Harmonic Analysis

Jacob Denson

August 2, 2018

Table Of Contents

Part I

Classical Fourier Analysis

Deep mathematical knowledge often arises hand in hand with the recognition of symmetry. Nowhere is this more clear than in the foundations of harmonic analysis, where we attempt to understand 'oscillating' mathematical objects, in their various forms. More particularly, we attempt to find transformations on certain classes of functions which are well behaved under linear shifts $f(x) \mapsto f(x + a)$. In the mid 18th century, problems in mathematical physics led D. Bernoulli, D'Alembert, Lagrange, and Euler to consider 2π periodic functions representable as a trigonometric series

$$f(t) = A + \sum_{n=1}^{\infty} B_n \cos(nt) + C_n \sin(kt)$$

In 1811, in his book, Théorie Analytique de la Chaleur, Joseph Fourier had the audacity to announce that *all* functions were representable in this form, and used it to solve the classical linear partial differential equations of physics. His conviction is the reason the classical theory of harmonic analysis is often named Fourier analysis, where we analyze the degree to which Fourier's proclamation holds, as well as it's paired statement on the real line, that a function f on the real line can be written as

$$f(t) = \int_{-\infty}^{\infty} A(\xi)\cos(\xi t) + B(\xi)\sin(\xi t) \, d\xi$$

for some functions A and B. In the 1820s, Poisson, Cauchy, and Dirichlet all attempted to form rigorous proofs that 'Fourier summation' holds for all functions. Their work is responsible for most of the modern subject of analysis we know today. The biggest question we will ask is how we interpret the convergence of Fourier series. Pointwise convergence is not enough to justify most analytic techniques, and what's more, under pointwise convergence, the representation of a function by Fourier series need not be unique. Uniform convergence is useful, but too hopeful to obtain for all functions; it need not even hold for crucial classes such as the family of continuous functions. This means we must introduce more subtle methods to measure convergence.

Chapter 1

Springs, Strings, and Symmetry

In this chapter, we look at the classical problems that causes Fourier and his peers to consider Fourier summation. Key to these ideas will be the oscillatory functions sine and cosine. Well before Fourier's time, Newton had encountered oscillation motion in the study of a spring. It is well known that solutions to the equation $\ddot{x} = -k^2 x$, are given by

$$x = A\cos(kt) + B\sin(kt) = C\cos(kt + \phi)$$

where the two representations are connected by the trigonometric equality

$$\cos(x + y) = \cos(x)\cos(y) - \sin(x)\sin(y)$$

ϕ is known as the **phase** of the oscillation, C is the **amplitude**, and $k/2\pi$ the **frequency**. No serious effort was required on our part of produce these equations. They were known to Newton, and to Hooke before him, and require only the basic methods of the calculus. But this equation is very important; it forebodes that trigonometric functions will occur over and over again in the study of oscillatory behaviour.

1.1   The Wave Equation

Fourier's main contribution to the foundation of harmonic analysis, namely, the existence of an expansion

$$f(t) = A + \sum_{n=1}^{\infty} B_n \cos(nt) + C_n \sin(kt)$$

for 2π periodic functions on the real line, originated in the theory of waves. Physical considerations lead Fourier to believe that the partial differential equation defining wave propogation was given by the **wave equation**

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

where $u(t,x)$ gives the amplitude of the wave at some spatial point x and time t. There is an obvious connection between the dynamics of waves and springs; both describe motion under the effects of a resistance from equilibrium. What makes the string's motion tricky to analyze is that the motion is infinite dimensional; the physical state of the string at any particular time is described by a function $u(t,\cdot)$ on the real line, which consists of a specifying the position of infinitely many points. This makes the analysis of the wave equation much harder than the harmonic oscillator.

The easiest version of this equation to solve is where we analyze a string tethered down between two endpoints, so that the wave propogation is one dimensional, where the partial differential equation can be written as

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

We assume, by changing our unit of distance, that u is tethered down at $x = 0$ and $x = \pi$, so we try and find solutions $u : [0,\pi] \times \mathbf{R} \to \mathbf{R}$ with $u(0,t) = u(\pi,t) = 0$. As can be observed in real life, strings can undergo harmonic motion, oscillating back and forth. Mathematically, this type of motion can be modelled by writing $u(x,t) = f(x)g(t)$, where $f(x)$ gives the initial profile of the string, and $g(t)$ is some oscillatory motion factor with $g(0) = 1$. Under this *variable separation*, the partial differential equation becomes an ordinary differential equation $f''(x)g(t) = f(x)g''(t)$. Unless f or g are equal to zero across the whole domain, this means there exists a value $\lambda \geqslant 0$ such that $f''(x) = -\lambda^2 f(x)$, $g''(x) = -\lambda^2 g(x)$ (the constant value must be negative in order for g to have oscillatory behaviour). It is easy to solve these equations, writing

$$f(x) = f(0)\cos(\lambda x) + \frac{f'(0)}{\lambda}\sin(\lambda x)$$

$$g(t) = g(0)\cos(\lambda t) + \frac{g'(0)}{\lambda}\sin(\lambda t)$$

Since we assume $f(0) = f(\pi) = 0$, we must have $\sin(\lambda\pi) = 0$, so that λ is an integer m. In conclusion, this means that

$$u(x,t) = A\sin(mx)\cos(\phi + mt) = \sin(mx)[B\cos(mt) + C\sin(mt)]$$

These functions are the harmonics. If you've ever learned to play music, these are the 'pure tones', which are combined together to product melodious sound.

It was Fourier who had the audacity to suggest that one could produce *all* solutions to the wave equation from these base tones. Since the wave equation is a *linear* partial differential equation, the set of solutions forms a linear class of functions, and we can therefore obtain a more complicated family of solutions to the wave equation of the form

$$u(t,x) = \sum_{m=1}^{n} \sin(mx)(A_m\cos(mt) + B_m\sin(mt))$$

Fourier said that these were *all* such solutions, provided we take $n \to \infty$, and consider an infinite series. Now given initial conditions $u(0,x) = f(x)$, we find

$$f(x) = \sum_{m=0}^{\infty} A_m\sin(mx)$$

and given an initial velocity function $\partial_t u(0,x) = g(x)$, by performing a formal differentiation, we should have

$$g(x) = \sum_{m=0}^{\infty} mB_m\sin(mx)$$

Thus in order to find the constants A_m, B_m which give the motion of a string in terms of harmonic frequencies, it suffices to decompose an arbitrary one dimensional function on $[0,\pi]$ into the sum of sinusoidal functions of differing frequency. The first problem of Fourier analysis is the investigation of the limits of this method; How do we obtain the coefficients of the sum from the function itself, and how can we ensure these coefficients reflect the original function?

1.2 The Fourier Coefficients of a Function

The question of obtaining the coefficients can be approached by a formal calculation. Suppose that a function f has an expansion

$$f(x) = \sum_{n=0}^{\infty} A_n \sin(nx)$$

Using the fact that the sin functions are orthogonal, in the sense that

$$\int_0^{\pi} \sin(mx)\sin(nx) = \begin{cases} 0 & m \neq n \\ \frac{\pi}{2} & m = n \end{cases}$$

We find that, ignoring issues of interchanging sums and integrals,

$$\int_0^{\pi} f(x)\sin(mx)dx = \int_0^{\pi} \sum_{n=0}^{\infty} A_n \sin(nx)\sin(mx)$$

$$= \sum_{n=0}^{\infty} \int_0^{\pi} A_n \sin(nx)\sin(mx) = \frac{\pi}{2} A_m$$

Given any function $f : [0, \pi] \to \mathbf{R}$, a reasonable candidate for the coefficients is

$$A_n = \frac{2}{\pi} \int_0^{\pi} f(x)\sin(mx)$$

These values will be known as the **Fourier coefficients** of the function f.

1.3 The Heat Equation

Suppose we have a two-dimensional region D, with a heat distribution fixed on the buondary, upon which temperature fluctuates in the interior. The equation modelling the evolution of the heat distribution over time is the **heat equation**

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

where Δu is the Laplacian operator. Here we are describing evolution over functions in two dimensions, which is also an infinite dimensional configuration.

To simplify again, we start by looking at only the steady state heat equations, those functions u satisfying $\Delta u = 0$, known as **harmonic functions**. Normally, we fix the boundary of a set C, and attempt to find a solution on the interior satisfying the boundary condition - physically, we fix a temperature on the boundary, wait for a long time, and see how the heat disperses on the interior. For now, let's consider functions on the unit disk **D**, with fixed values on the boundary S^1. In this domain, we can switch to polar coordinates, in which the Laplacian operator takes the form

$$\Delta u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}$$

We then apply the method of separation of coordinates. If $\Delta u = 0$, then

$$r^2\frac{\partial^2 u}{\partial r^2} + r\frac{\partial u}{\partial r} = -\frac{\partial^2 u}{\partial \theta^2}$$

Writing $u(r,\theta) = f(r)g(\theta)$, the equation above reads

$$r^2 f''(r)g(\theta) + rf'(r)g(\theta) = -f(r)g''(\theta)$$

$$\frac{r^2 f''(r) + rf'(r)}{f(r)} = -\frac{g''(\theta)}{g(\theta)}$$

This means that both sides are equal to a constant λ^2 for some λ (If the constant value was negative, g wouldn't be periodic). Solving these equations tells us

$$g''(\theta) = -\lambda g(\theta) \qquad r^2 f''(r) + rf'(r) - \lambda f(r) = 0$$

Then we have

$$g(\theta) = A\cos(\lambda\theta) + B\sin(\lambda\theta)$$

Since g is 2π periodic, we require λ to be an integer m. The equation for f can be solved when $m \neq 0$ to be

$$f(r) = Ar^m + Br^{-m}$$

and for physical reasons, we force $f(r)$ to be bounded at zero, so $B = 0$, and we find the only solutions with separable variables are

$$u(r,\theta) = [A\cos(m\theta) + B\sin(m\theta)]r^m = A\cos(m\theta - \phi)r^m$$

8

When $m = 0$, the solution is just constant, because the solutions to $rf''(r) + f'(r) = 0$ are described by the equation $f(r) = A\log(r) + B$, which is unbounded near the origin unless $A = 0$. After our previous work, we would hope that all solutions are of the form

$$u(r, \theta) = C + \sum_{m=1}^{\infty} [A_m \cos(m\theta) + B_m \sin(m\theta)]r^m$$

If we know the values of u at $r = 1$, then we may apply the same expansion technique of the wave equation, except now we are trying to expand a functions on $[-\pi, \pi]$ in sines *and* cosines. Noticing that sines and cosines remain orthogonal under integration on $[-\pi, \pi]$, and that

$$\int_{-\pi}^{\pi} \sin^2(mt) \, dt = \int_{-\pi}^{\pi} \cos^2(mt) \, dt = \pi$$

the coefficents of the expansion should be

$$A_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin(t) \, dt \qquad B_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos(t) \, dt$$

If we begin with a function on $[0, \pi]$, and enlarge the domain to $[-\pi, \pi]$ by making the function odd, then the B_m all vanish, and we obtain the same expansion as on $[0, \pi]$. Since an arbitrary function f on $[-\pi, \pi]$ can be written as the sum of odd and even functions, expansion on $[-\pi, \pi]$ in terms of sin and cos is no more general than an expansion on $[0, \pi]$, and for our analysis, choosing either method is up to style. However, in the next section, we will introduce an even more elegant notation, applying complex exponentials, which will make the problem the simplest possible.

1.4   The Fundamental Oscillator

We are working with 2π-periodic functions $f : \mathbf{R} \to \mathbf{R}$, and attempting to decompose them into summations of sines and cosines. Unfortunately, manipulations of sines and cosines are quite difficult to do homogenously. We now introduce a third periodic function which will encompass sines and cosines together. First, define the circle group $\mathbf{T}$ to be the set of complex numbers z with $|z| = 1$. Functions from $\mathbf{T}$ to $\mathbf{R}$ naturally correspond

to 2π-periodic functions; given $g : \mathbf{T} \to \mathbf{R}$, the correspondence is given by the equation

$$f(t) = g(e^{it})$$

Thus, when defining 2π periodic functions, we shall make no distinction between a function 'defined in terms of t' and a function 'defined in terms of z', after making the explicit identification $z = e^{it}$. Then an expansion of the form

$$f(t) = \sum_{k=0}^{\infty} A_k \cos(kt) + \sum B_k \sin(kt)$$

using Euler's identity $e^{it} = \cos t + i \sin t$, leads to an expansion

$$f(z) = \sum_{k=0}^{\infty} A_k \, \mathfrak{Re}[z^k] + B_k \, \mathfrak{Im}[z^k]$$
$$= \sum_{k=0}^{\infty} A_k \left(\frac{z^k + z^{-k}}{2} \right) - iB_k \left(\frac{z^k - z^{-k}}{2} \right) = \sum_{k=-\infty}^{\infty} C_k z^k$$

so a Fourier expansion on $[0, 2\pi]$ is really just a power series expansion on the circle in disguise. Thus expanding a real-valued function in e^{kit} is the same as expanding the function in terms of sines and cosines. The complex exponentials e^{kit} have the same orthogonality properties as sin and cos, so given a function f, the coefficients C_k can be found by the expansion

$$C_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-kit} dt$$

Thus a periodic function f gives rise to a function

$$\widehat{f}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-kit} dt = \fint_{-\pi}^{\pi} f(t) e^{-kit} \, dt$$

defined on $\mathbf{Z}$, called the **Fourier series** of f, which measures the average value of f when it is 'twisted' by an oscillation with frequency k. If f is a complex-valued function defined on the circle group, we define the integral

$$\int_{\mathbf{T}} f(z) \, dz = \frac{1}{2\pi} \int_{0}^{2\pi} f(e^{it}) \, dt$$

The notation for the Fourier transform then becomes

$$\widehat{f}(n) = \int_{S^1} f(z) z^{-n}$$

the most austere and elegant way to write the transform. The Fourier series representation in terms of complex exponentials will be our choice throughout the rest of these notes. We note, however, that no deep knowledge of the complex numbers was used here, or is used in any of the basic theory of Fourier analysis. For most purposes, the exponential is just a simple way to represent sums of sines and cosines.

1.5   Basic Properties of Fourier Series

One of the most important properties of the Fourier series is that it asks nicely in terms of certain linear transformations of the functions it acts upon. This is summarized in this table of properties.

Theorem 1.1. *The following holds for the Fourier series of a 2π periodic function f.*

- *If f is real-valued, then $\widehat{f}(-n) = \overline{\widehat{f}(n)}$.*

- *If we define the translation operator T_s by $(T_s f)(t) = f(t - s)$, and the frequency operator M_s by $(M_t f)(s) = e^{tis} f(s)$, then $\widehat{T_s f} = M_{-s}\widehat{f}$, and $\widehat{M_s f} = T_s \widehat{f}$. Thus a translation in the spatial domain corresponds to a modulation in the frequency domain, and vice versa.*

- *If f is odd, then $\widehat{f}(-n) = -\widehat{f}(n)$. In particular, if f is real-valued, then the Fourier coefficients of f are purely imaginary, and*

$$\widehat{f}(n) = -\oint_0^\pi f(t) \, \sin(nt)$$

- *If f is even, then $\widehat{f}(-n) = \widehat{f}(n)$, and if f is real-valued, then the Fourier ceofficients of f are real, and*

$$\widehat{f}(n) = \oint_0^\pi f(t) \, \cos(nt)$$

11

- *If we define $\overline{f}(x) = \overline{f(x)}$, then*

$$\widehat{\overline{f}}(n) = \overline{\widehat{f}(-n)}$$

- *If f is continuously differentiable, then $\widehat{f'}(n) = in\widehat{f}(n)$.*

These relations are all easy exercises in manipulating integrals over $\mathbf{T}$.

1.6 Examples of Expansions

Before we get to the real work, let's start by computing some Fourier series, to use as examples. We also illustrate the convergence properties of the series, which we shall look at in more detail later. The brunt of the calculation is left as an exercise.

Example. *Consider 'plucking' a string, by pinching a point p on a string of length π and moving it up. An equation modelling this type of configuration is*

$$f(x) = \begin{cases} \frac{x}{p} & : 0 \leqslant x \leqslant p \\ \frac{\pi - x}{\pi - p} & : p \leqslant x \leqslant \pi \end{cases}$$

To calculate Fourier coefficients, we extend f to be a 2π periodic odd function. Then, the Fourier coefficients are purely imaginary, with

$$\widehat{f}(n) = \frac{-i}{\pi} \int_0^\pi f(x)\sin(nx) = \frac{-i}{\pi}\left[\frac{1}{p}\int_0^p x\sin(nx) + \frac{1}{\pi - p}\int_p^\pi (\pi - x)\sin(nx) \right]$$

Integration by parts tells us that

$$\int_0^p f(x)\sin(nx) = \frac{\sin(np)}{n^2} - \frac{\cos(np)p}{n}$$

$$\int_p^\pi (\pi - x)\sin(nx) = \frac{\cos(np)(\pi - p)}{n} + \frac{\sin(np)}{n^2}$$

Putting these together, we find

$$\widehat{f}(n) = -i\frac{\sin(np)}{n^2 p(\pi - p)}$$

and so, in some sense, we can obtain the identity

$$f(t) = \sum_{k=-\infty}^{\infty} -i\frac{\sin(np)}{n^2 p(\pi - p)} e^{kit} = \sum_{k=1}^{\infty} \frac{2\sin(np)}{p(\pi - p)} \sin(nt)$$

Referring back to our discussion of the wave equation, this means that if we let the string go, and let it perturb back and forth, the motion will be described by the infinite series

$$\sum_{n=1}^{\infty} \frac{2\sin(np)}{n^2 p(\pi - p)} \sin(nx)\cos(nt)$$

This converges absolutely and uniformly across time and space.

Example. *Consider the function f, defined on $[0, \pi]$ by $f(x) = x(\pi - x)$, made odd so that the function is defined on $[-\pi, \pi]$. The Fourier series is then purely imaginary, and in fact,*

$$\hat{f}(n) = \begin{cases} \frac{4i}{\pi n^3} & n \text{ odd} \\ 0 & n \text{ even} \end{cases}$$

and we may write

$$f(x) \sim \sum_{n \text{ odd}} \frac{4i}{\pi n^3} [e^{nix} - e^{-nix}] = \sum_{n \text{ odd}} \frac{8}{\pi n^3} \sin(nx)$$

This sum converges absolutely and uniformly on the entire real line.

Example. *The tent function*

$$f(x) = \begin{cases} 1 - \frac{|x|}{\delta} & : |x| < \delta \\ 0 & : |x| \geqslant \delta \end{cases}$$

is even, and therefore has a purely real Fourier expansion

$$\hat{f}(0) = \frac{\delta}{2\pi} \qquad \hat{f}(n) = \frac{1 - \cos(n\delta)}{\delta \pi n^2}$$

so we obtain an expansion

$$f(x) \sim \frac{\delta}{2\pi} + \sum_{n \neq 0} \frac{1 - \cos(n\delta)}{\delta \pi n^2} e^{inx} = \frac{\delta}{2\pi} + 2\sum_{n=1}^{\infty} \frac{1 - \cos(n\delta)}{\delta \pi n^2} \cos(nx)$$

This sum also converges absolutely and uniformly.

13

Example. *Consider the characteristic function*

$$\chi_{(a,b)}(x) = \begin{cases} 1 & : x \in (a,b) \\ 0 & : x \notin (a,b) \end{cases}$$

Then

$$\widehat{\chi}_{(a,b)}(n) = \frac{1}{2\pi} \int_a^b e^{-inx} = \frac{e^{-ina} - e^{-inb}}{2\pi in}$$

Hence we may write

$$\chi_{(a,b)}(x) = \frac{b-a}{2\pi} + \sum_{n \neq 0} \frac{e^{-ina} - e^{-inb}}{2\pi in} e^{inx}$$

$$= \frac{b-a}{2\pi} + \sum_{n=1}^{\infty} \frac{\sin(nb) - \sin(na)}{\pi n} \cos(nx) + \frac{\cos(na) - \cos(nb)}{\pi n} \sin(nx)$$

This sum does not converge absolutely for any value of x (except when a and b are chosen trivially). To see this, note that

$$\left| \frac{e^{-inb} - e^{-ina}}{2\pi n} \right| = \left| \frac{1 - e^{in(b-a)}}{2\pi n} \right| \geq \left| \frac{\sin(n(b-a))}{2\pi n} \right|$$

so that it suffices to show $\sum |\sin(nx)| n^{-1} = \infty$ for every $x \notin \pi \mathbf{Z}$. This follows because enough of the values of $|\sin(nx)|$ are large, so that we may apply the divergence of $\sum n^{-1}$ become applicable. First, assume $x \in (0, \pi/2)$. If

$$m\pi - x/2 < nx < m\pi + x/2$$

for some $m \in \mathbf{Z}$, then

$$m\pi + x/2 < (n+1)x < m\pi + 3x/2 < (m+1)\pi - x/2$$

so that if $nx \in (-x/2, x/2) + \pi \mathbf{Z}$, $(n+1)x \notin (-x/2, x/2) + \pi \mathbf{Z}$. For y outside of $(-x/2, x/2) + \pi \mathbf{Z}$, we have $|\sin(y)| > |\sin(x/2)|$, and therefore for any n,

$$\frac{|sin(nx)|}{n} + \frac{|\sin((n+1)x)|}{n+1} > \frac{|\sin(x/2)|}{n+1}$$

14

and thus

$$\sum_{n=1}^{\infty} \frac{|\sin(nx)|}{n} = \sum_{n=1}^{\infty} \frac{|\sin(2nx)|}{2n} + \frac{|\sin((2n+1)x)|}{2n+1}$$

$$> |\sin(x/2)| \sum_{n=1}^{\infty} \frac{1}{2n+1} = \infty$$

In general, we may replace x with $x - k\pi$, with no effect to the values of the sum, so we may assume $0 < x < \pi$. If $\pi/2 < x < \pi$, then

$$\sin(nx) = \sin(n(\pi - x))$$

and $0 < \pi - x < \pi/2$, completing the proof, except when $x = \pi$, in which case

$$\sum_{n=1}^{\infty} \left| \frac{1 - e^{in\pi}}{2\pi n} \right| = \sum_{n \text{ even}} \left| \frac{1}{\pi n} \right| = \infty$$

Thus the convergence of a Fourier series need not be absolute.

Example. *We can often find formulas for certain fourier summations from taking the corresponding power series. For instance, the power series expansion*

$$\log\left(\frac{1}{1-x}\right) = \sum_{k=1}^{\infty} \frac{z^k}{k}$$

which converges pointwise for every $z \in \mathbf{D}$ but $z = 1$, implies that for $x \notin 2\pi\mathbf{Z}$,

$$\sum_{k=1}^{\infty} \frac{\cos(kx)}{k} = \mathfrak{Re}\left(\log\left(\frac{1}{1 - e^{ix}}\right)\right) = -\frac{1}{2}\log(2 - 2\cos(x))$$

$$\sum_{k=1}^{\infty} \frac{\sin(kx)}{k} = \mathfrak{Im}\left(\log\left(\frac{1}{1 - e^{ix}}\right)\right) = \arctan\left(\frac{\sin(x)}{1 - \cos(x)}\right)$$

where we agree that $\arctan(\pm\infty) = \pm\pi/2$. If a power series' radius of convergence exceeds 1, then it is likely that the corresponding Fourier series taken on the circle will be pleasant, whereas if the power series' radius is equal to 1, we can expect nasty behaviour on the boundary. In Complex analysis, one avoids talking about the boundary of the holomorphic function's definition, whereas in Fourier analysis we have to embrace the boundary points, which makes the theory a little more pathological.

15

Example. *If f is a **trigonometric polynomial**, meaning there are coefficients a_n such that*

$$f(t) = \sum_{n=-N}^{N} a_n e^{nit}$$

*then it is easy to see that $\widehat{f}(n) = a_n$. In particular, we will be interested in the analysis of the **Dirichlet kernel***

$$D_N(t) = \sum_{n=-N}^{N} e^{nit}$$

which assigns a unit mass to each integer frequency between $-N$ and N. By the geometric series summation formula, we may write the Dirichlet kernel as

$$1 + \sum_{n=1}^{N} e^{nit} + e^{-nit} = 1 + e^{it}\frac{e^{Nit}-1}{e^{it}-1} + e^{-it}\frac{e^{-Nit}-1}{e^{-it}-1}$$

$$= 1 + e^{it}\frac{e^{Nit}-1}{e^{it}-1} + \frac{e^{-Nit}-1}{1-e^{it}} = \frac{e^{(N+1)it}-e^{-Nit}}{e^{it}-1}$$

$$= \frac{e^{(N+1/2)it}-e^{-(N+1/2)it}}{e^{it/2}-e^{-it/2}} = \frac{\sin((N+1/2)t)}{\sin(t/2)}$$

D_N has average value $1/2\pi$ on $[-\pi, \pi]$, but the average value of $|D_N|$ becomes very large as N tends to ∞. We thus see that D_N has large oscillation, which causes D_N to be small but $|D_N|$ to be very large.

Example. *The **Poisson kernel** P_r is defined on $[-\pi, \pi]$, and for $0 \leqslant r < 1$, by the power series*

$$P_r(t) = \sum_{n=-\infty}^{\infty} r^{|n|} e^{int}$$

P_r arises in the discussion of heat distributions on the unit disk, because

$$P_r(t) = 1 + 2\sum_{n=1}^{\infty} r^n \cos(nt)$$

so P_r 'fits the mold' we gave of solutions to the heat equation, and in some sense, it is the fundamental solution, because all solutions can be rewritten in the form

$$\sum a_n r^{|n|} e^{int}$$

16

for some coefficients a_n, which is P_r with certain frequencies amplified. For $r < 1$, the sum defining P_r converges uniformly on the disk, and as such we obtain that $\widehat{P_r}(n) = r^{|n|}$, so we see that the high frequency parts of P_r decrease geometrically. We can also apply an (infinite) geometric series summation to obtain that

$$\sum r^{|n|} e^{int} = 1 + \frac{re^{it}}{1 - re^{it}} + \frac{re^{-it}}{1 - re^{-it}} = 1 + \frac{2r\cos t - 2r^2}{(1 - re^{it})(1 - re^{-it})}$$

$$= 1 + \frac{2r\cos t - 2r^2}{1 - 2r\cos t + r^2} = \frac{1 - r^2}{1 - 2r\cos t + r^2}$$

As $r \to 1$, the function concentrates at the origin, and there does not appear to be too much oscillation in the function, reflecting the fact that Poisson kernel is much better behaved and the Dirichlet kernel.

Chapter 2

Fourier Series Convergence

Let's focus in on the problem we introduced in the last chapter. For each integrable function $f : \mathbf{T} \to \mathbf{C}$, we have an associated **formal trigonometric series**

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

In some sense, f should be able to be approximated by the finite degree trigonometric polynomial

$$\sum_{n=-N}^{N} \widehat{f}(n)e^{inx}$$

At this point, we haven't deduced any reason for these sums to converge to f analytically. To understand the convergence, we define the m'th partial sum

$$S_m(f)(x) = \sum_{n=-m}^{m} \hat{f}(n)e^{inx}$$

The reason we sum from $-m$ to m is that if f is real valued, then $\hat{f}(n) = \overline{\hat{f}(n)}$, so the complex parts of the equation will cancel out and we find $S_m(f)$ will just be a sum of cosines and sines. The first relation we can expect is pointwise convergence; is it true that for every x,

$$\lim_{m\to\infty} S_m(f)(x) = f(x)$$

Perhaps if we're lucky, we'll get uniform convergence as well. Unfortunately, we will show that there are even examples of continuous periodic

functions whose partial sums diverge somewhere, so we must search for more exotic methods of convergence.

2.1 Do Functions Have Unique Series?

If the Fourier series of every function converged pointwise, we could conclude that if f and g have the same fourier coefficients, they must necessarily be equal. This is clearly not true, for if we alter a function at a point, the Fourier series, defined by averaging over the entire region, remains the same. Nonetheless, if a function is continuous editing the function at a point will break continuity, so we may have some hope of uniqueness of the expansion.

Theorem 2.1. *If $\widehat{f}(n) = 0$ for all n, then f vanishes wherever it is continuous.*

Proof. We shall prove this for real-valued functions. For every trigonometric polynomial $P(x) = \sum a_n e^{-nix}$, we have

$$\int_{-\pi}^{\pi} f(x)P(x)dx = 2\pi \sum a_n \widehat{h}(n) = 0$$

Suppose that f is continuous at zero, and assume without loss of generality that $f(0) > 0$. Pick δ such that if $|x| < \delta$, $|f(x)| > f(0)/2$. Consider the trigonometric polynomial

$$P(x) = \varepsilon + \cos x = \varepsilon + \frac{e^{ix} + e^{-ix}}{2}$$

where ε is small enough that $P(x) > A > 1$ for $|x| < \delta/2$, $P(x) > 0$ for $\delta/2 \leqslant |x| < \delta$, and $P(x) < B < 1$ for $|x| \geqslant \delta$. Consider the series of trigonometric polynomials

$$P_n(x) = (\varepsilon + \cos x)^n$$

For which we have

$$\left| \int_{-\pi}^{\pi} P_n(x)f(x)dx \right| \geqslant \int_{|x|<\delta} P_n(x)f(x)dx - \left| \int_{\delta \leqslant |x|} P_n(x)f(x)dx \right|$$

Now

$$\left| \int_{\delta \leqslant |x|} P_n(x)f(x)dx \right| \leqslant \int_{\delta \leqslant |x|} B^n |f(x)| \, dx \leqslant B^n \|f\|_1$$

19

whereas

$$\int_{|x|<\delta} P_n(x)f(x)dx = \int_{|x|<\delta/2} P_n(x)f(x) + \int_{\delta/2\leqslant|x|<\delta} P_n(x)f(x)$$

$$\geqslant \int_{|x|<\delta/2} P_n(x)f(x) \geqslant \delta A^n f(0)$$

and so we conclude

$$0 = \left|\int_{-\pi}^{\pi} P_n(x)f(x)dx\right| \geqslant \delta A^n f(0) - B^n\|f\|_1$$

Regardless of the values of $\|f\|_1, f(0)$, and δ, eventually, we find that the right hand side of the equation is positive, which is impossible. In general, if f is complex valued, then we may write $f = u + iv$, where

$$u(x) = \frac{f(x) + \overline{f(x)}}{2} \quad v(x) = \frac{f(x) - \overline{f(x)}}{2i}$$

The Fourier coefficients of $\overline{f}$ all vanish, because the coefficients of f vanish, and so we conclude the coefficients of u and v vanish. f is continuous at x if and only if it is continuous at u and v, and we know from the previous case this means that both u and v vanish at that point. $\square$

Corollary 2.2. *If $\widehat{f} = \widehat{g}$, where f and g are continuous, then $f = g$.*

Proof. Because $f - g$ is continuous with vanishing Fourier coefficients. $\square$

Later, we will see that this theorem can be generalized to not-necessarily continuous functions. Then the two Fourier series agree if and only if the two functions agree except on a 'small' set, that is, a set of measure zero.

Corollary 2.3. *If a continuous function f has absolutely convergent Fourier coefficients, then it's Fourier series converges uniformly to f.*

Proof. If $\sum|\widehat{f}(n)| < \infty$, then the functions $S_m(f)$ converge uniformly to a function g, which necessarily must be continuous. We may apply uniform convergence to conclude

$$\widehat{g}(n) = \lim_{m\to\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} S_m(f)(t)e^{-int} = \widehat{f}(n)$$

Hence $\widehat{f} = \widehat{g}$, so $f = g$. $\square$

2.2   Convergence In $C^2(\mathbf{T})$

The easiest place to verify convergence is in the space of continuously differentiable functions. It turns out that differentiability of a function corresponds to a asymptotic decay in the coefficients of the Fourier series. This decay can be used to show absolute convergence if the function is suitably differentiable.

Theorem 2.4. *If $f \in C^2(\mathbf{T})$, then $\widehat{f}(n) = O(n^{-2})$, and so f's Fourier series converges uniformly to f.*

Proof. We know that $\widehat{f''}(n) = in\widehat{f'}(n) = -n^2\widehat{f}(n)$. Since we have the elementary estimate

$$|\widehat{f''}(n)| = \left|\frac{1}{2\pi}\int_{-\pi}^{\pi}f''(t)e^{-int}\right| \leqslant \frac{\|f''\|_1}{2\pi}$$

This means $\widehat{f}(n) = O(n^{-2})$. Every series that is $O(n^{-2})$ converges absolutely by the comparison test, and so the corresponding Fourier series converges uniformly. But this means that it converges to a continuous function, and so it must converge to f. $\qquad\square$

More generally, if $f \in C^k(\mathbf{T})$, then $\widehat{f}(n) = O(n^{-k})$. This is a general idea relating to the fact that the smoothness of f relates to the decay of it's Fourier series. More generally, this is an instance of the general phenomenon that local properties of f represent themselves as coarse properties of it's Fourier series, and vice versa. Later on, we will show that we obtain uniform convergence even in $C^1(\mathbf{T})$, even if f is just Hölder continuous for $\alpha > 1/2$, in the sense that there is a constant C such that for any t, s,
$$|f(t) - f(s)| \leqslant C|t - s|^\alpha$$
which is slightly stronger than mere continuity.

2.3   Convolution and Kernel Methods

The notion of the convolution of two functions f and g is a key tool in Fourier analysis, both as a way to regularize functions, and as an operator

21

that transforms nicely when we take Fourier series. Given two functions f and g, we define

$$(f * g)(t) = \fint_{-\pi}^{\pi} f(s)g(t-s)$$

which we can think of smoothing g by taking it's averages in a neighbourhood of a point, according to a distribution with density f, and this is indeed the case if f is positive and integrates to 2π. The first of the following properties for convolution is an easy exercise in integral transformations.

Theorem 2.5. *Convolution has the following properties:*

- *Convolution is a commutative, associative, bilinear operation.*

- *$f * g$ is continuous*

- *$\widehat{f * g} = \hat{f}\hat{g}$.*

- *If f is in $C^1(\mathbf{T})$, then $f * g$ is C^1, and $(f * g)' = f' * g$. In particular, if f is $C^n(\mathbf{T})$, and g is in $C^m(\mathbf{T})$, then $f * g$ is in $C^{n+m}(\mathbf{T})$, so convolution is 'additively smoothing'.*

Proof. Assume that f and g are continuous functions on $\mathbf{T}$. Then

$$[(f * g)(t) - (f * g)(s)] = \frac{1}{2\pi} \int f(u)[g(t-u) - g(s-u)]$$

Since $\mathbf{T}$ is compact, g is uniformly continuous, and so we may find δ such that if $|t-s| < \delta$, then $|g(t-u) - g(s-u)| < \varepsilon$. But then the integral above is bounded by $(2\pi)^{-1}\varepsilon\|f\|_1$, and this gives the continuity of convolution. If f and g are general Riemann integrable functions, then there exists continuous functions f_k, and g_k with $\|f_k\|_\infty \leqslant \|f\|_\infty$, $\|g_k\|_\infty \leqslant \|g\|_\infty$, which converge to f and g in L^1. But then

$$(f * g) - (f_k * g_k) = (f - f_k) * g + f_k * (g - g_k)$$

Now

$$|((f - f_k) * g)(t)| = \frac{1}{2\pi}\left|\int (f - f_k)(s)g(t-s)\right| \leqslant \frac{1}{2\pi}\|g\|_\infty\|f - f_k\|_1$$

$$|f_k * (g-g_k)| = \frac{1}{2\pi}\left|\int f_k(t - s)(g - g_k)(s)\right| \leqslant \frac{1}{2\pi}\|f_k\|_\infty\|g-g_k\|_1 \leqslant \frac{1}{2\pi}\|f\|_\infty\|g-g_k\|_1$$

22

so we see $f_k * g_k$ converges uniformly to $f * g$. But since each $f_k * g_k$ is continuous, this means $f * g$ must also be continuous.

To obtain the product identity for the Fourier series, we can apply Fubini's theorem to write

$$
\begin{aligned}
\widehat{f * g}(n) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (f * g)(t) e^{-nit} \, dt \\
&= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(s) g(t - s) e^{-nit} \, ds \, dt \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \int_{-\pi}^{\pi} (L_{-s} g)(t) e^{-nit} \, dt \, ds \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) e^{-ins} \widehat{g}(n) \, ds \\
&= \widehat{f}(n) \widehat{g}(n)
\end{aligned}
$$

and this is exactly the identity required.

The fact that $f * g$ is differentiable, and that the derivative is $f' * g$ when f is differentiable, follows because, using the mean value theorem, that the partial sums

$$
\frac{f(x + h) - f(x)}{h}
$$

converge uniformly in x to $f'(x)$, hence

$$
(f * g)'(x) = \lim_{h \to 0} \int \frac{f(x + h - y) - f(x - y)}{h} g(x) = \int f'(x - y) g(x) = (f' * g)
$$

and this is the required property. $\qquad\square$

We know that suitably smooth functions have convergent Fourier series. If we want to establish the converence of the series corresponding to a function f, we might want to 'smooth' f by convolving it with a function g. Provided that $\widehat{g}$ is 'close' to 1, the above theorem says that $\widehat{f * g}$ will be close to $\widehat{g}$. If we can establish the convergence properties on the convolution $f * g$, then we can probably obtain results about f. The family of functions g we will consider for which $\widehat{g}$ approximates 1 are called **good kernels**. In particular, a good kernel is a sequence of functions K_n on $\mathbf{T}$, for which

$$
\fint_{-\pi}^{\pi} K_n(t) = 1
$$

23

the functions K_n are bounded in L^1, and for any $\delta > 0$,

$$\int_{\delta < |x|} |K_n(x)| \to 0$$

so the functions K_n become concentrated at the origin. The L^1 bounded-ness follows from the first condition if we assume the K_n are positive. As $n \to \infty$, we find that for any m, the fact that K_n vanishes outside of small neighbourhoods implies that

$$\widehat{K_n}(m) = \frac{1}{2\pi} \int K_n(x) e^{-mit} \to 1$$

so if f is any function, then the Fourier series of $f * K_n$ converges to the Fourier series of f pointwise. Even better, $f * K_n$ approximates f, because the 'averages' of f with respect to K_n become concentrated around f.

Theorem 2.6. *If f is any function, then $(f * K_n)(x) \to f(x)$ whenever f is continuous at x.*

Proof. We have

$$[(f * K_n)(x) - f(x)] = \frac{1}{2\pi} \int_{-\pi}^{\pi} [f(x-y) - f(x)] K_n(y) \, dy$$

Now fix δ such that $|f(x-y) - f(x)| < \varepsilon$ for $|y| < \varepsilon$. Then

$$
\begin{aligned}
|(f * K_n)(x) - f(x)| &= \frac{1}{2\pi} \int_{|y| < \varepsilon} |f(x-y) - f(x)||K_n(y)| \, dy \\
&+ \frac{1}{2\pi} \int_{|y| \geqslant \varepsilon} |f(x-y) - f(x)||K_n(y)| \, dy \\
&\leqslant \frac{\varepsilon}{2\pi} \|K_n\|_1 + \frac{2\|f\|_\infty}{2\pi} \int_{|y| \geqslant \varepsilon} |K_n(y)| \, dy \\
&\leqslant \frac{\varepsilon}{2\pi} \|K_n\|_1 + o(1)
\end{aligned}
$$

Taking δ smaller and smaller, we can make the two values as close as we desire, so we obtain pointwise convergence. Note that if f is continuous everywhere, then it is uniformly continuous, hence this bound is uniform, and $f * K_n$ therefore converges to f uniformly. $\square$

24

Recall the definition of the Dirichlet kernel

$$D_N(t) = \sum_{n=-N}^{N} e^{nit}$$

Because of the properties of convolution, we can represent the partial Fourier approximations of any function f by convolution with the Dirichlet kernel, in the sense that $S_n(f) = f * D_n$. If D_n was a good kernel, then we would obtain that the partial sums of S_n converge uniformly. This initially seems a good strategy, because $\oint D_N(t) = 1$. However, we find

$$\int_{-\pi}^{\pi} |D_n(t)| = \int_{-\pi}^{\pi} \left| \frac{\sin((N+1/2)t)}{\sin(t/2)} \right| = 2 \int_{0}^{\pi} \frac{|\sin((N+1/2)t)|}{\sin(t/2)}$$

$$\geqslant 4 \int_{0}^{\pi} \frac{|\sin((N+1/2)t)|}{t} = 4 \int_{0}^{(2N+1)\pi} \frac{|\sin(u/2)|}{u}$$

$$\geqslant \sum_{n=0}^{N-1} \frac{2}{n\pi} \int_{0}^{2\pi} \sin(u/2) = \frac{8}{\pi} \sum_{n=0}^{N-1} \frac{1}{n} = \Omega(\log N)$$

So the L^1 norm of D_n grows, albeit slowly, to ∞. This reflects the fact that D_n oscillates very frequently, and also that the pointwise convergence of the Fourier series is much more subtle than that provided by good kernels. However, in the next section we show that, if we interpret our convergence of series in a different manner, we get a family of good kernels, and therefore we obtain pointwise convergence for suitable reinterpretations of partial sums.

2.4   Countercultural Methods of Summation

The standard method of summation suffices for much of analysis. Given a sequence $a_0, a_1, \ldots$, we define the infinite sum as the limit of partial sums.

$$\sum_{k=0}^{\infty} a_k = \lim_{n \to \infty} \sum_{k=0}^{n} a_k$$

Similarily,

$$\sum_{k=-\infty}^{\infty} a_k = \sum_{k=0}^{\infty} a_k + a_{-k} = \lim_{n \to \infty} \sum_{k=-n}^{n} a_k$$

25

Some sums, like $\sum_{k=1}^{\infty} k$, obviously diverge, whereas other sums, like $\sum 1/n$, 'just' fail to converge because they grow suitably slowly towards infinity over time. Since the time of Euler, a new method of summation developed by Cesaro was introduced which 'regularized' certain terms by considering averaging the sums over time. Rather than considering limits of partial sums, we consider limits of averages of sums, known as Cesaro means. Letting $s_n = \sum_{k=0}^{n} a_k$, we define the Cesaro means

$$\sigma_n = \frac{s_0 + \cdots + s_n}{n + 1}$$

and then consider a sequence as Cesaro summable if the σ_n converge. If the normal summation exists, then the Cesaro limit exists, and is equal to the original sum. However, the Cesaro summation is stronger than normal convergence.

Example. *In the sense of Cesaro, we have*

$$1 - 1 + 1 - + \cdots = 1/2$$

which reflects the fact that the partials sums do 'converge', but to two different numbers 0 and 1, which the series oscillates between, and the Cesaro means average these two points of convergence out to give a single method of convergence.

Another notion of regularization sums emerged from Complex analysis, called Abel summation. Given a sequence $\{a_i\}$, we can consider the power series $\sum a_k r^k$. If this is well defined for $|r| < 1$, we can consider the Abel means $A_r = \sum a_k r^k$, and ask if $\lim_{r \to 1} A_r$ exists, which should be 'almost like' $\sum a_k$. If this limit exists, we call it the Abel sum of the sequence.

Example. *In the Abel sense, we have*

$$1 - 2 + 3 - 4 + 5 - \cdots = \frac{1}{4}$$

because

$$\sum_{k=0}^{\infty} (-1)^k (k+1) z^k = \frac{1}{(1+z)^2}$$

the coefficients here are $\Omega(N)$, so they can't be Cesaro summable.

26

Abel summation is even more general than Cesaro summation.

Theorem 2.7. *If a sequence is Cesaro summable, it is Abel summable, and to the same value.*

Proof. Let $\{a_i\}$ be a Cesaro summable sequence, which we may without loss of generality assume converges to 0. Now $(n+1)\sigma_n - n\sigma_{n-1} = s_n$, so

$$(1-r)^2 \sum_{k=0}^{n}(k+1)\sigma_k r^k = (1-r)\sum_{k=0}^{n} s_k r^k = \sum_{k=0}^{n} a_k r^k$$

As $n \to \infty$, the left side tends to a well defined value for $r < 1$, hence the same is true for $\sum_{k=0}^{n} a_k r^k$. Given $\varepsilon > 0$, let N be large enough that $|\sigma_n| < \varepsilon$ for $n > N$, and let M be a bound for all $|\sigma_n|$. Then

$$\left| (1-r)^2 \sum_{k=0}^{\infty}(k+1)\sigma_k r^k \right| \leqslant (1-r)^2 \left(\sum_{k=0}^{N}(k+1)|\sigma_k|r^k + \varepsilon \sum_{k=N+1}^{\infty}(k+1)r^k \right)$$

$$= (1-r)^2 \left(\sum_{k=0}^{N}(k+1)(|\sigma_k| - \varepsilon)r^k + \varepsilon \left[\frac{r^{n+1}}{1-r} + \frac{1}{(1-r)^2} \right] \right)$$

$$\leqslant (1-r)^2 M \sum_{k=0}^{N}(k+1)r^k + \varepsilon r^{n+1}(1-r) + \varepsilon$$

$$\leqslant (1-r)^2 M \frac{(N+1)(N+2)}{2} + \varepsilon r^{n+1}(1-r) + \varepsilon$$

Fixing N, and letting $r \to 1$, we may make the complicated sum on the end as small as possible, so the absolute value of the infinite sum is less than ε. Thus the Abel limit converges to zero. $\square$

Note that the Cesaro means of the Fourier series of f are given by

$$\sigma_N(f) = \frac{S_0(f) + \cdots + S_{N-1}(f)}{N} = f * \left(\frac{D_0 + \cdots + D_{N-1}}{N} \right) = f * F_N$$

The convergence properties of the Cesaro means therefore relate to the properties of the **Fejér kernel** F_N. We find that

$$F_N(x) = \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N} \right) e^{nit} = \frac{1}{N} \frac{\sin^2(Nx/2)}{\sin^2(x/2)}$$

so the oscillations of the Dirichlet kernel are slightly dampened, and as a result, F_N is a good kernel.

Theorem 2.8 (Fejér's Theorem). *The Fourier series of f is Cesaro summable to f at every point of continuity, and is uniformly Cesaro summable if f is continuous everywhere.*

Relating Abel summations to Fourier series requires a little bit more careful work, since we do not consider limits of finite sums. Note that the Abel sum is

$$A_r(f) = \sum_{n=-\infty}^{\infty} \widehat{f}(n) r^n e^{nit}$$

Using the fact that the partial sums of the *Poisson kernel*

$$P_r(t) = \sum r^{|n|} e^{nit}$$

converge uniformly, we can calculate that $A_r(f) = P_r * f$. Thankfully, we find P_r is a good kernel, and the Abel means work nicely with Fourier series.

Theorem 2.9. *The Fourier series of f is Abel summable to f at every point of continuity, and is uniformly Abel summable if f is continuous everywhere.*

2.5 Tchebychev Polynomials

If f is everywhere continuous, then for every ε Fejér's theorem says that we can find N such that $\|\sigma_N(f) - f\| \leqslant \varepsilon$. But

$$\sigma_N(f) = \frac{S_0(f) + \cdots + S_{N-1}(f)}{N}$$

is just a trigonometric polynomial, and so we have shown that with respect to the L^∞ norm, the space of trigonometric polynomials is dense in the space of all continuous functions. Now if f is a continuous function on $[0, \pi]$, then we can extend it to be even and 2π periodic, and then the trigonometric series $S_N(f)$ of f will be a cosine series, hence $\sigma_N(f)$ will also be a cosine series, and so for each ε, we can find N and coefficients $a_1, \ldots, a_N$ such that

$$|f(x) - \sum_{n=1}^{N} a_n \cos(nx)| < \varepsilon$$

Now we use a surprising fact. For each n, there exists a degree n polynomial T_n such that $\cos(nx) = T_n(\cos x)$. This is clear for $n = 0$ and $n = 1$. More generally, we can write

$$
\begin{aligned}
\cos((m+1)x) &= \cos((m+1)x) + \cos((m-1)x) - \cos((m-1)x) \\
&= \cos(mx + x) + \cos(mx - x) - \cos((m-1)x) \\
&= 2\cos x \cos(mx) - \cos((m-1)x)
\end{aligned}
$$

so the proof is complete by induction. It gives the relation

$$
T_{m+1}(x) = 2x T_m(x) - T_{m-1}(x)
$$

These polynomials are known as the Chebyshev polynomials.

Corollary 2.10 (Weirstrass)**.** *The polynomials are uniformly dense in $C[0,1]$.*

Proof. If f is a continuous function on $[0,1]$, we can define $g(t) = f(|\cos(t)|)$. Then g is even, and so for every $\varepsilon > 0$, we can find $a_1, \ldots, a_N$ such that

$$
\left| g(t) - \sum_{n=1}^{N} a_n \cos(nt) \right| = \left| g(t) - \sum_{n=1}^{N} a_n T_n(\cos t) \right| < \varepsilon
$$

But if $u = \cos(t)$, for $\cos(t) \geqslant 0$, this equation says

$$
\left| f(u) - \sum_{n=1}^{N} a_n T_n(u) \right| < \varepsilon
$$

and so we have uniformly approximated f by a polynomial. $\quad\square$

2.6 Returning to Harmonic Functions on the Disk

Using the Abel summability properties of functions, we can solve Laplace's equation on a disk uniquely given any boundary conditions.

Theorem 2.11. *If f is integrable on $\mathbf{T}$, then the function*

$$
u(re^{it}) = (f * P_r)(t)
$$

is a $C^2(\mathbf{D}^\circ)$ harmonic function such that at any point of continuity t of f,

$$
\lim_{r \uparrow 1} u(re^{it}) = f(t)
$$

If f is continuous everywhere, then this limit is uniform, and u is the unique $C^2(\mathbf{D}^\circ)$ function with these properties.

Proof. Note that because P_r is C^∞ for all $r < 1$ (the power series and its derivatives converge uniformly), $f * P_r$ is C^∞, and so

$$u(re^{it}) = \sum \widehat{f}(n) r^{|n|} e^{nit}$$

which converges uniformly in every bounded disk around the origin of radius less than one. This justifies the calculation

$$\frac{\partial^2 u}{\partial t^2} = -\sum \widehat{f}(n) n^2 r^{|n|} e^{nit}$$

$$\frac{\partial(f * P_r)}{\partial r} = \sum \widehat{f}(n) |n| r^{|n|-1} e^{nit} \qquad \frac{\partial(f * P_r)}{\partial r^2} = \sum \widehat{f}(n) |n| (|n| - 1) r^{|n|-2} e^{nit}$$

and now we find that on the interior of the unit disk,

$$\Delta u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u}{\partial t^2} = \sum [|n|(|n| - 1) + |n| - n^2] \widehat{f}(n) r^{|n|-2} e^{nit} = 0$$

Abel summability results imply that $u(re^{it}) \to f(t)$ at every point of continuity of f, and the convergence is uniform if f is continuous.

Now instead, suppose that u solves $\Delta u = 0$ in the interior of the unit disk, and converges uniformly to f at the boundary. Because u is in $C^2(\mathbf{D}^\circ)$, we have a uniform sum

$$u_r(t) = \sum a_n(r) e^{nit}$$

where

$$a_n(r) = \fint_{-\pi}^{\pi} u(re^{it}) e^{-int}$$

But then the $a_n(r)$ are differentiable, and

$$a_n'(r) = \fint_{-\pi}^{\pi} \frac{\partial u}{\partial r}(re^{it}) e^{-int} \qquad a_n''(r) = \fint_{-\pi}^{\pi} \frac{\partial^2 u}{\partial r^2}(re^{it}) e^{-int}$$

$$\fint_{-\pi}^{\pi} \frac{\partial^2 u}{\partial t^2}(re^{it}) e^{int} = -n^2 a_n(r)$$

and so $a_n'' + a_n'/r - n^2 a_n/r^2 = 0$, which is an ordinary differential equation whose only bounded solution is $a_n(r) = A_n r^n$. But now, if we let $r \to \infty$, we conclude that

$$A_n = \lim_{r \to 1} \fint_{-\pi}^{\pi} u(re^{it}) e^{-int} = \fint_{-\pi}^{\pi} u(e^{it}) e^{-int} = \widehat{f}(n)$$

so u is just $f * P_r$. $\qquad\qquad \square$

A final remark is that, if u is only required to converge to f *pointwise* on the boundary, then the function we found is no longer required to be unique. Below is an example of a function u which tends to zero pointwise on the boundary, yet does not vanish on the interior of the unit disk.

Example. *If P_r is the Poisson kernel, define a function*

$$u(r,t) = \frac{\partial P_r}{\partial t}$$

Then u is harmonic in the unit disk, because by commutativity of derivatives,

$$\Delta u = \frac{\partial}{\partial t}(\Delta P_r) = \frac{\partial}{\partial t}(0) = 0$$

We calculate

$$
\begin{aligned}
u(r,t) &= \sum_{n=1}^{\infty} inr^n[e^{int} - e^{-int}] \\
&= i\left[\frac{re^{it}}{(re^{it}-1)^2} - \frac{re^{-it}}{(re^{-it}-1)^2}\right] \\
&= i\left[\frac{re^{-it} + r^{-1}e^{it} - re^{it} - r^{-1}e^{-it}}{(re^{it}-1)^2(re^{-it}-1)^2}\right] \\
&= \frac{(r-r^{-1})\sin(t)}{(re^{it}-1)^2(re^{-it}-1)^2} \\
&= \frac{(r^2-1)\sin(t)}{r(re^{it}-1)^2(re^{-it}-1)^2}
\end{aligned}
$$

In this form, it is easy to see that for a fixed t, as $r \to 1$, $u(r,t) \to 0$. However, the denominator tells us this convergence isn't uniform.

2.7 L^2 **Convergence of Fourier Series**

Using the theory of inner product spaces, we can show that for all integrable functions the Fourier series of a function

2.8   A Divergent Fourier Series

Analysis was built to analyze continuous functions, so we would hope the method of fourier expansion would work for all continuous functions. Unfortunately, this is not so. The behaviour of the Dirichlet kernel away from the origin already tells us that the convergence of Fourier series is subtle. We shall take advantage of this to construct a continuous function with divergent fourier series at a point.

To start with, we shall consider the series

$$f(t) \sim \sum_{n \neq 0} \frac{e^{int}}{n}$$

where f is an odd function equaling $i(\pi - t)$ for $t \in (0, \pi]$. Such a function is nice to use, because its Fourier representation is simple, yet very close to diverging. Indeed, if we break the series into the pair

$$\sum_{n=1}^{\infty} \frac{e^{int}}{n} \qquad \sum_{n=-\infty}^{-1} \frac{e^{int}}{n}$$

Then these series no longer are the Fourier representations of a Riemann integrable function. For instance, if $g(t) \sim \sum_{n=1}^{\infty} \frac{e^{int}}{n}$, then the Abel means
$A_r(f)(t) =$

2.9   Conjugate Fourier Series

When f is a real-valued integrable function, then $\overline{\widehat{f}(-n)} = \widehat{f}(n)$. Thus we formally calculate that

$$\sum_{n=-\infty}^{\infty} \widehat{f}(n)e^{nit} = \mathrm{Re}\left(\widehat{f}(0) + 2 \sum_{n=1}^{\infty} \widehat{f}(n)e^{nit} \right)$$

This series defines an analytic function in the interior of the unit circle since the coefficients are bounded. Thus the sum is a harmonic function in the interior of the unit circle. The imaginary part of this sum is

$$\mathrm{Im}\left(\widehat{f}(0) + 2 \sum_{n=1}^{\infty} \widehat{f}(n)e^{nit} \right) = \mathfrak{Re}\left(-i \sum_{n=-\infty}^{\infty} \mathrm{sgn}(n)\widehat{f}(n)e^{nit} \right)$$

32

The right hand side is known as the conjugate series to the Fourier series $\widehat{f}(n)$. It is closely related to the study of a function $\tilde{f}$ known as the *conjugate function*.

Chapter 3

The Fourier Transform

In the last few chapters, we discussed the role of analyzing the frequency decomposition of a periodic function on the real line. In this chapter, we explore the ways in which we may extend this construction to perform frequency analysis of functions defined on the entire real line, and more generally, in higher dimensional Euclidean space. The only periodic trigonometric functions on $[0,1]$ on the real line had integer frequencies of the form $2\pi n$, whereas on the real line periodic functions can have frequencies corresponding to any real number. The analogue of the discrete Fourier series formula

$$f(x) = \sum_{k=-\infty}^{\infty} \widehat{f}(k)e(kx)$$

is the Fourier inversion formula

$$f(x) = \int_{-\infty}^{\infty} \widehat{f}(\xi)e(\xi x)\, d\xi$$

where for each real number ξ, we can define the quantity

$$\widehat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e(-\xi x)\, dx$$

The function $\widehat{f}$ is known as the **Fourier transform** of the function f. It is also denote by $\mathcal{F}(f)$, depending on whatever is convenient at the time. The role to which we can justify this formula is the main focus of this

chapter. Without too much more work, we will also analyze the Fourier transform on $\mathbf{R}^n$, which, given $f : \mathbf{R}^n \to \mathbf{R}^n$, considers the quantities

$$f(x) \sim \frac{1}{2\pi} \int_{\mathbf{R}^n} \widehat{f}(\xi) e(\xi \cdot x) \, d\xi \quad \widehat{f}(\xi) = \int_{\mathbf{R}^n} f(x) e(-\xi \cdot x) \, dx$$

where $\xi \in \mathbf{R}^n$, and $\xi \cdot x$ denotes the scalar product of ξ and x.

Perhaps the weakest criterion we can place on f in order to calculate the Fourier transform is that $f \in L^1(\mathbf{R})$. Indeed, in order to interpret the integral

$$\int_{-\infty}^{\infty} f(x) e(-\xi x) \, dx$$

in the Lebesgue sense, so that we can apply all the useful pointwise convergence theorems of that theory, we must assume the integrand is absolutely integrable, which means exactly that

$$\int_{-\infty}^{\infty} |f(x) e(-\xi x)| = \int_{-\infty}^{\infty} |f(x)| < \infty$$

thus we begin by considering the Fourier transform as an operator taking a function f in $L^1(\mathbf{R}^n)$ in the spatial domain and giving us a function $\widehat{f}$ in the frequency domain.

Theorem 3.1. *The mapping $f \mapsto \widehat{f}$ is a bounded linear operator from $L^1(\mathbf{R}^n)$ to $L^\infty(\mathbf{R}^n)$, because $\|\widehat{f}\|_\infty \leqslant \|f\|_1$. Furthermore, $\widehat{f} \in C_0(\mathbf{R}^n)$, in the sense that it is uniformly continuous if $f \in L^1(\mathbf{R}^n)$ and vanishes at ∞.*

Proof. The initial inequality is proven by taking absolutely values into an integral. This shows how weak the inequality is – it essentially removes all the oscillatory properties of a function we might be able to use to improve our bounds. Indeed, for any ξ, we have

$$|\widehat{f}(\xi)| = \left| \int f(x) e(-\xi \cdot x) \, dx \right| \leqslant \int |f(x)| |e(-\xi \cdot x)| \, dx = \|f\|_1$$

Now if χ_I is the characteristic function of an n dimensional box

$$I = [a_1, b_1] \times \cdots \times [a_n, b_n] = I_1 \times \cdots \times I_n$$

then

$$\widehat{\chi_I}(\xi) = \int_I e(-\xi \cdot x) = \prod_{k=1}^{n} \int_{a_k}^{b_k} e(-\xi_k x_k) = \prod_{k=1}^{n} \widehat{\chi_{I_k}}(\xi_k)$$

35

where

$$\widehat{\chi_{I_k}}(\xi_k) = \begin{cases} \frac{e(-\xi_k a_k) - e(-\xi_k b_k)}{2\pi i \xi_k} & \xi_k \neq 0 \\ b_k - a_k & \xi_k = 0 \end{cases}$$

We know that $\|\widehat{\chi_{I_k}}\|_\infty \leqslant \|\chi_{I_k}\|_1 = b_k - a_k$, and also $|\widehat{\chi_{I_k}}(\xi_k)| \leqslant |\pi \xi_k|^{-1}$. But this gives a global bound

$$|\widehat{\chi_I}(\xi)| \leqslant \frac{1}{\pi \max |\xi_k|} \prod_{k=1}^n \max(1, b_k - a_k) \leqslant \frac{n^{1/2}}{\pi |\xi|} \prod_{k=1}^n \max(1, b_k - a_k)$$

thus $|\widehat{\chi_I}(\xi)| = O(1/|\xi|)$, hence $\widehat{\chi_I} \in C_0(\mathbf{R}^n)$. As a product of uniformly continuous functions, $\widehat{\chi_I}$ is also uniformly continuous. Since the family $C_0(\mathbf{R}^n)$ form a vector space, this shows that all simple functions formed by characteristic functions over intervals have a uniformly continuous Fourier transform. But this vector space is dense in $L^1(\mathbf{R}^n)$, so if f is an arbitrary absolutely integrable function, we can find $f_1, f_2, \ldots$ converging to f in the L^1 norm, which implies $\widehat{f_n}$ converges to $\widehat{f}$ uniformly. If we assume $\widehat{f_n} \in C_0(\mathbf{R}^n)$, then we conclude $\widehat{f} \in C_0(\mathbf{R}^n)$ because $C_0(\mathbf{R}^n)$ is closed under uniform limits. $\qquad\square$

Remark. The theorem above implies that the *Fourier algebra* $\mathbf{A}(\mathbf{R}^n)$ of functions which are the Fourier transforms of functions in $L^1(\mathbf{R}^n)$ is contained within $C_0(\mathbf{R}^n)$. However, it is not the case that $\mathbf{A}(\mathbf{R}^n) = C_0(\mathbf{R}^n)$, and as of yet no current research has given a satisfying description of the elements of $\mathbf{A}(\mathbf{R}^n)$.

Elementary properties of integration give the following relations among the Fourier transforms of functions on $\mathbf{R}^n$. They connect to the translation invariance of the Lebesgue integral on $\mathbf{R}^n$:

- $\mathcal{F}$ is a linear operator from $L^1(\mathbf{R}^n)$ to $C_0(\mathbf{R}^n)$.

- If $\overline{f}(x) = \overline{f(x)}$ is the conjugate of a function f, then

$$(\overline{f})^\wedge(\xi) = \int \overline{f(x)} e(-x \cdot \xi) = \overline{\int f(x) e(x \cdot \xi)} = \overline{\widehat{f}(-\xi)}$$

- There is a duality between translation and frequency modulation. If $y \in \mathbf{R}^n$, we define the translational operator T_y and the frequency modulation operator M_y by

$$(T_y f)(x) = f(x - y) \qquad (M_y f)(x) = e(y \cdot x) f(x)$$

then $(T_y f)^\wedge = M_{-y}\widehat{f}$ and $(M_y f)^\wedge = T_y\widehat{f}$ because

$$\int f(x-y)e(-\xi \cdot x) = \int f(x)e(-\xi \cdot x + y) = \widehat{f}(\xi)e(-y \cdot \xi)$$

$$\int f(x)e(y \cdot x)e(-\xi \cdot x) = \int f(x)e(-(\xi - y) \cdot x)$$

- If $(\delta_a f)(x) = f(ax)$, for $a \in \mathbf{R}$, then

$$(\delta_a f)^\wedge(\xi) = \int f(ax)e(-x \cdot \xi) = a^{-n}\int f(y)e(-y \cdot (\xi/a)) = a^{-n}\widehat{f}(\xi/a)$$

- If $R \in O_n(\mathbf{R})$ is a rotation in n dimensional space, and we define $(Rf)(x) = f(Rx)$, then since $R^{-1}x \cdot \xi = x \cdot R\xi$,

$$(Rf)^\wedge(\xi) = \int f(Rx)e(-x \cdot \xi) = \int f(x)e(-R^{-1}x \cdot \xi)$$

$$= \int f(x)e(-x \cdot R\xi) = \widehat{f}(R\xi)$$

In particular, the negation rotation $x \mapsto -x$ implies that if f is symmetric, so $f(x) = f(-x)$, then $\widehat{f}(\xi) = \widehat{f}(-\xi)$, so $\widehat{f}$ is even. If f is odd, so $f(x) = -f(-x)$, we find that $\widehat{f}$ is odd. If f is a rotationally symmetric function, then so too is it's Fourier transform.

Thus the symmetries of the Lebesgue measure are naturally related in the symmetries of the Fourier transform. This is what makes the Fourier transform so useful. Another instance of this symmetry occurs in a connection with the differentiation of functions.

Theorem 3.2. *If $f \in L^1(\mathbf{R}^n)$, and $x_k f \in L^1(\mathbf{R}^n)$, then $\widehat{f}$ is differentiable in the kth variable, with*

$$(D_k\widehat{f})(\xi) = (-2\pi i x_k f)^\wedge(\xi)$$

Proof. Note that

$$\frac{\widehat{f}(\xi + he_k) - \widehat{f}(\xi)}{h} = \int f(x)e^{-2\pi i \xi x}\frac{e^{-2\pi i h x_k} - 1}{h} = \widehat{g_h}(\xi)$$

where

$$g_h(x) = f(x)\frac{e^{-2\pi i h x_k} - 1}{h}$$

37

As $h \to 0$, the dominated convergence theorem says that, since for h lying in a fixed, bounded interval,

$$\left| \frac{e^{-2\pi i h x_k} - 1}{h} \right| = O(1 + |x_k|)$$

we conclude that $\widehat{g_h}(\xi)$ converges pointwise to $(-2\pi i x_k f)^\wedge(\xi)$, which is the required theorem. □

Conversely, if the partial derivatives of a function are nice enough that we can take Fourier transforms, we obtain a dual result. We say f has a partial derivative $D_k f$ in L^1 if the functions

$$(\Delta_h f)(x) = \frac{f(x + h e_k) - f(x)}{h}$$

converge in L^1 to $D_k f$. Note that $D_k f$ might not always agree with the usual partial derivative in pathological cases, but does agree if f is continuously differentiable and has compact support.

Theorem 3.3. *If $f \in C_c^1(\mathbf{R}^n)$, then f has a partial derivative $D_k f$ in L^1.*

Proof. Because f has compact support, $D_k f$ is uniformly continuous. Then for each $\varepsilon > 0$, there is δ such that $|(D_k f)(x) - D_k(f)(y)| < \varepsilon$ if $|x - y| < \delta$. The mean value theorem implies that for any h, there exists s between 0 and h with

$$(\Delta_h f)(x) = \frac{f(x + h e_k) - f(x)}{h} = (D_k f)(x + s e_k)$$

It follows that if $|h| < \delta$, then

$$|(\Delta_h f)(x) - (D_k f)(x)| = |(D_k f)(x + s e_k) - (D_k f)(x)| < \varepsilon$$

Thus if f is supported on a compact set E, then $\|\Delta_h f - D_k f\|_1 < \varepsilon |E|$. □

Remark. If f no longer has compact support, but $D_k f$ vanishes rapidly at infinity, then we can normally still establish that $D_k f$ is the derivative of f in $L^1(\mathbf{R}^n)$. Indeed, suppose $|(D_k f)(x)| \leqslant g(|x|)$, where g is an increasing function with $\int_0^\infty t^{n-1} g(t) < \infty$, then surely $\Delta_h f$ converges to $D_k f$ in L^1

on any compact set, which implies that for any M, using the mean value theorem again,

$$\int_{\mathbf{R}^n} |(\Delta_h f)(x) - D_k f(x)| \, dx \leqslant o_M(1) + \int_{|x|>M} |(\Delta_h f)(x)| + |D_k f(x)|$$

$$\leqslant o_M(1) + O\left(\int_{|x|>M} g(|x| + |h|) \, dx\right) = o_M(1) + O\left(\int_M^\infty t^{n-1} g(t) \, dt\right)$$

If we choose M large enough that the big O term is $\leqslant \varepsilon$, then we find $\|\Delta_h f - D_k f\|_1 \leqslant \varepsilon + o_M(1)$, and taking $\varepsilon \to 0$ shows the convergence. This shows the derivatives exist if, for instance, f is a Schwarz function, since then $|D_k f(x)| \lesssim 1/(1 + |x|^{n+1})$.

Theorem 3.4. *If f has a partial derivative $D_k f$ in L^1, then $\widehat{D_k f}(\xi) = 2\pi i \xi_k \widehat{f}(\xi)$.*

Proof. It suffices to note that

$$\widehat{\Delta_h f}(\xi) = \frac{e^{2\pi i h \xi} - 1}{h} \widehat{f}(\xi)$$

Since $\Delta_h f \to f$ in L^1, $\widehat{\Delta_h f} \to \widehat{D_k f}$ uniformly, and in particular, converges to $\widehat{D_k f}$ pointwise, and the $\widehat{\Delta_h f}$ converge pointwise to $2\pi i \xi_k \widehat{f}(\xi)$. $\qquad\square$

These theorems have obvious extensions. If $P(X_1, \ldots, X_n)$ is a polynomial in n variables, then we let $P(D)$ denote the differential operator $P(D_1, \ldots, D_n)$. We therefore find that if $P(D)(f)$ exists in L^1, then

$$(P(D)(f))^\wedge = P(2\pi i \xi) \widehat{f}$$

and if $Pf \in L^1(\mathbf{R}^n)$, then

$$P(D)(\widehat{f}) = (P(-2\pi i \xi) f)^\wedge$$

Thus differentiability in the spatial domain corresponds to a polynomial decay in the frequency space, and vice versa.

3.1 Convolutions and Kernels

As with Fourier analysis on $\mathbf{T}$, convolution plays a major role in Fourier analysis on $\mathbf{R}^n$. If f and g are Lebesgue measurable, then

$$(x,y) \mapsto f(x-y)g(y)$$

is also Lebesgue measurable, because it can be seen as composed of the three maps $A \circ B \circ C$, where

$$A : (x,y) \mapsto (x-y,y) \quad B : (x,y) \mapsto (f(x),g(y)) \quad C : (x,y) \mapsto xy$$

C is continuous, hence Borel measurable. B is Lebesgue measurable, because if $E \times F$ is a product of Borel sets, then $B^{-1}(E \times F) = f^{-1}(E) \times g^{-1}(F)$ is the product of Lebesgue measurable sets, hence Lebesgue measurable. A is an automorphism of the Lebesgue σ algebra, since it has a Lebesgue measurable inverse $A^{-1}(x,y) = (x+y,y)$. Putting these facts together gives that $A \circ B \circ C$ is Lebesgue measurable. Furthermore, if $f,g \in L^1(\mathbf{R}^n)$, then Fubini's theorem guarantees that the *convolution* function

$$(f * g)(x) = \int f(x-y)g(y)\,dy$$

is commutative, defined almost everywhere, and is in $L^1(\mathbf{R}^n)$. Fubini's theorem also guarantees the operation is commutative and associative, and since $\|f * g\|_1 \leqslant \|f\|_1 \|g\|_1$, the operation turns $L^1(\mathbf{R}^n)$ into a Banach algebra.

Theorem 3.5. *If $f,g \in L^1(\mathbf{R}^n)$, then $\widehat{f * g} = \hat{f}\hat{g}$.*

Proof. By Fubini's theorem,

$$\widehat{f * g}(\xi) = \int\int f(y)g(x-y)e^{-x\cdot\xi}\,dx\,dy = \int \mathcal{F}(T_y g)(\xi)f(y)\,dy$$

$$= \hat{g}(\xi)\int f(y)e^{-y\cdot\xi}\,dy = \hat{g}(\xi)\hat{f}(\xi)$$

which is exactly what we wanted to verify. $\qquad\qquad\square$

In real analysis, it is proved that if $\{K_\delta\}$ is a good kernel, i.e. elements of $L^1(\mathbf{R}^n)$ satisfying $\int K_\delta(x)\,dx = 1$ and such that for any ε,

$$\lim_{\delta\to 0}\int_{|x|>\varepsilon}|K_\delta(x)| = 0$$

then $f * K_\delta$ converges to f in the L^1 norm as $\delta \to 0$. If in addition, the K_δ satisfy $|K_\delta(x)| \lesssim \delta^{-n}$, and $|K_\delta(x)| \lesssim \delta/|x|^{n+1}$, then $f * K_\delta$ converges to f almost everywhere. A common way to form such a kernel is to take some function $K \in L^1(\mathbf{R}^n)$ with $\int K = 1$ and to write $K_\delta(x) = \delta^{-n} K(x/\delta)$. The integrality condition is expressed by the fact that

$$\int_{|x|>\varepsilon} \delta^{-n} |K(x/\delta)| \, dx = \int_{|y|>\varepsilon/\delta} |K(y)| \, dy \to 0$$

So such a family is always a good kernel. If, in addition, K is bounded and has compact support, then we obtain an approximation to the identity. The advantage of this is that we can swap a general integrable function f in Fourier analysis with a suitably nice convolved function $f * K$, and this theorem shows that we can choose K to approximate f arbitrarily well. In particular, if $\widehat{k} = K$, and we define $k_\delta(x) = k(\delta x)$, then $\widehat{k_\delta} = K_\delta$.

3.2   Convergence Using Alternative Summation

As we might expect from the Fourier series theory, the formula

$$f(x) = \int \widehat{f}(\xi) e^{2\pi i \xi \cdot x} \, dx$$

does not hold for every integrable f, nor even for all continuous f, for $\widehat{f}$ may not be an integrable function. Nonetheless, this formula is 'very close to holding' for all integrable f, by interpreting the formula according to alternate summability methods.

Example. *Even if f is a non integrable function, the functions $f(x)e^{-\delta|x|}$ may be integrable, and so we can consider the Abel means*

$$A_\delta(f) = \int f(x) e^{-\delta|x|} \, dx$$

We say f is Abel summable if $A_\delta(f)$ converges as $\delta \to 0$. If f is integrable, then the dominated convergence theorem implies that $A_\delta(f) \to \int f(x)$. However, f may be Abel summable even if f is not integrable. For instance, if $f(x) = \sin(x)/x$, then f is not integrable, yet f is Abel summable to π over the real line. We shall find that the Fourier transform of f is Abel summable to f almost everywhere and in the L^1 norm.

41

Example. *Similarily, we can consider the Gauss sums*

$$G_\delta(f) = \int f(x) e^{-\delta|x|^2} \, dx$$

and we say f is Gauss summable to some value if $G_\delta(f)$ converges to a common value as $\delta \to 0$.

Example. *In basic calculus, the integral of a function f over the entire real line is defined as*

$$\int_{-\infty}^{\infty} f(x) \, dx = \lim_{N \to \infty} \int_{-N}^{N} f(x) \, dx$$

These integrals can be written as the integral of $f \chi_{[-N,N]}$, and so in a generalized sense, we can integrate a function f if $f \chi_{[-N,N]}$ is integrable for each N, and the integrals of these functions converge as $N \to \infty$. As an example, the function $f(x) = \sin(x)/x$ is non-integrable, yet we have

$$\lim_{N \to \infty} \int_{-N}^{N} \frac{\sin(x)}{x} \, dx = \pi$$

which follows because f oscillates between positive and negative values that get smaller and smaller over time.

More generally, if Φ_δ is a family of functions, then we can consider the 'Φ sums'

$$S_\delta(f) = \int f(x) \Phi_\delta(x) \, dx$$

and we say f is Φ summable to a value if $S_\delta(f)$ converges. In all the examples we will consider, we fix a function $\Phi \in C_0(\mathbf{R}^n)$ with $\Phi(0) = 1$, and define $\Phi_\delta(x) = \Phi(\delta x)$. In this case, whenever f is integrable, then $f(x)\Phi_\delta(x)$ converges to $f(x)$ for each x, and since these values are dominated by $|f|\|\Phi\|_\infty$, we conclude that f is Φ summable to it's actual sum, so Φ summation effectively generalizes integration. Using convolution, we will now show that for a large class of choices of Φ, the Φ sums

$$S_\delta\left(M_x \widehat{f}\right) = \int \widehat{f}(\xi) e(\xi \cdot x) \Phi(\delta \xi) \, d\xi$$

converge for almost all x to $f(x)$, and in the L^1 norm.

Theorem 3.6. *If $f, g \in L^1(\mathbf{R}^n)$,*

$$\int f(x)\widehat{g}(x)\, dx = \int \widehat{f}(\xi)g(\xi)\, dx$$

Proof. If $f, g \in L^1(\mathbf{R}^n)$, then $\widehat{f}$ and $\widehat{g}$ are bounded, continuous functions on $\mathbf{R}^n$. In particular, $\widehat{f}g$ and $f\widehat{g}$ are integrable. A simple of Fubini's theorem gives

$$\int f(x)\widehat{g}(x)\, dx = \int\int f(x)g(\xi)e^{-2\pi i \xi \cdot x}\, dx\, d\xi = \int g(\xi)\widehat{f}(\xi)\, d\xi$$

Thus the 'multiplication formula' holds. $\square$

If Φ is an integrable function, then we can apply the multiplication formula to conclude that

$$\int \widehat{f}(\xi)e(\xi \cdot x)\Phi(\delta\xi) = \int f(x)(M_x(\delta_\delta\Phi))^\wedge(x)$$

$$= \int f(x)\frac{\widehat{\Phi}\left(\frac{x-y}{\delta}\right)}{\delta^n}\, dx$$

If $K(x) = \widehat{\Phi}(-x)$, then we can summarize this calculation by saying

$$\int \widehat{f}(\xi)e(\xi \cdot x)\Phi(\delta\xi)\, d\xi = (f * K_\delta)(x)$$

Thus provided the K_δ form a good kernel, the Φ means converge to f in L^1, and if the K_δ form an approximation to the identity, then the Φ means converge to f almost everywhere, in particular, on the Lebesgue set of f, which includes all points of continuity of f.

In particular, suppose that such a Φ exists, with K_δ an approximation to the identity (this is most easy to see for the Fejér kernel below, since the corresponding K_δ is compactly supported). Then for any x in the Lebesgue set of f,

$$f(x) = \lim_{\delta \to 0} \int \widehat{f}(\xi)e(\xi \cdot x)\Phi(\delta x)$$

Note $\widehat{f}(\xi)e(\xi \cdot x)\Phi(\delta x)$ converge to $\widehat{f}(\xi)e(\xi \cdot x)$ pointwise. If we assume $\widehat{f}$ is integrable, then because of the bound

$$\left|\widehat{f}(\xi)e(\xi \cdot x)\Phi(\delta\xi)\right| \leqslant \|\Phi\|_\infty |\widehat{f}(\xi)|$$

43

we can apply the dominated convergence theorem to conclude that for any point x in the Lebesgue set of f,

$$f(x) = \int \widehat{f}(\xi) e(\xi \cdot x) = \lim_{\delta \to 0} \int \widehat{f}(\xi) e(\xi \cdot x) \Phi(\delta x) = \int \widehat{f}(\xi) e(\xi \cdot x)$$

and the limit also holds in the L^1 norm. Thus we have justified that the inversion theorem holds.

Theorem 3.7. *If f and $\widehat{f}$ are integrable, then for any x in the Lebesgue set,*

$$f(x) = \int \widehat{f}(\xi) e(\xi \cdot x) \, d\xi = \widecheck{\widehat{f}}(x)$$

where $\widecheck{f}$ is the inverse Fourier transform.

Remark. Using the same calculations as for the Fourier transform, one can show that if $f \in L^1(\mathbf{R}^n)$, then $\widecheck{f} \in C_0(\mathbf{R}^n)$. Thus we have argued that if f and it's Fourier transform are integrable, then f is equal to a $C_0(\mathbf{R}^n)$ function almost everywhere. So we can only really use the full Fourier inversion formula on a certain subclass of C_0 functions. However, this class is suitably dense in $L^1(\mathbf{R}^n)$, and so this formula doesn't cause us many problems. Indeed, if f satisfies the inversion formula, in the sense that there exists an integrable function F such that $f = \widecheck{F}$, and g is any L^1 function, then $f * g = (\widehat{g}F)^\vee$, so $f * g$ also 'satisfies the inversion formula'. This is another way we can obtain the general inversion formula, because if we can verify that a single pair kernel K satisfies the inversion formula, we can approximate an arbitrary function f by a family of functions $f * K_\delta$ which *does* satisfy the inversion formula. This is easily verified for the Fejér and Gauss kernel's below.

A simple corollary to the Fourier inversion theorem is that the Fourier transform is an injective linear operator. Indeed, if the Fourier transform of f vanishes, then it is certainly integrable, and the Fourier inversion theorem says that f is equal to zero almost everywhere. Thus if $\widehat{f} = \widehat{g}$, then $f = g$.

Example. *We obtain the Fejér kernel F_δ from the initial function*

$$F(x) = \left(\frac{\sin \pi x}{\pi x} \right)^2$$

44

We can see that the function F satisfies $\int F(x) \, dx = 1$ by applying contour integration. More generally, we will show that

$$\widehat{F}(\xi) = \begin{cases} 1 - |\xi| & : |\xi| \leqslant 1 \\ 0 & : |\xi| > 1 \end{cases}$$

Since F is an even function, $\widehat{F}$ is even, and so we may assume $\xi \geqslant 0$. We initially calculate

$$\widehat{F}(\xi) = \int_{-\infty}^{\infty} \left(\frac{\sin(\pi x)}{\pi x} \right)^2 e^{-2\pi i \xi x} \, dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \left(\frac{\sin x}{x} \right)^2 e^{-2i\xi x} \, dx$$

Now we have

$$(\sin z)^2 = \left(\frac{e^{iz} - e^{-iz}}{2i} \right)^2 = \frac{(2 - e^{2iz}) - e^{-2iz}}{4}$$

This means

$$\frac{(\sin z)^2}{z^2} e^{-2i\xi z} = \frac{2e^{-2i\xi z} - e^{-2(\xi+1)iz}) - e^{-2(\xi-1)iz}}{4z^2} = \frac{f_\xi(z) + g_\xi(z)}{4}$$

For $\xi \geqslant 0$, $f_\xi(z)$ is $O_\xi(1/|z|^2)$ in the lower half plane, because if $\text{Im}(z) \leqslant 0$,

$$|2e^{-2i\xi z} - e^{-2(\xi+1)z}| \leqslant 2e^{2\xi} + e^{2(\xi+1)} = O_\xi(1)$$

For $\xi \geqslant 1$, $g_\xi(z)$ is also $O_\xi(1/|z|^2)$ in the lower half plane, because

$$|e^{-2(\xi-1)iz}| \leqslant e^{2(\xi-1)}$$

Now since $(\sin x/x)^2 e^{-2i\xi x}$ can be extended to an entire function on the entire complex plane, which is bounded on any horizontal strip, we can apply Cauchy's theorem and take limits to conclude that

$$\widehat{F}(\xi) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{(\sin x)^2}{x^2} e^{-2i\xi x} \, dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{(\sin(x - iy)^2}{(x - iy)^2} e^{-2i\xi x - 2\xi y} \, dx$$

$$= \frac{1}{4\pi} \int_{-\infty}^{\infty} f_\xi(x - iy) + g_\xi(x - iy) \, dx$$

If $\xi \geqslant 1$, the functions f_ξ and g_ξ are both negligible in the lower half plane, and have no poles in the lower half plane, so if we let γ denote the curve of length

45

$2\pi n$ travelling anticlockwise along the lower semicircle with vertices $-n - iy$ and $n - iy$, then because $|z| \geqslant n$ on γ,

$$
\int_{-n}^{n} f_\xi(x - iy) + g_\xi(x - iy) \, dx = \int_\gamma f_\xi(z) + g_\xi(z) \, dz
$$

$$
= length(\gamma)\|f_\xi + g_\xi\|_{L^\infty(\gamma)}
$$

$$
= (2\pi n)O_\xi(1/n^2) = O_\xi(1/n)
$$

and so we conclude that

$$
\int_{-\infty}^{\infty} f_\xi(x - iy) + g_\xi(x - iy) \, dx = 0
$$

This means $\widehat{F}(\xi) = 0$. If $0 \leqslant \xi \leqslant 1$, then f_ξ is still small in the lower half plane, so we can conclude that

$$
\int_{-\infty}^{\infty} f_\xi(x - iy) \, dx = 0
$$

But g_ξ is now small in the upper half plane. For $Im(z) \geqslant -y$,

$$
|e^{-2(\xi-1)iz}| = |e^{2(1-\xi)iz}| \leqslant e^{2(1-\xi)y}
$$

so $g_\xi(z) = O_\xi(1/|z|^2)$ in the half plane above the line $\mathbf{R} - iy$. The only problem now is that g_ξ has a pole in this upper half plane, at the origin. Taking Laurent series here, we find that the residue at this point is $2i(\xi - 1)$. Thus, if we let γ be the curve obtained from travelling anticlockwise about the upper semicircle with vertices $-n - iy$ and $n - iy$, then $|z| \geqslant n - y$ on this curve, and the residue theorem tells us that

$$
\int_{-n}^{n} g_\xi(x - iy) \, dx + \int_\gamma g_\xi(z) \, dz = 2\pi i(2i(\xi - 1)) = 4\pi(1 - \xi)
$$

and we now find that, as with the evaluation of the previous case,

$$
\int_\gamma g_\xi(z) \, dz \leqslant (2\pi n)O_{\xi,y}(1/n^2) = O_{\xi,y}(1/n)
$$

Taking $n \to \infty$, we conclude

$$
\int_{-\infty}^{\infty} g_\xi(x - iy) \, dx = 4\pi(1 - \xi)
$$

and putting this all together, we conclude that $\widehat{F}(\xi) = 1 - \xi$. It is obviously true that $\int \widehat{F}(\xi) = 1$, because the graph of this function is just a triangle with a baselength 2 and heigh 1, so we obtain that for any $f \in L^1(\mathbf{R})$, $S_\delta(f)$ converges to f in the L^1 norm, and since the Fourier transform has compact support, it is easy to see it is an approximation to the identity, so that for any x in the Lebesgue set of f,

$$f(x) = \lim_{\delta \to 0} \int \widehat{f}(\xi) e(\xi \cdot x) \left(\frac{\sin(\delta \xi)}{\delta \xi} \right)^2 dx$$

It is interesting in this particular case to note that

$$\int_{-1}^{1} (1 - |\xi|) e^{2\pi i \xi x} \, d\xi = 2 \int_0^1 (1 - \xi) \cos(2\pi \xi x) \, d\xi$$

$$= 2 \left(\frac{(1 - \xi)\sin(2\pi \xi x)}{2\pi x} - \frac{\cos(2\pi \xi x)}{(2\pi x)^2} \Big|_0^1 \right)$$

$$= 2 \frac{1 - \cos(2\pi x)}{(2\pi x)^2} = \frac{\sin^2(\pi x)}{(\pi x)^2} = F(x)$$

which is exactly the inversion formula we want for all L^1 functions.

Example. *The Gauss kernel Φ_δ on $\mathbf{R}^n$ is obtained from the initial function $\Phi(x) = e^{-\pi |x|^2}$. Since this function breaks onto products of exponentials over each coordinate, it suffices to calculate the Fourier transform in one dimension, from which we can obtain the general transform by taking products. In the one dimensional case, since $\Phi'(x) = -2\pi x e^{-\pi x^2}$ is integrable, we conclude that $\widehat{\Phi}$ is differentiable, and*

$$(\widehat{\Phi})'(\xi) = (-2\pi i \xi \Phi)^\wedge(\xi) = i(\Phi')^\wedge(\xi) = i(2\pi i \xi)\widehat{\Phi}(\xi) = -2\pi \xi \widehat{\Phi}(\xi)$$

The uniqueness theorem for ordinary differential equations says that since

$$\widehat{\Phi}(0) = \int_{-\infty}^{\infty} e^{-\pi x^2} = 1 = \Phi(0)$$

We must have $\widehat{\Phi} = \Phi$. But this means that in all dimensions, $\Phi(x) = e^{-\pi |x|^2}$, then $\widehat{\Phi} = \Phi$. Since the Gauss kernel is one of the most well known approximations to the identity, we find that $G_\delta(f) \to f$ in L^1 for all $f \in L^1(\mathbf{R}^n)$, and for

all x in the Lebesgue set of f,

$$f(x) = \int \widehat{f}(\xi) e(\xi \cdot x) e^{-\pi |\xi|^2} d\xi$$

The scaling properties of the Fourier transform imply that

$$\widehat{\Phi_\delta}(\xi) = \widehat{\Phi}(\delta \xi) = e^{-\delta^2 \pi |\xi|^2}$$

Note that as $\delta \to 0$, Φ_δ becomes more and more peaked at the origin in the spatial domain, and it's Fourier transform becomes spread out. This is a instance of the Heisenberg uncertainty principle, which says that the Fourier transform of a localized function much be very well spread, and vice versa.

Example. *The Abel kernel A_δ on $\mathbf{R}^n$ is obtained from the initial function $A(x) = e^{-2\pi |x|}$. The calculation of the Fourier transform of this function indicates a useful principle in Fourier analysis: one can reduce expressions involving e^{-x} into expressions involving e^{-x^2} using the subordination principle. In particular, for $\beta > 0$ we have the formula*

$$e^{-\beta} = \int_0^\infty \frac{e^{-u}}{\sqrt{\pi u}} e^{-\beta^2/4u} \, du$$

We establish this by letting $v = \sqrt{u}$, so

$$\int_0^\infty \frac{e^{-u}}{\sqrt{\pi u}} e^{-\beta^2/4u} \, du = \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-v^2 - \beta^2/4v^2} \, dv = \frac{2e^{-\beta}}{\sqrt{\pi}} \int_0^\infty e^{-(v - \beta/2v)^2} \, dv$$

But the map $v \mapsto v - \beta/2v$ is measure preserving by Glasser's master theorem, so this integral is

$$\frac{2e^\beta}{\sqrt{\pi}} \int_0^\infty e^{-v^2} \, dv = e^\beta$$

In tandem with Fubini's theorem, this formula implies

$$\widehat{A}(\xi) = \int e^{-2\pi |x|} e^{-2\pi i \xi \cdot x} \, dx = \int \int_0^\infty \frac{e^{-u}}{\sqrt{\pi u}} e^{-|\pi x|^2/u} e^{-2\pi i \xi \cdot x} \, du \, dx$$

$$= \int_0^\infty \frac{e^{-u}}{\sqrt{\pi u}} \int e^{-|\pi x|^2/u} e^{-2\pi i \xi \cdot x} \, dx \, du = \int_0^\infty \frac{e^{-u}}{\sqrt{\pi u}} (\delta_{\sqrt{\pi/u}} \Phi)^\wedge(\xi) \, du$$

$$= \frac{1}{\pi^{(n+1)/2}} \int_0^\infty e^{-u} u^{(n-1)/2} e^{-u|\xi|^2} \, du$$

48

Setting $v = (1 + |\xi|^2)u$, we conclude that since by definition,

$$\int_0^\infty e^{-v} v^{(n-1)/2} = \Gamma\left(\frac{n+1}{2}\right)$$

$$\widehat{A}(\xi) = \frac{\Gamma((n+1)/2)}{[\pi(1 + |\xi|^2)]^{(n+1)/2}}$$

Thus the Abel mean is the Fourier inverse of the Poisson kernel on the upper half plane $\mathbf{H}^{n+1}$. In order to conclude this function is a good kernel, it now suffices to verify that

$$\int_{\mathbf{R}^n} \frac{d\xi}{(1 + |\xi|^2)^{(n+1)/2}} = \frac{\pi^{(n+1)/2}}{\Gamma((n+1)/2)}$$

The right hand side is half the surface area of the unit sphere in $\mathbf{R}^{n+1}$. Denoting this quantity by S_n, and switching to polar coordinates, we find that

$$\int_{\mathbf{R}^n} \frac{d\xi}{(1 + |\xi|^2)^{(n+1)/2}} = S_{n-1} \int_0^\infty \frac{r^{n-1}}{(1 + r^2)^{(n+1)/2}} \, dr$$

Setting $r = \tan u$, we find

$$\int_0^\infty \frac{r^{n-1}}{(1 + r^2)^{(n+1)/2}} \, dr = \int_0^{\pi/2} (\sin u)^{n-1} du$$

The theorem now follows from noticing that $S_{n-1}(\sin u)^{n-1}$ is the surface area of the $n - 1$ sphere obtained by slicing S^n with the hyperplane $x_n = \cos u$. Fubini's theorem implies that the integral is $S_n/2$, which is what we wanted to verify.

Theorem 3.8. *If f is an integrable, function continuous at the origin, and $\widehat{f} \geqslant 0$, then $\widehat{f}$ is integrable.*

Proof. This follows because

$$f(0) = \lim_{\delta \to 0} \int \widehat{f}(\xi) e^{-\delta|x|}$$

By Fatou's lemma,

$$f(0) = \lim_{\delta \to 0} \int \widehat{f}(\xi) e^{-\delta|x|} \geqslant \int \liminf_{\delta \to 0} \widehat{f}(\xi) e^{-\delta|x|} = \int \widehat{f}(\xi)$$

so $\widehat{f}$ is finitely integrable. $\qquad\qquad\square$

Note that this implies that we obtain the general inversion theorem, so in particular, it is only continuous functions, and functions almost everywhere equal to continuous functions, which can have non-negative Fourier transforms.

3.3   The L^2 Theory

Once of the most useful components of the L^2 theory of Fourier series is the Plancherel inequality is the Plancherel inequality, that

$$\sum |\widehat{f}(n)|^2 = \frac{1}{2\pi} \int_0^{2\pi} |f(x)|^2$$

On $\mathbf{R}^n$, we would like to justify that

$$\int |\widehat{f}(\xi)|^2 \, d\xi = \int |f(x)|^2 \, dx$$

However, on the non-compact Euclidean space, a general element of $L^2(\mathbf{R}^n)$ is not necessarily integrable, so we cannot take it's Fourier transform using the integral formula. Nonetheless, we can take the Fourier transform of an element of $L^1(\mathbf{R}^n) \cap L^2(\mathbf{R}^n)$, and we find the equation holds.

Theorem 3.9. *If $f \in L^1(\mathbf{R}^n) \cap L^2(\mathbf{R}^n)$, then $\|\widehat{f}\|_2 = \|f\|_2$.*

Proof. If $g(x) = \overline{f(-x)}$, then $\widehat{g}(\xi) = \overline{\widehat{f}(\xi)}$, so

$$(f * g)^\wedge(\xi) = \widehat{f}(\xi)\widehat{g}(\xi) = |\widehat{f}(\xi)|^2$$

Now we know $(f * g)^\wedge$ is integrable, hence for almost all x,

$$(f * g)(x) = \int |\widehat{f}(\xi)|^2 e(\xi \cdot x) \, d\xi$$

In particular, if this holds for $x = 0$ (if we choose f to be continuous), then we conclude

$$\int |f(x)|^2 \, dx = \int |\widehat{f}(\xi)|^2 \, d\xi$$

Then we can apply a density argument to get the result in general. $\square$

Though we cannot interpret the Fourier transform of an element of $L^2(\mathbf{R}^n)$ in terms of an integral, the theorem above provides a canonical way to define the Fourier transform of such an element. Since $L^1(\mathbf{R}^n) \cap L^2(\mathbf{R}^n)$ is dense in $L^2(\mathbf{R}^n)$, and we have shown the Fourier transform is continuous on this space in the L^2 norm (it is actually an isometry in this norm), the Hahn-Banach theorem implies that the Fourier transform extends to a unique isometry from $L^2(\mathbf{R}^n)$ to itself. In order to calculate the Fourier transform of a general element f of $L^2(\mathbf{R}^n)$, we approximate it by elements $f_n \in L^1(\mathbf{R}^n) \cap L^2(\mathbf{R}^n)$, and consider the L^2 limit of the functions $\widehat{f_n}$. The multiplication formula, by continuity, extends to elements of $L^2(\mathbf{R}^n)$, which implies that the Fourier transform is self adjoint. But every self adjoint isometry is unitary, so the Fourier transform is a unitary transformation of $L^2(\mathbf{R}^n)$. Of course, the L^2 extension of the inverse Fourier transform is the adjoint operator of the Fourier transform, since on $L^1(\mathbf{R}^n) \cap L^2(\mathbf{R}^n)$, we can apply Fubini's theorem to conclude

$$\int f(\xi)\widehat{g}(\xi) = \int f(\xi)\overline{\int g(x)e(-x\cdot\xi)\,dx}\,d\xi$$

$$= \int \left(\int f(\xi)e(x\cdot\xi)\right)\overline{g(x)}\,dx = \int \check{f}(x)g(x)\,dx$$

Thus the L^2 theory of the Fourier transform is the nicest place to perform Fourier calculation. It also illustrates an important point; useful equations like the multiplication formula and the adjoint calculation above can often be extended to general classes of functions after calculating on a suitably nice class by applying density and continuity arguments.

3.4   Schwartz Space and Tempered Distributions

We have already encountered the fact that Fourier transforms are well behaved under differentiation and multiplication by polynomials. If we let $\mathcal{S}(\mathbf{R}^n)$ denote a class of functions under which to study this phenomenon, it must be contained in $L^1(\mathbf{R}^n)$ and $C^\infty(\mathbf{R}^n)$, and also be closed under multiplication by polynomials. The differentiability and polynomial closure imply that the elements of $\mathcal{S}(\mathbf{R}^n)$ must have rapid decay properties: For any pair of multi-indices α and β,

$$\sup_{x\in\mathbf{R}^n} |x^\alpha||D_\beta f(x)| < \infty$$

51

We take this to be the definition of the **Schwartz space** $\mathcal{S}(\mathbf{R}^n)$: the elements of $C^\infty(\mathbf{R}^n)$ with the rapid decay property above. Examples of Schwartz functions include the Gaussian functions Φ_δ, and compactly supported C^∞ functions.

We have already justified that the Fourier transform maps Schwartz functions to Schwartz functions, and it is easy to see that the converse holds: the inverse Fourier transform also maps Schwartz functions to themselves. The L^2 theory of the Fourier transform then immediately implies that the Fourier transform is a bijective map from the Schwartz space to itself. A nice corollary is that if f and g are Schwartz, then $f * g$ is Schwartz, because it is the inverse Fourier transform of the Schwartz $\widehat{f}\widehat{g}$, and it is easy to see the product of two Schwartz functions is Schwartz.

We now give $\mathcal{S}(\mathbf{R}^n)$ the structure of a Fréchet space, such that the Fourier transform is a homeomorphism of this space. For any two indices α and β, we consider the seminorm $\|f\|_{\alpha\beta} = \|x^\alpha D_\beta f\|_\infty$. These seminorms obviously separate $\mathcal{S}(\mathbf{R}^n)$ (The seminorm for $\alpha = \beta = 0$ even separates the elements). The Schwartz space is complete with respect to this metric, for if $f_1, f_2, \ldots$ is a Cauchy sequence, then the seminorms $\|\cdot\|_{0\beta}$ give that the f_n converge uniformly to some function f, and the partial derivatives $D^\alpha f_n$ converge uniformly to $D^\alpha f$. Clearly also $\|f_n\|_{\alpha\beta} \to \|f\|_{\alpha\beta}$, from which we conclude that f is Schwartz. It is also to see that translations and modulations are continuous operators on the space, and the maps $f \mapsto x^\alpha D^\alpha f$ are also continuous. Furthermore, since the Fourier transform maps the norms $\|\cdot\|_{\alpha\beta}$ to scalar multiples of $\|\cdot\|_{\beta\alpha}$, it is easy to see the Fourier transform is continuous. The inverse transform has the same properties, so the Fourier transform is a homeomorphism.

Now we get to the interesting part of the theory. We have defined a homeomorphic linear transform from $\mathcal{S}(\mathbf{R}^n)$ to itself. The theory of functional analysis then says that we can define a dual map, which is a homeomorphism from the dual space $\mathcal{S}(\mathbf{R}^n)'$ to itself. Since the inclusion map $D(\mathbf{R}^n) \to \mathcal{S}(\mathbf{R}^n)$ from compactly supported C^∞ functions to rapidly decaying functions is continuous, and $D(\mathbf{R}^n)$ is dense in $\mathcal{S}(\mathbf{R}^n)$. The density implies that we have an injective, continuous map from $\mathcal{S}'(\mathbf{R}^n)$ to $D'(\mathbf{R}^n)$, so every element of the dual space can be identified with a distribution on space. We call such distributions **tempered**. They are precisely the linear functionals on $D(\mathbf{R}^n)$ which have a continuous extension to $\mathcal{S}(\mathbf{R}^n)$. This corresponds, in some sense to a growth restriction of the distribution at

infinity.

Example. *Any function in the L^p spaces can be viewed as a tempered distribution, since for any function $f \in L^p(\mathbf{R}^n)$, we can define the continuous map*

$$\Lambda_f(\phi) = \int f(x)\phi(x)\, dx$$

which is continuous because, by Hölder's inequality,

$$\left| \int f(x)\phi(x) \right| \leqslant \|f\|_p \|\phi\|_q$$

and $\|\phi\|_q$ is bounded by a sum of values of the form $\|x^\alpha \phi\|_\infty$ with the constants not depending on ϕ. More generally, if f satisfies $f(x)/(1 + |x|^2)^k \in L^p(\mathbf{R}^n)$ for some k, it is called a tempered function in L^p, and integration against this function acts as a tempered distribution. If $p = \infty$, then f is also known as a slowly increasing function.

Example. *If μ is a finite Borel measure, then the map*

$$\Lambda_\mu(\phi) = \int f(x)d\mu(x)$$

is a tempered distribution, and we identify the distribution with the measure μ. If μ is a measure such that $d\mu(x)/(1 + |x|^2)^k$ is finite for some k, then μ is known as a tempered measure, and also acts as a tempered distribution.

Example. *Every compactly supported distribution is tempered, because if Λ is supported on K, and ψ is a compactly supported bump function on K, then for a Schwarz function ϕ we can define $\Lambda(\phi) = \Lambda(\psi\phi)$, which is a continuous extension.*

Example. *The function $1/x$ is not locally integrable on $\mathbf{R}$, since it is not defined near the origin. However, we can associate the value with a distribution. If ϕ is a Schwartz function, we define the* **principal value**

$$p.v. \int_{-\infty}^{\infty} \frac{\phi(x)}{x}\, dx = \lim_{\varepsilon \to 0} \int_{|x| \geqslant \varepsilon} \frac{\phi(x)}{x}\, dx$$

Since $\int_{\varepsilon \leqslant |x| \leqslant 1} dx/x = 0$ for any $\varepsilon \leqslant 1$, we can write

$$\int_{|x| \geqslant \varepsilon} \frac{\phi(x)}{x} = \int_{|x| \geqslant 1} \frac{\phi(x)}{x} + \int_{\varepsilon \leqslant |x| \leqslant 1} \frac{\phi(x) - \phi(0)}{x}$$

53

Since ϕ has rapid decay, the first integral is well defined. Since ϕ is differentiable at the origin, the function $\phi(x) - \phi(0)/x$ has a removable singularity at the origin, so converges to a well defined quantity as $\varepsilon \to 0$. It is evident that

$$\left| p.v. \int_{-\infty}^{\infty} \frac{\phi(x)}{x} \, dx \right| \lesssim \|\phi\|_{\infty} + \|\phi'\|_{\infty}$$

so this functional is a tempered distribution, denoted by $p.v.(1/x)$. It is intimately connected to the theory of the Hilbert transform.

Using the same techniques as for distributions, the derivative $D^{\alpha}\Lambda$ of a tempered distribution Λ is tempered, as is $\phi\Lambda$, whenever ϕ is a Schwartz function, or $P\Lambda$, where P is a polynomial. To remind the reader, we think of a distribution Λ as corresponding to some function f, nice enough that some operation is well defined for f, such that

$$\Lambda(\phi) = \int f(x)\phi(x) \, dx$$

If we can justify an identity with respect to this operation which removes the reliance of regularity on f, we can normally swap f with a general distribution, and use it to define the operation on all distributions. We now apply this process to define the Fourier transform of a tempered distribution. The multiplication formula

$$\int \widehat{f}(x)g(x) = \int f(x)\widehat{g}(x)$$

provides the perfect way to define the Fourier transform of a distribution. It says that

$$\Lambda_{\widehat{f}}(g) = \Lambda_f(\widehat{g})$$

If we want to generalize the Fourier transform to be defined on distributions, we better have $\widehat{\Lambda_f} = \Lambda_{\widehat{f}}$ for all integrable f. In particular, this motivates us to define the general Fourier transform of a tempered distribution Λ as

$$\widehat{\Lambda}(\phi) = \Lambda\left(\widehat{\phi}\right)$$

If f is in L^1 or L^2, since the multiplication formula is true on these spaces, we see that $\widehat{\Lambda_f} = \Lambda_{\widehat{f}}$, so we're in good shape. Similarily, we can define the inverse Fourier transform

$$\widecheck{\Lambda}(\phi) = \Lambda\left(\widecheck{\phi}\right)$$

These are the dual maps of the Fourier transforms and it's inverse, so they are obviously homeomorphisms of the space of distributions.

Theorem 3.10. *If Λ is tempered, then*

$$(P(D)\Lambda)^\wedge = P(-2\pi i\xi)\widehat{\Lambda} \quad (P(-2\pi i\xi)\Lambda)^\wedge = P(D)\widehat{f}$$

Proof. We calculate

$$(P(D)\Lambda)^\wedge(\phi) = \Lambda(P(-D)\widehat{\phi}) = \Lambda((P(-2\pi i\xi)\phi)^\wedge) = P(\widehat{-2\pi i\xi})\Lambda(\phi)$$

and

$$(P(-2\pi i\xi)\Lambda)^\wedge(\phi) = \Lambda((P(-2\pi i\xi)\phi)^\wedge) = \Lambda(P(-D)\widehat{\phi}) = P(D)(\widehat{\Lambda})(\phi)$$

completing the calculation. $\square$

Example. *Consider the constant function 1. Then*

$$1(\phi) = \int \phi(x)\, dx$$

and so

$$\widehat{1}(\phi) = \int \widehat{\phi}(\xi)\, d\xi = \phi(0) = \delta(\phi)$$

so $\widehat{1} = \delta$. Similarily,

$$\widehat{\delta}(\phi) = \widehat{\phi}(0) = \int \phi(x)\, dx = 1(\phi)$$

Example. *If $P(-2\pi ix)$ is an arbitrary polynomial, then we know*

$$P(\widehat{-2\pi ix}) = P(\widehat{-2\pi ix}) \cdot 1 = P(D)\delta$$

which essentially provides us a way to compute the Fourier transform of any polynomial.

Example. *Let μ be a finite measure. We claim that $\widehat{\mu}$ is a bounded function, with*

$$\widehat{\mu}(\xi) = \int e(-x \cdot \xi)\, d\mu(x)$$

55

To see this, we note that if f is integrable, we can apply Fubini's theorem, calculating

$$\int \widehat{f}(x)d\mu(x) = \int\int f(\xi)e(-\xi \cdot x)d\xi \, d\mu(x)$$

$$= \int f(\xi)\int e(-\xi \cdot x)d\mu(x) \, d\xi$$

$$= \int f(\xi)\widehat{\mu}(\xi) \, d\xi$$

And this gives that the distribution $\widehat{\mu}$ is given by integration with respect to the function $\widehat{\mu}$. It is easy to check that $\|\widehat{\mu}\|_\infty \leqslant \|\mu\|$. Moreover, $\widehat{\mu}$ is continuous, since by the dominated convergence theorem,

$$\widehat{\mu}(\xi + h) = \int e(-x \cdot \xi)e(-x \cdot h)d\mu(x)$$

as $h \to 0$, the values of the function in the integral converge pointwise to $e(-x \cdot \xi)$, and so the continuity follows by the dominated convergence theorem.

Not being compactly supported, we cannot compute the convolution of tempered distributions with all C^∞ functions. Nonetheless, if ϕ is Schwartz, and Λ is tempered, then the definition $(\Lambda * \phi)(x) = \Lambda(T_x\phi^*)$ certainly makes sense, and gives a C^∞ function satisfying $D^\alpha(\Lambda * \phi) = (D^\alpha\Lambda) * \phi = \Lambda * (D^\alpha\phi)$. This function is slowly increasing, since it has polynomial growth. We know for some $N > 0$, for any Schwartz ϕ,

$$|\Lambda(\phi)| \lesssim \sup\{|x^\alpha||D^\beta\phi| : |\alpha|, |\beta| \leqslant N, k > 0\}$$

TODO: FINSIH THIS. Because $\Lambda * \phi$ is tempered, this means that we can consider the Fourier transform $\widehat{\Lambda * \phi}$. If ψ is compactly supported, then

$$(\Lambda * \phi)^\wedge \left(\widehat{\psi}\right) = s$$

TODO: FINISH, which proves $\widehat{\Lambda * \phi} = \widehat{\Lambda}\widehat{\phi}$.

3.5 Convolution Operators

It is know that if $T : D(\mathbf{R}^n) \to C^\infty(\mathbf{R}^n)$ is any continuous linear functional commuting with translations, it is given by convolution from some distribution. If this convolution is with respect to some tempered distribution, then the transformation extends from a map from $\mathcal{S}(\mathbf{R}^n)$ to $C^\infty(\mathbf{R}^n)$.

Studying the class of operators which commute with translations is very important because these operators occur again and again in Harmonic analysis. To begin with, with rely on a regularity result on the differentiation of functions in L^p spaces.

Lemma 3.11. *If $f \in L^p(\mathbf{R}^n)$, has derivatives in the L^p norm of all orders $\leqslant n + 1$, then f is almost everywhere equal to a continuous function g such that*

$$|g(0)| \lesssim \sum_{|\alpha| \leqslant n+1} \|D^\alpha f\|_p$$

where the hidden constant depends only on n and p.

Proof. s $\qquad\qquad\square$

Theorem 3.12. *If $T : L^p(\mathbf{R}^n) \to L^q(\mathbf{R}^n)$ is bounded, linear, and commutes with translations, then there exists a unique tempered distribution Λ such that $T(\phi) = \Lambda * \phi$ for all $\phi \in \mathcal{S}(\mathbf{R}^n)$*

Proof. If T commutes with translations, then for any Schwartz function ϕ, $T\phi$ has derivatives in the L^q norm of all orders, since $\Delta_{h,e}(T\phi) = T(\Delta_{h,e}\phi)$, and $\Delta_{h,e}\phi$ converges to $D_e\phi$ in the L^q norm, since the L^q norm is continuous in Schwartz space. In particular, we find $D^\alpha(T\phi) = T(D^\alpha\phi)$. Thus $T\phi$ is equal to a continuous function g_ϕ with

$$|g_\phi(0)| \lesssim \sum_{|\alpha| \leqslant n+1} \|D^\alpha(T\phi)\|_q = \sum_{|\alpha| \leqslant n+1} \|T(D^\alpha\phi)\|_q \leqslant \|T\| \sum_{|\alpha| \leqslant n+1} \|D^\alpha\phi\|_q$$

The map $\phi \mapsto g_\phi(0)$ is therefore continuous on $\mathcal{S}(\mathbf{R}^n)$, and therefore defines a tempered distribution Λ, and the fact that $T(\phi) = \Lambda * \phi$ then holds by the translation invariance of T. $\qquad\square$

Remark. It therefore follows that if $T : L^p(\mathbf{R}^n) \to L^q(\mathbf{R}^n)$ is bounded, linear, and commutes with translations, then for any Schwartz function ϕ, $T\phi$ is C^∞, and is slowly increasing, as is all of it's derivatives.

For each p and q, we will let (L^p, L^q) denote the space of tempered distributions which define a continuous linear map from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$, in the sense that the map $\phi \mapsto \Lambda * \phi$ is continuous as a map from $\mathcal{S}(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$, and, by the Hahn-Banach theorem, extends uniquely to a linear map on the whole space. In general, a characterization of such distributions is unknown except in a few situations.

Example. *The distributions in (L^2, L^2) are Fourier transforms of elements of $L^\infty(\mathbf{R}^n)$. The L^∞ norm of the element corresponds to the norm of the convolution operator. To see this, if Λ is a distribution, and Φ is the Gaussian distribution, $\Phi(x) = e^{-\pi|x|^2}$, then $\Lambda * \Phi$ is an L^2 function, since Φ is in L^2, and as such we conclude by the Plancherel theorem that $(\Lambda * \Phi)^\wedge = \Phi \Lambda^\wedge$ is an element of $L^2(\mathbf{R}^n)$. Thus we can think of $\widehat{\Lambda} = e^{\pi|x|^2}(\Lambda * \Phi)^\wedge$ as a function f. Plancherel's theorem implies that for any Schwarz function ϕ,*

$$\|f\widehat{\phi}\|_2 = \|\Lambda * \phi\|_2 \lesssim \|\phi\|_2 = \|\widehat{\phi}\|_2$$

But this means that $f \in L^\infty(\mathbf{R}^n)$, for if there is a set E of positive measure where $|f| \geqslant M$, we can find $\widehat{\phi}$ with $\widehat{\phi} = 1$ on E and with $\|\widehat{\phi}\|_2 = |E| + \varepsilon$, and then $\|f\widehat{\phi}\|_2 \geqslant M\|\widehat{\phi}\|_2$.

Example. *The distributions in (L^1, L^1) are precisely the finite Borel measures. The total variation of the measure corresponds to the norm of the convolution operator. It is clear that if μ is a Borel measure, then $\|\mu * \phi\|_1 \leqslant \|\mu\|_1\|\phi\|_1$. Conversely, if $\Lambda \in (L^1, L^1)$, and Φ_δ is the Gauss kernel, then we set $\Lambda_\delta = \Lambda * \Phi_\delta$. By assumption, Λ_δ is an L^1 function, and so Λ, $\|\Lambda_\delta\|_1 \lesssim \|\Phi_\delta\|_1 = 1$. This implies that the Λ_δ are uniformly bounded in L^1, so by the Banach Alaoglu theorem, since $L^1(\mathbf{R}^n)$ embeds itself in $M(\mathbf{R}^n)$, which is the dual of $C_0(\mathbf{R}^n)$, some subsequence of the Λ_δ converge weakly to some measure μ. We claim $\Lambda = \Lambda_\mu$. To prove this, fix some Schwartz function ϕ. If we let $\phi_\delta = \phi * \Phi_\delta$, then $D^\alpha \phi_\delta = (D^\alpha \phi) * \Phi_\delta$ converges uniformly to $D^\alpha \phi$, so ϕ_δ converges to ϕ in $\mathcal{S}(\mathbf{R}^n)$, and so $\Lambda(\phi)$ is the limit of $\Lambda(\phi_\delta)$. But*

$$\Lambda(\phi_\delta) = \Lambda(\Phi_\delta * \phi) = (\Lambda * (\Phi_\delta * \phi)^*)(0)$$
$$= ((\Lambda * \Phi_\delta) * \phi^*)(0)$$
$$= \Lambda_\delta(\phi)$$

and we know some subsequence converges to $\int \phi(x)d\mu(x)$. But we know that overall the values converge to $\Lambda(\phi)$, which implies

$$\Lambda(\phi) = \int \phi(x)d\mu(x)$$

Since ϕ was an arbitrary Schwartz function, we can now apply the density of $\mathcal{S}(\mathbf{R}^n)$ in $L^1(\mathbf{R}^n)$ to conclude that for any integrable function f,

$$\Lambda(f) = \int f(x)d\mu(x)$$

This classifies the (L^1, L^1) distributions.

We also have a duality theorem.

Theorem 3.13. *For any two (p, q), $(L^p, L^q) = (L^{q^*}, L^{p^*})$.*

Chapter 4

Applications

4.1   The Wirtinger Inequality on an Interval

Theorem 4.1. *Given* $f \in C^1[-\pi, \pi]$ *with* $\int_{-\pi}^{\pi} f(t)dt = 0$,

$$\int_{-\pi}^{\pi} |f(t)|^2 \leqslant \int_{-\pi}^{\pi} |f'(t)|^2$$

Proof. Consider the fourier series

$$f(t) \sim \sum a_n e^{nit} \quad f'(t) \sim \sum ina_n e^{nit}$$

Then $a_0 = 0$, and so

$$\int_{-\pi}^{\pi} |f(t)|^2 \, dt = 2\pi \sum |a_n|^2 \leqslant 2\pi \sum n^2 |a_n|^2 = \int_{-\pi}^{\pi} |f'(t)|^2 \, dt$$

equality holds here if and only if $a_i = 0$ for $i > 1$, in which case we find

$$f(t) = Ae^{nit} + \overline{A}e^{-nit} = B\cos(t) + C\sin(t)$$

for some constants $A \in \mathbf{C}$, $B, C \in \mathbf{R}$. $\qquad\square$

Corollary 4.2. *Given* $f \in C^1[a, b]$ *with* $\int_a^b f(t) \, dt = 0$,

$$\int_a^b |f(t)|^2 dt \leqslant \left(\frac{b-a}{\pi}\right)^2 \int_a^b |f'(t)|^2 \, dt$$

4.2 Energy Preservation in the String equation

Solutions to the string equation are
   If $u(t, x)$

4.3 Harmonic Functions

The study of a function f defined on the real line can often be understood by extending it's definition holomorphically to the complex plane. Here we will extend this tool, establishing that a large family of functions f defined on $\mathbf{R}^n$ can be understood by looking at a *harmonic* function on the upper half plane $\mathbf{H}^{n+1}$, which approximates f at it's boundary. This is a form of the Dirichlet problem, which asks, given a domain and a function on the domain's boundary, to find a function harmonic on the interior of the domain which 'agrees' with the function on the boundary, in one of several senses. As we saw in our study of harmonic functions on the disk in the study of Fourier series, we can study such harmonic functions by convolving f with an appropriate approximation to the identity which makes the function harmonic in the plane. In this case, we shall use the Poisson kernel for the upper half plane.

Theorem 4.3. *If $f \in L^p(\mathbf{R}^n)$, for $1 \leqslant p \leqslant \infty$, and $u(x, y) = (f * P_y)(x)$, where*

$$P_y(x) = \frac{\Gamma((n+1)/2)}{\pi^{(n+1)/2}} \frac{1}{(1 + |x|^2)^{(n+1)/2}}$$

then u is harmonic in the upper half plane, $u(x, y) \to f(x)$ for almost every x, and $u(\cdot, y)$ converges to f in L^p as $y \to 0$, with $\|u(\cdot, y)\|_p \leqslant \|f\|_p$. If, instead, f is a continuous and bounded function, then $u(\cdot, y)$ converges to f locally uniformly as $y \to 0$.

Proof. The almost everywhere convergence and convergence in norm follow from the fact that P_y is an approximation to the identity. The fact that u is harmonic follows because

$$\frac{\partial^2 u}{\partial x^2}(x, y) = (f * P_y'')(x) \qquad \frac{\partial^2 u}{\partial y^2} = (f *)$$

$\square$

Part II

Euclidean Harmonic Analysis

Here we employ the more modern techniques of functional analysis to understand the more advanced and technical portions of Fourier analysis, and related problems.

Chapter 5

Interpolation Theory

One of the most fundamental tools in the 'hard style' of mathematical analysis, involving explicit quantitative estimates on quantities that arises in basic methods of mathematics, is the theory of interpolation. The main goal of interpolation is to take two estimates, and blend them together to form a family of intermediate estimates. Often each estimate will focus on one component of the problem at hand (an estimate in terms of the decay of the function at ∞, an estimate involving the growth of the derivative, or the low frequency the function is, etc). By interpolating, we can optimize and obtain an estimate which simultaneously takes into account multiple features of the function.

The most basic way to interpolate is using the notion of convexity. Given two inequalities $A_0 \leqslant B_0$ and $A_1 \leqslant B_1$, for any parameter $0 \leqslant \theta \leqslant 1$, if we define the additive weighted averages

$$A_\theta = (1-\theta)A_0 + \theta A_1 \quad B_\theta = (1-\theta)B_0 + \theta B_1$$

Then we conclude $A_\theta \leqslant B_\theta$ for all θ. Similarily, we can consider the weighted multiplicative averages

$$A_\theta = A_0^{1-\theta} A_1^\theta \quad B_\theta = B_0^{1-\theta} B_1^\theta$$

in which case we still have $A_\theta \leqslant B_\theta$. Note that the additive averages are obtained by taking the unique linear function between two values, and the multiplicative averages are obtained by taking the unique log-linear function between two values. In particular, if A_θ is defined to be any convex function, then we have

$$A_\theta \leqslant (1-\theta)A_0 + \theta A_1$$

and if B_θ is logarithmically convex, so that $\log B_\theta$ is convex, then

$$B_\theta \leqslant B_0^{1-\theta} B_1^\theta$$

Thus convexity provides us with a more general way of interpolating estimates, which is what makes this property so useful in analysis, enabling us to simplify estimates.

Example. *For a fixed, measurable function f, the map $p \mapsto \|f\|_p$ is a log convex function. This statement is precisely Hölder's inequality, since the inequality*

$$\|f\|_{\theta p + (1-\theta)q} \leqslant \|f\|_p^\theta \|f\|_q^{1-\theta}$$

says

$$\left\| |f|^{\theta p} |f|^{(1-\theta)q} \right\|_1^{1/(\theta p + (1-\theta)q)} \leqslant \|f^{\theta p}\|_{1/\theta}^\theta \|f^{(1-\theta)q}\|_{1/(1-\theta)}^{1-\theta}$$

which is precisely Hölder's inequality. Note this implies that if $p_0 < p_\theta < p_1$, then $L^{p_0}(X) \cap L^{p_1}(X) \subset L^{p_\theta}(X)$.

Example. *For a fixed p, the* **weak L^p space** $L^{p,\infty}(X)$ *is the family of functions f for which the quantity*

$$\|f\|_{p,\infty} = \sup t F(t)^{1/p}$$

is finite, where $F(t) = ||f| \geqslant t|$. This weak L^p norm is log convex, because if $F(t) \leqslant A_0^{p_0}/t^{p_0}$, and $F(t) \leqslant A_1^{p_1}/t^{p_1}$, then we can apply scalar interpolation to conclude that if $p_\theta = (1-\alpha)p_0 + \alpha p_1$,

$$F(t) \leqslant \frac{A_0^{(1-\alpha)p_0} A_1^{\alpha p_1}}{t^{(1-\alpha)p_0 + \alpha p_1}} = \frac{A_\theta^{p_\theta}}{t^{p_\theta}}$$

where p_θ is the harmonic weighted average between p_0 and p_1, and A_θ the geometric weighted average. Using this argument, interpolating slightly to the left and right of p_θ, we can conclude that if $p_0 < p_\theta < p_1$, then $L^{p_0,\infty}(X) \cap L^{p_1,\infty}(X) \subset L^{p_\theta}(X)$.

Another major technique to perform an interpolation is to utilize the theory of complex analytic functions to obtain estimates. The core idea of this technique is to exploit the maximum principle, which says that bounding an analytic function at its boundary enables one to obtain bounds everywhere in the domain of the function. The next result, known as Lindelöf's theorem, is one of the fundamental examples of the application of complex analysis.

Theorem 5.1 (The Three Lines Lemma). *If f is a holomorphic function on the strip $S = \{s : \sigma \in [0,1]\}$ with the growth condition $|f(s)| \lesssim e^{e^{(\pi-\delta)|t|}}$, for some $\delta > 0$, then the function $\sigma \mapsto \sup f(\sigma + it)$ is log convex on $[0,1]$. In other words, if we have a bound $|f(it)| \leqslant B_0$ and $|f(1+it)| \leqslant B_1$, then we have a general estimate $|f(s)| \leqslant B_0^{1-\theta} B_1^{\theta}$.*

Proof. We can divide f by the holomorphic function $B_0^{1-s} B_1^s$, and since this new function satisfies the constraints of the theorem, it suffices to prove that if $|f(it)|, |f(1+it)| \leqslant 1$, then $|f(s)| \leqslant 1$ for all $s \in S$. If we assume first that $|f(s)| \to 0$ as $|t| \to \infty$, then, applying the maximum principle, we conclude that the maximum value on $\{s : |t| \leqslant T\}$ occurs for $\sigma \in \{0,1\}$, which gives us the bound for $|t| \leqslant T$. Taking $T \to \infty$ completes the proof. In general, if we define the function $g_\varepsilon(s) = \exp(\varepsilon i \exp(i[(\pi-\delta/2)s+\delta/4]))$, then the function $f(s)g_\varepsilon(s)g_\varepsilon(1-s) \to 0$ as $|t| \to \infty$, since

$$
\begin{aligned}
|g_\varepsilon(s)| &= \exp(-\varepsilon \mathrm{Im}(\exp(i[(\pi - \delta/2)s + \delta/4]))) \\
&= \exp(-\varepsilon e^{-(\pi-\delta/2)t} \sin(\delta/4 + (\pi - \delta/2)\sigma)) \\
&\leqslant \min\left(1, e^{-\varepsilon \sin(\delta/4)e^{-(\pi-\delta/2)t}}\right)
\end{aligned}
$$

Thus g_ε causes sharp decay for large values of t. $g_\varepsilon(1-s)$ then causes sharp decay for small values of t, so

$$
|g_\varepsilon(s)g_\varepsilon(1-s)| \leqslant e^{-\varepsilon \sin(\delta/4)e^{-(\pi-\delta/2)|t|}}
$$

We can thus apply the previous case to this function to conclude that $|f(s)g_\varepsilon(s)g_\varepsilon(1-s)| \leqslant 1$, and then taking $\varepsilon \to 0$ completes the claim. $\square$

Remark. The function $e^{-ie^{\pi i s}}$ shows that this theorem cannot be proven for all holomorphic functions. In particular, this means there is no family of holomorphic functions g_ε which decays faster than the functions constructed above as $|t| \to \infty$ and pointwise approximates the identity as $\varepsilon \to 0$.

Remark. Similar variants can be used to show that if f is a function on the upper half strip satisfying a growth condition, then the supremum over lines on the upper half strip is log convex. If f is a function on the annulus, we do not need to even restrict ourselves to a growth condition to conclude that the supremum over circles centred at the origin is log convex (a result known as the three circles lemma).

Example. *Here we show how we can use the three lines lemma to prove that the L^p norms are log convex. If $f = \sum a_n \chi_{E_n}$ is a simple function, then the function*

$$g(s) = \int |f|^s = \sum |a_n|^s |E_n|$$

is analytic in s, and satisfies the growth condition of the three lines lemma because each term of the sum is exponential in growth. Since $|g(s)| \leqslant |g(\sigma)|$, the three lines lemma implies that g is log convex on the real line. By normalizing the function f and the underlying measure, given p_0, p_1, we may assume $\|f\|_{p_0} = \|f\|_{p_1} = 1$, and it suffices to prove that $\|f\|_{p_\theta} \leqslant 1$ for all $p_\theta \in [p_0, p_1]$. But the log convexity of g guarantees this is true, since $|g(p)| = \|f\|_p^p$. A standard limiting argument then gives the inequality for all functions f.

As $p \to \infty$, the L^p norm excludes functions with large peaks, and as $p \to 0$, the L^p norm excludes functions with large tails. In particular, if $\|f\|_{p_0}, \|f\|_{p_1} < \infty$, then for any p_θ between p_0 and p_1,

$$\|f \chi_{|f| \leqslant 1}\|_{p_\theta}^{p_\theta} = \int_{|f| \leqslant 1} |f|^{p_\theta} \leqslant \int_{|f| \leqslant 1} |f|^{p_0} < \infty$$

$$\|f \chi_{|f| > 1}\|_{p_\theta}^{p_\theta} = \int_{|f| > 1} |f|^{p_\theta} \leqslant \int_{|f| > 1} |f|^{p_1} < \infty$$

so, applying the triangle inequality, we conclude that $\|f\|_{p_\theta} < \infty$. Furthermore, the inequalities show that we can split any function with finite L^{p_θ} norm into the sum of a function with finite L^{p_0} norm and another with finite L^{p_1} norm. In the case where $p_1 = \infty$, then clearly $f \chi_{|f| > 1}$ is bounded if f is bounded. This gives the inclusions

$$L^{p_0}(X) \cap L^{p_1}(X) \subset L^{p_\theta}(X) \subset L^{p_0}(X) + L^{p_1}(X)$$

For all p_θ.

5.1 Interpolation of Functions

5.2 Interpolation of Operators

In the theory of operator interpolation, we address the following situation. We are given a norm space X, which can be written as the sum of two other

norm spaces $Y+Z$, which perhaps have a different norm to X. We are given a (not necessarily bounded operator) $T : X + Y \to Z$. Given that $\|T\|_Y$ and $\|T\|_Z$ are finite, what tools enable us to conclude that $\|T\|_X$ is finite? In other cases, we are given a norm space X with a family of norms $\|\cdot\|_\alpha$, and given the boundedness of $T : X \to Y$ in $\|\cdot\|_{p_0}$ and $\|\cdot\|_{p_1}$, we want to obtain boundedness in $\|\cdot\|_{p_\theta}$ for all p_θ inbetween p_0 and p_1.

Theorem 5.2 (Riesz-Thorin). *Let $p_0, p_1 \in (0, \infty]$ and $q_0, q_1 \in [1, \infty]$. Consider a linear operator T mapping elements of $L^{p_0}(X)$ to $L^{q_0}(Y)$, and elements of $L^{p_1}(X)$ to $L^{q_1}(Y)$. Then for any $f \in L^{p_\theta}(X)$, where p_θ is the weighted harmonic average $1/p_\theta = (1-\theta)/p_0 + \theta/p_1$, $Tf \in L^{q_\theta}(Y)$, where q is a weighted harmonic average, and*

$$\|T\|_{p_\theta \to q_\theta} \leqslant \|T\|_{p_0 \to q_0}^{1-\theta} \|T\|_{p_1 \to q_1}^{\theta}$$

Proof. If $p_0 = p_1$, the proof follows by the log convexity of the L^p norms of a function. Thus we may assume $p_0 \neq p_1$, so p_θ is finite in any case of interest. By normalizing the measures on both spaces, we may assume $\|T\|_{p_0, q_0} = \|T\|_{p_1, q_1} = 1$, so it suffices to prove $\|T\|_{p_\theta, q_\theta} \leqslant 1$. Hölder's inequality implies

$$\left| \int_Y (Tf) g \right| \leqslant \|f\|_{p_0} \|g\|_{q_0'}, \|f\|_{p_1} \|g\|_{q_1'}$$

We now employ some complex interpolation to conclude that whenever f and g are simple functions,

$$\left| \int_Y (Tf) g \right| \leqslant \|f\|_{p_\theta} \|g\|_{q_\theta'}$$

To see this, we can normalize so $\|f\|_{p_\theta} = \|g\|_{q_\theta'} = 1$, and write $f = |f| \mathrm{sgn}(f)$. But if we now define

$$F(s) = \int_Y T(|f|^{(1-s)p_\theta/p_0 + sp_\theta/p_1} \mathrm{sgn}(f)) |g|^{(1-s)q_\theta'/q_0' + sq_\theta'/q_1'} \mathrm{sgn}(g)$$

Then an application of Lindelöf's theorem implies the result. Since g and f can be taken to approximate all elements of $L^{p_\theta}(X)$ and $L^{q_\theta'}(Y)$, we conclude that the theorem is true for all not-necessarily simple elements of these function spaces. But applying duality, this now implies thath $\|Tf\|_{q_\theta} \leqslant \|f\|_{p_\theta}$, which was what was required to be proven. $\qquad\square$

Given a linear operator T mapping simple functions on X to measurable functions on Y, we say T has *strong type* (p,q) if $\|Tf\|_q \leq \|f\|_p$ for any simple function f. The Hahn-Banach theorem then says T has an extension to a continuous operator from $L^p(X)$ to $L^q(Y)$, and, provided p is finite, or X has finite measure, this extension will be unique. If we consider the *strong type* diagram to be all points (p,q) where T is of strong type $(1/p, 1/q)$, then the Riesz-Thorin interpolation theorem implies that the intersection of the strong type diagram with $[0,\infty) \times [0,1]$ is convex, and the operator norm $\|T\|_{p,q}$ is a log convex function of $(1/p, 1/q)$.

Example. *If T is the identity map on functions on $[0,1]$, then because of the fact that $\|f\|_p \leq \|f\|_q$ for $p < q$,*

Example. *We know that if f and g are integrable functions, then $f * g$ is an integrable function, and $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$. If we apply Minkowski's integral inequality, then we find that if $f \in L^1(\mathbf{R}^n)$ and $g \in L^q(\mathbf{R}^n)$, then*

$$\|f * g\|_q = \left(\int |(f * g)(x)|^q \, dx \right)^{1/q} \leq \int \left(\int |f(y)g(x-y)|^q dx \right)^{1/q} dy$$

$$= \int |f(y)| \|g\|_q = \|f\|_1 \|g\|_q$$

This means $f * g \in L^q(\mathbf{R}^n)$, and so is, in particular, defined everywhere. Hölder's inequality implies that if $f \in L^q(\mathbf{R}^n)$ and $g \in L^{q^*}(\mathbf{R}^n)$, then

$$\left| \int f(y)g(x-y) \, dy \right| \leq \int |f(y-x)||g(x)| \leq \int |f(y)| \, dy \int |g(y)|^{q^*} \, dy = \|f\|_q \|g\|_{q^*}$$

Thus $\|f * g\|_\infty \leq \|f\|_q \|g\|_{q^*}$. But this means that the convolution operator $T_g : f \mapsto f * g$ is $(1,q)$, as well as (q^*, ∞) continuous. But now, connecting these two points, and applying the Riesz-Thorin theorem, we conclude that if $f \in L^p(\mathbf{R}^n)$, with $1/p = \theta/q^* + (1-\theta) = 1/q^* + (1-\theta)/q = 1 - 1/q + 1/r$, $(f * g) \in L^r(\mathbf{R}^n)$, and

$$\|f * g\|_r \leq \|f\|_p \|g\|_q$$

Note that we never used anything about $\mathbf{R}^n$ here other than it's translational structure, and as such Young's inequality continues to apply in the theory of any modular locally compact group. Indeed, this means if we swap μ with $\lambda\mu$, then we obtain that

$$\|f * g\|_r = \lambda^{1+1/r} = \lambda^{1/p+1/q} \|f\|_p \|g\|_p$$

which is a good way of remembering that we must have $1 + 1/r = 1/p + 1/q$.

Elias Stein noticed that the Riesz-Thorin theorem can be easily extended to allow bounds over an 'analytic family' of operators T_s, where $s \in S$, in the sense that for any two test functions f and g, $\langle T_s f, g \rangle$ is analytic in s, growing slower than doubly exponentially, and we have estimates

$$\|T_{it}f\|_{q_0} \leqslant B_t \|f\|_{p_0} \qquad \|T_{1+it}f\|_{q_1} \leqslant B_t \|f\|_{p_1}$$

where $B_t = O(e^{e^{(\pi-\delta)t}})$, then $\|T_\theta f\|_{q_\theta} \leqslant B_\theta \|f\|_{p_\theta}$.

Chapter 6

Oscillatory Integrals

The study of oscillatory integrals provides tools enabling one to obtain asymptotics for integrals of the form

$$I(\lambda) = \int e^{i\lambda\Phi(x)}\psi(x)\,dx$$

where Φ is the **phase**, and ψ is the amplitude, which are both fixed parameters, and λ is allowed to vary. Determining the asymptotics of these integrals as $\lambda \to \infty$ requires control over the cancelling properties of the integral, so certain decomposition techniques often employed, i.e. in the theory of singular integrals, fail completely here.

Example. *The most basic example, and most important, example of an oscillatory integral is the Fourier transform*

$$\widehat{f}(\xi) = \int e^{-2\pi i \xi \cdot x} f(x)\,dx$$

More generally, we can obtain another example by taking a finite measure μ, and considering it's Fourier transform

$$\widehat{\mu}(\xi) = \int e^{-2\pi i \xi \cdot x} d\mu(x)\,dx$$

One of the main applications of the theory of oscillatory integrals is to show that the Fourier transforms of certain measures have fast decay properties.

6.1 The Principle of Stationary Phase

6.1.1 Localization

The basic principle of the theory of stationary phase is that the asymptotics of the integral are determined by where $\nabla\Phi$ vanishes. One of the principles of oscillatory integrals is that the asymptotics of the theory depend only on the local properties of the function Φ.

Theorem 6.1. *If Φ and ψ are smooth, compactly supported functions with $\nabla\Phi(x) \neq 0$ for all x in the support of ψ, then $I(\lambda) \lesssim 1/\lambda^N$.*

Proof. Set $a = (\nabla\Phi)/|\nabla\Phi|^2$. Then for any smooth functions f and g, integrating by parts, we obtain that

$$\int \frac{a \cdot \nabla f(x)}{i\lambda} g(x)\, dx = -\int f(x) \frac{(\nabla \cdot (ag))(x)}{i\lambda}\, dx$$

Set

$$D(f) = (i\lambda)^{-1}(a \cdot \nabla f) \quad D^*(f) = -(i\lambda)^{-1}(\nabla \cdot (af))$$

Note that $D(e^{i\lambda\Phi}) = e^{i\lambda\Phi}$, so $D^N(e^{i\lambda\Phi}) = e^{i\lambda\Phi}$ for all integers N, and so

$$I(\lambda) = \int D^N(e^{i\lambda\Phi})\phi = \int e^{i\lambda\Phi}(D^*)^N(\psi)$$

Taking absolute values in the last integral gives that

$$|I(\lambda)| \leqslant \int |(D^*)^N(\psi)| \lesssim \frac{1}{\lambda^N}$$

which gives the required bounds. $\qquad\square$

Of course, if Φ changes suitably rapidly around a point x, in the sense that $\nabla\Phi$ is nonsingular, then as we increase λ, the oscillatory factor in the integral is allowed to oscillate at a fast enough rate that ψ is effectively constant, and so the integral has so much cancellation that we get rapid decay. Note, however, that this depends on ψ being effectively constant, i.e. smooth. If ψ and Φ are only C^N functions, then we can only get a $|\lambda|^{-N}$ decay rate. A particularly revealing example is where we take the Fourier

transform of the characteristic function of an interval $[a, b]$ (smooth, albeit at the two endpoints where we get a sharp jump), where

$$\int_a^b e^{-2\pi i \lambda x}\, dx = \frac{e^{2\pi i \lambda b} - e^{2\pi i \lambda a}}{2\pi i \lambda}$$

which has only a rate $1/|\lambda|$ decay. Note, however, if we are taking an oscillatory integral 'on an interval' $[a, b]$, where ψ and Φ are both C^N, and for all $n \leqslant N$, $\Phi^{(n)}(a) = \Phi^{(n)}(b)$ and $\psi^{(n)}(a) = \psi^{(n)}(b)$, then we can still get $1/|\lambda|^N$ decay using the same proof as above, except now the integration by parts must take into account the endpoints of the integral, which now cancel to a suitably high degree.

6.2   Scaling

To hint at the multidimensional theory, we now focus solely on the single-variable theory. This simplifies the situation considerably, and we shall find there is an essentially complete theory of such oscillatory integrals in one dimension. Note that we now have

$$D(f) = (i\lambda \phi')^{-1} f' \quad D^*(f) = -(i\lambda)^{-1}(f/\phi)'$$

In particular, we attempt to guess the asymptotic development of the oscillatory integral

$$\int e^{i\lambda \phi(x)} \psi(x)$$

where the derivatives of ϕ may vanish at suitable points, yet for a suitably high n, $\phi^{(n)}$ does not vanish on the support of ψ. In particular, suppose that we want to find the best constant α, such that there exists a constant C_n such that for any ϕ and interval $[a, b]$ such that $|\phi^{(n)}(x)| \geqslant 1$ on (a, b), then

$$\left| \int_a^b e^{i\lambda \phi(x)}\, dx \right| \leqslant C_n/|\lambda|^{\alpha}$$

If $\phi(x) = x^n$, then the change of variables $y = \beta x$ implies that if we have a best constant α which works for all ϕ, then we must have $\alpha = 1/k$. Indeed, this estimate, known to Van der Corput, says this result is true.

73

Theorem 6.2. *There exists a constant C_n such that if ϕ is smooth in (a,b) with $|\phi^{(n)} \geqslant 1|$ for all $x \in (a,b)$, then*

$$\left| \int_a^b e^{i\lambda\phi(x)} \right| \leqslant C_n \lambda^{-1/n}$$

where $n \geqslant 2$, or $n = 1$, under the extra assumptions that ϕ' is monotonic.

Proof. Consider first $n = 1$. Then, using the operator D, we have

$$\int_a^b e^{i\lambda\phi}\, dx = \int_a^b D(e^{i\lambda\phi})\, dx = \int_a^b e^{i\lambda\phi} D^*(1)\, dx + \frac{e^{i\lambda b}/\phi'(b) - e^{i\lambda a}/\phi'(a)}{i\lambda}$$

The boundary terms are collectively bounded by $2/|\lambda|$, and we can bound the integration term, using the monotonicity of ϕ', by

$$\left| \int_a^b e^{i\lambda\phi} D^*(1)\, dx \right| \leqslant |\lambda|^{-1} \int_a^b |(1/\phi')'|\, dx \leqslant |\lambda|^{-1} \left| \int_a^b (1/\phi')'\, dx \right|$$

$$= |\lambda|^{-1} \left| \frac{1}{\phi'(b)} - \frac{1}{\phi'(a)} \right| \leqslant |\lambda|^{-1}$$

so we can set $C_1 = 3$. We now prove the remaining inequalities by induction, using an integration by parts. Suppose that (by replacing ϕ with its negation if necessary) $\phi^{(n+1)}(x) \geqslant 1$ on $[a,b]$. Let x_0 be the point at which $|\phi^{(n)}(x)|$ is minimized. Without loss of generality, we may assume that $|\phi^{(n)}(x)$ TODO: FINISH THIS ARGUMENT, GIVES $C_n = 5 \cdot 2^{n-1} - 2$. $\qquad \square$

Remark. If $|\phi^{(n)}(x)| \geqslant \mu$, then $|\phi^{(n)}(x)/\mu| \geqslant 1$, and so substituting into the previous result establishes that $|I(\lambda)| \leqslant C_m/|\mu\lambda|^{1/n}$.

Remark. If $n = 1$, and ϕ' is not monotonic, we can choose ϕ to grow suitably slowly on intervals of the form $[0, \pi] + 2\pi n$, and ϕ to grow much faster on intervals of the form $[\pi, 2\pi] + 2\pi n$. It then follows that $\int_0^{2\pi N} \sin(\phi(x))$ is unbounded as we let $N \to \infty$, which prevents us from extending the result completely to the one dimensional case.

Continuing in the simpler, one dimensional case, we now consider a *nondegenerate* critical point x, where $\Phi'(x) = 0$, but $\Phi''(x) \neq 0$. A good instance of this occurs where $\Phi(x) = x^2$, where we find

$$\int e^{i\lambda x^2} \psi(x)\, dx = \sum_{k=0}^{N} C_k \lambda^{-1/2-k} + O(|\lambda|^{-3/2-N})$$

74

This is obtained by noting that the Fourier transform of the Gaussian implies

$$\int e^{-sx^2} \psi(x)\, dx = (\pi/s)^{1/2} \int e^{-\pi^2 \xi^2/s} \widehat{\psi}(\xi)\, d\xi$$

Since both sides are analytic, and they make sense for $\mathfrak{Re}(s) > 0$, we can take $s \to -i\lambda$ to conclude that

$$\int e^{i\lambda x^2} \psi(x)\, dx = \left(\frac{\pi}{i\lambda}\right)^{1/2} \int e^{-i\pi^2 \xi^2/\lambda} \widehat{\psi}(x)\, dx$$

Expanding the exponential e^{iu^2} gives the required bounds. Thus we can expect a critical, nondegenerate point to give a $O(\lambda^{-1/2})$ error bound.

Corollary 6.3. *If an amplitude ψ is present, then*

$$\left| \int_a^b e^{i\lambda\Phi(x)} \psi(x)\, dx \right| \leqslant 8 \left(\int_a^b |\psi'(x)|\, dx + |\psi(b)| \right) \lambda^{-1/2}$$

Proof. Integrating by parts, if $J(x) = \int_a^x e^{i\lambda\Phi(u)}\, du$, then

$$\int_a^b e^{i\lambda\Phi(x)} \psi(x)\, dx = J(b)\psi(b) - \int_a^b J(x)\psi'(x)\, dx$$

and then since $|J(x)| \leqslant 8\lambda^{-1/2}$, the proposition follows. TODO: ADDRESS MULTIDIMENSIONAL CASE. $\qquad\square$

Example. *The Bessel functions*

$$J_m(r) = \frac{1}{2\pi} \int_0^{2\pi} e^{ir\sin(x)} e^{-imx}\, dx$$

occur naturally in many areas of analysis. The definition of the functions can be seen as an oscillatory integral, with $\lambda = r$, $\Phi(x) = \sin(x)$, and $\psi(x) = e^{-imx}/2\pi$. Split $[0, 2\pi]$ into two intervals, the first upon which $\cos(x) \geqslant 1/\sqrt{2}$, the second where $\sin(x) \geqslant 1/\sqrt{2}$. On the first part, we may apply the corollary above to obtain a $O(r^{-1/2})$ bound, and on the second, we can apply the first to obtain a $O(r^{-1})$ bound. Summing these two bounds up gives the theorem.

6.3 Surface Carried Measures

Theorem 6.4. *If a hypersurface Σ has non-vanishing Gauss curvature at each point in the support of a surface carried measure μ, then*

$$|\widehat{\mu}(\xi)| = O(|\xi|^{-(d-1)/2})$$

s

6.4 Restriction Theorems

If $f \in L^p(\mathbf{R}^n)$, then the Hausdorff Young theorem says that $\widehat{f}$ is a function in $L^q(\mathbf{R}^n)$, where q is the dual of p. If $f \in L^1(\mathbf{R}^n)$, then $\widehat{f}$ is actually continuous, so you can meaningfully discuss the behaviour of the Fourier transform when restricted to low dimensional hypersurfaces, for instance, on a sphere of a fixed radius. However, in general $\widehat{f}$ will only be defined almost everywhere, and so it is unclear whether one can form a well defined restriction of the Fourier transform.

The general situation is as follows. If μ is a measure carried on a compact surface M, for a fixed p, does there exist an estimate

$$\|\widehat{f}\|_{L^q(M,\mu)} \lesssim \|f\|_{L^p(\mathbf{R}^n)}$$

for Schwartz functions f. If this is true, we can apply a density argument to show that the restriction operator $R(f) = \widehat{f}|_M$ uniquely extends to a well defined continuous linear operator from $L^p(\mathbf{R}^n)$ to $L^q(M,\mu)$.

We begin by determining a duality result to the restriction calculation. Assuming our functions are suitably regular, we calculate

$$\int_M (Rf)(\xi)\overline{g(\xi)}\, d\mu(\xi) = \int_M \left(\int_{\mathbf{R}^n} f(x) e^{-2\pi i \xi \cdot x}\, dx \right) \overline{g(\xi)}\, d\mu(\xi)$$

$$= \int_{\mathbf{R}^n} f(x) \overline{\int_M g(\xi) e^{2\pi i \xi \cdot x}\, d\mu(\xi)}\, dx$$

which implies the formal adjoint of the map R is the **extension operator**

$$(R^* f)(x) = \int_M e^{2\pi i \xi \cdot x} f(\xi) d\mu(\xi)$$

which extends a function in frequency space supported on M to a function on the entirety of phase space. By duality properties, R is continuous as an operator from $L^p(\mathbf{R}^n)$ to $L^q(M,\mu)$ if and only if R^* is continuous as an operator from $L^{q^*}(M,\mu)$ to $L^{p^*}(\mathbf{R}^n)$. We also calculate

$$((R^*R)f)(x) = \int_{\mathbf{R}^n} \left(\int_M e^{2\pi i\xi\cdot(x-y)}\,d\mu(\xi) \right) f(y)\,dy = (f*\breve{\mu})(x)$$

So if R is $(p,2)$ continuous, R^* is $(2,p^*)$ continuous, and so R^*R is (p,p^*) continuous. Conversely, if we know that R^*R is (p,p^*) continuous, then we find that for $f \in L^p(\mathbf{R}^n)$, Hölder's inequality implies

$$\|Rf\|^2_{L^2(M,\mu)} = (Rf, Rf)_M = ((R^*R)f,f)_{\mathbf{R}^d} \leqslant \|R^*R\|_{p\to p^*}\|f\|_p^2$$

and so we conclude that $\|R\|_{p\to 2} \leqslant \sqrt{\|R^*R\|_{p\to p^*}}$.

We now prove that R is $(2n+2/n+3,2)$ continuous, assuming that M has non-zero Gaussian curvature at each point. The previous paragram implies that it suffices to show that it is enough to show that R^*R is (p,p^*) continuous, where $p = (2n+2)/(n+3)$ and $p^* = (2n+2)/(n-1)$. Since

$$(R^*R)(f) = f*\breve{\mu}$$

We shall verify this using Stein's interpolation theorem. Consider the family of kernels k_s, where $k_s = \widecheck{K_s}$, and $K_s = \gamma_s |x_n - \varphi(x')|_+^{s-1}\varphi_0(x)$, where $\gamma_s = s(s+1)\dots(s+N)e^{s^2}$

Part III

Abstract Harmonic Analysis

The main property of spaces where Fourier analysis applies is symmetry – for a function **R**, we can translate and negate. On **R**^{n} we have not only translational symmetry but also rotational symmetry. It turns out that we can apply Fourier analysis to any 'space with symmetry'. That is, functions on an Abelian group. We shall begin with the study of finite abelian groups, where convergence questions disappear, and with it much of the analytical questions involved in the theory. We then proceed to generalize to a study of infinite abelian groups with topological structure.

Chapter 7

Finite Character Theory

Let us review our achievements so far. We have found several important families of functions on the spaces we have studied, and shown they can be used to approximate arbitrary functions. On the circle group $\mathbf{T}$, the functions take the form of the power maps $\phi_n : z \mapsto z^n$, for $n \in \mathbf{Z}$. The important properties of these functions is that

- The functions are orthogonal to one another.

- A large family of functions can be approximated by linear combinations of the power maps.

- The power maps are multiplicative: $\phi_n(zw) = \phi_n(z)\phi_n(w)$.

The existence of a family with these properties is not dependant on much more than the symmetry properties of $\mathbf{T}$, and we can therefore generalize the properties of the fourier series to a large number of groups. In this chapter, we consider a generalization to any finite abelian group.

The last property of the power maps should be immediately recognizable to any student of group theory. It implies the exponentials are homomorphisms from the circle group to itself. This is the easiest of the three properties to generalize to arbitrary groups; we shall call a homomorphism from a finite abelian group to $\mathbf{T}$ a **character**. For any abelian group G, we can put all characters together to form the character group $\Gamma(G)$, which forms an abelian group under pointwise multiplication $(fg)(z) = f(z)g(z)$. It is these functions which are 'primitive' in synthesizing functions defined on the group.

Example. *If μ_N is the set of Nth roots of unity, then $\Gamma(\mu_N)$ consists of the power maps $\phi_n : z \mapsto z^n$, for $n \in \mathbf{Z}$. Because*

$$\phi(\omega)^N = \phi(\omega^N) = \phi(1) = 1$$

we see that any character on μ_N is really a homomorphism from μ_N to μ_N. Since the homomorphisms on μ_N are determined by their action on this primitive root, there can only be at most N characters on μ_N, since there are only N elements in μ_N. Our derivation then shows us that the ϕ_N enumerate all such characters, which completes our proof. Note that since $\phi_n \phi_m = \phi_{n+m}$, and $\phi_n = \phi_m$ if and only if $n - m$ is divisible by N, this also shows that $\Gamma(\mu_N) \cong \mu_N$.

Example. *The group $\mathbf{Z}_N$ is isomorphic to μ_N under the identification $n \mapsto \omega^n$, where ω is a primitive root of unity. This means that we do not need to distinguish functions 'defined in terms of n' and 'defined in terms of ω', assuming the correspondance $n = \omega^n$. This is exactly the same as the correspondence between functions on $\mathbf{T}$ and periodic functions on $\mathbf{R}$. The characters of $\mathbf{Z}_n$ are then exactly the maps $n \mapsto \omega^{kn}$. This follows from the general fact that if $f : G \to H$ is an isomorphism of abelian groups, the map $f^* : \phi \mapsto \phi \circ f$ is an isomorphism from $\Gamma(H)$ to $\Gamma(G)$.*

Example. *If K is a finite field, then the set K^* of non-zero elements is a group under multiplication. A rather sneaky algebraic proof shows the existence of elements of K, known as primitive elements, which generate the multiplicative group of all numbers. Thus K is cyclic, and therefore isomorphic to μ_N, where $N = |K| - 1$. The characters of K are then easily found under the correspondence.*

Example. *For a fixed N, the set of invertible elements of $\mathbf{Z}_N$ form a group under multiplication, denoted $\mathbf{Z}_N^*$. Any character from $\mathbf{Z}_N^*$ is valued on the $\varphi(N)$'th roots of unity, because the order of each element in $\mathbf{Z}_N^*$ divides $\varphi(N)$. The groups are in general non-cyclic. For instance, $\mathbf{Z}_8^* \cong \mathbf{Z}_2^3$. However, we can always break down a finite abelian group into cyclic subgroups to calculate the character group; a simple argument shows that $\Gamma(G \times H) \cong \Gamma(G) \times \Gamma(H)$, where we identify (f, g) with the map $(x, y) \mapsto f(x)g(y)$.*

7.1 Fourier Analysis on Cyclic Groups

We shall start our study of abstract Fourier analysis by looking at Fourier analysis on μ_N. Geometrically, these points uniformly distribute them-

selves over $\mathbf{T}$, and therefore μ_N provides a good finite approximation to $\mathbf{T}$. Functions from μ_N to $\mathbf{C}$ are really just functions from $[n] = \{1,\dots,n\}$ to $\mathbf{C}$, and since μ_N is isomorphic to $\mathbf{Z}_N$, we're really computing the Fourier analysis of finite domain functions, in a way which encodes the translational symmetry of the function relative to translational shifts on $\mathbf{Z}_N$.

There is a trick which we can use to obtain quick results about Fourier analysis on μ_N. Given a function $f : [N] \to \mathbf{C}$, consider the N-periodic function on the real line defined by

$$g(t) = \sum_{n=1}^{N} f(n)\chi_{(n-1/2,n+1/2)}(t)$$

Classical Fourier analysis of g tells us that we can expand g as an infinite series in the functions $e(n/N)$, which may be summed up over equivalence classes modulo N to give a finite expansion of the function f. Thus we conclude that every function $f : [N] \to \mathbf{C}$ has an expansion

$$f(n) = \sum_{m=1}^{N} \widehat{f}(m)e(nm)$$

where $\widehat{f}(m)$ are the coefficients of the **finite Fourier transform** of f. This method certainly works in this case, but does not generalize to understand the expansion of general finite abelian groups.

The correct generalization of Fourier analysis is to analyze the set of complex valued 'square integrable functions' on the domain $[N]$. We consider the space V of all maps $f : [N] \to \mathbf{C}$, which can be made into an inner product space by defining

$$\langle f, g \rangle = \frac{1}{N} \sum_{n=1}^{N} f(n)\overline{g(n)}$$

We claim that the characters $\phi_n : z \mapsto z^n$ are orthonormal in this space, since

$$\langle \phi_n, \phi_m \rangle = \frac{1}{N} \sum_{k=1}^{N} \omega^{k(n-m)}$$

If $n = m$, we may sum up to find $\langle \phi_n, \phi_m \rangle = 1$. Otherwise we use a standard summation formula to find

$$\sum_{k=1}^{N} \omega^{k(n-m)} = \omega^{n-m}\frac{\omega^{N(n-m)} - 1}{\omega^{n-m} - 1}$$

82

Since $\omega^{N(n-m)} = 1$, we conclude the sum is zero. This implies that the ϕ_n are orthonormal, hence linearly independent. Since V is N dimensional, this implies that the family of characters forms an orthogonal basic for the space. Thus, for any function $f : [N] \to \mathbf{C}$, we have, if we set $\widehat{f}(m) = \langle f, \phi_m \rangle$, then

$$f(n) = \sum_{m=1}^{N} \langle f, \phi_m \rangle \phi_m(n) = \sum_{m=1}^{N} \widehat{f}(m) e(mn/N)$$

This calculation can essentially be applied to an arbitrary finite abelian group to obtain an expansion in terms of Fourier coefficients.

7.2   An Arbitrary Finite Abelian Group

It should be easy to guess how we proceed for a general finite abelian group. Given some group G, we study the character group $\Gamma(G)$, and how $\Gamma(G)$ represents general functions from G to $\mathbf{C}$. We shall let V be the space of all such functions from G to $\mathbf{C}$, and on it we define the inner product

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

If there's any justice in the world, these characters would also form an orthonormal basis.

Theorem 7.1. *The set $\Gamma(G)$ of characters is an orthonormal set.*

Proof. If e is a character of G, then $|e(a)| = 1$ for each a, and so

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

If $e \neq 1$ is a non-trivial character, then $\sum_{a \in G} e(a) = 0$. To see this, note that for any $b \in G$, the map $a \mapsto ba$ is a bijection of G, and so

$$e(b) \sum_{a \in G} e(a) = \sum_{a \in G} e(ba) = \sum_{a \in G} e(a)$$

83

Implying either $e(b) = 1$, or $\sum_{a \in G} e(a) = 0$. If $e_1 \neq e_2$ are two characters, then

$$\langle e_1, e_2 \rangle = \frac{1}{|G|} \sum_{a \in G} \frac{e_1(a)}{e_2(a)} = 0$$

since e_1/e_2 is a nontrivial character. $\qquad\square$

Because elements of $\Gamma(G)$ are orthonormal, they are linearly independent over the space of functions on G, and we obtain a bound $|\Gamma(G)| \leqslant |G|$. All that remains is to show equality. This can be shown very simply by applying the structure theorem for finite abelian groups. First, note it is true for all cyclic groups. Second, note that if it is true for two groups G and H, it is true for $G \times H$, because

$$\Gamma(G \times H) \cong \Gamma(G) \times \Gamma(H)$$

since a finite abelian group is a finite product of cyclic groups, this proves the theorem. This seems almost like sweeping the algebra of the situation under the rug, however, so we will prove the statement only using elementary linear algebra. What's more, these linear algebraic techniques generalize to the theory of unitary representations in harmonic analysis over infinite groups.

Theorem 7.2. *Let $\{T_1, \dots, T_n\}$ be a family of commuting unitary matrices. Then there is a basis $v_1, \dots, v_m \in \mathbf{C}^m$ which are eigenvectors for each T_i.*

Proof. For $n = 1$, the theorem is the standard spectral theorem. For induction, suppose that the $T_1, \dots, T_{k-1}$ are simultaneously diagonalizable. Write

$$\mathbf{C}^m = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_l}$$

where λ_i are the eigenvalues of T_k, and V_{λ_i} are the corresponding eigenspaces. Then if $v \in V_{\lambda_i}$, and $j < k$,

$$T_k T_j v = T_j T_k v = \lambda_i T_j v$$

so $T_j(V_{\lambda_i}) = V_{\lambda_i}$. Now on each V_{λ_i}, we may apply the induction hypotheis to diagonalize the $T_1, \dots, T_{k-1}$. Putting this together, we simultaneously diagonalize $T_1, \dots, T_k$. $\qquad\square$

This theorem enables us to prove the character theory in a much simpler manner. Let V be the space of complex valued functions on G, and define, for $a \in G$, the map $(T_a f)(b) = f(ab)$. V has an orthonormal basic consisting of the $\chi_a(b) = N[a = b]$, for $a \in G$. In this basis, we comcpute $T_a \chi_b = \chi_{ba^{-1}}$, hence T_a is a permutation matrix with respect to this basis, hence unitary. The operators T_a commute, since $T_a T_b = T_{ab} = T_{ba} = T_b T_a$. Hence these operators can be simultaneously diagonalized. That is, there is a family $e_1, \ldots, e_n \in V$ and $\lambda_{an} \in \mathbf{T}$ such that for each $a \in G$, $T_a e_n = \lambda_{an} f_n$. We may assume $e_n(1) = 1$ for each n by normalizing. Then, for any $a \in G$, we have $f_n(a) = f_n(a \cdot 1) = \lambda_{an} f_n(1) = \lambda_{an}$, so for any $b \in G$, $f_n(ab) = \lambda_{an} f_n(b) = f_n(a) f_n(b)$. This shows each f_n is a character, completing the proof. We summarize our discussion in the following theorem.

Theorem 7.3. *Let G be a finite abelian group. Then $\Gamma(G) \cong G$, and forms an orthonormal basis for the space of complex valued functions on G. For any function $f : G \to \mathbf{C}$,*

$$f(a) = \sum_{e \in \Gamma(G)} \langle f, e \rangle\, e(a) = \sum_{e \in \Gamma(G)} \hat{f}(e) e(a) \quad \langle f, g \rangle = \frac{1}{|G|} \sum_{a \in G} f(a) \overline{g(a)}$$

In this context, we also have Parseval's theorem

$$\|f(a)\|^2 = \sum_{e \in \hat{G}} |\hat{f}(e)|^2 \quad \langle f, g \rangle = \sum_{e \in \hat{G}} \hat{f}(e) \overline{\hat{g}(e)}$$

7.3   Convolutions

There is a version of convolutions for finite functions, which is analogous to the convolutions on $\mathbf{R}$. Given two functions f, g on G, we define a function $f * g$ on G by setting

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

The mapping $b \mapsto ab^{-1}$ is a bijection of G, and so we also have

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

For $e \in \Gamma(G)$,

$$\widehat{f * g}(e) = \frac{1}{|G|} \sum_{a \in G} (f * g)(a)\overline{e(a)}$$

$$= \frac{1}{|G|^2} \sum_{a,b \in G} f(ab)g(b^{-1})\overline{e(a)}$$

The bijection $a \mapsto ab^{-1}$ shows that

$$\widehat{f * g}(e) = \frac{1}{|G|^2} \sum_{a,b} f(a)g(b^{-1})\overline{e(a)e(b^{-1})}$$

$$= \frac{1}{|G|} \left(\sum_a f(a)\overline{e(a)} \right) \frac{1}{|G|} \left(\sum_b g(b)\overline{e(b)} \right)$$

$$= \hat{f}(e)\hat{g}(e)$$

In the finite case we do not need approximations to the identity, for we have an identity for convolution. Define $D : G \to \mathbf{C}$ by

$$D(a) = \sum_{e \in \Gamma(G)} e(a)$$

We claim that $D(a) = |G|$ if $a = 1$, and $D(a) = 0$ otherwise. Note that since $|G| = |\Gamma(G)|$, the character space of $\Gamma(G)$ is isomorphic to G. Indeed, for each $a \in G$, we have the maps $\hat{a} : e \mapsto e(a)$, which is a character of $\Gamma(G)$. Suppose $e(a) = 1$ for all characters e. Then $e(a) = e(1)$ for all characters e, and for any function $f : G \to \mathbf{C}$, we have $f(a) = f(1)$, implying $a = 1$. Thus we obtain $|G|$ distinct maps $\hat{a}$, which therefore form the space of all characters. It therefore follows from a previous argument that if $a \neq 1$, then

$$\sum_{e \in \Gamma(G)} e(a) = 0$$

Now $f * D = f$, because

$$\hat{D}(e) = \frac{1}{|G|} \sum_{a \in G} D(a)\overline{e(a)} = \overline{e}(1) = 1$$

D is essentially the finite dimensional version of the Dirac delta function, since it has unit mass, and acts as the identity in convolution.

7.4 (∗) The Fast Fourier Transform

The main use of the fourier series on μ_n is to approximate the Fourier transform on $\mathbf{T}$, where we need to compute integrals explicitly. If we have a function $f \in L^1(\mathbf{T})$, then f may be approximated in $L^1(\mathbf{T})$ by step functions of the form

$$f_n(t) = \sum_{k=1}^n a_k \mathbf{I}(x \in (2\pi(k-1)/n, 2\pi k/n))$$

And then $\widehat{f_n} \to \widehat{f}$ uniformly. The Fourier transform of f_n is the same as the Fourier transform of the corresponding function $k \mapsto a_k$ on $\mathbf{Z}_n$, and thus we can approximate the Fourier transform on $\mathbf{T}$ by a discrete computation on $\mathbf{Z}_n$. Looking at the formula in the definition of the discrete transform, we find that we can compute the Fourier coefficients of a function $f : \mathbf{Z}_n \to \mathbf{C}$ in $O(n^2)$ addition and multiplication operations. It turns out that there is a much better method of computation which employs a divide and conquer approach, which works when n is a power of 2, reducing the calculation to $O(n \log n)$ multiplications.

To see this, consider a particular division in the group $\mathbf{Z}_{2n}$. Given $f : \mathbf{Z}_{2n} \to \mathbf{C}$, define two functions $g, h : \mathbf{Z}_n \to \mathbf{C}$, defined by $g(k) = f(2k)$, and $h(k) = f(2k+1)$. Then g and h encode all the information in f, and if $v = e^{\pi i/n}$ is the canonical generator of $\mathbf{Z}_{2n}$, we have

$$\hat{f}(m) = \frac{\hat{g}(m) + \hat{h}(m)v^m}{2}$$

Because

$$\frac{1}{2n} \sum_{k=1}^n \left(g(k)\omega^{-km} + h(m)\omega^{-km}v^m \right) = \frac{1}{2n} \sum_{k=1}^n f(2k)v^{-2km} + f(2k+1)v^{-(2k+1)m}$$

$$= \frac{1}{2n} \sum_{k=1}^{2n} f(k)v^{-km}$$

This is essentially a discrete analogue of the Poission summation formula, which we will generalize later when we study the harmonic analysis of abelian groups. If $H(m)$ is the number of operations needed to calculate the Fourier transform of a function on μ_{2^n} using the above recursive formula, then the above relation tells us $H(2m) = 2H(m) + 3(2m)$.

87

If $G(n) = H(2^n)$, then $G(n) = 2G(n-1) + 32^n$, and $G(0) = 1$, and it follows that

$$G(n) = 2^n + 3 \sum_{k=1}^{n} 2^k 2^{n-k} = 2^n(1 + 3n)$$

Hence for $m = 2^n$, we have $H(m) = m(1 + 3\log(m)) = O(m \log m)$. Similar techniques show that one can compute the inverse Fourier transform in $O(m \log m)$ operations (essentially by swapping the root v with v^{-1}).

7.5   Dirichlet's Theorem

We now apply the theory of Fourier series on finite abelian groups to prove Dirichlet's theorem.

Theorem 7.4. *If m and n are relatively prime, then the set*

$$\{m + kn : k \in \mathbf{N}\}$$

contains infinitely many prime numbers.

An exploration of this requries the Riemann-Zeta function, defined by

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

The function is defined on $(1, \infty)$, since for $s > 1$ the map $t \mapsto 1/t^s$ is decreasing, and so

$$\sum_{n=1}^{\infty} \frac{1}{n^s} \leqslant 1 + \int_1^{\infty} \frac{1}{t^s} = 1 + \lim_{n \to \infty} \frac{1}{s-1} \left[1 - 1/n^{s-1}\right] = 1 + \frac{1}{s-1}$$

The series converges uniformly on $[1 + \varepsilon, N]$ for any $\varepsilon > 0$, so ζ is continuous on $(1, \infty)$. As $t \to 1$, $\zeta(t) \to \infty$, because $n^s \to n$ for each n, and if for a fixed M we make s close enough to 1 such that $|n/n^s - 1| < 1/2$ for $1 \leqslant n \leqslant M$, then

$$\sum_{n=1}^{\infty} \frac{1}{n^s} \geqslant \sum_{n=1}^{M} \frac{1}{n^s} = \sum_{n=1}^{M} \frac{1}{n} \frac{n}{n^s} \geqslant \frac{1}{2} \sum_{n=1}^{M} \frac{1}{n}$$

Letting $M \to \infty$, we obtain that $\sum_{n=1}^{\infty} \frac{1}{n^s} \to \infty$ as $s \to 1$.

The Riemann-Zeta function is very good at giving us information about the prime integers, because it encodes much of the information about the prime numbers.

Theorem 7.5. *For any $s > 1$,*

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

Proof. The general idea is this – we may write

$$\prod_{p \text{ prime}} \frac{1}{1 - p^s} = \prod_{p \text{ prime}} (1 + 1/p^s + 1/p^{2s} + \dots)$$

If we expand this product out formally, enumating the primes to be $p_1, p_2, \dots$, we find

$$\prod_{p \leqslant n} (1 + 1/p^s + 1/p^{2s} + \dots) = \sum_{n_1, n_2, \dots = 0}^{\infty} \frac{1}{p_1^{n_1}}$$

$\square$

Chapter 8

Topological Groups

In abstract harmonic analysis, the main subject matter is the **topological group**, a group G equipped with a topology which makes the operation of multiplication and inversion continuous. In the mid 20th century, it was realized that basic Fourier analysis could be generalized to a large class of groups. The nicest generalization occurs over the locally compact groups, which simplifies the theory considerably.

Example. *There are a few groups we should keep in mind for intuition in the general topological group.*

- *The classical groups $\mathbf{R}^n$ and $\mathbf{T}^n$, from which Fourier analysis originated.*

- *The group μ of roots of unity, rational numbers $\mathbf{Q}$, and cyclic groups $\mathbf{Z}_n$.*

- *The matrix subgroups of the general linear group $GL(n)$.*

- *The product $\mathbf{T}^\omega$ of Torii, occurring in the study of Dirichlet series.*

- *The product $\mathbf{Z}_2^\omega$, which occurs in probability theory, and other contexts.*

- *The field of p-adic numbers $\mathbf{Q}_p$, which are the completion of $\mathbf{Q}$ with respect to the absolute value $|p^{-m}q|_p = p^m$.*

8.1 Basic Results

The topological structure of a topological group naturally possesses large amounts of symmetry, simplifying the spatial structure. For any topolog-

ical group, the maps

$$x \mapsto gx \qquad x \mapsto xg \qquad x \mapsto x^{-1}$$

are homeomorphisms. Thus if U is a neighbourhood of x, then gU is a neighbourhood of gx, Ug a neighbourhood of xg, and U^{-1} a neighbourhood of x^{-1}, and as we vary U through all neighbourhoods of x, we obtain all neighbourhoods of the other points. Understanding the topological structure at any point reduces to studying the neighbourhoods of the identity element of the group.

In topological group theory it is even more important than in basic group theory to discuss set multiplication. If U and V are subsets of a group, then we define

$$U^{-1} = \{x^{-1} : x \in U\} \qquad UV = \{xy : x \in U, y \in V\}$$

We let $V^2 = VV$, $V^3 = VVV$, and so on.

Theorem 8.1. *Let U and V be subsets of a topological group.*

 (i) *If U is open, then UV is open.*

 (ii) *If U is compact, and V closed, then UV is closed.*

(iii) *If U and V are connected, UV is connected.*

(iv) *If U and V are compact, then UV is compact.*

Proof. To see that (i) holds, we see that

$$UV = \bigcup_{x \in V} Ux$$

and each Ux is open. To see (ii), suppose $u_i v_i \to x$. Since U is compact, there is a subnet u_{i_k} converging to y. Then $y \in U$, and we find

$$v_{i_k} = u_{i_k}^{-1}(u_{i_k} v_{i_k}) \to y^{-1}x$$

Thus $y^{-1}x \in V$, and so $x = yy^{-1}x \in UV$. (iii) follows immediately from the continuity of multiplication, and the fact that $U \times V$ is connected, and (iv) follows from similar reasoning. $\square$

91

Example. *If U is merely closed, then (ii) need not hold. For instance, in* **R**, *take* $U = \alpha\mathbf{Z}$, *and* $V = \mathbf{Z}$, *where* α *is an irrational number. Then* $U + V = \alpha\mathbf{Z} + \mathbf{Z}$ *is dense in* **R**, *and is hense not closed.*

There are useful ways we can construct neighbourhoods under the group operations, which we list below.

Lemma 8.2. *Let U be a neighbourhood of the identity. Then*

(1) *There is an open V such that* $V^2 \subset U$.

(2) *There is an open V such that* $V^{-1} \subset U$.

(3) *For any* $x \in U$, *there is an open V such that* $xV \subset U$.

(4) *For any x, there is an open V such that* $xVx^{-1} \subset U$.

Proof. (1) follows simply from the continuity of multiplication, and (2) from the continuity of inversion. (3) is verified because $x^{-1}U$ is a neighbourhood of the origin, so if $V = x^{-1}U$, then $xV = U \subset U$. Finally (4) follows in a manner analogously to (3) because $x^{-1}Ux$ contains the origin. $\square$

If $\mathcal{U}$ is an open basis at the origin, then it is only a slight generalization to show that for any of the above situations, we can always select $V \in \mathcal{U}$. Conversely, suppose that $\mathcal{V}$ is a family of subsets of a (not yet topological) group G containing e such that (1), (2), (3), and (4) hold. Then the family $\mathcal{V}' = \{xV : V \in \mathcal{V}, x \in G\}$ forms a subbasis for a topology on G which forms a topological group. If $\mathcal{V}$ also has the base property, then $\mathcal{V}'$ is a basis.

Theorem 8.3. *If K and C are disjoint, K is compact, and C is closed, then there is a neighbourhood V of the origin for which KV and CV is disjoint. If G is locally compact, then we can select V such that KV is precompact.*

Proof. For each $x \in K$, C^c is an open neighbourhood containing x, so by applying the last lemma recursively we find that there is a symmetric neighbourhood V_x such that $xV_x^4 \subset C^c$. Since K is compact, finitely many of the xV_x cover K. If we then let V be the open set obtained by intersecting the finite subfamily of the V_x, then KV is disjoint from CV. $\square$

Taking K to be a point, we find that any open neighbourhood of a point contains a closed neighbourhood. Provided points are closed, we can set C to be a point as well.

Corollary 8.4. *Every Kolmogorov topological group is Hausdorff.*

Related to this theorem is the

Theorem 8.5. *For any set $A \subset G$,*

$$\overline{A} = \bigcap_V AV$$

Where V ranges over the set of neighbourhoods of the origin.

Proof. If $x \notin \overline{A}$, then the last theorem guarantees that there is V for which $\overline{A}V$ and Ax are disjoint. We conclude $\bigcap AV \subset \overline{A}$. Conversely, any neighbourhood contains a closed neighbourhood, so that $\overline{A} \subset AV$ for a fixed V, and hence $\overline{A} \subset \bigcap AV$. $\qquad\square$

Theorem 8.6. *Every open subgroup of G is closed.*

Proof. Let H be an open subgroup of G. Then

$$\overline{H} = \bigcap_V HV$$

If W is a neighbourhood of the origin contained in H, then we find

$$\overline{H} \subset HW \subset H$$

so H is closed. $\qquad\square$

We see that open subgroups of a group therefore correspond to connected components of the group, so that connected groups have no proper open subgroups. This also tells us that a locally compact group is σ-compact on each of its components, for if V is a pre-compact neighbourhood of the origin, then $V^2, V^3, \ldots$ are all precompact, and $\bigcup_{k=1}^{\infty} V^k$ is an open subgroup of G, which therefore contains the component of e, and is σ-compact. Since the topology of a topological group is homogenous, we can conclude that all components of the group are σ compact.

8.2 Quotient Groups

If G is a topological group, and H is a subgroup, then G/H can be given a topological structure in the obvious way. The quotient map is open, because VH is open in G for any open set V, and if H is normal, G/H is also a topological group, because multiplication is just induced from the quotient map of $G \times G$ to $G/H \times G/H$, and inversion from G to G/H. We should think the quotient structure is pleasant, but if no conditions on H are given, then G/H can have pathological structure. One particular example is the quotient $\mathbf{T}/\mu_\infty$ of the torus modulo the roots of unity, where the quotient is lumpy.

Theorem 8.7. *If H is closed, G/H is Hausdorff.*

Proof. If $x \neq y \in G/H$, then xHy^{-1} is a closed set in G, not containing e, so we may conclude there is a neighbourhood V for which V and $VxHy^{-1}$ are disjoint, so VyH and VxH are disjoint. This implies that the open sets $V(xH)$ and $V(yH)$ are disjoint in G/H. $\qquad\square$

Theorem 8.8. *If G is locally compact, G/H is also.*

Proof. If $\{U_i\}$ is a basis of precompact neighbourhoods at the origin, then U_iH is a family of precompact neighbourhoods of the origin in G/H, and is in fact a basis, for if V is any neighbourhood of the origin, there is $U_i \subset \pi^{-1}(V)$, and so $U_iH \subset V$. $\qquad\square$

If G is a non-Hausdorff group, then $\overline{\{e\}} \neq \{e\}$, and $G/\overline{\{e\}}$ is Hausdorff. Thus we can get away with assuming all our topological groups are Hausdorff, because a slight modification in the algebraic structure of the topological group gives us this property.

8.3 Uniform Continuity

An advantage of the real line $\mathbf{R}$ is that continuity can be explained in a *uniform sense*, because we can transport any topological questions about a certain point x to questions about topological structure near the origin via the map $g \mapsto x^{-1}g$. We can then define a uniformly continuous function $f : \mathbf{R} \to \mathbf{R}$ to be a function possessing, for every $\varepsilon > 0$, a $\delta > 0$ such that if $|y| < \delta$, $|f(x+y) - f(x)| < \varepsilon$. Instead of having to specify a δ for every point

on the domain, the δ works uniformly everywhere. The group structure is all we need to talk about these questions.

We say a function $f : G \to H$ between topological groups is (left) uniformly continuous if, for any open neighbourhood U of the origin in H, there is a neighbourhood V of the origin in G such that for each x, $f(xV) \subset f(x)U$. Right continuity requires $f(Vx) \subset Uf(x)$. The requirement of distinguishing between left and right uniformity is important when we study non-commutative groups, for there are certainly left uniform maps which are not right uniform in these groups. If $f : G \to \mathbf{C}$, then left uniform continuity is equivalent to the fact that $\|L_x f - f\|_\infty \to 0$ as $x \to 1$, where $(L_x f)(y) = f(xy)$. Right uniform continuity requires $\|R_x f - f\|_\infty \to 0$, where $(R_x f)(y) = f(yx)$. R_x is a homomorphism, but L_x is what is called an antihomomorphism.

Example. *Let G be any Hausdorff non-commutative topological group, with sequences x_i and y_i for which $x_i y_i \to e$, $y_i x_i \to z \neq e$. Then the uniform structures on G are not equivalent.*

It is hopeless to express uniform continuity in terms of a new topology on G, because the topology only gives a local description of continuity, which prevents us from describing things uniformly across the whole group. However, we can express uniform continuity in terms of a new topology on $G \times G$. If $U \subset G$ is an open neighbourhood of the origin, let

$$L_U = \{(x,y) : yx^{-1} \in U\} \quad R_U = \{(x,y) : x^{-1}y \in U\}$$

The family of all L_U (resp. R_U) is known as the left (right) uniform structure on G, denoted $LU(G)$ and $RU(G)$. Fix a map $f : G \to H$, and consider the map

$$g(x,y) = (f(x), f(y))$$

from G^2 to H^2. Then f is left (right) uniformly continuous if and only if g is continuous with respect to $LU(G)$ and $LU(H)$ ($RU(G)$ and $RU(H)$). $LU(G)$ and $RU(G)$ are weaker than the product topologies on G and H, which reflects the fact that uniform continuity is a strong condition than normal continuity. We can also consider uniform maps with respect to $LU(G)$ and $RU(H)$, and so on and so forth. We can also consider uniform continuity on functions defined on an open subset of a group.

Example. *Here are a few examples of easily verified continuous maps.*

- *If the identity map on G is left-right uniformly continuous, then $LU(G) = RU(G)$, and so uniform continuity is invariant of the uniform structure chosen.*

- *Translation maps $x \mapsto axb$, for $a, b \in G$, are left and right uniform.*

- *Inversion is uniformly continuous.*

Theorem 8.9. *All continuous maps on compact subsets of topological groups are uniformly continuous.*

Proof. Let K be a compact subset of a group G, and let $f : K \to H$ be a continuous map into a topological group. We claim that f is then uniformly continuous. Fix an open neighbourhood V of the origin, and let V' be a symmetric neighbourhood such that $V'^2 \subset V$. For any x, there is U_x such that

$$f(x)^{-1} f(xU_x) \subset V'$$

Choose U'_x such that $U'^2_x \subset U_x$. The xU'_x cover K, so there is a finite subcover corresponding to sets $U'_{x_1}, \ldots, U'_{x_n}$. Let $U = U'_{x_1} \cap \cdots \cap U'_{x_n}$. Fix $y \in G$, and suppose $y \in x_k U'_{x_k}$. Then

$$
\begin{aligned}
f(y)^{-1} f(yU) &= f(y)^{-1} f(x_k) f(x_k)^{-1} f(yU) \\
&\subset f(y)^{-1} f(x_k) f(x_k)^{-1} f(x_k U x_k) \\
&\subset f(y)^{-1} f(x_k) V' \\
&\subset V'^2 \subset V
\end{aligned}
$$

So that f is left uniformly continuous. Right uniform continuity is proven in the exact same way. $\qquad \square$

Corollary 8.10. *All maps with compact support are uniformly continuous.*

Corollary 8.11. *Uniform continuity on compact groups is invariant of the uniform structure chosen.*

8.4 Ordered Groups

In this section we describe a general class of groups which contain both interesting and pathological examples. Let G be a group with an ordering

$<$ preserved by the group operations, so that $a < b$ implies both $ag < bg$ and $ga < gb$. We now prove that the order topology gives G the structure of a normal topological group (the normality follows because of general properties of order topologies).

First note, that $a < b$ implies $a^{-1} < b^{-1}$. This results from a simple algebraic trick, because

$$a^{-1} = a^{-1}bb^{-1} > a^{-1}ab^{-1} = b^{-1}$$

This implies that the inverse image of an interval (a, b) under inversion is (b^{-1}, a^{-1}), hence inversion is continuous.

Now let $e < b < a$. We claim that there is then $e < c$ such that $c^2 < a$. This follows because if $b^2 \geqslant a$, then $b \geqslant ab^{-1}$ and so

$$(ab^{-1})^2 = ab^{-1}ab^{-1} \leqslant ab^{-1}b = a$$

Now suppose $a < e < b$. If $\inf\{y : y > e\} = x > e$, then $(x^{-1}, x) = \{e\}$, and the topology on G is discrete, hence the continuity of operations is obvious. Otherwise, we may always find c such that $c^2 < b$, $a < c^{-2}$, and then if $c^{-1} < g, h < c$, then

$$a < c^{-2} < gh < c^2 < b$$

so multiplication is continuous at every pair (x, x^{-1}). In the general case, if $a < gh < b$, then $g^{-1}ah^{-1} < e < g^{-1}bh^{-1}$, so there is c such that if $c^{-1} < g', h' < c$, then $g^{-1}ah^{-1} < g'h' < g^{-1}bh^{-1}$, so $a < gg'h'h < b$. The set of gg', where $c^{-1} < g' < c$, is really just the set of $gc^{-1} < x < gc$, and the set of $h'h$ is really just the set of $c^{-1}h < x < ch$. Thus multiplication is continuous everywhere.

Example (Dieudonne). *For any well ordered set S, the dictionary ordering on $\mathbf{R}^S$ induces a linear ordering inducing a topological group structure on the set of maps from S to $\mathbf{R}$.*

Let us study Dieudonne's topological group in more detail. If S is a finite set, or more generally possesses a maximal element w, then the topology on $\mathbf{R}^S$ can be defined such that $f_i \to f$ if eventually $f_i(s) = f(s)$ for all $s < w$ simultaneously, and $f_i(w) \to f(w)$. Thus $\mathbf{R}^S$ is isomorphic (topologically) to a discrete union of a certain number of copies of $\mathbf{R}$, one for each tuple in $S - \{w\}$.

If S has a countable cofinal subset $\{s_i\}$, the topology is no longer so simple, but $\mathbf{R}^S$ is still first countable, because the sets

$$U_i = \{f : (\forall w < s_i : f(w) = 0)\}$$

provide a countable neighbourhood basis of the origin.

The strangest properties of $\mathbf{R}^S$ occur when S has no countable cofinal set. Suppose that $f_i \to f$. We claim that it follows that $f_i = f$ eventually. To prove by contradiction, we assume without loss of generality (by thinning the sequence) that no f_i is equal to f. For each f_i, find the largest $w_i \in S$ such that for $s < w_i$, $f_i(s) = f(s)$ (since S is well ordered, the set of elements for which $f_i(s) \neq f(s)$ has a minimal element). Then the w_i form a countable cofinal set, because if $v \in S$ is arbitrary, the f_i eventually satisfy $f_i(s) = f(s)$ for $s < v$, hence the corresponding w_i is greater than v_i. Hence, if $f_i \to f$ in $\mathbf{R}^S$, where S does not have a countable cofinal subset, then eventually $f_i = f$. We conclude all countable sets in $\mathbf{R}^S$ are closed, and this proof easily generalises to show that if S does not have a cofinal set of cardinality $\mathfrak{a}$, then every set of cardinality $\leqslant \mathfrak{a}$ is closed.

The simple corollary to this proof is that compact subsets are finite. Let $X = f_1, f_2, \ldots$ be a denumerable, compact set. Since all subsets of X are compact, we may assume $f_1 < f_2 < \ldots$ (or $f_1 > f_2 > \ldots$, which does not change the proof in any interesting way). There is certainly $g \in \mathbf{R}^S$ such that $g < f_1$, and then the sets $(g, f_2), (f_1, f_3), (f_2, f_4), \ldots$ form an open cover of X with no finite subcover, hence X cannot be compact. We conclude that the only compact subsets of $\mathbf{R}^S$ are finite.

Furthermore, the class of open sets is closed under countable intersections. Consider a series of functions

$$f_1 \leqslant f_2 \leqslant \cdots < h < \cdots \leqslant g_2 \leqslant g_1$$

Suppose that $f_i \leqslant k < h < k' \leqslant g_j$. Then the intersection of the (f_i, g_i) contains an interval (k, k') around h, so that the intersection is open near h. The only other possiblity is that $f_i \to h$ or $g_i \to h$, which can only occur if $f_i = h$ or $g_i = h$ eventually, in which case we cannot have $f_i < h$, $h < g_i$. We conclude the intersection of countably many intervals is open, because we can always adjust any intersection to an intersection of this form without changing the resulting intersecting set (except if the set is empty, in which case the claim is trivial). The general case results from noting that any open set in an ordered group is a union of intervals.

8.5 Topological Groups arising from Normal subgroups

Let G be a group, and $\mathcal{N}$ a family of normal subgroups closed under intersection. If we interpret $\mathcal{N}$ as a neighbourhood base at the origin, the resulting topology gives G the structure of a totally disconnected topological group, which is Hausdorff if and only if $\bigcap \mathcal{N} = \{e\}$. First note that $g_i \to g$ if g_i is eventually in gN, for every $N \in \mathcal{N}$, which implies $g_i^{-1} \in Ng^{-1} = g^{-1}N$, hence inversion is continuous. Furthermore, if h_i is eventually in hN, then $g_i h_i \in gNhN = ghN$, so multiplication is continuous. Finally note that $N^c = \bigcup_{g \neq e} gN$ is open, so that every open set is closed.

Example. *Consider $\mathcal{N} = \{\mathbf{Z}, 2\mathbf{Z}, 3\mathbf{Z}, \dots\}$. Then $\mathcal{N}$ induces a Hausdorff topology on $\mathbf{Z}$, such that $g_i \to g$, if and only if g_i is eventually in $g + n\mathbf{Z}$ for all n. In this topology, the series $1, 2, 3, \dots$ converges to zero!*

This example gives us a novel proof, due to Furstenburg, that there are infinitely many primes. Suppose that there were only finitely many, $\{p_1, p_2, \dots, p_n\}$. By the fundamental theorem of arithmetic,

$$\{-1, 1\} = (\mathbf{Z}p_1)^c \cap \cdots \cap (\mathbf{Z}p_n)^c$$

and is therefore an open set. But this is clearly not the case as open sets must contain infinite sequences.

Chapter 9

The Haar Measure

One of the reasons that we isolate locally compact groups to study is that they possess an incredibly useful object allowing us to understand functions on the group, and thus the group itself. A **left (right) Haar measure** for a group G is a Radon measure μ for which $\mu(xE) = \mu(E)$ for any $x \in G$ and measurable E ($\mu(Ex) = \mu(E)$ for all x and E). For commutative groups, all left Haar measures are right Haar measures, but in non-commutative groups this need not hold. However, if μ is a right Haar measure, then $\nu(E) = \mu(E^{-1})$ is a left Haar measure, so there is no loss of generality in focusing our study on left Haar measures.

Example. *The example of a Haar measure that everyone knows is the Lebesgue measure on* $\mathbf{R}$ *(or* $\mathbf{R}^n$*). It commutes with translations because it is the measure induced by the linear functional corresponding to Riemann integration on* $C_c^+(\mathbf{R}^n)$. *A similar theory of Darboux integration can be applied to linearly ordered groups, leading to the construction of a Haar measure on such a group.*

Example. *If G is a Lie group, consider a 2-tensor $g_e \in T_e^2(G)$ inducing an inner product at the origin. Then the diffeomorphism $f : a \mapsto b^{-1}a$ allows us to consider $g_b = f^*\lambda \in T_b^2(G)$, and this is easily verified to be an inner product, hence we have a Riemannian metric. The associated Riemannian volume element can be integrated, producing a Haar measure on G.*

Example. *If G and H have Haar measures μ and ν, then $G \times H$ has a Haar measure $\mu \times \nu$, so that the class of topological groups with Haar measures is closed under the product operation. We can even allow infinite products, provided that the groups involved are compact, and the Haar measures are normalized*

to probability measures. This gives us measures on F_2^ω and $\mathbf{T}^\omega$, which models the probability of an infinite sequence of coin flips.

Example. *dx/x is a Haar measure for the multiplicative group of positive real numbers, since*

$$\int_a^b \frac{1}{x} = \log(b) - \log(a) = \log(cb) - \log(ca) = \int_{ca}^{cb} \frac{1}{x}$$

If we take the multiplicative group of all non-negative real numbers, the Haar measure becomes $dx/|x|$.

Example. *$dxdy/(x^2 + y^2)$ is a Haar measure for the multiplicative group of complex numbers, since we have a basis of 'arcs' around the origin, and by a change of variables to polar coordinates, we verify the integral is changed by multiplication. Another way to obtain this measure is by noticing that $\mathbf{C}^\times$ is topologically isomorphic to the product of the circle group and the multiplicative group of real numbers, and hence the measure obtained should be the product of these measures. Since*

$$\frac{dxdy}{x^2 + y^2} = \frac{drd\theta}{r}$$

We see that this is just the product of the Haar measure on $\mathbf{R}^+$, dr/r, and the Haar measure on $\mathbf{T}$, $d\theta$.

Example. *The space $M_n(\mathbf{R})$ of all n by n real matrices under addition has a Haar measure dM, which is essentially the Lebesgue measure on $\mathbf{R}^{n^2}$. If we consider the measure on $GL_n(\mathbf{R})$, defined by*

$$\frac{dM}{det(M)^n}$$

To see this, note the determinant of the map $M \mapsto NM$ on $M_n(\mathbf{R})$ is $det(N)^n$, because we can view $M_n(\mathbf{R})$ as the product of $\mathbf{R}^n$ n times, multiplication operates on the space componentwise, and the volume of the image of the unit paralelliped in each $\mathbf{R}^n$ is $det(N)$. Since the multiplicative group of complex numbers $z = x + iy$ can be identified with the group of matrices of the form

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix}$$

101

and the measure on $\mathbf{C} - \{0\}$ then takes the form $dM/det(M)$. More generally, if G is an open subset of $\mathbf{R}^n$, and left multiplication acts affinely, $xy = A(x)y + b(x)$, then $dx/|det(A(x))|$ is a left Haar measure on G, where dx is Lebesgue measure.

It turns out that there is a Haar measure on any locally compact group, and what's more, it is unique up to scaling. The construction of the measure involves constructing a positive linear functional $\phi : C_c(G) \to \mathbf{R}$ such that $\phi(L_x f) = \phi(f)$ for all x. The Riesz representation theorem then guarantees the existence of a Radon measure μ which represents this linear functional, and one then immediately verifies that this measure is a Haar measure.

Theorem 9.1. *Every locally compact group G has a Haar measure.*

Proof. The idea of the proof is fairly simple. If μ was a Haar measure, $f \in C_c^+(G)$ was fixed, and $\phi \in C_c^+(G)$ was a function supported on a small set, and behaving like a step function, then we could approximate f well by translates of ϕ,

$$f(x) \approx \sum c_i (L_{x_i} \phi)$$

Hence

$$\int f(x) d\mu \approx \sum c_i \int L_{x_i} \phi = \sum c_i \int \phi$$

If $\int \phi = 1$, then we could approximate $\int f(x) d\mu$ as literal sums of coefficients c_i. Since μ is outer regular, and ϕ is supported on neighbourhoods, one can show $\int f(x) d\mu$ is the infimum of $\sum c_i$, over all choices of $c_i > 0$ and $\int \phi \geqslant 1$, for which $f \leqslant \sum c_i L_{x_i} \phi$. Without the integral, we cannot measure the size of the functions ϕ, so we have to normalize by a different factor. We define $(f : \phi)$ to be the infimum of the sums $\sum c_i$, where $f \leqslant \sum c_i L_{x_i} \phi$ for some $x_i \in G$. We would then have

$$\int f d\mu \leqslant (f : \phi) \int \phi d\mu$$

If k is fixed with $\int k = 1$, then we would have

$$\int f d\mu \leqslant (f : \phi)(\phi : k)$$

We cannot change k if we wish to provide a limiting result in ϕ, so we notice that $(f:g)(g:h) \leqslant (f:h)$, which allows us to write

$$\int f \, d\mu \leqslant \frac{(f:\phi)}{(k:\phi)}$$

Taking the support of ϕ to be smaller and smaller, this value should approximate the integral perfectly accurately.

Define the linear functional

$$I_\phi(f) = \frac{(f:\phi)}{(k:\phi)}$$

Then I_ϕ is a sublinear, monotone, function with a functional bound

$$(k:f)^{-1} \leqslant I_\phi(f) \leqslant (f:k)$$

Which effectively says that, regardless of how badly we choose ϕ, the approximation factor $(f:\phi)$ is normalized by the approximation factor $(k:\phi)$ so that the integral is bounded. Now we need only prove that I_ϕ approximates a linear functional well enough that we can perform a limiting process to obtain a Haar integral. If $\varepsilon > 0$, and $g \in C_c^+(G)$ with $g = 1$ on $\mathrm{supp}(f_1 + f_2)$, then the functions

$$h = f_1 + f_2 + \varepsilon g$$

$$h_1 = f_1/h \quad h_2 = f_2/h$$

are in $C_0^+(G)$, if we define $h_i(x) = 0$ if $f_i(x) = 0$. This implies that there is a neighbourhood V of e such that if $x \in V$, and y is arbitrary, then

$$|h_1(xy) - h_1(y)| \leqslant \varepsilon \quad |h_2(xy) - h_2(y)| < \varepsilon$$

If $\mathrm{supp}(\phi) \subset V$, and $h \leqslant \sum c_i L_{x_i} \phi$, then

$$f_j(x) = h(x)h_j(x) \leqslant \sum c_i \phi(x_i x) h_j(x) \leqslant \sum c_i \phi(x_i x) \left[h_j(x_i^{-1}) + \varepsilon \right]$$

since we may assume that $x_i x \in \mathrm{supp}(\phi) \subset V$. Then, because $h_1 + h_2 \leqslant 1$,

$$(f_1:\phi) + (f_2:\phi) \leqslant \sum c_j [h_1(x_j^{-1}) + \varepsilon] + \sum c_j [h_2(x_j^{-1}) + \varepsilon] \leqslant \sum c_j [1 + 2\varepsilon]$$

103

Now we find, by taking infinums, that

$$I_\phi(f_1) + I_\phi(f_2) \leqslant I_\phi(h)(1 + 2\varepsilon) \leqslant [I_\phi(f_1 + f_2) + \varepsilon I_\phi(g)][1 + 2\varepsilon]$$

Since g is fixed, and we have a bound $I_\phi(g) \leqslant (g : k)$, we may always find a neighbourhood V (dependant on f_1, f_2) for any $\varepsilon > 0$ such that

$$I_\phi(f_1) + I_\phi(f_2) \leqslant I_\phi(f_1 + f_2) + \varepsilon$$

if $\mathrm{supp}(\phi) \subset V$.

Now we have estimates on how well I_ϕ approximates a linear function, so we can apply a limiting process. Consider the product

$$X = \prod_{f \in C_0^+(G)} [(k : f)^{-1}, (k : f_0)]$$

a compact space, by Tychonoff's theorem, consisting of $F : C_c^+(G) \to \mathbf{R}$ such that $(k : f)^{-1} \leqslant F(f) \leqslant (f : k)$. For each neighbourhood V of the identity, let $K(V)$ be the closure of the set of I_ϕ such that $\mathrm{supp}(\phi) \subset V$. Then the set of all $K(V)$ has the finite intersection property, so we conclude there is some $I : C_c^+(G) \to \mathbf{R}$ contained in $\bigcap K(V)$. This means that every neighbourhood of I contains I_ϕ with $\mathrm{supp}(\phi) \subset V$, for all ϕ. This means that if $f_1, f_2 \in C_c^+(G)$, $\varepsilon > 0$, and V is arbitrary, there is ϕ with $\mathrm{supp}(\phi) \subset V$, and

$$|I(f_1) - I_\phi(f_1)| < \varepsilon \quad |I(f_2) - I_\phi(f_2)| < \varepsilon$$

$$|I(f_1 + f_2) - I_\phi(f_1 + f_2)| < \varepsilon$$

this implies that if V is chosen small enough, then

$$|I(f_1 + f_2) - (I(f_1) - I(f_2))| \leqslant 2\varepsilon + |I_\phi(f_1 + f_2) - (I_\phi(f_1) + I_\phi(f_2))| < 3\varepsilon$$

Taking $\varepsilon \to 0$, we conclude I is linear. Similar limiting arguments show that I is homogenous of degree 1, and commutes with all left translations. We conclude the extension of I to a linear functional on $C_0(G)$ is well defined, and the Radon measure obtained by the Riesz representation theorem is a Haar measure. $\qquad \square$

We shall prove that the Haar measure is unique, but first we show an incredibly useful regularity property.

Proposition 9.2. *If U is open, and μ is a Haar measure, then $\mu(U) > 0$. It follows that if f is in $C_c^+(G)$, then $\int f \, d\mu > 0$.*

Proof. If $\mu(U) = 0$, then for any $x_1, \ldots, x_n \in G$,

$$\mu\left(\bigcup_{i=1}^n x_i U\right) \leqslant \sum_{i=1}^n \mu(x_i U) = 0$$

If K is compact, then K can be covered by finitely many translates of U, so $\mu(K) = 0$. But then $\mu = 0$ by regularity, a contradiction. $\square$

Theorem 9.3. *Haar measures are unique up to a multiplicative constant.*

Proof. Let μ and ν be Haar measures. Fix a compact neighbourhood V of the identity. If $f, g \in C_c^+(G)$, consider the compact sets

$$A = \mathrm{supp}(f)V \cup V\mathrm{supp}(f) \qquad B = \mathrm{supp}(g)V \cup V\mathrm{supp}(g)$$

Then the functions $F_y(x) = f(xy) - f(yx)$ and $G_y(x) = g(xy) - g(yx)$ are supported on A and B. There is a neighbourhood $W \subset V$ of the identity such that $\|F_y\|_\infty, \|G_y\|_\infty < \varepsilon$ if $y \in W$. Now find $h \in C_c^+(G)$ with $h(x) = h(x^{-1})$ and $\mathrm{supp}(h) \subset W$ (take $h(x) = k(x)k(x^{-1})$ for some function $k \in C_c^+(G)$ with $\mathrm{supp}(k) \subset W$, and $k = 1$ on a symmetric neighbourhood of the origin). Then

$$\left(\int h \, d\mu\right)\left(\int f \, d\lambda\right) = \int h(y)f(x)d\mu(y)d\lambda(x)$$
$$= \int h(y)f(yx)d\mu(y)d\lambda(x)$$

and

$$\left(\int h \, d\lambda\right)\left(\int f \, d\mu\right) = \int h(x)f(y)d\mu(y)d\lambda(x)$$
$$= \int h(y^{-1}x)f(y)d\mu(y)d\lambda(x)$$
$$= \int h(x^{-1}y)f(y)d\mu(y)d\lambda(x)$$
$$= \int h(y)f(xy)d\mu(y)d\lambda(x)$$

105

Hence, applying Fubini's theorem,

$$\left| \int h d\mu \int f d\lambda - \int h d\lambda \int f d\mu \right| \leq \int h(y) |F_y(x)| d\mu(y) d\lambda(x)$$

$$\leq \varepsilon \lambda(A) \int h d\mu$$

In the same way, we find this is also true when f is swapped with g, and A with B. Dividing this inequalities by $\int h d\mu \int f d\mu$, we find

$$\left| \frac{\int f d\lambda}{\int f d\mu} - \frac{\int h d\lambda}{\int h d\mu} \right| \leq \frac{\varepsilon \lambda(A)}{\int f d\mu}$$

and this inequality holds with f swapped out with g, A with B. We then combine these inequalities to conclude

$$\left| \frac{\int f d\lambda}{\int f d\mu} - \frac{\int g d\lambda}{\int g d\mu} \right| \leq \varepsilon \left[\frac{\lambda(A)}{\int f d\mu} + \frac{\lambda(B)}{\int g d\mu} \right]$$

Taking ε to zero, we find $\lambda(A), \lambda(B)$ remain bounded, and hence

$$\frac{\int f d\lambda}{\int f d\mu} = \frac{\int g d\lambda}{\int g d\mu}$$

Thus there is a cosntant $c > 0$ such that $\int f d\lambda = c \int f d\mu$ for any function $f \in C_c^+(G)$, and we conclude that $\lambda = c\mu$. $\qquad\square$

The theorem can also be proven by looking at the translation invariant properties of the derivative $f = d\mu/d\nu$, where $\nu = \mu + \lambda$ (We assume our group is σ compact for now). Consider the function $g(x) = f(yx)$. Then

$$\int_A g(x) d\nu = \int_{yA} f(x) d\nu = \mu(yA) = \mu(A)$$

so g is derivative, and thus $f = g$ almost everywhere. Our interpretation is that for a fixed y, $f(yx) = f(x)$ almost everywhere with respect to ν. Then (applying a discrete version of Fubini's theorem), we find that for almost all x with respect to ν, $f(yx) = f(x)$ holds for almost all y. But this implies that there exists an x for which $f(yx) = f(x)$ holds almost everywhere. Thus for any measurable A,

$$\mu(A) = \int_A f(y) d\nu(y) = f(x)\nu(A) = f(x)\mu(A) + f(x)\nu(A)$$

106

Now $(1 - f(x))\mu(A) = f(x)\nu(A)$ for all A, implying (since $\mu, \nu \neq 0$), that $f(x) \neq 0, 1$, and so

$$\frac{1 - f(x)}{f(x)} \mu(A) = \nu(A)$$

for all A. This shows the uniqueness property for all σ compact groups. If G is an arbitrary group with two measures μ and ν, then there is c such that $\mu = c\nu$ on every component of G, and thus on the union of countably many components. If A intersects uncountably many components, then either $\mu(A) = \nu(A) = \infty$, or the intersection of A on each set has positive measure on only countably many components, and in either case we have $\mu(A) = \nu(A)$.

9.1 Fubini, Radon Nikodym, and Duality

Before we continue, we briefly mention that integration theory is particularly nice over locally compact groups, even if we do not have σ finiteness. This essentially follows because the component of the identity in G is σ compact (take a compact neighbourhood and its iterated multiples), hence all components in G are σ compact. The three theorems that break down outside of the σ compact domain are Fubini's theorem, the Radon Nikodym theory, and the duality between $L^1(X)$ and $L^\infty(X)$. We show here that all three hold if X is a locally compact topological group.

First, suppose that $f \in L^1(G \times G)$. Then the essential support of f is contained within countably many components of $G \times G$ (which are simply products of components in G). Thus f is supported on a σ compact subset of $G \times G$ (as a locally compact topological group, each component of $G \times G$ is σ compact), and we may apply Fubini's theorem on the countably many components (the countable union of σ compact sets is σ compact). The functions in $L^p(G)$, for $1 \leqslant p < \infty$, also vanish outside of a σ compact subset (for if $f \in L^p(G)$, $|f|^p \in L^1(G)$ and thus vanishes outside of a σ compact set). What's more, all finite sums and products of functions from these sets (in either variable) vanish outside of σ compact subsets, so we almost never need to explicitly check the conditions for satisfying Fubini's theorem, and from now on we apply it wantonly.s

Now suppose μ and ν are both Radon measures, with $\nu \ll \mu$, and ν is σ-finite. By inner regularity, the support of ν is a σ compact set E. By inner regularity, μ restricted to E is σ finite, and so we may find a Radon

107

Nikodym derivative on E. This derivative can be extended to all of G because ν vanishes on G.

Finally, we note that $L^\infty(X) = L^1(X)^*$ can be made to hold if X is not σ finite, but locally compact and Hausdorff, provided we are integrating with respect to a Radon measure μ, and we modify $L^\infty(G)$ slightly. Call a set $E \subset X$ **locally Borel** if $E \cap F$ is Borel whenever F is Borel and $\mu(F) < \infty$. A locally Borel set is **locally null** if $\mu(E \cap F) = 0$ whenever $\mu(F) < \infty$ and F is Borel. We say a property holds **locally almost everywhere** if it is true except on a locally null set. $f : X \to \mathbf{C}$ is **locally measurable** if $f^{-1}(U)$ is locally Borel for every borel set $U \subset \mathbf{C}$. We now define $L^\infty(X)$ to be the space of all functions bounded except on a locally null set, modulo functions that are locally zero. That is, we define a norm

$$\|f\|_\infty = \inf\{c : |f(x)| \leqslant c \text{ locally almost everywhere}\}$$

and then $L^\infty(X)$ consists of the functions that have finite norm. It then follows that if $f \in L^\infty(X)$ and $g \in L^1(X)$, then g vanishes outside of a σ-finite set Y, so $fg \in L^1(X)$, and if we let $Y_1 \subset Y_2 \subset \cdots \to Y$ be an increasing subsequence such that $\mu(Y_i) < \infty$, then $|f(x)| \leqslant \|f\|_\infty$ almost everywhere for $x \in Y_i$, and so by the monotone convergence theorem

$$\int |fg|d\mu = \lim_{Y_i \to \infty} \int_{Y_i} |fg|d\mu \leqslant \|f\|_\infty \int_{Y_i} |g|d\mu \leqslant \|f\|_\infty \|g\|_1$$

Thus the map $g \mapsto \int fg\,d\mu$ is a well defined, continuous linear functional with norm $\|f\|_\infty$. That $L^1(X)^* = L^\infty(X)$ follows from the decomposibility of the Carathéodory extension of μ, a fact we leave to the general measure theorists.

9.2 Unimodularity

We have thus defined a left invariant measure, but make sure to note that such a function is not right invariant. We call a group who's left Haar measure is also right invariant **unimodular**. Obviously all abelian groups are unimodular.

Given a fixed y, the measure $\mu_y(A) = \mu(Ay)$ is a new Haar measure on the space, hence there is a constant $\Delta(y) > 0$ depending only on y such that $\mu(Ay) = \Delta(y)\mu(A)$ for all measurable A. Since $\mu(Axy) = \Delta(y)\mu(Ay) =$

$\Delta(x)\Delta(y)\mu(A)$, we find that $\Delta(xy) = \Delta(x)\Delta(y)$, so Δ is a homomorphism from G to the multiplicative group of real numbers. For any $f \in L^1(\mu)$, we have

$$\int f(xy)d\mu(x) = \Delta(y^{-1}) \int f(x)d\mu(x)$$

If $y_i \to e$, and $f \in C_c(G)$, then $\|R_{y_i}f - f\|_\infty \to 0$, so

$$\Delta(y_i^{-1}) \int f(x)d\mu = \int f(xy_i)d\mu \to \int f(x)d\mu$$

Hence $\Delta(y_i^{-1}) \to 1$. This implies Δ, known as the unimodular function, is a continuous homomorphism from G to the real numbers. Note that Δ is trivial if and only if G is unimodular.

Theorem 9.4. *Any compact group is unimodular.*

Proof. $\Delta : G \to \mathbf{R}^*$ is a continuous homomorphism, hence $\Delta(G)$ is compact. But the only compact subgroup of $\mathbf{R}$ is trivial, hence Δ is trivial. $\square$

Let G^c be the smallest closed subgroup of G containing the commutators $[x,y] = xyx^{-1}y^{-1}$. It is verified to be a normal subgroup of G by simple algebras.

Theorem 9.5. *If G/G^c is compact, then G is unimodular.*

Proof. Δ factors through G/G^c since it is abelian. But if Δ is trivial on G/G^c, it must also be trivial on G. $\square$

The modular function relates right multiplication to left multiplcation in the group. In particular, if $d\mu$ is a Left Haar measure, then $\Delta^{-1}d\mu$ is a right Haar measure. Hence any right Haar measure is a constant multiple of $\Delta^{-1}d\mu$. Hence the measure $\nu(A) = \mu(A^{-1})$ has a value c such that for any function f,

$$\int \frac{f(x)}{\Delta(x)}d\mu(x) = c \int f(x)d\nu(x) = c \int f(x^{-1})d\mu$$

If $c \neq 1$, pick a symmetric neighbourhood U such that for $x \in U$, $|\Delta(x) - 1| \leqslant \varepsilon|c - 1|$. Then if $f > 0$

$$|c - 1|\mu(U) = |c\mu(U^{-1}) - \mu(U)| = \left|\int_U [\Delta(x^{-1}) - 1]d\mu(x)\right| \leqslant \varepsilon\mu(U)|c - 1|$$

A contradiction if $\varepsilon < 1$. Thus we have

$$\int f(x^{-1})d\mu(x) = \int \frac{f(x)}{\Delta(x)}d\mu(x)$$

A useful integration trick. When Δ is unbounded, then it follows that $L^p(\mu)$ and $L^p(\nu)$ do not consist of the same functions. There are two ways of mapping the sets isomorphically onto one another – the map $f(x) \mapsto f(x^{-1})$, and the map $f(x) \mapsto \Delta(x)^{1/p}f(x)$.

From now on, we assume a left invariant Haar measure is fixed over an entire group. Since a Haar measure is uniquely determined up to a constant, this is no loss of generality, and we might as well denote our integration factors $d\mu(x)$ and $d\mu(y)$ as dx and dy, where it is assumed that this integration is over the Lebesgue measure.

9.3   Convolution

If G is a topological group, then $C(G)$ does not contain enough algebraic structure to identify G – for instance, if G is a discrete group, then $C(G)$ is defined solely by the cardinality of G. The algebras we wish to study over G is the space $M(G)$ of all complex valued Radon measures over G and the space $L^1(G)$ of integrable functions with respect to the Haar measure, because here we can place a Banach algebra structure with an involution. We note that $L^1(G)$ can be isometrically identified as the space of all measures $\mu \in M(G)$ which are absolutely continuous with respect to the Haar measure. Given $\mu, \nu \in M(G)$, we define the convolution measure

$$\int \phi d(\mu * \nu) = \int \phi(xy)d\mu(x)d\nu(y)$$

The measure is well defined, for if $\phi \in C_c^+(X)$ is supported on a compact set K, then

$$\left| \int \phi(xy)d\mu(x)d\nu(y) \right| \leq \int_G \int_G \phi(xy)d|\mu|(x)d|\nu|(y)$$
$$\leq \|\mu\|\|\nu\|\|\phi\|_\infty$$

110

This defines an operation on $M(G)$ which is associative, since, by applying the associativity of G and Fubini's theorem.

$$\int \phi d((\mu * \nu) * \lambda) = \int \int \phi(xz) d(\mu * \nu)(x) d\lambda(z)$$

$$= \int \int \int \phi((xy)z) d\mu(x) d\nu(y) d\lambda(z)$$

$$= \int \int \int \phi(x(yz)) d\mu(x) d\nu(y) d\lambda(z)$$

$$= \int \int \phi(xz) d\mu(x) d(\nu * \lambda)(z)$$

$$= \int \phi d(\mu * (\nu * \lambda))$$

Thus we begin to see how the structure of G gives us structure on $M(G)$. Another example is that convolution is commutative if and only if G is commutative. We have the estimate $\|\mu * \nu\| \leqslant \|\mu\|\|\nu\|$, because of the bound we placed on the integrals above. $M(G)$ is therefore an involutive Banach algebra, which has a unit, the dirac delta measure at the identity.

As a remark, we note that involutive Banach algebras have nowhere as near a nice of a theory than that of C^* algebras. $M(G)$ cannot be renormed to be a C^* algebra, since every weakly convergent Cauchy sequence converges, which is impossible in a C^* algebra, except in the finite dimensional case.

A **discrete measure** on G is a measure in $M(G)$ which vanishes outside a countable set of points, and the set of all such measures is denoted $M_d(G)$. A **continuous measure** on G is a measure μ such that $\mu(\{x\}) = 0$ for all $x \in G$. We then have a decomposition $M(G) = M_d(G) \oplus M_c(G)$, for if μ is any measure, then $\mu(\{x\}) \neq 0$ for at most countably many points x, for

$$\|\mu\| \geqslant \sum_{x \in G} |\mu|(x)$$

This gives rise to a discrete measure ν, and $\mu - \nu$ is continuous. If we had another decomposition, $\mu = \psi + \phi$, then $\mu(\{x\}) = \psi(\{x\}) = \nu(\{x\})$, so $\psi = \nu$ by discreteness, and we then conclude $\phi = \mu - \nu$. $M_c(G)$ is actually a closed subspace of $M(G)$, since if $\mu_i \to \mu$, and $\mu_i \in M_c(G)$, and $\|\mu_i - \mu\| < \varepsilon$, then for any $x \in G$,

$$\varepsilon > \|\mu - \mu_i\| \geqslant |(\mu_i - \mu)(\{x\})| = |\mu(\{x\})|$$

Letting $\varepsilon \to 0$ shows continuity.

The convolution on $M(G)$ gives rise to a convolution on $L^1(G)$, where

$$(f * g)(x) = \int f(y)g(y^{-1}x)dy$$

which satisfies $\|f * g\|_1 \leqslant \|f\|_1 \|g\|_1$. This is induced by the identification of f with $f(x)dx$, because then

$$\int \phi(f(x)dx * g(x)dx) = \int \int \phi(yx)f(y)g(x)dydx$$
$$= \int \phi(y) \left(\int f(y)g(y^{-1}x)dx \right) dy$$

Hence $f d\mu * g d\mu = (f * g)d\mu$. What's more,

$$\|f\|_1 = \|f d\mu\|$$

If $\nu \in M(G)$, then we can still define $\nu * f \in L^1(G)$

$$(\nu * f)(x) = \int f(y^{-1}x)d\mu(y)$$

which holds since

$$\int \phi d(\nu * f\mu) = \int \phi(yx)f(x)d\nu(y)d\mu(x) = \int \phi(x)f(y^{-1}x)d\nu(y)d\mu(x)$$

If G is unimodular, then we also find

$$\int \phi d(f\mu * \nu) = \int \phi(yx)f(y)d\mu(y)d\nu(x) = \int \phi(x)f(y)d\mu(y)d\nu(y^{-1}x)$$

So we let $f * \mu(x) = \int f(y)d\mu(y^{-1}x)$.

Theorem 9.6. $L^1(G)$ and $M_c(G)$ *are closed ideals in* $M(G)$, *and* $M_d(G)$ *is a closed subalgebra.*

Proof. If $\mu_i \to \mu$, and each μ_i is discrete, the μ is discrete, because there is a countable set K such that all μ_i are equal to zero outside of K, so μ must also vanish outside of K (here we have used the fact that $M(G)$ is a Banach space, so that we need only consider sequences). Thus $M_d(G)$ is closed,

and is easily verified to be subalgebra, essentially because $\delta_x * \delta_y = \delta_{xy}$. If $\mu_i \to \mu$, then $\mu_i(\{x\}) \to \mu(\{x\})$, so that $M_c(G)$ is closed in $M(G)$. If ν is an arbitrary measure, and μ is continuous, then

$$(\mu * \nu)(\{x\}) = \int_G \mu(\{y\})d\nu(y^{-1}x) = 0$$

$$(\nu * \mu)(\{x\}) = \int_G \mu(\{y\})d\nu(xy^{-1}) = 0$$

so $M_c(G)$ is an ideal. Finally, we verify $L^1(G)$ is closed, because it is complete, and if $\nu \in M(G)$ is arbitrary, and if U has null Haar measure, then

$$(f\,dx * \nu)(U) = \int \chi_U(xy)f(x)dx\,d\nu(y) = \int_G \int_{y^{-1}U} f(x)dxd\nu(y) = 0$$

$$(\nu * f\,dx)(U) = \int \chi_U(xy)d\nu(x)f(y)dy = \int_G \int_{Ux^{-1}} f(y)dyd\nu(x) = 0$$

So $L^1(G)$ is a two-sided ideal. $\qquad\square$

If we wish to integrate by right multiplication instead of left multiplication, we find by the substitution $y \mapsto xy$ that

$$\begin{aligned}
(f * g)(x) &= \int f(y)g(y^{-1}x)dy \\
&= \int\int f(xy)g(y^{-1})dy \\
&= \int\int \frac{f(xy^{-1})g(y)}{\Delta(y)}dy
\end{aligned}$$

Observe that

$$f * g = \int f(y)L_{y^{-1}}g\,dy$$

which can be interpreted as a vector valued integral, since for $\phi \in L^\infty(\mu)$,

$$\int (f * g)(x)\phi(x)dx = \int f(y)g(y^{-1}x)\phi(x)dxdy$$

so we can see convolution as a generalized 'averaging' of translate of g with respect to the values of f. If G is commutative, this is the same as

the averaging of translates of f, but not in the noncommutative case. It then easily follows from operator computations $L_z(f * g) = (L_z f) * g$, and $R_z(f * g) = f * (R_z g)$, or from the fact that

$$(f * g)(zx) = \int f(y)g(y^{-1}zx)dy = \int f(zy)g(y^{-1}x)dy = [(L_z f) * g](x)$$

$$(f * g)(xz) = \int f(y)g(y^{-1}xz)dy = [f * (R_z g)](x)$$

Convolution can also be applied to the other L^p spaces, but we have to be a bit more careful with our integration.

Theorem 9.7. *If $f \in L^1(G)$ and $g \in L^p(G)$, then $f * g$ is defined for almost all x, $f * g \in L^p(G)$, and $\|f * g\|_p \leqslant \|f\|\|g\|_p$. If G is unimodular, then the same results hold for $g * f$, or if G is not unimodular and f has compact support.*

Proof. We use Minkowski's inequality to find

$$\|f * g\|_p = \left(\int \left| \int f(y)|g(y^{-1}x)dy \right|^p dx \right)^{1/p}$$

$$\leqslant \int |f(y)| \left(\int |g(y^{-1}x)|^p dx \right)^{1/p} dy$$

$$= \|f\|_1 \|g\|_p$$

If G is unimodular, then

$$\|g * f\|_p = \left(\int \left| \int g(xy^{-1})f(y)dy \right|^p dx \right)^{1/p}$$

and we may apply the same trick as used before.

If f has compact support K, then $1/\Delta$ is bounded above by $M > 0$ on K and

$$\|g * f\|_p = \left(\int \left| \int \frac{g(xy^{-1})f(y)}{\Delta(y)}dy \right|^p dx \right)^{1/p}$$

$$\leqslant \int \left(\int \left| \frac{g(xy^{-1})f(y)}{\Delta(y)} \right|^p dx \right)^{1/p} dy$$

$$= \|g\|_p \int_K \frac{|f(y)|}{\Delta(y)}d\mu(y)$$

$$\leqslant M\|g\|_p\|f\|_1$$

which shows that $g * f$ is defined almost everywhere. □

Theorem 9.8. *If G is unimodular, $f \in L^p(G)$, $g \in L^q(G)$, and $p = q^*$, then $f * g \in C_0(G)$ and $\|f * g\|_\infty \leqslant \|f\|_p \|g\|_q$.*

Proof. First, note that

$$|(f * g)(x)| \leqslant \int |f(y)||g(y^{-1}x)| dy$$

$$\leqslant \|f\|_p \left(\int |g(y^{-1}x)|^q dy \right)^{1/q}$$

$$= \|f\|_p \|g\|_q$$

For each x and y, applying Hölder's inequality, we find

$$|(f * g)(x) - (f * g)(y)| \leqslant \int |f(z)||g(z^{-1}x) - g(z^{-1}y)| dz$$

$$\leqslant \|f\|_p \left(\int |g(z^{-1}x) - g(z^{-1}y)|^q dz \right)^{1/q}$$

$$= \|f\|_p \left(\int |g(z) - g(zx^{-1}y)|^q dz \right)^{1/q}$$

$$= \|f\|_p \|g - R_{x^{-1}y}g\|_q$$

Thus to prove continuity (and in fact uniform continuity), we need only prove that $\|g - R_x g\|_q \to 0$ for $q \neq \infty$ as $x \to \infty$ or $x \to 0$. This is the content of the next lemma. □

We now show that the map $x \mapsto L_x$ is a continuous operation from G to the weak $*$ topology on the L_p spaces, for $p \neq \infty$. It is easily verified that translation is not continuous on L_∞, by taking a suitable bumpy function.

Theorem 9.9. *If $p \neq \infty$, then $\|g - R_x g\|_p \to 0$ and $\|g - L_x g\|_p \to 0$ as $x \to 0$.*

Proof. If $g \in C_c(G)$, then one verifies the theorem by using left and right uniform continuity. In general, we let $g_i \in C_c(G)$ be a sequence of functions converging to g in the L_p norm, and we then find

$$\|g - L_x g\|_p \leqslant \|g - g_i\|_p + \|g_i - L_x g_i\|_p + \|L_x(g_i - g)\|_p = 2\|g - g_i\|_p + \|g_i - L_x g_i\|_p$$

115

Taking i large enough, x small enough, we find $\|g - L_x g\|_p \to 0$. The only problem for right translation is the appearance of the modular function

$$\|R_x(g - g_i)\|_p = \frac{\|g - g_i\|_p}{\Delta(x)^{1/p}}$$

If we assume our x values range only over a compact neighbourhood K of the origin, we find that $\Delta(x)$ is bounded below, and hence $\|R_x(g-g_i)\|_p \to 0$, which effectively removes the problems in the proof. $\qquad\square$

Since the map is linear, we have verified that the map $x \mapsto L_x f$ is uniformly continuous in L^p for each $f \in L^p$. In the case where $p = \infty$, the same theorem cannot hold, but we have even better conditions that do not even require unimodularity.

Theorem 9.10. *If $f \in L^1(G)$ and $g \in L^\infty(G)$, then $f * g$ is left uniformly continuous, and $g * f$ is right uniformly continuous.*

Proof. We have

$$\|L_z(f * g) - (f * g)\|_\infty = \|(L_z f - f) * g\|_\infty \leqslant \|L_z f - f\|_1 \|g\|_\infty$$

$$\|R_z(g * f) - (g * f)\|_\infty = \|g * (R_z f - f)\|_\infty \leqslant \|g\|_\infty \|R_z f - f\|_1$$

and both integrals converge to zero as $z \to 1$. $\qquad\square$

The passage from $M(G)$ to $L^1(G)$ removes an identity from the Banach algebra in question (except if G is discrete), but there is always a way to approximate an identity.

Theorem 9.11. *For each neighbourhood U of the origin, pick a function $f_U \in (L^1)^+(G)$, with $\int \phi_U = 1$, $supp(f_U) \subset U$. Then if g is any function in $L^p(G)$,*

$$\|f_U * g - g\|_p \to 0$$

where we assume g is left uniformly continuous if $p = \infty$, and if f_U is viewed as a net with neighbourhoods ordered by inclusion. If in addition $f_U(x) = f_U(x^{-1})$, then $\|g * f_U - g\|_p \to 0$, where g is right uniformly continuous for $p = \infty$.

Proof. Let us first prove the theorem for $p \neq \infty$. If $g \in C_c(G)$ is supported on a compact K, and if U is small enough that $|g(y^{-1}x) - g(x)| < \varepsilon$ for $y \in U$, then because $\int_U f_U(y) = 1$, and by applying Minkowski's inequality, we find

$$\|f_U * g - g\|_p = \left(\int \left| \int f_U(y)[g(y^{-1}x) - g(x)]dy \right|^p dx \right)^{1/p}$$

$$\leqslant \int f_U(y) \left(\int |g(y^{-1}x) - g(x)|^p dx \right)^{1/p} dy$$

$$\leqslant 2\mu(K)\varepsilon \int f_U(y)dy \leqslant 2\mu(K)\varepsilon$$

Results are then found for all of L^p by taking limits. If g is left uniformly continuous, then we may find U such that $|g(y^{-1}x) - g(x)| < \varepsilon$ for $y \in U$ then

$$|(f_U * g - g)(x)| = \left| \int f_U(y)[g(y^{-1}x) - g(x)] \right| \leqslant \varepsilon$$

For right convolution, we find that for $g \in C_c(G)$, where $|g(xy) - g(x)| < \varepsilon$ for $y \in U$, then

$$\|g * f_U - g\|_p = \left(\int \left| \int g(y)f_U(y^{-1}x) - g(x)dy \right|^p dx \right)^{1/p}$$

$$= \left(\int \left| \int [g(xy) - g(x)]f_U(y)dy \right|^p dx \right)^{1/p}$$

$$\leqslant \int \left(\int |g(xy) - g(x)|^p dx \right)^{1/p} f_U(y)dy$$

$$\leqslant \mu(K)\varepsilon \int f_U(y)(1 + \Delta(y))dy$$

$$= \mu(K)\varepsilon + \mu(K)\varepsilon \int f_U(y)\Delta(y)dy$$

We may always choose U small enough that $\Delta(y) < \varepsilon$ for $y \in U$, so we obtain a complete estimate $\mu(K)(\varepsilon + \varepsilon^2)$. If g is right uniformly continuous, then choosing U for which $|g(xy) - g(x)| < \varepsilon$, then

$$|(g * f_U - g)(x)| = \left| \int [g(xy) - g(x)]f_U(y)dy \right| \leqslant \varepsilon$$

117

We will always assume from hereon out that the approximate identities in $L^1(G)$ are of this form. $\qquad\square$

We have already obtained enough information to characterize the closed ideals of $L^1(G)$.

Theorem 9.12. *If V is a closed subspace of $L^1(G)$, then V is a left ideal if and only if it is closed under left translations, and a right ideal if and only if it is closed under right translations.*

Proof. If V is a closed left ideal, and f_U is an approximate identity at the origin, then for any g,

$$\|(L_z f_U) * g - L_z g\|_1 = \|L_z(f_U * g - g)\|_1 = \|f_U * g - g\| \to 0$$

so $L_z g \in V$. Conversely, if V is closed under left translations, $g \in L^1(G)$, and $f \in V$, then

$$g * f = \int g(y) L_{y^{-1}} f \, dy$$

which is in the closed linear space of the translates of f. Right translation is verified very similarily. $\qquad\square$

9.4   The Riesz Thorin Theorem

We finalize our basic discussion by looking at convolutions of functions in $L^p * L^q$. Certainly $L^p * L^1 \subset L^p$, and $L^p * L^q \subset L^\infty$ for $q = p^*$. To prove general results, we require a foundational interpolation result.

Theorem 9.13. *For any $0 < \theta < 1$, and $0 < p, q \leqslant \infty$. If we define*

$$1/r_\theta = (1 - \theta)/p + \theta/q$$

to be the inverse interpolation of the two numbers. Then

$$\|f\|_{r_\theta} \leqslant \|f\|_p^{1-\theta} \|f\|_q^\theta$$

Proof. We apply Hölder's inequality to find

$$\|f\|_{r_\theta} \leqslant \|f\|_{p/(1-\theta)} \|f\|_{q/\theta} = \left(\int |f|^{p/(1-\theta)} \right)^{(1-\theta)/p} \left(\int |f|^{q/\theta} \right)^{\theta/q}$$

so it suffices to prove $\|f\|_{p/(1-\theta)} \leqslant \|f\|_p^{1-\theta}$, $\|f\|_{q/\theta} \leqslant \|f\|_q^{\theta}$.

The map $x \mapsto x^p$ is concave for $0 < p < 1$, so we may apply Jensen's inequality in reverse to conclude

$$\left(\int |f|^{p/(1-\theta)} \right)^{(1-\theta)/p} \leqslant \left(\int |f|^p \right)^{1/p}$$

$\square$

The Riesz Thorin interpolation theorem then implies $L^p * L^q \subset L^r$, for $p^{-1} + q^{-1} = 1 + r^{-1}$. However, these estimates only guarantee $L^1(G)$ is closed under convolution. If G is compact, then $L_p(G)$ is closed under convolution for all p (TODO). The L_p conjecture says that this is true if and only if G is compact. This was only resolved in 1990.

9.5 Homogenous Spaces and Haar Measures

The natural way for a locally compact topological group G to act on a locally compact Hausdorff space X is via a representation of G in the homeomorphisms of X. We assume the action is transitive on X. The standard example are the action of G on G/H, where H is a closed subspace. These are effectively all examples, because if we fix $x \in X$, then the map $y \mapsto yx$ induces a continuous bijection from G/H to X, where H is the set of all y for which $yx = x$. If G is a σ compact space, then this map is a homeomorphism.

Theorem 9.14. *If a σ compact topological group G has a transitive topological action on X, and $x \in X$, then the continuous bijection from G/G_x to X is a homeomorphism.*

Proof. It suffices to show that the map $\phi : G \to X$ is open, and we need only verify this for the neighbourhood basis of compact neighbourhoods V of the origin by properties of the action. G is covered by countably many translates $y_1 V, y_2 V, \ldots$, and since each $\phi(y_k V) = y_k \phi(V)$ is closed (compactness), we conclude that $y_k \phi(V)$ has non-empty interior for some y_k, and hence $\phi(V)$ has a non-empty interior point $\phi(y_0)$. But then for any $y \in V$, y is in the interior of $\phi(yVy_0^{-1}) \subset \phi(VVy_0^{-1})$, so if we fix a compact U, and find V with $V^3 \subset U$, we have shown $\phi(U)$ is open in X. $\square$

We shall say a space X is homogenous if it is homeomorphic to G/H for some group action of G over X. The H depends on our choice of basepoint x, but only up to conjugation, for if if we switch to a new basepoint y, and c maps x to y, then $ax = x$ holds if and only if $cac^{-1}y = y$. The question here is to determine whether we have a G-invariant measure on X. This is certainly not always possible. If we had a measure on $\mathbf{R}$ invariant under the affine maps $ax + b$, then it would be equal to the Haar measure by uniqueness, but the Haar measure is not invariant under dilation $x \mapsto ax$.

Let G and H have left Haar measures μ and ν respectively, denote the projection of G onto G/H as $\pi : G \to G/H$, and let Δ_G and Δ_H be the respective modular functions. Define a map $P : C_c(G) \to C_c(G/H)$ by

$$(Pf)(Hx) = \int_H f(xy)d\nu(y) = \int_H$$

this is well defined by the invariance properties of ν. Pf is obviously continuous, and $\operatorname{supp}(Pf) \subset \pi(\operatorname{supp}(f))$. Moreover, if $\phi \in C(G/H)$ we have

$$P((\phi \circ \pi) \cdot f)(Hx) = \phi(xH) \int_H f(xy)d\nu(y)$$

so $P((\phi \circ \pi) \cdot f) = \phi P(f)$.

Lemma 9.15. *If E is a compact subset of G/H, there is a compact $K \subset G$ with $\pi(K) = E$.*

Proof. Let V be a compact neighbourhood of the origin, and cover E by finitely many translates of $\pi(V)$. We conclude that $\pi^{-1}(E)$ is covered by finitely many of the translates, and taking the intersections of these translates with $\pi^{-1}(E)$ gives us the desired K. $\square$

Lemma 9.16. *A compact $F \subset G/H$ gives rise to a function $f \geqslant 0$ in $C_c(G)$ such that $Pf = 1$ on E.*

Proof. Let E be a compact neighbourhood containing F, and if $\pi(K) = E$, there is a function $g \in C_c(G)$ with $g > 0$ on K, and $\phi \in C_c(G/H)$ is supported on E and $\phi(x) = 1$ for $x \in F$, let

$$f = \frac{\phi \circ \pi}{Pg \circ \pi}g$$

120

Hence

$$Pf = \frac{\phi}{Pg}Pg = \phi$$

$\square$

Lemma 9.17. *If $\phi \in C_c(G/H)$, there is $f \in C_c(G)$ with $Pf = \phi$, and $\pi(\text{supp} f) = \text{supp}(\phi)$, and also $f \geqslant 0$ if $\phi \geqslant 0$.*

Proof. There exists $g \geqslant 0$ in $C_c(G/H)$ with $Pg = 1$ on $\text{supp}(\phi)$, and then $f = (\phi \circ \pi)g$ satisfies the properties of the theorem. $\square$

We can now provide conditions on the existence of a measure on G/H.

Theorem 9.18. *There is a G invariant measure ψ on G/H if and only if $\Delta_G = \Delta_H$ when restricted to H. In this case, the measure is unique up to a common factor, and if the factor is chosen, we have*

$$\int_G f \, d\mu = \int_{G/H} Pf \, d\psi = \int_{G/H} \int_H f(xy) d\nu(y) d\psi(xH)$$

Proof. Suppose ψ existed. Then $f \mapsto \int Pf d\psi$ is a non-zero left invariant positive linear functional on G/H, so $\int Pf d\psi = c \int f d\mu$ for some $c > 0$. Since $P(C_c(G)) = C_c(G/H)$, we find that ψ is determined up to a constant factor. We then compute, for $y \in H$,

$$\Delta_G(y) \int f(x) d\mu(x) = \int f(xy^{-1}) d\mu(x)$$

$$= \int_{G/H} \int_H f(xzy^{-1}) d\nu(z) d\psi(xH)$$

$$= \Delta_H(y) \int_{G/H} \int_H f(xz) d\nu(z) d\psi(xH)$$

$$= \Delta_H(y) \int f(x) d\mu(x)$$

Hence $\Delta_G = \Delta_H$. Conversely, suppose $\Delta_G = \Delta_H$. First, we claim if $f \in C_c(G)$ and $Pf = 0$, then $\int f d\mu = 0$. Indeed if $P\phi = 1$ on $\pi(\text{supp} f)$ then

$$0 = Pf(xH) = \int_H f(xy) d\nu(y) = \Delta_G(y^{-1}) \int_H f(xy^{-1}) d\nu(y)$$

121

so

$$0 = \int_G \int_H \Delta_G(y^{-1})\phi(x)f(xy^{-1})d\nu(y)d\mu(x)$$
$$= \int_H \int_G \phi(xy)f(x)d\mu(x)d\nu(y)$$
$$= \int_G P\phi(xH)f(x)d\mu(x)$$
$$= \int_G f(x)d\mu(x)$$

This implies that if $Pf = Pg$, then $\int_G f = \int_G g$. Thus the map $Pf \mapsto \int_G f$ is a well defined G invariant positive linear functional on $C_c(G/H)$, and we obtain a Radon measure from the Riesz representation theorem. $\square$

If H is compact, then Δ_G and Δ_H are both continuous homomorphisms from H to $\mathbf{R}^+$, so Δ_G and Δ_H are both trivial, and we conclude a G invariant measure exists on G/H.

9.6 Function Spaces In Harmonic Analysis

There are a couple other function spaces that are interesting in Harmonic analysis. We define $AP(G)$ to be the set of all almost periodic functions, functions $f \in L^\infty(G)$ such that $\{L_x f : x \in G\}$ is relatively compact in $L^\infty(G)$. If this is true, then $\{R_x f : x \in G\}$ is also relatively compact, a rather deep theorem. If we define $WAP(G)$ to be the space of weakly almost periodic functions (the translates are relatively compact in the weak topology). It is a deep fact that $WAP(G)$ contains $C_0(G)$, but $AP(G)$ can be quite small. The reason these function spaces are almost periodic is that in the real dimensional case, $AP(\mathbf{R})$ is just the closure of the set of all trigonometric polynomials.

Chapter 10

The Character Space

Let G be a locally compact group. A character on G is a *continuous* homomorphism from G to $\mathbf{T}$. The space of all characters of a group will be denoted $\Gamma(G)$.

Example. *Determining the characters of* $\mathbf{T}$ *involves much of classical Fourier analysis. Let* $f : \mathbf{T} \to \mathbf{T}$ *be an arbitrary continuous character. For each* $w \in \mathbf{T}$, *consider the function* $g(z) = f(zw) = f(z)f(w)$. *We know the Fourier series acts nicely under translation, telling us that*

$$\hat{g}(n) = w^n \hat{f}(n)$$

Conversely, since $g(z) = f(z)f(w)$,

$$\hat{g}(n) = f(w)\hat{f}(n)$$

Thus $(w^n - f(w))\hat{f}(n) = 0$ *for all* $w \in \mathbf{T}$, $n \in \mathbf{Z}$. *Fixing* n, *we either have* $f(w) = w^n$ *for all* w, *or* $\hat{f}(n) = 0$. *This implies that if* $f \neq 0$, *then* f *is just a power map for some* $n \in \mathbf{Z}$.

Example. *The characters of* $\mathbf{R}$ *are of the form* $t \mapsto e^{ti\xi}$, *for* $\xi \in \mathbf{R}$. *To see this, let* $e : \mathbf{R} \to \mathbf{T}$ *be an arbitrary character. Define*

$$F(x) = \int_0^x e(t)dt$$

Then $F'(x) = e(x)$. *Since* $e(0) = 1$, *for suitably small* δ *we have*

$$F(\delta) = \int_0^\delta e(t)dt = c > 0$$

and then it follows that

$$F(x + \delta) - F(x) = \int_x^{x+\delta} e(t)dt = \int_0^\delta e(x+t)dt = ce(x)$$

As a function of x, F is differentiable, and by the fundamental theorem of calculus,

$$\frac{dF(x+\delta) - F(x)}{dt} = F'(x+\delta) - F'(x) = e(x+\delta) - e(x)$$

This implies the right side of the above equation is differentiable, and so

$$ce'(x) = e(x+\delta) - e(x) = e(x)[e(\delta) - 1]$$

Implying $e'(x) = Ae(x)$ for some $A \in \mathbf{C}$, so $e(x) = e^{Ax}$. We require that $e(x) \in \mathbf{T}$ for all x, so $A = \xi i$ for some $\xi \in \mathbf{R}$.

Example. *Consider the group $\mathbf{R}^+$ of positive real numbers under multiplication. The map $x \mapsto \log x$ is an isomorphism from $\mathbf{R}^+$ and $\mathbf{R}$, so that every character on $\mathbf{R}^+$ is of the form $e^{is\log(x)} = x^{is}$, for some $s \in \mathbf{R}$. The character group is then $\mathbf{R}$, since $x^{is}x^{is'} = x^{i(s+s')}$.*

There is a connection between characters on G and characters on $L^1(G)$ that is invaluable to the generalization of Fourier analysis to arbitrary groups.

Theorem 10.1. *For any character $\phi : G \to \mathbf{C}$, the map*

$$\varphi(f) = \int \frac{f(x)}{\phi(x)}dx$$

is a non-zero character on the convolution algebra $L^1(G)$, and all characters arise this way.

Proof. The induced map is certainly linear, and

$$\varphi(f * g) = \int\int \frac{f(y)g(y^{-1}x)}{\phi(x)}dydx$$
$$= \int\int \frac{f(y)g(x)}{\phi(y)\phi(x)}dydx$$
$$= \int \frac{f(y)}{\phi(y)}dy \int \frac{g(x)}{\phi(x)}dx$$

124

Since ϕ is continuous, there is a compact subset K of G where $\phi > \varepsilon$ for some $\varepsilon > 0$, and we may then choose a positive f supported on K in such a way that $\varphi(f)$ is non-zero.

The converse results from applying the duality theory of the L^p spaces. Any character on $L^1(G)$ is a linear functional, hence is of the form

$$f \mapsto \int f(x)\phi(x)dx$$

for some $\phi \in L^\infty(G)$. Now

$$\int\int f(y)g(x)\phi(yx)dydx = \int\int f(y)g(y^{-1}x)\phi(x)dydx$$
$$= \int f(x)\phi(x)dx \int g(y)\phi(y)dy$$
$$= \int f(x)g(y)\phi(x)\phi(y)dxdy$$

Since this holds for all functions f and g in $L^1(G)$, we must have $\phi(yx) = \phi(x)\phi(y)$ almost everywhere. Also

$$\int \varphi(f)g(y)\phi(y)dy = \varphi(f * g)$$
$$= \int\int g(y)f(y^{-1}x)\phi(x)dydx$$
$$= \int\int (L_{y^{-1}}f)(x)g(y)\phi(x)dydx$$
$$= \int \varphi(L_{y^{-1}}f)g(y)dy$$

which implies $\varphi(f)\phi(y) = \varphi(L_{y^{-1}}f)$ almost everywhere. Since the map $\varphi(L_{y^{-1}}f)/\varphi(f)$ is a uniformly continuous function of y, ϕ is continuous almost everywhere, and we might as well assume ϕ is continuous. We then conclude $\phi(xy) = \phi(x)\phi(y)$. Since $\|\phi\|_\infty = 1$ (this is the norm of any character operator on $L^1(G)$), we find ϕ maps into $\mathbf{T}$, for if $\|\phi(x)\| < 1$ for any particular x, $\|\phi(x^{-1})\| > 1$. $\qquad\square$

Thus there is a one-to-one correspondence with $\Gamma(G)$ and $\Gamma(L^1(G))$, which implies a connection with the Gelfand theory and the character

theory of locally compact groups. This also gives us a locally compact topological structure on $\Gamma(G)$, induced by the Gelfand representation on $\Gamma(L^1(G))$. A sequence $\phi_i \to \phi$ if and only if

$$\int \frac{f(x)}{\phi_i(x)} dx \to \int \frac{f(x)}{\phi(x)} dx$$

for all functions $f \in L^1(G)$. This actually makes the map

$$(f, \phi) \mapsto \int \frac{f(x)}{\phi(x)} dx$$

a jointly continuous map, because as we verified in the proof above,

$$\hat{f}(\phi)\phi(y) = \widehat{L_y f}(\phi)$$

And the map $y \mapsto L_y f$ is a continuous map into $L^1(G)$. If $K \subset G$ and $C \subset \Gamma(G)$ are compact, this allows us to find open sets in G and $\Gamma(G)$ of the form

$$\{\gamma : \|1 - \gamma(x)\| < \varepsilon \text{ for all } x \in K\} \quad \{x : \|1 - \gamma(x)\| < \varepsilon \text{ for all } \gamma \in C\}$$

And these sets actually form a base for the topology on $\Gamma(G)$.

Theorem 10.2. *If G is discrete, $\Gamma(G)$ is compact, and if G is compact, $\Gamma(G)$ is discrete.*

Proof. If G is discrete, then $L^1(G)$ contains an identity, so $\Gamma(G) = \Gamma(L^1(G))$ is compact. Conversely, if G is compact, then it contains the constant 1 function, and

$$\hat{1}(\phi) = \int \frac{dx}{\phi(x)}$$

And

$$\frac{1}{\phi(y)}\hat{1}(\phi) = \int \frac{dx}{\phi(yx)} = \int \frac{dx}{\phi(x)} = \hat{1}(\phi)$$

So either $\phi(y) = 1$ for all y, and it is then verified by calculation that $\hat{1}(\phi) = 1$, or $\hat{1}(\phi) = 0$. Since $\hat{1}$ is continuous, the trivial character must be an open set by itself, and hence $\Gamma(G)$ is discrete. □

Given a function $f \in L^1(G)$, we may take the Gelfand transform, obtaining a function on $C_0(\Gamma(L^1(G)))$. The identification then gives us a function on $C_0(\Gamma(G))$, if we give $\Gamma(G)$ the topology induced by the correspondence (which also makes $\Gamma(G)$ into a topological group). The formula is

$$\widehat{f}(\phi) = \phi(f) = \int \frac{f(x)}{\phi(x)}$$

This gives us the classical correspondence between $L^1(\mathbf{T})$ and $C_0(\mathbf{Z})$, and $L^1(\mathbf{R})$ and $C_0(\mathbf{R})$, which is just the Fourier transform. Thus we see the Gelfand representation as a natural generalization of the Fourier transform. We shall also denote the Fourier transform by $\mathcal{F}$, especially when we try and understand it's properties as an operator. Gelfand's theory (and some basic computation) tells us instantly that

- $\widehat{f * g} = \widehat{f}\widehat{g}$ (The transform is a homomorphism).

- $\mathcal{F}$ is norm decreasing and therefore continuous: $\|\widehat{f}\|_\infty \leqslant \|f\|_1$.

- If G is unimodular, and $\gamma \in \Gamma(G)$, then $(f * \gamma)(x) = \gamma(x)\widehat{f}(\gamma)$.

Whenever we integrate a function with respect to the Haar measure, there is a natural generalization of the concept to the space of all measures on G. Thus, for $\mu \in M(G)$, we define

$$\widehat{\mu}(\phi) = \int \frac{dx}{\phi(x)}$$

which we call the **Fourier-Stieltjes transform** on G. It is essentially an extension of the Gelfand representation on $L^1(G)$ to $M(G)$. Each $\widehat{\mu}$ is a bounded, uniformly continuous function on $\Gamma(G)$, because the transform is still contracting, i.e.

$$\left| \int \frac{d\mu(x)}{\phi(x)} dx \right| \leqslant \|\mu\|$$

It is uniformly continuous, because

$$(L_\nu \widehat{\mu} - \widehat{\mu})(\phi) = \int \frac{1 - \nu(x)}{\nu(x)\phi(x)} d\mu(x)$$

127

The regularity of μ implies that there is a compact set K such that $|\mu|(K^c) < \varepsilon$. If $\nu_i \to 0$, then eventually we must have $|\nu_i(x) - 1| < \varepsilon$ for all $x \in K$, and then

$$|(L_\nu \widehat{\mu} - \widehat{\mu})(\phi)| \leqslant 2|\mu|(K^c) + \varepsilon\|\mu\| \leqslant \varepsilon(2 + \|\mu\|)$$

Which implies uniform continuity.

Let us consider why it is natural to generalize operators on $L^1(G)$ to $M(G)$. The first reason is due to the intuition of physicists; most of classical Fourier analysis emerged from physical considerations, and it is in this field that $L^1(G)$ is often confused with $M(G)$. Take, for instance, the determination of the electric charge at a point in space. To determine this experimentally, we take the ratio of the charge over some region in space to the volume of the region, and then we limit the size of the region to zero. This is the historical way to obtain the density of a measure with respect to the Lebesgue measure, so that the function we obtain can be integrated to find the charge over a region. However, it is more natural to avoid taking limits, and to just think of charge as an element of $M(\mathbf{R}^3)$. If we consider a finite number of discrete charges, then we obtain a discrete measure, whose density with respect to the Lebesgue measure does not exist. This doesn't prevent physicists from trying, so they think of the density obtained as shooting off to infinity at points. Essentially, we obtain the Dirac Delta function as a 'generalized function'. This is fine for intuition, but things seem to get less intuitive when we consider the charge on a subsurface of $\mathbf{R}^3$, where the 'density' is 'dirac'-esque near the function, where as measure theoretically we just obtain a density with respect to the two-dimensional Hausdorff measure on the surface. Thus, when physicists discuss quantities as functions, they are really thinking of measures, and trying to take densities, where really they may not exist.

There is a more austere explanation, which results from the fact that, with respect to integration, $L^1(G)$ is essentially equivalent to $M(G)$. Notice that if $\mu_i \to \mu$ in the weak-$*$ topology, then $\widehat{\mu}_i \to \widehat{\mu}$ pointwise, because

$$\int \frac{d\mu_i(x)}{\phi(x)} \to \int \frac{d\mu(x)}{\phi(x)}$$

(This makes sense, because weak-$*$ convergence is essentially pointwise convergence in $M(G)$). Thus the Fourier-Stietjes transform is continuous with respect to these topologies. It is the unique continuous extension of the Fourier transform, because

Theorem 10.3. $L^1(G)$ *is weak-* dense in* $M(G)$.

Proof. First, note that the Dirac delta function can be weak-* approximated by elements of $L^1(G)$, since we have an approximate identity in the space.

First, note that if $\mu_i \to \mu$, then $\mu_i * \nu \to \mu * \nu$, because

$$\int f \, d(\mu_i * \nu) = \int \int f(xy) \, d\mu_i(x) \, d\nu(y)$$

The functions $y \mapsto \int f(xy) \, d\mu_i(x)$ converge pointwise to $\int f(xy) \, d\mu(y)$. Since

$$\left| \int f(xy) \, d\mu_i(x) \right| \leqslant \|f\|_1 \|\mu_i\|$$

If i is taken large enough that $\qquad\qquad\qquad\qquad\qquad\qquad\square$

If $\phi_\alpha \to \phi$, in the sense that $\phi_\alpha(x) \to \phi(x)$ for all $x \in G$, then, because $\|\phi_\alpha(x)\| = 1$ for all x, we can apply the dominated convergence theorem on any compact subset K of G to conclude

$$\int_K \frac{d\mu(x)}{\phi_\alpha(x)} \to \int_K \frac{d\mu(x)}{\phi(x)}$$

It is immediately verified to be a map into $L^1(\Gamma(G))$, because

$$\int \left| \int \frac{d\mu(x)}{\phi(x)} \right| d\phi \leqslant \int \int \|\mu\|$$

The formula above immediately suggests a generalization to a transform on $M(G)$. For $\nu \in M(G)$, we define

$$\mathcal{F}(\nu)(\phi) = \int \frac{d\nu}{\phi}$$

If $\mathcal{G} : L^1(G) \to C_0(\Gamma(G))$ is the Gelfand transform, then the transform induces a map $\mathcal{G}^* : M(\Gamma(G)) \to L^\infty(G)$.

The duality in class-ical Fourier analysis is shown through the inversion formulas. That is, we have inversion functions

$$\mathcal{F}^{-1}(\{a_k\}) = \sum a_k e^{kit} \qquad \mathcal{F}^{-1}(f)(x) = \int f(t) e^{2\pi i x t}$$

which reverses the fourier transform on $\mathbf{T}$ and $\mathbf{R}$ respectively, on a certain subclass of L^1. One of the challenges of Harmonic analysis is trying to find where this holds for the general class of measurable functions.

The first problem is to determine surjectivity. We denote by $A(G)$ the space of all continuous functions which can be represented as the fourier transform of some function in $L^1(G)$. It is to even determine $A(\mathbf{T})$, the most basic example. $A(G)$ always separates the points of $\Gamma(G)$, by Gelfand theory, and if G is unimdoular, then it is closed under conjugation. If we let $g(x) = \overline{f(x^{-1})}$, we find

$$\mathcal{F}(g)(\phi) = \int \frac{g(x)}{\phi(x)} dx = \overline{\int \frac{f(x^{-1})}{\phi(x^{-1})} dx} = \overline{\int \frac{f(x)}{\phi(x)} dx} = \overline{\mathcal{F}(f)(\phi)}$$

so that by the Stone Weirstrass theorem $A(G)$ is dense in $C_0(\Gamma(L^1(G)))$.

Chapter 11

Banach Algebra Techniques

In the mid 20th century, it was realized that much of the analytic information about a topological group can be captured in various C^* algebras related to the group. For instance, consider the Gelfand space of $L^1(\mathbf{Z})$ is $\mathbf{T}$, which represents the fact that one can represent functions over $\mathbf{T}$ as sequences of numbers. Similarily, we find the characters of $L^1(\mathbf{R})$ are the maps $f \mapsto \widehat{f}(x)$, so that the Gelfand space of $\mathbf{R}$ is $\mathbf{R}$, and the Gelfand transform is the Fourier transform on this space. For a general G, we may hope to find a generalized Fourier transform by understanding the Gelfand transform on $L^1(G)$. We can also generalize results by extending our understanding to the class $M(G)$ of regular, Borel measures on G.

Chapter 12

Vector Spaces

If **K** is a closed, multiplicative subgroup of the complex numbers, then **K** is also a locally compact abelian group, and we can therefore understand **K** by looking at its dual group $\mathbf{K}^*$. The map $\langle x, y \rangle = xy$ is bilinear, in the set that it is a homomorphism in the variable y for each fixed x, and a homomorphism in the variable x for each y.

    If **K** is a subfield of the complex numbers, then **K** is also an abelian group under addition, and we can consider the dual group $\mathbf{K}^*$. The inner product $\langle x, y \rangle = xy$ gives a continuous bilinear map $\mathbf{K} \times \mathbf{K} \to \mathbf{C}$, and therefore we can define $x^* \in \mathbf{K}^*$ by $x^*(y) = \langle x, y \rangle$. If $x^*(y) = xy = 0$ for all y, then in particular $x^*(1) = x$, so $x = 0$. This means that the homomorphism $\mathbf{K} \to \mathbf{K}^*$ is injective.

Chapter 13

Interpolation of Besov and Sobolev spaces

An important class of operators arise as singular integrals, that is, they arise as convolution operators T given by $T(f) = f * K$, where K is an appropriate distribution. Taking Fourier transforms, these operators can also be defined by $\widehat{T(f)} = \hat{f}\hat{K}$. The function $\hat{K}$ is known as a **Fourier multiplier**, because it operates by multiplication on the frequencies of the function f. We say $\hat{K}$ is a **Fourier multiplier on** $L^p(\mathbf{R}^n)$ if T is a bounded map from $S(\mathbf{R}^n)$ to $L^p(\mathbf{R}^n)$, under the L^p norms. Such maps clearly extend uniquely to maps from $L^p(\mathbf{R}^n)$ to $L^p(\mathbf{R}^n)$, and so we can think of T as operating by convolution on the space of L^p functions. We will denote the space of all Fourier multipliers on L^p by M_p. We define the L^p norm on these distributions K, denoted $\|K\|_p$, to be the operator norm of the associated operator T.

Example. *Consider the space M_∞. If K is a distribution in M_∞, then $\|K\|_\infty < \infty$, and since convolution commutes with translations, in the sense that $f_h * K = (f * K)_h$, then*

$$\|K\|_\infty = \sup_{f \in L^\infty(\mathbf{R}^n)} \frac{|(f * K)(0)|}{\|f\|_\infty}$$

But then the map $f \mapsto (f * K)(0)$ is a bounded operator on the space of bounded continuous functions, and so the Riesz representation says there is a bounded Radon measure μ such that

$$(f * K)(0) = \int f(-y) \, d\mu(y)$$

But now we know

$$(f * K)(x) = (f_{-x} * K)(0) = \int f(x - y) d\mu(y) = (f * \mu)(x)$$

Thus M_∞ is really just the space of all bounded Radon measures, and

$$\|K\|_\infty = \sup_{f \in L^\infty(\mathbf{R}^n)} \frac{\left|\int f(y) \, d\mu(y)\right|}{\|f\|_\infty} = \|\mu\|_1$$

so M_∞ even has the same norm as the space of all bounded Radon measures. Note that it becomes a Banach algebra under convolution of distributions, since the convolution of two bounded Radon measures is a bounded Radon measure.

Theorem 13.1. *For any $1 \leqslant p \leqslant \infty$, and $q = p^*$, then $M_p = M_q$.*

Proof. Let $f \in L^p$, and $g \in L^q$, then Hölder's inequality gives

$$|(K * f * g)(0)| \leqslant \|K * f\|_p \|g\|_q \leqslant \|K\|_p \|f\|_p \|g\|_p$$

Thus $K * g \in L_q$, and that $K \in M_q$ with $\|K\|_q \leqslant \|K\|_p$. By symmetry, we find $\|K\|_p = \|K\|_q$. □

Example. *Consider M_2. If K is a distribution with $\|f * K\|_2 \leqslant A\|f\|_2$, then Parsevel's inequality implies that*

$$\|\widehat{f}\widehat{K}\|_2 = \|f * K\|_2 \leqslant A\|f\|_2 = A\|\widehat{f}\|_2$$

so for each $\widehat{f}$, TODO: PROVE THAT THIS IS REALLY JUST THE SPACE $L^\infty(\mathbf{R}^n)$, with the supremum norm. Note that this is also a Banach algebra under pointwise multiplication.

Using the Riesz-Thorin interpolation theorem, we find that if $1/p = (1-\theta)/p_0 + \theta/p_1$, then $\|K\|_p \leqslant \|K\|_{p_0}^{1-\theta} \|K\|_{p_1}^{\theta}$, when K lies in the three spaces. In particular, $\|K\|_p$ is a decreasing function of p for $1 \leqslant p \leqslant 2$, so we find $M_1 \subset M_p \subset M_q \subset M_2$ for $1 \leqslant p < q \leqslant 2$. In particular, all Fourier multipliers can be viewed as Fourier multipliers with respect to bounded, measurable functions on L^∞. Riesz interpolation shows that each M_p is a Banach algebra under multiplication in the frequency domain, or convolution in the spatial domain.

Theorem 13.2. *Let* $T : \mathbf{R}^n \to \mathbf{R}^m$ *be a surjective affine transformation. Then the endomorphism* T^* *on* $M_p(\mathbf{R}^n)$ *defined by* $(T^* f)(\xi) = f(T(\xi))$ *is an isometry, and if* T *is a bijection, so too is* T^*.

Proof. TODO $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

The next theorem is the main tool to prove results about Sobolev and Besov space. Note that it assumes $1 < p < \infty$, and cannot be applied for $p = 1$ or $p = \infty$. The proof relies on two lemmas, the first of which is used frequently later, and the second is used universally in modern harmonic analysis.

Lemma 13.3. *There exists a Schwartz function* φ *on* $\mathbf{R}^n$ *which is supported on the annulus*

$$\{\xi : 1/2 \leqslant |\xi| \leqslant 2\}$$

is positive for $1/2 < |\xi| < 2$, *and satisfies*

$$\sum_{k=-\infty}^{\infty} \varphi(2^{-k}\xi) = 1$$

for all $\xi \neq 0$.

Lemma 13.4 (Calderon-Zygmund Decomposition)**.** *Let* $f \in L^1(\mathbf{R}^n)$, *and* $\sigma > 0$. *Then there are pairwise almost disjoint cubes* $I_1, I_2, \ldots$ *with edges parallel to the coordinate axis and*

$$\sigma < \frac{1}{|I_n|} \int_{I_n} |f(x)| \, dx \leqslant 2^n \sigma$$

and with $|f(x)| \leqslant \sigma$ *for almost all* x *outside these cubes.*

Theorem 13.5 (The Mihlin Multiplier Theorem)**.** *Let m be a bounded function on* $\mathbf{R}^n$ *which is smooth except possibly at the origin, such that*

$$\sup_{\substack{\xi \in \mathbf{R}^n \\ |\alpha| \leqslant L}} |\xi|^{|\alpha|} |(D^\alpha m)(x)| < \infty$$

Then m is an L^p *Fourier multiplier for* $1 < p < \infty$.

13.1 Besov Spaces

Recall the Schwarz function φ used to prove the Mihlin multiplier theorem. We now define functions φ_k such that

$$\widehat{\varphi_n}(\xi) = \varphi(2^{-n}\xi) \qquad \hat{\psi}(\xi) = 1 - \sum_{n=1}^{\infty} \varphi(2^{-n}\xi)$$

Thus φ_n essentially covers the annulus $2^{n-1} \leqslant |\xi| \leqslant 2^{n+1}$, and the function ψ covers the remaining low frequency parts covered in the frequency ball of radius 2. We have

$$\varphi_n(\xi) = \widetilde{\varphi_{2^{-n}}}(\xi) = 2^{dn}\breve{\varphi}(2^n\xi)$$

Given $s \in \mathbf{R}$, and $1 \leqslant p, q \leqslant \infty$, we write

$$\|f\|_{pq}^s = \|\psi * f\|_p + \left(\sum_{n=1}^{\infty} (2^{sn}\|\varphi_k * f\|_p)^q \right)^{1/q}$$

The convolution $\varphi_n * f$ essentially captures the portion of f whose frequencies lie in the annulus $2^{n-1} \leqslant |\xi| \leqslant 2^{n+1}$

13.2 Proof of The Projection Result

As with Marstrand's projection theorem, we require an energy integral variant. Rather than considering the Riesz kernel on $\mathbf{R}^n$, we consider the kernel on balls

$$K_\alpha(x) = \frac{\chi_{B(0,R)}(x)}{|x|^\alpha}$$

where R is a fixed radius. If $\alpha < \beta$, and μ is measure supported on a β dimensional subset of $\mathbf{R}^n$, then $\mu * K_\alpha \in L^\infty(\mathbf{R}^d)$ because μ cancels out the singular part of K_α. Assuming $\beta < d$, we conclude $\mu * K_\alpha \in L^1(\mathbf{R}^d)$. Applying interpolation (TODO: Which interpolation), we conclude that $\nu * K_\rho$

136

Chapter 14

The Cap Set Problem

The cap set problem comes out of additive combinatorics, whose goal is to understand additive structure in some abelian group, typically the integers. For instance, we can think of a set A as being roughly closed under addition if $|A + A| = O(|A|)$. Over rings, we can study the interplay between additive and multiplicative structure. For instance, one conjecture of Erdös and Szemerédi says that if A is a finite subset of real numbers, then $\max(|A + A|, |A \cdot A|) \gtrsim |A|^{1+c}$ for some positive $c \in (0, 1)$. The best known c so far is $c \sim 1/3$, though it is conjectured that we can take c arbitrarily close to 1. This can be seen as a discrete version of the results of Bourgain and Edgar-Miller on the Hausdorff dimensions of Borel subrings.

Theorem 14.1 (Van Der Waerden - 1927). *For any positive integes r and k, there is N such that if the integers in $[1, N]$ are given an r coloring, then there is a monochromatic k term arithmetic progression.*

The coloring itself is not so important, more just the partitioning. We just pidgeonhole, using the density of k term arithmetic progressions. This problem suggests the Ramsey type problem of determining the largest set A of the integers $[1, N]$ which does not contain k term arithmetic progressions. Behrend's theorem says we can choose A to be on the order of $N \exp(-c\sqrt{\log N})$.

Theorem 14.2 (Roth - 1956). *If A is a set of integers in $[1, N]$ which is free of three term arithmetic progressions, then $|A| = O(N/\log \log N)$.*

Szemerédi proved that if A is free of k term arithmetic progressions, $|A| = o(N)$. If Erdös Turan, if $\sum_{x \in X} 1/x$ diverges, then X contains arbitrarily long arithmetic progressions. For now, we'll restrict our attention to three term arithmetic progressions. Heath and Brown showed that three term arithmetic progresisons are $O(N/(\log N)^c)$ for some constant c. In 2016, the best known bound was given by Bloom, given $O(N(\log\log N)^4/\log N)$.

One way we can simplify our problem is to note that avoiding three term arithmetic progressions is a local issue, so we can embed $[1, N]$ in $\mathbf{Z}/M\mathbf{Z}$ for suitably large M, and we lose none of the problems we had over the integers. A heuristic is that it is easier to solve these kind of problems in $\mathbf{F}_p^n$, where p is small and n is large, which should behave like $\{1, \ldots, p^n\}$. This leads naturally to the cap set problem.

Theorem 14.3 (Cap Set Problem). *What is the largest subset of $\mathbf{F}_3^n$ containing no three term arithmetic progressions?*

We look at $\mathbf{F}_3$ because it is the smallest case where three term arithmetic progressions become important.

Theorem 14.4 (Meschulam - 1995). *Let $A \subset \mathbf{F}_3^n$ be a cap set. Then $|A| = O(3^n/n)$. This is analogous to a $N/\log N$ case over the integers, giving evidence that the finite field case is easier.*

In 2012, Bateman and Katz showed $|A| = O(3^n/n^{1+\varepsilon})$ for some $c > 0$. This was a difficult proof. In 2016, there was a more significant breakthrough, which gave an easy proof using the polynomial method of an exponentially small bound of c^n, where $c < 4$, over $\mathbf{Z}/4\mathbf{Z}$, and a week later Ellenberg-Gijswijt used this argument in the $\mathbf{F}_3$ case to prove that if A is a capset in $\mathbf{F}_3$,then $|A| = O(c^n)$, for $c = 2.7551\ldots$.

The idea of the polynomial method is to take combinatorial information about some set, encode it as some algebraic structura information, and then apply the theory of polynomials to encode this algebraic information and use it to limit and enable certain properties to occur.

If V is the space of polynomials of degree d vanishing on a set A, then we know $\dim V \geqslant \dim \mathcal{P}_d - |A|$. This gives a lower bound on the size of A, whereas we want a lower bound. To get an upper bound, we take $|A|^c$ instead, which shows

$$\dim V \geqslant \dim \mathcal{P}_d + |A| - 3^n$$

whichs gives $|A| \leqslant 3^n + \dim V - \dim \mathcal{P}_d$. Now using linear algebra, we can find a polynomial P vanishing on A^c with support of cardinality greater than or equal to $\dim V$, hence

$$|A| \leqslant 3^n - \dim \mathcal{P}_d + \max |\mathrm{supp}(P)|$$

It follows that A is a cap set if and only if $x + y = 2z$, or $x + y + z = 0$ holds if and only if $x = y = z$. This is an algebraic property which says directly that A has no nontrivial three term arithmetic progressions. Thus for any $a_1, \ldots, a_m \in A$, $P(-a_i - a_j) = 0$ when $i \neq j$. Equivalently, this means $P(-a_i - a_j) \neq 0$ when $i = j$. This suggests we consider the $|A|$ by $|A|$ matrix M with $M_{ij} = P(-a_i - a_j)$. This is a diagonal matrix, with $M_{ii} = P(a_i)$. Thus the rank of this matrix is the dimension of the support of P, so it suffices to upper bound the rank of M. The key observation, where we now explicitly employ the fact that P is a polynomial, is that $P(-x - y)$ is a polynomial in $2n$ variables $x, y \in \mathbf{F}_3^n$,