When functions have no value(s): Delta functions and distributions

Steven G. Johnson, MIT course 18.303 notes

Created October 2010, updated October 11, 2011.

Abstract

These notes give a brief introduction to the motivations, concepts, and properties of *distributions*, which generalize the notion of functions $f(x)$ to allow derivatives of discontinuities, "delta" functions, and other nice things. This generalization is increasingly important the more you work with linear PDEs, as we do in 18.303. For example, Green's functions are extremely cumbersome if one does not allow delta functions. Moreover, solving PDEs with functions that are not classically differentiable is of great practical importance (e.g. a plucked string with a triangle shape is not twice differentiable, making the wave equation problematic with traditional functions). Any serious work with PDEs will eventually run into the concept of a "weak solution," which is essentially a version of the distribution concept.

1 What's wrong with functions?

The most familiar notion of a function $f(x)$ is a map from real numbers $\mathbb{R}$ to real numbers $\mathbb{R}$ (or maybe complex numbers $\mathbb{C}$); that is, for every x you have a value $y = f(x)$. Of course, you may know that one can define functions from any set to any other set, but at first glance it seems that $\mathbb{R} \to \mathbb{R}$ functions (and multi-dimensional generalizations thereof) are the best-suited concept for describing most physical quantities—for example, velocity as a function of time, or pressure or density as a function of position in a fluid, and so on. Unfortunately, such functions have some severe drawbacks that, eventually, lead them to be replaced in some contexts by another concept: *distributions* (also called *generalized functions*). What are these drawbacks?

1.1 No delta functions

For lots of applications, such as those involving PDEs and Green's functions, one would like to have a function $\delta(x)$ whose integral is "concentrated" at the point $x = 0$. That is, one would like the function $\delta(x) = 0$ for all $x \neq 0$, but with $\int \delta(x)dx = 1$ for any integration region that includes $x = 0$; this concept is called a "Dirac delta function" or simply a "delta function." $\delta(x)$ is usually the simplest right-hand-side for which to solve differential equations, yielding a Green's function. It is also the simplest way to consider physical effects that are concentrated within very small volumes or times, for which you don't actually want to worry about the microscopic details in this volume—for example, think of the concepts of a "point charge," a "point mass," a force plucking a string at "one point," a "kick" that "suddenly" imparts some momentum to an object, and so on. The problem is that there is no classical function $\delta(x)$ having these properties.

For example, one could imagine constructing this function as the limit:

$$\delta(x) = \lim_{\Delta x \to 0^+} \begin{cases} \frac{1}{\Delta x} & 0 \leq x < \Delta x \\ 0 & \text{otherwise} \end{cases} = \lim_{\Delta x \to 0^+} \delta_{\Delta x}(x)?$$

For any $\Delta x > 0$, the function $\delta_{\Delta x}(x)$ at right has integral $= 1$ and is zero except near $x = 0$. Unfortunately, the $\Delta x \to 0$ limit does not exist as an ordinary function: $\delta_{\Delta x}(x)$ approaches ∞ for $x = 0$, but of course ∞ is not a real number.

Informally, one often sees "definitions" of $\delta(x)$ that describe it as some mysterious object that is "not quite" a function, which $= 0$ for $x \neq 0$ but is undefined at $x = 0$, and which is "only really defined inside an integral" (where it "integrates" to 1).[1] This may leave you with a queasy feeling that $\delta(x)$ is somehow

---

[1] Historically, this unsatisfactory description is precisely how $\delta(x)$ first appeared, and for many years there was a corresponding cloud of guilt and uncertainty surrounding any usage of it. Most famously, an informal $\delta(x)$ notion was popularized by physicist Paul Dirac, who in his *Principles of Quantum Mechanics* (1930) wrote: "Thus $\delta(x)$ is not a quantity which can be generally used in mathematical analysis like an ordinary function, but its use must be confined to certain simple types of expression for which it is obvious that no inconsistency can arise." You know you are on shaky ground, in mathematics, when you are forced to appeal to the "obvious."

not real or rigorous (and therefore anything based on it may be suspect). For example, integration is an operation that is classically only defined for ordinary functions, so it may not even be clear (yet) what "$\int$" means when we write "$\int \delta(x)dx$".

1.2 Not all functions are differentiable

Most physical laws can be written in the form of derivatives, but lots of functions are not differentiable. Discontinuous functions arise all of the time at the interface between two materials (e.g. think of the density at the interface between glass and air). Of course, at a microscopic level you could argue that the quantum wavefunctions might be continuous, but one hardly wishes to resort to atoms and quantum mechanics every time a material has a boundary!

The classic example of a discontinuous function is the Heaviside step function:

$$S(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

The derivative $S'(x)$ is zero everywhere except at $x = 0$, where the derivative does not exist—the slope at $x = 0$ is "infinity." Notice that $H'(x)$ very much resembles $\delta(x)$, and $\int_\infty^x \delta(x')dx'$ would certainly look something like $S(x)$ if it existed (since the "integral" of δ should be 1 for $x > 0$)—this is not a coincidence, and it would be a shame not to exploit it!

A function doesn't need to be discontinuous to lack a derivative. Consider the function $|x|$: its derivative is $+1$ for $x > 0$ and -1 for $x < 0$, but at $x = 0$ the derivative doesn't exist. We say that $|x|$ is only "piecewise" differentiable.

Note that $S(x)$ is very useful for writing down all sorts of functions with discontinuities. $|x| = xS(x) - xS(-x)$, for example, and the $\delta_{\Delta x}(x)$ "box" function on the right-hand-side of the $\delta(x)$ "definition" above can be written $\delta_{\Delta x}(x) = [S(x) - S(x - \Delta x)]/\Delta x$.

1.3 Nagging worries about discrepancies at isolated points

When we try to do linear algebra with functions, we continually find ourselves worrying about excluding odd caveats and counterexamples that have to do with *finite* discrepancies at isolated points. For example, a Fourier series of a square-integrable function converges everywhere... except at isolated points of discontinuity [like the point $x = 0$ for $S(x)$, where a Fourier series would converge to 0.5]. As another example, $\langle u, v \rangle = \int \bar{u}v$ defines an inner product on functions and a norm $\|u\|^2 = \langle u, u \rangle > 0$ for

$u \neq 0$... except for $u(x)$ that are only nonzero for isolated points. And with functions like $S(x)$ there are all sorts of apparently pointless questions about what value to assign exactly at the discontinuity. It may likewise seem odd to care about what value to assign the slope of $|x|$ at $x = 0$; surely any value in $[-1, 1]$ should do?

1.4 Limits and derivatives cannot always be interchanged

In numerical and analytical PDE methods, we are continually writing functions as limits. A Fourier series is a limit as the number of terms goes to infinity. A finite-difference method solves the problem in the limit as $\Delta x \to 0$. We initially find the Green's function as the limit of the response to a box-like function that is nonzero outside of a width Δx, and then take the limit $\Delta x \to 0$. After all of these kinds of things, we eventually substitute our solution back into the PDE, and we assume that the limit of the solution is still a solution. In doing so, however, we usually end up interchanging the limits and the differentiation. For example, we usually differentiate a Fourier series by:

$$\frac{d}{dx}\sum_{n=1}^{\infty} a_n \sin(n\pi x) = \sum_{n=1}^{\infty} a_n \frac{d}{dx}\sin(n\pi x) = \sum_{n=1}^{\infty} n\pi a_n \cos(n\pi x),$$

but this *is not always true* if we interpret "=" in the usual sense of being true for every x. The general problem is that *one cannot always interchange limits and derivatives*, i.e. $\frac{d}{dx}\lim \neq \lim \frac{d}{dx}$. (Note that $\sum_{n=1}^{\infty}$ is really a limit $\lim_{N \to \infty} \sum_{n=1}^{N}$.)

A simple example of this is the function

$$f_\epsilon(x) = \begin{cases} -\epsilon & x < -\epsilon \\ x & -\epsilon \leq x \leq \epsilon \\ \epsilon & x > \epsilon \end{cases}, \qquad f'_\epsilon(x) = \begin{cases} 1 & -\epsilon < x < \epsilon \\ 0 & |x| > \epsilon \end{cases}.$$

The limit as $\epsilon \to 0$ of $f_\epsilon(x)$ is simply zero. But $f'_\epsilon(0) = 1$ for all ϵ, so

$$0 = \frac{d}{dx}\left(\lim_{\epsilon \to 0} f_\epsilon\right)\bigg|_{x=0} \neq 1 = \lim_{\epsilon \to 0}\left(\frac{d}{dx}f_\epsilon\bigg|_{x=0}\right).$$

Notice, however, that this mismatch only occurs at one isolated point $x = 0$. The same thing happens for Fourier series: differentiation term-by-term works *except at isolated points*. Thus, this problem returns to the complaint in the previous section 1.3.

1.5 *Too much information*

A function $f(x)$ gives us a value at *every point x*, but does this really correspond to a measurable quantity

in the physical universe? How would you measure the velocity at one *instant* in time, or the density at one *point* in a fluid? Of course, your measurement device could be very precise, and very small, and very fast, but in the end *all you can ever measure are averages* of $f(x)$ over a small region of space and/or time.

But an average is the same thing as an integral $\int f(x)$ over the averaging region. More generally, instead of averaging $f(x)$ uniformly in some region, we could average with some weights $\phi(x)$ (e.g. our device could be more sensitive to some points than others). Thus, the *only physical question we can ever ask about a function is the value of an integral*

$$\int_{-\infty}^{\infty} f(x)\phi(x)dx$$

of $f(x)$ against a *test function* $\phi(x)$.

But if all we can ever ask is such an integral, why are we worrying about isolated points? In fact, why do we even *define* $f(x)$ to have values at points at all? In a physical application of mathematics, perhaps we should only define things that we can measure! This insight fixes *every one of the problems above*, and leads to the concept of **distributions**.

2 Distributions

The old kind of function is a map from $\mathbb{R} \to \mathbb{R}$: given an x, we get a value $f(x)$. Following section 1.5, this is too much information; we can only ask for an "average" value given some weight function $\phi(x)$. So, we make a new definition of "function" that provides this information, and *only* this information:

- f is a rule that given any *test function* $\phi(x)$ returns a *number* $f\{\phi\}$.[2]

This new definition of a "function" is called a **distribution** or a **generalized function**. We are no longer allowed to ask the value at a *point* x. This will fix all of the problems with the old functions from above. However, we should be more precise about our definition. First, we have to specify what $\phi(x)$ can be:

- $\phi(x)$ is an ordinary function $\mathbb{R} \to \mathbb{R}$ (not a distribution) in some set $\mathcal{D}$. We require $\phi(x)$ to be *infinitely differentiable*. We also require $\phi(x)$ to be *nonzero* only in some finite region (the "support" of ϕ): ϕ is a smooth "bump" function.[3]

[The generalization to functions $\phi(\vec{x})$ for $\vec{x} \in \mathbb{R}^d$, with the distributions corresponding to d-dimensional integrals, is very straightforward, but we will stick to $d = 1$ for simplicity.] Second,[4] we require that $f\{\phi\}$ act like integration in that it must be *linear*:

- $f\{\alpha\phi_1 + \beta\phi_2\} = \alpha f\{\phi_1\} + \beta f\{\phi_2\}$ for any numbers $\alpha, \beta \in \mathbb{R}$ and any $\phi_1, \phi_2 \in \mathcal{D}$.

Thus, f is a linear map from $\mathcal{D} \to \mathbb{R}$, and the set of all distributions for a given set of test functions $\mathcal{D}$ is sometimes denoted $\mathcal{D}'$. There are two classes of distributions: **regular** and **singular** distributions.

2.1 Regular distributions from ordinary functions $f(x)$

The most obvious way to define a distribution $f\{\phi\}$ is simply an ordinary integral of an ordinary function: given an ordinary function $f(x)$, we can define the distribution:

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

This is called a **regular distribution**.

Not all ordinary functions $f(x)$ define regular distributions: we must have $\int f\phi$ finite for all $\phi \in \mathcal{D}$. This reduces to requiring that $\int_a^b |f(x)|dx < \infty$ for all intervals $[a, b]$ (f is "locally integrable").

2.2 Singular distributions & the delta function

Although the integral of an ordinary function is *one* way to define a distribution, it is not the *only* way. For example, the following distribution is a perfectly good one:

$$\delta\{\phi\} = \phi(0).$$

This rule is linear and continuous, there are no weird infinities, nor is there anything spooky or non-rigorous. Given a test function $\phi(x)$, the $\delta(x)$ *distribution* is simply the rule that gives $\phi(0)$ from each ϕ. $\delta(x)$, however, does not correspond to any ordinary function—it is not a regular distribution—so we call it a **singular distribution**.

- Notation: when we write "$\int \delta(x)\phi(x)dx$", we don't mean an ordinary integral, we really mean $\delta\{\phi\} = \phi(0)$.

---

[2]Many authors use the notation $\langle f, \phi \rangle$ instead of $f\{\phi\}$, but we are already using $\langle \cdot, \cdot \rangle$ for inner products and I don't want to confuse matters.

[3]Alternatively, one sometimes loosens this requirement to merely say that $\phi(x)$ must $\to 0$ quickly as $x \to \pm\infty$, and in particular that $x^n \phi(x) \to 0$ as $x \to \pm\infty$ for all integers $n \geq 0$.

[4]There is also a third requirement: $f\{\phi\}$ must be continuous, in that if you change $\phi(x)$ continuously the value of $f\{\phi\}$ must change continuously. In practice, you will never violate this condition unless you are trying to.

Furthermore, when we say $\delta(x - x')$, we just mean the distribution $\delta(x - x')\{\phi\} = \phi(x')$.

If we look at the "finite-Δx delta" approximation $\delta_{\Delta x}(x)$ from section 1.1, that defines the *regular* distribution:

$$\delta_{\Delta x}\{\phi\} = \frac{1}{\Delta x} \int_0^{\Delta x} \phi(x) dx,$$

which is just the average of $\phi(x)$ in $[0, \Delta x]$. Now, however, viewed as a distribution, the limit $\Delta x \to 0$ is perfectly well defined:[5] $\lim_{\Delta x \to 0} \delta_{\Delta x}\{\phi\} = \phi(0) = \delta\{\phi\}$, i.e. $\delta_{\Delta x} \to \delta$.

Of course, the δ distribution is not the only singular distribution, but it is the most famous one (and the one from which many other singular distributions are built). We will see more examples later.

2.3 Derivatives of distributions & differentiating discontinuities

How do we define the derivative f' of a distribution? Well, at the very least we want it to be the same as the ordinary derivative when f is a regular distribution $f(x)$ that happens to be differentiable in the ordinary sense. In that case, $f'\{\phi\} = \int f'(x)\phi(x) dx = -\int f(x)\phi'(x) dx = f\{-\phi'\}$, where we have integrated by parts and used the fact that $\phi(x)$ is zero outside a finite region to eliminate the boundary terms. This is such a nice result that we will use it to *define* the derivative of *any* distribution:

- The **distributional derivative** f' of $f\{\phi\}$ is given by the distribution $f'\{\phi\} = f\{-\phi'\}$, where $\phi'(x)$ is the ordinary derivative of $\phi(x)$.

(This is sometimes also called a **weak derivative**.) Since the test functions $\phi(x)$ were required to be infinitely differentiable, we have a remarkable consequence: *every distribution is infinitely differentiable* (in the distributional sense).

For example, since $\delta\{\phi\} = \phi(0)$, it immediately follows that the derivative of a delta function is the distribution $\delta'\{\phi\} = \delta\{-\phi'\} = -\phi'(0)$.

The most important consequence of this definition is that even **discontinuous functions are differentiable** as distributions, and their derivatives **give delta functions** for each discontinuity. Consider the regular distribution S defined by the step function $S(x)$:

$$S\{\phi\} = \int_{-\infty}^{\infty} S(x)\phi(x) dx = \int_0^{\infty} \phi(x) dx.$$

It immediately follows that the distributional derivative of the step function is

$$\begin{aligned} S'\{\phi\} &= S\{-\phi'\} = -\int_0^{\infty} \phi'(x) dx = -\phi(x)\big|_0^{\infty} \\ &= \phi(0) - \phi(\infty) = \phi(0). \end{aligned}$$

But this is exactly the same as $\delta\{\phi\}$, so we immediately conclude: $\boxed{S' = \delta}$.

Since any function with jump discontinuities can be written in terms of $S(x)$, we find that the derivative of any jump discontinuity gives a delta function multiplied by the magnitude of the jump. For example, consider:

$$f(x) = \begin{cases} x^2 & x < 3 \\ x^3 & x \geq 3 \end{cases} = x^2 + (x^3 - x^2)S(x - 3).$$

The distributional derivative works just like the ordinary derivative, except that $S' = \delta$, so[6]

$$f'(x) = 2x + (3x^2 - 2x)S(x - 3) + (3^3 - 3^2)\delta(x - 3)$$

$$= 18\delta(x - 3) + \begin{cases} 2x & x < 3 \\ 3x^2 & x \geq 3 \end{cases},$$

where of course by $f'(x)$ I mean the distribution $f'\{\phi\}$. It is common to be "casual" with notation in this way for distributions, treating them like ordinary functions, but you have to remember that you can't evaluate them at any point x, you can only evaluate them for test functions $\phi(x)$.

2.4 Isolated points

With ordinary functions, we had to make lots of caveats about isolated points. No more with distributions. The key point is that two *different* ordinary functions can define the *same* distribution. Consider, for example, the function

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

This is *not* a delta function: it is finite at $x = 0$, and is a perfectly acceptable function. It also defines a regular distribution:

$$f\{\phi\} = \int_{-\infty}^{\infty} f(x)\phi(x) dx = 0.$$

[5]We have used the fact that $\phi(x)$ is required to be continuous, from which one can show that nothing weird can happen with the average of $\phi(x)$ as $\Delta x \to 0$.

[6]I'm being a bit glib here. How do we know that the product rule works the same? Below, we will rigorously define what it means to multiply a distribution by a smooth function like $x^3 - x^2$, and the ordinary product rule will follow.

Think about it: no matter what $\phi(x)$ is, the integral must give zero because it is only nonzero (by a *finite* amount) at a single point, with zero area ("zero measure" for the pure-math folks). Thus, in the distribution sense, we can say perfectly rigorously that

$$f = 0$$

even though $f(x) \neq 0$ in the ordinary-function sense!

In general, any two ordinary functions that only differ (by *finite* amounts—not delta functions!) at *isolated points* (a "set of measure zero") define the *same* regular distribution. We no longer have to make caveats about isolated points—finite values at isolated points make *no difference* to a distribution.

For example, there are no more caveats about the Fourier series or Fourier transforms: they converge, period, for distributions.[7]

Also, there is no more quibbling about the value of things like $S(x)$ right at the point of discontinuities. It doesn't matter, for distributions. Nor is there quibbling about the derivative of things like $|x|$ right at the point of the slope discontinuity. It is an easy matter to show that the distributional derivative of $|x|$ is simply $2S(x) - 1$, i.e. it is the regular distribution corresponding to the function that is $+1$ for $x > 0$ and -1 for $x < 0$ (with the value at $x = 0$ being rigorously irrelevant).

2.5 Interchanging limits and derivatives

With a distribution, limits and (distributional) derivatives can always be interchanged. This is tremendously useful when talking about PDEs and convergence of approximations. In the distribution sense, the Fourier series can *always* be differentiated term-by-term, for example.

This is easy to prove. Suppose that the distributions $f_n \to f$ as $n \to \infty$. That is, for any $\phi(x)$, $f_n\{\phi\} \to f\{\phi\}$. Since this is true for *any* $\phi(x)$, it must be true for $-\phi'(x)$, and hence $f_n'\{\phi\} = f_n\{-\phi'\} \to f\{-\phi'\} = f'\{\phi\}$ as $n \to \infty$. Q.E.D.

3 Rules for distributions

For the most part, in 18.303, we will cheat a bit. We will treat things as ordinary functions whenever we can, using the ordinary operations, and only

switch to interpreting them as distributions when we run into difficulty (e.g. derivatives of discontinuities, etcetera). Since the rules for distribution operations are all defined to be consistent with those for functions in the case of regular distributions, this doesn't usually cause any trouble.

However, it is good to define some of the important operations on distributions precisely. All you have to do is to explain what the operation does to test functions, usually defined by analogy with $\int f(x)\phi(x)dx$ for regular distributions. Here are a few of the most basic operations:

- differentiation: $f'\{\phi\} = f\{-\phi\}$

- addition: $(f_1 + f_2)\{\phi\} = f_1\{\phi\} + f_2\{\phi\}$

- multiplication by smooth functions (including constants) $s(x)$: $[s(x) \cdot f]\{\phi\} = f\{s(x)\phi(x)\}$

- product rule for multiplication by smooth functions $s(x)$: $[s \cdot f]'\{\phi\} = [s \cdot f]\{-\phi'\} = f\{-s\phi'\} = f\{s'\phi - (s\phi)'\} = f\{s'\phi\} + f'\{s\phi\} = [s' \cdot f]\{\phi\} + [s \cdot f']\{\phi\} = [s' \cdot f + s \cdot f']\{\phi\}$.

- translation: $[f(x-y)]\{\phi(x)\} = f\{\phi(x+y)\}$

- scaling $[f(\alpha x)]\{\phi(x)\} = \frac{1}{\alpha}f\{\phi(x/\alpha)\}$

If you are not sure where these rules come from, just try plugging them into a regular distribution $f\{\phi\} = \int f(x)\phi(x)dx$, and you'll see that they work out in the ordinary way.

4 Problems with distributions

Unfortunately, distributions are not a free lunch; they come with their own headaches. There are two major difficulties, one of which is surmountable and the other is not:

- Boundary conditions: since distributions do not have values at individual points, it is not so easy to impose boundary conditions on the solutions if they are viewed as distributions—what does it mean to set $u(0) = 0$? There are ways around this, but they are a bit cumbersome, especially in more than one dimension.

- Multiplication: it is not generally meaningful to multiply distributions. The simplest example is the delta function: what would $\delta(x)^2$ be? $\delta_{\Delta x}(x)^2$ is okay, but its limit as $\Delta x \to 0$ does not exist even as a distribution (the amplitude goes as $1/\Delta x^2$ while the integral goes as Δx, so it diverges).

---

[7]Technically, we have to choose distributions with the right set of test functions. The right test functions for the general Fourier transform on the real line are those for which $x^n\phi(x) \to 0$ as $x \to \pm\infty$ for any $n > 0$, i.e. $\phi(x)$ vanishes faster than any polynomial. The resulting distributions are called *tempered distributions*, and are the domain of the Fourier transform.

For linear PDEs, lack of multiplication is not such a big problem, but it does mean that we need to be careful about Hilbert spaces: if we think of the solutions $u(x)$ as distributions, we have a problem because $\langle u, u \rangle$ may not be defined—the set of distributions does not form a Hilbert space. (Technically, we can make something called a *rigged Hilbert space* that includes distributions, but I don't want to go there.)

5 The weak form of a PDE

Suppose we have a linear PDE $\hat{A}u = f$. We want to allow f to be a delta function etcetera, but we still want to talk about boundary conditions, Hilbert spaces, and so on for u. There is a relatively simple compromise for linear PDEs, called the **weak form** of the PDE or the **weak solution**. This concept can roughly be described as requiring $\hat{A}u = f$ only in the "weak" sense of distributions (i.e., integrated against test functions, taking distributional derivatives in the case of a u with discontinuities), but requiring u to be in a more restrictive class, e.g. regular distributions corresponding to functions satisfying the boundary conditions in the ordinary sense but having some continuity and differentiability so that $\langle u, u \rangle$ and $\langle u, \hat{A}u \rangle$ are finite (i.e., living in an appropriate Sobolev space). Even this definition gets rather technical very quickly, especially if you want to allow delta functions for f (in which case u can blow up at the point of the delta for dimensions > 1). Nailing down precisely what spaces of functions and operators one is dealing with is where a lot of technical difficulty and obscurity arises in functional analysis. However, for practical engineering and science applications, we can get away with being a little careless in precisely how we define the function spaces because the weird counterexamples are usually obviously unphysical. The key insight of distributions is that what we care about is weak equality, not pointwise equality, and correspondingly we only need weak derivatives (integration by parts, or technically "bilinear forms").

6 Green's functions

Now that we have distributions, Green's functions are much easier to work with. Consider, for example, the Green's function $G(x, x')$ of $\hat{A} = -\frac{\partial^2}{\partial x^2}$ on $[0, L]$ with Dirichlet (zero) boundary conditions. If $\hat{A}u = f$ is to be solved by $u(x) = \int G(x, x')f(x')dx'$, then we must have $\hat{A}u = \int [\hat{A}G(x, x')]f(x')dx' = f(x)$, which

is true if

$$\hat{A}G(x, x') = \delta(x - x')$$

and the integrals are re-interpreted as evaluating a distribution.

What does this equation mean? For any $x \neq x'$ [or for any $\phi(x)$ with $\phi(x') = 0$] we must have $\hat{A}G(x, x') = 0 = -\frac{\partial^2}{\partial x^2}G(x, x')$, and this must mean that $G(x, x')$ is a straight line for $x < x'$ and $x > x'$. To satisfy, the boundary conditions, this straight line must pass through zero at 0 and L, and hence $G(x, x')$ must look like αx for $x < x'$ and $\beta(x - L)$ for $x > x'$ for some constants α and β.

$G(x, x')$ had better be continuous at $x = x'$, or otherwise we would get a delta function from the *first* derivative—hence $\alpha = \beta(x' - L)/x'$. The first derivative $\frac{\partial}{\partial x}G(x, x')$ then gives α for $x < x'$ and β for $x > x'$. What about the next derivative? Since $\frac{\partial}{\partial x}G(x, x')$ is discontinuous, it doesn't have an ordinary second derivative at $x = x'$, but as a distribution it is no problem: $\frac{\partial^2}{\partial x^2}G(x, x')$ is zero everywhere (the derivative of a constant) plus a delta function $\delta(x - x')$ multiplied by $\beta - \alpha$, the size of the jump. Thus, $-\frac{\partial^2}{\partial x^2}G(x, x') = \hat{A}G = (\alpha - \beta)\delta(x - x')$, and from above we must have $\alpha - \beta = 1$. Combined with the equation for α from continuity of G, we obtain $\beta = -x'/L$ and $\alpha = 1 - x'/L$, exactly the same as our result from class (which we got by a more laborious method).

Further reading

- I. M. Gel'fand and G. E. Shilov, *Generalized Functions, Volume I: Properties and Operations* (New York: Academic Press, 1964). [Out of print, but still my favorite book on the subject.]

- Robert S. Strichartz, *A Guide to Distribution Theory and Fourier Transforms* (Singapore: World Scientific, 1994).

- Jesper Lutzen, *The Prehistory of the Theory of Distributions* (New York: Springer, 1982). [A fascinating book describing the painful historical process that led up to distribution theory.]

Distributions in a nutshell

- Delta functions are okay. You can employ their informal description without guilt because there is a rigorous definition to fall back on in case of doubt.

    - An "integral" $\int_{-\infty}^{\infty} \delta(x - x')\phi(x)dx$ just means $\phi(x')$. Integrals over finite domains $\int_{\Omega} \delta(x - x')\phi(x)$ give $\phi(x')$ if x' is in the interior of Ω and 0 if x' is outside Ω, but are undefined (or at least, more care is required) if x' is on the boundary $d\Omega$.

- When in doubt about how to compute $f'(x)$, integrate by parts against a test function to see what $\int f'(x)\phi(x) = -\int f(x)\phi'(x)$ does (the "weak" or "distributional" derivative).

    - A derivative of a discontinuity at x' gives $\delta(x-x')$ multiplied by the size of the discontinuity [the difference $f(x'^{+}) - f(x'^{-})$], plus the ordinary derivative everywhere else.

    - This also applies to differentiating functions like $1/\sqrt{x}$ that have finite integrals but whose classical derivatives have divergent integrals—applying the weak derivative instead produces a well-defined distribution. [For example, this procedure famously yields $\nabla^2 \frac{1}{r} = -4\pi\delta(\vec{x})$ in 3d.]

- All that matters in the distribution (weak) sense is the integral of a function times a smooth, localized test function $\phi(x)$. Anything that doesn't change such integrals $\int f(x)\phi(x)$, like finite values of $f(x)$ at isolated points, doesn't matter. (That is, whenever we use "=" for functions we are almost always talking about *weak* equality.)

- Interchanging limits and derivatives is okay in the distribution sense. Differentiating Fourier series (and other expansions in infinite bases) term-by-term is okay.

- In practice, we only ever need to solve PDEs in the distribution sense (a "weak" solution): integrating the left- and right-hand sides against any test functions must give the same number, with all derivatives taken in the weak sense.

18.303 Linear Partial Differential Equations: Analysis and Numerics
Fall 2014