6 Generalized Functions

We've used separation of variables to solve various important second–order partial differential equations inside a compact domain $\Omega \subset \mathbb{R}^n$. But our approach should leave you with many questions. The method started by finding a particular solution of the form $\phi(\mathbf{x}) = X(x)Y(z)\cdots$, and then using the homogeneous boundary conditions to constrain (quantize) the allowed values of the separation constants that occur in such a solution. So far so good. We then used linearity of the p.d.e. to claim that a more general solution would be given by summing our particular solutions over all possible allowed values of the separation constants, with coefficients that were fixed by the boundary and/or initial conditions.

The trouble is that, generically, our boundary conditions required that we include an *infinite* number of terms in the sum. We should first of all worry about whether such an infinite sum converges everywhere within Ω, or for all times. But even if it does, can we be sure whether it converges to an actual solution of the differential equation we started with? That is, while each term in the infinite sum certainly obeys the p.d.e., perhaps the infinite sum itself does not: we've seen in chapter 1 that Fourier series sometimes converge to non–differentiable or even non–continuous functions and that term–by–term differentiation of a Fourier series makes convergence worse. And suppose we find the function our series converges to indeed is not twice differentiable. Should we worry? Could there be some sense in which we should still allow these non–differentiable solutions? But, if we do, can we be sure that we've really found the general solution to our problem? The uniqueness theorems we obtained in previous chapters always involved showing that the *integral* of some non-negative quantity was zero, and then concluding that the quantity itself must be zero. This might fail if, say, our series solution differs from the 'true' solution just on a set of measure zero.

You can certainly understand why mathematicians such as Laplace were so reluctant to accept Fourier's methods. Quite remarkably, the most fruitful way forward has turned out not to be to restrict Fourier to situations where everything converges to sufficiently differentiable functions that our concerns are eased, but rather to be *to generalize the very notion of a function itself* with the aim of finding the right class of object where Fourier's method always makes sense. *Generalized functions* were introduced in mathematics by Sobolev and Schwartz. They're designed to fulfill an apparently mutually contradictory pair of requirements: They are sufficiently well–behaved that they're infinitely differentiable and thus have a chance to satisfy partial differential equations, yet at the same time they can be arbitrarily singular – neither smooth, nor differentiable, nor continuous, nor even *finite* – if interpreted naively as 'ordinary functions'. Generalized functions have become a key tool in much of p.d.e. theory, and form a huge part of analysis.

If this formal theory is not your cup of tea, there's yet another reason to be interested in generalized functions. When we come to study inhomogeneous (driven) equations such as Poisson's equation $\nabla^2\phi = \rho$, physics considerations suggest that we'll be interested in cases where the source ρ is not a smooth function. For example, ρ might represent the charge density in some region Ω, and we may only have point–like charges. The total

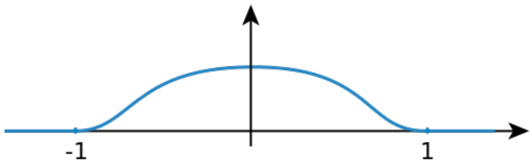

Figure 11. *The bump function $\Psi(x)$ is an example of a smooth test function with compact support.*

charge $Q = \int_\Omega \rho \, \mathrm{d}V$ is thus finite, but ρ vanishes everywhere except at finitely many points. Generalized functions will allow us to handle p.d.e.s with such singular source terms. In fact, the most famous generalized function was discovered in physics by Dirac before the analysts cottoned on, and generalized functions are often known as *distributions*, as a nod to the charge distribution example which inspired them.

6.1 Test functions and distributions

Non–examinable again!

To define a distribution, we must first choose a class of *test functions*. For $\Omega \subseteq \mathbb{R}^n$, the simplest class of test functions are infinitely smooth functions $\phi \in C^\infty(\Omega)$ that have compact support. That is, there exists a compact set $K \subset \Omega$ such that $\phi(x) = 0$ whenever $x \notin K$. A simple example of a test function in one dimension is

$$\Psi(x) \equiv \begin{cases} \mathrm{e}^{-1/(1-x^2)} & \text{when } |x| < 1 \\ 0 & \text{else} \end{cases} \tag{6.1}$$

which is shown in figure 11, but any infinitely smooth function with compact support will do. We let $\mathcal{D}(\Omega)$ denote the space of all such test functions.

Having chosen our class of test functions, we now define a distribution T to be a linear map $T : \mathcal{D}(\Omega) \to \mathbb{R}$, given by

$$T : \phi \mapsto T[\phi] \tag{6.2}$$

for $\phi \in \mathcal{D}(\Omega)$. The square bracket notation in $T[\phi]$ reminds us that T is not a function on Ω itself, but rather is a function *on the infinite dimensional space of test functions* on Ω.

The space of distributions with test functions in $\mathcal{D}(\Omega)$ is denoted $\mathcal{D}'(\Omega)$. It's again an infinite dimensional vector space, because we can add two distributions T_1 and T_2 together, defining the distribution $(T_1 + T_2)$ by

$$(T_1 + T_2)[\phi] \equiv T_1[\phi] + T_2[\phi] \tag{6.3}$$

for all $\phi \in \mathcal{D}(\Omega)$. Similarly, we can multiply a distribution by a constant, defining the distribution (cT) by

$$(cT_1)[\phi] \equiv c\,T_1[\phi] \tag{6.4}$$

for all $\phi \in \mathcal{D}(\Omega)$, and $c \in \mathbb{R}$. The multiplication on the *rhs* here is just the standard multiplication in $\mathbb{R}$. Finally, notice that while we can multiply distributions by smooth functions – if $\psi \in C^\infty(\Omega)$ and $T \in \mathcal{D}'(\Omega)$ then define the distribution (ψT) by $(\psi T)[\phi] := T[\psi\phi]$

using multiplication in $C^\infty(\Omega)$ – in general *there is no way to multiply two distributions together*.

The simplest type of distribution is just an ordinary function $f : \Omega \to \mathbb{R}$ that is locally integrable, meaning that its integral over any compact set converges[29]. To treat f as a distribution we must say how it acts on any test function $\phi \in \mathcal{D}(\Omega)$. To do so, we choose to define

$$f[\phi] := (f, \phi) = \int_\Omega f(x)\,\phi(x)\,\mathrm{d}V\,, \tag{6.5}$$

which is just the inner product of f with ϕ. This integral is guaranteed to be well–defined even when Ω is non–compact (say, the whole of $\mathbb{R}^n$) since ϕ has compact support and f is locally integrable. In particular, note that unlike the test functions, we do *not* require f itself to have compact support. Also notice that (6.5) certainly gives a *linear* map from $\mathcal{D}(\Omega)$ to $\mathbb{R}$, since if ϕ_1 and ϕ_2 are two test functions and c_1, c_2 are constants then

$$\begin{aligned} T_f[c_1\phi_1 + c_2\phi_2] &= (f, c_1\phi_1 + c_2\phi_2) = c_1(f, \phi_1) + c_2(f, \phi_2) \\ &= c_1 T_f[\phi_1] + c_2 T_f[\phi_2] \end{aligned} \tag{6.6}$$

by the linearity of the inner product in its second entry. Thus, in the case where the generalized function is just an ordinary function, the map $T_f : \mathcal{D}(\Omega) \to \mathbb{R}$ just corresponds to the usual inner product between functions.

By far the most important example of a generalized function that is not a function is the *Dirac delta*, written just δ. It is defined by

$$\delta[\phi] := \phi(0) \tag{6.7}$$

for all $\phi \in \mathcal{D}(\Omega)$, where 0 is the origin in $\mathbb{R}^n$. Note again that δ is indeed a linear map $\delta : \mathcal{D}(\Omega) \to \mathbb{R}$, with $\delta[c_1\phi_1 + c_2\phi_2] = c_1\phi_1(0) + c_2\phi_2(0)$, where the addition on the left is in the space of test functions, while the addition on the right is just addition in $\mathbb{R}$.

By analogy with the case where the generalized function is itself a function, it's often convenient to abuse notation and write

$$T[\phi] \overset{?!}{=} (T, \phi) = \int_\Omega T(x)\,\phi(x)\,\mathrm{d}V \tag{6.8}$$

even for general distributions. However, for a general distribution *the object $T(x)$ is not a function* – *i.e.*, there is no sense in which $T : \Omega \to \mathbb{R}$ – and to specify which distribution we're talking about, we have to give a separate definition of what value $T[\phi]$ actually takes. For example, it's common to write

$$\delta[\phi] = \int_\Omega \delta(x)\,\phi(x)\,\mathrm{d}V \tag{6.9}$$

for some object $\delta(x)$. However, $\delta(x)$ can't possibly be a genuine function, for if (6.9) is to be compatible with (6.7) for arbitrary test functions ϕ we need $\delta(x) = 0$ whenever $x \neq 0$.

[29] Thus, in one dimension, a function $1/x^2$ with a non–integrable singularity at $x = 0$ does not qualify as a distribution.

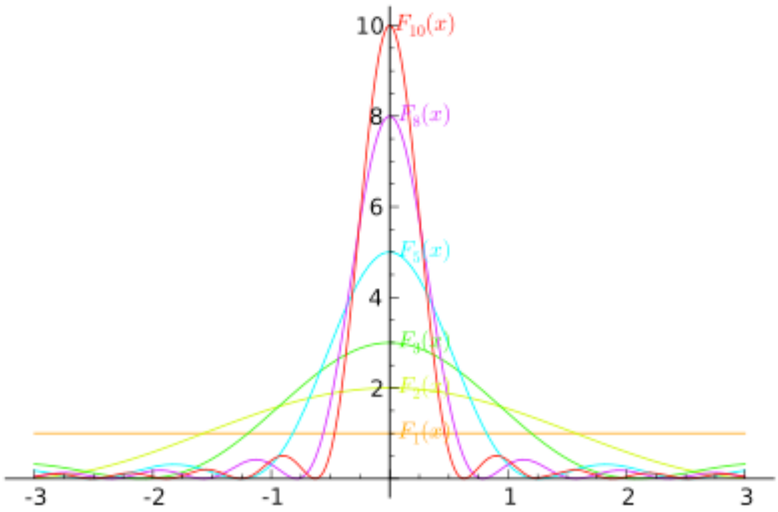

Figure 12. *The Fejér kernels $F_n(x)$ for $n = 1, 2, 3, 5, 8$ and 10. The limiting value of these kernels as $n \to \infty$ can be thought of as the Dirac δ-function.*

If not, we could get a non–zero value for our integral by choosing a test function whose support lies only in some small compact set away from the origin, in conflict with (6.7). On the other hand, if $\delta(x)$ does indeed vanish everywhere except at one point, the integral (6.9) cannot give the finite answer $\phi(0)$ if $\delta(x)$ takes any finite value at $x = 0$. So it isn't a genuine function in the sense of being a map from $\Omega \to \mathbb{R}$. Just to confuse you, $\delta(x)$ is ubiquitously known as the "Dirac δ-function".

One reason this abusive notation is convenient is that distributions can arise as the limit of a sequence of integrals of usual functions. For example, for any finite $n \in \mathbb{N}$ the function

$$F_n(x) = \begin{cases} \dfrac{1}{n+1}\dfrac{\sin^2[(n+1)x/2]}{\sin^2[x/2]} & \text{for } x \neq 0 \\ n+1 & \text{when } x = 0 \end{cases} \tag{6.10}$$

(which you may recognize as a Fejér kernel) is a perfectly respectable, finite and everywhere continuous function. In particular, for any finite n the integral $\int_{\mathbb{R}} F_n(x)\phi(x)\,\mathrm{d}x$ of the Fejér kernel times our compactly supported test function is well-defined. Now, as n increases, the $F_n(x)$ are increasingly concentrated around $x = 0$ as you can see from figure 12. Whilst the limiting value $\lim_{n\to\infty} F_n(x)$ of this sequence of functions does not exist (in particular, $\lim_{n\to\infty}(n+1)$ does not exist), the limiting value of the *integrals* is perfectly finite, and in fact

$$\lim_{n\to\infty} \int_{\mathbb{R}} F_n(x)\,\phi(x)\,\mathrm{d}x = \phi(0) \tag{6.11}$$

as we showed in section 1.6. Thus the Fejér kernels $\{F_n(x)\}$ form a sequence of functions whose limit can be understood as a distribution, which in this case we recognise as the Dirac δ-function.

The functions $F_n(x)$ are far from unique in having the Dirac δ as a limiting case. Other examples include the family of Gaussians (4.11) $G_n(x) = n\mathrm{e}^{-n^2x^2}/\sqrt{\pi}$, or the sinc functions

$S_n(x) = \sin(nx)/(\pi x)$, or the so–called 'top hat function'

$$P_n(x) = \begin{cases} \dfrac{n}{2} & \text{for } |x| < \dfrac{1}{n} \\ 0 & \text{otherwise,} \end{cases} \tag{6.12}$$

or any other sequence of functions each member of which has total area is 1, and which become increasingly spiked around the origin.

6.1.1 Differentiation of distributions

Now comes the magic. In order to use distributions in differential equations, we need to know what it means to differentiate them. In the case that our generalized function is just an ordinary function f, in one dimension we would have

$$T_{f'}[\phi] = (f', \phi) = \int_\Omega f'(x)\,\phi(x)\,\mathrm{d}x = -\int_\Omega f(x)\,\phi'(x)\,\mathrm{d}x = -(f, \phi') = -T_f[\phi'] \tag{6.13}$$

where the boundary term vanishes since ϕ had compact support inside Ω. For a generalized function T, we now *define*

$$T'[\phi] \equiv -T[\phi'] \qquad \text{for all } \phi \in \mathcal{D}(\Omega)\,. \tag{6.14}$$

Again, the idea here is that if we think of our distribution as coming from the limit of a sequence of integrals involving only 'ordinary' functions, by (6.13) this relation will hold for every member of the sequence, and so it will hold for the limiting value of the integrals. The definition also means that, provided we know what the distribution T does to all test functions, we also know what the distribution T' does.

As an example, consider again the Dirac δ defined by $\delta[\phi] = \phi(0)$. The derivative of the δ-function is given by (6.14) as

$$\delta'[\phi] = -\delta[\phi'] = -\phi'(0) \tag{6.15}$$

and so evaluates (minus) the derivative of test function at the origin. In turn, the Heaviside step function

$$\Theta(x) \equiv \begin{cases} 1 & \text{when } x > 0 \\ 0 & \text{when } x \leq 0\,. \end{cases} \tag{6.16}$$

is a function, whose integral over any compact set converges, so $\Theta(x)$ also defines a generalized function on $\mathbb{R}$ by $\Theta[\phi] = \int_{-\infty}^{\infty} \Theta(x)\phi(x)\,\mathrm{d}x = \int_0^\infty \phi(x)\,\mathrm{d}x$ which converges since ϕ has compact support. But while $\Theta(x)$ is patently not differentiable (or even continuous) as a function, it is perfectly differentiable if treated as a distribution. We have

$$\begin{aligned} \Theta'[\phi] &= -\Theta[\phi'] = -\int_{-\infty}^{\infty} \Theta(x)\,\frac{\partial\phi}{\partial x}\,\mathrm{d}x \\ &= -\int_0^\infty \frac{\partial\phi}{\partial x}\,\mathrm{d}x = \phi(0) - \phi(\infty) = \phi(0)\,, \end{aligned} \tag{6.17}$$

using the fact that ϕ has compact support. Since $\Theta'[\phi] = \phi(0) = \delta[\phi]$ holds for any test function ϕ, we can identify Θ' as the distribution δ.

We define higher derivatives of distributions similarly: since ϕ is infinitely differentiable we have

$$T^{(n)}[\phi] \equiv (-1)^n \, T[\phi^{(n)}] \tag{6.18}$$

and again this is determined once we know what T itself does. For example, all higher derivatives of the Dirac δ are given by

$$\delta^{(n)}[\phi] = (-1)^n \, \delta[\phi^{(n)}] = (-1)^n \, \phi^{(n)}(0) \,, \tag{6.19}$$

and the *rhs* makes sense since the test function $\phi(x)$ was infinitely differentiable. Notice that this definition pulls off a really remarkable trick: we've managed to define *all* the derivatives of an object such as δ that at first sight seems impossibly non–differentiable. Even more, our definition means that distributions inherit the excellent differentiability properties of the infinitely smooth test functions!

6.2 Properties of the Dirac δ

Since the Dirac δ is such an important distribution, it'll be worth our while examining it in somewhat greater detail. We first establish several properties that will come in handy later. These are easiest to obtain if one uses the integral expression (6.9) and manipulates the object $\delta(x)$ *as if* it were a genuine function. Such manipulations can be made rigorous in the deeper theory of distributions, but we'll content ourselves in this Methods course with what follows.

– Since $\delta(x)$ vanishes whenever $x \neq 0$ we can write

$$\delta[\phi] = \int_a^b \delta(x) \, \phi(x) \, \mathrm{d}x$$

  where $[a, b]$ is any interval containing the point $x = 0$. If $0 \notin [a, b]$ then the integral is zero.

– If $f : \Omega \to \mathbb{R}$ is continuous in a neighbourhood of the origin $0 \in \Omega$, then the distribution $(f\delta)$ obeys

$$(f\delta)[\phi] = \delta[f\phi] = f(0)\phi(0) = f(0) \times \delta[\phi] \tag{6.20}$$

  in accordance with our rule for multiplying distributions by smooth functions. In terms of the Dirac δ-'function' $\delta(x)$ we write this as $f(x)\,\delta(x) = f(0)\,\delta(x)$ using the idea that $\delta(x)$ vanishes whenever $x \neq 0$.

– For any $c \in \mathbb{R}$,

$$\int_{\mathbb{R}} \delta(cx) \, \phi(x) \, \mathrm{d}x = \frac{1}{|c|} \int_{\mathbb{R}} \delta(y) \, \phi(y/c) \, \mathrm{d}y = \frac{1}{|c|} \phi(0) \,, \tag{6.21}$$

  so we write $\delta(cx) = \delta(x)/|c|$.

– For any point $a \in \mathbb{R}$ we can define a translated Dirac δ by $\delta_a[\phi] = \phi(a)$. In terms of the δ-function, this is

$$\phi(a) = \int_{\mathbb{R}} \delta(y)\, \phi(y+a)\, \mathrm{d}y = \int_{\mathbb{R}} \delta(x-a)\, \phi(x)\, \mathrm{d}x \tag{6.22}$$

so that translating the δ-function shifts the point at which the test function is evaluated.

– For any continuously differentiable function $f : \Omega \to \mathbb{R}$ we have that $\delta(f(x))$ zero everywhere except at the zeros of f, so that an integral involving $\delta(f(x))$ times a test function receives no contributions outside an arbitrarily small neighbourhood of the zeros of f. In particular, if f has only simple zeros at points $\{x_1, x_2, \ldots, x_n\} \in \Omega$ then

$$\int_{\mathbb{R}} \delta(f(x))\, \phi(x)\, \mathrm{d}x = \sum_{i=1}^{n} \int_{x_i^-}^{x_i^+} \delta(f(x))\, \phi(x)\, \mathrm{d}x\,. \tag{6.23}$$

Using the fact that $f(x) \approx (x - x_i)\, f'(x_i)$ when x is near the root x_i, we have

$$\begin{aligned}\int_{\mathbb{R}} \delta(f(x))\, \phi(x)\, \mathrm{d}x &= \sum_{i=1}^{n} \left[\frac{1}{|f'(x_i)|} \int_{x_i^-}^{x_i^+} \delta(x - x_i)\, \phi(x)\, \mathrm{d}x \right] \\ &= \sum_{i: f(x_i)=0} \frac{\phi(x_i)}{|f'(x_i)|}\,,\end{aligned} \tag{6.24}$$

where the first equality used equation (6.21) and the second (6.22). For example, if $f(x) = x^2 - b^2$ then

$$\int_{\mathbb{R}} \delta(x^2 - b^2)\, \phi(x)\, \mathrm{d}x = \frac{1}{2|b|} \left[\phi(b) + \phi(-b) \right]\,. \tag{6.25}$$

The previous two results (6.21) & (6.22) can be understood as special cases of this one.

In all these expressions, the important point is that the integral is *localized* to an arbitrarily small neighbourhood of the zeros of the argument of the δ-function.

In physics the δ-function models *point sources* in a continuum. For example, suppose we have a unit point charge at $x = 0$ (in one dimension). Then its charge density $\rho(x)$ should satisfy $\rho(x) = 0$ for $x \neq 0$ and total charge $Q = \int \rho(x) = 1$. These are exactly the properties of the δ-function, so we set $\rho(x) = \delta(x)$ and the physical intuition is well modelled by the sequence in, say, (6.12).

In mechanics, δ-functions model *impulses*. Suppose a particle traveling with momentum $p = mv$ in one dimension. Newton's law gives $dp/dt = F$, so

$$p(t_2) - p(t_1) = \int_{t_1}^{t_2} F\, \mathrm{d}t\,.$$

If the particle is suddenly struck by a hammer at $t = 0$ then we might imagine that the force acts only over a vanishingly small time Δt with $|\Delta t| < 2\epsilon$ for some small ϵ, and yet results in a *finite* momentum change, say C. Then $\int_{-\epsilon}^{\epsilon} F\,\mathrm{d}t = \Delta p$ while F is nonzero only very near $t = 0$. In the limit of vanishing time interval Δt, $F(t) = C\,\delta(t)$ models the impulsive force. The δ-function was originally introduced by P.A.M. Dirac in the 1930s from considerations of position and momentum in quantum mechanics.

6.2.1 Eigenfunction expansion of the δ-function

The excellent differentiability of distributions also allows us to make sense of divergent Fourier series. Let's begin by computing the Fourier series of the Dirac δ-function. We consider $\delta(x)$ to be a distribution defined for $x \in [-L, L]$, and then formally write

$$\delta(x) = \sum_{n\in\mathbb{Z}} \hat{\delta}_n\,\mathrm{e}^{\mathrm{i}n\pi x/L} \qquad \text{with} \qquad \hat{\delta}_n = \frac{1}{2L}\int_{-L}^{L} \mathrm{e}^{-\mathrm{i}n\pi x/L}\,\delta(x)\,. \tag{6.26}$$

From the definition (6.7) we see that $\hat{\delta}_n = 1/2L$ for all n, so we find

$$\delta(x) = \frac{1}{2L}\sum_{n\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}n\pi x/L}\,. \tag{6.27}$$

What can this result possibly mean? On the one hand, the *lhs* is an object that doesn't really exist as a function, whereas on the right hand side we have a series that clearly diverges!

To find out, let

$$S_N\delta(x) \equiv \frac{1}{2L}\sum_{n=-N}^{N} \mathrm{e}^{\mathrm{i}n\pi x/L} \tag{6.28}$$

denote the partial Fourier sum. For any finite N, $S_N\delta(x)$ is a perfectly well behaved (even analytic) function of x. If we treat this function as a distribution, abusively also called $S_N\delta$, then for any test function $\phi \in \mathcal{D}([-L, L])$ we have

$$\begin{aligned}(S_N\delta)[\phi] = \int_{-L}^{L} S_N\delta(x)\,\phi(x)\,\mathrm{d}x \;&=\; \frac{1}{2L}\int_{-L}^{L}\left[\sum_{n=-N}^{N} \mathrm{e}^{\mathrm{i}n\pi x/L}\right]\phi(x)\,\mathrm{d}x \\ = \sum_{n=-N}^{N}\left[\frac{1}{2L}\int_{-L}^{L} \mathrm{e}^{\mathrm{i}n\pi x/L}\,\phi(x)\,\mathrm{d}x\right] \;&=\; \sum_{n=-N}^{N} \hat{\phi}_{-n}\,.\end{aligned} \tag{6.29}$$

In going to the second line we exchanged the finite sum and the integral, and then recognized the resulting integral as the definition of the Fourier coefficient ϕ_{-n} of our test function.

Now, the final sum in equation (6.29) is just the partial Fourier series of $\phi(x)$, evaluated at $x = 0$. Since the test function is everywhere smooth, its Fourier series converges everywhere. In particular, convergence at the origin means we have

$$\lim_{N\to\infty}\left(\sum_{n=-N}^{N} \hat{\phi}_{-n}\right) = \lim_{N\to\infty}\left(\sum_{n=-N}^{N} \hat{\phi}_{-n}\,\mathrm{e}^{-\mathrm{i}\pi n x/L}\right)\Bigg|_{x=0} = \phi(0)\,, \tag{6.30}$$

and therefore the limiting value of our distribution is

$$\lim_{N\to\infty}(S_N\delta)[\phi] = \phi(0) = \delta[\phi]\,. \tag{6.31}$$

It is in this sense that the partial sum $\frac{1}{2L}\sum_{n=-N}^{N} \mathrm{e}^{\mathrm{i}\pi x/L}$ converges to $\delta(x)$. Of course it could not converge as a function, because there simply is no such function "$\delta(x)$". Rather, it converges as a distribution.

As a related example, recall from section 1.5.4 that the sawtooth function given by the 2π-periodic extension of $f(x) = x$ for $x \in [-\pi, \pi)$ has Fourier series

$$2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\sin nx\,.$$

This infinite series does in fact converge for all $x \in \mathbb{R}$, albeit very slowly since the Fourier coefficients fall only as $1/n$. In fact, for $x \neq n\pi$ (with $n \in \mathbb{Z}$) it converges to the value of the sawtooth function, while it converges to zero when $x = n\pi$. Differentiating the series term–by–term leads to the series

$$2\sum_{n=1}^{\infty}(-1)^{n+1}\cos nx \;=\; \cos x - \cos 2x + \cos 3x - \cdots \tag{6.32}$$

which now diverges as a function. We can use the step function to write the sawtooth function as

$$f(x) = \; x + 2\pi\sum_{n=1}^{\infty}\Theta(x - n\pi) - 2\pi\sum_{n=0}^{\infty}\Theta(-x-\pi n) \tag{6.33}$$

for any $x \in \mathbb{R}$, where the step functions provide the jumps in the sawtooth. Using the fact that the derivative of the step function is the Dirac δ-function we have

$$f'(x) = 1 + 2\pi\sum_{n\in\mathbb{Z}}^{\infty}\delta(x - n\pi)\,, \tag{6.34}$$

understood as a distribution. This says that the sawtooth function has a constant gradient $(= 1)$ everywhere except at $x = n\pi$ for $n \in \mathbb{Z}$, where it has a δ-function spike. Following the same steps as above, you can check that

$$2\lim_{N\to\infty}\left(\sum_{n=1}^{N}(-1)^{n+1}\cos nx\right) = 1 + 2\pi\sum_{n\in\mathbb{Z}}^{\infty}\delta(x - n\pi) \tag{6.35}$$

so that the sequence of partial series $\sum_{n=1}^{N}(-1)^{n+1}\cos nx$ does converge as a distribution.

Of course, there's nothing particularly special about expanding the Dirac δ-function in a Fourier series; any other basis of orthogonal functions will do just as well. For $n \in \mathbb{Z}$ let $\{Y_n(x)\}$ be a complete set of orthonormal eigenfunctions of a Sturm–Liouville operator on the domain $[a, b]$, with weight function $w(x)$. For $\xi \in (a, b)$, the Dirac δ-function $\delta(x - \xi)$

is zero for the boundary points $x = a$ or $x = b$, and so the Sturm–Liouville operator will be self–adjoint. We thus expect to be able to expand

$$\delta(x-\xi) = \sum_{n\in\mathbb{Z}} c_n\, Y_n(x) \tag{6.36}$$

where the coefficient c_n is given by

$$c_n = \int_a^b Y_n^*(x)\,\delta(x-\xi)\,w(x)\,\mathrm{d}x = Y_n^*(\xi)\,w(\xi) \tag{6.37}$$

again from the definition of the δ-function. Using the fact that $\delta(x-\xi) = [w(x)/w(\xi)] \times \delta(x-\xi)$ we can write

$$\delta(x-\xi) = w(\xi)\sum_{n\in\mathbb{Z}} Y_n^*(\xi)\,Y_n(x) = w(x)\sum_{n\in\mathbb{Z}} Y_n^*(\xi)\,Y_n(x)\,. \tag{6.38}$$

This expansion is consistent with the sampling property, since if $g(x) = \sum_{m\in\mathbb{Z}}^{\infty} d_m\, Y_m(x)$ then, again exchanging the sums and integrals,

$$\begin{aligned}\int_a^b g^*(x)\,\delta(x-\xi)\,\mathrm{d}x &= \sum_{m,n\in\mathbb{Z}}\left[Y_n^*(\xi)\,d_m^*\int_a^b w(x)\,Y_m^*(x)\,Y_n(x)\,\mathrm{d}x\right]\\ &= \sum_{m\in\mathbb{Z}} d_m^* Y_m^*(\xi) = g^*(\xi)\end{aligned} \tag{6.39}$$

using the orthonormality of the Sturm–Liouville eigenfunctions.

We'll soon see that the eigenfunction expansion of the δ-function is intimately related to the eigenfunction expansion of the Green's function that we introduced in section 2.6. Our next task is to develop a theory of Green's functions for solving inhomogeneous ODEs.

6.3 Schwartz functions and tempered distributions

Non–examinable, yet again

The definition of distributions depends on a choice of class of test functions. Above, we considered test functions that are infinitely smooth and have compact support. These requirements ensured in particular that the integral in (6.5) was well–defined, and that the integration by parts in (6.13) received no boundary terms. Well–definedness and absence of boundary terms would be retained if we relax our requirement that test functions have compact support to the requirement simply that it decays sufficiently rapidly as $|x| \to \infty$.

We define a *Schwartz function* $\psi : \mathbb{R} \to \mathbb{C}$ to be an infinitely smooth function with the property that

$$\sup_{x\in\mathbb{R}} \left| x^m\, \psi^{(n)}(x)\right| < \infty \qquad \text{for all } m, n \in \mathbb{N}. \tag{6.40}$$

Thus both ψ and all its derivatives are bounded, and in particular vanish faster than any inverse power at infinity. A simple example of a Schwartz function is $\psi(x) = p(x)\,\mathrm{e}^{-x^2}$ with $p(x)$ any polynomial. We denote the space of Schwartz functions on Ω by $\mathcal{S}(\Omega)$ and, because of their excellent asymptotic properties, we can use Schwartz functions as our test functions in defining distributions. Note that any compactly supported test function $\phi \in \mathcal{D}(\Omega)$ certainly obeys the Schwartz conditions (6.40), so $\mathcal{D}(\Omega) \subset \mathcal{S}(\Omega)$ (and in fact one can show that $\mathcal{D}(\Omega)$ is dense in $\mathcal{S}(\Omega)$).

Just as we did with distributions and compactly supported test functions, we now define *tempered distributions* to be linear maps $T : \mathcal{S} \to \mathbb{R}$, and we write the space of tempered distributions as $\mathcal{S}'(\Omega)$. Now however $\mathcal{S}'(\Omega) \subset \mathcal{D}'(\Omega)$ so that there are fewer tempered distributions than distributions. For example, in order for even an *ordinary* function $g : \Omega \to \mathbb{R}$ to be admissible as a tempered distribution we'd need to require that

$$\lim_{|x|\to\infty} x^{-n}\, g(x) = 0 \qquad \text{for } \textit{some } n \in \mathbb{N}, \tag{6.41}$$

because the integral $(g, \psi) = \int_{\mathbb{R}} g(x)\, \psi(x)\, \mathrm{d}x$ will only exist if the good behaviour of ψ is not ruined by that of g. Thus the functions $x^3 + x$, e^{-x^2}, $\sin x$ and $x \ln |x|$ are all good tempered distributions, but the functions $1/x^2$ and e^{-x} are not. (*Exercise:* Show that the Dirac δ distribution is also a tempered distribution.)

Properties of tempered distributions may be obtained in exactly the same way as for distributions with compactly supported test functions. In particular, if $\psi \in \mathcal{S}(|\Omega)$ and $T : \mathcal{S}(\Omega) \to \mathbb{C}$ is a tempered distribution with $T : \psi \mapsto T[\psi]$, then the n^{th} derivative $T^{(n)}[\psi] := (-1)^n T[\psi^{(n)}]$ just as before. Tempered distributions come into their own in conjunction with Fourier transforms: The Fourier transform of a Schwartz function is again a Schwartz function, and this fact allows us to define the Fourier transform of any tempered distribution.