-1} P_n(x)P_n(y) = \frac{b_{N-1}[P_N(x)P_{N-1}(y) - P_{N-1}(x)P_N(y)]}{x-y}.$$

Exercise 2.14: Again suppose that the polynomials $P_n(x)$ obey the three-term recurrence relation

$$xP_n(x) = b_n P_{n+1}(x) + a_n P_n(x) + b_{n-1} P_{n-1}(x); \quad P_{-1}(x) = 0,\ P_0(x) = 1.$$

Consider the N-by-N tridiagonal matrix eigenvalue problem

$$\begin{bmatrix} a_{N-1} & b_{N-2} & 0 & 0 & \dots & 0 \\ b_{N-2} & a_{N-2} & b_{N-3} & 0 & \dots & 0 \\ 0 & b_{N-3} & a_{N-3} & b_{N-4} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & b_2 & a_2 & b_1 & 0 \\ 0 & \dots & 0 & b_1 & a_1 & b_0 \\ 0 & \dots & 0 & 0 & b_0 & a_0 \end{bmatrix} \begin{bmatrix} u_{N-1} \\ u_{N-2} \\ u_{N-3} \\ \vdots \\ u_2 \\ u_1 \\ u_0 \end{bmatrix} = x \begin{bmatrix} u_{N-1} \\ u_{N-2} \\ u_{N-3} \\ \vdots \\ u_2 \\ u_1 \\ u_0 \end{bmatrix}$$

a) Show that the eigenvalues x are given by the zeros x_ν, $\nu = 1, \dots, N$ of $P_N(x)$, and that the corresponding eigenvectors have components $u_n = P_n(x_\nu)$, $n = 0, \dots, N-1$.
b) Take the $x \to y$ limit of the Christoffel-Darboux formula from the preceding problem, and use it to show that the orthogonality and completeness relations for the eigenvectors can be written as

$$\begin{aligned} \sum_{n=0}^{N-1} P_n(x_\nu) P_n(x_\mu) &= w_\nu^{-1} \delta_{\nu\mu}, \\ \sum_{\nu=1}^{N} w_\nu P_n(x_\nu) P_m(x_\nu) &= \delta_{nm}, \quad n.m \le N-1, \end{aligned}$$

where $w_\nu^{-1} = b_{N-1} P'_N(x_\nu) P_{N-1}(x_\nu)$.
c) Use the original Christoffel-Darboux formula to show that, when the $P_n(x)$ are orthonormal with respect to the positive weight function $w(x)$, the normalization constants w_ν of this present problem coincide with the weights w_ν occurring in the Gauss quadrature rule. Conclude from this equality that the Gauss-quadrature weights are *positive*.

Exercise 2.15: Write the N-by-N tridiagonal matrix eigenvalue problem from the preceding exercise as $\mathbf{H}\mathbf{u} = x\mathbf{u}$, and set $d_N(x) = \det(x\mathbf{I} - \mathbf{H})$. Similarly define $d_n(x)$ to be the determinant of the n-by-n tridiagonal submatrix with $x - a_{n-1}, \dots, x - a_0$ along its principal diagonal. Laplace-develop the determinant $d_n(x)$ about its first row, and hence obtain the recurrence

$$d_{n+1}(x) = (x - a_n) d_n(x) - b_{n-1}^2 d_{n-1}(x).$$

Conclude that

$$\det(x\mathbf{I} - \mathbf{H}) = p_N(x),$$

where $p_n(x)$ is the monic orthogonal polynomial obeying

$$xp_n(x) = p_{n+1}(x) + a_n p_n(x) + b_{n-1}^2 p_{n-1}(x); \quad p_{-1}(x) = 0, \ p_0(x) = 1.$$

Exercise 2.16: Again write the N-by-N tridiagonal matrix eigenvalue problem from the preceding exercises as $\mathbf{Hu} = x\mathbf{u}$.

a) Show that the lowest and rightmost matrix element

$$\langle 0|(x\mathbf{I} - \mathbf{H})^{-1}|0\rangle \equiv (x\mathbf{I} - \mathbf{H})^{-1}_{00}$$

of the *resolvent matrix* $(x\mathbf{I} - \mathbf{H})^{-1}$ is given by a continued fraction $G_{N-1,0}(x)$ where, for example,

$$G_{3,z}(x) = \cfrac{1}{x - a_0 - \cfrac{b_0^2}{x - a_1 - \cfrac{b_1^2}{x - a_2 - \cfrac{b_2^2}{x - a_3 + z}}}}.$$

b) Use induction on n to show that

$$G_{n,z}(x) = \frac{q_n(x)z + q_{n+1}(x)}{p_n(x)z + p_{n+1}(x)},$$

where $p_n(x)$, $q_n(x)$ are the monic polynomial functions of x defined by the recurrence relations

$$\begin{aligned} xp_n(x) &= p_{n+1}(x) + a_n p_n(x) + b_{n-1}^2 p_{n-1}(x), \quad p_{-1}(x) = 0, \ p_0(x) = 1, \\ xq_n(x) &= q_{n+1}(x) + a_n q_n(x) + b_{n-1}^2 q_{n-1}(x), \quad q_0(x) = 0, \ q_1(x) = 1. \end{aligned}$$

b) Conclude that

$$\langle 0|(x\mathbf{I} - \mathbf{H})^{-1}|0\rangle = \frac{q_N(x)}{p_N(x)},$$

has a *pole* singularity when x approaches an eigenvalue x_ν. Show that the *residue* of the pole (the coefficient of $1/(x - x_n)$) is equal to the Gauss-quadrature weight w_ν for $w(x)$, the weight function (normalized so that $\int w\,dx = 1$) from which the coefficients a_n, b_n were derived.

Continued fractions were introduced by John Wallis in his *Arithmetica Infinitorum* (1656), as was the recursion formula for their evaluation. Today, when combined with the output of the next exercise, they provide the mathematical underpinning of the *Haydock recursion method* in the band theory of solids. Haydock's method computes $w(x) = \lim_{N\to\infty}\{\sum_\nu \delta(x-x_\nu)w_\nu\}$, and interprets it as the local density of states that is measured in scanning tunnelling microscopy.

Exercise 2.17: The Lanczos tridiagonalization algorithm. Let V be an N-dimensional complex vector space equipped with an inner product $\langle\ ,\ \rangle$ and let $H : V \to V$ be a hermitian linear operator. Starting from a unit vector $\mathbf{u}_0$, and taking $\mathbf{u}_{-1} = \mathbf{0}$, recursively generate the unit vectors $\mathbf{u}_n$ and the numbers a_n, b_n and c_n by

$$H\mathbf{u}_n = b_n\mathbf{u}_{n+1} + a_n\mathbf{u}_n + c_{n-1}\mathbf{u}_{n-1},$$

where the coefficients

$$\begin{aligned} a_n &\equiv \langle \mathbf{u}_n, H\mathbf{u}_n\rangle, \\ c_{n-1} &\equiv \langle \mathbf{u}_{n-1}, H\mathbf{u}_n\rangle, \end{aligned}$$

ensure that $\mathbf{u}_{n+1}$ is perpendicular to both $\mathbf{u}_n$ and $\mathbf{u}_{n-1}$, and

$$b_n = \|H\mathbf{u}_n - a_n\mathbf{u}_n - c_{n-1}\mathbf{u}_{n-1}\|,$$

a positive real number, makes $\|\mathbf{u}_{n+1}\| = 1$.

a) Use induction on n to show that $\mathbf{u}_{n+1}$, although only constructed to be perpendicular to the previous *two* vectors, is in fact (and in the absence of numerical rounding errors) perpendicular to *all* $\mathbf{u}_m$ with $m \le n$.
b) Show that a_n, c_n are *real*, and that $c_{n-1} = b_{n-1}$.
c) Conclude that $b_{N-1} = 0$, and (provided that no earlier b_n happens to vanish) that the $\mathbf{u}_n$, $n = 0, \ldots, N-1$, constitute an orthonormal basis for V, in terms of which H is represented by the N-by-N real-symmetric tridiagonal matrix $\mathbf{H}$ of the preceding exercises.

Because the eigenvalues of a tridiagonal matrix are given by the numerically easy-to-find zeros of the associated monic polynomial $p_N(x)$, the Lanczos algorithm provides a computationally efficient way of extracting the eigenvalues from a large sparse matrix. In theory, the entries in the tridiagonal $\mathbf{H}$ can be computed while retaining only $\mathbf{u}_n$, $\mathbf{u}_{n-1}$ and $H\mathbf{u}_n$ in memory at any one time. In practice, with finite precision computer arithmetic, orthogonality with the earlier $\mathbf{u}_m$ is eventually lost, and spurious or duplicated eigenvalues appear. There exist, however, stratagems for identifying and eliminating these fake eigenvalues.

The following two problems are "toy" versions of the *Lax pair* and *tau function* constructions that arise in the general theory of soliton equations. They provide useful practice in manipulating matrices and determinants.

Problem 2.18: The monic orthogonal polynomials $p_i(x)$ have inner products

$$\langle p_i, p_j\rangle_w \equiv \int p_i(x)p_j(x)w(x)\,dx = h_i\delta_{ij},$$

and obey the recursion relation

$$xp_i(x) = p_{i+1}(x) + a_i p_i(x) + b_{i-1}^2 p_{i-1}(x); \quad p_{-1}(x) = 0,\, p_0(x) = 1.$$

Write the recursion relation as

$$\mathbf{L}\mathbf{p} = x\mathbf{p},$$

where

$$\mathbf{L} \equiv \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \vdots \\ \dots & 1 & a_2 & b_1^2 & 0 \\ \dots & 0 & 1 & a_1 & b_0^2 \\ \dots & 0 & 0 & 1 & a_0 \end{bmatrix}, \quad \mathbf{p} \equiv \begin{bmatrix} \vdots \\ p_2 \\ p_1 \\ p_0 \end{bmatrix}.$$

Suppose that

$$w(x) = \exp\left\{-\sum_{n=1}^{\infty} t_n x^n\right\},$$

and consider how the $p_i(x)$ and the coefficients a_i and b_i^2 vary with the parameters t_n.

a) Show that

$$\frac{\partial \mathbf{p}}{\partial t_n} = \mathbf{M}^{(n)}\mathbf{p},$$

where $\mathbf{M}^{(n)}$ is some strictly upper triangular matrix - *i.e.* all entries on and below its principal diagonal are zero.

b) By differentiating $\mathbf{L}\mathbf{p} = x\mathbf{p}$ with respect to t_n show that

$$\frac{\partial \mathbf{L}}{\partial t_n} = [\mathbf{M}^{(n)}, \mathbf{L}].$$

c) Compute the matrix elements

$$\langle i|\mathbf{M}^{(n)}|j\rangle \equiv M_{ij}^{(n)} = \left\langle p_j, \frac{\partial p_i}{\partial t_n}\right\rangle_w$$

(note the interchange of the order of i and j in the $\langle\ ,\ \rangle_w$ product!) by differentiating the orthogonality condition $\langle p_i, p_j\rangle_w = h_i\delta_{ij}$. Hence show that

$$\mathbf{M}^{(n)} = (\mathbf{L}^n)_+$$

where $(\mathbf{L}^n)_+$ denotes the strictly upper triangular projection of the n'th power of $\mathbf{L}$ — *i.e.* the matrix $\mathbf{L}^n$, but with its diagonal and lower triangular entries replaced by zero.

Thus

$$\frac{\partial \mathbf{L}}{\partial t_n} = \left[(\mathbf{L}^n)_+, \mathbf{L}\right]$$

describes a family of deformations of the semi-infinite matrix $\mathbf{L}$ that, in some formal sense, preserve its eigenvalues x.

Problem 2.19: Let the monic polynomials $p_n(x)$ be orthogonal with respect to the weight function

$$w(x) = \exp\left\{-\sum_{n=1}^{\infty} t_n x^n\right\}.$$

Define the "tau-function" $\tau_n(t_1, t_2, t_3\ldots)$ of the parameters t_i to be the n-fold integral

$$\tau_n(t_1, t_2, \ldots) = \iint\cdots\int dx_x dx_2 \ldots dx_n \Delta^2(x) \exp\left\{-\sum_{\nu=1}^{n}\sum_{m=1}^{\infty} t_m x_\nu^m\right\}$$

where

$$\Delta(x) = \begin{vmatrix} x_1^{n-1} & x_1^{n-2} & \ldots & x_1 & 1 \\ x_2^{n-1} & x_2^{n-2} & \ldots & x_2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_n^{n-1} & x_n^{n-2} & \ldots & x_n & 1 \end{vmatrix} = \prod_{\nu<\mu}(x_\nu - x_\mu)$$

is the n-by-n Vandermonde determinant.

a) Show that

$$\begin{vmatrix} x_1^{n-1} & x_1^{n-2} & \ldots & x_1 & 1 \\ x_2^{n-1} & x_2^{n-2} & \ldots & x_2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_n^{n-1} & x_n^{n-2} & \ldots & x_n & 1 \end{vmatrix} = \begin{vmatrix} p_{n-1}(x_1) & p_{n-2}(x_1) & \ldots & p_1(x_1) & p_0(x_1) \\ p_{n-1}(x_2) & p_{n-2}(x_2) & \ldots & p_1(x_2) & p_0(x_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ P_{n-1}(x_n) & p_{n-2}(x_n) & \ldots & p_1(x_n) & p_0(x_n) \end{vmatrix}$$

b) Combine the identity from part a) with the orthogonality property of the $p_n(x)$ to show that

$$\begin{aligned} p_n(x) &= \frac{1}{\tau_n}\int dx_1 dx_2 \dots dx_n \Delta^2(x) \prod_{\mu=1}^{n}(x-x_\mu) \exp\left\{-\sum_{\nu=1}^{n}\sum_{m=1}^{\infty} t_m x_\nu^m\right\} \\ &= x^n \frac{\tau_n(t_1', t_2', t_3', \dots)}{\tau_n(t_1, t_2, t_3, \dots)} \end{aligned}$$

where

$$t_m' = t_m + \frac{1}{m x^m}.$$

Here are some exercises on distributions:

Exercise 2.20: Let $f(x)$ be a continuous function. Observe that $f(x)\delta(x) = f(0)\delta(x)$. Deduce that

$$\frac{d}{dx}[f(x)\delta(x)] = f(0)\delta'(x).$$

If $f(x)$ were differentiable we might also have used the product rule to conclude that

$$\frac{d}{dx}[f(x)\delta(x)] = f'(x)\delta(x) + f(x)\delta'(x).$$

Show, by evaluating $f(0)\delta'(x)$ and $f'(x)\delta(x) + f(x)\delta'(x)$ on a test function $\varphi(x)$, that these two expressions for the derivative of $f(x)\delta(x)$ are equivalent.

Exercise 2.21: Let $\varphi(x)$ be a test function. Show that

$$\frac{d}{dt}\left\{P\int_{-\infty}^{\infty}\frac{\varphi(x)}{(x-t)}\,dx\right\} = P\int_{-\infty}^{\infty}\frac{\varphi(x)-\varphi(t)}{(x-t)^2}\,dx.$$

Show further that the right-hand-side of this equation is equal to

$$-\left(\frac{d}{dx}P\left(\frac{1}{x-t}\right), \varphi\right) \equiv P\int_{-\infty}^{\infty}\frac{\varphi'(x)}{(x-t)}\,dx.$$

Exercise 2.22: Let $\theta(x)$ be the *step function* or *Heaviside distribution*

$$\theta(x) = \begin{cases} 1, & x > 0, \\ \text{undefined}, & x = 0, \\ 0, & x < 0. \end{cases}$$

By forming the weak derivative of both sides of the equation

$$\lim_{\varepsilon\to 0_+} \ln(x+i\varepsilon) = \ln|x| + i\pi\theta(-x),$$

conclude that

$$\lim_{\varepsilon\to 0_+} \left(\frac{1}{x+i\varepsilon}\right) = P\left(\frac{1}{x}\right) - i\pi\delta(x).$$

Exercise 2.23: Use induction on n to generalize exercise 2.21 and show that

$$\begin{aligned} \frac{d^n}{dt^n}\left\{P\int_{-\infty}^{\infty} \frac{\varphi(x)}{(x-t)}\,dx\right\} &= P\int_{-\infty}^{\infty} \frac{n!}{(x-t)^{n+1}}\left[\varphi(x) - \sum_{m=0}^{n-1} \frac{1}{m!}(x-t)^m \varphi^{(m)}(t)\right] dx, \\ &= P\int_{-\infty}^{\infty} \frac{\varphi^{(n)}}{x-t}\,dx. \end{aligned}$$

Exercise 2.24: Let the non-local functional $S[f]$ be defined by

$$S[f] = \frac{1}{4\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left\{\frac{f(x)-f(x')}{x-x'}\right\}^2 dx dx'$$

Compute the functional derivative of $S[f]$ and verify that it is given by

$$\frac{\delta S}{\delta f(x)} = \frac{1}{\pi}\frac{d}{dx}\left\{P\int_{-\infty}^{\infty} \frac{f(x')}{x-x'}\,dx'\right\}.$$

See exercise 6.10 for an occurence of this functional.