Chapter 9

The delta function and other distributions

9.1 Introduction

In this chapter we give a very informal introduction to *distributions*, also called *generalised functions*. We do two rather amazing things: we see how to differentiate a function with a jump discontinuity, and we develop a mathematical framework for point forces, masses, charges, sources etc. Furthermore, we find that these two ideas find their expression in the same mathematical object: the Dirac delta function.

When I learned proper real analysis for the first time, we spent ages agonising about continuity, left and right limits, one- and two-sided derivatives, and so on. The result was a lingering fear of pathological functions (continuous everywhere differentiable nowhere, that sort of thing) and associated technicalities. It came as a great relief to find (much later on, alas) that by getting away from the pointwise emphasis of introductory analysis one can give a beautifully consistent and holistic definition of the derivative of the *Heaviside function*[1]

$$\mathcal{H}(x) = \begin{cases} 1 & x > 0, \\ 0 & x \leq 0. \end{cases}$$

In pointwise mode, the best we can do with this function is to talk about the left and right limits of its derivative at the origin. Both these are equal to zero, but the function nevertheless gets up from 0 to 1. There must be something behind this!

The Heaviside function and its derivative, the delta function (or distribution), are ubiquitous in whole swathes of linear applied mathematics, not to mention discrete probability. They, and other distributions, are invaluable in developing an intuitive framework for modelling and its interaction with mathematics. Don't be inhibited about using them: your mistakes are unlikely to do worse than lead to inconsistencies (which I hope you are constantly on the look out for) and plainly wrong answers, rather than the deadly 'plausible but fallacious' solution.

[1] The value $\mathcal{H}(0) = 0$ has been assigned for consistency with probability, as we shall see; but for reasons that will shortly become clear it really doesn't matter what value we take.

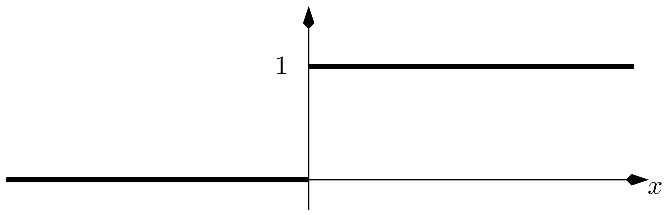

Figure 9.1: The Heaviside function $\mathcal{H}(x)$. Its derivative vanishes for all $x \neq 0$ but it still gets up from 0 to 1. How?

9.2 A point force on a stretched string; impulses

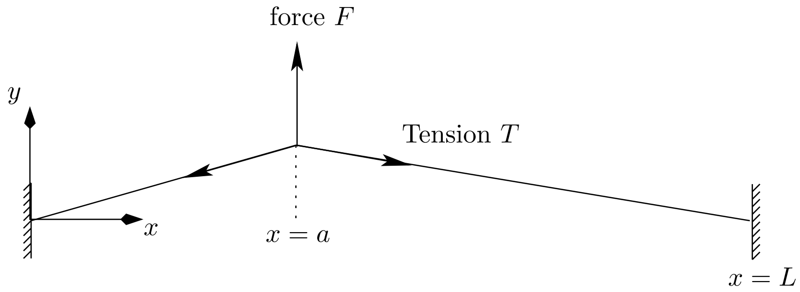

Figure 9.2: A string with a point force.

Let's start with a couple of motivating physical examples. We have all at some time worked out the displacement of a stretched string under the influence of a point force, as sketched in Figure 9.2. Under the standard assumptions that the string is effectively weightless, and that the force F (measured upwards, in the same direction as y) can be considered as acting at a point $x = a$ and only causes a small deflection, the equilibrium displacement $y(x)$ of the string satisfies

$$\frac{d^2y}{dx^2} = 0, \qquad 0 < x < a, \quad a < x < L, \tag{9.1}$$

with the force balance condition

Consistency check on the signs: $F > 0$ and dy/dx is negative to the right of a, positive to the left.

$$\left[T\frac{dy}{dx}\right]_{x=a-}^{x=a+} = -F. \tag{9.2}$$

Notice the implicit assumption that y itself is continuous at x_0 although its derivative is not.

Now we might ask, can we somehow put the force on the right-hand side of (9.1), and have the equilibrium conditions hold at $x = a$ as well? After all, if we have a distributed force per unit length $f(x)$ on the string, the usual force balance on a small element (see Figure 9.3) gives the equation[2]

$$T\frac{d^2y}{dx^2} = -f(x), \qquad 0 < x < L.$$

[2]You might wonder why there is a minus sign on the right. If we were to consider the unsteady motion of the string, Newton's Second Law in the form

$$\text{mass} \times \text{acceleration} = \text{force}$$

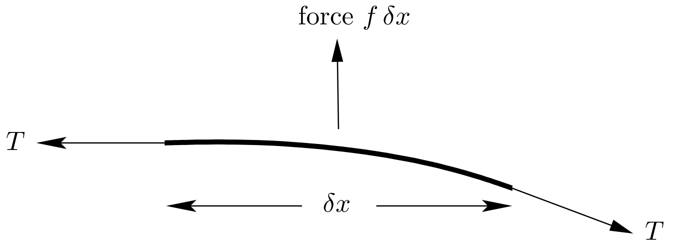

Figure 9.3: Force on an element of a string.

For example, when $f = -\rho g$, the gravitational force on a uniform wire of line density ρ, the displacement is a parabola (the small-displacement approximation to a catenary).

Can we devise some limiting process in which all the force becomes concentrated near $x = a$, with the total force $\int_0^L f(x)\,dx$ tending to F? A possible way to do this would be to take

Question expecting the answer yes.

$$f(x) = \begin{cases} F/2\epsilon & a-\epsilon < x < a+\epsilon, \\ 0 & \text{otherwise,} \end{cases}$$

and then to let $\epsilon \to 0$. But would we get the same answer if we took the limit of some other concentrated force density, and in any case how, exactly, are we to interpret the result of this limiting process?

In a very similar vein, recall the concept of an impulse in mechanics. In one-dimensional motion, the velocity v of a particle under a force $f(t)$ satisfies Newton's equation

$$m\frac{dv}{dt} = f(t),$$

from which

$$v(t) = v(0) + \frac{1}{m}\int_0^t f(s)\,ds.$$

If the force is very large but only lasts for a short time, say

$$f(t) = \begin{cases} I/\epsilon & 0 < t < \epsilon, \\ 0 & \text{otherwise,} \end{cases}$$

then we can integrate the equation of motion from $t = 0$ to $t = \epsilon$ to find

$$v(\epsilon) = \frac{1}{m}\int_0^\epsilon \frac{I}{\epsilon}\,dt = \frac{I}{m}.$$

gives

$$\rho\frac{\partial^2 y}{\partial t^2} = T\frac{\partial^2 y}{\partial x^2} + f,$$

leading to the minus sign in question. Many mathematicians, writing the wave equation as

$$\rho\frac{\partial^2 y}{\partial t^2} - T\frac{\partial^2 y}{\partial x^2} = f,$$

would write the equilibrium equation for the string as

$$-T\frac{d^2 y}{dx^2} = f(x).$$

Note the absence of minus signs in the impulse example that follows.

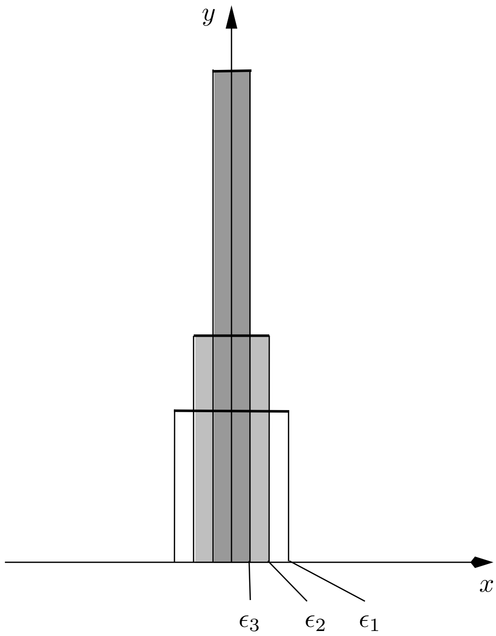

Figure 9.4: Three approximations to the delta function; $\epsilon_1 > \epsilon_2 > \epsilon_3 > 0$.

Notice that the wire slope has a jump discontinuity at a point force.

Letting $\epsilon \to 0$, we have the result of an *impulse* I: the velocity v changes discontinuously from 0 to I/m. Again, we can ask the question, can we put the limiting impulse directly into the equation of motion, rather than having to smooth it out and take a limit?

9.3 Informal definition of the delta and Heaviside functions

Obviously the answer to all our questions above is yes. The powerful and elegant theory of distributions allows us to model point forces and much more (dipoles, for example). However, the intuitive view of a point force (mass, charge, ...) as the limit of a distributed force turns out to be technically very cumbersome, and nowadays a more concise and general, but physically less intuitive, treatment is preferred. This oblique approach requires some groundwork, and we defer a brief self-contained description until Chapter 10. You will survive if you don't read it, although I recommend that you do: it is not technically demanding or complex.

In this chapter we concentrate on the intuitive approach to the delta function. Although this is not how the theory is nowadays developed, *it absolutely is how to visualise this central part of it.* Taking the examples of the previous section and stripping away the physical background, consider the functions

$$f_\epsilon(x) = \begin{cases} 1/2\epsilon & -\epsilon < x < \epsilon, \\ 0 & \text{otherwise.} \end{cases}$$

They are shown in Figure 9.4 for various values of ϵ. The following facts are obvious:

- $\displaystyle\int_{-\infty}^{\infty} f_\epsilon(x)\,dx = 1 \qquad \text{for all } \epsilon > 0;$
- for $x \neq 0, \qquad f_\epsilon(x) \to 0 \quad \text{as } \epsilon \to 0.$

The limiting 'function' is very strange indeed. It has a 'mass', or 'area under the graph', of 1, but that mass is all concentrated at $x = 0$. This is just what we need to model a point force, and even though we don't quite know how to interpret it rigorously, we provisionally christen the limit as the *delta function*, $\delta(x)$.

Two extremely useful properties of the delta function are now at least plausible. Firstly, as $\epsilon \to 0$,

$$\int_{-\infty}^{x} f_\epsilon(s)\,ds \to \begin{cases} 1 & x > 0, \\ 0 & x < 0, \end{cases}$$

and the right-hand side is the Heaviside function $\mathcal{H}(x)$ with its jump discontinuity at $x = 0$. So, we should have

For now, let's not worry what its value is at $x = 0$.

$$\int_{-\infty}^{x} \delta(s)\,ds = \mathcal{H}(x),$$

at least for $x \neq 0$. Furthermore, fingers crossed and appealing to the Fundamental Theorem of Calculus, we should conversely have

$$\frac{d}{dx}\mathcal{H}(x) = \delta(x).$$

That is, *delta functions let us differentiate functions with jump discontinuities.* The Heaviside function has a jump up of 1 at $x = 0$, and its derivative is $\delta(x)$, and by an obvious extension, the derivative of a function with a jump of A at $x = a$ contains a term $A\delta(x - a)$.

The second vital attribute of $\delta(x)$ is its 'sifting' property. Intuitively, for sufficiently smooth functions $\phi(x)$,

A proof is requested in the exercises.

$$\int_{-\infty}^{\infty} f_\epsilon(x)\phi(x)\,dx \to \phi(0) \quad \text{as } \epsilon \to 0,$$

simply because all the mass of $f_\epsilon(x)$, and hence of the product $f_\epsilon(x)\phi(x)$, becomes concentrated at the origin. So, we conjecture that we can make sense of the statement

$$\int_{-\infty}^{\infty} \delta(x)\phi(x)\,dx = \phi(0) \tag{9.3}$$

and, by a simple change of variable,

$$\int_{-\infty}^{\infty} \delta(x - a)\phi(x)\,dx = \phi(a)$$

for any real a.

These assertions are eminently plausible. However, if you stop to think how you might make them mathematically acceptable, difficulties start to appear.

Would we get the same results if we used a different approximating sequence $g_\epsilon(x)$? Do we need to worry about the value of $\mathcal{H}(0)$? Having differentiated $\mathcal{H}(x)$, can we define $d\delta/dx$? Clearly this last runs a big risk of being very dependent on the approximating sequence we use.

For example,

$$g_\epsilon(x) = \frac{1}{\epsilon\sqrt{2\pi}} e^{-\frac{1}{2}x^2/\epsilon}$$

as discussed in Section 9.4

For all these reasons, and more, the theory is best developed slightly differently, without the 'epsilonology'.[3] The clue lies in the sifting property. Using the fact that integration is a smoothing process, we can get away from the 'pointwise' view of functions which is so troublesome, and instead define distributions via *averaged* properties. An example is the integral (9.3), which leads to the definition of $\delta(x)$.[4] Before looking at this idea in more detail, we consider some examples.

9.4 Examples

9.4.1 A point force on a wire revisited

All our discussion suggests that we should model the point force F acting at $x = a$ by a term $F\delta(x - a)$ in the equilibrium equation for the displacement, and assume that the latter now holds for *all* x, so that

$$T\frac{d^2y}{dx^2} = -F\delta(x-a), \qquad 0 < x < L.$$

Assuming we believe that differentiation still makes sense.

We now know that this means that the left-hand side is the derivative of a function which jumps by F at $x = a$. But the left-hand side is also the derivative of $T\,dy/dx$. Thus, putting the delta function into the equilibrium equation leads *automatically* to the force balance

$$\left[T\frac{dy}{dx}\right]_{a-}^{a+} = -F,$$

and there is no need to state this separately.

9.4.2 Continuous and discrete probability.

We can interpret each of the approximations $f_\epsilon(x)$ of Figure 9.4 as the probability density of a random variable X_ϵ whose value is uniformly distributed on the interval $(-\epsilon, \epsilon)$. The mean of this distribution is 0 and its standard deviation is $\epsilon/\sqrt{3}$. As $\epsilon \to 0$, the random variable becomes equal to 1 with certainty, because its standard deviation tends to zero, and any random variable with zero standard deviation must be a constant. This suggests that we can interpret the delta function as the probability density 'function' of a variable whose probability of being equal to zero is 1. Likewise, the cumulative density function (distribution function) $F_{X_\epsilon}(x) = P(X_\epsilon < x)$ tends to the Heaviside function.[5]

[3]See [42] page 97 for this neologism.

[4]The process of generalisation by looking at a weaker (smoother) definition using an integral, rather than a pointwise definition, is common in analysis. A famous example in applied mathematics is the definition of weak solutions to hyperbolic conservation laws, which leads to the Rankine–Hugoniot relations for a shock.

[5]In this case the strict inequality in the definition of F_{X_ϵ} suggests that we should take $\mathcal{H}(0) = 0$. Looking in the books on my shelf, I find that there is no consensus in the probability

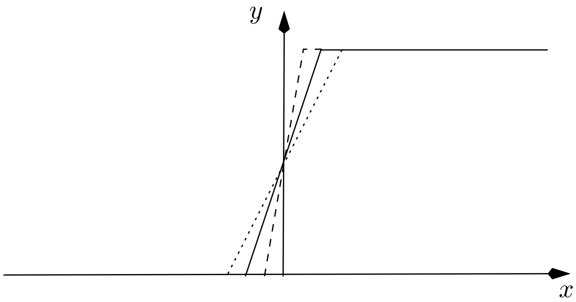

Figure 9.5: Cumulative density functions for the distributions of Figure 9.4.

In a similar vein, we can take approximations

$$g_\epsilon(x) = \frac{1}{\epsilon\sqrt{2\pi}} e^{-\frac{1}{2}x^2/\epsilon^2},$$

which are the density functions of normal random variables with mean zero and standard deviation ϵ. These also clearly tend to the delta function as $\epsilon \to 0$.

Now suppose we have a coin-toss random variable X taking values ± 1 with equal probability $\frac{1}{2}$. As X can only equal 1 or -1, all its probability mass is concentrated at these values: its density function is zero for $x \neq \pm 1$. The density of this random variable is

What is its distribution function?

$$f_X(x) = \tfrac{1}{2}\left(\delta(x+1) + \delta(x-1)\right).$$

In this way, we can unify continuous and discrete probability — at least when the number of discrete events is finite. The extension to infinitely many discrete events is much more difficult, and may require the tools of measure theory.

9.4.3 The fundamental solution of the heat equation

If we set $\epsilon = 2t$ in the functions g_ϵ of the previous section, we get the function

$$g(x,t) = \frac{1}{2\sqrt{\pi t}} e^{-x^2/4t}.$$

Direct differentiation shows that $g(x,t)$ satisfies the heat equation. As we saw above, as $t \downarrow 0$, $g(x,t) \to \delta(x)$. In summary, $g(x,t)$ satisfies the initial value problem

$$\frac{\partial g}{\partial t} = \frac{\partial^2 g}{\partial x^2}, \qquad t > 0, \quad -\infty < x < \infty,$$
$$g(x,0) = \delta(x).$$

This solution represents the evolution of a 'hot spot', a unit amount of heat

Note the infinite propagation speed of the heat: $t = 0$ is a (double) characteristic of the heat equation. Note also the very rapid decay in the solution as $|x|$ increases.

world whether to use $P(X < x)$ or $P(X \leq x)$ to define the distribution function (no wonder I can never remember). It is a matter of convention only, and would lead to corresponding conventional definitions of $\mathcal{H}(0)$. Another highly plausible definition is $\mathcal{H}(0) = \frac{1}{2}$, on the grounds that any Fourier series or transform inversion integral for a function with a jump converges to the average of the values on either side. This sort of hair splitting is one reason why the pointwise view of distributions is not really workable.

which at $t = 0$ is concentrated at $x = 0$.

With this solution, we can solve the more general initial value problem

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad t > 0, \quad -\infty < x < \infty,$$
$$u(x, 0) = u_0(x).$$

We first note that the initial data $u_0(x)$ can be written as

$$u_0(x) = \int_{-\infty}^{\infty} u_0(\xi)\delta(x - \xi)\, d\xi$$

by the picking-out property of the delta function. Now the evolution of a solution with initial data $\delta(x-\xi)$ is just $g(x-\xi, t)$ where g is as above. The integral over ξ amounts to superposing the initial data for these solutions, so that each point contributes a delta function weighted by $u_0(\xi)\, d\xi$. Because the heat equation is linear, we can superpose for $t > 0$ as well, so we have

Confirm that $u(x, t)$ satisfies the heat equation by differentiating under the integral sign.

$$\begin{aligned} u(x, t) &= \int_{-\infty}^{\infty} u_0(\xi) g(x - \xi, t)\, d\xi \\ &= \frac{1}{2\sqrt{\pi t}} \int_{-\infty}^{\infty} u_0(\xi) e^{-(x-\xi)^2/4t} d\xi. \end{aligned}$$

This solution has a physical interpretation as the superposition of elementary 'packets' of heat evolving independently.[6]

9.5 'Balancing the most singular terms'

If we have an equation involving 'ordinary' functions, and there is a singularity on one side, there must be a balancing singularity somewhere else. For example, we could never find coefficients a_n such that

$$\frac{1}{\sin x} = a_0 + a_1 x + a_2 x^2 + \cdots$$

because the left-hand side clearly has a $1/x$ (simple pole) singularity at $x = 0$. On the other hand there *is* an expansion

This is just the Laurent expansion.

$$\frac{1}{\sin x} = \frac{a_{-1}}{x} + a_0 + a_1 x + a_2 x^2 + \cdots ,$$

and furthermore we know that $a_{-1} = 1$ because $1/\sin x \sim 1/x$ as $x \to 0$. Thus, both sides have this singularity in their leading-order behaviour as $x \to 0$.

This is a simple but powerful idea, and it applies to distributions as well. In our naive approach, a delta function is a 'function' with a particular singularity at $x = 0$. Thus, if part (for example the right-hand side) of an equation contains a delta function as its 'most singular' term, there must be a balancing term somewhere else. For instance, when we write

$$\frac{dv}{dt} = \frac{I}{m}\delta(t),$$

ck and look at the point on a string in this light.

for the motion of a particle subject to a point force, there must be another singularity to balance the delta function. It can only be in dv/dt, so we know straightaway that v has a jump at $t = 0$; furthermore, we know that the magnitude of the jump is I/m, by 'comparing coefficients' of the delta functions. In this case it is trivial to find the balancing term, because there is only one candidate. Suppose, though, that the equation has a linear damping term:

$$m\frac{dv}{dt} = -mkv + I\delta(t),$$

where $k > 0$ is the damping coefficient. The balancing singularity is still in the derivative dv/dt, simply because dv/dt always has worse singularities than v itself. Going back, we can check: if dv/dt has a delta, then v has a jump, which is indeed less singular.

Differentiation makes matters worse, integration makes them better.

9.5.1 The Rankine–Hugoniot conditions

In Chapter 7 we looked briefly at the Rankine–Hugoniot conditions for a first order conservation law

$$\frac{\partial P}{\partial t} + \frac{\partial Q}{\partial x} = 0$$

where, for example, P is the density ρ of traffic and Q the flux $u\rho$. We saw that we can construct solutions in which P and Q have jump discontinuities across a shock at $x = S(t)$, provided that

$$\frac{dS}{dt} = \frac{[Q]}{[P]}.$$

We can interpret this condition as a balance of delta functions. If P has a (time-dependent) jump of magnitude $[P](t)$ at $x = S(t)$, we can (very informally) write

$$P(x,t) = [P](t)\mathcal{H}\left(x - S(t)\right) + \text{smoother part},$$

and similarly for $Q(x,t)$. Differentiating, we find

$$\begin{aligned}\frac{\partial P}{\partial t} &= -[P](t)\delta\left(x - S(t)\right)\frac{dS}{dt} + \text{less singular terms},\\ \frac{\partial Q}{\partial x} &= [Q](t)\delta\left(x - S(t)\right) + \text{less singular terms}.\end{aligned}$$

Adding these and balancing the coefficients of the delta functions, the Rankine–Hugoniot condition drops out.

9.5.2 Case study: cable-laying

In Chapter 4, we wrote down the model

$$\frac{dF_x}{ds} = -B_x, \quad \frac{dF_y}{ds} = -B_y + \rho_c g A = 0, \tag{9.4}$$

$$EAk^2\frac{d^2\theta}{ds^2} - F_x\sin\theta + F_y\cos\theta = 0, \tag{9.5}$$

[6] There is also an interpretation in terms of random walkers following Brownian Motion: see Exercise 9 on page 137.

where

$$(B_x, B_y) = \left(\rho_w g A \cos\theta + pA\frac{d\theta}{ds}\right)(-\sin\theta, \cos\theta). \tag{9.6}$$

for a cable being laid on a sea bed, where θ is the angle between the cable and the horizontal. We stated, on a rather intuitive basis, that the boundary conditions at $s = 0$ are $\theta = 0$ (no worries about this one) and $d\theta/ds = 0$, namely continuity of θ and $d\theta/ds$, since $\theta = 0$ for $s < 0$. We can now see why this is necessary. If $d\theta/ds$ is not continuous, then $d^2\theta/ds^2$ has a delta function discontinuity at $s = 0$. But then there is no balancing term in (9.5) since, loosely, (9.4) shows that both F_x and F_y are at least as continuous as B_x and B_y, and so from (9.6) they are no worse than $d\theta/ds$ with its assumed-for-a-contradiction jump discontinuity; we have duly obtained said contradiction.

Because there is a reaction force between the sea bed and the cable, and maybe some friction, we do not expect the right-hand sides of (9.4) to be continuous at $s = 0$.

9.6 Green's functions

9.6.1 Ordinary differential equations

The two-point boundary value problem[7]

$$\mathcal{L}_x y(x) = \frac{d}{dx}\left(p(x)\frac{dy}{dx}\right) + q(x)y = f(x), \qquad 0 < x < 1, \tag{9.7}$$

$$y(0) = y(1) = 0, \tag{9.8}$$

is standard. often arising in a separation of variables calculation in an exotic coordinate system. As a matter of terminology, we call the combination of $\mathcal{L}_x$, the interval on which it is applied, and the boundary conditions at the ends of this interval, the *differential operator* for this problem. Changing any of these changes the differential operator. The operator (9.7), (9.8) above is called *self-adjoint*, a term that will be made clearer later.

One of the first things that one does with problems of this kind is to show that they can be solved with the help of a *Green's function*. Provided that the homogeneous problem ($f(x) \equiv 0$) has no non-trivial solutions, the Green's function is the function $G(x, \xi)$ that satisfies

$$\mathcal{L}_\xi G(x, \xi) = 0, \qquad 0 < \xi < x, \quad x < \xi < 1, \tag{9.9}$$

$$G(x, 0) = G(x, 1) = 0, \tag{9.10}$$

with some rather opaque-seeming conditions at $\xi = x$:

$$[G]_{\xi=x-}^{\xi=x+} = 0, \qquad \left[p(\xi)\frac{dG}{d\xi}\right]_{\xi=x-}^{\xi=x+} = 1 \quad \left(\text{or} \quad \left[\frac{dG}{d\xi}\right]_{\xi=x-}^{\xi=x+} = \frac{1}{p(x)}\right). \tag{9.11}$$

If we can solve this problem, then we have a representation for $y(x)$ as

$$y(x) = \int_0^1 G(x, \xi) f(\xi)\, d\xi.$$

The elementary proof of this is by direct construction of the Green's function via variation of parameters, assuming the existence of appropriate solutions of

ke inverting a matrix **A**
ve $\mathbf{Ax} = \mathbf{b}$; see
se 7.

the homogeneous equation, and we do not describe it here. The point is that we need only calculate G once, and then we have the solution whatever we take for $f(x)$.[8] In this way, we can think of the operation of multiplying by the Green's function and integrating as the inverse of the differential operator $\mathcal{L}$.

This is all very well, but I don't think it gives a good intuitive feel for what the Green's function really *does*. Suppose, though, that we take the solution

$$y(x) = \int_0^1 G(x,\xi)f(\xi)\,d\xi \tag{9.12}$$

and apply $\mathcal{L}_x$ to it. Assuming that we can differentiate under the integral, we get

$$\begin{aligned}\mathcal{L}_x y(x) &= \int_0^1 \mathcal{L}_x G(x,\xi)f(\xi)\,d\xi \\ &= f(x).\end{aligned}$$

We recognise this: it is the sifting property. Whatever f we take, when we multiply $f(\xi)$ by $\mathcal{L}_x G(x,\xi)$ and integrate, we get $f(x)$. Thus, as a function (actually, a distribution) of x, $G(x,\xi)$ satisfies

$$\mathcal{L}_x G(x,\xi) = \delta(x-\xi),$$

that is

$$\frac{d}{dx}\left(p(x)\frac{dG}{dx}\right) + q(x)G = \delta(x-\xi).$$

Also, the boundary conditions $y(0) = y(1) = 0$ mean that we need to take

$$G(0,\xi) = G(1,\xi) = 0,$$

so that (9.12) satisfies the boundary conditions whatever $f(x)$ we take. In summary, as a function of x, the Green's function satisfies the differential equation with a delta-function on the right-hand side, and with the homogeneous version of the original boundary conditions.

This calculation tells us several things. Thinking physically, it tells us that *the Green's function is the response of the system to a point stimulus (force, charge, ...) at* $x = \xi$. The solution (9.12) is then just the response to $f(x)$, regarded as a superposition of point stimuli (the delta function at $x = \xi$) weighted by $f(\xi)\,d\xi$.

---

[7] The subscript to $\mathcal{L}$ tells you which variable to use. Strictly speaking, in much of the discussion to follow all the derivatives should be partial, but it seems to be conventional to stick to ordinary derivatives.

[8] A very common use of the Green's function is to turn a differential equation into an integral equation as a prelude to an iteration scheme to prove existence, uniqueness and regularity. Often the equation has a linear part and some nonlinearity as well, and we use the Green's function for the linear part. A simple example of this procedure is Picard's theorem for local existence and uniqueness of the solution to $d\mathbf{y}/dx = f(x,\mathbf{y})$, $\mathbf{y}(0) = \mathbf{y}_0$ for a set of first-order equations, where the first step is to write

$$\mathbf{y}(x) = \mathbf{y}_0 + \int_0^x f(\xi,\mathbf{y}(\xi))\,d\xi;$$

the only modification needed is to adapt the Green's function methodology to cater for initial value problems, as described in Exercise 4.

Looking more mathematically, if we expand $\mathcal{L}_x G(x,\xi)$ as

$$\mathcal{L}_x G(x,\xi) = p(x)\frac{d^2G}{dx^2} + \text{lower order derivatives},$$

we see by balancing the most singular terms (the highest derivatives) that d^2G/dx^2 must have a delta function, scaled by $p(x)$, at $x=\xi$. That is,

$$[G]_{x=\xi^-}^{x=\xi+} = 0, \qquad \left[p(x)\frac{dG}{dx}\right]_{x=\xi^-}^{x=\xi+} = 1 \quad \left(\text{or} \quad \left[\frac{dG}{dx}\right]_{x=\xi^-}^{x=\xi+} = \frac{1}{p(\xi)}\right).$$

This should ring a bell. It is the same as the 'opaque' jump conditions (9.11), except that it refers to the x–dependence of $G(x,\xi)$ instead of the ξ–dependence. Indeed, comparing the original definition of G given in (9.9)–(9.11) and recalling that $G(0,\xi) = G(1,\xi) = 0$, we see that the two formulations are identical except that x and ξ are swapped. That is, we have established that, for self-adjoint problems,

Note that $\delta(x-\xi) = \delta(\xi - x)$.

$$G(x,\xi) = G(\xi,x)$$

and that

$$\mathcal{L}_\xi G(x,\xi) = \delta(\xi - x).$$

We are now in a position to tie together the x and ξ dependence of $G(x,\xi)$. Consider the integral

$$\int_0^1 y(\xi)\mathcal{L}_\xi G(x,\xi) - G(x,\xi)\mathcal{L}_\xi y(\xi)\, d\xi.$$

Inserting the right-hand sides of the differential equations for G and y, we get

$$\begin{aligned}\int_0^1 y(\xi)\mathcal{L}_\xi G(x,\xi) - G(x,\xi)\mathcal{L}_\xi y(\xi)\, d\xi &= \int_0^1 y(\xi)\delta(\xi - x) - G(x,\xi)f(\xi)\, d\xi \\ &= y(x) - \int_0^1 G(x,\xi)f(\xi)\, d\xi.\end{aligned}$$

On the other hand, integrating the same expression by parts, we get

$$\begin{aligned}\int_0^1 y(\xi)\mathcal{L}_\xi G(x,\xi) - G(x,\xi)\mathcal{L}_\xi y(\xi)\, d\xi &= \int_0^1 y(\xi)\left(\frac{d}{d\xi}\left(p(\xi)\frac{dG}{d\xi}\right) + q(\xi)G(x,\xi)\right) \\ &\qquad - G(x,\xi)\left(\frac{d}{d\xi}\left(p(\xi)\frac{dy}{d\xi}\right) + q(\xi)y(\xi)\right)\, d\xi \\ &= \left[y(\xi)p(\xi)\frac{dG}{d\xi} - G(x,\xi)p(\xi)\frac{dy}{d\xi}\right]_0^1 \\ &\qquad - \int_0^1 p(\xi)\frac{dy}{d\xi}\frac{dG}{d\xi} - p(\xi)\frac{dG}{d\xi}\frac{dy}{d\xi}\, d\xi \\ &= 0.\end{aligned}$$

Thus we retrieve the solution

$$y(x) = \int_0^1 G(x,\xi)f(\xi)\, d\xi.$$

unction of x the
ntial equation for G is
${}_xG = 0$ with the same
ary conditions as those

This calculation is really the key to the whole procedure. It tells us that the differential equation and boundary conditions (that is, the differential operator) for G as a function of ξ must be such that we can integrate by parts and get zero (so in the second line of our calculation, we must have zero multiplying dy/dx, about which we know nothing at the endpoints).

Non-self-adjoint problems

For a self-adjoint problem, such as those discussed thus far, G is symmetric and the two operators, for y and G, are the same. Now suppose that we have a more general problem, such as

$$\mathcal{L}_x y(x) = a(x)\frac{d^2y}{dx^2} + b(x)\frac{dy}{dx} + c(x)y = f(x),$$

with the boundary conditions (sometimes called *primary boundary conditions*

$$\alpha_0 y(0) + \beta_0 y'(0) = 0, \qquad \alpha_1 y(1) + \beta_1 y'(1) = 0.$$

(to save ink, $y' = dy/dx$). We aim to find a differential operator for G which allows us to follow the calculation above as closely as possible. That is, we want to find a combination of derivatives $\mathcal{L}^*$ such that, as a function of ξ, $G(x,\xi)$ satisfies

$$\mathcal{L}^*_\xi G(x,\xi) = \delta(x-\xi),$$

with appropriate boundary conditions. We can then integrate by parts as above; and provided that

$$\int_0^1 y(\xi)\mathcal{L}^*_\xi G(x,\xi) - G(x,\xi)\mathcal{L}_\xi y(\xi)\, d\xi = 0,$$

we have the answer

$$y(x) = \int_0^1 G(x,\xi)f(\xi)\, d\xi.$$

For the general operator just introduced, the new operator, called the *adjoint operator*, is given by

$$\mathcal{L}^* v(x) = \frac{d^2}{dx^2}\left(a(x)v(x)\right) - \frac{d}{dx}\left(b(x)v(x)\right) + c(x)v(x),$$

with the *adjoint boundary conditions*

$$a(0)\big(\alpha_0 v(0) + \beta_0 v'(0)\big) + \beta_0\big(a'(0) - b(0)\big)v(0) = 0,$$
$$a(1)\big(\alpha_1 v(1) + \beta_1 v'(1)\big) + \beta_1\big(a'(1) - b(1)\big)v(1) = 0,$$

as you will find out by doing Exercise 5.

You might very reasonably ask why we bother with the adjoint when all we need to do is differentiate the answer

$$y(x) = \int_0^1 G(x,\xi)f(\xi)\, d\xi$$

under the integral sign to show that

$$\mathcal{L}_x y = \int_0^1 \mathcal{L}_x G(x,\xi) f(\xi)\, d\xi \\ = f(x),$$

so that

$$\mathcal{L}_x G(x,\xi) = \delta(x-\xi)$$

with no mention of adjoints at all. An aesthetic reason is the mathematical structure uncovered (compare vector spaces and their duals), but a compelling practical reason is that if the primary boundary conditions are *inhomogeneous*, for example $y(0) = y_0 \neq 0$, $y(1) = y_1 \neq 0$, then only the adjoint calculation works (try it!).

One can take all this a great deal further, both making it more rigorous and looking at more general problems. I recommend reading the relevant parts of [32] or [54] if you want to do this; we are moving on to a brief look at partial differential equations.

9.6.2 Partial differential equations

Much of the theory we have just seen can be generalised to linear partial differential equations. This is so much vaster a topic that it is only feasible to discuss one example in detail, the Green's function for Poisson's equation, which is probably the closest in spirit to the two-point boundary value problems we have been discussing so far. We then briefly mention two other canonical problems, for the heat equation and the wave equation.

We first have to generalise the delta function. In our informal style, this is easy: we just say that for $\mathbf{x} \in \mathbb{R}^n$, the delta function $\delta(\mathbf{x})$ is such that

$$\int_{\mathbb{R}^n} \delta(\mathbf{x})\phi(\mathbf{x})\, d\mathbf{x} = \phi(\mathbf{0})$$

for all smooth functions $\phi(\mathbf{x})$. As before, we can think of this as a limiting process in which the delta function is the limit of a family of functions whose mass becomes more and more concentrated near the origin.[9] Thinking about how the integral is calculated, say in two dimensions with $d\mathbf{x} = dx\, dy$, we may also write

$$\delta(\mathbf{x}) = \delta(x)\delta(y),$$

and similarly in three or more variables.

Think of some physical interpretations for u, and then for the Green's function G.

Now suppose that we have to solve the problem

$$\mathcal{L}_{\mathbf{x}} u(\mathbf{x}) = \nabla^2 u(\mathbf{x}) = f(\mathbf{x})$$

in some region D, with the homogeneous Dirichlet boundary condition

$$u(\mathbf{x}) = 0 \qquad \text{on} \quad \partial D.$$

The Laplacian is self-adjoint ($\mathcal{L} = \mathcal{L}^*$) ...

We choose the Green's function to satisfy

[9] They might, but need not, be radially symmetric; we might, but won't, worry about how to define integrals in n dimensions.

$$\mathcal{L}_\xi G(\mathbf{x},\xi) = \delta(\xi - \mathbf{x})$$

and look at the integral

$$\begin{aligned}\int_D u(\xi)\mathcal{L}_\xi G(\mathbf{x},\xi) - G(\mathbf{x},\xi)\mathcal{L}_\xi u(\mathbf{x})\, d\xi &= \int_D u(\xi)\nabla_\xi^2 G(\mathbf{x},\xi) - G(\mathbf{x},\xi)\nabla_\xi^2 u(\mathbf{x})\, d\xi \\ &= \int_D u(\xi)\delta(\xi-\mathbf{x}) - G(\mathbf{x},\xi)f(\xi)\, d\xi \\ &= u(\mathbf{x}) - \int_D G(\mathbf{x},\xi)f(\xi)\, d\xi.\end{aligned}$$

On the other hand, using Green's theorem, we have

$$\int_D u(\xi)\nabla^2 G(\mathbf{x},\xi) - G(\mathbf{x},\xi)\nabla^2 u(\mathbf{x})\, d\xi \tag{9.13}$$

$$= \int_{\partial D} u(\xi)\mathbf{n}\cdot\nabla_\xi G(\mathbf{x},\xi) - G(\mathbf{x},\xi)\mathbf{n}\cdot\nabla_\xi u(\mathbf{x})\, dS_\xi$$

$$= 0, \tag{9.14}$$

. . . because $u\nabla^2 G - G\nabla^2 u$ is a divergence and can be integrated (a generalisation of integration by parts).

provided that we take $G(\mathbf{x},\xi) = 0$ for $\xi \in \partial D$, where we do not know the normal derivative of u. Putting these together, we have

$$u(\mathbf{x}) = \int_D G(\mathbf{x},\xi)f(\xi)\, d\xi.$$

It is an easy generalisation to account for nonzero Dirichlet data $u(\mathbf{x}) = g(\mathbf{x})$ on ∂D: we just get an extra known term in (9.14).

Two more things should be said about this calculation. The first is that we have not yet said anything about the nature of the singularity of $G(\mathbf{x},\xi)$ at $\mathbf{x} = \xi$ (in one space dimension, as we saw above, the first derivative of G has a jump and G itself is continuous). Knowing as we do that line (in two dimensions) or point (in three) charges generate electric fields which are solutions of Laplace's equation, we should not be surprised to see logs in two dimensions and inverse distances in three. This is confirmed by a simple version of the calculation we have just done.[10] In $\mathbb{R}^3$ for example, take $\xi = \mathbf{0}$ and suppose that

Or line/point masses and their gravitational potentials, fluid sources and their velocity potentials, or heat sources and their steady-state temperature fields.

$$\nabla^2 G = \delta(\mathbf{x}) \tag{9.15}$$

in the whole space. Clearly, then, G is radially symmetric: $G = G(r)$ where $r = |\mathbf{x}|$. That means that

$$G(r) = A + \frac{B}{r}$$

and if we want $G \to 0$ as $r \to \infty$, we take $A = 0$. Now use the divergence theorem on the left-hand side of (9.15), integrating over a sphere of radius r centred at $\mathbf{x} = \mathbf{0}$. The left-hand side gives a surface integral equal to $-4\pi B/r$ and the volume integral of the delta function on the right is equal to 1. We conclude that the singular behaviour of $G(\mathbf{x},\xi)$ near $\mathbf{x} = \xi$ is

The meaning of $\sim$ and $O(1)$ is explained in Chapter 12.

$$G(\mathbf{x},\xi) \sim -\frac{1}{4\pi|\mathbf{x}-\xi|} + O(1),$$

and in two dimensions the corresponding result is

Can you now answer the marginal question after equation (4) on page 33?

$$G(\mathbf{x}, \xi) \sim \frac{1}{2\pi} \log |\mathbf{x} - \xi| + O(1).$$

The second point to make about the Green's function for the Laplacian is that it has a natural physical interpretation. The singular part we have just discussed gives us the electric potential due to a point charge (or whatever) with no boundaries. The remaining part, $G + 1/(4\pi|\mathbf{x} - \xi|)$, is known as the *regular part* of the Green's function and it gives the potential due to the image charge system induced by the boundary condition $G = 0$ on ∂D. Indeed, almost all the Green's functions for which explicit formulas are available are constructed by the method of images (possibly with the help of conformal maps).

The heat and wave equations

You can safely ignore this section, but have a look if you have seen the classical treatments of these problems.

To round off, let's look quickly at two other equations, the heat and wave equations in two space variables. Let us look at the simplest initial-value problem for the heat equation, on the whole line, namely

$$\mathcal{L}_{x,t} u = \frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0, \qquad -\infty < x < \infty, \quad t > 0,$$
$$u(x, 0) = u_0(x).$$

By any of a variety of methods (for example, the Fourier transform in x), we obtain the solution in the form

$$u(x,t) = \frac{1}{2\sqrt{\pi t}} \int_{-\infty}^{\infty} u_0(\xi) e^{-(x-\xi)^2/4t} \, d\xi.$$

It is no surprise that this is closely related to the Green's function. The adjoint to the forward heat equation is the backward heat equation, and as a function of ξ and τ (the analogue here of ξ above), $G(x,t;\xi,\tau)$ satisfies

$$\mathcal{L}^*_{\xi,\tau} G = \frac{\partial G}{\partial \tau} + \frac{\partial^2 G}{\partial \xi^2} = \delta(\xi - x)\delta(\tau - t),$$

and, remembering the fundamental solution of the forward heat equation (see Exercise 8 and reversing time,

Two minus signs from the exponent cancel.

$$G(x,t;\xi,\tau) = \frac{1}{2\sqrt{\pi(t-\tau)}} e^{-(x-\xi)^2/4(t-\tau)}.$$

The usual integration in the form

$$\int_{-\infty}^{\infty} \int_0^t u\mathcal{L}_{\xi,\tau} G - G\mathcal{L}_{\xi,\tau} u \, d\tau \, d\xi$$

then yields precisely the solution we derived earlier. It is an exercise to generalise this result to the heat equation with a source term, $\mathcal{L}u = f(x,t)$; you will get a

[10] In the more classical treatment of Green's functions, you see essentially this calculation when you integrate $u\nabla^2 G - G\nabla^2 u$ over a region consisting of D with a sphere of radius ϵ around $\mathbf{x} = \xi$ removed. There, the singular behaviour of G is prescribed (and looks mysterious: why this form?), whereas here it emerges naturally.

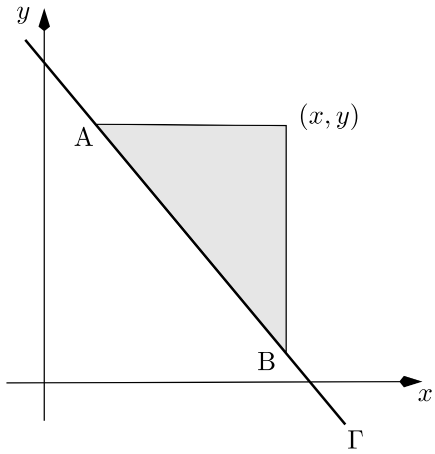

Figure 9.6: Domain of integration for the Riemann function for the wave equation.

double integral involving the product of G and f which has the simple physical interpretation of being a superposition of solutions of initial value problems starting at different times. Do it and see.

For the inhomogeneous wave equation in the canonical form

$$\mathcal{L}u = \frac{\partial^2 u}{\partial x \partial y} = f(x, y),$$

with Cauchy data u and $\partial u/\partial n$ given on a non-characteristic curve Γ, we proceed in the same spirit but differently in detail. One of the differences of detail is that the the Green's function is now usually called a Riemann function, and we denote it by $R(x, y; \xi, \eta)$. The differential operator $\partial^2 / \partial x \partial y$ is self-adjoint, but we have to consider the direction of information flow carefully (see Figure 9.6). When we solve

$$\mathcal{L}^* R = \frac{\partial^2 G}{\partial \xi \partial \eta} = \delta(\xi - x)\delta(\eta - y),$$

we look for a solution valid for $\xi < x$, $\eta < y$. Then the 'usual' integral

$$\int u\mathcal{L}^* R - R\mathcal{L}u$$

is taken over the characteristic triangle shaded in Figure 9.6, and after use of Green's theorem yields the solution in terms of an integral along Γ from A to B and an integral over the shaded triangle.

The Riemann function for the wave operator is particularly simple:

Differentiate it and see.

$$R(x, y; \xi, \eta) = \mathcal{H}(x - \xi)\mathcal{H}(y - \eta),$$

i.e. it is equal to 1 in the quadrant $\xi < x$, $\eta < y$ and zero elsewhere. It yields the familiar D'Alembert solution (see [42]). Unfortunately this is a rare explicit example. Although it is not hard to prove that the Riemann function exists, only for a very few hyperbolic equations can it be found in closed form.

Sources and further reading

The material on Green's functions is just a small step into Sturm–Liouville and Hilbert–Schmidt theory and eigenfunction expansions/transform methods. If you want to explore further, [25] gives a straightforward account of the theory for ordinary differential equations, [42] present an informal introduction to the corresponding material for partial differential equations, and the excellent [54] contains a more thorough account.

9.7 Exercises

1. **Truncated random variables.** Suppose that X is a continuous random variable taking values in $(-\infty, \infty)$, for example Normal. The *truncated* variable Y is defined by

$$Y = \begin{cases} X & \text{if } X < a \\ a & \text{if } X \geq a. \end{cases}$$

   What are its distribution and density functions?

2. **A useful identity.** Interchange the order of integration (draw a picture of the region of integration) to show that

$$\int_0^x \int_0^\xi f(s)\,ds\,d\xi = \int_0^x (x-\xi)f(\xi)\,d\xi.$$

   Generalise to reduce an n–fold repeated integral of a function of a single variable to a single integral.

3. **Green's function for a stretched string.** Integrate twice to find the solution of the two-point boundary value problem

$$\frac{d^2y}{dx^2} = f(x), \quad 0 < x < 1, \qquad y(0) = y(1) = 0$$

   in the form

$$y(x) = \int_0^1 G(x,\xi) f(\xi)\,d\xi.$$

   Verify that if you differentiate twice under the integral sign and use the jump conditions at $\xi = x$ you recover the original problem.

4. **Green's function for an initial value problem.**

   Use the result of Exercise 2 to show that the solution of the initial value problem

$$\frac{d^2y}{dx^2} = f(x), \quad 0 < x < 1, \qquad y(0) = \frac{dy}{dx}(0) = 0 \tag{9.16}$$

   is

$$y(x) = \int_0^x (x-\xi) f(\xi)\,d\xi.$$

Now pick $X > x$ and write this answer in the form

$$y(x) = \int_0^X G(x,\xi) f(\xi)\, d\xi;$$

what is G? Show that G satisfies

$$\frac{d^2G}{d\xi^2} = \delta(x-\xi), \quad 0 < \xi < X, \qquad G = \frac{dG}{d\xi} = 0 \quad \text{at } x = X.$$

Verify by differentiating under the integral sign that your answer satisfies the original problem. What is the adjoint problem (differential equation and boundary conditions) to the original problem (9.16)?

This kind of Green's function is the ordinary differential equation analogue of the Riemann function for a hyperbolic equation.

5. **Adjoint of a differential operator.** Suppose that

$$\mathcal{L}_x y = a(x)\frac{d^2y}{dx^2} + b(x)\frac{dy}{dx} + c(x)y,$$

with

$$\alpha_0 y(0) + \beta_0 y'(0) = 0, \qquad \alpha_1 y(1) + \beta_1 y'(1) = 0.$$

Show that the adjoint is

$$\mathcal{L}_x^* v = \frac{d^2}{dx^2}\left(a(x)v\right) - \frac{d}{dx}\left(b(x)v\right) + c(x)v$$

with

$$a(0)\big(\alpha_0 v(0) + \beta_0 v'(0)\big) + \beta_0\big(a'(0) - b(0)\big)v(0) = 0,$$
$$a(1)\big(\alpha_1 v(1) + \beta_1 v'(1)\big) + \beta_1\big(a'(1) - b(1)\big)v(1) = 0,$$

in either or both of the following ways.

(a) Show that $y\mathcal{L}_x^* v - v\mathcal{L}_x y$ can be integrated by parts as in the text;

The easy way if you know the answer.

(b) Write

$$\mathcal{L}_x^* v = A(x)\frac{d^2v}{dx^2} + B(x)\frac{dv}{dx} + C(x)v$$

What you might do if you didn't know the answer and couldn't guess it.

and hack away at the integration by parts (start by integrating the highest derivatives) until everything has been integrated. Whenever terms crop up that can't be integrated up, set them equal to zero to find A, B and C, and similarly determine the adjoint boundary conditions.

Hence verify that, for self-adjoint operators, $\mathcal{L}_x y$ is of the form

$$\mathcal{L}_x y = \frac{d}{dx}\left(p(x)\frac{dy}{dx}\right) + q(x)y$$

for some functions $p(x)$ and $q(x)$, while the boundary conditions are as above. Also show that periodic boundary conditions, $y(0) = y(1)$, $y'(0) = y'(1)$, give a self-adjoint operator as long as $p(0) = p(1)$.

What is the adjoint operator if $\mathcal{L}y = d^2y/dx^2$, $0 < x < 1$, and the boundary conditions for y are $y(0) = y(1) + y'(1)$, $y'(0) = 0$?

6. **The Fredholm Alternative: linear algebra and two-point boundary value problems.** Suppose that $\mathbf{A}$ is an $n \times n$ matrix, and we want to solve the linear equations

$$\mathbf{Ay} = \mathbf{f}$$

for the vector $\mathbf{y}$ given $\mathbf{f}$. Show that, if $\mathbf{y}_1$ and $\mathbf{y}_2$ are two solutions, then their difference is an eigenvector of $\mathbf{A}$ with eigenvalue 0.

We know that if the rank of $\mathbf{A}$ is n, then $\mathbf{A}$ is invertible, its determinant (equal to the product of the eigenvalues) is nonzero, and the solution $\mathbf{y}$ exists and is unique. Suppose now that the rank of $\mathbf{A}$ is $n-1$, so that the null space of $\mathbf{A}$ has dimension 1 and precisely one eigenvalue of $\mathbf{A}$ is zero. That is, there are vectors $\mathbf{v}$ and $\mathbf{w}$, unique up to multiplication by a scalar, such that

If $\mathbf{A}$ is symmetric, then $\mathbf{v} = \mathbf{w}$.

$$\mathbf{Av} = \mathbf{0}, \qquad \mathbf{w}^\top \mathbf{A} = \mathbf{0}^\top;$$

they are the right and left eigenvectors of $\mathbf{A}$ with eigenvalue 0. Put another way, the corresponding homogeneous system $\mathbf{Ay} = \mathbf{0}$ has the nontrivial solution $c\mathbf{v}$ for any scalar c.

Premultiply $\mathbf{Ay} = \mathbf{f}$ by $\mathbf{w}^\top$ to show that

- **Either** $\mathbf{w}^\top \mathbf{f} = 0$, in which case the solution exists but is only unique up to addition of scalar multiples of $\mathbf{v}$;
- **Or** $\mathbf{w}^\top \mathbf{f} \neq 0$, in which case no solution exists at all.

Illustrate by finding the value of f_2 for which the equations

$$\begin{pmatrix} 1 & -1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ f_2 \end{pmatrix}$$

have any solution at all; interpret geometrically.

This result is known as the *Fredholm Alternative*. It applied, *mutatis mutandis*, to two-point boundary value problems. For example, consider

$$\mathcal{L}_x y = \frac{d^2 y}{dx^2} + \alpha^2 y = f(x), \quad 0 < x < 1, \qquad y(0) = y(1) = 0 \qquad (9.17)$$

(the analogue of $\mathbf{Ay} = \mathbf{f}$). Show that the corresponding homogeneous problem $\mathcal{L}_x y = 0$ has only the trivial solution $y = 0$ unless $\alpha = m\pi$ for integral m (the analogue of $\mathbf{A}$ having zero for an eigenvalue). Find the corresponding eigenfunctions (analogous to $\mathbf{v}$ and $\mathbf{w}$, here equal as $\mathcal{L}_x$ is self-adjoint). Suppose that $\alpha = \pi$. Multiply (9.17) by the corresponding eigenfunction and integrate by parts to show that there is only a solution to (9.17) if

$$\int_0^1 f(x) \sin \pi x \, dx = 0,$$

the analogue of $\mathbf{w}^\top \mathbf{f} = 0$. Generalise to the case of any (not necessarily self-adjoint) second order differential operator.

Of course, this is not a coincidence. One could take a two-point boundary value problem and discretise it using finite difference approximations to the derivatives; the result would be a set of linear equations whose solvability or otherwise should, as $n \to \infty$, be the same as that of the original continuous problem.

7. **Matrix inversion.** In this question, we develop the matrix analogue of the calculation of Section 9.6.1 involving the Green's function for a two-point boundary value problem for an ordinary differential equation. For clarity, we use the summation convention (see page 22) throughout.

   Suppose that the matrix equation $\mathbf{A}\mathbf{y} = \mathbf{f}$ (in which $\mathbf{A}$ is not necessarily symmetric) is written in component form as

   $$A_{ij}y_j = f_i \qquad \text{(identify this with } \mathcal{L}_x y = f\text{)}.$$

   Let the inverse matrix $\mathbf{A}^{-1}$ have components $(A^{-1})_{ij} = G_{ij}$, so that from $\mathbf{y} = \mathbf{A}^{-1}\mathbf{f}$ we have

   $$y_i = G_{ij}f_j \qquad \text{(identify with } y(x) = \int_0^1 G(x,\xi)f(\xi)\,d\xi\text{)}.$$

   Let δ_{ij} be the Kronecker delta, the discrete analogue of the delta function. Show that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ and $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ are written

   That is, $\delta_{ij} = 0$ if $i \neq j$, $\delta_{ij} = 1$ if $i = j$. What is δ_{ii}?

   $$G_{ij}A_{jk} = \delta_{ik} \qquad \text{(identify with } \mathcal{L}_x G = \delta(x-\xi)\text{)},$$
   $$A_{ij}G_{jk} = \delta_{ik}.$$

   Take the transpose of the last equation to identify it with $\mathcal{L}_\xi^* G = \delta(\xi - x)$. Lastly, take the dot product with the vector (y_k) to show that

   Note that, just as $\delta(x-\xi) = \delta(\xi - x)$, so $\delta_{ij} = \delta_{ji}$.

   $$0 = A_{ij}G_{jk}y_k - G_{ij}A_{jk}y_k = y_i - G_{ij}f_j;$$

   identify this with the calculation involving $\int y\mathcal{L}^*G - G\mathcal{L}y$.

8. **The fundamental solution of the heat equation.** Show that the heat equation

   $$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

   has similarity solutions of the form $u(x,t) = t^\alpha f(x/\sqrt{t})$ for all α and find the ordinary differential equation satisfied by f. Show that

   $$\int_{-\infty}^{\infty} u(x,t)\,dx$$

   is independent of t when $\alpha = -\frac{1}{2}$, use the result of Exercise 1 of the next chapter to show that in this case $u(x,0) \propto \delta(x)$, and hence find the fundamental solution of the heat equation.

9. **Brownian Motion.** A particle performs the standard drunkard's random walk on the real line, in which in timestep i, of length δt, it moves by $X_i = \pm\delta x$ with equal probability $\frac{1}{2}$. It starts from the origin and the increments are independent. Define

   $$W_n = \sum_1^n X_i.$$

   Show that $\mathbb{E}[W_n] = 0$, $\mathrm{var}[W_n] = n\delta x^2/\delta t$. Now let $n \to \infty$ with $n\delta t = t$ fixed and $\delta x = \sqrt{\delta t}$. Call the limiting process (assuming it exists!) W_t. Use the Central Limit Theorem to show that

   This scaling is the simplest that allows proper time variation yet keeps the variance of the limit finite.

- For each $t > 0$, W_t has the Normal distribution with mean zero and variance t.

Show also that

- $W_0 = 0$.
- For each $0 \leq s < t$, $W_t - W_s$ is independent of W_s.

The resulting stochastic process is called *Brownian Motion* and it is central to modern analysis of financial markets. Give a heuristic argument that the sample paths (realisations, graphs of the random walk) are continuous in t but not differentiable.

Now let $p(x,t)$ be the probability density function of many such random walks (as a function of position x for each t). Go back to the discrete random walk and, as in the discussion of Poisson processes in Chapter 7, condition on one step to write down

$$p(x, t + \delta t) = \tfrac{1}{2}\left(p(x - \delta x, t) + p(x + \delta x, t)\right).$$

Expand the right-hand side in a Taylor series and use $\delta x = \sqrt{\delta t}$ to show that

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2}.$$

Explain why $p(x, 0) = \delta(x)$ and hence find $p(x,t)$ (see Exercise 8).

The $\frac{1}{2}$ in front of the second derivative in the heat equation is a diagnostic feature for a probabilist as distinct from a 'physical' applied mathematician.

10. **Regular part of the Green's function for the Laplacian.**

A horizontal membrane stretched over a region D is stretched to tension T and a normal force f per unit area is then applied. The displacement (which, like the force, is measured vertically upwards) is zero on the boundary ∂D. Show that the displacement $u(x,y)$ of the membrane satisfies

$$T\nabla^2 u = -f \quad \text{in} \quad D, \qquad u = 0 \quad \text{on} \quad \partial D.$$

Do not worry about the infinite displacement!

Suppose that $f(x,y) = \delta(\mathbf{x} - \xi)$ where $\mathbf{x} = (x,y)$ and $\xi = (\xi, \eta)$ is known. How is $u(x, y; \xi, \eta)$ related to the Green's function for the Laplacian in D?

Can you find a dimensionless parameter to quantify this modelling assumption?

Now suppose that the force is due to a very heavy ball which is free to roll around, and that it is in equilibrium at ξ. Suppose that we model its effect by that of a point force. Take a small square centred on $\mathbf{x} = \xi$ and resolve forces in the $x-$ and y–directions to show that the gradient of the regular part of G vanishes at $\mathbf{x} = \xi$. Do you think there is always just one such equilibrium point? If not, when might you have one and when more than one?

"What's the word beginning with D which means distribution? Oh, distribution."

Chapter 10

Theory of distributions

The time has come to look at the theoretical underpinning of the delta function and its relatives. You may choose not to read this chapter, but I promise that it is not complex or technically demanding. We begin with a few (as few as we can get away with) necessary definitions.

10.1 Test functions

We noted earlier that the proper way to approach $\delta(x)$ was by thinking of the result of multiplying a suitably smooth function $\phi(x)$ and integrating to get $\phi(0)$. The first step in setting up a robust framework is to define a class of 'suitably smooth' functions, called *test functions*. We say that $\phi(x)$ is a test function if

- $\phi(x)$ is a C^∞ function. That is, it has derivatives of all orders at each point $x \in \mathbb{R}$.
- $\phi(x)$ has *compact support*: that is, it vanishes outside some interval (a, b). (The support is the closure of the set where ϕ is non-zero.)

Because every derivative of ϕ is itself differentiable, the derivatives are all continuous and bounded.

The first of these requirements makes these functions very smooth indeed.[1] This high degree of regularity guarantees a trouble-free ride for the theory, the reason being that if $\phi(x)$ is a test function, then so are all its derivatives.

We should note that test functions do exist (and that we never need to know much more than this: they are a background tool). The easiest way to see this is to construct one, using the famous example of a function which has derivatives of all orders, and hence a Taylor series, at $x = 0$, but which is not equal to the sum of its Taylor series. That is, look at

See Exercise 5 on page 148. Perhaps those pathological real-analysis examples were more useful than I thought.

$$\Phi(x) = \begin{cases} 0 & x \le 0, \\ e^{-1/x} & x > 0, \end{cases}$$

which vanishes for $x \le 0$, is positive for $x > 0$, and is C^∞. The only thing wrong with this function is that it does not have compact support. To fix this

[1] Roughly speaking, only real analytic functions (defined as equal to the sum of a convergent Taylor series) are smoother, and they can never be test functions because they cannot have compact support (why not?).

up, just multiply by, say, $\Phi(1-x)$:

$$\phi(x) = \Phi(x)\Phi(1-x)$$

is a perfectly good test function with support on the interval $(0,1)$.

We also need a definition of convergence for a sequence of test functions $\{\phi_n(x)\}$. We say that $\phi_n(x) \to 0$ as $n \to \infty$ if

- $\phi_n(x)$ and all its derivatives $\phi_n^{(m)}(x)$ tend to zero, uniformly in both x and m;
- There is an interval (a,b) containing the support of all the ϕ_n.

The first of these is an incredibly strong form of convergence: the ϕ_n have no room to wriggle at all. The second stops them from running away to infinity as n increases.

The only other thing to say about test functions is that we shall denote them by lower case Greek letters, usually ϕ or ψ.

10.2 The action of a test function

Suppose that $f(x)$ is an integrable[2] function (we denote such functions by lower case Roman letters f, g, etc.). We define the *action* of f on a test function $\phi(x)$ by

$$\langle f, \phi \rangle = \int_{-\infty}^{\infty} f(x)\phi(x)\,dx.$$

It's also a bit like an inner product: but note that f and ϕ lie in different spaces.

So, this action is a kind of weighted average of $f(x)$. If we know the action of f on all test functions, we should know all about f itself (a bit like recovering a probability distribution from its moments). The action, regarded as a map from the space of test functions to $\mathbb{R}$, satisfies the usual linearity properties, such as

$$\langle f, a\phi + b\psi \rangle = a\langle f, \phi \rangle + b\langle f, \psi \rangle,$$

for real constants a, b. Also, if $\phi_n(x) \to 0$ in the sense defined above, then $\langle f, \phi_n \rangle \to 0$ as a sequence of real numbers.

10.3 Definition of a distribution

In defining distributions, we use the very mathematical idea of taking things we already know about, here functions, and dropping some of their properties while retaining others in order to obtain something broader or more general. In this way, we see that distributions are indeed 'generalised functions', despite the inexplicable reluctance of some to use the term.

As foreshadowed above, the properties that we want to keep are those to do with the action of a function on a test function; that is, we keep the 'smoothing' idea of averaging while quietly dropping all worries about pointwise definition. We do this is such a way that all the properties of distributions are *consistent* with the corresponding properties of (say) piecewise continuous functions.

Measurable functions would be better, but that requires too much machinery.

[2]We sidestep the question of what we mean by this, exactly. Piecewise continuous will do for now, or locally Lebesgue integrable.

Then, all such functions are subsumed within the larger class of distributions.

The two properties that we keep are those we reached at the end of the previous section: linearity and continuity. We define: *a distribution $\mathcal{D}$ is a continuous linear map from the space of test functions to $\mathbb{R}$, denoted by*

$$\mathcal{D} : \phi \mapsto \langle \mathcal{D}, \phi \rangle \in \mathbb{R}.$$

The result of the map, $\langle \mathcal{D}, \phi \rangle$, is known as the *action* of $\mathcal{D}$ on ϕ. We say that two distributions are equal if their action is the same for all test functions.

The properties of linearity and continuity are as above:

$$\langle \mathcal{D}, a\phi + b\psi \rangle = a\langle \mathcal{D}, \phi \rangle + b\langle \mathcal{D}, \psi \rangle,$$

for real constants a, b, and

$$\text{if} \quad \phi_n(x) \to 0 \quad \text{as} \quad n \to \infty, \quad \text{then} \quad \langle \mathcal{D}, \phi_n \rangle \to 0.$$

Evidently any piecewise continuous function $f(x)$ corresponds to a distribution $\mathcal{D}_f$ with the obvious action $\langle \mathcal{D}_f, \phi \rangle = \langle f, \phi \rangle$. Indeed, we normally don't bother to write $\mathcal{D}_f$, but just use f itself. This is an example of the consistency referred to above.

We shall mostly use the letter style of $\mathcal{D}$, $\mathcal{H}$ to denote distributions, unless they already have a name. The set of test functions is often called [script D, need typeface for this] and the set of distributions is then written [script D prime]. Sometimes we write $\mathcal{D}(x)$ to emphasise the dependence on x; the dependence is of course in the test functions, but it's quite OK, and indeed a good idea, to think of distributions as depending on x as well.

Example: the delta function. There could be no better example than the delta distribution, δ or $\delta(x)$. It is defined as a distribution by its action on a test function $\phi(x)$:

$$\langle \delta, \phi \rangle = \phi(0).$$

We could also have written

$$\langle \delta(x), \phi(x) \rangle = \phi(0).$$

You should check carefully that this action does indeed define a distribution satisfying the properties above. Again, it is OK, and indeed a good idea, to think intuitively of the action of the delta function as

$$\langle \delta, \phi \rangle = \int_{-\infty}^{\infty} \delta(x)\phi(x)\, dx.$$

However, you should always use the formal definition to prove anything about $\delta(x)$ or any other distribution.

10.4 Further properties of distributions

If our distributions are to be useful, we need to give them some more properties. We assume that, if $\mathcal{D}$ and $\mathcal{E}$ are distributions, a is a real constant, $\phi(x)$ is a test function and $\Phi(x)$ is a C^∞ function (not necessarily a test function), then there are new distributions $\mathcal{D} + \mathcal{E}$, $a\mathcal{D}$, $\mathcal{D}(x - a)$ and $\mathcal{D}(ax)$ such that

- $\langle \mathcal{D} + \mathcal{E}, \phi \rangle = \langle \mathcal{D}, \phi \rangle + \langle \mathcal{E}, \phi \rangle$;

- $\langle a\mathcal{D}, \phi\rangle = a\langle \mathcal{D}, \phi\rangle$;
- $\langle \mathcal{D}(x-a), \phi(x)\rangle = \langle \mathcal{D}(x), \phi(x+a)\rangle$;
- $\langle \mathcal{D}(ax), \phi(x)\rangle = \dfrac{1}{|a|}\langle \mathcal{D}, \phi(x/a)\rangle$.
- $\langle \Phi(x)\mathcal{D}(x), \phi(x)\rangle = \langle \mathcal{D}(x), \Phi(x)\phi(x)\rangle$.

Note how we slip in and o stating the x–dependence explicitly.

Watch out for the modulus sign.

Note that $\Phi(x)\phi(x)$ is a test function even if $\Phi(x)$ is not.

You should check all these when $\mathcal{D}$ corresponds to an integrable function $f(x)$; it will give you intuition as to why the definitions have been made in this way. Note in particular that from the third definition, we have

$$\begin{aligned}\langle \delta(x-a), \phi(x)\rangle &= \langle \delta(x), \phi(x+a)\rangle \\ &= \phi(a).\end{aligned}$$

As expected, we have recovered the sifting property of the delta function.

10.5 The derivative of a distribution

One more idea completes our introduction to the distributional framework. If we want to make sense of ideas such as $d^2y/dx^2 = \delta(x-\xi)$, we had better have a definition of the derivative of a distribution. Again, consistency with ordinary functions provides the way in. If $f(x)$ is differentiable, with derivative $f'(x)$, then integrating by parts we calculate the action of $f'(x)$:

What properties of test functions do we use here?

$$\begin{aligned}\langle f'(x), \phi(x)\rangle &= \int_{-\infty}^{\infty} f'(x)\phi(x)\,dx \\ &= f(x)\phi(x)\big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(x)\phi'(x)\,dx \\ &= -\langle f(x), \phi'(x)\rangle.\end{aligned}$$

Notice how the compact support of the test function takes care of $f(x)\phi(x)\big|_{-\infty}^{\infty}$.

We define the *derivative $\mathcal{D}'$ of a distribution* $\mathcal{D}$ in terms of its action by

$$\langle \mathcal{D}', \phi\rangle = -\langle \mathcal{D}, \phi'\rangle$$

(note that $\phi'(x)$ is also a test function). The point is that although we do not know about $\mathcal{D}'$, we do know about $\mathcal{D}$, so we can calculate $\langle \mathcal{D}, \phi'\rangle$ and hence $\langle \mathcal{D}', \phi\rangle$.

For example, let us show that $\mathcal{H}'(x) = \delta(x)$. We define the Heaviside function $\mathcal{H}(x)$ by its action:

$$\langle \mathcal{H}, \phi\rangle = \int_0^{\infty} \phi(x)\,dx;$$

this is entirely consistent with our view of $\mathcal{H}(x)$ as the unit step function since

$$\mathcal{H}(x)\phi(x) = \begin{cases} 0 & x \le 0, \\ \phi(x) & x > 0.\end{cases}$$

Now consider the action of $\mathcal{H}'(x)$:

$$\begin{aligned}\langle \mathcal{H}'(x), \phi(x)\rangle &= -\langle \mathcal{H}(x), \phi'(x)\rangle \\ &= -\int_0^\infty \phi'(x)\,dx \\ &= \phi(0) \\ &= \langle \delta(x), \phi(x)\rangle.\end{aligned}$$

Since their actions are identical, we conclude that $\mathcal{H}'(x) = \delta(x)$ (as distributions).

We can extend this definition recursively, to give action of the m–th derivative of $\mathcal{D}$ as

$$\langle \mathcal{D}^{(m)}(x), \phi(x)\rangle = (-1)^m \langle \mathcal{D}, \phi^{(m)}(x)\rangle$$

for $m = 1, 2, 3, \ldots$. Because every derivative of a test function is a test function, we see that distributions have derivatives of all orders too, an example of the technical simplicity of this theory.

10.6 Extensions of the theory of distributions

We conclude with an overview (a glimpse, really) of two vital extensions of the theory just outlined.

10.6.1 More variables

It is a very straightforward business to define distributions in the context of functions of several variables. We first define test functions to have compact support and to be C^∞ in all their arguments. Then, we define distributions as continuous linear maps from this space of test functions to $\mathbb{R}$. In particular, the delta function satisfies

$$\langle \delta(\mathbf{x}), \phi(\mathbf{x})\rangle = \phi(\mathbf{0}).$$

The partial derivatives of a distribution $\mathcal{D}(\mathbf{x})$ are defined recursively using the formula

$$\langle \frac{\partial \mathcal{D}}{\partial x_i}, \phi\rangle = -\langle \mathcal{D}, \frac{\partial \phi}{\partial x_i}\rangle.$$

Again, D has derivatives of all orders, and because the mixed partial derivatives of the test functions are always equal, so are the mixed partials of $\mathcal{D}$. Thus, identities such as $\nabla \wedge \nabla \mathcal{D} \equiv \mathbf{0}$ are automatically true for distributions. The whole theory is splendidly robust, and we need have no qualms at all about writing down equations such as $\nabla^2 G = \delta(\mathbf{x} - \xi)$.

10.6.2 Fourier transforms

Space does not permit a full description of the theory of Fourier transforms of distributions in one or more variables. Nevertheless, here is an outline. For technical reasons, we use a slightly different class of test functions, which are still C^∞ but no longer have compact support. Instead, they and all their derivatives decay faster than any power of x as $x \to \pm\infty$. In principle, this defines a different class of distributions (known as *tempered distributions* — the compact support ones are *Schwartz*[3] *distributions*), but we won't notice the difference.

See Exercise 12 on page 150 to see why this would not be so for compact support test functions.

The new test functions can be shown to have the nice property that if $\phi(x)$ is a test function then so is its Fourier transform; this is why we use this class of test functions. We write the transform as[4]

$$\hat{\phi}(k) = \int_{-\infty}^{\infty} \phi(x) e^{ikx}\,dx.$$

This is just the usual Fourier transform; we write the inverse transform as

$$\check{\psi}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \psi(k) e^{-ikx}\,dk,$$

and we recall the standard results

$$\widehat{\frac{d\phi}{dx}} = -ik\hat{\phi}, \qquad \widehat{x\phi} = -i\frac{d\hat{\phi}}{dk},$$

the first of which is established by integration by parts and the second by differentiation under the integral sign.

Let's see what the action of the Fourier transform of an ordinary function is on a test function. The Fourier transform of a tempered distribution $\mathcal{D}$ is then defined to be consistent with this; as ever, we look at its action and transfer the work to the test function. A formal calculation gives

You might want to write this out, swapping the dummy variables x and k in the second line.

$$\begin{aligned}\langle \hat{f}, \phi\rangle &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(x)e^{ikx}dx\right)\phi(k)\,dk \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} \phi(k)e^{ikx}dk\right) f(x)\,dx \\ &= \langle f, \hat{\phi}\rangle.\end{aligned}$$

We therefore define

$$\langle \hat{\mathcal{D}}, \phi\rangle = \langle \mathcal{D}, \hat{\phi}\rangle,$$

and similarly we define the inverse by

Check this one for an ordinary function.

$$\langle \check{\mathcal{D}}, \phi\rangle = \langle \mathcal{D}, \check{\phi}\rangle.$$

Notice how important it is that $\hat{\phi}$ should be a test function too. If it were not, we could not be confident that some of these actions are defined at all. Notice too that the factors of 2π don't appear here: they are all hidden in the inverse of ϕ.

Using these deceptively simple formulas, we can prove that the Fourier transform of the derivative $\mathcal{D}' = d\mathcal{D}/dx$ is $-ik\hat{\mathcal{D}}$:

Line 1 is the definition of the transform; line 2 is the distributional derivative; line 3 is a standard identity; in line 4 we swap x for k and shift it to the first argument of the action.

[3]Rather to my surprise, Schwartz, who invented the theory in 1944, died as recently as the time of writing. A fearless opponent of political and military oppression and a great mathematician, his support was the interval $(1915, 2002)$.

[4]Beware: notations differ, both in the signs in the exponent and in the placement of the 2π which can appear in the exponent, or symmetrically as $1/\sqrt{2\pi}$ multiplying both the transform and its inverse. The definition here is probably the commonest among applied mathematicians.

$$\begin{aligned}\langle \widehat{\mathcal{D}'}, \phi\rangle &= \langle \mathcal{D}', \hat{\phi}\rangle \\ &= -\langle \mathcal{D}, d\hat{\phi}/dk\rangle \\ &= -\langle \mathcal{D}, \widehat{ix\phi}\rangle \\ &= \langle -ik\hat{\mathcal{D}}, \phi\rangle\end{aligned}$$

as required. It is an exercise for you to prove that the transform of $x\mathcal{D}$ is $-id\hat{\mathcal{D}}/dk$.

We end this section by finding the transforms of $\delta(x)$ and 1. (Yes, 1 has a Fourier transform in this theory; so do x, $|x|$, etc.).[5] The transform of $\delta(x)$ must surely be 1: informally,

Very informally, because e^{ikx} is not a test function, although one could 'truncate' it by multiplying by a test function which is small for $|x| > R$ and taking $R \to \infty$.

$$\int_{-\infty}^{\infty} \delta(x)e^{ikx}dx = e^{ik0} = 1.$$

Formally,

$$\begin{aligned}\langle \hat{\delta}, \phi\rangle &= \langle \delta, \hat{\phi}\rangle \\ &= \hat{\phi}(0) \\ &= \int_{-\infty}^{\infty} \phi(x)\,dx \\ &= \langle 1, \phi\rangle\end{aligned}$$

so we do indeed have

$$\hat{\delta}(k) = 1.$$

For the inverse, we have

$$\begin{aligned}\check{\delta} &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \delta(k)e^{-ikx}dk \\ &= \frac{1}{2\pi},\end{aligned}$$

so taking the transform of both sides, remembering that $(\check{\delta})\hat{} = \delta$, we get

$$\hat{1}(k) = 2\pi\delta(k).$$

You may like to show this from the formal definitions alone, using the fact that for test functions $\langle 1, \check{\phi}\rangle = 2\pi\langle 1, \hat{\phi}\rangle$.

The heat equation

We conclude with an example: it's one we have seen before but we do it in a different way. Consider the initial value problem for the heat equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad -\infty < x < \infty, \quad t > 0,$$
$$u(x, 0) = \delta(x).$$

[5] The transforms of sums of delta functions are the characteristic functions of discrete random variables.

This time we'll take a Fourier transform in x. The equation for $\hat{u}(k,t)$ is

$$\frac{\partial \hat{u}}{\partial t} = -k^2 \hat{u}, \qquad -\infty < k < \infty, \quad t > 0,$$
$$u(x,0) = \hat{\delta}(k) = 1.$$

The solution is

$$\hat{u}(k,t) = e^{-k^2 t},$$

and inversion by any of a number of methods (see Exercise 14 on page 150) yields the answer

$$u(x,t) = \frac{1}{2\sqrt{\pi t}} e^{-x^2/4t}.$$

Sources and further reading

The theory of distributions in its modern form was developed by Schwartz [53]; the epsilonological approach is exemplified by Lighthill's book [38]. My description of the modern theory is heavily based on the very approachable book by Richards & Youn [50] (my main quibble with that book is the intrusive 2π in the exponent of the Fourier Transform).

If the idea of extending our definition of functions to make sense of the result

$$\int_{-1}^{1} \frac{dx}{x^2} = -2$$

appeals to you then you should definitely read [50].

10.7 Exercises

1. **Constructing delta functions from continuous functions I: by the Lebesgue Dominated Convergence Theorem.** Suppose that $f(x) \in L^1$ is continuous and $\int_{-\infty}^{\infty} f(x)\,dx = 1$. Take a test function $\phi(x)$ and show that, as $\epsilon \to 0$,

$$I_\epsilon = \int_{-\infty}^{\infty} \frac{1}{\epsilon} f\left(\frac{x}{\epsilon}\right) \phi(x)\,dx \to \phi(0),$$

as follows. First show that

$$I_\epsilon = \int_{-\infty}^{\infty} f(s)\phi(\epsilon s)\,ds.$$

Next, show that

$$|f(s)\phi(\epsilon s)| < M\,|f(s)|$$

for some constant $M > 0$, that if $f(s) \in L^1$ then $f(s)\phi(\epsilon s) \in L^1$, and that, for each s, $f(s)\phi(\epsilon s) \to f(s)\phi(0)$ as $\epsilon \to 0$. Deduce from the Dominated Convergence Theorem that you can justify interchanging the limit and the integral:

$$\lim_{\epsilon \to 0} \int_{-\infty}^{\infty} f(s)\phi(\epsilon s)\,ds = \phi(0).$$

2. **Constructing delta functions from continuous functions II: by splitting the range of integration.** If you don't know about Lebesgue integration, derive the following slightly weaker result. Suppose that $f(x)$ is any continuous function with

$$\int_{-\infty}^{\infty} f(x)\,dx = 1, \qquad \int_{-\infty}^{\infty} |f(x)|\,dx < \infty, \quad \int_{-\infty}^{\infty} |xf(x)|\,dx < \infty.$$

Take a test function $\phi(x)$ and show that, as $\epsilon \to 0$,

$$I_\epsilon = \int_{-\infty}^{\infty} \frac{1}{\epsilon} f\left(\frac{x}{\epsilon}\right) \phi(x)\, dx \to \phi(0),$$

as follows. First write $x = \epsilon s$ in the integral and split the range of integration up to get

$$I_\epsilon = \int_{-\infty}^{-1/\sqrt{\epsilon}} + \int_{-1/\sqrt{\epsilon}}^{1/\sqrt{\epsilon}} + \int_{1/\sqrt{\epsilon}}^{\infty} f(s)\phi(\epsilon s)\, ds.$$

Noting that $|\phi(x)|$ is bounded and using the idea that if $|h| < c$, $|\int gh| \le \int |gh| \le c \int |g|$, show that the first and third integrals tend to zero as $\epsilon \to 0$ because f is integrable. For the inner integral, expand $\phi(\epsilon s)$ using Taylor's theorem to get

$$\int_{-1/\sqrt{\epsilon}}^{1/\sqrt{\epsilon}} f(s)\Big(\phi(0) + \epsilon s \phi'\left(\xi(s)\right)\Big)\, ds$$

where $\xi(s)$ lies between 0 and s. Show that the first term in this integral tends to what we want and, noting that $|\phi'|$ is bounded, that the second tends to zero as $\epsilon \to 0$.

3. **Delta sequences.** Consider the functions

$$f_n(x) = \frac{n}{\pi\left(1 + n^2 x^2\right)} \quad \text{and} \quad g_n(x) = \frac{\sin nx}{\pi x}.$$

Sketch them and show that $f_n(x)$ tends to $\delta(x)$ as $n \to \infty$, in the distributional sense, so for any test function $\phi(x)$,

$$\langle f_n, \phi \rangle \to \phi(0)$$

as $n \to \infty$. Use the method of Exercise 3, but be careful when estimating the integrals as $f_n(x)$ does not satisfy all the conditions of that question. Repeat for $g_n(x)$.

This might suggest that if $\delta_n(x)$ is a sequence tending to $\delta(x)$ then $\delta_n(0) \to \infty$. Construct a piecewise constant example to show that this is false.

4. **Discrete and continuous sources.** Suppose that $u(\mathbf{x})$ is a classical solution of $\nabla^2 u = f(\mathbf{x})$ in $\mathbb{R}^n$, $n \ge 2$, where $f(\mathbf{x})$ is smooth and has compact support, and appropriate growth conditions at infinity are assumed. Let $\phi(\mathbf{x})$ be a test function. Use Green's theorem in the form

$$\int_D v\nabla^2 w - w\nabla^2 v = \int_{\partial D} v\frac{\partial w}{\partial n} - w\frac{\partial v}{\partial n}, \qquad (10.1)$$

where D is a region containing the support of f, to show that

$$\langle u, \nabla^2\phi \rangle = \langle f, \phi \rangle.$$

Now suppose that we approximate $f(\mathbf{x})$ by delta functions, defining the sequence of distributions

$$\mathcal{F}_n = \sum_1^n \alpha_n \delta(\mathbf{x} - \mathbf{x}_n)$$

and taking the limit $n \to \infty$ in such a way that all the weights α_n tend to zero but

$$\langle \mathcal{F}_n, \phi \rangle \to \langle f, \phi \rangle$$

for all test functions ϕ. Also let u_n be the solution of $\nabla^2 u_n = \mathcal{F}_n$. Show that

$$\langle u_n, \nabla^2 \phi \rangle = \langle \mathcal{F}_n, \phi \rangle,$$

and deduce that $u_n \to u$ (as a distribution). Interpret this result in terms of the gravitational potential due to a finite mass distribution (or in electrostatic terms).

5. **The function $e^{-1/x}$.** Consider

$$\Phi(x) = \begin{cases} 0 & x \le 0, \\ e^{-1/x} & x > 0. \end{cases}$$

Remember that $X^n e^{-X} \to 0$ as $X \to \infty$ for all N.

Show that for $x > 0$ its n–th derivative $\Phi^{(n)}(x)$ is a polynomial in $1/x$ times $e^{-1/x}$, and hence that $\lim_{x \downarrow 0} \Phi^{(n)}(x) = 0$. Deduce that the Taylor coefficients of this function are all zero. Does the complex function $e^{-1/z}$ have a Taylor series at $z = 0$? If not, what does it have?

6. **The distribution $\delta(ax)$.** Show from the interpretation as an integral that

$$\delta(ax) = \frac{1}{|a|}\delta(x).$$

7. **Derivatives of the delta function.** Show carefully, using the definition of a distributional derivative, that, if $\Psi(x)$ is a smooth (C^∞) function and $\mathcal{D}$ a distribution, then $(\mathcal{D}\Psi)' = \mathcal{D}'\Psi + \mathcal{D}\Psi'$ (Leibniz). Deduce that

$$x^n \delta^{(m)}(x) = \begin{cases} 0 & m < n, \\ \dfrac{(-1)^n m!}{(m-n)!}\delta^{(m-n)}(x) & m \ge n \end{cases}$$

($\delta^{(m)} = m$th derivative). What is $x\delta(x)$? Show that $\delta(x) = -x\delta'(x)$.

8. **Convergence of series of distributions.** We say that a sequence $\{\mathcal{D}_n\}$ of distributions converges to $\mathcal{D}$ if

$$\langle \mathcal{D}_n, \phi \rangle \to \langle \mathcal{D}, \phi \rangle$$

for all test functions $\phi(x)$. This is an incredibly tolerant form of convergence, because our definition of convergence of a sequence of test functions is so stringent: show that if $\mathcal{D}_n \to \mathcal{D}$, then the same applies to all the derivatives, so that $\mathcal{D}_n^{(m)} \to \mathcal{D}^{(m)}$. Show also that you can differentiate a convergent series of distributions term by term.

Find the Fourier series of the sawtooth function

$$f(x) = \begin{cases} \dfrac{1}{2} - \dfrac{x}{2\pi} & 0 < x < \pi, \\ -\dfrac{1}{2} - \dfrac{x}{2\pi} & -\pi < x < 0. \end{cases}$$

Now differentiate both sides, noting that the jumps of 1 in $f(x)$ at $x = 2n\pi$ contribute delta functions $\delta(x - 2n\pi)$, to establish the result

$$\sum_{n=-\infty}^{\infty} \delta(x - 2n\pi) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \cos mx,$$

an identity which makes classical nonsense but perfect distributional sense.

Note: it can be shown that every distribution $\mathcal{D}$ is the distributional limit of a sequence of test functions (which are C^∞). So the set of distributions is not unboundedly diverse.

9. **Derivative of a distribution.** Let $\mathcal{D}(x)$ be a distribution. Show (by considering its action) that

Remember that $\langle \mathcal{D}(x+h), \phi(x) \rangle = \langle \mathcal{D}(x), \phi(x-h) \rangle$.

$$\mathcal{D}'(x) = \lim_{h \to 0} \frac{\mathcal{D}(x+h) - \mathcal{D}(x)}{h}.$$

Use the right-hand side of this equation to confirm (again by considering the action) that $\delta(x) = \mathcal{H}'(x)$.

10. **Dipoles.** The derivative of the delta function, $\delta'(x)$, is known as a (one-dimensional) *dipole*, which you can think of as the limit as $\epsilon \to 0$ of a positive delta function at $x = \epsilon$ and a negative one at $x = 0$ (see Exercise 9). What is its action on a test function $\psi(x)$?

In hydrodynamics, a mass dipole aligned with the x–axis is obtained as the limit of point (in two dimensions, line) sources of strength q at $(\pm\epsilon, 0, 0)$, keeping the product $m = 2\epsilon q$ constant as $\epsilon \to 0$. Explain why the velocity potential for inviscid irrotational flow with a point source at the origin satisfies

Notation clash! ϕ is not a test function here.

$$\nabla^2 \phi = q\delta(\mathbf{x})$$

and deduce that if there is a dipole as above at the origin, the potential satisfies

$$\nabla^2 \phi = m \frac{\partial \delta}{\partial x}.$$

(The right-hand side may also be written as $\delta'(x)\delta(y)\delta(z)$ in three dimensions, or $\delta'(x)\delta(y)$ in two.) Hence calculate the potential for a dipole and sketch the streamlines in two dimensions. Show that the potential $U(r\cos\theta + a^2 \cos\theta / r)$ for flow past a cylinder consists of a uniform flow plus a dipole.

Interpret these results in terms of electric charges. (Whereas point charges generate electric fields, because there are no magnetic monopoles, the basic generator of magnetic fields is the infinitesimal current loop, giving a dipole field with lines of force similar to those of a bar magnet. Higher-order derivatives, called multipoles, are important in, for example, the analysis of the far field of radio transmitters.)

11. **Vector distributions.** [NEED BOLD CALLiGRAPHIC FONT HERE for the vector distributions, and bold for the vector ϕ in (b).] Develop the following two ways of defining vector-valued distributions in $\mathbb{R}^3$. In both cases aim to establish the identities $\nabla \cdot \nabla \wedge \mathcal{D} \equiv 0$, $\nabla \wedge \nabla \mathcal{D} \equiv \mathbf{0}$ for vector

and scalar distributions $\mathcal{D}$ and $\mathcal{D}$ respectively. You will need to establish variants of Green's theorem in order to define the action of the operators div and curl by integration by parts.

(a) Take scalar test functions $\phi(\mathbf{x})$ and define their action on a vector function $\mathbf{v}(\mathbf{x})$ as the vector

$$\langle \mathbf{v}, \phi \rangle = \int_{\mathbb{R}^{\nvDash}} \mathbf{v}(\mathbf{x})\phi(\mathbf{x})\, d\mathbf{x}.$$

Then define a vector-valued distribution $\mathcal{D}$ as a continuous linear map from the space of test functions to $\mathbb{R}^3$ consistently with this action.

(b) Use vector test functions ϕ and the action

$$\langle \mathbf{v}, \phi \rangle = \int_{\mathbb{R}^3} \mathbf{v} \cdot \phi\, d\mathbf{x}.$$

12. **Open support test functions.** To get an idea why compact support test functions do not lead to a good theory for the distributional Fourier transform, work out the Fourier transform of

It's not hard: just integrate.

$$f(x) = \begin{cases} 1 & -1 < x < 1, \\ 0 & \text{otherwise,} \end{cases}$$

and observe that, unlike $f(x)$, $\hat{f}(k)$ does not have compact support. (Although $f(x)$ is not a test function, a similar result would hold if it were.) Now look at the definition of the Fourier transform to see why compact support test functions are not useful here.

13. **Commutation of the Fourier transform and its inverse.** Show directly from the definitions that if $\mathcal{D}$ is a distribution with Fourier transform $\hat{\mathcal{D}}$, then

$$(\hat{\mathcal{D}})\check{} = (\check{\mathcal{D}})\hat{} = \mathcal{D},$$

assuming that this holds for test functions.

14. **The inverse of e^{-k^2t}.** Find the inverse of $\hat{u}(k,t) = e^{-k^2t}$ in the following two ways.

(a) Write down the inversion integral and complete the square in the exponent; then, thinking of the integral as a contour integral in the complex k–plane, move the integration contour to the line $\operatorname{Im} k = -x/2t$ (check that the endpoint contributions vanish) and evaluate a standard real integral, using the result $\int_{-\infty}^{\infty} e^{-s^2} ds = \sqrt{\pi}$.

(b) Show that $\partial\hat{u}/\partial k = -2kt\hat{u}$, then use the standard identities for the transforms of $\partial u/\partial x$ and xu to obtain a similar ordinary differential equation for u; solve this and choose the 'constant of integration' (which is actually a function of t) to set $\int_{-\infty}^{\infty} u(x,t)\, dx = 1$ for all t (which is easy to show from the original problem).

15. **The pseudofunction $1/x$.** Obviously, $1/x$ is defined for $x \neq 0$ as an ordinary function. Its definition for all $x \in \mathbb{R}$ is achieved by defining its action on a test function $\phi(x)$:

$$\langle 1/x, \phi(x) \rangle = \lim_{\epsilon \to 0} \langle 1/x, \phi(x) \rangle_\epsilon,$$

where

$$\langle 1/x, \phi(x)\rangle_\epsilon = \int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\infty} \frac{\phi(x)}{x}\,dx;$$

the limiting integral, denoted by

$$⨍_{-\infty}^{\infty} \frac{\phi(x)}{x}\,dx,$$

is called a *Cauchy principal value integral.* Note that the small interval $(-\epsilon, \epsilon)$ that we remove before integrating and taking the limit $\epsilon \to 0$ is symmetric about $x = 0$.

Show that the limit exists for all test functions $\phi(x)$. Show directly from the distributional definitions that

$$\frac{1}{x} = \frac{d}{dx}\log|x|;$$

that is, show that

$$\langle d\log|x|/dx, \phi(x)\rangle = -\langle 1/x, d\phi/dx\rangle$$

by considering the same statement with $\langle\cdot,\cdot\rangle$ replaced by $\langle\cdot,\cdot\rangle_\epsilon$ and letting $\epsilon \to 0$.

Show also (for future reference) that

$$⨍_{-1}^{1} \frac{dx}{x} = 0. \tag{10.2}$$

16. **The Fourier transform of $\mathcal{H}(x)$.** A distribution $\mathcal{D}(x)$ is called *odd* if the result of its action gives $\mathcal{D}(-x) = -\mathcal{D}(x)$, and *even* if $\mathcal{D}(-x) = \mathcal{D}(x)$. Show that $\delta(x)$ is even. Show also that $x\delta(x) = 0$. If $\mathcal{H}(x)$ is the Heaviside function, show that $\widetilde{\mathcal{H}}(x) = \mathcal{H}(x) - \frac{1}{2}$ is odd.

    Show that the Fourier transform of a real-valued odd function is a purely imaginary odd function of k, and deduce (or assert) that the same applies to distributions.

    Since $\mathcal{H}'(x) = \delta(x)$, taking the Fourier transform gives

    $$-ik\widehat{\mathcal{H}} = \hat{\delta} = 1.$$

    However, before dividing through by k, we must realise that we can add $ck\delta(k)$ $(= 0)$ to the right-hand side, where c is an as yet unspecified complex constant. By considering instead the transform of the odd distribution $\widetilde{\mathcal{H}}(x)$, and recalling that $\hat{1} = 2\pi\delta(k)$, show that

    $$\widehat{\mathcal{H}}(k) = -\frac{1}{ik} + \pi\delta(k).$$

    Note that $\widehat{\mathcal{H}}$ requires the definition of $1/k$ introduced in the previous exercise.

17. **The Fourier transform of $\mathcal{H}(x)$ again.** Here are two more ways of calculating $\widehat{\mathcal{H}}(k)$.

(i) Consider

$$\int_0^{1/\epsilon} e^{ikx} dx = -\frac{1}{ik} + \frac{e^{ik/\epsilon}}{ik}.$$

The first part is already in the answer, so the second part must tend to $\pi\delta(k)$ as $\epsilon \to 0$. Write

$$\frac{e^{ik/\epsilon}}{ik} = \frac{\sin(k/\epsilon)}{k} - i\frac{\cos(k/\epsilon)}{k}$$

and note that the real part has been shown (in Exercise 3) to give $\pi\delta(k)$. It remains to show that the principal value integral

$$⨍_{-\infty}^{\infty} \frac{\cos(k/\epsilon)\phi(k)}{k}\, dk \to 0$$

as $\epsilon \to 0$ for any test function ϕ. Write ϕ as the sum of its even and odd parts and note that we need only consider the odd part of ϕ as the integral of the even part vanishes by symmetry. Now proceed as in earlier exercises, splitting the range of integration into $|k| > \sqrt{\epsilon}$ and $|k| < \sqrt{\epsilon}$ and dealing with each separately. Alternatively, don't bother with the odd/even split, and just use (10.2) for the inner integral.

They are $\frac{1}{2}(\phi(k) \pm \phi(-k))$; show that both of these are test functions.

Use the decay properties of the test function to justify use of the Riemann–Lebesgue lemma for the outer integrals, and expand ϕ in a Taylor series for the inner one.

(ii) Consider the Fourier transform of

$$\mathcal{H}_\epsilon(x) = \mathcal{H}(x)e^{-\epsilon x},$$

which clearly exists for $\epsilon > 0$; show that it is

$$\widehat{\mathcal{H}}_\epsilon(k) = \frac{1}{\epsilon - ik}.$$

Note that this 'does the right sort of thing' as $\epsilon \to 0$: it tends to $-1/ik$ for $k \neq 0$, and to infinity for $k = 0$.

Writing

$$\frac{1}{\epsilon - ik} + \frac{1}{ik} = \frac{\epsilon}{\epsilon^2 + k^2} - \frac{i\epsilon^2}{k(\epsilon^2 + k^2)},$$

show that as $\epsilon \to 0$ the action of the right-hand side on a test function tends to that of $\pi\delta(k)$. (You will need to interpret the second term as a principal value integral; use the results of Exercise 3.)

18. **More Fourier transforms.** What are the Fourier transforms of

$$x, \quad x^n, \quad |x|,$$

for integral $n > 0$?

Remember that $\widehat{(xf)} = -id\hat{f}/dk$.

"You can always make infinity smaller by multiplying by h."