Notes on Distributions

Peter Woit
Department of Mathematics, Columbia University
woit@math.columbia.edu

April 7, 2020

1 Introduction

This is a set of notes written for the 2020 spring semester Fourier Analysis class, covering material on distributions. This is an important topic not covered in Stein-Shakarchi. These notes will supplement two textbooks you can consult for more details:

- *A Guide to Distribution Theory and Fourier Transforms* [2], by Robert Strichartz. The discussion of distributions in this book is quite comprehensive, and at roughly the same level of rigor as this course. Much of the motivating material comes from physics.
- *Lectures on the Fourier Transform and its Applications* [1], by Brad Osgood, chapters 4 and 5. This book is wordier than Strichartz, has a wealth of pictures and motivational examples from the field of signal processing. The first three chapters provide an excellent review of the material we have covered so far, in a form more accessible than Stein-Shakarchi.

One should think of distributions as mathematical objects generalizing the notion of a function (and the term "generalized function" is often used interchangeably with "distribution"). A function will give one a distribution, but many important examples of distributions do not come from a function in this way. Some of the properties of distributions that make them highly useful are:

- Distributions are always infinitely differentiable, one never needs to worry about whether derivatives exist. One can look for solutions of differential equations that are distributions, and for many examples of differential equations the simplest solutions are given by distributions.
- So far we have only defined the Fourier transform for a very limited class of functions ($\mathcal{S}(\mathbf{R})$, the Schwartz functions). Allowing the Fourier transform to be a distribution provides a definition that makes sense for a wide variety of functions. The definition of the Fourier transform can be extended so that it takes distributions to distributions.

- When studying various kernel functions, we have often been interested in what they do as a parameter is taken to a limit. These limits are not functions themselves, but can be interpreted as distributions. For example, consider the heat kernel on $\mathbf{R}$

$$H_{t,\mathbf{R}}(x) = \frac{1}{\sqrt{4\pi t}} e^{-\frac{x^2}{4t}}$$

  Its limiting value as $t \to 0^+$ is not a function, but it would be highly convenient to be able to treat it as one. This limit is a distribution, often called the "δ-function", and this turns out to have a wide range of applications. It was widely used among engineers and physicists before the rigorous mathematical theory that we will study was developed in the middle of the twentieth century.

2 Distributions: definition

The basic idea of the theory of distributions is to work not with a space of functions, but with the dual of such a space, the linear functionals on the space. The definition of a distribution is thus:

Definition (Distributions)**.** *A distribution is a continuous linear map*

$$\varphi \to \langle f, \varphi \rangle \in \mathbf{C}$$

for φ in some chosen class of functions (the "test functions").

To make sense of this definition we need first to specify the space of test functions. There are two standard choices:

- The space $\mathcal{D}(\mathbf{R})$ of smooth (infinitely differentiable) functions that vanish outside some bounded set.
- The space $\mathcal{S}(\mathbf{R})$ of Schwartz functions that we have previously studied.

In this course we will use $\mathcal{S}(\mathbf{R})$ as our test functions, and denote the space of distributions as $\mathcal{S}'(\mathbf{R})$. Distributions with this choice of test functions are conventionally called "tempered" distributions, but we won't use that terminology since our distributions will always be tempered distributions. The advantage of this choice will be that one can define the Fourier transform of an element of $\mathcal{S}'(\mathbf{R})$ so it also is in $\mathcal{S}'(\mathbf{R})$ (which would not have worked for the choice $\mathcal{D}(\mathbf{R})$). The definition of $\mathcal{S}(\mathbf{R})$ was motivated largely by this property as a space of test functions.

The tricky part of our definition of a distribution is that we have not specified what it means for a linear functional on the space $\mathcal{S}(\mathbf{R})$ to be continuous. For a vector space V, continuity of a linear map $f : V \to \mathbf{C}$ means that we can make

$$|f(v_2) - f(v_1)| = |f(v_2 - v_1)|$$

arbitrarily small by taking $v_2 - v_1$ arbitrarily close to zero, but this requires specifying an appropriate notion of the size of an element $v \in V$, where $V = \mathcal{S}(\mathbf{R})$. We need a notion of size of a function which takes into account the size of the derivatives of the function and the behavior at infinity. Roughly, one wants to say that $\varphi \in \mathcal{S}(\mathbf{R})$ is of size less than ϵ if for all x, k, l

$$|x^k(\frac{d}{dx})^l\varphi(x)| < \epsilon$$

For a more detailed treatment of this issue, see section 6.5 of [2]. We will take much the same attitude as Strichartz: when we define a distribution as a linear functional, it is not necessary to check continuity, since it is hard to come up with linear functionals that are not continuous.

3 Distributions: examples

For any space of functions on $\mathbf{R}$, we can associate to a function f in this space the linear functional

$$\varphi \in \mathcal{S}(\mathbf{R}) \rightarrow \langle f, \varphi \rangle = \int_{-\infty}^{\infty} f\varphi dx$$

as long as this integral is well-defined. We will be using the same notation "f" for the function and for the corresponding distribution. This works for example for $f \in \mathcal{S}(\mathbf{R})$, so

$$\mathcal{S}(\mathbf{R}) \subset \mathcal{S}'(\mathbf{R})$$

It also works for the space $\mathcal{M}(\mathbf{R})$ of moderate decrease functions used by Stein-Shakarchi, and even for spaces of functions that do not fall off at infinity, as long as multiplication by a Schwartz function causes sufficient fall off at infinity for the integral to converge (this will be true for instance for functions with polynomial growth).

An example of a function that does not fall off at $\pm\infty$ is

$$f(x) = e^{-2\pi i x}$$

The corresponding distribution is the linear functional on Schwartz functions that evaluates the Fourier transform of the function at $p = 0$

$$\varphi \rightarrow \langle f, \varphi \rangle = \int_{-\infty}^{\infty} e^{-2\pi i p x}\varphi(x)dx = \widehat{\varphi}(p)$$

For an example of a function that does not give an element of $\mathcal{S}'(\mathbf{R})$, consider

$$f(x) = e^{x^2}$$

If you integrate this against the Schwartz function

$$\varphi(x) = e^{-x^2}$$

you would get

$$\langle f, \varphi \rangle = \int_{-\infty}^{\infty} e^{x^2} e^{-x^2} dx = \int_{-\infty}^{\infty} 1 dx$$

which is not finite. To make sense of functions like this as distributions, one needs to use as test functions not $\mathcal{S}(\mathbf{R})$, but $\mathcal{D}(\mathbf{R})$, which gives a larger space of distributions $\mathcal{D}'(\mathbf{R})$ than the tempered distributions $\mathcal{S}'(\mathbf{R})$. Since functions in $\mathcal{D}(\mathbf{R})$ are identically zero if one goes far enough out towards $\pm\infty$, the integral will exist no matter what growth properties at infinity f has.

Now for some examples of distributions that are not functions:

- *The Dirac "delta function" at* x_0: this is the linear functional that evaluates a function at $x = x_0$

  $$\delta_{x_0} : \varphi \to \langle \delta_{x_0}, \varphi \rangle = \varphi(x_0)$$

  An alternate notation for this distribution is $\delta(x - x_0)$, and the special case $x_0 = 0$ will sometimes be written as δ or $\delta(x)$. Although this is a distribution and not a function, we will use the terminology "delta function" to refer to it. The delta function will often be manipulated as if it were a usual sort of function, one just has to make sure that when this is done there is a sensible interpretation in the language of distributions. For instance, one often writes $\langle \delta_{x_0}, \varphi \rangle$ as an integral, as in

  $$\int_{-\infty}^{\infty} \delta(x - x_0)\varphi(x)dx = \varphi(x_0)$$

- *Limits of functions*: many limits of functions are not functions but are distributions. For example, consider the limiting behavior of the heat kernel as $t > 0$ goes to 0

  $$\lim_{t \to 0^+} H_{t,\mathbf{R}}(x) = \lim_{t \to 0^+} \frac{1}{\sqrt{4\pi t}} e^{-\frac{x^2}{4t}}$$

  We have seen that $H_{t,\mathbf{R}}$ is a good kernel, so that

  $$\begin{aligned}\lim_{t \to 0^+} (\varphi * H_{t,\mathbf{R}})(x_0) &= \lim_{t \to 0^+} \int_{-\infty}^{\infty} \varphi(x) \frac{1}{\sqrt{4\pi t}} e^{-\frac{(x-x_0)^2}{4t}} dx \\ &= \varphi(x_0)\end{aligned}$$

  whiich shows that, as distributions

  $$\lim_{t \to 0^+} H_{t,\mathbf{R}}(x - x_0) = \delta(x - x_0)$$

  Another example of the same sort is the limit of the Poisson kernel used to find harmonic functions on the upper half plane:

  $$\lim_{y \to 0^+} P_y(x) = \lim_{y \to 0^+} \frac{1}{\pi} \frac{y}{x^2 + y^2}$$

which satisfies

$$\lim_{y\to 0^+} (\varphi * P_y)(x_0) = \varphi(x_0)$$

so can also be identified with the delta function distribution

$$\lim_{y\to 0^+} P_y(x - x_0) = \delta(x - x_0)$$

- *Dirac comb*: for a periodic version of the Dirac delta function, on can define the "Dirac comb" by

$$\delta_{\mathbf{Z}} : \varphi \to \langle \delta_{\mathbf{Z}}, \varphi \rangle = \sum_{n=-\infty}^{\infty} \varphi(n)$$

  This is the limiting value of the Dirichlet kernel (allow x to take values on all of $\mathbf{R}$, and rescale to get periodicity 1)

$$\lim_{N\to\infty} D_N(x) = \lim_{N\to\infty} \sum_{n=-N}^{N} e^{2\pi i n x} = \delta_{\mathbf{Z}}$$

- *Principal value of a function*: given a function f that is singular for some value of x, say $x = 0$, one can try to use it as a distribution by defining the integral as a "principal value integral". One defines the principal value distribution for f by

$$PV(f) : \varphi \to \langle PV(f), \varphi \rangle = \lim_{\epsilon\to 0^+} \int_{|x|>\epsilon} f(x)\varphi(x)dx$$

  This gives something sensible in many cases, for example for the case $f = \frac{1}{x}$

$$\langle PV\left(\frac{1}{x}\right), \varphi \rangle = \lim_{\epsilon\to 0^+} \int_{|x|>\epsilon} \frac{1}{x}\varphi(x)dx$$

  Here the integral gives a finite answer since it is finite for regions away from $x = 0$, and for the region $[-1, 1]$ including $x = 0$ one has

$$\lim_{\epsilon\to 0^+} (\int_{-1}^{-\epsilon} \frac{1}{x}\varphi(x)dx + \int_{\epsilon}^{1} \frac{1}{x}\varphi(x)dx) = \lim_{\epsilon\to 0^+} \int_{\epsilon}^{1} \frac{1}{x}(\varphi(x) - \varphi(-x))dx$$

  which is well-defined since

$$\lim_{x\to 0} \frac{\varphi(x) - \varphi(-x)}{x} = 2\varphi'(0)$$

- *Limits of functions of a complex variable*: Another way to turn $\frac{1}{x}$ into a distribution is to treat x as a complex variable and consider

$$\frac{1}{x + i0} \equiv \lim_{\epsilon\to 0^+} \frac{1}{x + i\epsilon}$$

Here we have moved the pole in the complex plane from 0 to $-i\epsilon$, allowing for a well-defined integral along the real axis. We will show later that this gives a distribution which is a combination of our previous examples:

$$\frac{1}{x+i0} = PV\left(\frac{1}{x}\right) - i\pi\delta(x)$$

Note that there is also a complex conjugate distribution

$$\frac{1}{x-i0} = PV\left(\frac{1}{x}\right) + i\pi\delta(x)$$

and one has

$$\delta(x) = \frac{1}{2\pi i}\left(\frac{1}{x-i0} - \frac{1}{x+i0}\right)$$

Since

$$\frac{1}{2\pi i}\left(\frac{1}{x-i\epsilon} - \frac{1}{x+i\epsilon}\right) = \frac{1}{\pi}\frac{\epsilon}{x^2+\epsilon^2}$$

this is consistent with our previous identification of the delta function with the limit of the Poisson kernel.

4 The derivative of a distribution

Many of the operators that one is used to applying to functions can also be applied to distributions. In this section we'll define the derivative of a distribution, but will begin by considering more general linear operators.

4.1 The transpose of a linear transformation

As a general principle from linear algebra, given a linear transformation T acting on a vector space V, one can define a "transpose" map T^t that acts as a linear transformation on the dual vector space V^* (warning: this can be a bit confusing since we are considering linear transformations of a space of linear maps). Here V^* is the vector space of linear maps $V \to \mathbf{C}$. If we have

$$T : v \in V \to Tv \in V$$

then the transpose operator

$$T^t : l \in V^* \to T^t l \in V^*$$

is defined by taking $T^t l$ to be the linear map given by

$$(T^t l)v = l(Tv)$$

For the case of V a finite dimensional vector space with a basis, so V can be identified with column vectors, then V^* can be identified with row vectors. A linear transformation is then given by a matrix, and the transpose transformation is given by the transpose matrix.

Applying this to the infinite dimensional vector space $V = \mathcal{S}(\mathbf{R})$, for any linear transformation T on $\mathcal{S}(\mathbf{R})$ we can define a transpose linear functional T^t taking a distribution $f \in \mathcal{S}'(\mathbf{R})$ to a new one $T^t f$ defined by

$$\langle T^t f, \varphi \rangle = \langle f, T\varphi \rangle$$

(here $\varphi \in \mathcal{S}(\mathbf{R})$).

A simple example of a linear transformation on $\mathcal{S}(\mathbf{R})$ is multiplication by a function ψ (one needs ψ to be in some class of functions such that $\psi\varphi \in \mathcal{S}(\mathbf{R})$ when $\varphi \in \mathcal{S}(\mathbf{R})$, which we won't try to otherwise characterize). Calling this linear transformation M_ψ one can multiply distributions f by functions ψ, with the result $M_\psi^t f$ given by

$$\langle M_\psi^t f, \varphi \rangle = \langle f, \psi\varphi \rangle$$

which we will just write as ψf. Writing distributions as integrals, this just says that

$$\int_{-\infty}^{\infty} (\psi f)\varphi dx = \int_{-\infty}^{\infty} f(\psi\varphi) dx$$

4.2 Translations

An important linear transformation that acts on functions, in particular Schwartz functions, is translation by a constant a:

$$(T_\tau \varphi)(x) = \varphi(x + a)$$

The transpose transformation on distributions is given by

$$\langle T_a^t f, \varphi \rangle = \langle f, T_a \varphi \rangle$$

If f is actually a function, then

Claim. *For $f \in \mathcal{S}'(\mathbf{R})$ a function, i.e.*

$$\langle f, \varphi \rangle = \int_{-\infty}^{\infty} f\varphi \; dx$$

$T_a^t f$ is also a function, given by

$$T_a^t f(x) = f(x - a)$$

Proof.

$$\begin{aligned} \langle T_\tau^t f, \varphi \rangle &= \langle f, T_\tau \varphi \rangle \\ &= \int_{-\infty}^{\infty} f(x)\varphi(x + \tau) dx \\ &= \int_{-\infty}^{\infty} f(x - \tau)\varphi(x) dx \end{aligned}$$

□

This shows that there's a rather confusing sign to keep track of, and that if one wants the definition of a translation on a distribution to match the definition of translation on a function one should define translations on distributions by

$$T_a f = T^t_{-a} f$$

so

$$\langle T_a f, \varphi\rangle = \langle f, T_{-a}\varphi\rangle$$

For an example of how this works for a distribution that is not a function, the translation of a delta function is given by

$$\langle T_a \delta, \varphi\rangle = \langle \delta, T_{-a}\varphi\rangle = \varphi(0-a) = \varphi(-a)$$

so

$$T_a\delta = \delta_{-a}$$

4.3 The derivative

The derivative is also a linear operator, the infinitesimal version of a translation, given by

$$\frac{d}{dx} = \lim_{a\to 0}\frac{T_a - I}{a}$$

where I is the identity operator. On distributions, because of the sign issue noted above, infinitesimal translation on distributions should be defined by

Definition (Derivative of a distribution)**.** *The derivative of the distribution* $f \in \mathcal{S}'\mathbf{R}$ *is the linear functional*

$$\frac{df}{dx} : \varphi \to \langle \frac{df}{dx}, \varphi\rangle = \langle f, -\frac{d}{dx}\varphi\rangle$$

This definition is consistent with the usual derivative when f is a function, since, using integration by parts

$$\int_{-\infty}^{\infty}(\frac{d}{dx}f)\varphi dx = \int_{-\infty}^{\infty}\frac{d}{dx}(f\varphi)dx - \int_{-\infty}^{\infty} f\frac{d}{dx}\varphi dx$$

and the first term on the right hand side vanishes since $f\varphi$ goes to zero at $\pm\infty$ since $\varphi \in \mathcal{S}(\mathbf{R})$.

Note something quite remarkable about this definition: distributions are always infinitely differentiable (since Schwartz functions are). In particular, this allows one to often treat functions with discontinuities as distributions, with all derivatives well-defined.

As an example consider the Heaviside or step function

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

This is not a continuous or differentiable function, but can be interpreted as the distribution H such that

$$\langle H, \varphi \rangle = \int_{-\infty}^{\infty} H(x)\varphi(x)dx = \int_0^{\infty} \varphi(x)dx$$

One can compute the derivative of H as a distribution, finding (shifting to the "prime" notation for differentiation):

Claim.

$$H' = \delta$$

Proof.

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

□

The delta function distribution δ also has a derivative, δ', given by

$$\langle \delta', \varphi \rangle = \langle \delta, -\varphi' \rangle = -\varphi(0)$$

One can keep taking derivatives, each time getting the evaluation at 0 functional for a higher derivative, with an alternating sign:

$$\langle \delta^{(n)}, \varphi \rangle = (-1)^n \varphi^{(n)}(0)$$

More generally, for delta functions supported at x_0 one has

$$\langle \delta^{(n)}(x - x_0), \varphi \rangle = (-1)^n \varphi^{(n)}(x_0)$$

For another sort of example, one can interpret the function (singular at $x = 0$) $\ln|x|$ as a distribution (see pages 289-90 of [1] for an argument that the integrals needed to make sense of this as a distribution exist), and then compute its derivative, finding:

Claim.

$$\frac{d}{dx} \ln|x| = PV\left(\frac{1}{x}\right)$$

Proof. Using

$$\frac{d}{dx}(\varphi(x) \ln|x|) = \varphi'(x) \ln|x| + \frac{\varphi(x)}{x}$$

and integration by parts we have

$$\begin{aligned}\int_{\epsilon}^{\infty} \frac{\varphi(x)}{x} dx &= \varphi(x) \ln|x| \Big|_{\epsilon}^{\infty} - \int_{\epsilon}^{\infty} \varphi'(x) \ln|x| \\ &= -\varphi(\epsilon) \ln \epsilon - \int_{\epsilon}^{\infty} \varphi'(x) \ln|x|\end{aligned}$$

and similarly

$$\int_{-\infty}^{-\epsilon} \frac{\varphi(x)}{x} dx = \varphi(x) \ln|x| \Big|_{-\infty}^{-\epsilon} - \int_{-\infty}^{-\epsilon} \varphi'(x) \ln|x| dx$$
$$= \varphi(-\epsilon) \ln \epsilon - \int_{-\infty}^{-\epsilon} \varphi'(x) \ln|x| dx$$

so

$$\langle PV\left(\frac{1}{x}\right), \varphi\rangle = \lim_{\epsilon\to 0^+} \left((\varphi(-\epsilon) - \varphi(\epsilon)) \ln \epsilon - \int_{-\infty}^{-\epsilon} \varphi'(x) \ln|x| dx - \int_{\epsilon}^{\infty} \varphi'(x) \ln|x| dx \right)$$

But one can define

$$\psi(x) = \int_0^1 \varphi'(xt) dt = \frac{\varphi(x) - \varphi(0)}{x}$$

and use this to show that

$$\varphi(x) = \varphi(0) + x\psi(x)$$

and $\psi(x)$ is non-singular. Then

$$\langle PV\left(\frac{1}{x}\right), \varphi\rangle = \lim_{\epsilon\to 0^+} \left(-2\psi(0)\epsilon \ln \epsilon - \int_{-\infty}^{-\epsilon} \varphi'(x) \ln|x| dx - \int_{\epsilon}^{\infty} \varphi'(x) \ln|x| dx \right)$$
$$= - \int_{\infty}^{\infty} \varphi'(x) \ln|x| dx$$
$$= \langle \frac{d}{dx} \ln|x|, \varphi\rangle$$

□

One can extend the definition of ln to the complex plane, by taking $z = re^{i\theta}$, and

$$\ln z = \ln r + i\theta$$

The problem with this definition is that θ is only defined up to multiples of 2π. One can make $\ln z$ well-defined by taking $-\pi < \theta < \pi$, although then ln has a jump discontinuity at $\theta = \pm\pi$. For real values x, both positive and negative, one has

$$\lim_{\eta\to 0^+} \ln(x \pm i\eta) = \ln|x| \pm i\pi H(-x)$$

As a distribution, this has derivative

$$\frac{d}{dx} \lim_{\eta\to 0^+} \ln(x \pm i\eta) = PV\left(\frac{1}{x}\right) \mp i\pi\delta(x)$$

This is one way of justifying the claim made earlier in the examples of distributions that

$$\lim_{\eta\to 0^+} \frac{1}{x \pm i\eta} = PV\left(\frac{1}{x}\right) \mp i\pi\delta(x)$$

There are other approaches to justifying this formula. In particular one can use contour integration methods and the residue theorem from complex analysis, or use

$$\frac{1}{x+i\eta}=\frac{x}{x^2+\eta^2}\mp i\frac{\eta}{x^2+\eta^2}$$

Integrating this against a Schwartz function and taking the limit as $\eta\rightarrow 0^+$, the first term gives $PV\left(\frac{1}{x}\right)$, the second gives $\mp i\pi\delta(x)$.

5 The Fourier transform of a distribution

Warning: while chapter 5 of [2] contains a very good discussion of the Fourier transform of a distribution, it uses a different normalization of the Fourier transform than the one we are using. In that book, the normalization and notation are:

$$\mathcal{F}\varphi=\widehat{\varphi}(p)=\int_{-\infty}^{\infty}e^{ipx}dx$$

$$\mathcal{F}^{-1}\varphi=\check{\varphi}(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-ipx}dp$$

The difference in normalization is just that Strichartz's p is 2π times our p.

Since the Fourier transform operator $\mathcal{F}$ is a linear operator on $\mathcal{S}(\mathbf{R})$, we can define the Fourier transform on distributions as the transpose operator, and will use the same symbol $\mathcal{F}$:

Definition (Fourier transform of a distribution)**.** *The Fourier transform of a distribution $f\in\mathcal{S}'(\mathbf{R})$ is the distribution*

$$\mathcal{F}f:\varphi\rightarrow\langle\mathcal{F}f,\varphi\rangle=\langle f,\mathcal{F}\varphi\rangle$$

This definition makes sense precisely because the space of Schwartz functions $\mathcal{S}(\mathbf{R})$ was defined so that $\varphi\in\mathcal{S}(\mathbf{R})$ implies $\mathcal{F}\varphi\in\mathcal{S}(\mathbf{R})$.

Turning to examples, let's first compute the Fourier transform of the delta function:

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

So the Fourier transform of the delta function supported at 0 is the constant function 1.

$$\mathcal{F}\delta=1$$

Note that the Fourier transform here takes a distribution that is not a function (the delta function) to a function, but a function that does not fall off at $\pm\infty$ and so is not in $\mathcal{S}(\mathbf{R})$).

To check Fourier inversion in this case

$$\langle \mathcal{F}1, \varphi \rangle = \langle 1, \mathcal{F}\varphi \rangle = \langle 1, \widehat{\varphi} \rangle = \int_{-\infty}^{\infty} \widehat{\varphi}(p)dp = \varphi(0)$$

so

$$\mathcal{F}1 = \delta$$

and we see that using distributions the constant function now has a Fourier transform, but one that is a distribution, not a function. In the physics literature you may find the above calculation in the form

$$\mathcal{F}1 = \int_{-\infty}^{\infty} e^{-2\pi ipx}dx = \delta(x)$$

For the case of a delta function supported at a, one has

$$\begin{aligned}
\langle \mathcal{F}\delta_a, \varphi \rangle &= \langle \delta_a, \mathcal{F}\varphi \rangle \\
&= \widehat{\varphi}(a) \\
&= \int_{-\infty}^{\infty} e^{-2\pi iax}\varphi \; dx \\
&= \langle e^{-2\pi iax}, \varphi \rangle
\end{aligned}$$

so we have the equality of distributions

$$\mathcal{F}\delta_a = e^{-2\pi iax}$$

Similarly

$$\begin{aligned}
\langle \mathcal{F}e^{2\pi ixa}, \varphi \rangle &= \langle e^{2\pi ixa}, \mathcal{F}\varphi \rangle \\
&= \int_{-\infty}^{\infty} e^{2\pi ixa}\mathcal{F}\varphi \; dx \\
&= \mathcal{F}^{-1}(\mathcal{F}\varphi)(a) \\
&= \varphi(a) = \langle \delta_a, \varphi \rangle
\end{aligned}$$

so

$$\mathcal{F}e^{2\pi ixa} = \delta_a$$

Fourier inversion holds true for Schwartz functions, so it also holds true for distributions, since

$$\begin{aligned}
\langle f, \varphi \rangle &= \langle f, \mathcal{F}\mathcal{F}^{-1}\varphi \rangle \\
&= \langle \mathcal{F}f, \mathcal{F}^{-1}\varphi \rangle \\
&= \langle \mathcal{F}^{-1}\mathcal{F}f, \varphi \rangle
\end{aligned}$$

This shows that

$$\mathcal{F}^{-1}\mathcal{F}f = f$$

as distributions.

In the previous section we saw that, for consistency with the definition of translation on functions, we should define the action of translation on distributions by

$$\langle T_a f, \varphi\rangle = \langle f, T_{-a}\varphi\rangle$$

where

$$(T_{-a}\varphi)(x) = \varphi(x-a)$$

As for functions, the Fourier transform then turns translations into multiplication by a phase. This is because

$$\langle \mathcal{F}(T_a f), \varphi\rangle = \langle T_a f, \mathcal{F}\varphi\rangle = \langle f, T_{-a}(\mathcal{F}\varphi)\rangle$$

but

$$\begin{aligned}
T_{-a}(\mathcal{F}\varphi) &= (\mathcal{F}\varphi)(p-a)\\
&= \int_{-\infty}^{\infty} e^{-2\pi i(p-a)x}\varphi\ dx\\
&= \int_{-\infty}^{\infty} e^{-2\pi ipx}(e^{2\pi iax}\varphi)\ dx\\
&= \mathcal{F}(e^{2\pi iax}\varphi)
\end{aligned}$$

so

$$\begin{aligned}
\langle \mathcal{F}(T_a f), \varphi\rangle &= \langle f, \mathcal{F}(e^{2\pi iax}\varphi))\rangle\\
&= \langle \mathcal{F}f, e^{2\pi iax}\varphi\rangle\\
&= \langle e^{2\pi iap}\mathcal{F}f, \varphi\rangle
\end{aligned}$$

Thus we see that, as distributions, we get the same relation as for functions:

$$\mathcal{F}(T_a f) = e^{2\pi iap}\mathcal{F}f$$

Taking the derivative with respect to a at $a=0$ gives

$$\mathcal{F}(f') = 2\pi ip\mathcal{F}f$$

the same relation between derivatives and the Fourier transform as in the function case. The Fourier transform takes a constant coefficient differential operator to multiplication by a polynomial.

Applying this to the distribution $f=\delta'$, the derivative of the delta function, gives

$$\mathcal{F}\delta' = 2\pi ip\mathcal{F}\delta = 2\pi ip$$

Taking more derivatives gives higher powers of p, and we see that, as distributions, polynomials in p now have Fourier transforms, given by linear combinations of derivatives of the delta function.

For another example, one can show (this is in exercises 4.12 and 4.13 of [1] on your problem set) that the Heaviside function has, as a distribution, a Fourier transform given by

$$\mathcal{F}H = \frac{1}{2\pi i} PV\left(\frac{1}{p}\right) + \frac{1}{2}\delta$$

One way to see this is to use the facts that $H' = \delta$ and $\mathcal{F}\delta = 1$ and the relation above for the Fourier transform of derivative to show that

$$1 = \mathcal{F}\delta = \mathcal{F}H' = 2\pi i p \mathcal{F}H$$

If one could just divide distributions by distributions, one would have

$$\mathcal{F}H = \frac{1}{2\pi i p}$$

but one does need to be more careful, interpreting the right hand side as a principal value distribution and noting that $\mathcal{F}H$ can also have a contribution from a distribution like δ supported only at $p = 0$ (so gives 0 when multiplied by $2\pi i p$).

If one defines a variant of the Heaviside function by

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

which satisfies

$$H + H^- = \text{sgn}, \quad H - H^- = 1$$

(here sgn is the sign function), then one can show that

$$\mathcal{F}H^- = \frac{1}{2\pi i} PV\left(\frac{1}{p}\right) - \frac{1}{2}\delta$$

As expected

$$\mathcal{F}H - \mathcal{F}H^- = \mathcal{F}1 = \delta$$

and one finds the Fourier transform of the sign function

$$\mathcal{F}H + \mathcal{F}H^- = \mathcal{F}\text{sgn} = \frac{1}{\pi i} PV\left(\frac{1}{p}\right)$$

Another sort of linear transformation that one can perform on $\mathcal{S}(\mathbf{R})$ is the rescaling

$$\varphi \to (S_a\varphi)(x) = \varphi(ax)$$

for a a non-zero constant. For functions f one has

$$\int_{-\infty}^{\infty} f(ax)\varphi(x)dx = \int_{-\infty}^{\infty} f(u)\frac{1}{|a|}\varphi(\frac{u}{a})du$$

where $u = ax$ and the absolute value is needed to get the sign right when a is negative. One can define the rescaling transformation on distributions in a way compatible with functions by

$$\begin{aligned}\langle S_\alpha f, \varphi\rangle &= \langle f, \frac{1}{|a|}\varphi(\frac{u}{a})\rangle \\ &= \langle f, \frac{1}{|a|}S_{\frac{1}{a}}\varphi\rangle\end{aligned}$$

As an example, taking $f = \delta$ gives

$$\langle S_a\delta, \varphi\rangle = \langle \delta, \frac{1}{|a|}S_{\frac{1}{a}}\varphi\rangle = \frac{1}{|a|}\varphi(\frac{0}{a}) = \frac{1}{|a|}\varphi(0) = \langle \frac{1}{|a|}\delta, \varphi\rangle$$

so

$$S_a\delta = \frac{1}{|a|}\delta$$

As a final example, consider the function

$$f(x) = e^{isx^2}$$

This does not fall off at $\pm\infty$, but interpreted as a distribution it will have a Fourier transform. More generally, one can use our earlier computations of the Fourier transform of a Gaussian to show that

$$\mathcal{F}(e^{-zx^2}) = \sqrt{\frac{\pi}{z}}e^{-\frac{\pi^2 p^2}{z}} = \sqrt{\frac{\pi}{z}}e^{-\frac{\bar{z}\pi^2 p^2}{|z|^2}}$$

where $z = t + is$. This makes sense as a calculation about functions for $t > 0$, as an equality of distributions for $z = -is$, thought of as the limit as $t \to 0^+$. So, as distributions

$$\mathcal{F}(e^{isx^2}) = \sqrt{\frac{\pi}{-is}}e^{-i\frac{\pi^2 p^2}{s}}$$

There is one subtlety here: the need to decide which of the two possible square roots of the first factor should be taken. This ambiguity can be resolved by noting that for $t > 0$ where one has actual functions there is an unambiguous choice, and in the limit $t \to 0^+$ one has:

$$\sqrt{\frac{\pi}{-is}} = \sqrt{\frac{\pi}{|s|}}\begin{cases} e^{i\frac{\pi}{4}}, s > 0 \\ e^{-i\frac{\pi}{4}}, s < 0\end{cases}$$

6 Convolution of a function and a distribution

Recall that for functions $\varphi_1, \varphi_2 \in \mathcal{S}(\mathbf{R})$ one can define the convolution product

$$(\varphi_1 * \varphi_2)(x) = \int_{-\infty}^{\infty}\varphi_1(x-y)\varphi_2(y)dy$$

$\varphi_1 * \varphi_2$ is also in $\mathcal{S}(\mathbf{R})$ and the Fourier transform takes convolution product to pointwise product, satisfying

$$\mathcal{F}(\varphi_1 * \varphi_2) = (\mathcal{F}\varphi_1)(\mathcal{F}\varphi_2)$$

In general, one can't make sense of the product of two distributions in $\mathcal{S}'(\mathbf{R})$, so one also in general cannot define a convolution product of two distributions. One can however define a sensible convolution product of Schwartz functions and distributions

$$\psi * f, \ \psi \in \mathcal{S}(\mathbf{R}), \ f \in \mathcal{S}'(\mathbf{R})$$

For a first guess at how to define this one can try, as for usual mutiplication of a distribution by a function, the transpose of multiplication by a function

$$\langle \psi * f, \varphi \rangle = \langle f, \psi * \varphi \rangle$$

This definition however will not match with the definition when f is a function, due to a similar minus sign problem as in the translation operator case. Using the above definition when f is a function, one would find

$$\begin{aligned}\langle \psi * f, \varphi \rangle &= \int_{-\infty}^{\infty} (\int_{-\infty}^{\infty} \psi(x-y) f(y) dy) \varphi(x) dx \\ &= \int_{-\infty}^{\infty} (\int_{-\infty}^{\infty} \psi(x-y) \varphi(x) dx) f(y) dx\end{aligned}$$

The problem is that the inner integral in the last equation is not $\psi * \varphi$. The sign of the argument of $\psi(x-y)$ is wrong, it should be $\psi(y-x)$. To fix this, define an operator T_-on functions that changes the sign of the argument

$$(T_-\varphi)(x) = \varphi(-x)$$

and then define

Definition (Convolution of a function and a distribution)**.** *The convolution of a function $\psi \in \mathcal{S}(\mathbf{R})$ and a distribution f is the distribution given by*

$$\langle \psi * f, \varphi \rangle = \langle f, (T_-\psi) * \varphi \rangle$$

With this definition, we have

Claim. *For $\psi \in \mathcal{S}(\mathbf{R})$ and $f \in \mathcal{S}'(\mathbf{R})$ we have*

$$\mathcal{F}(\psi * f) = (\mathcal{F}\psi)\mathcal{F}f$$

This is an equality of distributions, with the right hand side distribution the product of a function and a distribution.

Proof.

$$\begin{aligned}
\langle \mathcal{F}(\psi * f), \varphi \rangle &= \langle \psi * f, \mathcal{F}\varphi \rangle \\
&= \langle f, (T_- \psi) * \mathcal{F}\varphi \rangle \\
&= \langle f, \mathcal{F}\mathcal{F}^{-1}((T_- \psi) * \mathcal{F}\varphi) \rangle \\
&= \langle \mathcal{F}f, \mathcal{F}^{-1}((T_- \psi) * \mathcal{F}\varphi) \rangle \\
&= \langle \mathcal{F}f, (\mathcal{F}\psi)(\mathcal{F}^{-1}\mathcal{F}\varphi) \rangle \\
&= \langle \mathcal{F}f, (\mathcal{F}\psi)\varphi \rangle \\
&= \langle (\mathcal{F}\psi)(\mathcal{F}f), \varphi \rangle
\end{aligned}$$

□

A remarkable aspect of a this definition of convolution is that it implies that, while f is a distribution that may not be a function, $\psi * f$ is always a function:

Claim. *For f a distribution and $\psi \in S(\mathbf{R})$, $\psi * f$ is a function, given by*

$$\langle f, T_x T_- \psi \rangle$$

Proof.

$$\begin{aligned}
\langle \psi * f, \varphi \rangle &= \langle f, (T_- \psi) * \varphi \rangle \\
&= \langle f, \int_{-\infty}^{\infty} \psi(x-y)\varphi(x)dx \rangle \\
&= \langle f, \int_{-\infty}^{\infty} (T_x T_- \psi)\varphi(x)dx \rangle \\
&= \int_{-\infty}^{\infty} \langle f, T_x T_- \psi \rangle \varphi(x)dx
\end{aligned}$$

□

For a simple example, take $f = \delta$, the delta function distribution. Then

$$(\psi * \delta)(x) = \langle \delta, T_x T_- \psi \rangle = \psi(x-y)\Big|_{y=0} = \psi(x)$$

so

$$\psi * \delta = \psi$$

Note that this is consistent with the fact that $\mathcal{F}\delta = 1$ since

$$\mathcal{F}(\psi * \delta) = (\mathcal{F}\psi)(\mathcal{F}\delta) = \psi$$

For derivatives of the delta function one finds for the first derivative

$$\begin{aligned}
(\psi * \delta')(x) &= \langle \delta', T_x T_- \psi\rangle \\
&= -\langle \delta, \frac{d}{dy}\psi(x-y)\rangle \\
&= -\langle \delta, -\frac{d}{dx}\psi(x-y) \\
&= \langle \delta, \frac{d}{dx}\psi(x-y)\rangle \\
&= \frac{d}{dx}\psi
\end{aligned}$$

and in general

$$\psi * \delta^{(n)} = \frac{d^n}{dx^n}\psi$$

Another example of the use of convolution for distributions is the Hilbert transform. Recall that in the last section we found that the Fourier transform of the sign function is given by

$$\mathcal{F}\mathrm{sgn} = \frac{1}{\pi i} PV\left(\frac{1}{x}\right)$$

If we define the Hilbert transform on Schwartz functions by

$$\psi \to \mathcal{H}\psi(x) = \psi * \frac{1}{\pi} PV\left(\frac{1}{t}\right) = \frac{1}{\pi} PV\left(\frac{\psi(x-t)}{t}\right)$$

then, on Fourier transforms this is the operator

$$\begin{aligned}
\widehat{\psi} \to \mathcal{F}(\mathcal{H}\psi) &= (\mathcal{F}\psi)\mathcal{F}(\frac{1}{\pi} PV\left(\frac{1}{t}\right)) \\
&= (\mathcal{F}\psi)\mathcal{F}\mathcal{F}i\mathrm{sgn} \\
&= (\mathcal{F}\psi)i\mathrm{sgn}(-p) \\
&= -i\mathrm{sgn}(p)\widehat{\psi}(p)
\end{aligned}$$

7 Distributional solutions of differential equations

It turns out to be very useful to look for solutions to differential equations in a space of distributions rather than some space of functions. This is true even if one is only interested in solutions that are functions, since interpreting these functions as distributions removes the need to worry about their differentiability (since distributions are infinitely differentiable), or the existence of their Fourier transforms (the Fourier transform of a function may be a distribution).

We'll reconsider the solution of the heat equation

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

where we now think of the variable t as parametrizing a distribution on the space $\mathbf{R}$ with coordinate x. We'll write such a distribution as u_t. We say that u_t solves the heat equation if, for all $\varphi \in \mathcal{S}(\mathbf{R})$

$$\langle \frac{d}{dt}u_t, \varphi\rangle = \langle u_t'', \varphi\rangle = \langle u_t, \varphi''\rangle$$

This can be solved by much the same Fourier transform methods as in the function case. The relation of the Fourier transform and derivatives is the same, so we can again turn the heat equation into an equation for the Fourier transform $\widehat{u}_t$ of u_t:

$$\frac{d}{dt}\widehat{u}_t = (2\pi i p)^2\widehat{u}_t = -4\pi^2 p^2 \widehat{u}_t$$

with solution

$$\widehat{u}_t = \widehat{u}_0 e^{-4\pi^2 p^2 t}$$

This can be used to find u_t by the inverse Fourier transform

$$u_t = \mathcal{F}^{-1}(\widehat{u}_0 e^{-4\pi^2 p^2 t})$$

In our earlier discussion of this equation, we needed the initial data $\widehat{u}_0$ to be a Schwartz function, but now it can be something much more general, a distribution. In particular this allows functional initial data that does not fall off quickly at $\pm\infty$, or even distributional initial data.

For example, one could take as initial data a delta-function distribution at $x = 0$, i.e.

$$u_0 = \delta, \quad \widehat{u}_0 = 1$$

Then

$$u_t = \mathcal{F}^{-1} e^{-4\pi^2 p^2 t}$$

which is just the heat kernel $H_{t,\mathbf{R}}$.

In the general case one can, as for functions,write the solution as a convolution

$$u_t = u_0 * H_{t,\mathbf{R}}$$

(since the Fourier transform $\widehat{u}_t$ is the product of the Fourier transform $\widehat{u}_0$ and the Fourier transform of $H_{t,\mathbf{R}}$). Note that the fact that the convolution of a function and a distribution is a function implies the remarkable fact that, starting at $t = 0$ with initial data that is a distribution, too singular to be a function, for any $t > 0$, no matter how small, the solution to the heat equation will be a function. This property of the heat equation is often called a "smoothing property."

The same argument can be used to solve the Schrödinger equation

$$\frac{\partial}{\partial t}\psi = \frac{i\hbar}{2m}\frac{\partial^2}{\partial x^2}u$$

taking ψ_t to be a t-dependent distribution on a one-dimensional space $\mathbf{R}$. One finds

$$\widehat{\psi}_t = \widehat{\psi}_0 e^{-i\frac{\hbar}{2m}(4\pi^2p^2t)}$$

and

$$\psi_t = \psi_0 * S_t$$

where

$$S_t = \mathcal{F}^{-1}(e^{-i\frac{\hbar}{2m}(4\pi^2p^2t)})$$

is a distribution called the Schrödinger kernel. Recall that we earliier studied essentially this distribution, in the form of the distribution $\mathcal{F}(e^{isx^2})$. In this case, the fact that the convolution of a function and a distribution is a function tells us that if our solution is a function at $t = 0$, it is a function for all t. Unlike the heat equation case, we can't so simply take initial data to be a distribution, since then the solution will be the convolution product of one distribution with another.

References

[1] Osgood, Brad G., Lectures on the Fourier Transform and its Applications. American Mathematical Society, 2019.

[2] Strichartz, Robert S., A Guide to Distribution Theory and Fourier Transforms. World Scientific, 2003.