Fourier Analysis Notes, Spring 2020

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

September 3, 2020

Contents

Chapter 1

Introduction

These are notes from the second half of a spring 2020 Fourier analysis class, written up since the class turned into an online class for the second half of the semester due to the COVID pandemic. The course to some degree followed the textbook [3], with additional material on distributions from other sources.

The first part of the course discussed the basic theory of Fourier series and Fourier transforms, with the main application to finding solutions of the heat equation, the Schrödinger equation and Laplace's equation. For the Fourier series, we roughly followed chapters 2, 3 and 4 of [3], for the Fourier transform, sections 5.1 and 5.2 . An alternate more detailed source that is not qute as demanding on the students is the first half of the book by Howell, [1]. A quick summary of this material follows.

1.1 Fourier series

The subject of Fourier series deals with complex-valued periodic functions, or equivalently, functions defined on a circle. Taking the period or circumference of the circle to be 2π, the Fourier coefficients of a function are

$$\widehat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}d\theta$$

and the Fourier series for the function is

$$\sum_{n=-\infty}^{\infty} \widehat{f}(n)e^{in\theta}$$

A surprisingly difficult central problem of the subject is that of pointwise-convergence. For what conditions on f does the Fourier series at θ converge to $f(\theta)$? It turns out that differentiability of f is sufficient to imply pointwise-convergence (which we proved), but continuity is not sufficient (we did not prove this). We also looked at examples of the convergence behavior at a jump discontinuity (the "Gibbs phenomenon").

One can consider the Hermitian inner products

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) \overline{g(\theta)} d\theta$$

on functions and on sets of Fourier coefficients

$$\langle \{\widehat{f}\}, \{\widehat{g}\} \rangle = \sum_{n=-\infty}^{\infty} \widehat{f}(n) \overline{\widehat{g}(n)}$$

and ask instead about norm-convergence of the Fourier series (i.e. does the norm of the difference between f and its Fourier series converge to zero?). The answer here is simpler than that for pointwise convergence. Functions with finite norm have norm-convergent Fourier series, with the map taking f to the set $\{\widehat{f}\}$ of its Fourier coefficients a unitary (inner-product preserving) isomorphism. The discussion of this subject provided an opportunity to explain the limitations of the Riemann integral and advertise (without defining it) the Lebesgue integral.

The convolution product $f * g$ on periodic functions was defined, showing that it corresponds to the pointwise product on Fourier coefficients. Given a function $g(\theta)$ on the circle, Fourier series were used to find solutions of partial differential equations:

- $$f(t, \theta) = g * \mathcal{H}_t(\theta)$$

  with $\mathcal{H}_t(\theta)$ the heat kernel solves the heat equation

  $$\frac{\partial}{\partial t} f = \frac{\partial^2}{\partial \theta^2} f$$

  for $f(0, \theta) = g(\theta)$ and $t \geq 0$

- $$f(t, \theta) = g * S_t(\theta)$$

  with $S_t(\theta)$ the Schrödinger kernel (an imaginary time version of the heat kernel) solves the Schrödinger equation

  $$i\hbar \frac{\partial}{\partial t} f = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial \theta^2} f$$

  for $f(0, \theta) = g(\theta)$ and all t.

- $$f(r, \theta) = g * P_r(\theta)$$

  with $P_r(\theta)$ the Poisson kernel solves the Laplace equation

  $$\Delta f = 0$$

  for $f(1, \theta) = g(\theta)$ and $0 \leq r \leq 1$

1.2 The Fourier transform

Turning from functions on the circle to functions on $\mathbf{R}$, one gets a more symmetrical situation, with the Fourier coefficients of a function f now replaced by another function on $\mathbf{R}$, the Fourier transform $\widetilde{f}$, given by

$$\widetilde{f}(p) = \int_{-\infty}^{\infty} f(x)e^{-2\pi ipx}dx$$

The analog of the Fourier series is the integral

$$f(x) = \int_{-\infty}^{\infty} \widetilde{f}(p)e^{2\pi ipx}dx$$

The problem of convergence of the Fourier series is replaced by the problem of understanding which functions f have a well-defined Fourier transform $\widetilde{f}$, and for which such $\widetilde{f}$ can f be recovered by the second integral above (this is the problem of "Fourier inversion").

We took a different approach to this problem than the textbook [3], first showing that Fourier inversion holds for a certainly highly restricted class $\mathcal{S}(\mathbf{R})$ of functions, the Schwartz functions. A function f is in $\mathcal{S}(\mathbf{R})$ if it and all its derivatives exist for all x and fall off faster than any power of x. Functions more general than Schwartz functions (i.e. slower fall-off at $\pm\infty$, lack of derivatives or discontinuity for some values of x) will be treated as distributions, a topic not covered in [3] but discussed in detail later in these notes.

For the Fourier transform one again can define the convolution $f * g$ of two functions, and show that under Fourier transform the convolution product becomes the usual product

$$(\widetilde{fg})(p) = \widetilde{f}(p)\widetilde{g}(p)$$

The Fourier transform takes differentiation to multiplication by $2\pi ip$ and one can as in the Fourier series case use this to find solutions of the heat and Schrödinger equations (with $\theta \in S^1$ replaced by $x \in \mathbf{R}$), as well as solutions to the Laplace equation in the upper half-plane.

Chapter 2

The Poisson Summation Formula, Theta Functions, and the Zeta Function

2.1 The Poisson summation formula and some applications

Given a Schwartz function $f \in \mathcal{S}(\mathbf{R})$ on the real number line $\mathbf{R}$, you can construct a periodic function by taking the infinite sum

$$F_1(x) = \sum_{n=-\infty}^{\infty} f(x+n)$$

This has the properties

- The sum converges for any x, since f is falling off at $\pm\infty$ faster than any power.
- $F_1(x)$ is periodic with period 1

  $$F_1(x+1) = F_1(x)$$

  This is because going from x to $x+1$ just corresponds to a shift of indexing by 1 in the defining sum for $F_1(x)$.

Since $F_1(x)$ is a periodic function, you can treat it by Fourier series methods. Note that the periodicity here is chosen to be 1, not 2π, so you need slightly different formulas. For a function g with period 1 whose Fourier series is pointwise convergent, you have

$$\widehat{g}(n) = \int_0^1 g(x) e^{-i2\pi nx} dx$$

$$g(x) = \sum_{n=-\infty}^{\infty} \hat{g}(n) e^{i2\pi nx}$$

If you compute the Fourier coefficients of $F_1(x)$ you find

$$\begin{aligned}
\widehat{F_1}(m) &= \int_0^1 \sum_{n=-\infty}^{\infty} f(x+n) e^{-i2\pi mx} dx \\
&= \sum_{n=-\infty}^{\infty} \int_0^1 f(x+n) e^{-i2\pi mx} dx \\
&= \sum_{n=-\infty}^{\infty} \int_n^{n+1} f(x+n) e^{-i2\pi mx} dx \\
&= \int_{-\infty}^{\infty} f(x) e^{-i2\pi mx} dx \\
&= \widehat{f}(m)
\end{aligned}$$

Theorem (Poisson Summation Formula)**.** *If $f \in \mathcal{S}(\mathbf{R})$*

$$\sum_{n=-\infty}^{\infty} f(x+n) = \sum_{n=-\infty}^{\infty} \widehat{f}(n) e^{i2\pi nx}$$

Proof. The left hand side is the definition of $F_1(x)$, the right hand side is its expression as the sum of its Fourier series. □

What most often gets used is the special case $x = 0$, with the general case what you get from this when translating by x:

Corollary. *If $f \in \mathcal{S}(\mathbf{R})$*

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \widehat{f}(n)$$

This is a rather remarkable formula, relating two completely different infinite sums: the sum of the values of f at integer points and the sum of the values of its Fourier transform at integer points.

2.1.1 The heat kernel

The Poisson summation formula relates the heat kernel on $\mathbf{R}$ and on S^1. Recall that the formula for the heat kernel on $\mathbf{R}$ is

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

with Fourier transform

$$\widehat{H}_{t,\mathbf{R}}(p) = e^{-4\pi^2 p^2 t}$$

Applying the Poisson summation formula to $H_{t,\mathbf{R}}$ gives

$$\sum_{n=-\infty}^{\infty} H_{t,\mathbf{R}}(x+n) = \sum_{n=-\infty}^{\infty} e^{-4\pi^2 n^2 t} e^{2\pi i n x} = H_{t,S^1}(x) \tag{2.1}$$

where H_{t,S^1} is the heat kernel on S^1.

Recall that earlier in the class we claimed that H_{t,S^1} was a "good kernel" and thus, for continuous functions $f(\theta)$ on the circle

$$\lim_{t \to 0^+} f * H_{t,S^1}(\theta) = f(\theta)$$

At the time we were unable to prove this, but using the above relation with the simpler $H_{t,\mathbf{R}}$ in equation 2.1, we can now show that H_{t,S^1} has the desired three properties:

- $$\int_0^1 H_{t,S^1}(x)dx = 1$$

  (the only contribution to the integral is from the $n = 0$ term).

- $$H_{t,S^1}(x) > 0$$

  (since $H_{t,\mathbf{R}} > 0$).

- To show that $H_{t,S^1}(x)$ concentrates at $x = 0$ as $t \to 0^+$, use

  $$H_{t,S^1}(x) = H_{t,\mathbf{R}}(x) + \sum_{|n| \geq 1} H_{t,\mathbf{R}}(x+n)$$

  The first term is a Gaussian that has the concentration property for $t \to 0^+$. The second term goes to 0 as $t \to 0^+$ for $|x| \leq \frac{1}{2}$ (see Stein-Shakarchi, pages 157).

2.1.2 The Poisson kernel

Recall that, given a Schwartz function f on $\mathbf{R}$, one could construct a harmonic function $u(x, y)$ on the upper half plane with that boundary condition by taking

$$u(x, y) = f * P_{y,\mathbf{R}}(x)$$

where the Poisson kernel in this case is

$$P_{y,\mathbf{R}} = \frac{1}{\pi} \frac{y}{x^2 + y^2}$$

which has Fourier transform

$$\widehat{P}_{y,\mathbf{R}}(p) = e^{-2\pi |p| y}$$

For continuous functions f on the circle bounding the unit disk, a unique harmonic function $u(r,\theta)$ on the disk with that boundary condition is given by

$$u(r,\theta) = f * P_{r,S^1}(\theta)$$

where the Poisson kernel in this case is

$$P_{r,S^1}(\theta) = \sum_{n=-\infty}^{\infty} r^{|n|} e^{in\theta} = \frac{1-r^2}{1-2r\cos\theta + r^2}$$

Applying Poisson summation to $P_{y,\mathbf{R}}$ one gets a relation between these two kernels

$$\begin{aligned}\sum_{n=-\infty}^{\infty} P_{y,\mathbf{R}}(x+n) &= \sum_{n=-\infty}^{\infty} \widehat{P}_{y,\mathbf{R}}(n) e^{2\pi i n x} \\ &= \sum_{n=-\infty}^{\infty} e^{-2\pi|n|y} e^{2\pi i n x} \\ &= P_{e^{-2\pi y},S^1}(2\pi x)\end{aligned}$$

which is the Poisson kernel on the disk, with $r = e^{-2\pi y}$ and $\theta = 2\pi x$.

2.2 Theta functions

One can define a fascinating class of functions called "theta functions", with the simplest example

$$\theta(s) = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 s}$$

Here s is real and the sum converges for $s > 0$, but one can also take $s \in \mathbf{C}$, with $\mathrm{Re}(s) > 0$. Applying the Poisson summation formula to the Schwartz function

$$f(x) = e^{-\pi s x^2}, \quad \widehat{f}(p) = \frac{1}{\sqrt{s}} e^{-\pi \frac{p^2}{s}}$$

gives

$$\begin{aligned}\theta(s) &= \sum_{n=-\infty}^{\infty} f(n) \\ &= \sum_{n=-\infty}^{\infty} \widehat{f}(n) \\ &= \frac{1}{\sqrt{s}} \sum_{n=-\infty}^{\infty} e^{-\pi \frac{n^2}{s}} \\ &= \frac{1}{\sqrt{s}} \theta(\frac{1}{s})\end{aligned}$$

which is often called the "functional equation" of the theta function. We will later see that this can be used to understand the properties of the zeta function in number theory.

2.2.1 The Jacobi theta function

This section is about a more general theta function, called the Jacobi theta function. It is getting a bit far from the material of this course, but I wanted to write it up here so that you can see the connection to the heat and Schrödinger equations on the circle.

Definition (Jacobi theta function)**.** *The Jacobi theta function is the function of two complex variables given by*

$$\Theta(z,\tau) = \sum_{n=-\infty}^{\infty} e^{i\pi n^2 \tau} e^{i2\pi n z}$$

This sum converges for $z, \tau \in \mathbf{C}$, with τ in the upper half plane. Note that the heat kernel is given by

$$H_{t,\mathbf{R}}(x) = \Theta(x, i4\pi t)$$

and is well-defined for $t > 0$. The Schrödinger kernel is given by

$$S_t(x) = \Theta(x, \frac{\hbar}{2m}\frac{t}{\pi})$$

which (to stay in the upper half plane of definition), really should be defined as

$$S_t(x) = \lim_{\epsilon \to 0^+} \Theta(x, \frac{\hbar}{2m}\frac{t+i\epsilon}{\pi})$$

The Jacobi theta function has the following properties:

- Two-fold periodicity in z (up to a phase, for fixed τ). Clearly

$$\Theta(z+1,\tau) = \Theta(z,\tau)$$

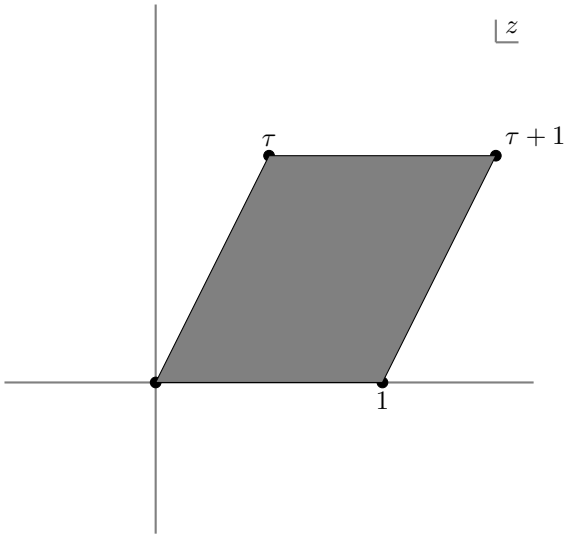
One also has

$$\Theta(z+\tau,\tau) = \Theta(z,\tau) e^{-\pi i \tau} e^{-2\pi i z}$$

since

$$\begin{aligned}
\Theta(z+\tau,\tau) &= \sum_{n=-\infty}^{\infty} e^{\pi i n^2 \tau} e^{2\pi i n(z+\tau)} \\
&= \sum_{n=-\infty}^{\infty} e^{\pi i (n^2+2n)\tau} e^{2\pi i n z} \\
&= \sum_{n=-\infty}^{\infty} e^{\pi i (n+1)^2 \tau} e^{-\pi i \tau} e^{2\pi i n z} \\
&= \sum_{n=-\infty}^{\infty} e^{\pi i (n+1)^2 \tau} e^{-\pi i \tau} e^{2\pi i (n+1) z} e^{-2\pi i z} \\
&= \Theta(z,\tau) e^{-\pi i \tau} e^{-2\pi i z}
\end{aligned}$$

The value of Φ is determined everywhere in terms of its values on the fundamental region below:



- Periodicity in τ

$$\Theta(z,\tau+2) = \Theta(z,\tau)$$

since

$$\begin{aligned}\Theta(z,\tau+2) &= \sum_{n=-\infty}^{\infty} e^{\pi i n^2(\tau+2)} e^{2\pi i n z} \\ &= \sum_{n=-\infty}^{\infty} e^{2\pi i n^2} e^{\pi i n^2 \tau} e^{2\pi i n z} \\ &= \sum_{n=-\infty}^{\infty} e^{\pi i n^2 \tau} e^{2\pi i n z} \\ &= \Theta(z,\tau)\end{aligned}$$

- The following property under inversion in the τ plane, which follows from the functional equation for $\theta(s)$.

$$\Theta(z,-\frac{1}{\tau}) = \sqrt{\frac{\tau}{i}} e^{\pi i \tau z^2} \Theta(z\tau,\tau)$$

2.3 The Riemann zeta function

Recall that we have shown that one can evaluate the sums

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}, \quad \sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90}$$

using Fourier series methods. A central object in number theory is the

Definition (Riemann zeta function)**.** *The Riemann zeta function is given by*

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

For $s \in \mathbf{R}$, this converges for $s > 1$.

One can evaluate $\zeta(s)$ not just at $s = 2, 4$, but at s any even integer (see problem sets) with result

$$\zeta(2n) = \frac{(-1)^{n+1}}{2(2n)!} B_{2n} (2\pi)^{2n}$$

Here B_n are the Bernoulli numbers, which can be defined as the coefficients of the power series expansion

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} B_n \frac{x^n}{n!}$$

There is no known formula for the values of $\zeta(s)$ at odd integers.

The zeta function contains a wealth of information about the distribution of prime numbers. Using the unique decomposition of an integer into primes, one can show

$$\begin{aligned}\zeta(s) &= \sum_{n=1}^{\infty} \frac{1}{n^s} \\ &= \prod_{\text{primes p}} (1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots \\ &= \prod_{\text{primes p}} \frac{1}{1 - p^{-s}}\end{aligned}$$

One can consider the zeta function for complex values of s, in which case the sum defining it converges in the half-plane $\mathrm{Re}(s) > 1$. This function of s can be uniquely extended as a complex valued function to parts of the complex plane with $\mathrm{Re}(s) \leq 1$ by the process of "analytic continuation". This can be done by finding a solution of the Cauchy-Riemann equations which matches the values of $\zeta(s)$ for values of s where the sum in the definition converges, but also exists for other values of s. This analytically extended zeta function then has the following properties:

- $\zeta(s)$ has a well-defined analytic continuation for all s except $s = 1$. There is a pole at $s = 1$ with residue 1, meaning $\zeta(s)$ behaves like $\frac{1}{1-s}$ near $s = 1$.

- $$\zeta(0) = -\frac{1}{2}$$

  Note that if you tried to define $\zeta(0)$ using the sum, this would imply

  $$-\frac{1}{2} = 1 + 1 + 1 + \cdots$$

- At negative integers

  $$\zeta(-n) = \begin{cases} 0 & n \text{ even} \\ (-1)^n \frac{B_{n+1}}{n+1} & n \text{ odd} \end{cases}$$

  In particular, $\zeta(-1) = -\frac{1}{12}$ which motivates claims one sometimes sees that

  $$-\frac{1}{12} = 1 + 2 + 3 + 4 + \cdots$$

2.3.1 The Mellin transform

To prove the functional equation for the zeta function, we need to relate it to the theta function, and will do this using

Definition (Mellin transform)**.** *The Mellin transform of a function $f(x)$ is the function*

$$(\mathcal{M}f)(s) = \int_0^\infty f(x)x^s \frac{dx}{x}$$

Note that the Mellin transform is the analog of the Fourier transform one gets when one replaces the additive group $\mathbf{R}$ with the multiplicative group of positive real numbers. The analogy goes as follows:

- For the Fourier transform, we are using behavior of functions under the transformation

  $$f(x) \to f(x+a)$$

  where $a \in \mathbf{R}$.

  For the Mellin transform, we are using behavior of functions under the transformation

  $$f(x) \to f(ax)$$

  where $a \in \mathbf{R}$ is positive.

- In the Fourier transform case, the function e^{ipx} behaves simply (multiplication by a scalar) under the transformation:

  $$e^{ipx} \to e^{ip(x+a)} = e^{ipa}e^{ipx}$$

  In the Mellin transform case, the function x^s behaves simply (multiplication by a scalar) under the transformation:

  $$x^s \to (ax)^s = a^s x^s$$

- In the Fourier transform case, the integral

  $$\int_{-\infty}^\infty (\cdot)dx$$

  is invariant under the transformation, since $d(x+a) = dx$..

  In the Mellin transform case, the integral

  $$\int_0^\infty (\cdot)\frac{dx}{x}$$

  is invariant under the transformation, since

  $$\frac{d(ax)}{ax} = \frac{dx}{x}$$

Using the Mellin transform, one can define the gamma function by

Definition (Gamma function)**.** *The gamma function is*

$$\Gamma(s) = (\mathcal{M}e^x)(s) = \int_0^\infty e^{-x}x^{s-1}dx$$

The gamma function generalizes the factorial, satisfying $\Gamma(n) = n!$ for n a positive integer. This is because one has (integrating by parts)

$$\begin{aligned}\Gamma(s+1) &= \int_0^\infty e^{-x}x^s dx \\ &= -x^s e^{-x}\,\Big|_0^\infty + s\int_0^\infty e^{-x}x^{s-1}dx \\ &= s\Gamma(s)\end{aligned}$$

and

$$\Gamma(1) = \int_0^\infty e^{-x}dx = -e^{-x}\,\Big|_0^\infty = 1$$

If one allows s to be a complex variable, the definition of $\Gamma(s)$ as an integral converges for $\mathrm{Re}(s) > 0$. One can extend the region of definition of $\Gamma(s)$ to the entire complex plane using the relation

$$\Gamma(s) = \frac{1}{s}\Gamma(s+1)$$

This satisfies the Cauchy-Riemann equations and is the analytic continuation of $\Gamma(s)$ from the region of $\mathbf{C}$ where it is defined by an integral. This definition does imply poles at the non-positive integers $s = 0, -1, -2, \ldots$.

For another example of the Mellin transform, one can take the transform of a Gaussian and get a gamma function:

$$\begin{aligned}(\mathcal{M}e^{-x^2})(s) &= \int_0^\infty e^{-x^2}x^{s-1}dx \\ &= \int_0^\infty e^{-y}(y^{\frac{1}{2}})^{s-1}\frac{dy}{2\sqrt{y}} \\ &= \frac{1}{2}\int_0^\infty e^{-y}y^{\frac{s}{2}-1}dy \\ &= \frac{1}{2}\Gamma(\frac{s}{2})\end{aligned}$$

where in the second line we have made the substitution

$$y = x^2, \quad dy = 2xdx, \quad dx = \frac{dy}{2\sqrt{y}}$$

2.3.2 The zeta function and the Mellin transform of the theta function

It turns out that the zeta function is closely related to the Mellin transform of the theta function. In this section we will show this, in the next use the

functional equation of the theta function to give a functional equation for the zeta function.

Recall the definition of the theta function

$$\theta(x) = \sum_{n=-\infty}^{\infty} e^{-n^2 \pi x}$$

We will work with a slight variant of this, defining

$$w(x) = \sum_{n=1}^{\infty} e^{-n^2 \pi x} = \frac{1}{2}(\theta(x) - 1)$$

Taking the Mellin transform one finds

$$\begin{aligned}(\mathcal{M}w)(s) &= \int_0^{\infty} w(x) x^{s-1} dx \\ &= \sum_{n=1}^{\infty} \int_0^{\infty} e^{-\pi n^2 x} x^{s-1} dx\end{aligned}$$

One can do these integrals with the substitution

$$u = \pi n^2 x, \quad dx = \frac{du}{\pi n^2}$$

with the result

$$\begin{aligned}\int_0^{\infty} e^{-\pi n^2 x} x^{s-1} dx &= \int_0^{\infty} e^{-u} (\frac{u}{\pi n^2})^{s-1} \frac{du}{\pi n^2} \\ &= (\frac{1}{\pi n^2})^{s-1} \frac{1}{\pi n^2} \int_0^{\infty} e^{-u} u^{s-1} du \\ &= \frac{1}{\pi^s n^{2s}} \Gamma(s)\end{aligned}$$

and one finally gets for the Mellin transform of $w(x)$

$$(\mathcal{M}w)(s) = \sum_{n=1}^{\infty} \frac{1}{\pi^s} \frac{1}{n^{2s}} \Gamma(s) \tag{2.2}$$

$$= \frac{1}{\pi^s} \Gamma(s) \zeta(2s+1) \tag{2.3}$$

2.3.3 The functional equation for the zeta function

Finally we would like to use the relation between the zeta function and the Mellin transform of the theta function, together with the functional equation of the theta function to show

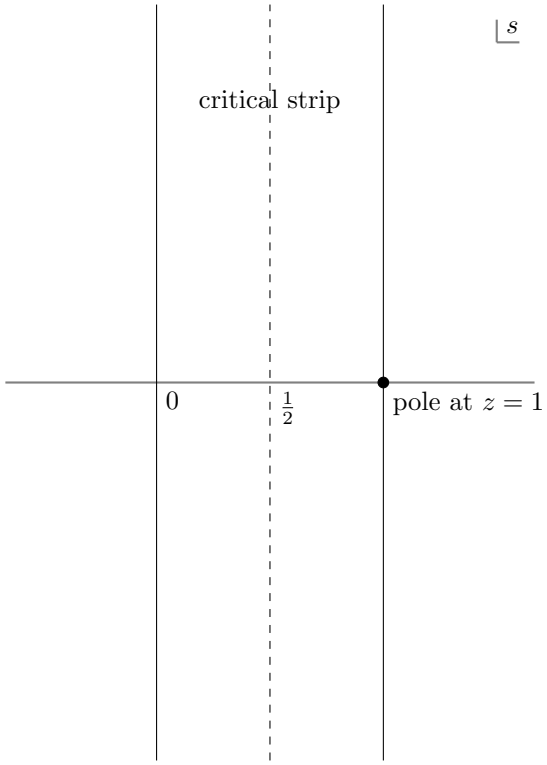
Theorem (Functional equation of the zeta function)**.** *If we define*

$$\Lambda(s) = \pi^{-\frac{s}{2}}\Gamma(\frac{s}{2})\zeta(s)$$

then

$$\Lambda(s) = \Lambda(1-s)$$

This theorem allows us to extend our definition of $\Lambda(s)$ (and thus $\zeta(s)$) from the region $\mathrm{Re}(s) > 1$ where it is defined by a convergent infinite sum to the region $\mathrm{Re}(s) < 0$, giving us an analytic continuation of $\Lambda(s)$ to this region. The most mysterious behavior of $\Lambda(s)$ is in the "critical strip" $0 < \mathrm{Re}(s) < 1$ where no definition other than as an analytic continuation exists.



Perhaps the most famous unproved conjecture in mathematics is

Conjecture (The Riemann Hypothesis)**.** *The zeros of* $\Lambda(s)$ *all lie on the centerline of the critical strip, where the real part of* s *is* $\frac{1}{2}$*.*

We will of course not prove the Riemann hypothesis, but will prove the functional equation for the zeta function, by showing that the functional equation of the theta equation implies

$$(\mathcal{M}w)(s) = (\mathcal{M}w)(\frac{1}{2} - s) \tag{2.4}$$

This implies (using equation 2.2) that

$$\pi^{-s}\Gamma(s)\zeta(2s) = \pi^{s-\frac{1}{2}}\Gamma(\frac{1}{2}-s)\zeta(1-2s)$$

which implies (changing s to $\frac{s}{2}$) that

$$\pi^{-\frac{s}{2}}\Gamma(\frac{s}{2})\zeta(s) = \pi^{-\frac{1-2}{2}}\Gamma(\frac{1-s}{2})\zeta(1-s)$$

and thus

$$\Lambda(s) = \Lambda(1-s)$$

To show equation 2.4, we will use the fact that the functional equation for the theta function

$$\theta(s) = \frac{1}{\sqrt{s}}\theta(\frac{1}{s})$$

implies

$$\begin{aligned} w(x) &= \frac{1}{2}(\theta(x)-1) \\ &= \frac{1}{2}(\frac{1}{\sqrt{x}}\theta(\frac{1}{x})-1) \\ &= \frac{1}{2}(\frac{1}{\sqrt{x}}(2w(\frac{1}{x})+1)-1) \\ &= \frac{1}{\sqrt{x}}w(\frac{1}{x}) + \frac{1}{2}\frac{1}{\sqrt{x}} - \frac{1}{2} \end{aligned}$$

Breaking the integral for the Mellin transform of $w(x)$ into two parts and using the above relation for $w(x)$ in the first part we find

$$\begin{aligned} (\mathcal{M}w)(s) &= \int_0^1 w(x)x^{s-1}dx + \int_1^\infty w(x)x^{s-1}dx \\ &= \int_0^1 (\frac{1}{\sqrt{x}}w(\frac{1}{x}) + \frac{1}{2}\frac{1}{\sqrt{x}} - \frac{1}{2})dx + \int_1^\infty w(x)x^{s-1}dx \\ &= \int_1^\infty u^{-1-s}(u^{\frac{1}{2}}w(u) - \frac{1}{2} + \frac{1}{2}u^{\frac{1}{2}})du + \int_1^\infty w(x)x^{s-1}dx \\ &= \int_1^\infty u^{-\frac{1}{2}-s}w(u)du - \frac{1}{2s} - \frac{1}{1-2s} + \int_1^\infty w(x)x^{s-1}dx \\ &= \int_1^\infty w(u)(u^{-\frac{1}{2}-s} + u^{s-1})du - \frac{1}{2s} - \frac{1}{2(\frac{1}{2}-s)} \end{aligned}$$

This is symmetric under the interchange of s and $\frac{1}{2}-s$, which gives equation 2.4. In the second step we used the substitution

$$u = \frac{1}{x}, \quad du = -x^{-2}dx = -u^2dx$$

Chapter 3

Distributions

3.1 Introduction

These notes will cover distributions, an important topic not discussed in Stein-Shakarchi. Two textbooks you can consult for more details are:

- *A Guide to Distribution Theory and Fourier Transforms* [4], 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* [2], 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.

3.2 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 [4]. 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 very hard to come up with linear functionals that are not continuous.

3.3 Distributions: examples

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

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

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).

Warning! *The notation $\langle f, \varphi \rangle$ that we are using to define a distribution is NOT the inner product on functions when f is a function. One could try to reconcile notations and make this work out, but for better or worse we are not doing so. Two reasons are*

- *For complex valued functions our convention (opposite to standard physics one) is that the Hermitian inner product is complex anti-linear in the second variable (i.e. there is a conjugation on the second function in an inner product). If we put this property in our definition of a distribution, it would lead to ongoing problems keeping track of that conjugation.*
- *In the case of functions with well-defined inner product (L^2 functions), the inner product gives an isomorphism between the function space and its dual space, both of which are $L^2(\mathbf{R})$. The inner product has most of the properties of the finite-dimensional Hermitian inner product, for instance the dual of the dual is the original space, and*

$$\langle v_2, v_1 \rangle = \overline{\langle v_1, v_2 \rangle}$$

The situation with distributions is very different: the dual $\mathcal{S}'(\mathbf{R})$ is a different, much larger space than $\mathcal{S}(\mathbf{R})$, in no sense isomorphic to it. The notation for distributions is inherently asymmetric, with what goes on the left (a distribution) very different than what goes on the right (a Schwartz function).

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 $p = 0$

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

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

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

which 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, one 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.

3.4 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.

3.4.1 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

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

3.4.2 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

$$T_a f = T_{-a}^t f$$

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

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

3.4.3 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

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

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 [2] 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

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

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

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

□

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 $\eta \to 0^+$, the first term gives $PV\left(\frac{1}{x}\right)$, the second gives $\mp i\pi\delta(x)$.

3.5 The Fourier transform of a distribution

Warning: while chapter 5 of [4] 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*

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

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:

$$\mathcal{F}(T_a f) = e^{2\pi i a p}\mathcal{F}f$$

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

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

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

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

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 [2] 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

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

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

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

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:

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

3.6 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 multiplication 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

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

If we define the Hilbert transform on Schwartz functions ψ 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

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

3.7 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:

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

with solution

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

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

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

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

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

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

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

and

$$\psi_t = \psi_0 * S_t$$

where

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

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.

Chapter 4

Higher dimensions

4.1 Introduction

So far in this course we have been discussing Fourier analysis for functions of a single variable: functions on $\mathbf{R}$ in the Fourier transform case, periodic with period 2π in the Fourer series case. In this part of the course we'll first generalize to higher dimensions, then apply Fourier analysis techniques to study partial differential equations in higher dimensions. Unlike the last part of the course (distributions), the material we're covering here is generally well-described in the course textbook [3], with the notes here covering much the same material with less detail although a slightly different point of view (unlike [3], we'll sometimes work with distributions).

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

4.2.1 Fourier series for $d > 1$

Consider a function $f(\theta_1, \theta_2, \cdots, \theta_d)$ of d variables, equivalently periodic on $\mathbf{R}^d$ with period 2π in each variable, or defined only for $-\pi < \theta_j \leq \pi$. We can think of such a function as defined on the circle S^1 in the case $d = 1$, in general on a product $S^1 \times \cdots \times S^1$ of d circles. We'll define Fourier series as the obvious generalization of the the $d = 1$ case. The Fourier coefficients of such a function will depend on integers $n_1, n_2, \cdots, n_d$ and be given by

$$\widehat{f}(n_1, n_2, \cdots, n_d) = \frac{1}{(2\pi)^d} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} e^{-i(n_1\theta_1 + n_2\theta_2 + \cdots n_d\theta_d)} f(\theta_1, \theta_2, \cdots, \theta_d) d\theta_1 d\theta_2 \cdots \theta_d$$

Just like in the $d = 1$ case, some condition on the function f is needed that makes these integrals well defined.

One can ask whether Fourier inversion is true, in the sense that

$$f(\theta_1, \theta_2, \cdots, \theta_d) = \sum_{n_1=-\infty}^{\infty} \cdots \sum_{n_d=-\infty}^{\infty} \widehat{f}(n_1, n_2, \cdots, n_d) e^{i(n_1\theta_1 + n_2\theta_2 + \cdots + n_d\theta_d)}$$

and there are similar theorems as in the d=1 case. If one considers just mean-square convergence, one finds that square-integrable functions satisfy Fourier inversion, and as Hilbert spaces

$$L^2(S^1 \times \cdots \times S^1) = \ell^2(\mathbf{Z}^d)$$

(the Fourier transform preserves the inner product and the Parseval formula holds).

The situation with point-wise convergence is much worse in higher dimensions, since one has a d-fold infinite sum, with the sum in principle depending on the order of summation. However, the examples we saw of resummation and definitions of the sum that are given by limits of convolution with a "good" kernel still work fine. For instance, the obvious generalization of the heat kernel on the circle has the properties needed to ensure that

$$\lim_{t \to 0^+} (f * H_{t, S^1 \times \cdots \times S^1})(\theta_1, \theta_2, \cdots, \theta_d) = f(\theta_1, \theta_2, \cdots, \theta_d)$$

4.2.2 The Fourier transform for $d > 1$

The Fourier transform also generalizes in a straightforward way to $d > 1$ dimensions. Using the vector notation

$$\mathbf{x} = (x_1, x_2, \cdots, x_d), \quad \mathbf{p} = (p_1, p_2, \cdots, p_d), \quad \mathbf{x} \cdot \mathbf{p} = x_1 p_1 + x_2 p_2 + \cdots x_d p_d$$

one defines

$$\widehat{f}(\mathbf{p}) = \mathcal{F} f(\mathbf{p}) = \int_{\mathbf{R}^d} f(\mathbf{x}) e^{-2\pi i \mathbf{x} \cdot \mathbf{p}} dx_1 \cdots dx_d$$

and would like to prove an inversion formula

$$f(\mathbf{x}) = (\mathcal{F}^{-1} \widehat{f})(\mathbf{x}) = \int_{\mathbf{R}^d} f(\mathbf{p}) e^{2\pi i \mathbf{x} \cdot \mathbf{p}} dp_1 \cdots dp_d$$

Just as in the Fourier series case, it turns out that if one just considers mean-squared convergence, using Lebesgue integration and appropriately defining the Fourier transform so as to give

$$\mathcal{F} : L^2(\mathbf{R}^d) \to L^2(\mathbf{R}^d)$$

then Fourier inversion holds ($\mathcal{F}^{-1}\mathcal{F} = 1$) and $\mathcal{F}$ is an isomorphism of Hilbert spaces, so preserves inner products (Plancherel theorem), in particular

$$\int_{\mathbf{R}^d} |f(\mathbf{x})|^2 dx_1 dx_2 \cdots dx_d = \int_{\mathbf{R}^d} |\widehat{f}(\mathbf{p})|^2 dp_1 dp_2 \cdots dp_d$$

As in the $d = 1$ case, we'll proceed by first working with a special class of well-behaved functions, the Schwartz functions. In words the definition in higher dimensions is the same as in $d = 1$: a function f is in the Schwartz space $\mathcal{S}(\mathbf{R}^d)$ if f is smooth (C^∞) and f and all its derivatives fall off at $\pm\infty$ faster than any power. See the textbook [3], page 180 for a slightly more precise version of $\mathcal{S}(\mathbf{R}^d)$. We'll treat less well-behaved functions as distributions, in a space $\mathcal{S}'(\mathbf{R}^d)$ of linear functionals on $\mathcal{S}(\mathbf{R}^d)$.

Many of the properties of the Fourier transform in $d > 1$ are much the same as in $d = 1$ and proven by essentially the same arguments. In particular

- The Fourier transform of a function in $\mathcal{S}(\mathbf{R}^d)$ is in $\mathcal{S}(\mathbf{R}^d)$.
- The Fourier transform on distributions in $\mathcal{S}'(\mathbf{R}^d)$ is defined as the transpose of the Fourier transform on functions in $\mathcal{S}(\mathbf{R}^d)$ and takes distributions to distributions.
- Fourier transformation takes translation by a vector $\mathbf{a}$ to multiplication by the function $e^{2\pi i \mathbf{p}\cdot\mathbf{a}}$.
- Fourier transformation takes the partial derivatives $\boldsymbol{\nabla} = (\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_d})$ to multiplication by $2\pi i\mathbf{p}$.

4.3 Rotations and the Fourier transform

In any dimension, one can define rotations as those linear transformations that preserve the inner product:

Definition. *A rotation of* $\mathbf{R}^d$ *is a linear map*

$$R : \mathbf{x} \in \mathbf{R}^d \to R\mathbf{x} \in \mathbf{R}^d$$

such that

$$R\mathbf{x} \cdot R\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$$

These are the linear transformations that preserve lengths ($|\mathbf{x}|^2 = |R\mathbf{x}|^2$) and angles. Such linear transformations form a group, meaning

- The composition of two rotations is a rotation.
- There is a unit, the identity transformation I.
- Any rotation R has an inverse R^{-1} such that R composed with R^{-1} is the identity I.

The group of rotations in d dimensions is called $O(d)$, with the O for "orthogonal". A rotation R gives a linear transformation $\mathcal{R}$ on functions on $\mathbf{R}^d$

$$\mathcal{R} : f(\mathbf{x}) \to \mathcal{R}f = f(R\mathbf{x})$$

We define the action of rotations on distributions using the transpose, with an inverse so that the definition agrees with the definition on functions. If f is a distribution, then $\mathcal{R}f$ is the distribution given by

$$\langle \mathcal{R}f, \varphi \rangle = \langle f, \mathcal{R}^{-1}\varphi \rangle$$

For each rotation R we now have two linear transformations on $\mathcal{S}(\mathbf{R}^d)$ (and on $\mathcal{S}'(\mathbf{R}^d)$): the Fourier transform $\mathcal{F}$ and the rotation action $\mathcal{R}$. These commute

Claim. *For $f \in \mathcal{S}(\mathbf{R}^d)$*

$$\mathcal{R}\mathcal{F} = \mathcal{F}\mathcal{R}$$

Proof.

$$\begin{aligned}
\mathcal{R}\mathcal{F}f &= \widehat{f}(R\mathbf{p}) \\
&= \int_{\mathbf{R}^d} e^{2\pi i \mathbf{x}\cdot R\mathbf{p}} f(\mathbf{x}) dx_1 dx_2 \cdots dx_d \\
&= \int_{\mathbf{R}^d} e^{2\pi i (R\mathbf{x}'\cdot R\mathbf{p})} f(R\mathbf{x}') |det R| dx'_1 dx'_2 \cdots dx'_d \\
&= \int_{\mathbf{R}^d} e^{2\pi i \mathbf{x}'\cdot \mathbf{p}} f(R\mathbf{x}') dx'_1 dx'_2 \cdots dx'_d \\
&= \widehat{f(R\mathbf{x})}(\mathbf{p}) = \mathcal{F}\mathcal{R}f
\end{aligned}$$

Here in the third line we have used the substitution $\mathbf{x} = R\mathbf{x}'$. □

A radial function will be a function on $\mathbf{R}^d$ that only depends on the distance to the origin, so is invariant under rotations:

Definition. *A radial function is a function satisfying*

$$\mathcal{R}f = f$$

for all rotations R.

The commutativity of $\mathcal{R}$ and $\mathcal{F}$ imply

Claim. *The Fourier transform of a radial function is radial.*

Proof. If f is radial, then $\mathcal{R}f = f$ and by commutativity of $\mathcal{R}$ and $\mathcal{F}$

$$\mathcal{R}\mathcal{F}f = \mathcal{F}\mathcal{R}f = \mathcal{F}f$$

□

In $d = 1$, $O(1)$ is the two element group $\mathbf{Z}_2 = \{I, -I\}$, with one element taking $x \to x$ and the other taking $x \to -x$. In this case a radial function is just an even function, and we have seen previously that the Fourier transform of an even function is even.

4.3.1 Two dimensions

In two dimensions one has (as sets)

$$O(2) = SO(2) \times \mathbf{Z}_2$$

meaning that any rotation R can be decomposed into the product of

- An element R_θ in the subgroup $SO(2)$ given by counter-clockwise rotations by angles θ.
- An element of the two element group $\mathbf{Z}_2$ with one element the identity, the other a reflection about an axis, e.g. the reflection $R_- : (x_1, x_2) \to (x_1, -x_2)$ about the x_1 axis.

Two different ways one can work with rotations in two dimensions are

- Identify $\mathbf{R}^2 = \mathbf{C}$ using $z = x_1 + ix_2$. Then elements of the $SO(2)$ subgroup act by
$$z \to R_\theta z = e^{i\theta} z$$
and the reflection R_- acts by conjugation ($z \to \overline{z}$).
- Using matrices, the action of an element of $SO(2)$ is given by
$$R_\theta \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$
and the action of reflection is
$$R_- \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$
One can characterize the rotation matrices as all two by two matrices M that satisfy the condition $MM^T = I$ (here M^T is the transpose of M). One can check that this is the condition on matrices corresponding to the condition that as an action on vectors they preserve the inner product. Since the determinants will satisfy
$$\det(MM^T) = \det M^2 = 1$$
one has $\det M = \pm 1$. The group of these matrices breaks up into a component with determinant 1 (this is $SO(2)$) and a component with determinant -1 (these are a product of an element of $SO(2)$ and a reflection).

To study radial functions in two dimensions, it is convenient to change to polar coordinates, and for a radial function write

$$f(x_1, x_2) = f(r, \theta) = f(r) = f(|\mathbf{x}|)$$

For its Fourier transform (which will also be radial) write

$$\widehat{f}(p_1, p_2) = \widehat{f}(|\mathbf{p}|)$$

If we compute this Fourier transform we find (using the fact that $\mathbf{p} \cdot \mathbf{x} = |\mathbf{p}| r \cos\theta$, where θ is the angle between $\mathbf{x}$ and $\mathbf{p}$)

$$\begin{aligned}\widehat{f}(|\mathbf{p}|) &= \int_{-\pi}^{\pi} \int_0^{\infty} f(r) e^{-2\pi i |\mathbf{p}| p r \cos\theta} r dr d\theta \\ &= \int_0^{\infty} f(r) r \left(\int_{-\pi}^{\pi} e^{2\pi i |\mathbf{p}| r \sin\theta} d\theta \right) dr\end{aligned}$$

Here we have used $-\cos\theta = \sin(\theta - \frac{\pi}{2})$.

The function in parentheses is a Bessel function, often written

$$J_0(r) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ir\sin\theta} d\theta$$

More generally, $J_n(r)$ is the n'th Fourier coefficient of the function $e^{ir\sin\theta}$ so

$$J_n(r) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ir\sin\theta} e^{-in\theta} d\theta$$

We see that in two dimensions the Fourier transform of a radial function f can be written as

$$\widehat{f}(|\mathbf{p}|) = 2\pi \int_0^{\infty} r f(r) J_0(2\pi |\mathbf{p}| r) dr$$

4.3.2 Three dimensions

In any number of dimensions, rotations are the linear transformations that, written as matrices M, satisfy the condition

$$M^T M = I$$

where M^T is the transpose matrix. To see that this is the same as the condition of preserving the inner product (the dot product), note that

$$\mathbf{x} \cdot \mathbf{y} = \begin{pmatrix} x_1, \cdots, x_d \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix}$$

and

$$M\mathbf{x} \cdot M\mathbf{y} = \begin{pmatrix} x_1 & \cdots & x_d \end{pmatrix} M^T M \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix}$$

As noted in the last section $M^T M = I$ implies that $\det M = \pm 1$. The group $O(d)$ breaks up into two components: a subgroup $SO(d)$ of orientation preserving rotations (those with determinant $+1$ and a component of rotations that change orientation, those with determinant -1. If you think of the matrix M as a collection of d column vectors, the condition $M^T M = I$ says that

- Since the off-diagonal elements of I are zero, the dot product of two different column vectors is zero, so they are all pair-wise orthogonal.
- Since the diagonal elements of I are 1, the dot product of each column vector with itself is 1, so the row vectors are not just orthogonal, but orthonormal.

Elements of $O(d)$ can thus be characterized by a set of d orthonormal vectors in $\mathbf{R}^d$, this is just an orthonormal basis in R^d.

In the case $d = 2$, elements of $O(2)$ are given by choosing for the first column a vector on the unit circle in $\mathbf{R}^2$

$$\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$$

and for the second column one of the two perpendicular unit vectors

$$\pm \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$$

The positive sign gives determinant $+1$ and thus elements of $SO(2)$, the negative sign gives the other component of $O(2)$. In the case $d = 3$, explicit parametrizations are more complicated, but you can construct all elements of $O(3)$ by

- Pick a first unit vector in $\mathbf{R}^3$. These lie on the unit sphere $S^2 \subset \mathbf{R}^3$ and can be parametrized by two angles.
- Pick a second unit vector in $\mathbf{R}^3$, perpendicular to the first. If you take the first unit vector to point to the North pole of S^2, this second one will lie on a circle S^1 that is the equator of the sphere. These are parametrized by a third angle.
- Pick a third unit vector in $\mathbf{R}^3$, perpendicular to each of the first two. There are two possible choices with opposite sign, one of which will give determinant 1 and an element of $SO(3)$, the other will give determinant -1 and an element of the other component of $O(3)$.

We see that elements of $SO(3)$ can be parametrized by three angles, and are given by a choice of an element of S^2 and an element of S^1. It however is not true that $SO(3) = S^2 \times S^1$ as a space since there is a subtlety: the S^1 is not fixed, but depends on your first choice of an element of S^2.

A standard way to explicitly parametrize elements of $SO(3)$ is by three Euler angles ϕ, θ, ψ, writing a rotation matrix as

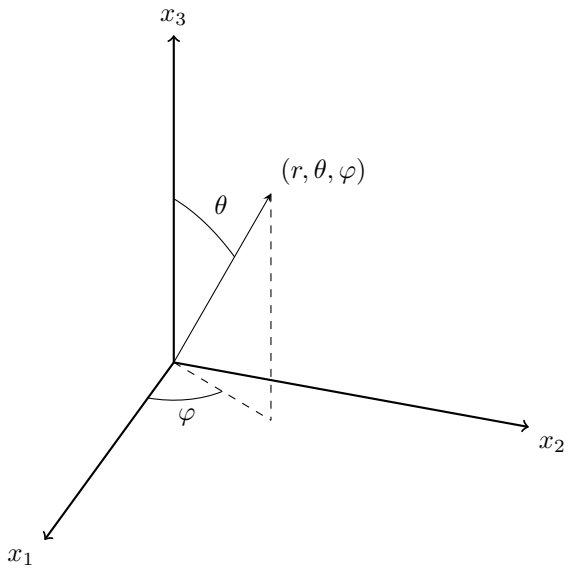
$$\begin{pmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

This product of matrices corresponds to (read from right to left) the composition of

- A rotation by angle φ about the 3-axis.
- A rotation by angle θ about the 1-axis.
- A rotation by angle ψ about the (new) 3-axis.

To do calculations in $\mathbf{R}^2$ for which one wants to exploit rotational symmetry, one uses polar coordinates. For problems in $\mathbf{R}^3$ involving rotational symmetry, the standard choice is spherical coordinates r, θ, φ for a vector, where

- r is the distance to the origin.
- θ is the angle between the 3-axis and the vector.
- φ is the angle between the 1-axis and the projection of the vector to the $1-2$ plane.



The relation between x_1, x_2, x_3 and r, θ, φ coordinates is given by

$$\begin{aligned} x_1 &= r \sin\theta \cos\varphi \\ x_2 &= r \sin\theta \sin\varphi \\ x_3 &= r \cos\theta \end{aligned}$$

Recall that integrals in spherical coordinates are given by

$$\int_{\mathbf{R}^3} F(x_1, x_2, x_3) dx_1 dx_2 dx_3 = \int_0^\infty \int_0^\pi \int_0^{2\pi} F(r, \theta, \varphi) r^2 \sin\theta d\varphi d\theta dr$$

where the range of integration is a full 2π for φ, and half as large (0 to π) for θ. Doing the integral for the Fourier transform in spherical coordinates, for the

case of a radial function $f(r,\theta,\varphi) = f(r)$, $\widehat{f}$ will be radial also, only depending on $|\mathbf{p}|$

$$\widehat{f}(|\mathbf{p}|) = \int_0^\infty f(r) \left(\int_0^\pi \int_0^{2\pi} e^{-2\pi i \mathbf{p}\cdot\mathbf{x}} \sin\theta d\varphi d\theta \right) r^2 dr$$

To evaluate the integral in parentheses, use the fact that it is independent of the direction of $\mathbf{p}$, so you might as well take $\mathbf{p}$ in the 3-direction

$$\mathbf{p} = |\mathbf{p}|\mathbf{e}_3$$

If we denote the unit vector in the $\mathbf{x}$ direction by $\widehat{\mathbf{r}}$, then

$$\mathbf{x} = r\widehat{\mathbf{r}}$$

and

$$\mathbf{p}\cdot\mathbf{x} = r|\mathbf{p}|\cos\theta$$

The integral in parentheses then becomes

$$\begin{aligned}
\int_0^\pi \int_0^{2\pi} e^{-2\pi i r|\mathbf{p}|\cos\theta} \sin\theta d\varphi d\theta =& 2\pi \int_0^\pi e^{-2\pi i r|\mathbf{p}|\cos\theta} \sin\theta d\theta \\
=& 2\pi \int_{-1}^1 e^{2\pi i r|\mathbf{p}|u} du \\
=& 2\pi \frac{1}{2\pi i r|\mathbf{p}|} e^{2\pi i r|\mathbf{p}|u} \Big|_{-1}^1 \\
=& \frac{2}{r|\mathbf{p}|} \sin(2\pi r|\mathbf{p}|)
\end{aligned}$$

(in the second step the substitution is $u = -\cos\theta$). Our final formula for the Fourier transform of a radial function in three dimensions is

$$\widehat{f}(|\mathbf{p}|) = \frac{2}{\mathbf{p}|} \int_0^\infty f(r) \sin(2\pi r|\mathbf{p}|) r dr$$

Chapter 5

Wave Equations

5.1 Introduction

In this part of the course we'll begin by quickly discussing the heat and Schrödinger equations in higher dimensions (where the behavior is very similar to that in $d = 1$), then go on to discuss the wave equation

$$\frac{\partial^2}{\partial t^2} u = c^2 \nabla^2 u$$

where $u(t, \mathbf{x})$ is a function of a scalar time variable t and a spatial vector variable $\mathbf{x}$ (c is a constant). Since this is second-order in time, it exhibits some different behavior that that of the first-order heat and Schrödinger equations. The wave equation is discussed both in the introductory section 1, chapter 1 of [3], as well as in section 3 of chapter 6.

5.2 The heat and Schrödinger equations in higher dimensions

I won't write out here the formulas for using the Fourier transform to solve the heat and Schrödinger equations in arbitrary dimensions, since these are essentially the same as in the $d = 1$ case or a fairly obvious generalization. For these equations we see that if the initial value data at $t = 0$ is radial, the solution will be radial for all t. In the Schrödinger equation case, physicists refer to such rotationally symmetric solutions as "s-wave" solutions.

5.3 The wave equation in $d = 1$

In one spatial dimension, we are looking for functions of two variables $u(t, x)$ that satisfy

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

Note that if u is of the form $F(x \pm ct)$ for some function F of a single variable $y = x \pm ct$, then

$$\frac{\partial F}{\partial y} = \frac{\partial F}{\partial x} = \pm \frac{\partial F}{\partial (ct)} = \pm \frac{1}{c} \frac{\partial F}{\partial t}$$

so

$$\frac{\partial^2 F}{\partial x^2} = \frac{\partial}{\partial x} \left(\pm \frac{1}{c} \frac{\partial F}{\partial t} \right) = (\pm)(\pm) \frac{1}{c^2} \frac{\partial^2 F}{\partial t^2} = \frac{1}{c^2} \frac{\partial^2 F}{\partial t^2}$$

and one gets solutions to the wave equation for any function F. One reason to consider distributional solutions is that then this same argument works for things like F non-differentiable.

To apply Fourier analysis methods, as in the case of the heat or Schrödinger equation, Fourier transform in the x variable and think of t as a parameter. The derivatives in x become multiplication by $2\pi ip$ and one gets the following equation for the Fourier transform in x of $u(t, x) = u_t(x)$:

$$\begin{aligned} \frac{d^2}{dt^2} \widehat{u}_t(p) &= (-2\pi ip)^2 c^2 \widehat{u}_t(p) \\ &= -4\pi^2 c^2 p^2 \widehat{u}_t(p) \end{aligned}$$

This has a general solution

$$\widehat{u}_t(p) = A(p) \cos(2\pi c|p|t) + B(p) \sin(2\pi c|p|t)$$

where $A(p), B(p)$ are arbitrary functions of p.

The $A(p), B(p)$ can be determined from initial data

$$u(0, x) = f(x), \quad \frac{\partial u}{\partial t}\Big|_{t=0} = g(x)$$

or the Fourier transforms of this initial data

$$\widehat{u}_0(p) = \widehat{f}(p), \quad \frac{\partial}{\partial t} \widehat{u}_t \Big|_{t=0} = \widehat{g}(p)$$

Note that a solution is determined by not one but two arbitrary functions since the wave equation is second order in time, so one needs as initial data not just u at $t = 0$, but also its first time derivative at $t = 0$.

Setting $t = 0$ in the general solution one finds

$$\widehat{u}_0(p) = \widehat{f}(p) = A(p)$$

Differentiating the general solution gives

$$\frac{d}{dt} \widehat{u}_t(p) = A(p)(-2\pi c|p|) \sin(2\pi c|p|t) + B(p)(2\pi c|p|) \cos(2\pi c|p|t)$$

At $t = 0$ the right hand side of this equation is

$$B(p)(2\pi c|p|)$$

so

$$B(p) = \frac{\widehat{g}(p)}{2\pi c|p|}$$

In terms of the initial data

$$\widehat{u}_t(p) = \widehat{f}(p)\cos(2\pi c|p|t) + \frac{\widehat{g}(p)}{2\pi c|p|}\sin(2\pi c|p|t)$$

and

$$\frac{d}{dt}\widehat{u}_t(p) = 2\pi c|p|\widehat{f}(p)\sin(2\pi c|p|t) + \widehat{g}(p)\cos(2\pi c|p|t)$$

One can find the solution as the inverse Fourier transform of $\widehat{u}_t(p)$.

If the initial data $f(x), g(x)$ are in $\mathcal{S}(\mathbf{R})$, so are their Fourier transforms as well as the functions

$$\widehat{f}(p)\cos(2\pi c|p|t), \quad \frac{\widehat{g}(p)}{2\pi c|p|}\sin(2\pi c|p|t)$$

The solution $u(t,x)$ is the inverse transform of the sum of these, so also in $\mathcal{S}(\mathbf{R})$. Note that, unlike in the heat equation case, this will be true for all t, not just $t \geq 0$.

As in the heat equation and Schrödinger cases, the derivation given above works not just for functions, but for distributions, and by the same argument, for initial data $f, g \in \mathcal{S}'(\mathbf{R})$ one gets a solution $u(t,x)$ that is a distribution in the x variable, for each t. For some examples, consider

- $$u(t,x) = \cos(x+ct)$$

  This is a so-called plane wave, moving with speed c. For all t it is not in $\mathcal{S}(\mathbf{R})$, but is in $\mathcal{S}'(\mathbf{R})$. In particular, the initial data is

  $$u(0,x) = \cos x$$
  $$\frac{\partial u}{\partial t}\Big|_{t=0} = -c\sin x$$

- A distributional solution with initial data concentrated at $x = 0$, for instance

  $$u(0,x) = f(x) = \delta(x), \quad \widehat{f} = 1$$
  $$\frac{\partial u}{\partial t}\Big|_{t=0} = 0 = \widehat{g}$$

  will give the solution

$$\begin{aligned} u(t,x) &= \int_{-\infty}^{\infty} \cos(2\pi c|p|t)e^{2\pi ipx}dp \\ &= \frac{1}{2}\int_{-\infty}^{\infty}(e^{2\pi ic|p|t} + e^{-2\pi ic|p|t})e^{2\pi ipx}dp \\ &= \frac{1}{2}\int_{-\infty}^{\infty}(e^{2\pi ip(x+ct)} + e^{-2\pi ip(x-ct)})dp \\ &= \frac{1}{2}(\delta(x+ct) + \delta(x-ct)) \end{aligned}$$

(in this calculation at the second step the absolute values around p can be removed since the terms swap as $p \to -p$, and one needs to give the calculation an appropriate interpretation in terms of distributions since the integrals are not well defined).

Define a function $W_t(x)$ of x, parametrized by t, such that its Fourier transform is

$$\widehat{W_t}(p) = \frac{\sin(2\pi|p|ct)}{2\pi|p|c} = \frac{\sin(2\pi pct)}{2\pi pc} = (e^{2\pi ipct} - e^{-2\pi ipct})\frac{1}{4i\pi pc}$$

(one can remove the absolute values on $|p|$ since the sign changes top and bottom cancel). One then has

$$\frac{\partial \widehat{W_t}}{\partial t} = \cos(2\pi pct)$$

and can write the Fourier transform of a solution as

$$\widehat{u_t}(p) = \widehat{f}(p)\frac{\partial \widehat{W_t}}{\partial t} + \widehat{g}(p)\widehat{W_t}(p)$$

The solution is then given in terms of convolution by

$$u(t,x) = f * \frac{\partial W_t}{\partial t} + g * W_t$$

Calculating $W_t(x)$ gives

Claim.

$$W_t(x) = \frac{1}{2c}(H(x+ct) - H(x-ct))$$

where $H(x)$ is the Heaviside function.

Proof. Recall that the Fourier transform of $f(x+a)$ is $e^{2\pi ipa}$ times the Fourier transform of $f(x)$. This implies

$$\mathcal{F}(\frac{1}{2c}(H(x+ct) - H(x-ct)) = \frac{1}{2c}(e^{2\pi ipct} - e^{-2\pi ipct})\mathcal{F}H$$

Recall that

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

so

$$\mathcal{F}(\frac{1}{2c}(H(x+ct) - H(x-ct)) = (e^{2\pi ipct} - e^{-2\pi ipct})(\frac{1}{4i\pi c}PV\left(\frac{1}{p}\right) + \frac{1}{2}\delta)$$

The term involving the delta function cancels since it is multiplied by something that vanishes at $p = 0$, and one sees that one gets (giving the $\frac{1}{p}$ the principal value interpretation) the formula we used to define $\widehat{W_t}$. □

$W_t(x)$ is thus the function given by

$$W_t(x) = \begin{cases} 0, \ x < -ct \\ \frac{1}{2c}, \ -ct < x < ct \\ 0, \ x > ct \end{cases}$$

and its derivative is given by

$$\frac{\partial W}{\partial t} = \frac{1}{2}(\delta(x+ct) + \delta(x-ct))$$

Then the formula for $u(t,x)$ in terms of initial value data has two parts

$$\begin{aligned} u_f(t,x) = f * \frac{\partial W_t}{\partial t} &= \int_{-\infty}^{\infty} f(x-y)\frac{1}{2}(\delta(y+ct) + \delta(y-ct))dy \\ &= \frac{1}{2}(f(x-ct) + f(x+ct)) \end{aligned}$$

and

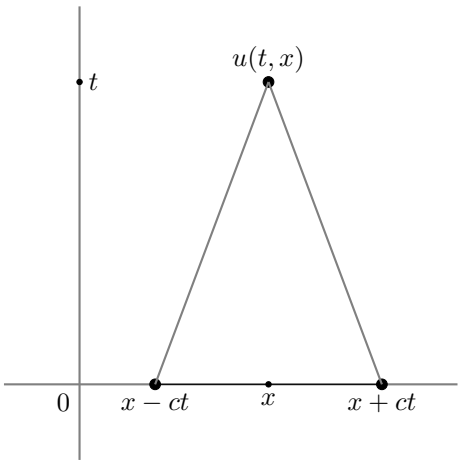
$$\begin{aligned} u_g(t,x) = g * W_t &= \int_{-\infty}^{\infty} g(x-y)W_t(y)dy \\ &= \frac{1}{2c}\int_{-ct}^{ct} g(x-y)dy \\ &= \frac{1}{2c}\int_{x-ct}^{x+ct} g(u)du \end{aligned}$$

In the last step we have used the substitution $u = x - y$.

The formula

$$u(t,x) = \frac{1}{2}(f(x-ct) + f(x+ct)) + \frac{1}{2c}\int_{x-ct}^{x+ct} g(u)du$$

is known as d'Alembert's formula. It implements the Huygens principle, which says that the value of u at (t,x) only depends on initial value data that can reach this point from $t = 0$ by traveling at velocity $\leq c$, so coming from the interval $[x-ct, x+ct]$. The first contribution comes from signals traveling at exactly velocity c, to the right or left, while the second comes from signals traveling at velocity less than c.



For another point of view on the relation of solutions to the initial data, note that $u(t,x)$ is a sum of two sorts of solutions

- Solutions with $f=0, g\neq 0$, given by $u=g*W_t$
- Solutions v that are derivatives of the first kind of solutions, given by $v=\frac{\partial u}{\partial t}=g*\frac{\partial W_t}{\partial t}$ (now g gives the initial value of the solution, the initial value of its time derivative is 0).

One can characterize solutions to the wave equation by a conserved quantity, the energy

Definition (Energy)**.** *The energy of a solution $u(t,x)$ is*

$$E(t)=\frac{1}{2}\int_{-\infty}^{\infty}\left(|\frac{\partial u}{\partial t}|^2+c^2|\frac{\partial u}{\partial x}|^2\right)dx$$

The first term is called the kinetic energy, the second the potential energy.

Claim.

$$\frac{dE}{dt}=0$$

Proof. Applying the Plancherel theorem relating the norm-squared of a function and its Fourier transform, as well as using the formula for the Fourier transform of a derivative, one finds that, in terms of the Fourier transform $\widehat{u}(t,p)$ the energy is

$$\begin{aligned}E(t)&=\frac{1}{2}\int_{-\infty}^{\infty}\left(|\frac{\partial\widehat{u}}{\partial t}|^2+c^2|(2\pi ip)^2\widehat{u}(t,p)|^2\right)dp\\&=\frac{1}{2}\int_{-\infty}^{\infty}\left(|\frac{\partial\widehat{u}}{\partial t}|^2+4\pi^2c^2p^2|\widehat{u}(t,p)|^2\right)dp\\&=\frac{1}{2}\int_{-\infty}^{\infty}(4\pi^2c^2p^2|\widehat{f}(p)|^2+|\widehat{g}(p)|^2)dp\end{aligned}$$

where we have used

$$\widehat{u}(t,p) = \widehat{f}(p)\cos(2\pi c|p|t) + \frac{\widehat{g}(p)}{2\pi c|p|}\sin(2\pi c|p|t)$$

and

$$\frac{d}{dt}\widehat{u}(t,p) = -2\pi c|p|\widehat{f}(p)\sin(2\pi c|p|t) + \widehat{g}(p)\cos(2\pi c|p|t)$$

From the above we see that, expressed in terms of the Fourier transform $\widehat{u}(t,p)$, the energy E of a solution is time independent. Another application of the Plancherel formula shows that E is given in terms of the initial value data as

$$E = \frac{1}{2}\int_{-\infty}^{\infty}\left(c^2|\frac{\partial f}{\partial x}|^2 + |g|^2\right)dx$$

□

Note that the energy is only defined and a conserved quantity for solutions $u(t,x) \in \mathcal{S}(\mathbf{R})$. For distributions, you can't necessarily make sense of the norm-squared of the solutions (for example, the delta-function distribution has no well-defined square). One can use conservation of energy to show that the solution we have found to the wave equation with given initial data is unique, see Problem 3 in Chapter 6 of [3] for details.

5.4 The wave equation in $d = 3$

In an arbitrary number of dimensions the wave equation is

$$\frac{\partial^2}{\partial t^2}u = c^2\nabla^2 u$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \cdots + \frac{\partial^2}{\partial x_d^2}$$

Fourier transforming with respect to the variables x_j gives the following equation for the Fourier transform $\widehat{u}$

$$\begin{aligned}\frac{d^2}{dt^2}\widehat{u}(t,p) &= (2\pi i\mathbf{p})\cdot(2\pi i\mathbf{p})c^2\widehat{u}(t,p)\\ &= -4\pi^2c^2|\mathbf{p}|^2\widehat{u}_t(t,p)\end{aligned}$$

The solution is essentially the same as in the $d = 1$ case, with $|p|$ replaced by $|\mathbf{p}|$

$$\widehat{u}(t,\mathbf{p}) = \widehat{f}(\mathbf{p})\cos(2\pi c|\mathbf{p}|t) + \frac{\widehat{g}(\mathbf{p})}{2\pi c|\mathbf{p}|}\sin(2\pi c|\mathbf{p}|t)$$

and this can be written

$$\widehat{u}(t,\mathbf{p}) = \widehat{f}(\mathbf{p})\frac{\partial\widehat{W_t}}{\partial t} + \widehat{g}(\mathbf{p})\widehat{W_t}(\mathbf{p})$$

The solution is then given in terms of convolution by

$$u(t,x) = f * \frac{\partial W_t}{\partial t} + g * W_t$$

In any dimension

$$\widehat{W}(t,\mathbf{p}) = \frac{\sin(2\pi|\mathbf{p}|ct)}{2\pi|\mathbf{p}|c}$$

and

$$W(t,\mathbf{x}) = \int_{\mathbf{R}^d} \frac{\sin(2\pi|\mathbf{p}|ct)}{2\pi|\mathbf{p}|c} e^{2\pi i \mathbf{x}\cdot\mathbf{p}} dp_1 \cdots dp_d$$

Note that the fact that $\widehat{W}(t,\mathbf{p})$ only depends on the length of $\mathbf{p}$ implies that $W(t,\mathbf{x})$ is radial. We can apply the $d = 3$ formula for the Fourier transform of a radial function to the calculation of $W(t,\mathbf{x})$, giving ($r = |\mathbf{x}|, p = |\mathbf{p}|$)

$$\begin{aligned}
W(t,r) &= \frac{2}{r}\int_0^\infty \frac{1}{2\pi cp}\sin(2\pi pct)\sin(2\pi rp)pdp \\
&= \frac{1}{\pi rc}\int_0^\infty \sin(2\pi pct)\sin(2\pi rp)dp \\
&= \frac{1}{2\pi rc}\int_{-\infty}^\infty \left(\frac{1}{2i}\right)(e^{i2\pi pct} - e^{-i2\pi pct})\left(\frac{1}{2i}\right)(e^{i2\pi pr} - e^{-i2\pi pr})dp \\
&= \frac{1}{2\pi rc}\int_{-\infty}^\infty \frac{-1}{4}(e^{i2\pi p(r+ct)} + e^{-i2\pi p(r+ct)} - e^{i2\pi p(r-ct)} - e^{-i2\pi p(r-ct)})dp \\
&= \frac{1}{4\pi rc}(\delta(r-ct) - \delta(r+ct))
\end{aligned}$$

(in the second step we used the fact that the integrand was an even function).

Assuming we are studying solutions for $t > 0$, then

$$W(t,r) = \frac{1}{4\pi rc}\delta(r-ct)$$

The solution with initial value data

$$u(0,\mathbf{x}) = 0, \quad \frac{\partial u}{\partial t}(0,\mathbf{x}) = g(\mathbf{x})$$

will be given by

$$\begin{aligned}
u(t,\mathbf{x}) &= g * W(t,r) \\
&= \int_{\mathbf{R}^3} g(\mathbf{x}-\mathbf{y})\frac{1}{4\pi|\mathbf{y}|c}\delta(|\mathbf{y}|-ct)dy_1dy_2dy_3 \\
&= \int_{\mathbf{R}^3} g(\mathbf{x}+\mathbf{y})\frac{1}{4\pi|\mathbf{y}|c}\delta(|\mathbf{y}|-ct)dy_1dy_2dy_3 \\
&= \int_{\mathbf{R}^3} g(\mathbf{x}+r\widehat{r})\frac{1}{4\pi rc}\delta(r-ct)r^2\sin\theta dr d\theta d\varphi \\
&= \frac{1}{4\pi c^2 t}\int_{S^2} g(\mathbf{x}+ct\widehat{r})d\sigma
\end{aligned}$$

where in the second step we used symmetry of the integral under $\mathbf{y} \to -\mathbf{y}$.

We'll write this (following [3]) as

$$u(t, \mathbf{x}) = tM_t(g)$$

where

$$M_t(g) = \frac{1}{4\pi c^2 t^2} \int_{S^2} g(\mathbf{x} + ct\widehat{r}) d\sigma$$

can be thought of as the average of the function g over a sphere of radius ct centered at $\mathbf{x}$.

Recall that the general solution is given by adding an $f = 0, g \neq 0$ solution to a time derivative of such a solution (giving the $f \neq 0, g = 0$ case), so

$$u(t, \mathbf{x}) = \frac{\partial}{\partial t}(tM_t(f)) + tM_t(g)$$

This is known as Kirchoff's formula. In this case, the Huygens principle says that the solution at $t, \mathbf{x}$ depends just on initial conditions on the sphere of points at distance ct from $\mathbf{x}$.

5.5 The wave equation in $d = 2$

Finding an explicit formula for the solutions to the wave equation in the $d = 2$ case is less straightforward than in $d = 3$, because the formula for the Fourier transform of a radial function is more complicated, involving a Bessel function. One can however proceed as follows (method of "descent"):

- Extend the two-dimensional initial value data

  $$f(x_1, x_2), \quad g(x_1, x_2)$$

  to three-dimensional initial value data, constant in the x_3 direction

  $$\widetilde{f}(x_1, x_2, x_3) = f(x_1, x_2), \quad \widetilde{g}(x_1, x_2, x_3) = g(x_1, x_2)$$

  Note that this is not in $\mathcal{S}(\mathbf{R}^3)$ even if the two-dimensional data is in $\mathcal{S}(\mathbf{R}^2)$, but it will be in $\mathcal{S}'(\mathbf{R}^3)$.

- Use the three-dimensional Kirchoff formula, getting an answer that only depends on x_1, x_2, so gives solutions for $d = 2$.

As in the three-dimensional case, we'll first do the calculation for $f = 0$, so

$$\begin{aligned} u(t, x_1, x_2, x_3) &= tM_t(\widetilde{g}) \\ &= t\frac{1}{4\pi c^2 t^2} \int_0^{2\pi} \int_0^{\pi} \widetilde{g}(\mathbf{x} + ct\widehat{r}) c^2 t^2 \sin\theta d\theta d\varphi \end{aligned}$$

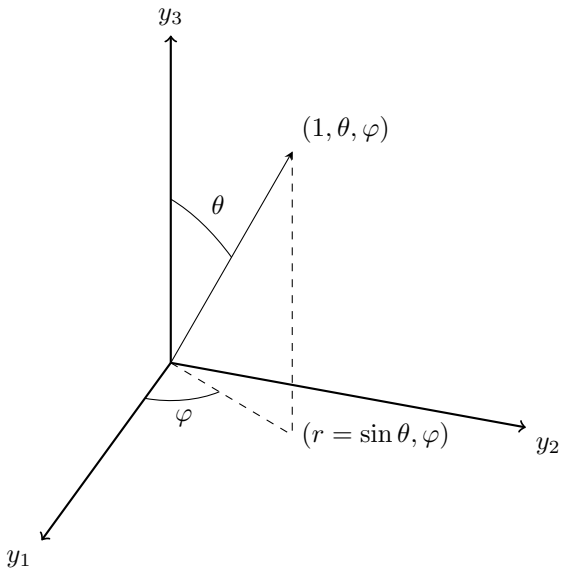
Here

$$\widehat{r} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$$

is the unit vector in the $\mathbf{x}$ direction. We evaluate this integral in two steps, first doing the top hemisphere, then the bottom hemisphere. In the top hemisphere, the θ integration goes from 0 to $\pi/2$, so we want to evaluate the integral over the top half of the unit sphere

$$t\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\frac{\pi}{2}}\widetilde{g}(\mathbf{x}+ct\widehat{r})\sin\theta d\theta d\varphi$$

We can do the integral instead on the projection to the unit disk. When we project down, our coordinates θ, φ on the sphere become a pair of polar coordinates (r, φ) on the unit disk in the (y_1, y_2) plane, with $r = \sin\theta$.



Since $dr = \cos\theta d\theta$, the area element in the integral over the sphere becomes

$$\sin\theta d\theta d\varphi = \frac{\sin\theta}{\cos\theta}drd\varphi = \frac{r}{\sqrt{1-r^2}}drd\varphi$$

If we now change to rectangular coordinates y_1, y_2 on the disk, the area element is

$$\frac{1}{\sqrt{1-|\mathbf{y}|^2}}dy_1dy_2$$

Since $\widetilde{g}$ only depends on $g(y_1, y_2)$, our integral can be written as the following integral over the unit disk

$$t\frac{1}{4\pi}\int_{|\mathbf{y}|<1}g(\mathbf{x}+ct\mathbf{y})\frac{1}{\sqrt{1-|\mathbf{y}|^2}}dy_1dy_2$$

Doing the same thing for the bottom hemisphere of the unit sphere gives the same result. Adding them together we have found that the solution is given by

$$u(t, x_1, x_2) = t\widetilde{M}_t(g)$$

where

$$\widetilde{M}_t(g) = \frac{1}{2\pi}\int_{|\mathbf{y}|<1} g(x_1 + cty_1, x_2 + cty_2)\frac{1}{\sqrt{1-y_1^2-y_2^2}}dy_1dy_2$$

The general solution will be given by

$$u(t, x_1, x_2) = \frac{\partial}{\partial t}(t\widetilde{M}_t(f)) + t\widetilde{M}_t(g)$$

Note that this depends on the initial value data f, g for all points in the disk of radius ct, not just points on the boundary at distance ct from x_1, x_2, showing that the behavior in dimensions 2 and 3 is quite different.

Chapter 6

The finite Fourier transform

6.1 Introduction

In this final section of the course we'll discuss a topic which is in some sense much simpler than the cases of Fourier series for functions on S^1 and the Fourier transform for functions on $\mathbf{R}$, since it will involve functions on a finite set, and thus just algebra and no need for the complexities of analysis. This topic has important applications in the approximate computation of Fourier series (which we won't cover), and in number theory (which we'll say a little bit about).

6.2 The group $\mathbf{Z}(N)$

What we'll be doing is simplifying the topic of Fourier series by replacing the multiplicative group

$$S^1 = \{z \in \mathbf{C} : |z| = 1\}$$

by the finite subgroup

$$\mathbf{Z}(N) = \{z \in \mathbf{C} : z^N = 1\}$$

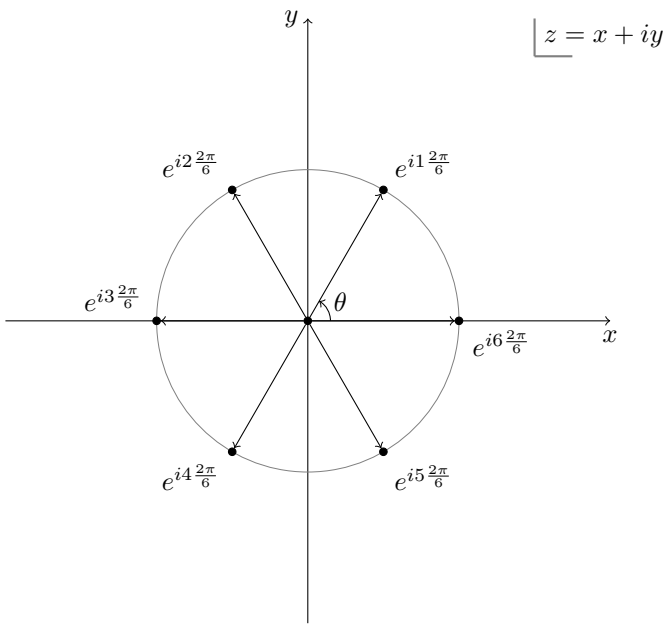
If we write $z = re^{i\theta}$, then

$$\begin{aligned} z^N = 1 &\implies r^N e^{i\theta N} = 1 \\ &\implies r = 1, \quad \theta N = k2\pi \ (k \in \mathbf{Z}) \\ &\implies \theta = \left(\frac{k}{N} 2\pi\right)_{\text{mod } 2\pi} \qquad k = 0, 1, 2, \cdots, N-1 \end{aligned}$$

The group $\mathbf{Z}(N)$ is thus explicitly given by the set

$$\mathbf{Z}(N) = \{1, e^{i\frac{2\pi}{N}}, e^{i2\frac{2\pi}{N}}, \cdots, e^{i(N-1)\frac{2\pi}{N}}\}$$

Geometrically, these are the points on the unit circle one gets by starting at 1 and dividing it into N sectors with equal angles $\frac{2\pi}{N}$. For $N = 6$ one has



The set $\mathbf{Z}(N)$ is a group, with

- identity element 1 ($k = 0$)
- inverse
$$(e^{2\pi i \frac{k}{N}})^{-1} = e^{-2\pi i \frac{k}{N}} = e^{2\pi i \frac{(N-k)}{N}}$$
- multiplication law
$$e^{ik\frac{2\pi}{N}} e^{il\frac{2\pi}{N}} = e^{i(k+l)\frac{2\pi}{N}}$$

One can equally well write this group as the additive group $(\mathbf{Z}/N\mathbf{Z}, +)$ of integers mod N, with isomorphism

$$\begin{aligned} \mathbf{Z}(N) &\leftrightarrow \mathbf{Z}/N\mathbf{Z} \\ e^{ik\frac{2\pi}{N}} &\leftrightarrow [k]_N \\ 1 &\leftrightarrow [0]_N \\ e^{-ik\frac{2\pi}{N}} &\leftrightarrow -[k]_N = [-k]_N = [N-k]_N \end{aligned}$$

6.3 Fourier analysis on $\mathbf{Z}(N)$

An abstract point of view on the theory of Fourier series is that it is based on exploiting the existence of a particular orthonormal basis of functions on the group S^1. These basis elements are eigenfunctions of the linear transformations

given by rotations. The orthonormal basis elements are the $e_m = e^{im\theta},\ m \in \mathbf{Z}$, recalling that

$$\langle e_n, e_m \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{in\theta} e^{-im\theta} d\theta = \delta_{n,m}$$

Rotation by an angle ϕ acts on the circle S^1 by

$$\theta \to \theta + \phi$$

and on functions on the circle by the linear transformation

$$f(\theta) \to (T_\phi f)(\theta) = f(\theta + \phi)$$

(note that the rotation transformation on S^1 itself is not a linear transformation, since S^1 is not a linear space). The functions e_n are eigenfunctions of T_ϕ, since

$$(T_\phi e_n)(\theta) = e^{in(\theta+\phi)} = e^{in\phi} e^{in\theta} = e^{in\phi} e_n$$

We would like to do the same thing for functions on $\mathbf{Z}(N)$: find an orthonormal set of such functions that are eigenvalues for the action of the set $\mathbf{Z}(N)$ on itself by discrete rotations. We'll write a complex-valued function on $\mathbf{Z}(N)$ as

$$F : [k] \in \mathbf{Z}(N) \to F(k) \in \mathbf{C},\ F(k) = F(k+N)$$

For inner product we'll take

$$\langle F, G \rangle = \sum_{k=0}^{N-1} F(k) \overline{G(k)}$$

so

$$||F||^2 = \sum_{k=0}^{N-1} |F(k)|^2$$

With these choices we have

Claim. *The functions $e_l : \mathbf{Z}(N) \to \mathbf{C}$ given by*

$$e_l(k) = e^{i2\pi \frac{lk}{N}}$$

for $l = 0.1, 2, \cdots, N-1$ satisfy

$$\langle e_l, e_m \rangle = N \delta_{l,m}$$

so the functions

$$e_l^* = \frac{1}{\sqrt{N}} e_l$$

are orthonormal. They form a basis since there are N of them and the space of functions on $\mathbf{Z}(N)$ is N-dimensional.

Proof. First define

$$W_N = e^{i\frac{2\pi}{N}}$$

then

$$\begin{aligned}\langle e_l, e_m\rangle &= \sum_{k=0}^{N-1} e^{i\frac{2\pi}{N}lk} e^{-i\frac{2\pi}{N}mk} \\ &= \sum_{k=0}^{N-1} (W_N)^{(l-m)k}\end{aligned}$$

If $l = m$ this is a sum of N 1's, so

$$\langle e_l, e_m\rangle = N$$

If $l \neq m$, let $q = W_N^{l-m}$. Then the sum is

$$\langle e_l, e_m\rangle = 1 + q + q^2 + \cdots + q^{N-1} = \frac{1-q^N}{1-q}$$

but this is 0 since $q^N = (W_N)^N = 1$. □

Our analog of the Fourier series

$$F(\theta) = \sum_{n=-\infty}^{\infty} a_n e^{in\theta}$$

for a function F on S^1 will be writing a function on $\mathbf{Z}(N)$ in terms of the orthonormal basis $\{e_n^*\}$, as

$$\begin{aligned}F(k) &= \sum_{n=0}^{N-1} \langle F, e_n^*\rangle e_n^* \\ &= \sum_{n=0}^{N-1} \langle F, e_n^*\rangle \frac{1}{\sqrt{N}} e^{i2\pi\frac{nk}{N}}\end{aligned}$$

The analog of the Fourier coefficients

$$a_n = \widehat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} F(\theta) e^{-in\theta}$$

will be the finite set of numbers

$$\begin{aligned}\widehat{F}(n) &= \frac{1}{\sqrt{N}}\langle F, e_n^*\rangle = \frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} F(k)\frac{1}{\sqrt{N}} e^{-i2\pi\frac{kn}{N}} \\ &= \frac{1}{N}\sum_{k=0}^{N-1} F(k) e^{-i2\pi\frac{kn}{N}}\end{aligned}$$

for $n = 0, 1, 2, \cdots, N-1$. Here the Fourier inversion theorem is automatic, just the usual fact that for finite dimensional vector spaces the coefficients of a vector with respect to an othonormal basis are given by the inner products of the vector with the basis elements.

For another perspective on this, note that there are two distinguished orthonormal bases for functions on $\mathbf{Z}(N)$

- the N functions of k given by

$$\delta_{kl} = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}$$

  for $l = 0, 1, 2, \cdots, N-1$.

- the N functions of k given by

$$\frac{1}{\sqrt{N}} e^{i2\pi \frac{kl}{N}}$$

  for $l = 0, 1, 2, \cdots, N-1$.

The Fourier transform for $\mathbf{Z}(N)$ that takes

$$\mathcal{F} : \{F(0), F(1), \cdots, F(N-1)\} \to \{\widehat{F}(0), \widehat{F}(1), \cdots \widehat{F}(N-1)\}$$

is just the change of basis matrix between the above two bases. It can be written as an $N \times N$ complex matrix.

The Plancherel (or Parseval) theorem in this case is automatic from linear algebra: in the complex case, a change of basis between two orthonormal bases is given by a unitary matrix. Note that the way we have defined things, the coefficients with respect to the second orthonormal basis are given by the function $\sqrt{N}\widehat{F}$, not $\widehat{F}$, so the theorem says that

$$\sum_{k=0}^{N-1} |F(k)|^2 = \sum_{k=0}^{N-1} |\sqrt{N}\widehat{F}(k)|^2 = N \sum_{k=0}^{N-1} |\widehat{F}(k)|^2$$

Just as for Fourier series and transforms, one can define a convolution product, in this case by

$$(F * G)(k) = \sum_{l=0}^{N-1} F(k-l)G(l)$$

and show that the Fourier transform takes the convolution product to the usual point-wise product.

6.4 Fourier analysis on commutative groups

The cases that we have seen of groups $G = S^1, \mathbf{R}, \mathbf{Z}(N)$, are just special cases of a general theory that works for any commutative group, i.e. any set with an

associative, commutative ($ab = ba$) multiplication, with an identity element and inverses. When the set is finite, this general theory is very straightforward, but for infinite sets like S^1 and $\mathbf{R}$ one needs to take into account more complicated issues (e.g. those of analysis that we have run into).

The general theory starts with the definition

Definition (Group character)**.** *A character of a group G is a function*

$$e : G \to \mathbf{C}^*$$

such that

$$e(ab) = e(a)e(b)$$

(one says that e is a "homomorphism"). Here $\mathbf{C}^*$ is the multiplicative group of non-zero elements of $\mathbf{C}$.

When G is a finite group, all elements will have finite order ($a^n = 1$ for some n) and thus

$$e(a^n) = e(a)^n = 1$$

so characters will take as values not general non-zero complex numbers, but n'th roots of unity, so in the subgroup $U(1) \subset \mathbf{C}^*$ of elements of the form $e^{i\theta}$. Such characters will be called "unitary characters".

For the case of $G = \mathbf{Z}(N)$, the

$$e_l(k) = e^{i2\pi \frac{lk}{N}}$$

are characters, since

$$e_l(k)e_l(m) = e_l(k+m)$$

We will denote the set of unitary characters of a group G by $\widehat{G}$, and we have

Claim. *$\widehat{G}$ is a commutative group. It will be called the "character group" of G.*

Proof. $1 \in \widehat{G}$ is the identity function $e(a) = 1$, multiplication is given by

$$(e_1 \cdot e_2)(a) = e_1(a)e_2(a)$$

and the inverse of a character e is given by

$$e^{-1}(a) = (e(a))^{-1}$$

□

Some of the examples we have seen so far of pairs G and $\widehat{G}$ are

- The group $G = \mathbf{Z}(N)$, with elements $k = 0, 1, \cdots, N-1$, has character group $\widehat{G} = \mathbf{Z}(N)$, which has elements e_l for $l = 0, 1, \cdots, N-1$ given by the functions
$$e_l(k) = e^{i2\pi \frac{kl}{N}}$$

- The group $G = S^1$, with elements $e^{i\theta}$, has character group $\widehat{G} = \mathbf{Z}$, with the integer n corresponding to the function

$$e_n(\theta) = e^{in\theta}$$

- The group $G = \mathbf{R}$. with elements x, has character group $\widehat{G} = \mathbf{R}$, which has elements e_p for $p \in \mathbf{R}$ given by the functions

$$e_p(x) = e^{i2\pi px}$$

For finite groups one can define an inner product on $\widehat{G}$ by

$$\langle e, e' \rangle = \frac{1}{|G|} \sum_{a \in G} e(a)\overline{e'(a)}$$

where e, e' are characters of G. These have the property

Claim. *Distinct elements of $\widehat{G}$ are orthonormal.*

Proof. For $e = e'$, one has

$$\langle e, e \rangle = \frac{1}{|G|} \sum_{a \in G} e(a)\overline{e(a)} = \frac{1}{|G|} \sum_{a \in G} 1 = 1$$

For $e \neq e'$ one has

$$\begin{aligned} \langle e, e' \rangle &= \frac{1}{|G|} \sum_{a \in G} e(a)\overline{e'(a)} \\ &= \frac{1}{|G|} \sum_{a \in G} e(a)(e'(a))^{-1} \end{aligned}$$

Picking an element $b \in G$ such that $e(b) \neq e'(b)$ (possible since e, e' are different functions) and using the fact that multiplication of all elements by b is just a relabeling of the group elements, the above gives

$$\begin{aligned} &= \frac{1}{|G|} \sum_{a \in G} e(ba)(e'(ba))^{-1} \\ &= \frac{1}{|G|} \sum_{a \in G} e(b)e(a)(e'(a))^{-1}(e'(b))^{-1} \\ &= e(b)(e'(b))^{-1} \frac{1}{|G|} \sum_{a \in G} e(a)(e'(a))^{-1} \end{aligned}$$

So we have shown

$$\langle e, e' \rangle = e(b)(e'(b))^{-1} \langle e, e' \rangle$$

but by assumption we have

$$e(b)(e'(b))^{-1} \neq 1$$

so we must have $\langle e, e' \rangle = 0$. □

Somewhat harder to prove is that the unitary characters are complete, giving a basis for functions on G. For a proof of this, see pages 233-234 of [3].

We can now define the Fourier transform for any finite commutative group G

Definition. *The Fourier transform for a finite commutative group G takes the function f on G to the function on $\widehat{G}$*

$$\mathcal{F}f = \widehat{f}(e) = \langle f, e\rangle = \frac{1}{|G|}\sum_{a\in G} f(a)\overline{e(a)}$$

The Fourier inversion formula again is just the expansion of the function in the orthnormal basis of characters

$$f = \sum_{e\in\widehat{G}} \widehat{f}(e)e = \mathcal{F}^{-1}f$$

$\mathcal{F}$ is just the transformation of basis matrix between the basis of characters and the basis of "δ-functions" (= 1 on one element, 0 on the others). It will be given by a unitary $|G| \times |G|$ matrix, implying a Plancherel/Parseval theorem

$$||f||^2 = \frac{1}{|G|}\sum_{a\in G} |f(a)|^2 = \sum_{e\in\widehat{G}} |\widehat{f}(e)|^2 = ||\widehat{f}||^2$$

One can define a convolution product by

$$f * g(a) = \frac{1}{|G|}\sum_{b\in G} f(ab^{-1})g(b)$$

which will satisfy

$$\widehat{f * g} = \widehat{f}\widehat{g}$$

Note that

$$e * e' = \begin{cases} e \ if \ e = e' \\ 0 \ if \ e \neq e' \end{cases}$$

So

$$f \to f * e$$

is a projection map, onto the subspace of functions ϕ on G that are eigenvectors for the linear transformation

$$\phi(a) \to (T_b\phi)(a) = \phi(ba)$$

with eigenvalue $e(b)$.

Given two groups G_1 and G_2, one can form a new group, the product group $G_1 \times G_2$. This is the group with elements pairs

$$(a_1, a_2), \quad a_1 \in G_1, \ a_2 \in G_2$$

and multiplication law

$$(a_1, a_2)(b_1, b_2) = (a_1 a_2, b_1 b_2)$$

It is not hard to show that for finite commutative groups the character groups satisfy

$$\widehat{G_1 \times G_2} = \widehat{G_1} \times \widehat{G_2}$$

We won't cover this in this course since it would take us too far afield into algebra, but in a standard abstract algebra course you will learn a theorem that says that any finite commutative group G is isomorphic to the product group

$$\mathbf{Z}(N_1) \times \mathbf{Z}(N_2) \times \cdots \times \mathbf{Z}(N_k)$$

for some positive integers $N_1, N_2, \ldots, N_k$. Its character group $\widehat{G}$ will then by the above be

$$\widehat{\mathbf{Z}(N_1)} \times \widehat{\mathbf{Z}(N_2)} \times \cdots \times \widehat{\mathbf{Z}(N_k)}$$

We see that for a general finite commutative group G and $\widehat{G}$ will be isomorphic and the example we have worked out of $\mathbf{Z}(N)$ is fundamental: the general case is just a product of these for different N. In the next section though, we will see that given an interesting finite commutative group, finding its decomposition into $\mathbf{Z}(N)$ factors can be quite non-trivial.

One generalization of these ideas is to the case of general commutative groups, not necessarily finite. Here one can get into quite complicated questions in analysis, some of which we have seen in the cases $G = S^1$ and $G = \mathbf{R}$. An even larger generalization is to the case of non-commutative groups. To get an analog there for Fourier analysis, one needs to consider not just characters

$$e : G \to U(1)$$

but more general maps

$$\pi : G \to U(n)$$

satisfying the homomorphism property

$$\pi(a)\pi(b) = \pi(ab)$$

where $U(n)$ is the group of unitary $n \times n$ matrices. Identifying all such π is a difficult problem, but once one does so, for each such π one gets an $n \times n$ matrix-valued function on G. The analog of Fourier analysis in this case is the decomposition of arbitrary functions on G in terms of the functions given by the matrix elements of these matrix-valued functions.

6.5 Fourier analysis on $\mathbf{Z}^*(q)$

We'll now turn to some applications of the finite Fourier transform in number theory. These are based on considering not the additive structure on $\mathbf{Z}(N)$, but

the multiplicative structure. One can define a multiplication on the set $\mathbf{Z}(N)$ by

$$[l]_N \cdot [m]_N = [lm]_N$$

This product is commutative, associative, and there's an identity element ($[1]_N$). This makes $\mathbf{Z}(N)$ an example of what algebraists call a "ring." The problem though is that many elements of $\mathbf{Z}(N)$ have no multiplicative inverse. For example, if one takes $N = 4$ and looks for an integer m such that

$$([2]_4)^{-1} = [m]_4$$

the integer m must satisfy

$$[2]_4 \cdot [m]_4 = [2m]_4 = [1]_4$$

But this can't possibly work since $2m$ is even and 1 is odd.

We can however go ahead and define a group by just taking the elements of $\mathbf{Z}(N)$ that do have a multiplicative inverse (such elements are called "units" of the ring $\mathbf{Z}(N)$).

Definition. *The group $\mathbf{Z}^*(q)$ is the set of $[l]_q$ of elements of $\mathbf{Z}(q)$ that have a multiplicative inverse, with the group law multiplication mod q.*

For an alternate characterization of $\mathbf{Z}^*(q)$, first recall that every positive integer $N > 1$ can be uniquely factored in primes, meaning

$$N = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

For any integers a and b we can define

$$\gcd(a, b)$$

to be the largest integer that divides both a and b. If $\gcd(a, b) = 1$ we say that "a and b are relatively prime". This is equivalent to saying that they have no prime factors in common. The group $\mathbf{Z}^*(q)$ could instead have been defined as

$$\mathbf{Z}^*(q) = \{[l]_q \in \mathbf{Z}(q) : \gcd(l, q) = 1\}$$

To see one direction of the equivalence with the other definition, that an element $[l]_q \in \mathbf{Z}(q)$ having an inverse implies that l and q are relatively prime, start by assuming they aren't relatively prime, which means

$$l = pn_1, \quad q = pn_2$$

for some prime p and integers n_1, n_2. In order to find an inverse of $[l]_q$, we need to find an integer m such that

$$[lm]_q = [1]_q$$

which means that

$$lm = nq + 1$$

for some integer n. Under the assumption $\gcd(l, q) \neq 1$ we find that we need to solve

$$pn_1 m = npn_2 + 1$$

which implies

$$n_1 m = nn_2 + \frac{1}{p}$$

but the left hand side is an integer, while the right hand side is a non-trivial fraction. The contradiction implies there is no inverse.

The number of elements of $\mathbf{Z}^*(q)$ is the number of elements of $\{1, 2, \cdots, q-1\}$ that are relatively prime to q. This number defines the Euler ϕ or "totient" function, i.e.

$$\phi(q) = |\mathbf{Z}^*(q)|$$

If q is a prime p, then $1, 2, \ldots, p-1$ are relatively prime to p and all $p-1$ elements of the set

$$[1]_p, [2]_p, \cdots [p-1]_p$$

have multiplicative inverses. In this case one can show (although we won't do it here) that the multiplicative group $\mathbf{Z}^*(p)$ is isomorphic to the additive group $\mathbf{Z}(p-1)$. As noted before, there is a general theorem that any finite commutative group is isomorphic to a product of groups $\mathbf{Z}(N_j)$ for various N_j.

For any given q that isn't prime, finding the integers N_j and the isomorphism for the case of $\mathbf{Z}^*(q)$ is a non-trivial problem. One can easily though work out what happens for small q. For example, if $q = 4$ we find that $\mathbf{Z}^*(4)$ has two elements

$$[1]_4, [3]_4$$

(since $1, 3$ are the only integers from 0 to 3 relatively prime to 4). One can see that there is an isomorphism of groups between $\mathbf{Z}^*(4)$ and $\mathbf{Z}(2)$ given by

$$[1]_4 \leftrightarrow ([0]_2, +)$$

$$[3]_4 \leftrightarrow ([1]_2, +)$$

since the first of these is the identity, the second an element that squares to the identity

$$[3]_4 \cdot [3]_4 = [9]_4 = [1]_4 \leftrightarrow [1]_2 + [1]_2 = [2]_2 = [0]_2$$

By the general theory, the character group $\widehat{\mathbf{Z}^*(q)}$ has $\phi(q)$ elements. These can be thought of as functions $e(k)$ on the classes $[k]_q$, which are only non-zero. on the k that are relatively prime to q. In number theory these characters appear as "Dirichlet characters", defined for all integers m by

$$\chi_e(m) = \begin{cases} e([m]_q), & \gcd(m, q) = 1 \\ 0, & \gcd(m, q) \neq 1 \end{cases}$$

By the homomorphism property of characters, these are multiplicative functions on $\mathbf{Z}$, satisfying

$$\chi_e(nm) = \chi_e(n)\chi_e(m)$$

6.6 The zeta function, primes and Dirichlet's theorem

Recall that earlier on in this course we studied the zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

using Poisson summation to derive the functional equation for $\zeta(s)$. One of the main applications of the zeta function is to the study of the distribution of prime numbers, based on

Claim (Euler product formula)**.** *For $s > 1$*

$$\prod_{primes\ p} \frac{1}{1-p^{-s}} = \zeta(s)$$

Proof. The geometric series formula gives

$$\frac{1}{1-\frac{1}{p^s}} = 1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \cdots$$

Taking the infinite product

$$\prod_{\text{primes } p_j} (1 + \frac{1}{p_j^s} + \frac{1}{p_j^{2s}} + \frac{1}{p_j^{3s}} + \cdots)$$

for $p_1 < p_2 < \cdots$, and writing this out as a sum of terms, one gets all terms of the form

$$\frac{1}{p_1^{n_1}} \frac{1}{p_2^{n_2}} \cdots \frac{1}{p_k^{n_k}}$$

with coefficient 1, each raised to the power s. By unique factorization of integers this sum is the same sum as

$$\sum_{n=1}^{\infty} \frac{1}{n^s}$$

□

This can be used to give a proof that there are an infinite number of primes (of course there is a much simpler proof by contradiction: multiply all primes and add 1). The argument is that we know that

$$\lim_{s \to 1^+} \zeta(s) = 1 + \frac{1}{2} + \frac{1}{3} + \cdots = \infty$$

but by the Euler product formula, this is

$$\lim_{s \to 1^+} \prod_{\text{primes } p} \frac{1}{1-p^{-s}}$$

which can only be infinite if the number of primes is infinite. For a more difficult example, there's

Claim. *The sum*

$$\sum_{primes\ p} \frac{1}{p}$$

diverges.

Proof. Taking the logarithm of the Euler product formula

$$\prod_{\text{primes } p} \frac{1}{1-p^{-s}} = \zeta(s)$$

one finds

$$\ln(\zeta(s)) = -\sum_p \ln(1-p^{-s})$$

Using the power series expansion

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

one has the inequality (for $|x| < \frac{1}{2}$)

$$\begin{aligned} |\ln(1+x) - x| &< \frac{x^2}{2}(1 + |x| + |x|^2 + \cdots) \\ &< \frac{x^2}{2}(1 + \frac{1}{2} + (\frac{1}{2})^2 \cdots) \\ &= \frac{x^2}{2}(\frac{1}{1-\frac{1}{2}}) = x^2 \end{aligned}$$

So

$$|\ln(1-p^{-s}) - p^{-s}| < p^{-2s}$$

which implies that in the sum above one can replace $\ln(1-p^{-s})$ by $\frac{1}{p^s}$, changing the result by at most

$$\sum_p \frac{1}{p^{2s}}$$

But

$$\sum_p \frac{1}{p^{2s}} < \sum_{n=1}^{\infty} \frac{1}{n^{2s}} = \zeta(2s)$$

and

$$\lim_{s\to 1^+} \zeta(2s) = \zeta(2) = \frac{\pi^2}{6}$$

which is finite. So

$$\lim_{s\to 1^+} \sum_p \frac{1}{p^s}$$

is infinite, since it differs by less than a finite constant from

$$\lim_{s\to 1^+} \ln(\zeta(s))$$

which we know to be infinite.

□

For more subtle aspects of the distribution of primes, one can define a generalization of the zeta function, which uses the characters of $\mathbf{Z}^*(q)$. These are called Dirichlet L-functions, and defined by

Definition (Dirichlet L-function)**.** *For χ a Dirichlet character, the corresponding Dirichlet L-function is*

$$L(\chi, s) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$$

By much the same argument as for the Euler product formula, these satisfy a product formula

$$L(\chi, s) = \prod_{\text{primes } p} \frac{1}{1 - \chi(p)p^{-s}}$$

By studying the logarithm of this and its limit as $s \to 1^+$, and exploiting Fourier analysis on $\mathbf{Z}^*(q)$, one can show

Theorem (Dirichlet's theorem)**.** *For a fixed l and q, gcd$(l, q) \neq 1$*

$$\lim_{s \to 1^+} \sum_{primes \ p:[p]_q = l} \frac{1}{p^s}$$

diverges, which implies that there is an infinite number of primes in any arithmetic progression (sequence of the form $\{l, l + q, l + 2q, \cdots\}$).

Proof. The proof is rather complicated and we won't have time to go through it this semester. It can be found in chapter 8 of [3]. □

Bibliography

[1] Howell, Kenneth B., Principles of Fourier Analysis. CRC Press, 2016.

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

[3] Stein, Elias M. and Shakarchi, Rami, Fourier Analysis: An Introduction. Princeton University Press, 2003.

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