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 Application* [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 (\(S(\mathbb{R})\), 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 $\mathbb{R}$

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

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 \textit{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:

\textbf{Definition (Distributions).} A distribution is a continuous linear map

$$\varphi \rightarrow \langle f, \varphi \rangle \in \mathbb{C}$$

for $\varphi$ in some chosen class of functions (the \textit{“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 $\mathcal{D}(\mathbb{R})$ of smooth (infinitely differentiable) functions that vanish outside some bounded set.

- The space $\mathcal{S}(\mathbb{R})$ of Schwartz functions that we have previously studied.

In this course we will use $\mathcal{S}(\mathbb{R})$ as our test functions, and denote the space of distributions as $\mathcal{S}'(\mathbb{R})$. Distributions with this choice of test functions are conventionally called \textit{“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 $\mathcal{S}'(\mathbb{R})$ so it also is in $\mathcal{S}'(\mathbb{R})$ (which would not have worked for the choice $\mathcal{D}(\mathbb{R})$). The definition of $\mathcal{S}(\mathbb{R})$ 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 $\mathcal{S}(\mathbb{R})$ to be continuous. For a vector space $V$, continuity of a linear map $f : V \to \mathbb{C}$ 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 \( V = \mathcal{S}(\mathbb{R}) \). 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 \( \varphi \in \mathcal{S}(\mathbb{R}) \) is of size less than \( \epsilon \) if for all \( x, k, l \)

\[
|x^k \left( \frac{d}{dx} \right)^l \varphi(x)| < \epsilon
\]

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 \( \mathbb{R} \), we can associate to a function \( f \) in this space the linear functional

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

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 \( f \in \mathcal{S}(\mathbb{R}) \), so

\[
\mathcal{S}(\mathbb{R}) \subset \mathcal{S}'(\mathbb{R})
\]

It also works for the space \( \mathcal{M}(\mathbb{R}) \) 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

\[
f(x) = e^{-2\pi ix}
\]

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

\[
\varphi \rightarrow \langle f, \varphi \rangle = \int_{-\infty}^{\infty} e^{-2\pi ix} \varphi(x) dx = \hat{\varphi}(0)
\]

For an example of a function that does not give an element of \( \mathcal{S}'(\mathbb{R}) \), 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 \( \mathcal{S}(\mathbb{R}) \), but \( \mathcal{D}(\mathbb{R}) \), which gives a larger space of distributions \( \mathcal{D}'(\mathbb{R}) \) than the tempered distributions \( \mathcal{S}'(\mathbb{R}) \). Since functions in \( \mathcal{D}(\mathbb{R}) \) 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 \( x = x_0 \)
  \[ \delta_{x_0} : \varphi \rightarrow \langle \delta_{x_0}, \varphi \rangle = \varphi(x_0) \]

  An alternate notation for this distribution is \( \delta(x - x_0) \), and the special case \( x_0 = 0 \) will sometimes be written as \( \delta \) or \( \delta(x) \). 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 \( t \rightarrow 0^+ \)
  \[ \lim_{t \rightarrow 0^+} H_{t,R}(x) = \lim_{t \rightarrow 0^+} \frac{1}{\sqrt{4\pi t}} e^{-\frac{x^2}{4t}} \]

  We have seen that \( H_{t,R} \) is a good kernel, so that
  \[ \lim_{t \rightarrow 0^+} (\varphi \ast H_{t,R})(x_0) = \lim_{t \rightarrow 0^+} \int_{-\infty}^{\infty} \varphi(x) \frac{1}{\sqrt{4\pi t}} e^{-\frac{(x-x_0)^2}{4t}} \, dx \]
  \[ = \varphi(x_0) \]

  which shows that, as distributions
  \[ \lim_{t \rightarrow 0^+} H_{t,R}(x - x_0) = \delta(x - x_0) \]

  Another example of the same sort is the limit of the Poisson kernel used to find harmonic functions on the upper half plane:
  \[ \lim_{y \rightarrow 0^+} P_y(x) = \lim_{y \rightarrow 0^+} \frac{1}{\pi} \frac{y}{x^2 + y^2} \]

  which satisfies
  \[ \lim_{y \rightarrow 0^+} (\varphi \ast P_y)(x_0) = \varphi(x_0) \]
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, one can define the “Dirac comb” by

\[
\delta_Z : \varphi \to \langle \delta_Z, \varphi \rangle = \sum_{n = -\infty}^{\infty} \varphi(n)
\]

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

\[
\lim_{N \to \infty} D_N(x) = \lim_{N \to \infty} \sum_{n = -\infty}^{\infty} e^{2\pi inx} = \delta_Z
\]

- **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

\[
\lim_{\epsilon \to 0^+} \left( \int_{-1}^{-\epsilon} \frac{1}{x} \varphi(x)dx + \int_{\epsilon}^{1} \frac{1}{x} \varphi(x)dx \right) = \lim_{\epsilon \to 0^+} \int_{\epsilon}^{1} \frac{1}{x} (\varphi(x) - \varphi(-x))dx
\]

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

\[
\frac{1}{x + i\epsilon} \equiv \lim_{\epsilon \to 0^+} \frac{1}{x + i\epsilon}
\]
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

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

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

5 The Fourier transform of a distribution

6 Convolution and distributional solutions of differential equations

7 Fourier series and the Fourier transform for $d > 1$

8 Radial functions and distributions

References
