

Harmonic Analysis

Jacob Denson

August 7, 2019

Table Of Contents

I	Classical Fourier Analysis	2
1	Springs, Strings, and Symmetry	4
1.1	The Wave Equation	4
1.2	Obtaining the Fourier Coefficients	6
1.3	The Heat Equation	8
1.4	Complex Exponentials	10
1.5	Basic Properties of Fourier Series	11
1.6	Examples of Expansions	13
2	Fourier Series	17
2.1	Do Functions Have Unique Series?	18
2.2	Quantitative Bounds on Fourier Coefficients	20
2.3	Asymptotic Decay of Fourier Series	23
2.4	Convolution and Kernel Methods	24
2.5	The Dirichlet Kernel	29
2.6	Countercultural Methods of Summation	31
2.7	Fejer's Theorem	32
2.8	The De la Vallée Poisson Kernel	33
2.9	Pointwise Behaviour at Discontinuity Points	35
2.10	Tchebychev Polynomials	37
2.11	Exponential Sums and Equidistribution	38
2.12	Abel Summation and Harmonics on the Disk	40
2.13	The Isoperimetric Inequality	42
2.14	Heat Propagation Into the Ground	42
2.15	Seafaring with Fourier	43

3	The Fourier Transform	45
3.1	Convergence Using Alternative Summation	51
3.2	The L^2 Theory	63
3.3	The Hausdorff-Young Inequality	65
3.4	The Poisson Summation Formula	66
3.5	The Gibbs Phenomenon	66
3.6	Sums of Random Variables	66
3.7	Transforms of Holomorphic Functions	68
3.8	Characteristic Functions	69
4	Finite Character Theory	70
4.1	Fourier Analysis on Cyclic Groups	71
4.2	An Arbitrary Finite Abelian Group	73
4.3	Convolutions	75
4.4	The Fast Fourier Transform	77
4.5	Dirichlet's Theorem	78
5	Applications	80
5.1	The Wirtinger Inequality on an Interval	80
5.2	Energy Preservation in the String equation	81
5.3	Harmonic Functions	81
II	Euclidean Harmonic Analysis	82
6	Monotone Rearrangement Invariant Norms	84
6.1	The L^p norms	85
6.2	Weak and Lorentz Norms	94
6.3	Dyadic Layer Cake Decompositions	98
6.4	Normability of the Lorentz Spaces	105
6.5	Metric Structure of Norms	109
7	Schwartz Space and Tempered Distributions	111
7.1	Convolution Operators	119
8	Riemann Theory of Trigonometric Series	122
8.1	Convergence in L^p and the Hilbert Transform	123
8.2	A Divergent Fourier Series	127
8.3	Conjugate Fourier Series	128

9	Interpolation Theory	129
9.1	Convex Interpolation	129
9.2	Complex Interpolation	130
9.3	Real Interpolation	132
9.4	Interpolation of Operators	132
10	Oscillatory Integrals	136
10.1	The Principle of Stationary Phase	137
10.1.1	Localization	137
10.2	Scaling	138
10.3	Surface Carried Measures	141
10.4	Restriction Theorems	141
III	Abstract Harmonic Analysis	143
11	Topological Groups	145
11.1	Basic Results	145
11.2	Quotient Groups	148
11.3	Uniform Continuity	149
11.4	Ordered Groups	151
11.5	Topological Groups arising from Normal subgroups	153
12	The Haar Measure	155
12.1	Fubini, Radon Nikodym, and Duality	162
12.2	Unimodularity	163
12.3	Convolution	165
12.4	The Riesz Thorin Theorem	173
12.5	Homogenous Spaces and Haar Measures	174
12.6	Function Spaces In Harmonic Analysis	177
13	The Character Space	178
14	Banach Algebra Techniques	186
15	Vector Spaces	187

16 Interpolation of Besov and Sobolev spaces	188
16.1 Besov Spaces	191
16.2 Proof of The Projection Result	191
17 The Cap Set Problem	192
18 Appendix: Interpolation Theory	195

Part I

Classical Fourier Analysis

Deep mathematical knowledge often arises hand in hand with the characterization of symmetry. Nowhere is this more clear than in the foundations of harmonic analysis, where we attempt to understand mathematical ‘signals’ by the ‘frequencies’ from which they are composed. 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 his book, *Théorie Analytique de la Chaleur*, published in 1811, Joseph Fourier had the audacity to announce that *all* functions were representable in this form, and used it to solve linear partial differential equations in 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 its 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. And in order to interpret the validity of Fourier summation, we will need to utilize all the convergence techniques developed during this time. Under pointwise convergence, the representation of a function by Fourier series need not be unique. Uniform convergence is useful, but a function is not uniformly summable for all functions, even if they are continuous! 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 caused Fourier and his peers to consider Fourier summation. Well before Fourier, Newton and Hooke studied oscillatory motion on the spring. It is well known that solutions to the equation $\ddot{x} = -k^2x$ are given by the equation

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

The coefficient ϕ is known as the *phase* of the oscillation, C the *amplitude*, and $k/2\pi$ the *frequency*. This equation forbodes the fact that trigonometric equations are fundamental to the study of oscillatory behaviour.

1.1 The Wave Equation

Fourier's main contribution to the foundation of harmonic analysis, namely, for 2π periodic functions $f(t)$ on the real line, the existence of an expansion

$$f(t) = C + \sum_{n=1}^{\infty} A_n \cos(nt) + B_n \sin(nt),$$

originated when studying the motion of waves. Physical considerations lead Fourier and the mathematicians of his time to believe that the partial differential equation defining wave propagation was given by the wave equation $\Delta u = \ddot{u}$, 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 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 on the real line, which has infinitely many configuration points. This makes the analysis of the wave equation much harder than the harmonic oscillator.

The easiest version of the wave equation is where we analyze a string tethered down between two endpoints, so that the wave propagation is one dimensional. In this setting, the equation reads $u_{xx} = \ddot{u}$. 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 $u : [0, \pi] \times \mathbf{R} \rightarrow \mathbf{R}$ 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 wave equation can be written

$$f''(x)g(t) = f(x)g''(t).$$

This means that unless f or g are equal to zero across their entire domain, $f''(x)/f(x) = g''(t)/g(t)$ equal a constant independant of x and t . This constant must be negative in order for g and f to be oscillatory. Thus there exists $\lambda \geq 0$ such that

$$f''(x) = -\lambda^2 f(x) \quad \text{and} \quad g''(t) = -\lambda^2 g(t).$$

It is easy to solve these equations, writing

$$\begin{aligned} f(x) &= f(0) \cos(\lambda x) + \frac{f'(0)}{\lambda} \sin(\lambda x), \\ g(t) &= g(0) \cos(\lambda t) + \frac{g'(0)}{\lambda} \sin(\lambda t). \end{aligned}$$

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

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

These functions are the harmonics. In music, these are the 'pure tones', which are combined together to produce 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

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

Daniel Bernoulli had suggested this before Fourier, but it was Fourier who believed that these were *all* such solutions, provided we take $n \rightarrow \infty$, and consider an infinite series. This means that if $u(0, x) = f(x)$, then

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

Given an initial velocity function $\dot{u}(0, x) = g(x)$, by performing a formal differentiation, we should also expect

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

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 functions $f : [0, \pi] \rightarrow \mathbf{R}$ and $g : [0, \pi] \rightarrow \mathbf{R}$ into the sum of sinusoidal functions of integer 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 what conditions ensure these coefficients reflect the original function?

1.2 Obtaining the Fourier Coefficients

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 various frequencies of sine functions are orthogonal to one another, 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,

$$\begin{aligned} \int_0^\pi f(x) \sin(mx) dx &= \int_0^\pi \sum_{n=0}^\infty A_n \sin(nx) \sin(mx) \\ &= \sum_{n=0}^\infty \int_0^\pi A_n \sin(nx) \sin(mx) = \frac{\pi A_m}{2}. \end{aligned}$$

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

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

These values are the Fourier coefficients of a function $f : [0, \pi] \rightarrow \mathbf{R}$.

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

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

We can calculate Fourier coefficients by an integration by parts, i.e.

$$A_n = \frac{2}{\pi} \left(\int_0^p \frac{x \sin(nx)}{p} dx + \int_p^\pi \frac{\pi-x}{\pi-p} \sin(nx) dx \right).$$

Now

$$\begin{aligned} \int \frac{x \sin(nx)}{p} &= \frac{\sin(nx)}{pn^2} - \frac{x \cos(nx)}{pn}, \\ \int \frac{\pi-x}{\pi-p} \sin(nx) &= -\frac{\pi-x}{\pi-p} \frac{\cos(nx)}{n} - \frac{1}{\pi-p} \frac{\sin(nx)}{n^2}. \end{aligned}$$

Thus

$$\begin{aligned} A_n &= \frac{2}{\pi} \left(\left(\frac{\sin(np)}{pn^2} - \frac{\cos(np)}{n} \right) + \left(\frac{\cos(np)}{n} + \frac{\sin(np)}{(\pi-p)n^2} \right) \right) \\ &= \frac{2 \sin(np)}{p(\pi-p)} \frac{1}{n^2}. \end{aligned}$$

Thus if we pluck a string, and let it perturb back and forth, the motion will be described by the infinite series

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

This converges absolutely and uniformly across time and space.

1.3 The Heat Equation

Suppose we have a two-dimensional region D , with a heat distribution fixed on the boundary, upon which temperature fluctuates in the interior. The equation modelling the evolution of the heat distribution over time is the *heat equation* $\Delta u = \dot{u}$. Here we are describing evolution over functions in two dimensions, which like the wave equation, also described dynamics of an infinite dimensional configuration.

To simplify again, we start by looking at the steady state heat equations, i.e. those functions u satisfying $\Delta u = 0$, known as *harmonic functions*. Normally, we fix the values of u on the boundary of some set, and attempt to determine the behaviour of u on the interior. Physically, this means we consider a system which fixes a temperature on the boundary, waits a long time, and then we look at how the heat has dispersed on the interior. For now, let's consider functions on the unit disk \mathbf{D} , with fixed values on the boundary. In this domain, we can switch to polar coordinates, in which the Laplacian equation becomes

$$\Delta u = u_{rr} + u_r/r + u_{\theta\theta}/r^2.$$

We now apply a separation of coordinates. If $\Delta u = 0$, then we know

$$r^2 u_{rr} + r u_r = -u_{\theta\theta}.$$

Writing $u(r, \theta) = f(r)g(\theta)$, the equation above can be rearranged to read

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

This means that both sides are equal to a constant λ^2 for some λ . The constant is positive, for if the constant value was negative, g wouldn't be periodic, and therefore wouldn't reflect a function on the unit disk. Solving these equations tells us

$$g''(\theta) = -\lambda g(\theta) \quad \text{and} \quad r^2 f''(r) + r f'(r) - \lambda f(r) = 0.$$

Thus 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 read $f(r) = Ar^m + Br^{-m}$. For physical reasons, we force $f(r)$ to be bounded at zero, which implies $B = 0$. As a result, we conclude that the only solutions to Laplace's equation with separable variables are

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

When $m = 0$, the solution is just constant, because the solutions to the equation $r f''(r) + f'(r) = 0$ are described by the equation $f(r) = A \log(r) + B$. Since this is unbounded at the origin, we discard it as a solution. Like with the wave equation, we would hope 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 when $r = 1$, then we may apply the same expansion technique of the wave equation, except now we are trying to expand a function on $[-\pi, \pi]$ in sines and cosines. Noticing that sines and cosines are orthogonal under integration on $[-\pi, \pi]$, and that

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

a formal calculation indicates the coefficients should be given by

$$A_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin(mt) dt \quad \text{and} \quad B_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos(mt) dt.$$

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 $[0, \pi]$. Thus the analysis of functions on $[0, \pi]$ is the same as the analysis of odd functions on $[-\pi, \pi]$. This means the analysis of functions $f : [0, \pi] \rightarrow \mathbf{R}$ reduces to the analysis of functions $f : [-\pi, \pi] \rightarrow \mathbf{R}$.

1.4 Complex Exponentials

In the situation of analyzing a function $f : [-\pi, \pi] \rightarrow \mathbf{R}$, there is a useful additional choice of notation, which replaces the sine and cosine with a single family of trigonometric functions. Given an integer n , we consider the 2π periodic function complex-valued function $e_n(t) = e^{nit}$. Euler's formula says that

$$e_n(t) = \cos(nt) + i \sin(nt).$$

This implies that if $n > 0$,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e_n(-t) dt = \frac{A_m - iB_m}{2} \quad \text{and} \quad \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e_{-n}(-t) dt = \frac{A_m + iB_m}{2}.$$

Thus the values of $\{A_m, B_m : m \geq 0\}$ can be recovered from the values of

$$\left\{ C_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e_n(-t) dt : n \in \mathbf{Z} \right\}.$$

In particular, we formally have

$$\sum_{m=0}^{\infty} A_m \cos(mt) + B_m \sin(mt) = \sum_{n=-\infty}^{\infty} C_n e_n(t).$$

This choice of coefficients is the most standard in modern Fourier analysis.

To summarize, we have shown a 2π periodic function $f : \mathbf{R} \rightarrow \mathbf{C}$ gives rise to a function $\hat{f} : \mathbf{Z} \rightarrow \mathbf{C}$, defined by

$$\hat{f}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e_n(-t) dt,$$

which measures the average value of f when it is ‘twisted’ by an oscillation with frequency n . And now the Fourier summation formula takes the form

$$f(t) = \sum_{n=-\infty}^{\infty} \hat{f}(n)e_n(t).$$

As an alternate notation, \hat{f} is also denoted by $\mathcal{F}(f)$. The Fourier series representation in terms of complex exponentials will be our choice throughout the rest of these notes. No deep knowledge of the complex numbers is used here. 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 the coefficients are controlled by reasonable transformations. Most basic, but unappreciated property of the Fourier transform is *linearity*: For any two functions $f : [-\pi, \pi] \rightarrow \mathbf{R}$ and $g : [-\pi, \pi] \rightarrow \mathbf{R}$, the linearity of the integral formula defining the Fourier transform implies

$$\widehat{f+g} = \hat{f} + \hat{g}.$$

Linearity is *essential* to most methods in this book, and the field of non-linear functionals still has many fundamental unanswered problems. Fourier series are also stable under various other transformations which occur in analysis, which makes the transform useful. We summarize these properties here:

- If, for $f : [-\pi, \pi] \rightarrow \mathbf{R}$, we define $\bar{f}(x) = \overline{f(x)}$, then

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

- If $f : [-\pi, \pi] \rightarrow \mathbf{R}$ is a real-valued function, then

$$\hat{f}(-n) = \overline{\hat{f}(n)}.$$

Formally, this means that if $\hat{f}(n) = A_n + iB_n$, then

$$\begin{aligned}\sum_{n \in \mathbf{Z}} \hat{f}(n) e_n(t) &= \hat{f}(0) + \sum_{n=1}^{\infty} \hat{f}(n) e_n(t) + \overline{\hat{f}(n) e_n(t)} \\ &= \hat{f}(0) + 2 \sum_{n=1}^{\infty} \operatorname{Re}(\hat{f}(n) e_n(t)) \\ &= \hat{f}(0) + 2 \sum_{n=1}^{\infty} A_n \cos(nt) - B_n \sin(nt).\end{aligned}$$

Thus the Fourier expansion of a real-valued function is a real-coefficient expansion in sines and cosines.

- Define the translation operator T_s by $(T_s f)(t) = f(t - s)$, and the frequency modulation operator M_s by $(M_s f)(t) = e_s(t) f(t)$. Then we have

$$\widehat{T_s f} = M_{-s} \hat{f} \quad \text{and} \quad \widehat{M_s f} = T_s \hat{f}.$$

Thus a translation in the spatial domain corresponds to a modulation in the frequency domain, and vice versa. One can see this using the Fourier summation formula, by the formal calculation

$$\begin{aligned}(T_s f)(t) &= \sum_{n \in \mathbf{Z}} A_n e_n(t - s) = \sum_{n \in \mathbf{Z}} (A_n e_n(-s)) e_n(t), \\ (M_s f)(t) &= e_s(t) \sum_{n \in \mathbf{Z}} A_n e_n(t) = \sum_{n \in \mathbf{Z}} A_n e_{n+s}(t) = \sum_{n \in \mathbf{Z}} A_{n-s} e_n(t).\end{aligned}$$

- If f is odd, then

$$\hat{f}(-n) = -\hat{f}(n).$$

In particular, if f is a real-valued, odd function, then the Fourier coefficients of f are purely imaginary, and

$$\hat{f}(n) = \frac{1}{\pi i} \int_0^\pi f(t) \sin(nt).$$

- If f is even, then $\hat{f}(-n) = \hat{f}(n)$. If f is real-valued, then the Fourier coefficients of f are real, and

$$\hat{f}(n) = \frac{1}{\pi} \int_0^\pi f(t) \cos(nt).$$

- If f has a derivative f' , then $(f')^\wedge(n) = in\hat{f}(n)$. A formal calculation using the Fourier summation formula hints at this, by letting

$$\frac{d}{dt} \sum_{n=-\infty}^{\infty} A_n e_n(t) = \sum_{n=-\infty}^{\infty} A_n \frac{de_n(t)}{dt} = \sum_{n=-\infty}^{\infty} (niA_n) e_n(t).$$

All but the last relations can be proved by easy exercises in manipulating periodic integrals, when f is a 2π periodic measurable function with

$$\int_{-\pi}^{\pi} |f(x)| dx < \infty.$$

The space of all such functions will be denoted by $L^1(\mathbf{T})$, which is a Banach space under the norm

$$\|f\|_{L^1(\mathbf{T})} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)| dx.$$

The latter property involving the derivative of f holds by an easy integration by parts. This proof therefore holds whenever f has a weak derivative in $L^1(\mathbf{T})$.

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 these series, which we shall look at in more detail later.

Example. Consider the function $f : [0, \pi] \rightarrow \mathbf{R}$ defined by $f(x) = x(\pi - x)$. Then a series of integration by parts gives that

$$\int x(\pi - x) \sin(nx) = \frac{x(\pi - x) \cos(nx)}{n} + \frac{(\pi - 2x) \sin(nx)}{n^2} - \frac{2 \cos(nx)}{n^3}.$$

Thus

$$\frac{2}{\pi} \int_0^{\pi} x(\pi - x) \sin(nx) = \frac{4(1 - \cos(n\pi))}{n^3} = \begin{cases} \frac{8}{\pi n^3} & n \text{ odd,} \\ 0 & n \text{ even.} \end{cases}$$

Thus we have a formal representation

$$f(x) = \sum_{n \text{ odd}} \frac{8}{\pi n^3} \sin(nx).$$

This sum converges absolutely and uniformly for $x \in [0, \pi]$. If we extend the domain of f to $[-\pi, \pi]$ by making f odd, then

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

In this case, we still have

$$f(x) = \sum_{\substack{n \text{ odd} \\ n > 0}} \frac{4}{\pi i n^3} [e_n(x) - e_n(-x)] = \sum_{n \text{ odd}} \frac{8}{\pi n^3} \sin(nx).$$

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

Example. The tent function

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

is even, and therefore has a purely real Fourier expansion

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

Thus we obtain an expansion

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

This sum also converges absolutely and uniformly.

Example. Consider the characteristic function

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

Then

$$\hat{\chi}_{(a,b)}(n) = \frac{1}{2\pi} \int_a^b e_n(-x) = \frac{e_n(-a) - e_n(-b)}{2\pi i n}.$$

Hence we may write

$$\begin{aligned} \chi_{(a,b)}(x) &= \frac{b-a}{2\pi} + \sum_{n \neq 0} \frac{e_n(-a) - e_n(-b)}{2\pi i n} e_n(x) \\ &= \frac{b-a}{2\pi} + \sum_{n=1}^{\infty} \frac{\sin(nb) - \sin(na)}{\pi n} \cos(nx) + \frac{\cos(na) - \cos(nb)}{\pi n} \sin(nx). \end{aligned}$$

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

$$\left| \frac{e_n(-b) - e_n(-a)}{2\pi n} \right| = \left| \frac{1 - e_n(b-a)}{2\pi n} \right| \geq \left| \frac{\sin(n(b-a))}{2\pi n} \right|,$$

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

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

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

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

Thus 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 ,

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

This means

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

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

$$\sum_{n=1}^{\infty} \left| \frac{1 - e_n(\pi)}{2\pi n} \right| = \sum_{n \text{ even}} \left| \frac{1}{\pi n} \right| = \infty.$$

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. This is because if we set $z = e^{it}$, then

$$\sum_{n=-\infty}^{\infty} a_n e^{nit} = \sum_{n=-\infty}^{\infty} a_n z^n$$

becomes a Laurent series in z . For instance, we have a power series expansion

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

This converges pointwise for every $z \in \mathbf{D}$ but $z = 1$. Thus for $x \notin 2\pi\mathbf{Z}$,

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{\cos(kx)}{k} &= \Re \left(\log \left(\frac{1}{1 - e(x)} \right) \right) = -\frac{1}{2} \log(2 - 2\cos(x)), \\ \sum_{k=1}^{\infty} \frac{\sin(kx)}{k} &= \Im \left(\log \left(\frac{1}{1 - e(x)} \right) \right) = \arctan \left(\frac{\sin(x)}{1 - \cos(x)} \right). \end{aligned}$$

Here we agree that $\arctan(\pm\infty) = \pm\pi/2$. One can check that, indeed, the Fourier series of these two functions corresponds precisely to these summations. 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 because this is the domain we are interested in. This makes the theory a little more pathological.

Chapter 2

Fourier Series

Let's focus in on the problem we introduced in the last chapter. We write $\mathbf{T} = \mathbf{R}/2\pi\mathbf{Z}$, so that a function $f : \mathbf{T} \rightarrow \mathbf{C}$ is a complex-valued 2π periodic function on the real line. The Lebesgue measure on \mathbf{R} induces a natural Borel measure on \mathbf{T} , which we normalize so that \mathbf{T} has total measure one. Given a point $x \in \mathbf{T}$, we let $|x| = \min_{n \in \mathbf{Z}} |x + 2\pi n|$. For each integrable function $f : \mathbf{T} \rightarrow \mathbf{C}$, we have an associated *formal trigonometric series*

$$\sum_{n=-\infty}^{\infty} \hat{f}(n)e_n(x).$$

In some sense, f should be able to be approximated by the trigonometric polynomials obtained by truncating this series. However, at this point, we haven't deduced any reason for these sums to converge to f analytically.

To understand the convergence, we define the partial sums

$$(S_N f)(x) = \sum_{n=-N}^N \hat{f}(n)e_n(x).$$

The reason we sum from $-N$ to N is so that if f is real-valued, then the complex parts of $\hat{f}(n)$ and $\hat{f}(-n)$ will cancel out for each n , and so $S_N(f)$ will be a sum of cosines and sines, as in classical Fourier analysis.

An initial hope is that the Fourier series of a function converges pointwise, i.e. that for every x , $\lim_{N \rightarrow \infty} S_N(f)(x) = f(x)$. In fact, for periodic, continuously differentiable functions, we'll actually get uniform convergence. But even if we try and look at continuous periodic functions, the

Fourier summation can diverge. Thus we look for more exotic forms of convergence, and different quantitative descriptions of the asymptotic behaviour of the Fourier series.

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

Proof. We shall prove this first for real-valued functions, and without loss of generality, for the origin. For every trigonometric polynomial $g(x) = \sum a_n e_n(-x)$, we have

$$\int_{-\pi}^{\pi} f(x)g(x)dx = 2\pi \sum a_n \hat{f}(n) = 0.$$

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

$$g(x) = \varepsilon + \cos x = \varepsilon + \frac{e(x) + e(-x)}{2}.$$

where ε is small enough that $g(x) > A > 1$ for $|x| < \delta/2$, $P(x) > 0$ for $\delta/2 \leq |x| < \delta$, and $g(x) < B < 1$ for $|x| \geq \delta$. The series of trigonometric polynomials $g_n(x) = (\varepsilon + \cos x)^n$ satisfy

$$\left| \int_{\mathbb{T}} g_n(x)f(x)dx \right| \geq \int_{|x| < \delta} g_n(x)f(x)dx - \left| \int_{\delta \leq |x|} g_n(x)f(x)dx \right|.$$

Hölder's inequality then guarantees that as $n \rightarrow \infty$,

$$\left| \int_{\delta \leq |x|} g_n(x)f(x)dx \right| \lesssim B^n = o(1).$$

whereas

$$\left| \int_{|x|<\delta} g_n(x) f(x) dx \right| \geq \int_{|x|<\delta/2} g_n(x) f(x) \gtrsim A^n.$$

which increases exponentially fast as $n \rightarrow \infty$. Thus we conclude

$$0 = \left| \int_{-\pi}^{\pi} g_n(x) f(x) dx \right| \gtrsim A^n - o(1).$$

For suitably large values of n , the right hand side is positive, whereas the left hand side is zero, which is impossible. By contradiction, we conclude $f(0) = 0$. 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 $f, g \in C(\mathbf{T})$ and $\widehat{f} = \widehat{g}$, then $f = g$.*

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

Corollary 2.3. *If $f \in C(\mathbf{T})$ and $\widehat{f} \in L^1(\mathbf{Z})$, $S_N f \rightarrow f$ uniformly as $N \rightarrow \infty$.*

Proof. If $\sum |\widehat{f}(n)| < \infty$, then we calculate

$$|(S_{N+M}f)(x) - (S_N f)(x)| \leq \sum_{k=N+1}^M |\widehat{f}(k)|.$$

Since the series \widehat{f} is absolutely summable, for any ε , there is sufficiently large N , such that the quantity above is bounded by ε . Thus the functions $\{S_N f\}$ are a Cauchy sequence in $L^\infty(\mathbf{T})$, and therefore converge uniformly to some function Sf , which necessarily must be continuous. Uniform convergence also implies we can interchange integrals to conclude

$$\widehat{Sf}(n) = \lim_{N \rightarrow \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} S_N(f)(t) e_n(-t) dt = \widehat{f}(n).$$

Thus $\widehat{Sf} = \widehat{f}$, so $Sf = f$ since both functions are continuous. \square

Later we show that if $f \in C^m(\mathbf{T})$, then as $|n| \rightarrow \infty$, $\hat{f}(n) = O(1/n^m)$. In particular, if $m \geq 2$ this means $S_N f \rightarrow f$ uniformly. If $f \in C^\infty(\mathbf{T})$, then the partial sums $S_N f$ and all of their derivatives converge to f at a rate of $O_M(1/N^M)$ for *any* integer M . Conversely, take a sequence a_n , for $n \in \mathbf{Z}$, with $a_n = O_M(1/|n|^M)$ for all M . If we set $f_N = \sum_{|n| \leq N} a_n e_n$, then f_N converges uniformly to some f . But one can verify that the m 'th derivative of the f_N also converge uniformly to some function, which therefore must be the m 'th derivative of f . Thus $f \in C^\infty(\mathbf{T})$, and $\hat{f}(n) = a_n$. This means there is a perfect duality between infinitely differentiable functions and arbitrarily fast decaying sequences of coefficients. In more advanced contexts, like the theory of distributions, this duality is very useful for studying the Fourier transform.

2.2 Quantitative Bounds on Fourier Coefficients

Using classical analysis, the last section essentially details all the results we can obtain about the Fourier series of a function. Of course, in practical contexts most of the functions dealt with are arbitrarily smooth, so the picture seems rather complete. Nonetheless, we still want to study the qualitative properties of the Fourier series in terms of the qualitative properties of functions. For instance, does the Fourier series of functions which are globally small converge faster than functions which are only small on average. In modern terms, do we get faster convergence rates if $\|f\|_{L^\infty(\mathbf{T})}$ is small rather than just if $\|f\|_{L^1(\mathbf{T})}$ is small. Does the convergence get faster if we consider the convergence with respect to an L^p norm rather than an L^∞ norm? Thus we want to understand the behaviour of the Fourier series with respect to a family of *norm spaces*. Of course, these norms are defined in a more general space of measurable functions, and to apply functional analysis arguments it is essential to ‘complete’ the picture of these norms, so we will find that many of the arguments, initially invented to study smooth functions, also work naturally with arbitrary integrable functions.

These quantitative estimates are still interesting even if we knew everything there was to know about the pointwise convergence of Fourier series. A series of functions may converge to a function pointwise, whereas none of these individual functions may look like the function they converge to at all. So we may want to look at quantitative measures of how globally similar two functions are. For instance, we might want to know how the

Fourier series of a function changes under a small perturbation, and this corresponds to understanding quantitative estimates about the operators S_N , or the rate of convergence of $S_N f$ to f , which might be required if we want to determine how large N must be to approximate f .

Example. *If we consider a square wave χ_I for some interval I , then the techniques of the following section allow us to prove that $\|\chi_I - S_N \chi_I\|_{L^2(\mathbf{T})} = O(1/\sqrt{N})$, independently of I . This means that if we want to simulate square waves with a musical instrument up to some square mean error ε , then we will need $\Omega(\varepsilon^{-2})$ different notes to represent the sound accurately. A piano with 88 keys can only approximate square waves slightly better than a recorder with 5 holes. If f is $C^{m+1}(\mathbf{T})$, then $\|f - S_N f\|_{L^2(\mathbf{T})} = O(1/N^{m/2})$, so we require significantly less notes to simulate this sound, i.e. $\Omega(\varepsilon^{-2/m})$. In this case a piano can simulate these sounds much more accurately.*

One initial equation which might summarize how well behaved the Fourier series is with respect to suitable norms would be to obtain an estimate of the form $\|\hat{f}\|_{L^q(\mathbf{Z})} \lesssim \|f\|_{L^p(\mathbf{T})}$ for particular values of p and q . If this was established, we could conclude that the Fourier series is stable in the L^q norm under small perturbations in the L^p norm. The first inequality we give is trivial, but is certainly tight, i.e. for $f(t) = e_n(t)$.

Theorem 2.4. *For any f , $\|\hat{f}\|_\infty \leq \|f\|_1$.*

Proof. We just take absolute values into the oscillatory integral defining the Fourier coefficients, calculating that for any n ,

$$|\hat{f}(n)| = \left| \frac{1}{2\pi} \int f(t) e_n(-t) \right| \leq \frac{1}{2\pi} \int |f(t)| = \|f\|_1,$$

which was the required bound. \square

This proof doesn't really take any features of the Fourier coefficient into account other than that they are defined in terms of an oscillatory integral. The same bound would hold for any oscillatory integral. But it makes sense that the bound is still tight in terms of the quantities given, because the magnitude of the Fourier series gives oscillatory information which is completely annihilated in the phase space when taking an absolute value in the L^p norm. The only L^p space where we can get a completely satisfactory bound is for $p = 2$, where we can use Hilbert space techniques.

Theorem 2.5. For any function f , $\|\hat{f}\|_{L^2(\mathbf{Z})} \leq \|f\|_{L^2(\mathbf{T})}$.

Proof. With respect to the normalized inner product on the space $L^2(\mathbf{T})$, the calculations of the last chapter tell us that the exponentials are an orthonormal family of functions, in the sense that for distinct n and m , $(e_n, e_m) = 0$, and $(e_n, e_n) = 1$. Since $\hat{f}(n) = (f, e_n)$, the theorem then becomes an instance of Bessel's inequality. \square

The exponentials e_n are actually an orthonormal basis for $L^2(\mathbf{T})$, which can either be proven immediately by using the Stone Weierstrass theorem to show trigonometric polynomials are dense in $L^2(\mathbf{T})$, or by results we prove later on in these notes. This tells us that the Fourier expansion is an orthonormal basis expansion, and so Parseval's equality tells us that the inequality in the theorem above is actually an *equality*, i.e.

$$\|\hat{f}\|_{L^2(\mathbf{Z})} = \|f\|_{L^2(\mathbf{T})}.$$

This correspondence makes the Hilbert space $L^2(\mathbf{T})$ often the best place to understand Fourier expansion techniques, and general results are often achieved by reduction to this well understood case. For instance, the inequality above, combined with the trivial inequality, is easily interpolated using the Riesz-Thorin technique to give the Hausdorff Young inequality.

Theorem 2.6. For any f , and $1 \leq p \leq 2$, $\|\hat{f}\|_{L^q(\mathbf{Z})} \leq \|f\|_{L^p(\mathbf{T})}$.

It might be surprising to note that the Hausdorff Young inequality essentially completes the bounds on the Fourier series with respect to the L^p norms. There is no interesting result one can obtain for $p > 2$ other than the obvious inequality

$$\|\hat{f}\|_{L^2(\mathbf{Z})} \leq \|f\|_{L^p(\mathbf{T})}.$$

Thus we see that it is easier to control the magnitude of Fourier coefficients in terms of the width of the original function, rather than controlling the width of the Fourier coefficient in terms of the height of the original function. This makes sense, because the L^p norm of f measures fairly different aspects of the function than the L^q norm of \hat{f} . It is only in the case of the L^2 norm where results are precise, and where p is small that we can take a trivial bound, that we get an inequality like the Hausdorff Young result.

2.3 Asymptotic Decay of Fourier Series

The next result, known as Riemann-Lebesgue lemma, shows that the Fourier series of any integrable function decays, albeit arbitrarily slowly. The proof we give is an instance of an important principle in Functional analysis that we will use over and over again. Suppose for each n , we have a bounded operator $T_n : X \rightarrow Y$ between Banach spaces, and we want to show that for each $x \in X$, $\lim T_n(x) = T(x)$, where T is another bounded operator. The principle is that if it is obvious that $\lim T_n(x') = T(x')$ for a dense family of points $X' \subset X$, and the operators T_n are *uniformly* bounded, then for any x , as $n \rightarrow \infty$ we have

$$\begin{aligned} \|T_n(x) - T(x)\| &\leq \|T_n(x) - T_n(x')\| + \|T_n(x') - T(x')\| + \|T(x') - T(x)\| \\ &\lesssim \|x - x'\| + o_{x'}(1). \end{aligned}$$

If we choose x' such that $\|x - x'\| \leq \varepsilon$, then for n large enough we find that $\|T_n(x) - T(x)\| \lesssim \varepsilon$. Since ε was arbitrary, this means that $\lim_{n \rightarrow \infty} T_n(x) = T(x)$. The advantage of the principle is that it is suitably abstract, and can thus be used very flexibly. But the disadvantage is that it is a very soft analytical argument, and cannot be used to obtain results on the rate of convergence of $T_n(x)$ to $T(x)$.

Lemma 2.7. *If $f \in L^1(\mathbf{T})$, then $\widehat{f}(n) = o(1)$ as $|n| \rightarrow \infty$.*

Proof. We claim the lemma is true for the characteristic function of an interval χ_I . If $I = [a, b]$, then

$$\widehat{\chi_I}(n) = \frac{1}{2\pi} \int_a^b e_n(-t) = \frac{i(e_n(-b) - e_n(-a))}{2\pi n} = O(1/n).$$

By linearity of the integral, the Fourier transform of any step function vanishes at ∞ . But if $\Lambda_n(f) = \widehat{f}(n)$, then $\|\Lambda_n f\|_{L^\infty(\mathbf{Z})} \leq \|f\|_{L^1(\mathbf{T})}$, which shows that the functionals Λ_n are uniformly bounded as maps from $L^1(\mathbf{T})$ to $L^\infty(\mathbf{Z})$. Since $\lim_{|n| \rightarrow \infty} \Lambda_n(f) = 0$ for any step function f , and the step functions are dense in $L^1(\mathbf{T})$, we conclude that $\lim_{|n| \rightarrow \infty} \Lambda_n(f) = 0$ for all $f \in L^1(\mathbf{T})$. \square

Even though the Fourier series of any step function decays at a rate $O(1/n)$, it is *not* true that a general Fourier series decays at a rate of $O(1/n)$.

And in fact, one can provide examples of functions whose Fourier series decay at an arbitrarily slow rate. However, for smooth functions, we can obtain a decay rate. This is an instance of a general result related to the duality between decay and smoothness in phase and frequency space.

Theorem 2.8. *If $f \in C^m(\mathbf{T})$, then $|\hat{f}(n)| \leq n^{-m} \|f^{(m)}\|_{L^1(\mathbf{T})}$.*

Proof. Applying the derivative law for Fourier series m times, we find that

$$|\hat{f}(n)| = n^{-m} |\widehat{f^{(m)}}(n)| \leq n^{-m} \|f^{(m)}\|_{L^1(\mathbf{T})}. \quad \square$$

Suppose that μ is a measure on \mathbf{T} with finite variation, i.e. $\mu \in M(\mathbf{T})$. Then one can define the Fourier series of μ by

$$\hat{\mu}(n) = \int_{-\pi}^{\pi} e_n(-x) d\mu(x).$$

If μ is absolutely continuous with respect to the Lebesgue measure, with $d\mu = f dx$, then $\hat{\mu} = \hat{f}$, so this is an extension of the Fourier series to measures. One can verify that

$$\|\hat{\mu}\|_{L^\infty(\mathbf{Z})} \leq \|\mu\|_{M(\mathbf{T})}.$$

If δ is the Dirac delta measure at the origin, i.e. $\mu(E) = 1$ if $0 \in E$, and $\mu(E) = 0$ otherwise, then for all n ,

$$\hat{\delta}(n) = 1.$$

Thus the Fourier series of δ has no decay at all. One can view this as saying functions are ‘smoother’ than measures, and therefore have a Fourier decay, albeit one that is qualitative rather than quantitative.

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

$$(f * g)(t) = \frac{1}{2\pi} \int_{\mathbf{T}} f(s) g(t-s).$$

Thus we smear the values of g with respect to a density function f . The next lemma helps us quantitatively understand the smearing process.

Lemma 2.9. *For any $1 \leq p < \infty$, and $f \in L^p(\mathbf{T})$, $\lim_{h \rightarrow 0} T_h f = f$ in $L^p(\mathbf{T})$.*

Proof. If f is $C^1(\mathbf{T})$, then $|f(x+h) - f(x)| \lesssim h$ uniformly in x , implying that $\|T_h f - f\|_{L^p(\mathbf{T})} \leq \|T_h f - f\|_{L^\infty(\mathbf{T})} \lesssim h$, and so $T_h f \rightarrow f$ in all the spaces $L^p(\mathbf{T})$. We have $\|T_h f\|_{L^p(\mathbf{T})} = \|f\|_{L^p(\mathbf{T})}$, so the T_h are a bounded family of operators, and since $C^1(\mathbf{T})$ is dense in $L^p(\mathbf{T})$ for $1 \leq p < \infty$, we conclude that $\lim_{h \rightarrow 0} T_h f = f$ for all $f \in L^p(\mathbf{T})$. \square

Theorem 2.10. *Convolution has the following properties:*

- If $f \in L^p(\mathbf{T})$ and $g \in L^q(\mathbf{T})$, for $1/p + 1/q = 1$, then $f * g$ is uniformly continuous.
- If $f \in L^p(\mathbf{T})$ and $g \in L^q(\mathbf{T})$, and if we define r so that $1/r = 1/p + 1/q - 1$, with $1 \leq r \leq \infty$, then $f * g$ is defined by the integral formula almost everywhere, and

$$\|f * g\|_{L^r(\mathbf{T})} \leq \|f\|_{L^p(\mathbf{T})} \|g\|_{L^q(\mathbf{T})}.$$

This is known as Young's inequality for convolutions.

- Convolution is a commutative, associative, bilinear operation.
- If $f, g \in L^1(\mathbf{T})$, then $\widehat{f * g} = \widehat{f} \widehat{g}$.
- If f has a weak derivative f' in $L^1(\mathbf{T})$, then $f * g$ has a weak derivative in $L^1(\mathbf{T})$, and $(f * g)' = f' * g$. Thus convolution is 'additively smoothing'. In particular, if $f \in C^k(\mathbf{T})$ and $g \in C^l(\mathbf{T})$, then $f * g \in C^{k+l}(\mathbf{T})$.
- If f is supported on E , and g on F , then $f * g$ is supported on $E + F$.

Proof. Suppose $f \in L^p(\mathbf{T})$, and $g \in L^q(\mathbf{T})$, then

$$\begin{aligned} |(f * g)(x-h) - (f * g)(x)| &\leq \int |f(x-h-y) - f(x-y)| |g(y)| dy \\ &\leq \|f_h - f\|_{L^p(\mathbf{T})} \|g\|_{L^q(\mathbf{T})}. \end{aligned}$$

The right hand side is a bound independent of x and converges to zero as $h \rightarrow 0$, so $f * g$ is uniformly continuous. Applying Hölder's inequality

again gives that $\|f * g\|_{L^\infty(\mathbf{T})} \leq \|f\|_{L^p(\mathbf{T})} \|g\|_{L^q(\mathbf{T})}$. If $f \in L^p(\mathbf{T})$, and $g \in L^1(\mathbf{T})$, we use Minkowski's inequality to conclude that

$$\begin{aligned} \|f * g\|_{L^p(\mathbf{T})} &= \left(\int \left| \int f(x-y)g(y) dy \right|^p dx \right)^{1/p} \\ &\leq \int \left(\int |f(x-y)g(y)|^p dx \right)^{1/p} dy \\ &= \int g(y) \|f\|_{L^p(\mathbf{T})} dy = \|f\|_{L^p(\mathbf{T})} \|g\|_{L^1(\mathbf{T})}. \end{aligned}$$

Thus $f * g$ is finite almost everywhere. The inequality also implies that

$$\|f * g\|_{L^p(\mathbf{T})} \leq \|f\|_{L^1(\mathbf{T})} \|g\|_{L^p(\mathbf{T})}$$

if $f \in L^1(\mathbf{T})$, and $g \in L^p(\mathbf{T})$. But now implying Riesz-Thorin interpolation gives the general Young's inequality. Elementary applications of change of coordinates and Fubini's theorem establish the commutativity and associativity of convolution for functions $f, g \in L^1(\mathbf{T})$. But $L^1(\mathbf{T}) \cap L^p(\mathbf{T})$ is dense in $L^p(\mathbf{T})$. Since $f * g = g * f$ for a dense family of functions, and convolution is continuous from $L^p(\mathbf{T}) \times L^q(\mathbf{T}) \rightarrow L^r(\mathbf{T})$, we obtain the identity holds everywhere. Similarly, one can apply Fubini's theorem to obtain associativity for $f, g, h \in L^1(\mathbf{T})$, and then apply density. To obtain the product identity for the Fourier series, we can apply Fubini's theorem to write

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

and this is exactly the identity required. To calculate the weak derivative of $f * g$, we fix $\phi \in C^\infty(\mathbf{T})$, and calculate using two applications of Fubini's

theorem that

$$\begin{aligned}
\int_{\mathbf{T}} (f' * g)(x) \phi(x) dx &= \int_{\mathbf{T}} \int_{\mathbf{T}} f'(x-y) g(y) \phi(x) dy dx \\
&= \int_{\mathbf{T}} g(y) \int_{\mathbf{T}} f'(x-y) \phi(x) dx dy \\
&= - \int_{\mathbf{T}} g(y) \int_{\mathbf{T}} f(x-y) \phi'(x) dx dy \\
&= - \int_{\mathbf{T}} \left(\int_{\mathbf{T}} g(y) f(x-y) dy \right) \phi'(x) dx \\
&= - \int_{\mathbf{T}} (f * g)(x) \phi'(x) dx.
\end{aligned}$$

If $f = 0$ a.e outside E , and $g = 0$ a.e. outside F , then $(f * g)(x)$ can be nonzero only when there is a set G of positive measure such that for any $y \in G$, $f(y) \neq 0$ and $g(x-y) \neq 0$. But this means that $E \cap G \cap (x-F)$ has positive measure, so that there is $y \in E$ such that $x-y \in F$, meaning that $x \in E + F$. \square

We know that suitably smooth functions have convergent Fourier series. The advantage of convolution is if we want to study the properties of a function f , convolution with a smooth function g gives a smooth function, and provided \widehat{g} is close to 1, $\widehat{f * g}$ will be close to \widehat{f} . If we can establish the convergence properties on the convolution $f * g$, then we can probably obtain results about f . From the frequency side, $\sum \widehat{f}(n) e_n$ might not converge, but $\sum a_N(n) \widehat{f}(n) e_n$ might converge for a suitably fast decaying sequence $a_N(n)$, and if $a_N(n) \rightarrow 1$ as $N \rightarrow \infty$, this sequence might reflect properties of the original sequence.

The functions g above are called **good kernels**. A good kernel is a sequence of integrable functions K_N on \mathbf{T} bounded in L^1 norm, for which

$$\frac{1}{2\pi} \int_{\mathbf{T}} K_N(t) dt = 1.$$

so that integration against K_N operates essentially like an average, and for any $\delta > 0$, as $N \rightarrow \infty$,

$$\int_{\delta < |t|} |K_N(t)| dt \rightarrow 0,$$

so the functions K_N become concentrated at the origin as $N \rightarrow \infty$. If in addition, we have an estimate $|K_N(x)| \lesssim N$, we say it is an **approximation to the identity**.

Theorem 2.11. *Let K_N be a good kernel. Then*

- $(K_N * f)(x) \rightarrow f(x)$ for any continuity point x of f .
- $(K_N * f) \rightarrow f$ uniformly if $f \in C(\mathbf{T})$, and $K_N * f$ converges to f in $L^p(\mathbf{T})$ if $f \in L^p(\mathbf{T})$, for $1 \leq p < \infty$.
- If K_N is an approximation to the identity, $(K_N * f)(x) \rightarrow f(x)$ for x in the Lebesgue set of f .

Proof. The operators $T_N f = K_N * f$ are uniformly bounded as operators on $L^p(\mathbf{T})$. Basic analysis shows that $(K_N * f)(x) \rightarrow f(x)$ at each point x where f is continuous, and converges uniformly to f if f is in $C(\mathbf{T})$. But a density argument allows us to conclude that $K_N * f \rightarrow f$ in $L^p(\mathbf{T})$ for each $f \in L^p(\mathbf{T})$, for $1 \leq p < \infty$. To obtain pointwise convergence, we calculate

$$|(K_N * f)(x) - f(x)| \leq \frac{1}{2\pi} \int_{\mathbf{T}} |f(x-y) - f(x)| |K_N(y)| dy.$$

Let $A(\delta) = \delta^{-1} \int_{|y| < \delta} |f(x-y) - f(x)|$. Then as $\delta \rightarrow 0$, $A(\delta) \rightarrow 0$ because x is in the Lebesgue set of f . And we find that for each k , since $|K_N(y)| \lesssim N$,

$$\int_{2^k/N < |x| < 2^{k+1}/N} |f(x-y) - f(x)| |K_N(y)| \lesssim \frac{A(2^{k+1}/N)}{2^{k+1}}.$$

Thus we have a bound

$$|(K_N * f)(x) - f(x)| \lesssim \sum_{k=0}^{\infty} \frac{A(2^k/N)}{2^k}.$$

Because f is integrable, A is continuous, and hence bounded for all $\delta \leq \pi$. This means that

$$|(K_N * f)(x) - f(x)| \lesssim \sum_{k=0}^M \frac{A(2^k/N)}{2^k} + \|A\|_{\infty} \sum_{k=M}^{\infty} \frac{1}{2^k} = \sum_{k=0}^M \frac{A(2^k/N)}{2^k} + O(1/2^M).$$

For any fixed M , the finite sum tends to zero as $N \rightarrow \infty$, so we obtain that $|(K_N * f)(x) - f(x)| = o(1) + O(1/2^M)$. Taking $M \rightarrow \infty$ proves the result. \square

2.5 The Dirichlet Kernel

We calculate that

$$(S_N f)(t) = \sum_{n=-N}^N \hat{f}(n) e_n(t) = \frac{1}{2\pi} \int f(x) \left(\sum_{n=-N}^N e_n(t-x) \right) dx.$$

The bracketed part of the final term in the equation is independant of the function f , and is therefore key to understanding the behaviour of the sums S_N . We call it the Dirichlet kernel D_N , defined as

$$D_N(t) = \sum_{n=-N}^N e_n(t).$$

Thus $S_N f = f * D_N$, so analyzing this convolution gives results about the sums of Fourier series.

Theorem 2.12. *For any integer N and $t \in \mathbf{R}$,*

$$D_N(t) = \frac{\sin((N+1/2)t)}{\sin(t/2)}.$$

Proof. By the geometric series summation formula, we may write

$$\begin{aligned} D_N(t) &= 1 + \sum_{n=1}^N e_n(t) + e_n(-t) = 1 + e(t) \frac{e_N(t) - 1}{e(t) - 1} + e(-t) \frac{e_N(-t) - 1}{e(-t) - 1} \\ &= 1 + e(t) \frac{e_N(t) - 1}{e(t) - 1} + \frac{e_N(-t) - 1}{1 - e(t)} = \frac{e_{N+1}(t) - e_N(-t)}{e(t) - 1} \\ &= \frac{e_{N+1/2}(t) - e_{N+1/2}(-t)}{e_{1/2}(t) - e_{1/2}(-t)} = \frac{\sin((N+1/2)t)}{\sin(t/2)}. \end{aligned}$$

Thus as $N \rightarrow \infty$, D_N oscillates highly rapidly. □

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

$(2\pi)^{-1} \int D_N(t) = 1$. However, we find

$$\begin{aligned} \int_{-\pi}^{\pi} |D_N(t)| &= \int_{-\pi}^{\pi} \left| \frac{\sin((N+1/2)t)}{\sin(t/2)} \right| = 2 \int_0^{\pi} \frac{|\sin((N+1/2)t)|}{\sin(t/2)} \\ &\geq 4 \int_0^{\pi} \frac{|\sin((N+1/2)t)|}{t} = 4 \int_0^{(2N+1)\pi} \frac{|\sin(u/2)|}{u} \\ &\geq \sum_{n=0}^{N-1} \frac{2}{n\pi} \int_0^{2\pi} \sin(u/2) = \frac{8}{\pi} \sum_{n=0}^{N-1} \frac{1}{n} = \Omega(\log N). \end{aligned}$$

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. In fact, a simple functional analysis argument shows that pointwise convergence of Fourier series fails for continuous functions.

Theorem 2.13. *There exists $f \in C(\mathbf{T})$ such that $(S_N f)(0)$ diverges as $N \rightarrow \infty$.*

Proof. If we consider the linear operators $\Lambda_N f = (S_N f)(0) = (f * D_N)(0)$ as maps from $C(\mathbf{T})$ to $L^1(\mathbf{T})$, and if we let f be a continuous function approximating $\text{sgn}(D_N)$, then we can obtain a sequence f_N such that $|\Lambda_N f_N| = \Omega(\log N) \|f_N\|_{\infty}$. This implies that $\|\Lambda_N\| \rightarrow \infty$ as $N \rightarrow \infty$. The uniform boundedness principle thus implies that there exists a *single* function $f \in C(\mathbf{T})$ such that $\sup |\Lambda_N f| = \infty$, so $(S_N f)(0)$ diverges as $N \rightarrow \infty$. \square

One way around this is to consider only functions f which provide a suitable dampening condition on the oscillation of D_N . The result we find reveals the surprising fact that the local behaviour of a function at a point is sufficient to determine pointwise convergence of a Fourier series, which is defined in terms of the integral of the function on its entire domain. Nonetheless, having some smoothness properties of a function can fix this problem for the Dirichlet kernel, thus obtaining some form of pointwise convergence.

Theorem 2.14. *If $f \in L^1(\mathbf{T})$ is Lipschitz at x , then $(S_N f)(x) \rightarrow f(x)$.*

Proof. We write

$$(S_N f)(x) - f(x) = (f * D_N)(x) - f(x) = \frac{1}{2\pi} \int D_N(y) [f(x-y) - f(x)].$$

We have

$$D_N(y) = \frac{\sin((N + 1/2)y)}{\sin(y/2)} = \frac{\sin(Ny) \cos(y/2)}{\sin(y/2)} + \cos(Ny),$$

so

$$(S_N f)(x) - f(x) = \frac{1}{2\pi} \int \left[\frac{y \cos(y/2)}{\sin(y/2)} \sin(Ny) + y \cos(Ny) \right] \frac{f(x-y) - f(x)}{y} dy.$$

The function

$$g(y) = \frac{f(x-y) - f(x)}{y}$$

is uniformly bounded for small values of y , and bounded by the sum of two integrable functions for large y . Thus g is integrable, hence $yg(y)$ and $g(y) \cos(y/2)(y/\sin(y/2))$ are also integrable. But applying the Riemann Lebesgue lemma implies that as $N \rightarrow \infty$, these integrals vanish, which implies $(S_N f)(x) \rightarrow f(x)$ as $N \rightarrow \infty$. \square

Of course, applying various better decay rates leads to a more uniform version of this theorem. The decay of the Fourier series depends on the decay of the Fourier coefficients of $yg(y)$ and $g(y) \cos(y/2)(y/\sin(y/2))$. In particular, if these coefficients is $O(|n|^{-m})$, then the convergence rate is also $O(|n|^{-m})$. If this decay rate is independent of x for suitable values of x , the convergence will be uniform over these values of x .

2.6 Countercultural Methods of Summation

We now interpret our convergence of series according to a different kernel, so we do get a family of good kernels, and therefore we obtain pointwise convergence for suitable reinterpretations of partial sums. One reason why the Dirichlet kernel fails to be a good kernel is that the Fourier coefficients of the kernel have a sharp drop – the coefficients are either equal to one or to zero. If we mollify, then we will obtain a family of good kernels. And the best way to do this is to alter our summation methods slightly.

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

$$\frac{s_0 + \cdots + s_n}{n+1},$$

A sequence is Cesaro summable to some value if these averages 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 $1 - 1 + 1 - 1 + \cdots = 1/2$, 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 $\lim_{r \rightarrow 1} A_r$ 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 $1 - 2 + 3 - 4 + 5 - \cdots = 1/4$, 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.

2.7 Fejer’s Theorem

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

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

so the oscillations of the Dirichlet kernel are slightly dampened, and as a result, F_N is an approximation to the identity.

Theorem 2.15 (Fejér's Theorem). *For any f ,*

- $(\sigma_N f)(x) \rightarrow f(x)$ for all x in the Lebesgue set of f .
- $\sigma_N f \rightarrow f$ uniformly if $f \in C(\mathbf{T})$.
- $\sigma_N f \rightarrow f$ in the L^p norm for $1 \leq p < \infty$.

Corollary 2.16. *If $\hat{f} = 0$, then $f = 0$ almost everywhere.*

Proof. If $\hat{f} = 0$, then $\sigma_N f = 0$ for all N . But $\sigma_N f \rightarrow f$ in L^1 , which means that $f = 0$ in $L^1(\mathbf{T})$, so $f = 0$ almost everywhere. \square

If we look at the Fourier expansion of the trigonometric polynomial $\sigma_N(f)$, we see that

$$\sigma_N f = \sum_{n=-N}^N \frac{N - |n|}{N} \hat{f}(n) e_n.$$

Thus the Fourier coefficients are slowly added to the expansion, rather than a sharp cutoff as with ordinary Dirichlet summation. This is one reason for the nice convergence properties the kernel has as compared to the Dirichlet kernel.

2.8 The De la Vallée Poisson Kernel

By taking a kernel halfway between the Dirichlet kernel and the Fejer kernel, we can actually obtain important results about ordinary summation. For two integers $M > N$, we define

$$\sigma_{N,M}(f) = \frac{M\sigma_M(f) - N\sigma_N(f)}{M - N}.$$

If we take a look at the Fourier expansion of $\sigma_{n,m}f$, we find

$$\sigma_{N,M}f = \sum_{n=-M}^M \frac{M-|n|}{M-N} e_n - \sum_{n=-N}^N \frac{N-|n|}{M-N} e_n = S_N f + \sum_{|n|=N+1}^M \frac{M-|n|}{M-N} e_n.$$

So we still have a slow decay in the Fourier coefficients. And as a result, if we look at the associated De la Vallée Poisson kernel, we find that a suitable subsequence is an approximation to the identity. In particular, for any fixed integer k , the sequence $\sigma_{kN,(k+1)N}$ leads to a good kernel. More interestingly, if the Fourier coefficients of f have some decay, then the De la Vallée does not differ that much from the ordinary sum, which gives useful results.

Theorem 2.17. *If $\hat{f}(n) = O(|n|^{-1})$, then for any integers N and k , if $kN \leq M < (k+1)N$, then*

$$\|\sigma_{kN,(k+1)N}f - S_M f\|_{L^\infty(\mathbf{T})} \lesssim 1/k.$$

Where the implicit constant is independant of N and k .

Proof. We just calculate that, since the Poisson sum has essentially the same weight for low term coefficients as the sum $S_M f$,

$$\|\sigma_{kN,(k+1)N}f - S_M f\|_{L^\infty(\mathbf{T})} \lesssim \sum_{kN \leq |n| < (k+1)N} |\hat{f}(n)| \lesssim \sum_{n=kN}^{(k+1)N} \frac{1}{n} \leq \frac{N}{kN} = \frac{1}{k}. \quad \square$$

Corollary 2.18. *If f is a function with $\hat{f}(n) = O(|n|^{-1})$,*

- $S_N f$ converges to f in the L^p norm for $1 \leq p < \infty$.
- $S_N f$ converges uniformly to f if $f \in C(\mathbf{T})$.
- $(S_N f)(x) \rightarrow f(x)$ for each Lebesgue point x of f .

Proof. The idea is quite simple. Fix N . Given any ε , we can use the last theorem to find k large enough such that if $kN \leq M < k(N+1)$,

$$\|\sigma_{kN,(k+1)N}f - S_M f\|_{L^\infty(\mathbf{T})} \leq \varepsilon.$$

But this gives the first and second result, up to perhaps a ε of error. The latter result is given by similar techniques. \square

In particular, if $f \in C(\mathbf{T})$, and has a weak derivative in $L^1(\mathbf{T})$, then $S_N f \rightarrow f$ uniformly.

2.9 Pointwise Behaviour at Discontinuity Points

We have established results about the pointwise convergence of the Fourier series at continuity points. We now discuss the behaviour of the series at ‘simple’ discontinuity points. First, an integral example.

Example. Consider the sawtooth function defined on $[-\pi, \pi]$ by $s(t) = t$, and then made periodic on the entire real line. We can easily calculate the Fourier series here, obtaining that

$$s(t) = i \sum_{n \neq 0} \frac{(-1)^n e_n(t)}{t} = -2 \sum_{n=1}^{\infty} (-1)^n \frac{\sin(nt)}{n}.$$

Our previous results imply that since s is smooth almost everywhere, for $t \notin \pi + 2\pi\mathbf{Z}$, $(S_N s)(t) \rightarrow f(t)$. If m is given, we calculate

$$(S_N s)(\pi + 2\pi m) = -2 \sum_{n=1}^N (-1)^n \frac{\sin(n\pi)}{n} = 0.$$

So $(S_N s)(\pi + 2\pi m) \rightarrow 0$. Notice that 0 is neither $\lim_{t \downarrow -\pi} s(t) = -\pi$, nor $\lim_{t \uparrow \pi} s(t) = \pi$, but the value halfway between the two.

The convergence at the continuity of the sawtooth function is characteristic of a large class of discontinuities.

Theorem 2.19. If $\hat{f}(n) = O(|n|^{-1})$, and $f(t_0-)$ and $f(t_0+)$ exist, then

$$(S_N f)(t_0) \rightarrow \frac{f(t_0-) + f(t_0+)}{2}.$$

Proof. The idea of our proof is to break f into a nice continuous function, and the sawtooth function, where we already understand the convergence of Fourier series. Without loss of generality, let $t_0 = \pi$. Define $g(t) = f(t) + (2\pi)^{-1}(f(\pi+) - f(\pi-))s(t)$ on $(-\pi, \pi)$, where s is the sawtooth function. Then

$$\lim_{t \uparrow \pi} g(t) = \lim_{t \downarrow -\pi} g(t) = \frac{f(\pi+) + f(\pi-)}{2}.$$

Thus g can be defined at $\pi + 2\pi\mathbf{Z}$ so it is continuous there. Now we find $|\hat{g}| \lesssim |\hat{f}| + |\hat{s}| = O(|n|^{-1})$, and so

$$(S_N g)(\pi) \rightarrow \frac{f(\pi+) + f(\pi-)}{2}.$$

We also have $(S_N s)(\pi) \rightarrow 0$. Thus

$$(S_N f)(\pi) = (S_N g)(\pi) - (S_N s)(\pi) \rightarrow \frac{f(\pi+) + f(\pi-)}{2}. \quad \square$$

This isn't the end of our discussion about points of discontinuity. There is an interesting phenomenon which occurs locally around the point of discontinuity. If f is continuous locally around a discontinuity point t_0 , $S_N f \rightarrow f$ pointwise locally around t_0 . Thus, being continuous, $S_N f$ must 'jump' from $(S_N f)(t_0-)$ to $(S_N f)(t_0+)$ locally around t_0 . Interestingly enough, we find that the jump is not precise, the jump is overshoot and then must be corrected to the left and right of t_0 . This is known as the *Gibb's phenomenon*, after the man who clarified the reason for why this phenomenon occurred in physical measurements where first thought to be a defect in the equipment used to take the measurements.

Theorem 2.20. *Given f with finitely many discontinuity points, and with $\hat{f}(n) = O(|n|^{-1})$, in particular one at t_0 , we find*

$$\lim_{N \rightarrow \infty} (S_N f)(t_0 \pm \pi/N) = f(t_0 \pm) \pm C \cdot \frac{f(t_0) - f(t_0-)}{2},$$

where

$$C = 2\pi \int_0^\pi (\sin x)/x \approx 16.610.$$

Proof. First consider the jump function s . Then

$$(S_N s)(\pi + \pi/N) = -2 \sum_{n=1}^N \frac{\sin(n\pi/N)}{n} = -2\pi \sum_{n=1}^N \frac{1}{N} \left(\frac{\sin(n\pi/N)}{\pi n/N} \right).$$

Here we're just taking averages of values of $\sin(x)/x$ at $x = \pi/N, x = 2\pi/N$, and so on and so forth. Thus is a Riemann sum, so as $N \rightarrow \infty$, we get that

$$(S_N s)(\pi + \pi/N) \rightarrow -2\pi \int_0^\pi \frac{\sin x}{x}.$$

The same calculations give

$$(S_N s)(\pi - \pi/N) \rightarrow 2\pi \int_0^\pi \frac{\sin x}{x}.$$

In general, given f , we can write $f = g + \sum \lambda_j h_j$, where g is continuous, and h_j is a translate of the sawtooth function. Then $S_N g$ converges to g uniformly, and $S_N h_j \rightarrow 0$ for all h_j uniformly in an interval outside of their discontinuity point. To see this, we note that an integration by parts gives

$$\left| \int_{-\pi}^{\pi} D_N(y)[s(x-y) - s(x)] dy \right| \leq |G_N(x - \pi)|,$$

where $G_N(y) = -i \sum_{|n| \leq N} e_n(t)/n$, so $G'_N = D_N$. It now suffices to show $G_N(x - \pi) \rightarrow 0$ outside a neighbourhood of π . But if $A(u, t) = \sum_{|n| \leq u} e_n(t)$, summation by parts gives

$$\sum_{|n| \leq N} \frac{e_n(t)}{n} = \frac{A(N, t)}{N} + \int_1^N \frac{A(u, t)}{u^2}.$$

Now a simple geometric sum shows $A(u, t) \lesssim 1/|e(t) - 1|$, so provided $d(t, 2\pi\mathbf{Z})$ is bounded below, the quantity above tends to zero uniformly. This gives the required result. \square

2.10 Tchebychev Polynomials

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

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

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

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

Thus we have the relation $T_{m+1}(x) = 2xT_m(x) - T_{m-1}(x)$. These polynomials are known as **Tchebyshev polynomials**, enabling us to move between ‘periodic coordinates’ and standard Euclidean coordinates.

Corollary 2.21 (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, \dots, 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 $x = \cos t$, for $\cos t \geq 0$, this equation says

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

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

2.11 Exponential Sums and Equidistribution

The next result uses Fourier analysis to characterize the asymptotic distribution of a certain sequence a_1, a_2, \dots . In particular, it is most useful in determining when this distribution is distributed when we consider $2\pi a_1, 2\pi a_2, \dots$ as elements of \mathbf{T} , i.e. so we only care about the fractional part of the numbers, or in other terms their behaviour modulo one. We say the sequence is *uniformly distributed* if for any interval $I \subset \mathbf{T}$, $\#\{2\pi a_n \in I : n \leq N\} \sim N|I|$ as $N \rightarrow \infty$. By approximating continuous functions by step functions, this implies that if $f : \mathbf{T} \rightarrow \mathbf{C}$ is continuous, then

$$\frac{f(2\pi a_1) + \dots + f(2\pi a_N)}{N} \rightarrow \frac{1}{2\pi} \int_{\mathbf{T}} f(t) dt.$$

It is the right hand side to which we can apply Fourier summation to obtain a very useful condition. We let $S_N f$ denote the left hand side of the equation, and Tf the right hand side.

Theorem 2.22 (Weyl Condition). *A sequence $a_1, a_2, \dots \in \mathbf{T}$ is uniformly distributed if and only if for every n , as $N \rightarrow \infty$, $e_n(2\pi a_1) + \dots + e_n(2\pi a_N) = o(N)$.*

Proof. The condition in the theorem implies that for any trigonometric polynomial f , $S_N f \rightarrow Tf$. The S_N are uniformly bounded as functions on $L^\infty(\mathbf{T})$, and T is a bounded functional on this space as well. But this means that $\lim S_N f = Tf$ for all f in $C(\mathbf{T})$, since this equation holds on the dense subset of trigonometric polynomials. \square

This technique enables us to completely characterize the equidistribution behaviour of arithmetic sequences. Given a particular γ , we consider the equidistribution of the sequence $\gamma, 2\gamma, \dots$, which depends on the irrationality of γ .

Example. *Let γ be an arbitrary real number. Then for any n , if $e_n(2\pi\gamma) \neq 1$,*

$$\sum_{m=1}^N e_n(2\pi m\gamma) = \frac{e_n(2\pi(N+1)\gamma) - 1}{e_n(2\pi\gamma) - 1} \lesssim 1 = o(N).$$

If γ is an irrational number, then $e_n(2\pi\gamma) \neq 1$ for all n , which implies that $\gamma, 2\gamma, \dots$ is equidistributed. Conversely, if $e_n(2\pi\gamma) = 1$ for some n , we have

$$\sum_{m=1}^N e_n(a_m) = N.$$

which is not $o(N)$, so the sequence $\gamma, 2\gamma, \dots$ is not equidistributed. If γ is rational, there certainly is n such that $n\gamma \in \mathbf{Z}$, and so $e_n(2\pi\gamma) = 1$.

On the other hand, it is still an open research to characterize, for which γ the sequence $\gamma, \gamma^2, \gamma^3, \dots$ is equidistributed. Here is an example showing that there are γ for which the sequence is not equidistributed.

Example. *Let γ be the golden ratio $(1 + \sqrt{5})/2$. Consider the sequence*

$$a_n = \left(\frac{1 + \sqrt{5}}{2} \right)^n + \left(\frac{1 - \sqrt{5}}{2} \right)^n = b_n + c_n.$$

Then one checks that a_n is a kind of Fibonacci sequence, with $a_{n+1} = a_n + a_{n-1}$, and initial conditions $a_0 = 2$, $a_1 = 1$. One checks that c_n is always negative for odd n , and positive for even n , and tends to zero as $n \rightarrow \infty$. Since a_n is an integer, this means that $d(b_n, \mathbf{Z}) = d(\gamma^n, \mathbf{Z}) \rightarrow 0$. But this means that the average distribution of the γ^n modulo one is concentrated at the origin.

2.12 Abel Summation and Harmonics on the Disk

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

$$A_r(f) = \sum_{n=-\infty}^{\infty} \hat{f}(n) r^n e_n(t).$$

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

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

converge uniformly over \mathbf{T} , 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. To see this, we can apply an infinite geometric series summation to obtain that

$$\begin{aligned} \sum r^{|n|} e_n(t) &= 1 + \frac{re(t)}{1 - re(t)} + \frac{re(-t)}{1 - re(-t)} = 1 + \frac{2r \cos t - 2r^2}{(1 - re(t))(1 - re(-t))} \\ &= 1 + \frac{2r \cos t - 2r^2}{1 - 2r \cos t + r^2} = \frac{1 - r^2}{1 - 2r \cos t + r^2}. \end{aligned}$$

As $r \rightarrow 1$, the function concentrates at the origin, because as $r \rightarrow 1$, if $\delta \leq |t| \leq \pi$, then $1 - \cos t$ is bounded away from the origin, so

$$\begin{aligned} \left| \frac{1 - r^2}{1 - 2r \cos t + r^2} \right| &= \left| \frac{1 + r}{(1 + (1 - 2 \cos t)r) + 2(1 - \cos t)r^2/(1 - r)} \right| \\ &= O\left(\frac{1 - r}{1 - \cos t}\right) = O_\delta(1 - r). \end{aligned}$$

Thus the oscillation in the Poisson kernel cancels out as $r \rightarrow 1$, and so the Poisson kernel is a good kernel.

Theorem 2.23. *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.*

The Poisson kernel is not a trigonometric polynomial, and therefore not quite as easy to work with as the Féjer kernel. However, it is the real part of the Cauchy kernel

$$\frac{1 + re(t)}{1 - re(t)}.$$

and therefore links the study of trigonometric series and the theory of analytic functions. Furthermore, it is very important to solving the problem of heat distribution on a disk. Using the Abel summability properties of functions, we can solve Laplace's equation on a disk uniquely given any boundary conditions.

Theorem 2.24. *If f is integrable on \mathbf{T} , then in polar coordinates, the function $u(r, \theta) = (f * P_r)(\theta)$ is the unique $C^2(\mathbf{D}^\circ)$ harmonic function such that at any point of continuity θ of f , $\lim_{\theta \rightarrow 1} u(r, \theta) = f(\theta)$. If f is continuous everywhere, then this limit is uniform, and in this case u is the unique $C^2(\mathbf{D}^\circ)$ function with these properties.*

Proof. Because P_r is C^∞ for all $r < 1$, $f * P_r$ is C^∞ , and so

$$u(r, \theta) = \sum \hat{f}(n) r^{|n|} e_n(\theta),$$

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

$$u_{\theta\theta} = - \sum \hat{f}(n) n^2 r^{|n|} e_n(t),$$

$$(f * P_r)_r = \sum \hat{f}(n) |n| r^{|n|-1} e_n(t) \quad (f * P_r)_{rr} = \sum \hat{f}(n) |n| (|n| - 1) r^{|n|-2} e_n(t).$$

Now we find that on the interior of the unit disk,

$$\Delta u = u_{rr} + u_r/r + u_{\theta\theta}/r^2 = \sum [|n|(|n| - 1) + |n| - n^2] \hat{f}(n) r^{|n|-2} e_n(t) = 0.$$

Abel summability results imply that $u(re(t)) \rightarrow f(t)$ 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 $u_r(t) = \sum a_n(r)e_n(t)$. The $a_n(r)$ are defined as an integral formula, and thus are bounded, differentiable functions, with

$$a'_n(r) = \frac{1}{2\pi} \int_{\mathbf{T}} u_r(r, \theta) e_n(-t) \quad a''_n(r) = \frac{1}{2\pi} \int_{\mathbf{T}} u_{rr}(r, \theta) e_n(-t),$$

$$\frac{1}{2\pi} \int_{\mathbf{T}} \ddot{u}(r, \theta) e_n(t) = -n^2 a_n(r).$$

So $a''_n + a'_n/r - n^2 a_n/r^2 = 0$, which is an ordinary differential equation whose only bounded solution is $a_n(r) = A_n r^n$. If we let $r \rightarrow \infty$, we conclude that

$$A_n = \lim_{r \rightarrow 1} \frac{1}{2\pi} \int_{\mathbf{T}} u(re(t)) e_n(-t) = \frac{1}{2\pi} \int_{\mathbf{T}} u(e(t)) e_n(-t) = \hat{f}(n).$$

Thus u is just $f * P_r$. □

2.13 The Isoperimetric Inequality

2.14 Heat Propagation Into the Ground

Let us consider an application of the Fourier series taken from Fourier's original work. Consider heat moving from above ground to below ground, and vice versa. If we let $H(t, y)$ denote the temperature at a depth y into the ground at time t , for $y > 0$. Assuming that the material of the ground is homogenous, by choosing appropriate units, the differential equation becomes $H_t = H_{yy}$, a variant of the heat equation. We assume that the heat at the surface changes periodically over the days and seasons, so

$$H(t, 0) = A \cos(2\pi t/D) + B \cos(2\pi t/Y) + C,$$

where A, B, C are arbitrary constants, D is the length of a day, and Y is the length of a year, so $Y = 365D$. In our calculation, we assume the regularity condition that $H \in L^\infty[0, \infty)^2$, so the temperature does not magnify infinitely at large depths or large times.

To solve this equation, we use two tricks: linearity, and Fourier series. We can solve the heat equation by solving the three heat equations with initial conditions $H_D(t, 0) = \cos(2\pi t/D)$, $H_Y(t, 0) = \cos(2\pi t/Y)$, and $H_C(t, 0) = 1$, and then obtain a general solution by letting $H = AH_D +$

$BH_Y + CH_C$. The third equation is easiest: we let $H_C(t, y) = 1$ for all t and y . To solve the other equations, we can use variable separation. Assuming H_D and H_Y are bounded, this forces

$$H_D(t, y) \propto \cos((2\pi/D)t - (\pi/D)^{1/2}y)e^{-(\pi/D)^{1/2}y},$$

$$H_Y(t, y) \propto \cos((2\pi/Y)t - (\pi/Y)^{1/2}y)e^{-(\pi/Y)^{1/2}y}.$$

Thus the temperature in the ground splits into the daily heating effect, the seasonal heating effect, and a constant temperature. From these equations we get several interesting qualitative properties. As we go deeper into the ground, the temperature decays at a rate inversely dependant on the length of time, so even at small depths, the daily temperature becomes negligible, and only the seasonal temperature is important. Experimentally, determining the constants in our equation, we determine this happens about half a foot into the ground. Next, the deeper we go in the ground, the more a ‘time lag’ exists, where the seasonal temperature back in time has now travelled to the temperature at the current point in the ground. Experimentally, we determine that about 2-3 metres below ground, the temperature lags by six months. Fourier mentions this is a good depth to build a wine cellar.

2.15 Seafaring with Fourier

Here we discuss two problems in seafaring that can be solved quite accurately with Fourier analysis, first done by Kelvin in the late 1800s. Consider first the problem in determining the error of compass measurement on a ship when taking an initial bearing at harbor travelling. Thus for each angle θ , we consider an error $g(\theta)$ such that if, at an angle θ , we take a measurement $f(\theta)$, then $f(\theta) = \theta + g(\theta)$. Often g is up to 20 degrees, but it will suffice to know g up to an angle of two or three degrees, since other systematic errors in travel disturb the angle the ship actually travels by this amount anyway. And thus experimentally we find it suffices to approximate g by a degree four trigonometric polynomial, i.e. we substitute g for an approximate value

$$g_1(\theta) = A_0 + A_1 \cos \theta + B_1 \sin \theta + A_2 \cos(2\theta) + B_2 \sin(2\theta).$$

We can obtain measurements $g(\theta)$ for certain values of θ by locating landmarks, and 6 measurements suffice to uniquely identify g_1 from all other degree five trigonometric polynomials.

Another seafaring problem is to determine the future height of the tide. We expect the height of the tides to be due to periodic forces in nature. If $h(t)$ is the height of the tide, we might expect by linearity of the wave equation that $h(t) = h_1(t) + h_2(t) + \dots$, where $h_1(t)$ is the height with relation to the rotation of the earth and the moon, $h_2(t)$ the rotation of the earth and the sun, and so on and so forth to more negligible values. Each h_k is periodic with some period ω_k . If we assume that each h_k is a trigonometric polynomial, then there is a way to reduce the calculation of the coefficients to a certain integral formula which one can approximate by taking samples of the height of the tides over time. Unfortunately, one must take a large number of samples to obtain this integral formula, but Kelvin designed one of the first automated calculators to approximate this.

Theorem 2.25. *If $h(t) = \sum_{n=1}^N A_n \cos(\omega_n t)$, where the ω_n are all different periods, then for any S ,*

$$A_n = \frac{2}{T} \lim_{n \rightarrow \infty} \int_S^{S+T} h(t) \cos(\omega_n t) dt.$$

Proof. We just change variables. If $2\pi N/\omega_n < T \leq 2\pi(N+1)/\omega_n$,

$$\begin{aligned} \int_S^{S+T} h(t) \cos(\omega_n t) dt &= N \int_0^{2\pi/\omega_n} h(t) \cos(\omega_n t) dt + O(1) \\ &= \frac{N}{\omega_n} \int_0^{2\pi} \frac{1}{N} \sum_{k=1}^N h(S + t/\omega_n + 2\pi k/\omega_n) \cos(t) dt + O(1). \end{aligned}$$

The average in the integral causes all oscillation not corresponding to ω_n to cancel out, so

$$\frac{2}{T} \int_S^{S+T} h(t) \cos(\omega_n t) dt \rightarrow \frac{2}{\omega_n} \left(A_n \int_0^{2\pi} \cos^2(t) + B_n \int_0^{2\pi} \cos(t) \sin(t) \right). \quad \square$$

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 for not necessarily periodic functions on the 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

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

is the Fourier inversion formula

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

where for each real number ξ , we define

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

The function \hat{f} is known as the **Fourier transform** of the function f . It is also denoted by $\mathcal{F}(f)$. 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 $f : \mathbf{R}^n \rightarrow \mathbf{R}^n$, considers the quantities

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

for $\xi \in \mathbf{R}^n$, where $e(t) = \exp(2\pi it)$ for any $t \in \mathbf{R}$.

Later, we will interpret the Fourier transform in a very general manner for a very arbitrary class of functions. But first we must interpret the Fourier transform as a Lebesgue integral, and the weakest assumptions we can make in order to do this are that f is an integrable function, i.e. that $f \in L^1(\mathbf{R}^d)$. During arguments, we can often assume additional regularity properties of f , and then apply density arguments to get the result in general. Most of the properties of the Fourier transform are exactly the same as for Fourier series. The only new phenomenon in the basic theory is that the Fourier transform of an integrable function is continuous.

Theorem 3.1. *For any $f \in L^1(\mathbf{R}^d)$, $\|\widehat{f}\|_{L^\infty(\mathbf{R}^d)} \leq \|f\|_{L^1(\mathbf{R}^d)}$, and $\widehat{f} \in C_0(\mathbf{R}^d)$.*

Proof. For any $\xi \in \mathbf{R}^d$,

$$|\widehat{f}(\xi)| = \left| \int f(x) e(-\xi \cdot x) dx \right| \leq \int |f(x)| |e(-\xi \cdot x)| dx = \|f\|_{L^1(\mathbf{R}^d)}.$$

If χ_I is the characteristic function of an n dimensional box, i.e.

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

then

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

where

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

L'Hopital's rule shows $\widehat{\chi_{I_k}}$ is a continuous function. We also have the upper bound

$$\widehat{\chi_{I_k}}(\xi_k) \lesssim_{I_k} (1 + |\xi_k|)^{-1}$$

for all $\xi_k \in \mathbf{R}$, which implies that

$$\widehat{\chi_I}(\xi) = \prod \widehat{\chi_{I_k}}(\xi_k) \lesssim_I \prod \frac{1}{1 + |\xi_k|} \lesssim_n \frac{1}{1 + |\xi|}.$$

Thus $\widehat{\chi_I}(\xi) \rightarrow 0$ as $|\xi| \rightarrow \infty$. But this implies the Fourier transform of any step function is continuous and vanishes at ∞ . Since step functions are dense in $L^1(\mathbf{R}^d)$, a density argument then gives the result for all integrable functions. \square

Remark. The space

$$\mathbf{A}(\mathbf{R}^d) = \left\{ \widehat{f} : f \in L^1(\mathbf{R}^d) \right\}$$

is called the *Fourier algebra*. The last theorem shows $\mathbf{A}(\mathbf{R}^d) \subset C_0(\mathbf{R}^d)$, but it is *not* the case that $\mathbf{A}(\mathbf{R}^d) = C_0(\mathbf{R}^d)$. As of yet, not current research has given a satisfactory description of the elements of $\mathbf{A}(\mathbf{R}^d)$.

Lemma 3.2. *For any $0 \leq a < b < \infty$, independantly of a and b ,*

$$\left| \int_a^b \frac{\sin x}{x} \right| = O(1).$$

Proof. Since $|\sin(x)/x| \leq 1$ for all x , we may assume $b > 1$, for otherwise we obtain a trivial bound. This also implies

$$\left| \int_a^b \frac{\sin x}{x} dx \right| \leq 1 + \left| \int_1^b \frac{\sin x}{x} dx \right|.$$

An integration by parts then shows that

$$\left| \int_1^b \frac{\sin x}{x} dx \right| \leq \left| \left(\cos 1 - \frac{\cos b}{b} \right) \right| + \left| \int_1^b \frac{\cos x}{x^2} dx \right| \lesssim 1. \quad \square$$

Theorem 3.3. $\mathbf{A}(\mathbf{R}) \neq C_0(\mathbf{R})$. *In particular, $\mathbf{A}(\mathbf{R})$ does not contain any odd functions g in $C_0(\mathbf{R})$ such that*

$$\limsup_{b \rightarrow \infty} \left| \int_1^b \frac{g(\xi)}{\xi} d\xi \right| = \infty.$$

Proof. Suppose $f \in L^1(\mathbf{R})$, and $\hat{f} \in C_0(\mathbf{R})$ is an odd function. Then we know

$$\hat{f}(\xi) = -i \int_{-\infty}^{\infty} f(x) \sin(2\pi \xi x) dx.$$

Thus an application of Fubini's theorem shows that

$$\left| \int_1^b \frac{\hat{f}(\xi)}{\xi} d\xi \right| = \left| \int_{-\infty}^{\infty} f(x) \left(\int_1^b \frac{\sin(2\pi \xi x)}{\xi} d\xi \right) dx \right|.$$

But

$$\left| \int_1^b \frac{\sin(2\pi \xi x)}{\xi} d\xi \right| = \left| \int_{2\pi x}^{2\pi bx} \frac{\sin \xi}{\xi} d\xi \right| \lesssim 1.$$

Thus we obtain that

$$\left| \int_1^b \frac{\hat{f}(\xi)}{\xi} d\xi \right| \lesssim \|f\|_{L^1(\mathbf{R})}. \quad \square$$

Elementary properties of integration give the following relations among the Fourier transforms of functions on \mathbf{R}^d . They are strongly related to the translation invariance of the Lebesgue integral on \mathbf{R}^d :

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

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

If f is real, the formula above says $\hat{f}(\xi) = \overline{\hat{f}(-\xi)}$, and so if we define $a(\xi) = \operatorname{Re}(\hat{f}(\xi))$, $b(\xi) = \operatorname{Im}(\hat{f}(\xi))$, then formally we have

$$\int_{-\infty}^{\infty} \hat{f}(\xi) e(\xi x) d\xi = 2 \int_0^{\infty} a(\xi) \cos(2\pi \xi x) - b(\xi) \sin(2\pi \xi x) d\xi.$$

Thus the Fourier representation formula expresses the function f as an integral in sines and cosines.

- There is a duality between translation and frequency modulation. For $y \in \mathbf{R}^d$, we define $(T_y f)(x) = f(x - y)$. If $\xi \in \mathbf{R}^d$, then we define

$(M_\xi f)(x) = e(\xi \cdot x)f(x)$. We then find that

$$\begin{aligned}\widehat{T_y f}(\xi) &= \int f(x-y)e(-\xi \cdot x) dx \\ &= e(-\xi \cdot y) \int f(x)e(-\xi \cdot x) dx = (M_{-y}\widehat{f})(\xi).\end{aligned}$$

and

$$\widehat{M_\xi f}(\eta) = \int e(\xi \cdot x)f(x)e(-\eta \cdot x) dx = \widehat{f}(\eta - \xi) = (T_\xi \widehat{f})(\eta).$$

Thus we conclude $\mathcal{F} \circ T_y = M_{-y} \circ \mathcal{F}$, and $\mathcal{F} \circ M_\xi = T_\xi \circ \mathcal{F}$.

- Let $T : \mathbf{R}^d \rightarrow \mathbf{R}^d$ be an invertible linear transformation. Then a change of variables $y = Tx$ gives

$$\begin{aligned}\widehat{f \circ T}(\xi) &= \int f(Tx)e(-\xi \cdot x) dx \\ &= \frac{1}{|\det(T)|} \int f(y)e(-\xi \cdot T^{-1}y) dy \\ &= \frac{1}{|\det(T)|} \int f(y)e(-T^{-t}\xi \cdot y) dy \\ &= \frac{1}{|\det(T)|} (\widehat{f} \circ T^{-t})(\xi).\end{aligned}$$

Thus we conclude that if $T^*(f) = f \circ T$, then $\mathcal{F} \circ T^* = |\det(T)|^{-1} (T^{-t})^* \circ \mathcal{F}$.

- As a special case of the theorem above, if $a \in \mathbf{R}$ and $(D_a f)(x) = f(ax)$, then

$$\widehat{D_a f}(\xi) = a^{-d} \widehat{f}(\xi/a)$$

If we dilate by a small value of a , then the values of f are traced over more slowly, so $D_a f$ has smaller frequencies. But the magnitude of these frequencies is increase to compensate.

- If $R \in O_n(\mathbf{R})$, then $\widehat{f \circ R}(\xi) = \widehat{f}(R\xi)$, i.e. $\mathcal{F} \circ R^* = R^* \circ \mathcal{F}$. In particular, if f is a radial function, so $f \circ R = f$ for any R , then $\widehat{f}(R\xi) = \widehat{f}(\xi)$ for any $R \in O_n(\mathbf{R})$, so \widehat{f} is also a radial function. If f is even, so $f(x) = f(-x)$ for all x , then $\widehat{f}(\xi) = \widehat{f}(-\xi)$ for all ξ , so \widehat{f} is even. Similarly, if f is odd, then \widehat{f} is odd.

- Given $f, g \in L^1(\mathbf{R}^d)$, we define the convolution

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

This convolution possesses precisely the same properties as convolution on \mathbf{T} . Most importantly for us,

$$\mathcal{F}(f * g) = \mathcal{F}(f) \cdot \mathcal{F}(g),$$

so convolution in phase space is just a product in frequency space.

Just as with Fourier series, we have a duality between decay of a function and smoothness of it's transform. We say f has a **strong derivative** f_k in $L^p(\mathbf{R}^d)$ if the family of functions

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

converge in $L^p(\mathbf{R}^d)$ to f_k . Essentially, this means that the approximations of f to it's derivative quantitatively converge in the mean. If f has a strong derivative in $L^p(\mathbf{R}^d)$, then f_k is actually differentiable almost everywhere. However, even if f has a pointwise partial derivative f_k , the differences $\Delta_h f$ may not converge to f_k fast enough to conclude that f has a strong derivative. It is fairly easy to prove using the mean value theorem that if f converges to f_k in the L^∞ norm, and f has compact support, then f has a strong derivative in all other L^p spaces. If f is not compactly supported, but decays rapidly at ∞ , then it is often the case that the L^p derivative will be the same as the classical derivative. In particular, this is true of a Schwartz function.

Theorem 3.4. *If $f \in L^1(\mathbf{R}^d)$, and $x_k f \in L^1(\mathbf{R}^d)$, then \hat{f} has a strong derivative in the L^∞ norm, and $\hat{f}_k(\xi) = -2\pi i(x_k f)^\wedge(\xi)$.*

Proof. Note that a change of variables implies

$$(\Delta_h \hat{f})(\xi) = \int f(x) \frac{e(-hx_k) - 1}{h} e(-\xi \cdot x) dx = \hat{g}_h(\xi),$$

where

$$g_h(x) = f(x) \frac{e(hx_k) - 1}{h}.$$

Note that

$$\left| \frac{e(hx_k) - 1}{h} \right| = O(1 + |x_k|).$$

Since $x_k f$ is integrable, we can apply the dominated convergence theorem. Because $(e(hx_k) - 1)/h$ tends to $-2\pi i x_k f(x)$ as $h \downarrow 0$, the function g_h tends to $-2\pi i x_k f$ in $L^1(\mathbf{R}^d)$. Taking Fourier transforms, we conclude that $\Delta_h \hat{f} = \hat{g}_h$ converges uniformly to $(-2\pi i x_k f)^\wedge(\xi)$. \square

Remark. In particular, the Fourier transform of a compactly supported function lies in $C^\infty(\mathbf{R}^d)$, and has strong derivatives of all orders.

Theorem 3.5. *If f has a strong derivative f_k in the L^1 norm, $\hat{f}_k(\xi) = 2\pi i \xi_k \hat{f}(\xi)$.*

Proof. It suffices to note that

$$\widehat{\Delta_h f}(\xi) = \frac{e(h\xi_k) - 1}{h} \hat{f}(\xi).$$

Since $\Delta_h f \rightarrow f_k$ in L^1 , $\widehat{\Delta_h f} \rightarrow \hat{f}_k$ uniformly, and in particular, converges to \hat{f}_k pointwise. But we know $\widehat{\Delta_h f}$ converges pointwise to $2\pi i \xi_k \hat{f}(\xi)$. \square

3.1 Convergence Using Alternative Summation

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

$$f(x) = \int \hat{f}(\xi) e(\xi \cdot x) dx$$

does not hold for every integrable f , nor even for all continuous f . In particular, the Fourier transform of f need not even lie in $L^1(\mathbf{R}^d)$, so the integral formula may not even make sense. Nonetheless, just as with Fourier series, one can obtain general results by ‘dampening’ the integration.

Example. *Even if f is a non integrable function, the functions $f(x)e^{-\delta|x|}$ may be integrable for $\delta > 0$. We say f is Abel summable to a value A if*

$$\lim_{\delta \rightarrow 0} \int f(x) e^{-\delta|x|} dx = A$$

For each $\delta > 0$ and $f \in L^1(\mathbf{R}^d)$, we let

$$(A_\delta f)(x) = \int \hat{f}(\xi) e(\xi \cdot x) e^{-\delta|\xi|} d\xi.$$

If $f \in L^1(\mathbf{R}^d)$, then the dominated convergence theorem implies that

$$\int f(x) e^{-\delta|x|} dx \rightarrow \int f(x) dx.$$

so f is Abel summable. 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.

Example. Similarly, we can consider the Gauss sums

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

We say f is Gauss summable to if these values converge as $\delta \rightarrow 0$. For $f \in L^1(\mathbf{R}^d)$, we let

$$(G_\delta f)(x) = \int \hat{f}(\xi) e(\xi \cdot x) e^{-\delta|\xi|^2} d\xi.$$

Example. For $d = 1$, we can also consider the Fejér sums

$$(\sigma_\delta f)(x) = \int_{-\infty}^{\infty} \hat{f}(\xi) e(\xi \cdot x) \left(\frac{\sin(\delta\pi\xi)}{\delta\pi\xi} \right)^2 d\xi.$$

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

$$\int_{-\infty}^{\infty} f(x) dx = \lim_{t \rightarrow \infty} \int_{-t}^t f(x) dx.$$

These integrals can be written as the integral of $f \chi_{[-t,t]}$, and so in a generalized sense, we can integrate a function f if $f \chi_{[-t,t]}$ is integrable for each N , and the integrals of these functions converge as $t \rightarrow \infty$. Thus we study

$$(S_R f)(x) = \int_{-R}^R \hat{f}(\xi) e(\xi \cdot x) d\xi.$$

Abel summability is more general than the piecewise limit integral considered in the last example, as the next lemma proves.

Lemma 3.6. Suppose $f \in L^1_{loc}(\mathbf{R}^d)$, that

$$\lim_{t \rightarrow \infty} \int_{-t}^t f(x) dx$$

exists, and that $f(x)e^{-\delta x^2}$ is absolutely integrable for each $\delta > 0$. Then f is Abel summable, and

$$\lim_{\delta \rightarrow 0} \int_{-\infty}^{\infty} f(x) e^{-\delta |x|^2} dx = \lim_{t \rightarrow \infty} \int_{-t}^t f(x) dx.$$

Proof. Let

$$\lim_{t \rightarrow \infty} \int_{-t}^t f(x) dx = A.$$

For each $x \geq 0$, write

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

Then F is continuous, and $F(x) \rightarrow A$ as $x \rightarrow \infty$. We know that $F'(x) = f(x) + f(-x)$, and an integration by parts gives for each $s > 0$,

$$\int_{-s}^s f(x) e^{-\delta x^2} dx = \int_0^s [f(x) + f(-x)] e^{-\delta x^2} dx = F(s) e^{-\delta s^2} + 2\delta \int_0^s x F(x) e^{-\delta x^2} dx.$$

Taking $s \rightarrow \infty$, using the fact that F is bounded so that $F(s) e^{-\delta s^2} \rightarrow 0$, we conclude

$$\int f(x) e^{-\delta x^2} dx = 2\delta \int_0^{\infty} x F(x) e^{-\delta x^2} dx.$$

Given $\varepsilon > 0$, fix t such that $|F(s) - A| \leq \varepsilon$ for $s \geq t$. Then

$$\begin{aligned} \left| \int f(x) e^{-\delta x^2} dx - A \right| &\leq 2\delta \left| \int_0^t x F(x) e^{-\delta x^2} dx \right| \\ &\quad + 2\delta \varepsilon \left| \int_t^{\infty} x e^{-\delta x^2} dx \right| \\ &\quad + \left| 2\delta A \int_t^{\infty} x e^{-\delta x^2} dx - A \right|. \end{aligned}$$

The first and second components of this upper bound can each be made smaller than ε for small enough δ . And

$$2\delta \int_t^{\infty} x e^{-\delta x^2} dx = e^{-\delta t^2}$$

So the 3rd term is equal to

$$|A| |1 - e^{-\delta t^2}|$$

and for small enough δ , we can also bound this by ε . Thus we have shown for small enough δ that

$$\left| \int f(x) e^{-\delta x^2} dx - A \right| \leq 3\varepsilon.$$

It now suffices to take $\varepsilon \rightarrow 0$. □

Abel summation is even more general than Gauss summation.

Lemma 3.7. *If f is Gauss summable, and $f(x)e^{-\delta|x|}$ is absolutely integrable for each $\delta > 0$, then f is Abel summable, and*

$$\lim_{\delta \rightarrow 0} \int f(x) e^{-\delta|x|^2} dx = \lim_{\delta \rightarrow 0} \int f(x) e^{-\delta|x|} dx.$$

Proof. Let

$$\lim_{\delta \rightarrow 0} \int f(x) e^{-\delta|x|^2} dx = A.$$

If there existed constants c_n and λ_n such that $e^{-\delta|x|} = \sum c_n e^{-(\lambda_n \delta |x|)^2}$, this theorem would be easy. This is not exactly true, but we do have the *subordination principle*, which says

$$e^{-\delta|x|} = \int_0^\infty \frac{e^{-u}}{\sqrt{\pi u}} e^{-\delta^2|x|^2/4u} du.$$

This formula, which is proved using basic complex analysis, is shown later on in this chapter. Applying Fubini's theorem, this means that

$$\int f(x) e^{-\delta|x|} = \int_0^\infty \frac{e^{-u}}{\sqrt{\pi u}} \int f(x) e^{-\delta^2|x|^2/4u} dx du.$$

For any fixed $t > 0$, we certainly have

$$\lim_{\delta \rightarrow 0} \int_t^\infty \frac{e^{-u}}{\sqrt{\pi u}} \int f(x) e^{-\delta^2|x|^2/4u} dx du = A \int_t^\infty \frac{e^{-u}}{\sqrt{\pi u}}$$

And this is equal to $A(1 + o(1))$ as $t \rightarrow 0$. And now we calculate

$$\int_0^t \frac{e^{-u}}{\sqrt{\pi u}} \int f(x) e^{-\delta^2|x|^2/4u} du \leq \left\| \frac{e^{-u}}{\sqrt{\pi u}} \right\|_{L^1[0,t]} \left\| \int f(x) e^{-\delta^2|x|^2/4u} \right\|_{L^\infty[0,t]}$$

The left norm tends to zero as $t \rightarrow 0$. And as $u \downarrow 0$, the dominated convergence theorem implies that

$$\int f(x) e^{-\delta|x|^2/4u} \rightarrow 0.$$

This completes the proof. \square

For any family of functions Φ_δ , we can consider the ‘ Φ sums’

$$\int f(x) \Phi_\delta(x) d\xi$$

and the corresponding Fourier transform operators

$$S_\delta(f, \Phi)(x) = \int \hat{f}(\xi) e(\xi \cdot x) \Phi_\delta(\xi) d\xi.$$

We say f is Φ summable to a value if

$$\int f(x) \Phi_\delta(x) d\xi$$

converges. In all the examples we will consider, we construct Φ sums by fixing a function $\Phi \in C_0(\mathbf{R}^n)$ with $\Phi(0) = 1$, and defining $\Phi_\delta(x) = \Phi(\delta x)$. When this is the case $f(x) \Phi_\delta(x)$ converges to $f(x)$ pointwise for each x as $\delta \rightarrow 0$. Thus if $f \in L^1(\mathbf{R}^d)$, the dominated convergence theorem implies that f is Φ summable to its usual integral. We now use these summability kernels to understand the Fourier summation formula.

Theorem 3.8 (The Multiplication Formula). *If $f, g \in L^1(\mathbf{R}^n)$,*

$$\int f(x) \hat{g}(x) dx = \int \hat{f}(\xi) g(\xi) d\xi.$$

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

$$\int f(x) \hat{g}(x) dx = \int \int f(x) g(\xi) e(-\xi \cdot x) dx d\xi = \int g(\xi) \hat{f}(\xi) d\xi. \quad \square$$

If Φ is integrable, then the multiplication formula shows

$$\begin{aligned} S_\delta(f, \Phi) &= \int \hat{f}(\xi) e(\xi \cdot x) \Phi(\delta \xi) d\xi \\ &= \int f(x) (M_x(\delta_\delta \Phi))^\wedge(x) dx = \delta^{-n} \int f(x) \cdot \hat{\Phi}\left(\frac{x-y}{\delta}\right) dx. \end{aligned}$$

Thus if we define $K_\delta^\Phi(x) = \delta^{-n} \hat{\Phi}(-x/\delta)$, then $S_\delta(f, \Phi) = K_\delta^\Phi * f$. Thus we have expressed the Fourier summation operators as convolution operations.

We now recall some notions of convolution kernels that help us approximate functions. Recall that if a family of kernels $\{K_\delta\}$ satisfies

- For any $\delta > 0$,

$$\int K_\delta(\xi) d\xi = 1.$$

- The values $\{\|K_\delta\|_{L^1(\mathbf{R}^n)}\}$ are uniformly bounded in δ .

- For any $\varepsilon > 0$,

$$\lim_{\delta \rightarrow 0} \int_{|\xi| \geq \varepsilon} |K_\delta(\xi)| d\xi \rightarrow 0.$$

then the family forms a *good kernel*. If this is the case, then $f * K_\delta$ converges to f in the L^p norms if $f \in L^p(\mathbf{R}^n)$, and converges to f uniformly if f is continuous and bounded. If we have the stronger conditions that

- For any $\delta > 0$,

$$\int K_\delta(\xi) d\xi = 1.$$

- $\|K_\delta\|_{L^\infty(\mathbf{R}^d)} \lesssim 1/\delta^d$.

- For any $\delta > 0$ and $\xi \in \mathbf{R}^d$,

$$|K_\delta(\xi)| \lesssim \frac{\delta}{|x|^{d+1}}.$$

then the family $\{K_\delta\}$ is an approximation to the identity, and so $(K_\delta * f)(x)$ converges to $f(x)$ for any x in the Lebesgue set of f .

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

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

Using contour integration, we now show

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

Since this functions is compactly supported, with total mass one, it is easy to see the corresponding Kernel K_δ^F are an approximation to the identity. Thus $\sigma_\delta f$ converges to f in all the manners described above.

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

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

Now we have

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

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

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

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

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

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

$$\begin{aligned} \hat{F}(\xi) &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{(\sin x)^2}{x^2} e^{-2i\xi x} dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{(\sin(x-iy))^2}{(x-iy)^2} e^{-2i\xi x - 2\xi y} dx \\ &= \frac{1}{4\pi} \int_{-\infty}^{\infty} f_\xi(x-iy) + g_\xi(x-iy) dx. \end{aligned}$$

If $\xi \geq 1$, 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 $2\pi n$ travelling anticlockwise along the lower semicircle with vertices $-n - iy$ and $n - iy$, then because $|z| \geq n$ on γ ,

$$\begin{aligned} \int_{-n}^n f_\xi(x - iy) + g_\xi(x - iy) dx &= \int_\gamma f_\xi(z) + g_\xi(z) dz \\ &= \text{length}(\gamma) \|f_\xi + g_\xi\|_{L^\infty(\gamma)} \\ &= (2\pi n) O_\xi(1/n^2) = O_\xi(1/n), \end{aligned}$$

and so we conclude that

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

This means $\hat{F}(\xi) = 0$. If $0 \leq \xi \leq 1$, 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 $\text{Im}(z) \geq -y$,

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

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 $|z| \geq n - y$ 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,

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

Taking $n \rightarrow \infty$, we conclude

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

and putting this all together, we conclude that $\hat{F}(\xi) = 1 - \xi$.

Example. In the next paragraph, we calculate that if $\Phi(x) = e^{-\pi|x|^2}$, then $\hat{\Phi} = \Phi$. Thus if we define the Weierstrass kernel by

$$W_{\delta}(\xi) = \delta^{-d} e^{-\pi|x|^2/\delta^2},$$

then $G_{\delta}(f) = W_{\delta} * f$. Since the family $\{W_{\delta}\}$ is an approximation to the identity, this shows $G_{\delta}(f)$ converges to f in all the appropriate senses.

Since Φ 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 $\hat{\Phi}$ is differentiable, and

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

The uniqueness theorem for ordinary differential equations says that since

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

Thus we must have $\hat{\Phi} = \Phi$.

Example. The Fourier transform of the function $e^{-|x|}$ is the Poisson kernel

$$P(\xi) = \frac{\Gamma((d+1)/2)}{(\pi(1+|\xi|^2))^{(d+1)/2}}$$

Later on we show the corresponding scaled kernel $\{P_{\delta}\}$ is an approximation to the identity, and thus $A_{\delta}f = P_{\delta} * f$ converges to f in all appropriate senses.

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

$$\begin{aligned} \hat{A}(\xi) &= \int e^{-2\pi|x|} e^{-2\pi i \xi \cdot x} dx = \int \int_0^\infty \frac{e^{-u}}{\sqrt{\pi u}} e^{-|\pi x|^2/u} e^{-2\pi i \xi \cdot x} du dx \\ &= \int_0^\infty \frac{e^{-u}}{\sqrt{\pi u}} \int e^{-|\pi x|^2/u} e^{-2\pi i \xi \cdot x} dx du = \int_0^\infty \frac{e^{-u}}{\sqrt{\pi u}} (\delta_{\sqrt{\pi/u}} \Phi)^\wedge(\xi) du \\ &= \frac{1}{\pi^{(n+1)/2}} \int_0^\infty e^{-u} u^{(n-1)/2} e^{-u|\xi|^2} du \end{aligned}$$

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

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

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

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 $\{P_\delta\}$ 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.

Example. We note that

$$\int_{-R}^R e(-\xi x) dx = \frac{e(-\xi R) - e(\xi R)}{-2\pi i \xi} = \frac{\sin(2\pi \xi R)}{\pi \xi}.$$

so the Fourier transform of $\chi_{[-R,R]}$ is the Dirichlet kernel

$$D_R(\xi) = \frac{\sin(2\pi \xi R)}{\pi \xi}$$

We note that $D_R \notin L^1(\mathbf{T})$. Thus D_R is not a good kernel, which makes the convergence rates of $S_R f$ more subtle.

Thus we now know there are a large examples of functions $\Phi \in C_0(\mathbf{R}^d)$ with $\Phi(0) = 1$, and such that for any x in the Lebesgue set of f ,

$$f(x) = \lim_{\delta \rightarrow 0} \int \hat{f}(\xi) e(\xi \cdot x) \Phi(\delta \xi) d\xi.$$

If \hat{f} is integrable, then the bound $|\hat{f}(\xi) e(\xi \cdot x) \Phi(\delta \xi)| \leq \|\Phi\|_\infty |\hat{f}(\xi)|$ implies that we can use the dominated convergence theorem to conclude that for any point x in the Lebesgue set of f ,

$$f(x) = \lim_{\delta \rightarrow 0} \int \hat{f}(\xi) e(\xi \cdot x) \Phi(\delta \xi) d\xi = \int \hat{f}(\xi) e(\xi \cdot x) d\xi$$

Thus the inversion theorem holds pointwise almost everywhere.

Theorem 3.9. *If f and \hat{f} , then for any x in the Lebesgue set of f ,*

$$f(x) = \int \hat{f}(\xi) e(\xi \cdot x) d\xi.$$

We define, for any integrable $f : \mathbf{R}^n \rightarrow \mathbf{R}$, the *inverse* Fourier transform

$$\check{f}(x) = \int f(\xi) e(\xi \cdot x) d\xi$$

The inverse transform is also denoted by $\mathcal{F}^{-1}(f)$. The last theorem says that \mathcal{F}^{-1} really is the inverse operator to the operator \mathcal{F} , at least on the set of functions f where \hat{f} is integrable. In particular, this is true if f has strong derivatives in the L^1 norm for any multi-index $|\alpha| \leq n+1$, and so the Fourier inversion formula holds for sufficiently smooth functions.

Corollary 3.10. *If $f \in C(\mathbf{R})$ is integrable and $\hat{f} \in L^1(\mathbf{R})$, $S_R f \rightarrow f$ uniformly.*

Proof. Since $f \in C(\mathbf{R})$ The dominated convergence theorem implies that for each $x \in \mathbf{R}$,

$$f(x) = \int_{\mathbf{R}^n} \hat{f}(\xi) e(\xi \cdot x) = \lim_{R \rightarrow \infty} \int_{-R}^R \hat{f}(\xi) e(\xi \cdot x) = \lim_{R \rightarrow \infty} (S_R f)(x).$$

And

$$\int_{|x| \geq R} \hat{f}(\xi) e(\xi \cdot x) \leq \|\hat{f}\|_{L^1(\mathbf{R})}.$$

so the pointwise convergence is uniform. □

Remark. This theorem also generalizes to \mathbf{R}^n . Here, the operators S_R are no longer canonically defined, but if we consider any increasing nested family of sets B_R with $\lim B_R = \mathbf{R}^n$, then the corresponding operators

$$S_R f = \int_{B_R} \hat{f}(\xi) e(\xi \cdot x)$$

also converge uniformly to f .

Corollary 3.11. *The map $\mathcal{F} : L^1(\mathbf{R}^d) \rightarrow C_0(\mathbf{R}^d)$ is injective.*

Proof. If $\hat{f} = 0$, then \hat{f} is certainly integrable. But this means that the Fourier inversion theorem can apply, giving that for almost every point x ,

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi) e(\xi \cdot x) d\xi = 0.$$

Thus $f = 0$ almost everywhere. \square

This corollary is underestimated in utility. Even if the Fourier inversion theorem doesn't hold, we can still view the Fourier transform as another way to represent a function, since the Fourier transform does not lose any information. For instance, it can be used very easily to verify identities involving convolutions.

Corollary 3.12. *For any δ_1, δ_2 ,*

$$W_{\delta_1+\delta_2} = W_{\delta_1} * W_{\delta_2} \quad \text{and} \quad P_{\delta_1+\delta_2} = P_{\delta_1} * P_{\delta_2}.$$

Proof. We recall that

$$W_{\delta_1+\delta_2} = \mathcal{F}(e^{-(\delta_1+\delta_2)|x|^2}).$$

But $e^{-(\delta_1+\delta_2)|x|^2} = e^{-\delta_1|x|^2} e^{-\delta_2|x|^2}$ breaks into a product, which allows us to calculate

$$\mathcal{F}(e^{-\pi\delta_1|x|^2} e^{-\pi\delta_2|x|^2}) = \mathcal{F}(e^{-\pi\delta_1|x|^2}) * \mathcal{F}(e^{-\pi\delta_2|x|^2}) = W_{\delta_1} * W_{\delta_2}.$$

Thus $W_{\delta_1} * W_{\delta_2} = W_{\delta_1+\delta_2}$. Similarly, $P_{\delta_1+\delta_2}$ is the Fourier transform of $e^{-(\delta_1+\delta_2)|x|}$, which breaks into a product, whose individual Fourier transforms are P_{δ_1} and P_{δ_2} . \square

3.2 The L^2 Theory

One integral component of Fourier series on $L^2(\mathbf{T})$ is Plancherel's equality

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

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

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

However, on the non-compact Euclidean space, a general element of $L^2(\mathbf{R}^n)$ is not necessarily integrable, so we cannot take its 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 3.13. *If $f \in L^1(\mathbf{R}^n) \cap L^2(\mathbf{R}^n)$, then $\|\hat{f}\|_2 = \|f\|_2$.*

Proof. The theorem is an easy consequence of the multiplication formula, since

$$|\hat{f}(\xi)|^2 = \hat{f}(\xi) \overline{\hat{f}(\xi)},$$

and

$$\overline{\hat{f}(\xi)} = \widehat{\overline{f(-\xi)}} = \widehat{\overline{f}(\xi)}.$$

This implies

$$\int |\hat{f}(\xi)|^2 d\xi = \int \hat{f}(\xi) \overline{\hat{f}(\xi)} d\xi = \int f(x) \overline{\widehat{\overline{f}(\xi)}} dx = \int |f(x)|^2 dx. \quad \square$$

A simple interpolation argument leads to the following corollary, which is a variant of the Hausdorff-Young inequality for functions on \mathbf{R}^n .

Corollary 3.14. *If $f \in L^1(\mathbf{R}^n) \cap L^p(\mathbf{R}^n)$ for $1 \leq p \leq 2$, then*

$$\|\hat{f}\|_{L^q(\mathbf{R}^n)} \leq \|f\|_{L^p(\mathbf{R}^n)}.$$

where $2 \leq q \leq \infty$ is the conjugate of p .

Though the integral formula of an element of $L^2(\mathbf{R}^n)$ does not make sense, the bounds above provide a canonical way to define the Fourier transform of an element of $L^p(\mathbf{R}^n)$, for $1 \leq p \leq 2$. The space $L^1(\mathbf{R}^n) \cap L^p(\mathbf{R}^n)$ is a dense subset of $L^p(\mathbf{R}^n)$, so we can use the Hahn-Banach theorem to define the Fourier transform $\mathcal{F} : L^p(\mathbf{R}^n) \rightarrow L^q(\mathbf{R}^n)$ as the *unique* bounded operator agreeing with the integral formula on the common domain. The extended Fourier transform on $L^2(\mathbf{R}^n)$ is still unitary, because the multiplication formula extends to $L^2(\mathbf{R}^n)$, so that

$$(\mathcal{F}(f), g) = \int \hat{f}(\xi) \overline{\hat{g}(\xi)} d\xi = \int f(x) \overline{\widehat{g}(-\xi)} dx = (f, \mathcal{F}^{-1}(g)).$$

Thus the adjoint of \mathcal{F} is \mathcal{F}^{-1} , which means exactly that \mathcal{F} is unitary.

Remark. Later on, we will justify the definition of the Fourier transform on an element of $L^p(\mathbf{R}^n)$, for $2 \leq p \leq \infty$, using the theory of distributions.

3.3 The Hausdorff-Young Inequality

For functions on \mathbf{T} , it is unclear how to provide examples which show why the Hausdorff-Young inequality cannot be extended to give results for $p > 2$. Over \mathbf{R} , we can provide examples which explicitly indicate the tightness of the appropriate constants.

Example. Given an integrable function f , let $f_r(x) = f(rx)$. Then we find $\widehat{f}_r(\xi) = r^{-n} \widehat{f}(\xi/r)$, and so

$$\|f_r\|_{L^p(\mathbf{R}^n)} = r^{-n/p} \|f\|_{L^p(\mathbf{R}^n)} \quad \text{and} \quad \|\widehat{f}_r\|_{L^q(\mathbf{R}^n)} = r^{n/q-n} \|\widehat{f}\|_{L^q(\mathbf{R}^n)}.$$

In order for a bound to hold in terms of p and q uniformly for all values of r , we need $r^{-n/p} = r^{n/q-n}$, which means $1/q + 1/p = 1$, so p and q must be conjugates of one another.

Example. Consider the family of functions $f_s(x) = s^{-n/2} e^{-\pi|x|^2/s}$, where $s = 1 + it$ for some $t \in \mathbf{R}$. One can easily calculate that $\widehat{f}_s(\xi) = e^{-\pi s|\xi|^2}$. We calculate

$$\|f_s\|_{L^p(\mathbf{R}^n)} = |s|^{-n/2} \left(\int e^{-(p/|s|^2)\pi|x|^2} dx \right)^{1/p} = |s|^{n/p-n/2} p^{-n/p}$$

whereas $\|\widehat{f}_s\|_q = q^{-n/2}$. Thus to be able compare the two quantities as $t \rightarrow \infty$, we need $n/p - n/2 \leq 0$, so $p \leq 2$. As $t \rightarrow \infty$, $|f_s(x)| \sim t^{-n/2} e^{-\pi|x/t|^2}$, so the t gives us a decay in f_s . However, when we take the Fourier transform the t only corresponds to oscillatory terms. Thus we need $p \leq 2$ so that the decay in t isn't too important in relation to the overall width of the function.

The Hausdorff-Young inequality shows that the Fourier transforms narrowly supported functions into a function with small magnitude. But the example above shows that the Fourier transform is not so good at transforming functions with small magnitude into functions which are narrowly supported, because the Fourier transform can absorb the small magnitude into an oscillatory property not reflected in the norms. Some kind of way of measuring oscillation needs to be considered to get a tighter control on the function. Of course, in hindsight, we should have never expected too much control of the Fourier transform in terms of the L^p norms, since the Fourier transform measures the oscillatory nature of the input

function, and oscillatory properties of a function in phase space are not very well reflected in the L^p norms, except when applying certain orthogonality properties with an L^2 norm, or destroying the oscillation with an L^∞ norm.

3.4 The Poisson Summation Formula

3.5 The Gibbs Phenomenon

3.6 Sums of Random Variables

We now switch to an application of harmonic analysis to studying sums of random variables probability theory. If X is a random vector, its probabilistic information is given by its distribution on \mathbf{R}^n , which can be seen as a measure \mathbf{P}_X on \mathbf{R}^n , with $\mathbf{P}_X(E) = \mathbf{P}(X \in E)$. Given two independent random vectors X and Y , \mathbf{P}_{X+Y} is the convolution $\mathbf{P}_X * \mathbf{P}_Y$ between the measures \mathbf{P}_X and \mathbf{P}_Y , in the sense that

$$\mathbf{P}_{X+Y}(E) = \int \chi_E(x+y) d\mathbf{P}_X(x) d\mathbf{P}_Y(y)$$

If $d\mathbf{P}_X = f_X \cdot dx$ and $d\mathbf{P}_Y = f_Y \cdot dx$, then $d(\mathbf{P}_X * \mathbf{P}_Y) = (f_X * f_Y) \cdot dx$ is just the normal convolution of functions. This is why harmonic analysis becomes so useful when analyzing sums of independent random variables.

It is useful to express the Fourier transform in a probabilistic language. Given a random variable X ,

$$\widehat{\mathbf{P}}_X(\xi) = \int e^{i\xi \cdot x} d\mathbf{P}_X(x)$$

Thus the natural Fourier transform of a random vector X is the **characteristic function** $\varphi_X(\xi) = \mathbf{E}(e^{i\xi \cdot X})$. It is a continuous function for any random variable X . We can also express the properties of the Fourier transform in a probabilistic language.

Lemma 3.15. *Let X and Y be independent random variables. Then*

- $\varphi_X(0) = 1$, and $|\varphi_X(\xi)| \leq 1$ for all ξ .

- (Symmetry) $\varphi_X(\xi) = \overline{\varphi_X(-\xi)}$.
- (Convolution) $\varphi_{X+Y} = \varphi_X \varphi_Y$.
- (Translation and Dilation) $\varphi_{X+a}(\xi) = e^{ia \cdot \xi} \varphi_X(\xi)$, and $\varphi_{\lambda X}(\xi) = \varphi_X(\lambda \xi)$.
- (Rotations) If $R \in O(n)$ is a rotation, then $\varphi_{R(X)}(\xi) = \varphi_X(R(X))$.

Using the Fourier inversion formula, if φ_X is integrable, then X is a continuous random variable, with density

$$f(x) = \int e^{-i\xi x} \varphi_X(\xi) d\xi$$

In particular, if $\varphi_X = \varphi_Y$, then X and Y are identically distributed. This already gives interesting results.

Theorem 3.16. *If X and Y are independant normal distributions, then $aX + bY$ is normally distributed.*

Proof. Since $\varphi_{aX+bY}(\xi) = \varphi_X(a\xi)\varphi_Y(b\xi)$, it suffices to show that the product of two such characteristic functions is the characteristic function of a normal distribution. If X has mean μ and covariance matrix Σ , then $X \cdot \xi$ has mean $\mu \cdot \xi$ and variance $\xi^T \Sigma \xi$, and one calculates that $\mathbf{E}[e^{i\xi \cdot X}] = e^{-i\mu \cdot \xi - \xi^T \Sigma \xi / 2}$ using similar techniques to the Fourier transform of a Gaussian. One verifies that the class of functions of the form $e^{-i\mu \cdot \xi - \xi^T \Sigma \xi / 2}$ is certainly closed under multiplication and scaling, which completes the proof. \square

Now we can prove the celebrated central limit theorem. Note that if

Theorem 3.17. *Let X_1, \dots, X_N be independant and identically distributed with mean zero and variance σ^2 . If $S_N = X_1 + \dots + X_N$, then*

$$\mathbf{P}(S_N \leq \sigma \sqrt{N} t) \rightarrow \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-y^2/2} dy$$

Proof. We calculate that

$$\varphi_{S_N/\sigma\sqrt{N}}(\xi) = \varphi_X(\xi/\sigma\sqrt{N})^N$$

Define $R_n(x) = e^{ix} - 1 - (ix) - (ix)^2/2 - \cdots - (ix)^n/n!$. Then because of oscillation and the fundamental theorem of calculus,

$$|R_0(x)| = \left| i \int_0^x e^{iy} dy \right| \leq \min(2, |x|)$$

Next, since $R'_{n+1}(x) = iR_n$,

$$R_{n+1}(x) = i \int_0^x R_n(y) dy$$

This gives that $|R_n(x)| \leq \min(2|x|^n/n!, |x|^{n+1}/(n+1)!)$. In particular, we conclude

$$|\varphi_X(\xi) - 1 - \sigma^2 \xi^2/2| = |\mathbf{E}(R_2(\xi X))| \leq \mathbf{E}|R_2(\xi X)| \leq |\xi|^2 \mathbf{E}(\min(|X|^2, |\xi X|^3/6))$$

By the dominated convergence theorem, as $\xi \rightarrow 0$, $\varphi_X(\xi) = 1 - \xi^2 \sigma^2/2 + o(\xi^2)$. But this means that

$$\varphi_{S_N/\sigma\sqrt{N}}(\xi) = (1 - \xi^2/2N + o(\xi^2/\sigma^2 N))^N = \exp(-\xi^2/2)$$

This implies the random variables converge weakly to a normal distribution. \square

3.7 Transforms of Holomorphic Functions

An interesting thing about the Fourier transform on the real line is that we can apply both real-variable techniques and complex analytic techniques to the study. Just as smoothness of a function corresponded to polynomial decay in its Fourier transform, we will find that analyticity corresponds to an exponential decay. We let $S_a = \{z \in \mathbf{C} : |\operatorname{Im}(z)| < a\}$ denote the horizontal strip.

Theorem 3.18. *Let f be holomorphic on S_a , integrable on each horizontal line contained in the strip, and such that $f(z) \rightarrow 0$ as $|\operatorname{Re}(z)| \rightarrow \infty$. Then we find $|\hat{f}(\xi)| \lesssim_b e^{-2\pi b|\xi|}$ for any $b < a$.*

Proof. For any $b < a$, R , and $\xi > 0$, consider the contour γ_R obtained from the rectangle $-R, R, -R - ib$, and $R - ib$. As $R \rightarrow \infty$, the integral along the

vertical lines of the rectangle tends to zero as $R \rightarrow \infty$, so we conclude that

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx &= \int_{-\infty}^{\infty} f(x - ib) e^{-2\pi i (x - ib) \xi} dx \\ &= e^{-2\pi i b \xi} \int_{-\infty}^{\infty} f(x - ib) e^{-2\pi i \xi x} dx = e^{-2\pi i b \xi} \hat{f}_b(\xi) \end{aligned}$$

where $f_b(x) = f(x - ib)$. Thus it suffices to show $|\hat{f}_b(\xi) - \hat{f}(\xi)| \lesssim_b 1$. But this follows because we certainly have $\|f_b - f\|_{\infty} < \infty$ since both functions are bounded, which completes the proof in this case. A similar estimate for $\xi < 0$ also gives the general result. \square

It follows that \hat{f} has exponential decay if f satisfies the hypothesis of the theorem. Thus we can always apply the inverse Fourier transform to conclude

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

In this case, we can actually *prove* the equation using complex analysis.

Theorem 3.19. *A*

Proof. As in the last theorem, the sign of ξ matters. We write

$$\int_{-\infty}^{\infty} \hat{f}(\xi) e^{i\xi x} d\xi$$

\square

3.8 Characteristic Functions

Chapter 4

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 N th 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 $f : G \rightarrow H$ 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 $(x, y) \mapsto f(x)g(y)$.

4.1 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 $f : [N] \rightarrow \mathbf{C}$, 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 $f : [N] \rightarrow \mathbf{C}$ has an expansion

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

where $\hat{f}(m)$ 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 $f : [N] \rightarrow \mathbf{C}$, 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

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

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 $f : [N] \rightarrow \mathbf{C}$, we have, if we set $\hat{f}(m) = \langle f, \phi_m \rangle$, then

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

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

4.2 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 4.1. *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

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

Implies 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. \square

Because elements of $\Gamma(G)$ are orthonormal, they are linearly independent over the space of functions on G , and we obtain a bound $|\Gamma(G)| \leq |G|$. 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 4.2. *Let $\{T_1, \dots, T_n\}$ be a family of commuting unitary matrices. Then there is a basis $v_1, \dots, v_m \in \mathbb{C}^m$ 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

$$\mathbb{C}^m = V_{\lambda_1} \oplus \dots \oplus V_{\lambda_l}$$

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 hypothesis to diagonalize the T_1, \dots, T_{k-1} . Putting this together, we simultaneously diagonalize T_1, \dots, T_k . \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 basis consisting of the $\chi_a(b) = N[a = b]$, for $a \in G$. In this basis, we compute $T_a \chi_b = \chi_{ba^{-1}}$, 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 $e_1, \dots, e_n \in V$ 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 4.3. *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 $f : G \rightarrow \mathbf{C}$,*

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

In this context, we also have Parseval's theorem

$$\|f(a)\|^2 = \sum_{e \in \hat{G}} |\hat{f}(e)|^2 \quad \langle f, g \rangle = \sum_{e \in \hat{G}} \hat{f}(e) \overline{\hat{g}(e)}$$

4.3 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 $e \in \Gamma(G)$,

$$\begin{aligned}\widehat{f * g}(e) &= \frac{1}{|G|} \sum_{a \in G} (f * g)(a) \overline{e(a)} \\ &= \frac{1}{|G|^2} \sum_{a, b \in G} f(ab) g(b^{-1}) \overline{e(a)}\end{aligned}$$

The bijection $a \mapsto ab^{-1}$ shows that

$$\begin{aligned}\widehat{f * g}(e) &= \frac{1}{|G|^2} \sum_{a, b} f(a) g(b^{-1}) \overline{e(a) e(b^{-1})} \\ &= \frac{1}{|G|} \left(\sum_a f(a) \overline{e(a)} \right) \frac{1}{|G|} \left(\sum_b g(b) \overline{e(b)} \right) \\ &= \widehat{f}(e) \widehat{g}(e)\end{aligned}$$

In the finite case we do not need approximations to the identity, for we have an identity for convolution. Define $D : G \rightarrow \mathbb{C}$ 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 $f : G \rightarrow \mathbb{C}$, 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

$$\widehat{D}(e) = \frac{1}{|G|} \sum_{a \in G} D(a) \overline{e(a)} = \overline{e(1)} = 1$$

D is essentially the finite dimensional version of the Dirac delta function, since it has unit mass, and acts as the identity in convolution.

4.4 The Fast Fourier Transform

The main use of the fourier series on μ_n in applied mathematics 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 $\hat{f}_n \rightarrow \hat{f}$ 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 $f : \mathbf{Z}_n \rightarrow \mathbf{C}$ 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. Before this process was discovered, calculation of Fourier transforms was seen as a computation to avoid wherever possible.

To see this, consider a particular division in the group \mathbf{Z}_{2n} . Given $f : \mathbf{Z}_{2n} \rightarrow \mathbf{C}$, define two functions $g, h : \mathbf{Z}_n \rightarrow \mathbf{C}$, defined by $g(k) = f(2k)$, and $h(k) = f(2k+1)$. Then g and h encode all the information in f , and if $v = e(\pi/n)$ is the canonical generator of \mathbf{Z}_{2n} , we have

$$\hat{f}(m) = \frac{\hat{g}(m) + \hat{h}(m)v^m}{2}$$

Because

$$\begin{aligned} \frac{1}{2n} \sum_{k=1}^n \left(g(k)\omega^{-km} + h(k)\omega^{-km}v^m \right) &= \frac{1}{2n} \sum_{k=1}^n f(2k)v^{-2km} + f(2k+1)v^{-(2k+1)m} \\ &= \frac{1}{2n} \sum_{k=1}^{2n} f(k)v^{-km} \end{aligned}$$

This is essentially a discrete analogue of the Poisson 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 $H(2m) = 2H(m) + 3(2m)$. 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 ν with ν^{-1}).

4.5 Dirichlet's Theorem

We now apply the theory of Fourier series on finite abelian groups to prove Dirichlet's theorem.

Theorem 4.4. *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 requires 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

$$\sum_{n=1}^{\infty} \frac{1}{n^s} \leq 1 + \int_1^{\infty} \frac{1}{t^s} = 1 + \lim_{n \rightarrow \infty} \frac{1}{s-1} [1 - 1/n^{s-1}] = 1 + \frac{1}{s-1}$$

The series converges uniformly on $[1 + \varepsilon, N]$ for any $\varepsilon > 0$, so ζ is continuous on $(1, \infty)$. As $t \rightarrow 1$, $\zeta(t) \rightarrow \infty$, because $n^s \rightarrow n$ 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 $1 \leq n \leq M$, then

$$\sum_{n=1}^{\infty} \frac{1}{n^s} \geq \sum_{n=1}^M \frac{1}{n^s} = \sum_{n=1}^M \frac{1}{n} \frac{n}{n^s} \geq \frac{1}{2} \sum_{n=1}^M \frac{1}{n}$$

Letting $M \rightarrow \infty$, we obtain that $\sum_{n=1}^{\infty} \frac{1}{n^s} \rightarrow \infty$ as $s \rightarrow 1$.

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 4.5. *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, enumerating the primes to be p_1, p_2, \dots , we find

$$\prod_{p \leq n} (1 + 1/p^s + 1/p^{2s} + \dots) = \sum_{n_1, n_2, \dots = 0}^{\infty} \frac{1}{p_1^{n_1}}$$

□

Chapter 5

Applications

5.1 The Wirtinger Inequality on an Interval

Theorem 5.1. Given $f \in C^1[-\pi, \pi]$ with $\int_{-\pi}^{\pi} f(t) dt = 0$,

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

Proof. Consider the fourier series

$$f(t) \sim \sum a_n e_n(t) \quad f'(t) \sim \sum i n a_n e_n(t)$$

Then $a_0 = 0$, and so

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

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

$$f(t) = A e_n(t) + \bar{A} e_n(-t) = B \cos(t) + C \sin(t)$$

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

Corollary 5.2. Given $f \in C^1[a, b]$ with $\int_a^b f(t) dt = 0$,

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

5.2 Energy Preservation in the String equation

Solutions to the string equation are

If $u(t, x)$

5.3 Harmonic Functions

The study of a function f defined on the real line can often be understood by extending its 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 its 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 5.3. *If $f \in L^p(\mathbf{R}^n)$, for $1 \leq p \leq \infty$, 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, $u(x, y) \rightarrow f(x)$ for almost every x , and $u(\cdot, y)$ converges to f in L^p as $y \rightarrow 0$, with $\|u(\cdot, y)\|_{L^p(\mathbf{R}^n)} \leq \|f\|_{L^p(\mathbf{R}^n)}$. If, instead, f is a continuous and bounded function, then $u(\cdot, y)$ converges to f locally uniformly as $y \rightarrow 0$.

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

$$u_{xx}(x, y) = (f * P_y'')(x) \quad u_{yy} = (f *)$$

□

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. Rather than focusing on the qualities of particular functions, we employ key tools from functional analysis. But different from basic functional analysis, where soft analysis is key, in harmonic analysis we are most interested in hard estimates. Thus we are interested in quantitative estimates on the properties of operators on function spaces. Key to this is the employment of certain operator norms which quantify a control on the qualitative properties of functions, be they their height, their width, their derivative, and many other more quantifiable properties.

Chapter 6

Monotone Rearrangement Invariant Norms

Often in analysis, a problem occurs where one must understand how a quantitative control on the properties of a function, e.g. its oscillation or growth at ∞ , implies a control on the properties of a related object, i.e. the decay of its derivative. The most analytically sound way to understand these controls is with respect to a *norm*, a quantitative measure of some property of a function which behaves well under the operations of addition and multiplication. Here we discuss the *rearrangement invariant* norms, those norms such that if $\Phi : Y \rightarrow X$ is a measure preserving bijection of the measure spaces X and Y , then f and $f \circ \Phi$ should have the same norm. Furthermore, we study *monotone* norms, so if $f \leq g$, the norm of f should be smaller than the norm of g .

In an abstract measure space X , there are two very useful classes of functions:

- The *indicator functions* $\mathbf{I}_E(x) = \mathbf{I}(x \in E)$, for a measurable set E .
- The *step functions* $f = \sum_{i=1}^n a_i \mathbf{I}_{E_i}$, for disjoint sets E_i .

With respect to any of the norms we encounter in this chapter, the vector space of step functions will be dense, and we will often use these functions to test the behaviour of certain operators.

Really, the only properties a monotone, translation-invariant norm can reflect are the functions typical amplitude, or *height*, and its *width*, or the measure of the set where the majority of the amplitude is distributed.

These two properties are not precisely defined, and so the rearrangement invariant norms help quantify them in various ways. In a particular problem, one picks the norm best emphasizing the features useful in a problem, often chosen by optimizing over a large family of norms.

6.1 The L^p norms

The most important family of measurable function are those which are integrable, i.e. such that

$$\int |f(x)| dx < \infty.$$

This condition already places a control over the height and width of f ; the function cannot be too large, except on a small set, and also must have the large majority of it's mass on a small set. The condition

$$\int |f(x)|^2 dx < \infty$$

also places a certain amount of control over the height and width of a function f , but now the height is controlled more tightly than the width, since the portions of the domain where f is large are magnified in $|f(x)|^2$, and the portions of the domain where f is small are made more negligible. More generally, for any $p > 0$, the condition

$$\int |f(x)|^p dx < \infty \tag{6.1}$$

controls the height and width of a function, and we define the space $L^p(X)$ as the family of measurable functions $f : X \rightarrow \mathbb{C}$ satisfying (6.1). The norm

$$\|f\|_p = \left(\int |f|^p \right)^{1/p}$$

provides a quantity measuring f 's width and height. The importance of taking the p 'th root is useful for the stability of the norm, so that $\|\lambda f\|_p = |\lambda| \|f\|_p$ for any constant λ .

Example. If $f(x) = |x|^{-\alpha}$ for $x \in \mathbf{R}^d$, then integration by radial coordinates shows that

$$\int_{\varepsilon \leq |x| \leq M} |f(x)|^p dx \sim_d \int_{\varepsilon}^M r^{d-1-p\alpha} dr = \frac{M^{d-p\alpha} - \varepsilon^{d-p\alpha}}{d-p\alpha}.$$

This quantity remains finite as $\varepsilon \rightarrow 0$ if and only if $d > p\alpha$, and finite as we let $M \rightarrow \infty$ if and only if $d < p\alpha$. Thus if $p < d/\alpha$, f is locally in L^p , or in $L^p(B)$ for every bounded $B \subset \mathbf{R}^d$. Conversely, if $p > d/\alpha$, then for every domain B with $d(B, 0) > 0$, $f \in L^p(B)$.

The last example shows that control on the L^p norm of a function for large values of p prevents the formation of large singularities of the function, and control of the norm for small values of p ensures that functions have large decay at infinity, in a rough sense.

Example. If $f = A\chi_E$, and we set $H = |A|$ and $W = |E|$, then $\|f\|_p = W^{1/p}H$. As $p \rightarrow \infty$, the value of $\|f\|_p$ depends more and more on H , and less on W , and in fact $\lim_{p \rightarrow \infty} \|f\|_p = H$. If $f = \sum A_n \chi_{E_n}$, and $|A_m|$ is the largest constant from all other values A_n , then as p becomes large, $|A_m|^p$ overwhelms all other terms. We calculate that as $p \rightarrow \infty$,

$$\|f\|_p = \left(\sum |E_n| |A_n|^p \right)^{1/p} = |A_m|^p (|E_m| + o(1))^{1/p} = |A_m| (1 + o(1)).$$

This implies $\|f\|_p \rightarrow |A_m|$ as $p \rightarrow \infty$. But as $p \rightarrow 0$, the power of $1/p$ works against us, so that $\|f\|_p$ rarely converges to any universal quantity. We can conclude that $\lim_{p \rightarrow 0} \|f\|_p^p = \sum |E_n|$, so the width of the function becomes more and more emphasized. In the case where $\sum |E_i| = 1$, we can use L'Hopitals rule as $p \rightarrow 0$ to conclude that

$$\lim_{p \rightarrow 0} \log \|f\|_p = \lim_{p \rightarrow 0} \frac{\log(\sum |a_i|^p |E_i|)}{p} = \sum |E_i| \log |a_i| = \int_{|f(x)| > 0} \log |f(x)| dx.$$

Taking exponentials on both sides, we conclude that

$$\lim_{p \rightarrow 0} \|f\|_p = \exp \left(\int_{|f(x)| > 0} \log |f(x)| dx \right).$$

However, the fact that $\log |f(x)|$ can take negative values makes this norm difficult to work with, unless $f(x)$ is suitably granular.

As $p \rightarrow \infty$, the width of a function is disregarded completely by the L^p norm, motivating the definition of the L^∞ norm; Given a measurable f , we define $\|f\|_\infty$ to be the smallest number such that $|f| \leq \|f\|_\infty$ almost surely. We then define $L^\infty(X)$ to be the space of measurable functions f for which $\|f\|_\infty < \infty$. We have already shown $\|f\|_p \rightarrow \|f\|_\infty$ if s is a simple function, and the density of such functions, combined with some functional analysis, shows that if $f \in L^p(X) \cap L^\infty(X)$ for any value of p , $\|f\|_q \rightarrow \|f\|_\infty$ as $q \rightarrow \infty$. Sometimes, authors also define $\|f\|_0 = |\text{supp } f| = |\{x : f(x) \neq 0\}|$, but this is really a ‘zeroeth power’ of the L^0 norm. Because of this, most of the nice norm properties of the L^p spaces are broken. For instance, $\|\cdot\|_0$ is no longer homogenous, i.e. if $\lambda \neq 0$, $\|\lambda f\|_0 = \|f\|_0$.

Example. Let $p < q$, and suppose $f \in L^p(X) \cap L^q(X)$. For any $r \in (p, q)$, the L^r norm emphasizes the height of f less than the L^q norm, and emphasizes the width of f less than the L^p norm. In particular, we find that for any $\lambda \geq 0$,

$$\begin{aligned} \|f\|_r^r &= \int_{\mathbf{R}} |f(x)|^r dx = \int_{|f(x)| \leq 1} |f(x)|^r dx + \int_{|f(x)| > 1} |f(x)|^r dx \\ &\leq \int_{|f(x)| \leq 1} |f(x)|^p dx + \int_{|f(x)| > 1} |f(x)|^q dx \\ &\leq \|f\|_p^p + \|f\|_q^q < \infty. \end{aligned}$$

In particular, this shows $f \in L^r(X)$.

Remark. The bound obtained in the last example can be improved by using scaling symmetries. For any $A > 0$,

$$\|f\|_r^r = \frac{\|Af\|_r^r}{A^r} \leq \frac{\|Af\|_p^p + \|Af\|_q^q}{A^r} \leq \frac{A^p \|f\|_p^p + A^q \|f\|_q^q}{A^r}.$$

If $1/r = \theta/p + (1-\theta)/q$, and we set $A = \|f\|_q^{q/(p-q)} / \|f\|_p^{p/(p-q)}$, then the above inequality implies $\|f\|_r \leq 2 \|f\|_p^\theta \|f\|_q^{1-\theta}$, which is a homogenous equality. The constant 2 can be removed in the equation using the *tensor power trick*. If we consider the function on $(\mathbf{R}^d)^n$ defined by $f^{\otimes n}(x_1, \dots, x_n) = f(x_1) \dots f(x_n)$, then $\|f^{\otimes n}\|_r = \|f\|_r^n$, and so

$$\|f\|_r = \|f^{\otimes n}\|_r^{1/n} \leq (2 \|f^{\otimes n}\|_p \|f^{\otimes n}\|_q)^{1/n} = 2^{1/n} \|f\|_p \|f\|_q.$$

We can then take $n \rightarrow \infty$ to conclude that $\|f\|_r \leq \|f\|_p \|f\|_q$.

The argument in the last remark is an instance of *real interpolation*; In order to conclude some fact about a function which lies ‘between’ two other functions we know how to deal with, we split the function up into two parts lying in the other spaces, deal with them separately, and then put them back together to get some equality. One can then apply various symmetry considerations (homogeneity and the tensor power trick being two examples) to eliminate extraneous constants unneeded to the problem. We now also show how to prove this inequality using convexity, which illustrates another core technique of the field. In the next theorem, $1/\infty = 0$.

Theorem 6.1 (Hölder). *If $0 < p, q \leq \infty$ and $1/p + 1/q = 1/r$, $\|fg\|_r \leq \|f\|_p \|g\|_q$.*

Proof. The case where p or q is ∞ is left as an exercise to the reader. In the other case, by moving around exponents, we may simplify to the case where $r = 1$. The theorem depends on the log convexity inequality, such that for $A, B \geq 0$ and $0 \leq \theta \leq 1$, $A^\theta B^{1-\theta} \leq \theta A + (1 - \theta)B$. But since the logarithm is concave, we calculate

$$\log(A^\theta B^{1-\theta}) = \theta \log A + (1 - \theta) \log B \leq \log(\theta A + (1 - \theta)B),$$

and we can then exponentiate. To prove Hölder’s inequality, by scaling f and g , which is fine by homogeneity, we may assume that $\|f\|_p = \|g\|_q = 1$. Then we calculate

$$\begin{aligned} \|fg\|_1 &= \int |f(x)| |g(x)| = \int |f(x)|^{p/p} |g(x)|^{q/q} \\ &\leq \int \frac{|f(x)|^p}{p} + \frac{|g(x)|^q}{q} = \frac{1}{p} + \frac{1}{q} = 1 = \|f\|_p \|g\|_q. \end{aligned}$$

If $p = \infty$, $q = 1$, then the inequality is trivial, since we have the pointwise inequality $|f(x)g(x)| \leq \|f\|_\infty |g(x)|$ almost everywhere, which we can then integrate. \square

Remark. Note that $A^\theta B^{1-\theta} \leq \theta A + (1 - \theta)B$ is an *equality* if and only if $A = B$, or $\theta \in \{0, 1\}$. In particular, following through the proof above shows that if $\|f\|_p = \|g\|_q = 1$, we must have $|f(x)|^{1/p} = |g(x)|^{1/q}$ almost everywhere. In general, this means Hölder’s inequality is sharp if and only if $|f(x)|^{1/p}$ is a constant multiple of $|g(x)|^{1/q}$.

The next inequality is known as the *triangle inequality*.

Corollary 6.2. Given f, g , and $p \geq 1$, $\|f + g\|_p \leq \|f\|_p + \|g\|_p$.

Proof. The inequality when $p = 1$ is obtained by integrating the inequality $|f(x) + g(x)| \leq |f(x)| + |g(x)|$, and the case $p = \infty$ is equally trivial. When $1 < p < \infty$, by scaling we can assume that $\|f\|_p + \|g\|_p = 1$. Then we can apply Hölder's inequality combined with the $p = 1$ case to conclude

$$\begin{aligned} \int |f(x) + g(x)|^p &\leq \int |f(x)| |f(x) + g(x)|^{p-1} + \int |g(x)| |f(x) + g(x)|^{p-1} \\ &\leq \|f\|_p \|f + g\|_q^{p-1} + \|g\|_p \|f + g\|_q^{p-1} = \|f + g\|_p^{p-1} \end{aligned}$$

Thus $\|f + g\|_p^p \leq \|f + g\|_p^{p-1}$, and simplifying gives $\|f + g\|_p \leq 1$. \square

Remark. Suppose $\|f + g\|_p = \|f\|_p + \|g\|_p$. Following through the proof given above shows that both applications of Hölder's inequality must be sharp. And this is true if and only if $|f(x)|^p$ and $|g(x)|^p$ are scalar multiples of $|f(x) + g(x)|^p$ almost everywhere. But this means $|f(x)|$ and $|g(x)|$ are scalar multiples of $|f(x) + g(x)|$. If $|f(x)| = A|f(x) + g(x)|$ and $|g(x)| = B|f(x) + g(x)|$. If $g \neq 0$, this implies there is C such that $|f(x)| = C|g(x)|$ for some $C > 0$. Thus we can write $f(x) = Ce^{i\theta(x)}g(x)$, and we must have

$$\|f + g\|_p^p = \int |1 + Ce^{i\theta(x)}|^p |g(x)|^p = (1 + C)^p \int |g(x)|^p$$

so $|1 + Ce^{i\theta(x)}| = |1 + C|$ almost everywhere but this can only be true if $e^{i\theta(x)} = 1$ almost everywhere, so $f = Cg$. Thus the triangle inequality is only sharp if f and g are positive scalar multiples of one another.

This discussion leads to a useful heuristic: Unless f and g are 'aligned' in a certain way, the triangle inequality is rarely sharp. For instance, if f and g have disjoint support, we calculate that

$$\|f + g\|_p = \left(\|f\|_p^p + \|g\|_p^p \right)^{1/p}$$

For $p > 1$, this is always sharper than the triangle inequality.

Remark. If $p < 1$, then the proof above no longer works, and in fact, is no longer true. In fact, if f and g are non-negative functions, then we actually have the *anti* triangle inequality

$$\|f + g\|_p \geq \|f\|_p + \|g\|_p.$$

Theorem 6.3. *If $p \geq 1$, then for any positive functions f_1, \dots, f_N ,*

$$(\|f_1\|_p^p + \dots + \|f_N\|_p^p)^{1/p} \leq \|f_1 + \dots + f_N\|_p \leq \|f_1\|_p + \dots + \|f_N\|_p. \quad (6.2)$$

If $p \leq 1$, then the inequality reverses, i.e. for any positive functions f_1, \dots, f_N ,

$$\|f_1\|_p + \dots + \|f_N\|_p \leq \|f_1 + \dots + f_N\|_p \leq (\|f_1\|_p^p + \dots + \|f_N\|_p^p)^{1/p} \quad (6.3)$$

Proof. The upper bound in 6.2 is just obtained by applying the triangle inequality iteratively. To obtain the lower bound, we note that for $A_1, \dots, A_N \geq 0$,

$$(A_1 + \dots + A_N)^p \geq A_1^p + \dots + A_N^p,$$

One can prove this from induction from the inequality $(A_1 + A_2)^p \geq A_1^p + A_2^p$, which holds when $A_2 = 0$, and the derivative of the left hand side is greater than the right hand side for all $A_2 \geq 0$. But then setting $A_k = f_k$ and then integrating gives

$$\|f_1 + \dots + f_N\|_p^p \geq \|f_1\|_p^p + \dots + \|f_N\|_p^p.$$

Now assume $0 < p < 1$. We begin by proving the lower bound in 6.3. We can assume $N = 2$, and $\|f_1\|_p + \|f_2\|_p = 1$, and then it suffices to show $\|f_1 + f_2\|_p \geq 1$. For any $\theta \in (0, 1)$, and $A, B \geq 0$, concavity implies

$$(A + B)^p = (\theta(A/\theta) + (1 - \theta)(B/(1 - \theta)))^p \geq \theta^{1-p}A^p + (1 - \theta)^{1-p}B^p.$$

Thus setting $A = f_1(x)$, $B = f_2(x)$, and $\theta = \|f_1\|_p$, so that $1 - \theta = \|f_2\|_p$, and then integrating, we find

$$\|f_1 + f_2\|_p^p \geq \theta + (1 - \theta) = 1.$$

On the other hand, the inequality $(A_1 + \dots + A_N)^p \leq A_1^p + \dots + A_N^p$, which holds for $A_1, \dots, A_N \geq 0$, can be applied with $f_k = A_k$ and integrated to yield

$$\|f_1 + \dots + f_N\|_p^p \leq \|f_1\|_p^p + \dots + \|f_N\|_p^p. \quad \square$$

Thus the triangle inequality is not satisfied for the L^p norms when $p < 1$. This is one of the deficiencies which leads the L^p theories for $0 < p < 1$ to be rather deficient when compared to the case with $p \geq 1$. This can be fixed in many ways, most notably to harmonic analysts, using the theory of Hardy spaces. On the other hand, we do have a *quasi* triangle inequality.

Theorem 6.4. For $f_1, \dots, f_N \in L^p(X)$, with $0 < p < 1$,

$$\|f_1 + \dots + f_N\|_p \leq N^{1/p-1}(\|f_1\|_p + \dots + \|f_N\|_p).$$

Proof. By Hölder's inequality applied to sums,

$$\|f_1 + \dots + f_N\|_p \leq (\|f_1\|_p^p + \dots + \|f_N\|_p^p)^{1/p} \leq N^{1/p-1}(\|f_1\|_p + \dots + \|f_N\|_p).$$

Thus $L^p(X)$ is a quasinorm space. \square

Remark. When $p < 1$, the space $L^p(X)$ is *not* normable. To see why, we look at the topological features of $L^p(X)$. Fix $\varepsilon > 0$, and let C be a convex set containing all functions f with $\|f\|_p < \varepsilon$. Thus, in particular, C contains all step functions $H\mathbf{I}_E$ where $H|E|^{1/p} < \varepsilon$. But if we now find a countable sequence of disjoint sets $\{E_k\}$, each with positive measure, and for each k , define $H_k = (\varepsilon/2)|E_k|^{-1/p}$, then for any N , the function

$$f_N = (H_1/N)\mathbf{I}_{E_1} + \dots + (H_N/N)\mathbf{I}_{E_N}$$

lies in C , and

$$\|f_N\|_p = (1/N)(H_1^p|E_1| + \dots + H_N^p|E_N|)^{1/p} = (\varepsilon/2)N^{1/p-1}$$

as $N \rightarrow \infty$, the L^p norm of f_N becomes unbounded. In particular, this means that we have proven that every bounded convex subset of $L^p(X)$ has empty interior, and a norm space certainly does not have this property.

As we have mentioned, as $p \rightarrow \infty$, the L^p norm excludes functions with large peaks, or large height, and as $p \rightarrow 0$, the L^p norm excludes functions with large tails, or large width. They form a continuously changing family of functions as p ranges over the positive numbers. In general, there is no inclusion of $L^p(X)$ in $L^q(X)$ for any p, q , except in two circumstances which occur often enough to be mentioned.

Example. If X is a finite measure space, and $0 < p \leq q \leq \infty$, $L^p(X) \subset L^q(X)$. Hölder's inequality implies $\|f\|_p = \|f\chi_X\|_p \leq \|f\|_q \|X\|^{1/p-1/q}$. Taking $q \rightarrow \infty$, we conclude $\|f\|_p \leq \|X\|^{1/p} \|f\|_\infty$. On the other hand, we can rearrange the inequality to read more symmetrically as $\|f\|_p \|X\|^{-1/p} \leq \|f\|_q \|X\|^{-1/q}$. In particular, the L^p norms increase as p increases if X is a probability space.

Example. On the other hand, suppose the measure space is granular, in the sense that there is $\varepsilon > 0$ such that either $|E| = 0$ or $|E| \geq \varepsilon$ for any measurable set E . Then $L^p(X) \supset L^q(X)$ for $0 < p \leq q \leq \infty$. First we check the $q = \infty$ case, which follows by the trivial estimate

$$\int |f(x)|^p \geq \varepsilon \|f\|_\infty^p,$$

so $\|f\|_\infty \leq \|f\|_p \varepsilon^{-1/p}$. But then applying log convexity, if $p \leq q < \infty$, we can write $1/q = \theta/p$ for $0 < \theta \leq 1$, and then log convexity shows

$$\|f\|_q = \|f\|_p^\theta \|f\|_\infty^{1-\theta} \leq \varepsilon^{-(1-\theta)/p} \|f\|_p = \varepsilon^{-1/p-1/q} \|f\|_p.$$

If $\varepsilon = 1$, which occurs if $X = \mathbf{Z}$, then the L^p norms are decreasing in p .

Remark. We can often use such results in spaces which are not granular by coarsening the sigma algebra. For instance, the Lebesgue measure is ε^d granular over the sigma algebra generated by the length ε cubes whose corner's lie on the lattice $(\mathbf{Z}/\varepsilon)^d$, and if a function is measurable with respect to such a σ algebra we call the function ε granular.

Remark. If we let $X = \{1, \dots, N\}$, then X is both finite and granular, so all L^p norms are comparable. In particular, if $p \leq q$,

$$\|f\|_q \leq \|f\|_p \leq N^{1/p-1/q} \|f\|_q.$$

The left hand side of this inequality becomes sharp when f is concentrated at a single point, i.e. $f(n) = \mathbf{I}(n=1)$. On the other hand, the left hand side becomes sharp when f is constant, i.e. $f(n) = 1$ for all n .

Theorem 6.5. If p_θ lies between p_0 and p_1 , then

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

Proof. If $\|f\|_{p_0}, \|f\|_{p_1} < \infty$, then for any p_θ between p_0 and p_1 ,

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

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

Applying the triangle inequality, we conclude that $\|f\|_{p_\theta} < \infty$. In the case where $p_1 = \infty$, then $f\chi_{|f|>1}$ is bounded, and must have finite support if $p_0 < \infty$, which shows this integral is bounded. Note the inequalities above 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. \square

This theorem is important in the study of interpolation theory, because if we have two linear operators T_{p_0} defined on $L^{p_0}(X)$ and T_{p_1} on $L^{p_1}(X)$, and they agree on $L^{p_0}(X) \cap L^{p_1}(X)$, then there is a unique linear operator T_{p_θ} on $L^{p_\theta}(X)$ which agrees with these two functions, and we can consider the boundedness of such a function with respect to the L^{p_θ} norms.

Finally, we note the duality property for the L^p spaces. Given any values of p and q with $1/p + 1/q = 1$, Hölder's inequality implies that if $\|f\|_p < \infty$ and $\|g\|_q < \infty$, then fg is integrable. In particular, the map $\lambda : f \mapsto \int fg$ is a linear functional on L^p . Hölder's inequality implies that $\|\lambda\| \leq \|g\|_q$. But this is actually an equality.

Theorem 6.6. *For any function f with $1 \leq p < \infty$,*

$$\|f\|_p = \sup \left\{ \int f(x)g(x) : \|g\|_q = 1 \right\}.$$

If the underlying measure space is σ finite, then this claim also holds for $p = \infty$.

Proof. Suppose that $1 < p < \infty$. Given f , we define

$$g(x) = \frac{1}{\|f\|_p^{p-1}} \operatorname{sgn}(f(x)) |f(x)|^{p-1}.$$

If $\|f\|_p < \infty$, then

$$\|g\|_q^q = \frac{1}{\|f\|_p^{pq-q}} \int |f(x)|^{pq-q} = \frac{1}{\|f\|_p^p} \|f\|_p^p = 1,$$

and

$$\int f(x)g(x) = \frac{1}{\|f\|_p^{p-1}} \int |f(x)|^p = \|f\|_p.$$

On the other hand, suppose $\|f\|_p = \infty$. Then there exists a sequence of step functions $s_1 \leq s_2 \leq \dots \rightarrow |f|$, and so $\|s_k\|_p \rightarrow \infty$. For each k , find a function $g_k \geq 0$ with $\|g_k\|_q = 1$, and $\int g_k(x)s_k(x) \geq \|s_k\|_p/2$. Then

$$\int g_k(x) \operatorname{sgn}(f(x))f(x) = \int g_k(x)|f(x)| \geq \int g_k(x)s_k(x) \geq \|s_k\|_p/2 \rightarrow \infty,$$

which completes the proof.

Now suppose $p = 1$. Then, given any f , we can define

$$g(x) = \operatorname{sgn}(f(x)).$$

Then $\|g\|_\infty = 1$, and

$$\int f(x)g(x) = \int |f(x)| = \|f\|_1.$$

Note this holds even if $\|f\|_1 = \infty$.

Finally, suppose $p = \infty$. Given any f , fix $\varepsilon > 0$. Then we can find a set E with $0 < |E| < \infty$ such that $|f(x)| \geq \|f\|_\infty - \varepsilon$ for $x \in E$. If $g(x) = \operatorname{sgn}(f(x))\mathbf{I}_{E/|E|}$, then $\|g\|_1 = 1$, and

$$\int f(x)g(x) = \frac{1}{|E|} \int_E |f(x)| \geq \|f\|_\infty - \varepsilon.$$

Taking $\varepsilon \rightarrow 0$ completes the claim. \square

6.2 Weak and Lorentz Norms

The weak L^p norms are obtained as a slight ‘refinement’ of the L^p norms. The first idea to understand these norms is that any translation invariant property of a function f should express itself as a property of its distribution function $F : [0, \infty) \rightarrow [0, \infty)$ defined by $F(t) = |\{|f| > t\}|$. Then F is right-continuous and decreasing. Of course, for other functions g, h, \dots , the corresponding distribution functions are denoted G, H, \dots .

Lemma 6.7. *Given a function f and g , $\alpha \in \mathbf{C}$, and $t, s > 0$, then*

- *If $|g| \leq |f|$, then $G \leq F$.*
- *If $g = \alpha f$, then $G(t) = F(t/|\alpha|)$.*
- *If $h = f + g$, then $H(t + s) \leq F(t) + G(s)$.*
- *If $h = fg$, then $H(ts) \leq F(t) + G(s)$.*

Proof. The first point is obvious, as is the second point. The third point follows because if $|f(x) + g(x)| \geq t + s$, then either $|f(x)| \geq t$ or $|g(x)| \geq s$. Finally, if $|f(x)g(x)| \geq ts$, then $|f(x)| \geq t$ or $|g(x)| \geq s$. \square

Theorem 6.8. *If ϕ is an increasing, differentiable function on the real line with $\phi(0) = 0$, then*

$$\int_X \phi(|f(x)|) = \int_0^\infty \phi'(t) F(t) dt$$

Proof. An application of Fubini's theorem is all that is needed to show

$$\begin{aligned} \int_X \phi(|f(x)|) dx &= \int_X \int_0^{|f(x)|} \phi'(t) dt dx \\ &= \int_0^\infty \phi'(t) \int_{|f(x)| > t} dx du \\ &= \int_0^\infty \phi'(t) F(t) dt. \end{aligned} \quad \square$$

As a special case we find

$$\|f\|_p = \left(p \int_0^\infty F(t) t^p \frac{dt}{t} \right)^{1/p}.$$

Thus in order for $\|f\|_p$ to be finite, we need $F(t)t^p$ to be integrable with respect to the Haar measure on \mathbf{R}^+ . For this to be true, $F(t)$ must tend to zero slightly faster than $1/t^p$. Indeed, we find

$$F(t) = |\{|f|^p > t^p\}| \leq \frac{1}{t^p} \int |f|^p = \frac{\|f\|_p^p}{t^p},$$

a fact known as *Chebyshev's inequality*. But if we change the integrality condition to the condition that $F(t)t^p \in L^q(\mathbf{R}^+)$ for some $0 < q \leq \infty$, we obtain a different condition, and this leads to the definition of the *Lorentz norms*.

The most basic Lorentz norm is the *weak L^p norm*, denoted by $\|f\|_{p,\infty}$, which is the smallest value A such that $F(t) \leq (A/t)^p$. We let $L^{p,\infty}(X)$ denote the space of all functions f for which $\|f\|_{p,\infty} < \infty$. By Chebyshev's inequality, $L^p(X) \subset L^{p,\infty}(X)$. The reason that the value A occurs within the brackets is so that the norm is homogenous; if $g = \alpha f$, and $\|f\|_{p,\infty} = A$, then

$$G(t) = F(t/|\alpha|) \leq \left(\frac{A|\alpha|}{t} \right)^p,$$

so $\|\alpha f\|_{p,\infty} = |\alpha| \|f\|_p$. The weak norms do not satisfy a triangle inequality, but they do satisfy a quasitriangle inequality. If $g = f_1 + \cdots + f_N$, with $\|f_k\|_{p,\infty} = A_k$, then

$$\begin{aligned} G(t) &\leq F_1(t/N) + \cdots + F_N(t/N) \\ &\leq (NA_1/t)^p + \cdots + (NA_N/t)^p = \frac{N^p(A_1^p + \cdots + A_N^p)}{t^p}. \end{aligned}$$

Thus $\|G\|_{p,\infty} \leq N(A_1^p + \cdots + A_N^p)^{1/p}$. If $p \geq 1$, then this is upper bounded by $N(A_1 + \cdots + A_N)$, so

$$\|G\|_{p,\infty} \leq N(\|f_1\|_{p,\infty} + \cdots + \|f_N\|_{p,\infty}).$$

If $p \leq 1$, we instead bound $(A_1^p + \cdots + A_N^p)^{1/p}$ by $N^{1/p}(A_1 + \cdots + A_N)$, so that

$$\|G\|_{p,\infty} \leq N^{1+1/p}(\|f_1\|_{p,\infty} + \cdots + \|f_N\|_{p,\infty}).$$

The weak norm provides another monotone translation invariant measure, and it often comes up when finer tuning is needed in certain interpolation arguments, especially when dealing with maximal functions.

Example. If $f = H\mathbf{I}_E$, with $|E| = W$, then

$$F(t) = W \cdot \mathbf{I}_{[0,H]}.$$

Thus

$$\|f\|_{p,\infty} = \left(\sup_{0 \leq t < H} W t^p \right)^{1/p} = W^{1/p} H^p = \|f\|_p.$$

If $f = H_1\mathbf{I}_{E_1} + H_2\mathbf{I}_{E_2}$, with $|E_1| = W_1$ and $|E_2| = W_2$, with $H_1 \leq H_2$, then

$$F(t) = \begin{cases} W_1 + W_2 & : t < H_1, \\ W_2 & : t < H_2, \\ 0 & : \text{otherwise.} \end{cases}$$

Thus

$$\|f\|_{p,\infty} = \left(\max((W_1 + W_2)H_1^p, W_2H_2^p) \right)^{1/p} = \max((W_1 + W_2)^{1/p}H_1, W_2^{1/p}H_2).$$

Example. The function $f(x) = 1/|x|^\alpha$ does not lie in any $L^p(\mathbf{R}^d)$, but lies in $L^{p,\infty}$ precisely when $p = d/\alpha$, since

$$|\{1/|x|^{p\alpha} > t\}| = \left| \left\{ |x| \leq \frac{1}{t^{1/p\alpha}} \right\} \right| \propto_d \frac{1}{t^{d/p\alpha}}.$$

Thus the weak L^p norms are able to more precisely determine the distribution of certain functions.

In general, we define the Lorentz spaces $L^{p,q}(X)$ as the space of functions with finite Lorentz norm

$$\|f\|_{p,q} = q^{1/q} \|tF^{1/p}\|_{L^q(\mathbf{R}^+)} = \left(\int_0^\infty q(tF(t)^{1/p})^q \frac{dt}{t} \right)^{1/q}.$$

Thus $\|f\|_{p,\infty}$ gives the weak L^p norm, and $\|f\|_{p,p} = \|f\|_p$. The Lorentz norm often occurs with other constants attached, but this only scales the Lorentz norm, and is rarely relevant to the problem at hand, since we normally only desire bounds of the form $\|f\|_{p,q} \lesssim_{p,q} A$, for some values A . The main case are interested in which we haven't already encountered is when $q = 1$, for which

$$\|f\|_{p,1} = \int_0^\infty F(t)^{1/p} dt.$$

We shall show that $\|f\|_{p,\infty} \lesssim_{p,q} \|f\|_{p,p} \lesssim_{p,q} \|f\|_{p,1}$ for all functions f , so that as we decrease the value of q , the norm $\|\cdot\|_{p,q}$ becomes a stronger measurement.

The fact that multiplying a function by a constant dilates the distribution implies that the Lorentz norm is homogenous. We do not have a triangle inequality for the Lorentz norms, but we have a quasi triangle inequality.

Theorem 6.9. All the Lorentz spaces satisfy a quasi triangle inequality. More precisely, if $p, q \geq 1$, then

$$\|f_1 + \cdots + f_N\|_{p,q} \leq N^{1+[p \leq 1](1/p-1)+[q \leq 1](1/q-1)} (\|f_1\|_{p,q} + \cdots + \|f_N\|_{p,q}).$$

Proof. We calculate that if $g = f_1 + \cdots + f_N$,

$$\begin{aligned}\|g\|_{p,q} &= \left(q \int_0^\infty \left[t G(t)^{1/p} \right]^q \frac{dt}{t} \right)^{1/q} \\ &\leq \left(q \int_0^\infty \left[t (F_1(t/N) + \cdots + F_N(t/N))^{1/p} \right]^q \frac{dt}{t} \right)^{1/q} \\ &\leq N \left(q \int_0^\infty \left[t (F_1(t) + \cdots + F_N(t))^{1/p} \right]^q \frac{dt}{t} \right)^{1/q}.\end{aligned}$$

If $p \geq 1$, then

$$(F_1(t) + \cdots + F_N(t))^{1/p} \leq F_1(t)^{1/p} + \cdots + F_N(t)^{1/p},$$

whereas if $p \leq 1$, Hölder's inequality shows

$$(F_1(t) + \cdots + F_N(t))^{1/p} \leq N^{1/p-1} \left(F_1(t)^{1/p} + \cdots + F_N(t)^{1/p} \right),$$

If $q \geq 1$, the two inequalities above, combined with the triangle inequality in $L^q(\mathbf{R}^+)$, yield the required bound. If $q \leq 1$, then we apply the quasitriangle inequality. \square

6.3 Dyadic Layer Cake Decompositions

An important trick to utilizing Lorentz norms is by utilizing a layer cake decomposition. The dyadic layer cake decompositions enable us to understand a function by breaking it up into parts upon which we can control the height or width of a function. The idea is quite simple. Given a function f , we can define $f_k(x) = f(x) \mathbf{I}(2^{k-1} \leq |f(x)| \leq 2^k)$. Then $f = \sum_{k \in \mathbf{Z}} f_k$, and we have control over the height of each f_k . This is the *vertical layer cake decomposition*. On the other hand, we can also define $H_k = \inf\{t : F(t) \leq 2^k\}$. Then if $f_k(x) = f(x) \mathbf{I}(H_k \leq |f(x)| \leq H_{k-1})$. Then we have explicit control over the width of f_k , supported on a set of measure at most 2^k , and some, non-explicit control over its height. This is the *horizontal layer cake decomposition*.

Theorem 6.10. *There exists constants A_1, \dots, A_5 , all comparable up to a constant depending on p and q , such that*

$$1. \|f\|_{p,q} \leq A_1.$$

2. We can write $f = \sum_{k \in \mathbf{Z}} f_k$, where $2^{k-1} \leq f_k(x) \leq 2^k$ for all x , f_k is supported on a set with width W_k , and

$$\left(\sum_{k \in \mathbf{Z}} [2^k W_k^{1/p}]^q \right)^{1/q} \leq A_2.$$

3. We can write $|f(x)| \leq \sum_{k \in \mathbf{Z}} 2^k \mathbf{I}_{E_k}$, where $|E_k| = W_k$ and

$$\left(\sum_{k \in \mathbf{Z}} [2^k W_k^{1/p}]^q \right)^{1/q} \leq A_3.$$

4. We can write $f(x) = \sum_{k \in \mathbf{Z}} f_k$, where f_k is supported on a set with measure $W_k \leq 2^k$, $H_{k+1} \leq |f(x)| \leq H_k$ on the support of f , and

$$\left(\sum_{k \in \mathbf{Z}} [H_k 2^{k/p}]^q \right)^{1/q} \leq A_4.$$

5. We can write $|f(x)| \leq \sum_{k \in \mathbf{Z}} H_k \mathbf{I}_{E_k}$, where $|E_k| \leq 2^k$, and

$$\left(\sum_{k \in \mathbf{Z}} [H_k 2^{k/p}]^q \right)^{1/q} \leq A_5.$$

Proof. It is obvious that we can always select $A_3 \leq A_2$ and $A_5 \leq A_4$. Next, we bound A_2 in terms of A_1 by performing a vertical layer cake decomposition on f . If we write $f = \sum_{k \in \mathbf{Z}} f_k$, then f_k is supported on a set with measure $W_k = F(2^{k-1}) - F(2^k) \leq F(2^{k-1})$, and so

$$\begin{aligned} \sum_{k \in \mathbf{Z}} [2^k W_k^{1/p}]^q &\leq \sum_{k \in \mathbf{Z}} [2^k F(2^{k-1})^{1/p}]^q \\ &\lesssim_q \sum_{k \in \mathbf{Z}} [2^{k-1} F(2^k)^{1/p}]^q \\ &\lesssim \sum_{k \in \mathbf{Z}} \int_{2^{k-1}}^{2^k} [t F(t)^{1/p}]^q \frac{dt}{t} \lesssim_q \|f\|_{p,q}^q \leq A_1^q. \end{aligned}$$

Thus $A_2 \lesssim_q A_1$. Next, we bound A_4 in terms of A_1 . Perform a horizontal layer cake decomposition, writing $f = \sum f_k$, where f_k is supported on a set with measure $W_k \leq 2^k$, and $H_{k+1} \leq |f_k(x)| \leq H_k$. Then a telescoping sum shows

$$\begin{aligned} H_k 2^{k/p} &= \left(\sum_{m=0}^{\infty} (H_{k+m}^q - H_{k+m+1}^q) 2^{kq/p} \right)^{1/q} \\ &\lesssim_q \left(\sum_{m=0}^{\infty} \int_{H_{k+m+1}}^{H_{k+m}} [t 2^{k/p}]^q \frac{dt}{t} \right)^{1/q} \\ &\leq \left(\sum_{m=0}^{\infty} \int_{H_{k+m+1}}^{H_{k+m}} [t F(t)^{1/p}]^q \frac{dt}{t} \right)^{1/q} \end{aligned}$$

Thus

$$\left(\sum_{k \in \mathbf{Z}} [H_k 2^{k/p}]^q \right)^{1/q} \leq \left(\int_0^\infty [t F(t)^{1/p}]^q \frac{dt}{t} \right)^{1/q} \lesssim_q A_1.$$

Thus $A_4 \lesssim_q A_1$. It remains to bound A_1 by A_5 and A_3 . Given A_3 , we can write $|f(x)| \leq \sum 2^k \mathbf{1}_{E_k}$, where $|E_k| \leq W_k$. We then find

$$F(2^k) \leq \sum_{m=1}^{\infty} W_{k+m}.$$

Thus

$$\int_{2^{k-1}}^{2^k} [t F(t)^{1/p}]^q \frac{dt}{t} \lesssim \left[2^k \left(\sum_{m=0}^{\infty} W_k \right)^{1/p} \right]^q.$$

Thus if $q \leq p$,

$$\begin{aligned}
\|f\|_{p,q} &\lesssim_q \left(\sum_{k \in \mathbf{Z}} \left[2^k \left(\sum_{m=0}^{\infty} W_{k+m} \right)^{1/p} \right]^q \right)^{1/q} \\
&\leq \left(\sum_{k \in \mathbf{Z}} \sum_{m=0}^{\infty} \left[2^k W_{k+m}^{1/p} \right]^q \right)^{1/q} \\
&\leq \left(\sum_{m=0}^{\infty} 2^{-qm} \sum_{k \in \mathbf{Z}} \left[2^{k+m} W_{k+m}^{1/p} \right]^q \right)^{1/q} \\
&\leq \left(A_3^q \sum_{m=0}^{\infty} 2^{-mq} \right)^{1/q} \lesssim_q A_3.
\end{aligned}$$

If $q \geq p$, we can employ the triangle inequality for $l^{q/p}$ to write

$$\begin{aligned}
\|f\|_{p,q} &\lesssim_q \left(\sum_{k \in \mathbf{Z}} \left[2^k \left(\sum_{m=0}^{\infty} W_{k+m} \right)^{1/p} \right]^q \right)^{1/q} \\
&\leq \left(\sum_{m=0}^{\infty} \left(\sum_{k \in \mathbf{Z}} 2^{kq} W_{k+m}^{q/p} \right)^{p/q} \right)^{1/p} \\
&\leq \left(A_3^p \sum_{m=0}^{\infty} 2^{-mq} \right)^{1/p} \lesssim_{p,q} A_3.
\end{aligned}$$

The bound of A_1 in terms of A_5 involves the same ‘shifting’ technique, and is left to the reader. \square

Remark. Heuristically, the theorem above says that if $f = f_1 + \dots + f_N$, where f_k is supported on a set with width approximately W_k , and on its support $|f_k(x)| \sim H_k$, and if either the heights or widths grow as powers of two, then

$$\|f\|_{p,q} \lesssim_{p,q} \left(H_1^q W_1^{q/p} + \dots + H_N^q W_N^{q/p} \right)^{1/q}.$$

Thus the $L^{p,q}$ norm has little interaction between elements of the sum when the sum occurs over dyadically different heights or width.

Remark. Suppose that on the support of f , $A \leq |f(x)| \leq AN$. If we perform a dyadic layer cake decomposition, we can write f as the sum of $O(\log N)$ functions $f_K, f_{K+1}, \dots, f_{K+\log N}$. In particular, this means that for $q_1 < q_2$,

$$\left(\sum_{k=K}^{K+\log N} 2^{kq_1} W_k^{q_1/p} \right)^{1/q_1} \leq (\log N)^{1/q_1 - 1/q_2} \left(\sum_{k=K}^{K+\log N} 2^{kq_2} W_k^{q_2/p} \right)^{1/q_2}.$$

Thus the $L^{p,q}$ norms differ by a multiplicative power of $\log N$.

Remark. Conditions (2), (3), (4), and (5) are all defined with respect to an l^q norm. In particular, this means that $L^{p,q_1} \subset L^{p,q_2}$ if $q_1 \leq q_2$.

Here is a simple consequence, due to O'Neil.

Theorem 6.11. *If $0 < p_1, p_2, p < \infty$ and $0 < q_1, q_2, q < \infty$ with*

$$1/p = 1/p_1 + 1/p_2 \quad \text{and} \quad 1/q = 1/q_1 + 1/q_2,$$

then

$$\|fg\|_{p,q} \lesssim_{p_1,p_2,q_1,q_2} \|f\|_{p_1,q_1} \|g\|_{p_2,q_2}.$$

Proof. Without loss of generality, assume $\|f\|_{p_1,q_1} = \|g\|_{p_2,q_2} = 1$. Perform horizontal layer cake decompositions of f and g , writing $|f| \leq \sum_{k \in \mathbf{Z}} H_k \mathbf{I}_{E_k}$ and $|g| \leq \sum_{k \in \mathbf{Z}} H'_k \mathbf{I}_{F_k}$, where $|E_k|, |F_k| \leq 2^k$. Then

$$|fg| \leq \sum_{k,k' \in \mathbf{Z}} H_k H'_k \mathbf{I}_{E_k \cap F_{k'}}$$

For each fixed k , $|E_{k+m} \cap F_m| \leq 2^m$, and so

$$\begin{aligned} \left\| \sum_{m \in \mathbf{Z}} H_{k+m} H'_m \mathbf{I}_{E_{k+m} \cap F_m} \right\|_{p,q} &\lesssim_{p,q} \left(\sum_{m \in \mathbf{Z}} [H_{k+m} H'_m 2^{m/p}]^q \right)^{1/q} \\ &= \left(\sum_{m \in \mathbf{Z}} \left[(H_{k+m} 2^{m/p_1}) (H'_m 2^{m/p_2}) \right]^q \right)^{1/q} \\ &\leq \left(\sum_{m \in \mathbf{Z}} [H_{k+m} 2^{m/p_1}]^{q_1} \right)^{1/q_1} \left(\sum_{m \in \mathbf{Z}} [H'_m 2^{m/p_2}]^{q_2} \right)^{1/q_2} \lesssim_{p,q,p_1,q_1,p_2,q_2} 2^{-k/p_1} \end{aligned}$$

Summing over $k > 0$ gives that

$$\left\| \sum_{k \geq 0} \sum_{m \in \mathbf{Z}} H_{k+m} H'_m \mathbf{I}_{E_{k+m} \cap F_m} \right\| \lesssim_{p,q,p_1,q_1,p_2,q_2} 1$$

By the quasitriangle inequality, it now suffices to obtain a bound

$$\left\| \sum_{k < 0} \sum_{m \in \mathbf{Z}} H_{k+m} H'_m \mathbf{I}_{E_{k+m} \cap F_m} \right\|_{p,q}.$$

This is done in much the same way, but instead using the bound $|E_{k+m} \cap F_m| \leq 2^{k+m}$. \square

Corollary 6.12. *If $p > 1$ and $q > 0$, $L^{p,q}(X) \subset L^{1,loc}(X)$.*

Proof. Let E have finite measure and let $f \in L^{p,q}(X)$. Then the Hölder's inequality for Lorentz spaces shows

$$\|f\|_{L^1(E)} = \|\mathbf{I}_E f\|_1 \lesssim_{p,q} |E|^{1-1/p} \|f\|_{p,q} < \infty. \quad \square$$

Finally, we consider the duality of the $L^{p,q}$ norms.

Theorem 6.13. *Let $1 < p < \infty$ and $1 \leq q \leq \infty$. Then if $f \in L^{p,q}(X)$,*

$$\|f\|_{p,q} \sim \sup \left\{ \int f g : \|g\|_{p',q'} \leq 1 \right\}.$$

Proof. The upper bound is an easy consequence of Hölder's inequality. For the case $q = \infty$, we actually prove

$$\|f\|_{p,\infty} \gtrsim_p \sup \left\{ |E|^{-1/p'} \left| \int_E f \right| : |E| < \infty \right\},$$

If $F(t_0) \geq (A/t_0)^p$ for some t_0 , then setting $E_\theta = \{x : |f(x)| = r e^{i\theta}, r > t_0\}$, we find

$$\int_0^{2\pi} |E_\theta| \geq (A/t_0)^p.$$

In particular, if $\varepsilon > 0$ is fixed, pigeonholing shows we can find a set

$$E = \bigcup_{\eta=\theta}^{\theta+\pi/4} E_\theta$$

such that $|E| \geq (A/t_0)^p/8$. Thus

$$\begin{aligned} |E|^{-1/p'} \left| \int_E f(x) dx \right| &\geq |E|^{-1/p'} \operatorname{Re} \left(\int_E e^{-i\theta} f(x) dx \right) \\ &\gtrsim |E|^{-1/p'} \int_E |f(x)| dx \gtrsim_p t_0 \cdot |E|^{1/p} \gtrsim A. \end{aligned}$$

Now we consider the case with $q < \infty$. Without loss of generality, we may assume $\|f\|_{p,q} = 1$. We may perform a vertical layer cake decomposition, writing $f = \sum_{k \in \mathbb{Z}} f_k$, where $2^{k-1} \leq |f_k(x)| \leq 2^k$, is supported on a set with width W_k , and

$$\left((2^k W_k^{1/p})^q \right) \sim_{p,q} 1.$$

Define $a_k = 2^k W_k^{1/p}$, and set $g = \sum_{k \in \mathbb{Z}} g_k$, where $g_k(x) = a_k^{q-p} \operatorname{sgn}(f_k(x)) |f_k(x)|^{p-1}$. Then

$$\begin{aligned} \int f(x) g(x) &= \sum_{k \in \mathbb{Z}} \int f_k(x) g_k(x) = \sum_{k \in \mathbb{Z}} a_k^{q-p} \int |f_k(x)|^p \\ &\gtrsim_p \sum_{k \in \mathbb{Z}} a_k^{q-p} W_k 2^{kp} = \sum_{k \in \mathbb{Z}} a_k^q \gtrsim_{p,q} 1. \end{aligned}$$

We therefore need to show that $\|g\|_{p',q'} \lesssim 1$. We note $|g_k(x)| \lesssim a_k^{q-p} 2^{kp}$, and has width W_k . This gives a decomposition of g , but neither the height nor the widths necessarily in powers of two. Still, we can fix this since the heights increase exponentially; define

$$H_k = \sup_{l \geq 0} a_{k-l}^{q-p} 2^{kp} 2^{-lp/2}.$$

Then $|g_k(x)| \lesssim_{p,q} H_k$, and $H_{k+1} \geq 2^{p/2} H_k$. In particular, if we pick m such that $2^{mp/2} \geq 1$, then for any $l \leq m$, the sequence H_{km+l} , as k ranges over values, increases dyadically, and so by the quasitriangle inequality for the

$L^{p',q'}$ norm, and then the triangle inequality in l^q , we find

$$\begin{aligned}
\|g\|_{p',q'} &\lesssim_{m,p,q} \left(\sum [H_k W_k^{1/p'}]^{q'} \right)^{1/q'} \\
&\lesssim \left(\sum_{k \in \mathbb{Z}} \left[\left(\sup_{l \geq 0} a_{k-l}^{q-p} 2^{kp} 2^{-lp/2} \right) (a_k 2^{-k})^{p-1} \right]^{q'} \right)^{1/q'} \\
&\lesssim_p \left(\sum_{k \in \mathbb{Z}} \left[a_k^{p-1} \sum_{l=0}^{\infty} a_{k-l}^{q-p} 2^{-lp/2} \right]^{q'} \right)^{1/q'} \\
&\lesssim \sum_{l=0}^{\infty} 2^{-lp/2} \left(\sum_{k \in \mathbb{Z}} [a_k^{p-1} a_{k-l}^{q-p}]^{q'} \right)^{1/q'}.
\end{aligned}$$

Applying's Hölder's inequality shows

$$\begin{aligned}
\left(\sum_{k \in \mathbb{Z}} [a_k^{p-1} a_{k-l}^{q-p}]^{q'} \right)^{1/q'} &\leq \left(\sum_{k \in \mathbb{Z}} a_k^q \right)^{(p-1)/q} \left(\sum_{k \in \mathbb{Z}} a_{k-l}^q \right)^{(q-p)/q} \\
&\lesssim_{p,q} \|f\|_{p,q}^{q-1} \lesssim_{p,q} 1. \quad \square
\end{aligned}$$

Remark. The technique given in this proof shows that if $f = \sum f_k$, where f_k is supported on a set with measure W_k , and $|f_k(x)| \leq H_k$, then provided $H_k \sim 2^{ck}$ for any k , we can still perform a dyadic layer cake decomposition when measures the $L^{p,q}$ norms.

6.4 Normability of the Lorentz Spaces

Though the Lorentz norms do not satisfy the triangle inequality, the space $L^{p,q}(X)$ is still a 'Banach-able' space when $p > 1$, and $q \geq 1$. First off, the standard proof shows the norm gives a complete quasimetric, since a Cauchy sequence in the $L^{p,q}$ norm converges to a function almost everywhere, which is easily verified to have finite $L^{p,q}$ norm. The easiest way to define a norm is to define the *non-decreasing rearrangement* of a function f as $f^*(t) = \inf\{s : F(s) \leq t\}$. Then f^* is non-increasing, and continuous on the right, and $f^*(t) > s$ if and only if $t < F(s)$, so

$$|\{f^*(t) > \alpha\}| = |\{0 < t < F(\alpha)\}| = F(\alpha)$$

so f and f^* have the same distribution. Also, $F(f^*(t)) \leq t$, and $(\alpha f)^*(t) = |\alpha| f^*(t)$. Furthermore,

$$\|f\|_{p,q} = \left(\frac{q}{p} \int_0^\infty (t^{1/p} f^*(t))^q \frac{dt}{t} \right)^{1/q},$$

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

Lemma 6.14. *For any measurable set E ,*

$$\int_E |f(x)| dx \leq \int_0^{|E|} f^*(t) dt.$$

and

$$\int_{\{|f(x)| > t\}} |f(x)| dx = \int_0^{F(t)} f^*(t) dt.$$

Proof. If $g \leq f$, $g^* \leq f^*$. Thus $(\chi_E f)^* \leq f^*$, so

$$\int_E |f(x)| dx = \int \chi_E |f(x)| = \int_0^\infty (\chi_E f)^*(t) dt = \int_0^{|E|} (\chi_E f)^*(t) dt \leq \int_0^{|E|} f^*(t) dt.$$

On the other hand, $(\chi_E f)^* = f^*$ when $E = \{|f(x)| > t\}$, which gives the second equality. \square

For a function f and $t > 0$, we define a family of averages

$$m(t) = \frac{1}{t} \int_0^\infty f^*(t) dt.$$

For any fixed $t > 0$, the map $f \mapsto m(t)$ is a norm. Provided our measure space is non-atomic, we have

$$m(t) = \sup_{|E| \leq t} \int_E |f(x)| d\mu.$$

We define

$$\|f\|_{p,q} = \left(\frac{q}{p} \int_0^\infty [t^{1/p} m(t)]^q \frac{dt}{t} \right)^{1/q}$$

and

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

For $q \geq 1$, each of these functions is a norm, simply because the function m is a norm. On the other hand, since f^* is decreasing, $f^*(t) \leq m(t)$ for all t , which shows $\|f\|_{p,q} \geq \|f\|_{p,q}$. If $p = 1$ and $q < \infty$, if $\|f\|_{1,q} < \infty$, then $f = 0$, so these norms are effectively useless. If $q = \infty$, then $\|f\|_{1,\infty} = \|f\|_1$, and therefore doesn't measure the correct norm. But in all other cases, i.e. for $p > 1$ and $q \geq 1$, the norm is comparable to the $L^{p,q}$ norm.

Theorem 6.15.

$$\|f\|_{p,q} \leq \frac{p}{p-1} \|f\|_{p,q}.$$

Proof. We utilize a *Hardy's inequality* technique, which shows that the L^p norm of the averages of a function is comparable to the L^p norm of the function. Applying Minkowski's integral inequality, we conclude that

$$\begin{aligned} \left(\frac{q}{p} \int_0^\infty [t^{1/p} m(t)]^q \frac{dt}{t} \right)^{1/q} &= \left(\frac{q}{p} \int_0^\infty \left(\int_0^1 t^{1/p} f^*(ts) ds \right)^q \frac{dt}{t} \right)^{1/q} \\ &\leq \int_0^1 \left(\int_0^\infty \frac{q}{p} (t^{1/p} f^*(ts))^q \frac{dt}{t} \right)^{1/q} ds \\ &\leq \left(\int_0^1 s^{-1/p} ds \right) \left(\frac{q}{p} \int_0^\infty (t^{1/p} f^*(t))^q \frac{dt}{t} \right)^{1/q} \\ &\leq \frac{1}{1-1/p} \|f\|_{p,q} = \frac{p}{p-1} \|f\|_{p,q}. \end{aligned}$$

For $q = \infty$, and $t > 0$, we have

$$\begin{aligned} t^{1/p} m(t) &= t^{1/p-1} \int_0^t f^*(s) ds \\ &\leq (\sup_{s>0} s^{1/p} f^*(s)) t^{1/p-1} \int_0^t t^{-1/p} ds \\ &= \frac{1}{1-1/p} \|f\|_{p,\infty} = \frac{p}{p-1} \|f\|_{p,\infty}. \end{aligned}$$

since t was arbitrary, this gives the required bound. \square

Theorem 6.16. *If $p > 1$, there exist a norm on $L^{p,\infty}(X)$, which induces the same topology on $L^{p,\infty}(X)$, for which $L^{p,\infty}(X)$ is a Banach space. If $0 < p \leq 1$, $L^{p,\infty}(X)$ is metrizable.*

Proof. Let $|E| < \infty$, and $\|f\|_{p,\infty} < \infty$. Then if $r < p$,

$$\begin{aligned}
\int_E |f(x)|^r dx &\leq r \int_0^\infty \min(|E|, F(t)) t^r \frac{dt}{t} \\
&\leq r \int_0^\infty \min(|E|, \|f\|_{p,\infty}^p t^{-p}) t^r \frac{dt}{t} \\
&\leq r \left(\int_0^{\|f\|_{p,\infty}^p |E|^{-1/p}} |E| t^r \frac{dt}{t} + \int_{\|f\|_{p,\infty}^p |E|^{-1/p}}^\infty \|f\|_{p,\infty}^p t^{r-p} \frac{dt}{t} \right) \\
&\leq \frac{p}{p-r} \|f\|_{p,\infty}^r |E|^{1-r/p}.
\end{aligned}$$

Thus on a finite measure space X , $L^p(X) \subset L^{p,\infty}(X) \subset L^q(X)$. For a fixed $r < p$, define

$$\|f\|_{p,\infty} = \sup_{0 < |E| < \infty} |E|^{-1/r+1/p} \left(\int_E |f(x)|^r dx \right)^{1/r}.$$

Our calculations above show

$$\|f\|_{p,\infty} \leq \left(\frac{p}{p-r} \right)^{1/r} \|f\|_{p,\infty}.$$

Conversely, if we set $E = \{|f(x)| > \alpha\}$, we obtain that

$$|E| \leq \frac{1}{\alpha^r} \int_E |f(x)|^r \leq \|f\|_{p,\infty}^r |E|^{1-r/p}$$

so

$$|E| \leq \frac{\|f\|_{p,\infty}^p}{\alpha^p},$$

which implies $\|f\|_{p,\infty} \leq \|f\|_{p,\infty}$. Thus these two quantities are always comparable. But it is obvious that $\|\alpha f\| = |\alpha| \|f\|$, and $\|f+g\|_{p,\infty} \leq \|f\|_{p,\infty} + \|g\|_{p,\infty}$, so $L^{p,\infty}(X)$ has the structure of a Banach space under the norm $\|\cdot\|_{p,\infty}$. \square

Remark. When $p = 1$, $q > 1$, the space $L^{p,q}(X)$ is not normable. We prove this when $X = \mathbf{R}$. Indeed, for each k , let

$$f_k(x) = \frac{1}{|x-k|}.$$

Then $\|f_k\|_{1,\infty} \lesssim 1$ is bounded independently of k . If $\|\cdot\|$ was a norm giving the same topology as $L^{p,q}(X)$, then if $f = \sum_{k=-N}^N f_k$, the triangle inequality would imply $\|f\| \lesssim N$. But if $|x| \leq N$,

$$f(x) = \sum_{k=-N}^N \frac{1}{|x-k|} = \int_{1 \leq |x-y| \leq N} \frac{dy}{y} = \log(x+N)$$

6.5 Metric Structure of Norms

The geometric interest of these quantitative norms are that we can define a ‘distance’ between two functions f and g by the magnitude of a norm of the function $f - g$. Thus, for instance, the $L^1(X)$ can be given a geometric structure by defining the distance between two functions f and g as $\|f - g\|_p$. The fact that we can treat functions as points in an ‘infinite dimensional geometry’ is the fundamental foundation of the field of functional analysis. For $p \geq 1$, these norms do actually satisfy the triangle inequality, giving rise to a metric. For $0 < p < 1$, the norms give rise to a quasimetric.

One reason to discuss norms in the context of measure theory is that it enlightens us on the sense that the Lebesgue integral is that it is a ‘completion’ of the Riemann integral. We have already seen this principle in the discussion of the metric $d(E, F) = |E \Delta F|$, for which the family of Lebesgue measurable sets ‘complete’ the metric on Jordan measurable sets or elementary sets. Norms enable us to see this principle in a deeper context, for a norm gives rise to a metric on the family of measurable functions, and the Riemann integrable functions in this metric are dense in this space, so that the Lebesgue functions ‘fill the holes’ with respect to any of the given norms. The triangle inequality for the norms says that the norm works like a metric on space.

The triangle inequality is only ever sharp if f and g are correlated in some manner. For instance, if the supports of f and g are disjoint, then we find

$$\|f + g\|_p^p = \int (f + g)^p = \int f^p + g^p = \|f\|_p^p + \|g\|_p^p$$

so $\|f + g\|_p = (\|f\|_p^p + \|g\|_p^p)^{1/p}$, for any $0 < p < \infty$, with the natural infinite extension being $\|f + g\|_\infty = \max(\|f\|_\infty, \|g\|_\infty)$. In fact, the only time where $\|f + g\|_p = \|f\|_p + \|g\|_p$ is when f and g are positive scalar multiples of one

another (up to a set of measure zero). To see this, note that if equality held, then we would have, except on a set of measure zero, $|f + g|^p = \theta^{1-p}|f|^p + (1-\theta)^{1-p}|g|^p$. We claim that this forces $g = (1-\theta)/\theta f$ (TODO).

If $0 < p < 1$, the triangle inequality isn't satisfied, which can be seen in the case of the l_p norm on \mathbf{R}^n , which isn't convex. However, the discrete sum implies we have a pointwise estimate $|A + B|^p \leq |A|^p + |B|^p$, so that we have $\|f + g\|_p^p \leq \|f\|_p^p + \|g\|_p^p$. In fact, for $p < 1$ we have

$$\frac{1}{N^{1/p}} \left\| \sum_{n=1}^N f_n \right\|_p \leq \left\| \frac{1}{N} \sum_{n=1}^N |f_n| \right\|_p \leq \frac{1}{N} \sum_{n=1}^N \|f_n\|_p$$

To prove this, since $1/p \geq 1$, we can apply convexity to conclude

$$\left(\frac{1}{N} \int \left(\sum_{n=1}^N |f_n| \right)^p \right)^{1/p} \leq \left(\frac{1}{N} \int \sum_{n=1}^N |f_n|^p \right)^{1/p} \leq \frac{1}{N} \sum_{n=1}^N \left(\int |f_n|^p \right)^{1/p}$$

This is 'almost' as good as a triangle inequality for the purposes of defining a geometric space. We call such a quantity a quasinorm.

The next most important inequality in the analysis of measurable functions is Hölder's inequality, which enables us to quantify the norm of the product of two functions in terms of the norm of the original two functions.

Theorem 6.17. *If $0 < p, q, r \leq \infty$, and $1/p + 1/q = 1/r$, then $\|fg\|_r \leq \|f\|_p \|g\|_q$.*

Proof. We may assume p, q , and r are finite, for the other cases are trivial. By normalization, we may assume $\|f\|_p = \|g\|_q = 1$. If $\theta = r/q$, $1 - \theta = r/p$, then we must show

$$\int |f|^{p(1-\theta)} |g|^{q\theta} \leq 1$$

But this follows by convexity of the map $\theta \mapsto A^{1-\theta} B^\theta$, for any A, B , we can apply a pointwise estimate, and integrate. \square

Remark. The symmetry of the inequality under the homogenities of functions explains why the two must occur the same number of times on both sides of the equation. The homogeneity under the scaling of the underlying measure is the reason why $1/p + 1/q = 1/r$.

Chapter 7

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 $\mathcal{S}(\mathbf{R}^d)$ denote a class of functions under which to study this phenomenon, it must be contained in $L^1(\mathbf{R}^d)$ and $C^\infty(\mathbf{R}^d)$, and also be closed under multiplication by polynomials. The differentiability and polynomial closure imply that the elements of $\mathcal{S}(\mathbf{R}^d)$ must have rapid decay properties: For any non-negative integer m and multi-index α , there exists a constant $C_{\alpha,\beta}$ such that

$$|f_\alpha(x)| \leq \frac{C_{\alpha,m}}{1 + |x|^m}.$$

We take this as a *definition* of the space $\mathcal{S}(\mathbf{R}^d)$. That is, for each non-negative integer n and m , we consider the seminorm

$$\|f\|_{n,m} = \sup_{|\beta| \leq n} \|(1 + |x|)^m f_\beta\|_{L^\infty(\mathbf{R}^d)}.$$

We then consider

$$\mathcal{S}(\mathbf{R}^d) = \left\{ f : \mathbf{R}^d \rightarrow \mathbf{R} : \text{for all } n, m, \|f\|_{n,m} < \infty \right\}.$$

Elements of $\mathcal{S}(\mathbf{R}^d)$ are known as *Schwartz functions*, and $\mathcal{S}(\mathbf{R}^d)$ is known as the *Schwartz space*. The seminorms naturally give $\mathcal{S}(\mathbf{R}^d)$ the structure of a Fréchet space. Sometimes, it is more convenient to use the equivalent family of seminorms $\|f\|_{\alpha,\beta} = \|x^\alpha f_\beta\|_{L^\infty(\mathbf{R}^n)}$, because x^α often behaves more

nicely under various operations. It is obvious that $\mathcal{S}(\mathbf{R}^d)$ is separated by the seminorms defined on it, because $\|\cdot\|_{L^\infty(\mathbf{R}^d)} = \|\cdot\|_{0,0}$ is a norm used to define the space. We now show the choice of seminorms make the space complete.

Theorem 7.1. $\mathcal{S}(\mathbf{R}^d)$ is a complete metric space.

Proof. Let $\{f_1, f_2, \dots\}$ be a Cauchy sequence with respect to the seminorms. This implies that for each integer m , and multi-index α , the sequence of functions $(1 + |x|)^m (f_k)_\alpha$ is Cauchy in $L^\infty(\mathbf{R}^d)$. Since $L^\infty(\mathbf{R}^d)$ is complete, there are functions $g_{m,\alpha}$ such that $(1 + |x|)^m (f_k)_\alpha$ converges uniformly to $g_{m,\alpha}$. If we set $f = g_{0,0}$, then it is easy to see using the basic real analysis of uniform continuity that f is infinitely differentiable, and $(1 + |x|)^m f_\alpha = g_{m,\alpha}$. This shows that $f \in C^\infty(\mathbf{R}^d)$. The sequence $\{f_k\}$ is bounded in $\mathcal{S}(\mathbf{R}^d)$, since it is Cauchy. And since $\|f_k - f\|_{n,m} \rightarrow 0$ for each n and m , this implies that $\|f\|_{n,m} < \infty$ for each n and m . Thus $f \in \mathcal{S}(\mathbf{R}^d)$, and $f_k \rightarrow f \in \mathcal{S}(\mathbf{R}^d)$. \square

Example. The Gaussian function $\phi : \mathbf{R}^d \rightarrow \mathbf{R}$ defined by $\phi(x) = e^{-|x|^2}$ is Schwartz. For any multi-index α , there is a polynomial P_α of degree at most $|\alpha|$ such that $\phi_\alpha = P_\alpha \phi$; this can be established by a simple induction. But this means that for each fixed α , $|P_\alpha(x)| \lesssim 1 + |x|^{|\alpha|}$. Since $e^{-|x|^2} \lesssim 1/(1 + |x|)^{m+|\alpha|}$ for any fixed m and α , we find

$$|(1 + |x|)^m \phi_\alpha| \leq (1 + |x|)^m |P_\alpha \phi| \lesssim \frac{1 + |x|^{|\alpha|+m}}{1 + |x|^{|\alpha|+m}} = 1.$$

Since m and α were arbitrary, this shows ϕ is Schwartz.

Example. The space $C_c^\infty(\mathbf{R}^d)$ consists of all compactly supported C^∞ functions. If $f \in C_c^\infty(\mathbf{R}^d)$, then f is Schwartz. This is because for each α and m , $(1 + |x|)^m f_\alpha$ is a continuous function vanishing outside a compact set, and is therefore bounded.

Because of the sharp control we have over functions in $\mathcal{S}(\mathbf{R}^d)$, almost every analytic operation we want to perform on $\mathcal{S}(\mathbf{R}^d)$ is continuous. To show that an operator T on $\mathcal{S}(\mathbf{R}^d)$ is bounded, it suffices to show that for each n and m , there is n', m' such that $\|Tf\|_{n,m} \lesssim_{n,m} \|f\|_{n',m'}$. For a functional $\Lambda : \mathcal{S}(\mathbf{R}^d) \rightarrow \mathbf{R}$, it suffices to show that there exists n and m such that $|\Lambda f| \lesssim \|f\|_{n,m}$. The minimal such choice of n is known as the **order** of

Λ . We normally do not care about the constant behind the operators for these norms, since the norms are not translation invariant and therefore highly sensitive to the positions of various operations. We really just care about proving the existence of such a constant.

Lemma 7.2. *If g is a function, with g and all its derivatives subpolynomial, then the map $f \mapsto gf$ is a bounded operator on $\mathcal{S}(\mathbf{R}^d)$.*

Proof. Fix n , and find values A and M , depending only on n and g , such that for any $|\beta| \leq n$,

$$|g_\beta(x)| \leq A \cdot (1 + |x|)^M.$$

Consider $|\alpha| \leq n$. Then the Leibnitz formula implies that

$$\begin{aligned} (1 + |x|)^m |(gf)_\alpha| &\leq 2^{|\alpha|} \sum_{\beta \leq \alpha} (1 + |x|)^m |g_\beta f_{\alpha-\beta}| \\ &\leq A \cdot 2^n \sum_{\beta \leq \alpha} (1 + |x|)^{m+M} |f_{\alpha-\beta}| \\ &\leq A \cdot 4^n \|f\|_{n,m+M}. \end{aligned}$$

Thus $\|gf\|_{n,m} \lesssim_{n,m} \|f\|_{n,m+M}$, which implies the operator is bounded. \square

If f and g are Schwartz functions, and $|\alpha| \leq n$, then the Leibnitz formula again implies that if $k + k' = m$, then

$$(1 + |x|)^m (gf)_\alpha \lesssim_n \sum_{\beta \leq \alpha} (1 + |x|)^m g_\beta f_{\alpha-\beta} \lesssim_n \|f\|_{n,k} \|g\|_{n,k'}$$

Thus $\|gf\|_{n,m} \lesssim_{n,m} \|f\|_{n,k} \|g\|_{n,k'}$, so $(f, g) \mapsto fg$ is a continuous bilinear operator on $\mathcal{S}(\mathbf{R}^d) \times \mathcal{S}(\mathbf{R}^d)$. Most importantly, we have shown the product of two Schwartz functions is Schwartz.

Theorem 7.3. *The following sublinear operators are all bounded on $\mathcal{S}(\mathbf{R}^n)$.*

- For each $h \in \mathbf{R}^n$, the translation operator $(T_h f)(x) = f(x - h)$.
- For each $\xi \in \mathbf{R}^n$, the modulation operator $(M_\xi f)(x) = e(i\xi \cdot x) f(x)$.
- The L^p norms $\|f\|_{L^p(\mathbf{R}^n)}$, for $1 \leq p \leq \infty$.
- The Fourier transform.

Furthermore, the Fourier transform is an isomorphism of $\mathcal{S}(\mathbf{R}^n)$.

Proof. We leave all but the last point as exercises. Here it will be convenient to use the norms $\|\cdot\|_{\alpha,\beta}$ as well as the norms $\|\cdot\|_{n,m}$. If $|\alpha| \leq m$, $|\beta| \leq n$, then we can use the Leibnitz formula to conclude that

$$\begin{aligned} |\xi^\alpha \mathcal{F}(f)_\beta| &\lesssim_{\alpha,\beta} \mathcal{F}((x^\beta f)_\alpha) \\ &\lesssim_{\alpha,\beta} \max_{\gamma \leq \alpha \wedge \beta} |\mathcal{F}(x^{\beta-\gamma} f_{\alpha-\gamma})| \\ &\lesssim_{\alpha,\beta} \max_{\gamma \leq \alpha \wedge \beta} \|x^{\beta-\gamma} f_\gamma\|_{L^1(\mathbf{R}^d)} \\ &\leq \max_{\gamma \leq \alpha \wedge \beta} \|(1+|x|)^{|\beta|} f_\gamma\|_{L^1(\mathbf{R}^d)} \lesssim \|f\|_{|\alpha|,|\beta|+d+1}. \end{aligned}$$

Thus \mathcal{F} is a bounded linear operator on $\mathcal{S}(\mathbf{R}^d)$. Since all Schwartz functions are arbitrarily smooth, the Fourier inversion formula applies to all Schwartz functions, and so \mathcal{F} is a bijective bounded linear operator with inverse \mathcal{F}^{-1} . The open mapping theorem then immediately implies that \mathcal{F}^{-1} is bounded. \square

Corollary 7.4. *If f and g are Schwartz, then $f * g$ is Schwartz.*

Proof. Since $f * g = \mathcal{F}^{-1}(\mathcal{F}(f)\mathcal{F}(g))$, this fact follows from the fact that the product of two Schwartz functions is Schwartz. \square

Now we get to the interesting part of the theory. We have defined a homeomorphic linear transform from $\mathcal{S}(\mathbf{R}^d)$ to itself. The theory of functional analysis then says that we can define a dual map, which is a homeomorphism from the dual space $\mathcal{S}(\mathbf{R}^d)^*$ to itself. Note the inclusion map $C_c^\infty(\mathbf{R}^d) \rightarrow \mathcal{S}(\mathbf{R}^d)$ is continuous, and $C_c^\infty(\mathbf{R}^d)$ is dense in $\mathcal{S}(\mathbf{R}^d)$. This implies that we have an injective, continuous map from $\mathcal{S}^*(\mathbf{R}^d)$ to $(C_c^\infty(\mathbf{R}^d))^*$, so every functional on the Schwarz space can be identified with a distribution. We call such distributions **tempered**. They are precisely the linear functionals on $C_c^\infty(\mathbf{R}^d)$ which have a continuous extension to $\mathcal{S}(\mathbf{R}^d)$. Intuitively, this corresponds to an asymptotic decay condition.

Example. For any $f \in L^1_{loc}(\mathbf{R}^d)$, we define f^* to be the distribution

$$f^*(\phi) = \int f(x)\phi(x) dx$$

But this distribution is not always tempered. If $f \in L^p(\mathbf{R}^d)$ for some p , then, applying Hölder's inequality, we obtain that

$$|f^*(\phi)| \leq \|f\|_{L^p(\mathbf{R}^d)} \|\phi\|_{L^q(\mathbf{R}^d)}.$$

Since $\|\cdot\|_{L^q(\mathbf{R}^d)}$ is a continuous norm on $\mathcal{S}(\mathbf{R}^d)$, this shows f^* is bounded. More generally, if $f \in L^1_{loc}(\mathbf{R}^d)$, and $f(x)(1+|x|)^{-m}$ is in $L^p(\mathbf{R}^d)$ for some m , then f^* is a tempered distribution. If $p = \infty$, such a function is known as **slowly increasing**.

Example. For any Radon measure, μ , we can define a distribution

$$\mu^*(\phi) = \int \phi(x) d\mu(x)$$

But this distribution is not always tempered. If $|\mu|$ is finite, the inequality $\|\mu^*(\phi)\| \leq \|\mu\| \|\phi\|_{L^\infty(\mathbf{R}^d)}$ gives boundedness. More generally, if μ is a measure such that $|\mu(x)|/(1+|x|^\alpha)$ is finite for some k , then μ is known as a **tempered measure**, and also acts as a tempered distribution, since

$$|\mu^*(\phi)| \leq \|\mu(x)/(1+|x|^\alpha)\| \|\phi\|_{L^\infty(\mathbf{R}^d, 1+|x|^\alpha)}.$$

Example. Suppose Λ is a distribution support on a compact set K . Then Λ is tempered, since if ψ is a compactly supported bump function on K , then for any Schwarz function ϕ we can define $\Lambda(\phi) = \Lambda(\psi\phi)$, which is continuous.

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**

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

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

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

Since ϕ has rapid decay, the first integral is well defined. Since ϕ is differentiable at the origin, the second integral is bounded for all $\varepsilon \geq 0$. But this means that

$$\lim_{\varepsilon \rightarrow 0} \int_{|x| \geq \varepsilon} \frac{\phi(x)}{x} = \int_{|x| \geq 1} \frac{\phi(x)}{x} + \int_{|x| \leq 1} \frac{\phi(x) - \phi(0)}{x}$$

Thus it is evident that the principal value exists, and

$$\left| p.v. \int_{-\infty}^{\infty} \frac{\phi(x)}{x} dx \right| \lesssim \|\phi\|_{L^1(\mathbf{R})} + \|\phi'\|_{L^\infty(\mathbf{R})}$$

so this functional is a tempered distribution of order 1, and is 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 Λ_α of a tempered distribution Λ is tempered, as is $\phi\Lambda$, whenever ϕ is a Schwartz function, or $f\Lambda$, where f is a polynomial.

To remind the reader, we think of a distribution Λ as corresponding to arbitrarily regular function 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

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

provides the perfect situation. It says that for $f \in L^1(\mathbf{R}^d)$, $\Lambda_{\mathcal{F}(f)}(g) = \Lambda_f(\mathcal{F}(g))$. If we want to generalize the Fourier transform to be defined on distributions, we better have $\mathcal{F}(\Lambda_f) = \Lambda_{\mathcal{F}(f)}$ for all integrable f . In particular, this motivates us to define the general Fourier transform of a tempered distribution Λ as $\mathcal{F}(\Lambda)(\phi) = \Lambda(\mathcal{F}(\phi))$. Similarly, we can define the inverse Fourier transform, which are the dual maps of the Fourier transforms and it's inverse, so they are obviously homeomorphisms of the space of tempered distributions.

Theorem 7.5. *If Λ is tempered,*

$$\mathcal{F}(\Lambda_\alpha) = (-2\pi i\xi)^\alpha \mathcal{F}(\Lambda) \quad \text{and} \quad \mathcal{F}((-2\pi i\xi)^\alpha \Lambda) = \mathcal{F}(\Lambda)_\alpha.$$

Example. *Consider the constant function 1. Then*

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

and so

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

so $\hat{1}$ is the Dirac delta distribution δ . Similarly,

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

so the Fourier transform of the Dirac delta function is the constant 1 function.

Example. We know $((-2\pi ix)^\alpha)^\wedge = ((-2\pi ix)^\alpha \cdot 1)^\wedge = \delta_\alpha$, which essentially provides us a way to compute the Fourier transform of any polynomial.

Theorem 7.6. If μ is a finite measure, $\hat{\mu}$ is a uniformly continuous bounded function with $\|\hat{\mu}\|_{L^\infty(\mathbf{R}^d)} \leq \|\mu\|$, and

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

The function $\hat{\mu}$ is also smooth if μ has moments of all orders, i.e. $\int |x|^k d\mu(x) < \infty$ for all $k > 0$.

Proof. Let $\phi \in \mathcal{S}(\mathbf{R}^d)$. We then calculate that

$$\hat{\mu} \cdot \phi$$

If f is integrable, we can apply Fubini's theorem to conclude

$$\begin{aligned} \int \hat{f}(x) d\mu(x) &= \int \int f(\xi) e(-\xi \cdot x) d\xi d\mu(x) \\ &= \int f(\xi) \int e(-\xi \cdot x) d\mu(x) d\xi \\ &= \int f(\xi) \hat{\mu}(\xi) d\xi \end{aligned}$$

This gives that the distribution $\hat{\mu}$ is given by integration with respect to the required function. It is easy to check that $\|\hat{\mu}\|_{L^\infty(\mathbf{R}^d)} \leq \|\mu\|$. Moreover, $\hat{\mu}$ is continuous, since by the dominated convergence theorem,

$$\hat{\mu}(\xi + \eta) - \hat{\mu}(\xi) = \int e(-\xi \cdot x) (e(-\eta \cdot x) - 1) d\mu(x)$$

as $h \rightarrow 0$, the values of the function in the integral converge pointwise to 0, and so the uniform continuity follows by the dominated convergence theorem.

To show $\hat{\mu}$ is smooth if it has all moments, we calculate that if $|\eta| = 1$, and $t \in \mathbf{R}$, then

$$\frac{\hat{\mu}(\xi + t\eta) - \hat{\mu}(\xi)}{t} = \int e(-\xi \cdot x)(e(-t\eta \cdot x) - 1) d\mu(x).$$

□

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 ,

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

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

Example. It is often useful to know the Fourier transform of the radial functions $f(x) = 1/|x|^\alpha$ on \mathbf{R}^d , for $\alpha < d$, so that the function is locally integrable. Since f satisfies a multiplicative symmetry $f(tx) = t^{-\alpha} f(x)$, the multiplicative Haar measure dt/t become very useful in the analysis of this function. We calculate that the multiplicative convolution of this character against a Gaussian gives, for $\alpha > 0$, by a change of variables,

$$\int_0^\infty t^\alpha e^{-\pi t^2 |x|^2} \frac{dt}{t} = \frac{1}{2\pi^{\alpha/2} |x|^\alpha} \int_0^\infty s^{\alpha/2} e^{-s} \frac{ds}{s} = \frac{\Gamma(\alpha/2)}{2\pi^{\alpha/2} |x|^\alpha}$$

Let $g(x)$ denote the left hand side of the equation. Then if ϕ is an arbitrary

Schwarz function, for $\alpha < d$, using Fubini's theorem,

$$\begin{aligned}
\int g(x) \phi^\vee(x) dx &= \int \int \int_0^\infty t^\alpha e^{-\pi t^2 |x|^2} \phi(\xi) e(\xi \cdot x) \frac{dt}{t} d\xi dx \\
&= \int \phi(\xi) \int_0^\infty t^\alpha \int e^{-\pi t^2 |x|^2} e(\xi \cdot x) dx \frac{dt}{t} d\xi \\
&= \int \phi(\xi) \int_0^\infty t^{\alpha-d} e^{-\pi |\xi|^2 / t^2} \frac{dt}{t} d\xi \\
&= \int \frac{\Gamma((d-\alpha)/2)}{2\pi^{(d-\alpha)/2} |\xi|^{d-\alpha}} \phi(\xi) d\xi
\end{aligned}$$

Thus, putting these two calculations together, we conclude

$$\frac{\Gamma(\alpha/2)}{2\pi^{\alpha/2}} (1/|x|^\alpha)^\wedge = \frac{\Gamma((d-\alpha)/2)}{2\pi^{(d-\alpha)/2} |\xi|^{d-\alpha}}$$

which can be simplified to

$$(1/|x|^\alpha)^\wedge = \frac{\Gamma((d-\alpha)/2)}{\Gamma(\alpha/2)} \pi^{\alpha-d/2} \frac{1}{|\xi|^{d-\alpha}}$$

Thus the Fourier transforms $|x|^{-\alpha}$ for $\alpha \in (0, d)$ map into themselves. For $\alpha \geq d$, we need to work with principal values TODO:

7.1 Convolution Operators

It is known that if $T : D(\mathbf{R}^d) \rightarrow C^\infty(\mathbf{R}^d)$ 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 $\mathcal{S}(\mathbf{R}^d)$ to $C^\infty(\mathbf{R}^d)$. 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, we rely on a regularity result on the differentiation of functions in L^p spaces.

Lemma 7.7. *If $f \in L^p(\mathbf{R}^d)$, has derivatives in the L^p norm of all orders $\leq d+1$, then f is almost everywhere equal to a continuous function g such that*

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

where the hidden constant depends only on n and p .

Proof. s □

Theorem 7.8. *If $T : L^p(\mathbf{R}^d) \rightarrow L^q(\mathbf{R}^d)$ is bounded, linear, and commutes with translations, then there exists a unique tempered distribution Λ such that $T(\phi) = \Lambda * \phi$ for all $\phi \in \mathcal{S}(\mathbf{R}^d)$*

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

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

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

Remark. It therefore follows that if $T : L^p(\mathbf{R}^d) \rightarrow L^q(\mathbf{R}^d)$ 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 its 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 $L^p(\mathbf{R}^d)$ to $L^q(\mathbf{R}^d)$, in the sense that the map $\phi \mapsto \Lambda * \phi$ is continuous as a map from $\mathcal{S}(\mathbf{R}^d)$ to $L^q(\mathbf{R}^d)$, 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 $L^\infty(\mathbf{R}^d)$. 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 $(\Lambda * \Phi)^\wedge = \Phi^\wedge \Lambda^\wedge$ is an element of $L^2(\mathbf{R}^d)$. Thus we can think of $\hat{\Lambda} = e^{\pi|x|^2}(\Lambda * \Phi)^\wedge$ as a function f . Plancherel's theorem implies that for any Schwarz function ϕ ,*

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

But this means that $f \in L^\infty(\mathbf{R}^d)$, for if there is a set E of positive measure where $|f| \geq M$, we can find $\hat{\phi}$ with $\hat{\phi} = 1$ on E and with $\|\hat{\phi}\|_2 = |E| + \varepsilon$, and then $\|f\hat{\phi}\|_2 \geq M\|\hat{\phi}\|_2$. Note that over $L^2(\mathbf{T})$, the only convolution operators are given by the distributions given by a Fourier series with bounded coefficients.

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 $\|\mu * \phi\|_1 \leq \|\mu\|_1 \|\phi\|_1$. 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 $\Lambda, \|\Lambda_\delta\|_1 \lesssim \|\Phi_\delta\|_1 = 1$. This implies that the Λ_δ are uniformly bounded in L^1 , so by the Banach Alaoglu theorem, since $L^1(\mathbf{R}^d)$ embeds itself in $M(\mathbf{R}^d)$, which is the dual of $C_0(\mathbf{R}^d)$, 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 $\mathcal{S}(\mathbf{R}^d)$, and so $\Lambda(\phi)$ is the limit of $\Lambda(\phi_\delta)$. But

$$\begin{aligned} \Lambda(\phi_\delta) &= \Lambda(\Phi_\delta * \phi) = (\Lambda * (\Phi_\delta * \phi)^*)(0) \\ &= ((\Lambda * \Phi_\delta) * \phi^*)(0) \\ &= \Lambda_\delta(\phi) \end{aligned}$$

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 $\mathcal{S}(\mathbf{R}^d)$ in $L^1(\mathbf{R}^d)$ 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 7.9. For any two (p, q) , $(L^p, L^q) = (L^{q^*}, L^{p^*})$.

Chapter 8

Riemann Theory of Trigonometric Series

Using the techniques of measure theory, we can actually prove that the Fourier series is essentially the unique way of representing a function on any part of its domain as a trigonometric series.

Lemma 8.1. *For any sequence u_n and set E of finite measure,*

$$\lim_{n \rightarrow \infty} \int_E \cos^2(nx + u_n) dx = |E|/2$$

Proof. We have

$$\cos^2(nx + u_n) = \frac{1 + \cos(2nx + 2u_n)}{2} = \frac{1}{2} + \frac{\cos(2nx)\cos(2u_n) - \sin(2nx)\sin(2u_n)}{2}$$

Since $\cos(2u_n)$ and $\sin(2u_n)$ are bounded, we have $\int \chi_E(x) \cos(2nx)$ and $\int \chi_E(x) \sin(2nx) \rightarrow 0$ as $n \rightarrow \infty$, and the same is true for the latter component of the sum since $\cos(2u_n)$ and $\sin(2u_n)$ are bounded, we conclude that

$$\int_E \cos^2(nx + u_n) = \int \chi_E(x) \cos^2(nx + u_n) = |E|/2$$

completing the proof. \square

Theorem 8.2 (Cantor-Lebesgue Theorem). *If, for some pair of sequences a_0, a_1, \dots and b_0, b_1, \dots are chosen such that*

$$\sum_{n=0}^{\infty} a_n \cos(2\pi nx) + b_n \sin(2\pi nx)$$

converges on a set of positive measure in $[0, 1]$, then $a_n, b_n \rightarrow 0$.

Proof. Let E be the set of points upon which the trigonometric series converges. We write $a_n \cos(2\pi nx) + b_n \sin(2\pi nx) = r_n \cos(nx + c_n)$. The result of the theorem is then precisely that $r_n \rightarrow 0$. If this is not true, then we must have $\cos(nx + c_n) \rightarrow 0$ for every $x \in E$. In particular, the dominated convergence theorem implies that

$$\lim_{n \rightarrow \infty} \int_E \cos(nx + c_n)^2 dx = 0$$

Yet we know this tends to $|E|/2$ as $n \rightarrow \infty$, which is a contradiction. \square

TODO: EXPAND ON THIS FACT.

8.1 Convergence in L^p and the Hilbert Transform

We now move onto a more 20th century viewpoint on Fourier series, namely, those to do with operator theory. Under this viewpoint, the properties of convergence are captured under the boundedness of certain operators on function spaces, allowing us to use the modern theory of functional analysis to it's full extent on our problems. However, unlike in most of basic functional analysis, where we assume all operators we encounter are bounded to begin with, in harmonic analysis we more often than not are given an operator defined only on a subset of spaces, and must prove the continuity of such an operator to show it is well defined on all of space. We will illustrate this concept through the theory of the circular Hilbert transform, and its relation to the norm convergence of Fourier series.

A **Fourier multiplier** is a linear transform T associated with a given sequence of scalars λ_n , for $n \in \mathbf{Z}$. It is defined for any trigonometric polynomial $f = \sum_{|n| \leq N} c_n e_n$ as $Tf = \sum_{|n| \leq N} \lambda_n c_n e_n$. The trigonometric polynomials are dense in $L^p(\mathbf{T})$, for each $p < \infty$. An important problem is determining whether T is therefore figuring out whether the operator can be extended to a *continuous operator* on the entirety of L^p . Because the trigonometric polynomials are dense in L^p , in the light of the Hahn Banach theorem it suffices to prove an inequality of the form $\|Tf\| \lesssim \|f\|$. Here are some examples of Fourier operators we have already seen.

Example. The truncation operator S_N is the transform associated with the scalars $\lambda_n = \mathbb{I}[|n| \leq N]$. The truncation is continuous, since for any integrable function f , the Fourier coefficients are uniformly bounded by $\|f\|_1$, so $\|S_N f\|_1 \leq N\|f\|_1$. Similarly, the F   er truncation σ_N associated to the multipliers $\lambda_N = \mathbb{I}[|n| \leq N](1 - |n|/N)$ is continuous on all integrable functions. These operators are easy to extend precisely because the nonzero multipliers have finite support.

Example. In the case of the Abel sum, A_r , associated with $\lambda_n = r^{|n|}$, A_r extends in a continuous way to all integrable functions, since

$$|A_r f| = \left| \sum r^{|n|} \widehat{f}(n) e_n(t) \right| \leq \|f\|_1 \sum r^{|n|} = \|f\|_1 \left(1 + \frac{2}{1-r} \right)$$

Thus the map is bounded.

To understand whether the truncations $S_N f$ of f converge to f in the L^p norms, rather than pointwise, we turn to the analysis of an operator which is the core of the divergence issue, known as the **Hilbert transform**. It is a Fourier multiplier operator H associated with the coefficients

$$\lambda_n = \frac{\operatorname{sgn}(n)}{i} = \begin{cases} +1/i & n > 0 \\ 0 & n = 0 \\ -1/i & n < 0 \end{cases}$$

Because

$$\mathbb{I}[|n| \leq N] = \frac{\operatorname{sgn}(n+N) - \operatorname{sgn}(n-N)}{2} + \frac{\mathbb{I}[n=N] + \mathbb{I}[n=-N]}{2}$$

we conclude

$$S_n f = \frac{i(e_{-n} H(e_n f) - e_n H(e_{-n} f))}{2} + \frac{\widehat{f}(n) e_n + \widehat{f}(-n) e_{-n}}{2}$$

Since the operators $f \mapsto \widehat{f}(n) e_n$ are bounded in all the L^p spaces since they are continuous in $L^1(\mathbb{T})$, we conclude that the operators S_n are uniformly bounded as endomorphisms on $L^p(\mathbb{T})$ provided that H is bounded as an operator from $L^p(\mathbb{T})$ to $L^q(\mathbb{T})$. Since $S_n f$ converges to f in L^p whenever f is a trigonometric polynomial, this would establish that $S_n f$ converges to

f in the L^p norm for any function f in $L^p(\mathbf{T})$. Later on, as a special case of the Hilbert transform on the real line, we will be able to prove that H is a bounded operator on $L^p(\mathbf{T})$ for all $1 < p < \infty$, and as a result, we find that $S_N f \rightarrow f$ in L^p for all such p . Unfortunately, H is not bounded from $L^1(\mathbf{T})$ to itself, and correspondingly, $S_N f$ does not necessarily converge to f in the L^1 norm for all integrable f .

For now, we explore some more ideas in how we can analyze the Hilbert transform via convolution, the dual of Fourier multipliers. The fact that $\widehat{f * g} = \widehat{f} \widehat{g}$ implies that if there is an integrable function g whose Fourier coefficients corresponds to the multipliers of an operator T , then $f * g = T f$ for any trigonometric polynomial f , and by the continuity of convolution, this is the unique extension of the Fourier multiplier operator. In the theory of distributions, one generalizes the family of objects one can take the Fourier series from integrable functions to a more general family of objects, such that every sequence of Fourier coefficients is the Fourier series of some *distribution*. One can take the convolution of any such distribution Λ with a C^∞ function f , and so one finds that $\Lambda * f = T f$ for any trigonometric polynomial f . There is a theorem saying that *all* continuous translation invariant operators from $L^p(\mathbf{T})$ to $L^q(\mathbf{T})$ are given by convolution with a Fourier multiplier operator. In practice, we just compute the convolution kernel which defines the Fourier multiplier, but it is certainly a satisfying reason to justify the study of Fourier multipliers. For instance, a natural question is to ask which Fourier multipliers result in bounded operations in space.

Theorem 8.3. *A Fourier multiplier is bounded from $L^2(\mathbf{T})$ to itself if and only if the coefficients are bounded.*

Proof. If a Fourier multiplier is given by λ_n , then for some trigonometric polynomial f ,

$$\|T f\|_2^2 = \sum |\widehat{T f}(n)|^2 = \sum |\lambda_n|^2 |\widehat{f}(n)|^2$$

If the λ_n are bounded, then we can obtain from this formula the bound

$$\|T f\|_2^2 \leq \max |\lambda_n| \|f\|_2^2$$

Conversely, if $T f$ is bounded, then

$$|\lambda_n|^2 = \|T(e_n)\|_2^2 \leq \|T\|^2$$

so the λ_n are bounded. □

Corollary 8.4. *The Hilbert transform is a bounded endomorphism on $L^2(\mathbf{T})$. Note that we already know that $S_N f \rightarrow f$ in the L^2 norm.*

The terms of the Hilbert transform cannot be considered the Fourier coefficients of any integrable function. Indeed, they don't vanish as $n \rightarrow \infty$. Nonetheless, we can use Abel summation to treat the Hilbert transform as convolution with an appropriate operator. For $0 < r < 1$, consider, for $z = e^{it}$,

$$K_r(z) = \sum_{n \in \mathbf{Z}} \frac{\text{sgn}(n)}{i} r^{|n|} z^n = K * P_r$$

Since we know the Hilbert transform is continuous in $L^2(\mathbf{T})$, we can conclude that, in particular, for any C^∞ function f ,

$$Hf = \lim_{r \rightarrow 1} K * (P_r * f) = \lim_{r \rightarrow 1} (K * P_r) * f = \lim_{r \rightarrow 1} K_r * f$$

So it suffices to determine the limit of the K_r . We find that

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{(rz)^n - (r\bar{z})^n}{i} &= \frac{r}{i} \left(\frac{1}{\bar{z} - r} - \frac{1}{z - r} \right) = \frac{r}{i} \frac{z - \bar{z}}{|z|^2 - 2r\text{Re}(z) + r^2} \\ &= \frac{2r \sin(t)}{1 - 2r \cos(t) + r^2} = \frac{4r \sin(t/2) \cos(t/2)}{(1-r)^2 + 4r \sin^2(t/2)} \\ &= \cot(t/2) + O\left(\frac{(1-r)^2}{t^3}\right) \end{aligned}$$

Thus $K_r(t)$ tends to $\cot(t/2)$ locally uniformly away from the origin. But

$$K_r(t) = \frac{4r \sin(t/2) \cos(t/2)}{(1-r)^2 + 4r \sin^2(t/2)} = O\left(\frac{t}{(1-r)^2}\right)$$

If f is any C^∞ function on \mathbf{T} , then

$$\left| \int_{|t| \geq \varepsilon} [K_r(t) - \cot(t/2)] f(t) dt \right| \lesssim (1-r)^2 \|f\|_\infty \int_{|t| \geq \varepsilon} \frac{dt}{|t|^3} \lesssim \frac{(1-r)^2 \|f\|_\infty}{\varepsilon^2}$$

$$\begin{aligned} \left| \int_{|t| < \varepsilon} K_r(t) f(t) dt \right| &\leq \int_0^\varepsilon |K_r(t)| |f(t) - f(-t)| dt \\ &\lesssim \int_0^\varepsilon |t K_r(t)| |f'(0)| dt \lesssim \frac{|f'(0)|}{(1-r)^2} \int_0^\varepsilon t^2 dt \lesssim \|f'\|_\infty \frac{\varepsilon^3}{(1-r)^2} \end{aligned}$$

$$\left| \int_{|t| < \varepsilon} \cot(t/2) f(t) dt \right| \lesssim \int_0^\varepsilon \frac{|f(t) - f(-t)|}{t} dt \lesssim \varepsilon f'(0)$$

Thus

$$\left| \int K_r(t) f(t) dt - \int \cot(t/2) f(t) dt \right| \lesssim \frac{(1-r)^2}{\varepsilon^2} \|f\|_\infty + \left(\frac{\varepsilon^3}{(1-r)^2} + \varepsilon \right) \|f'\|_\infty$$

Choosing $\varepsilon = (1-r)^\alpha$ for some $2/3 < \alpha < 1$ shows that for sufficiently smooth f ,

$$(Hf)(x) = \lim_{r \rightarrow 1} \int \cot(t/2) f(x-t) dt$$

8.2 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

$$f(t) \sim \sum_{n \neq 0} \frac{e_n(t)}{n}$$

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

$$\sum_{n=1}^{\infty} \frac{e_n(t)}{n} \quad \sum_{n=-\infty}^{-1} \frac{e_n(t)}{n}$$

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

8.3 Conjugate Fourier Series

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

$$\sum_{n=-\infty}^{\infty} \hat{f}(n) e_n(t) = \operatorname{Re} \left(\hat{f}(0) + 2 \sum_{n=1}^{\infty} \hat{f}(n) e_n(t) \right)$$

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

$$\operatorname{Im} \left(\hat{f}(0) + 2 \sum_{n=1}^{\infty} \hat{f}(n) e_n(t) \right) = \Re \left(-i \sum_{n=-\infty}^{\infty} \operatorname{sgn}(n) \hat{f}(n) e_n(t) \right)$$

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

Chapter 9

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.

9.1 Convex Interpolation

The most basic way to interpolate is using the notion of convexity. Given two inequalities $A_0 \leq B_0$ and $A_1 \leq B_1$, for any parameter $0 \leq \theta \leq 1$, if we define the additive weighted averages $A_\theta = (1 - \theta)A_0 + \theta A_1$ and $B_\theta = (1 - \theta)B_0 + \theta B_1$, then we conclude $A_\theta \leq B_\theta$ for all θ . Similarly, we can consider the weighted multiplicative averages $A_\theta = A_0^{1-\theta} A_1^\theta$ and $B_\theta = B_0^{1-\theta} B_1^\theta$, in which case we still have $A_\theta \leq B_\theta$. 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 $A_\theta \leq (1 - \theta)A_0 + \theta A_1$, and if B_θ is logarithmically convex, so that $\log B_\theta$ is convex, then $B_\theta \leq B_0^{1-\theta} B_1^\theta$. 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

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

says

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

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

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

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 $L^{p_0,\infty}(X) \cap L^{p_1,\infty}(X) \subset L^{p_\theta}(X)$.

9.2 Complex Interpolation

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 9.1 (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 $|f(it)| \leq B_0$ and $|f(1 + it)| \leq B_1$, then we have a general estimate $|f(s)| \leq B_0^{1-\theta} B_1^\theta$.*

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 $|f(it)|, |f(1 + it)| \leq 1$, then $|f(s)| \leq 1$ for all $s \in S$. If we assume first that $|f(s)| \rightarrow 0$ as $|t| \rightarrow \infty$, then, applying the maximum principle, we conclude that the maximum value on $\{s : |t| \leq T\}$ occurs for $\sigma \in \{0, 1\}$, which gives us the bound for $|t| \leq T$. Taking $T \rightarrow \infty$ 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 $f(s)g_\varepsilon(s)g_\varepsilon(1 - s) \rightarrow 0$ as $|t| \rightarrow \infty$, since

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

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

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

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

Remark. The function $e^{-ie^{\pi is}}$ 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 $|t| \rightarrow \infty$ and pointwise approximates the identity as $\varepsilon \rightarrow 0$.

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 $|g(s)| \leq |g(\sigma)|$, 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 $\|f\|_{p_\theta} \leq 1$ 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 .

9.3 Real Interpolation

The idea of real interpolation is to break functions into various parts, and then applying estimates we have obtained from endpoints to obtain an estimate on the original function. The decomposition normally depends on parameters which can be optimized to give a good result for an estimate of the function in question. Real interpolation is much more flexible than complex interpolation, but complex interpolation obtains cleaner bounds. The classic example of real interpolation is the Marcinkiewiz interpolation theorem.

Theorem 9.2. Let T be a sublinear operator defined on $L^{p_0}(X) + L^{p_1}(X)$, and suppose.

Proof. Let $f \in L^{p_\theta}(X)$ be given. If we fix a λ , then we can write $f = f \chi_E + f \chi_{E^c}$, where $E = \{x : |f(x)| \geq \lambda\}$. Then $|E| < \infty$, and so

$$\|f \chi_E\|_{p_0} \leq |E|^{1/p - 1/p_0} \|f\|_p$$

□

9.4 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) $T : X + Y \rightarrow Z$. 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 $T : X \rightarrow Y$ 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 9.3 (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*

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

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 $\|T\|_{p_\theta, q_\theta} \leq 1$. Hölder's inequality implies

$$\left| \int_Y (Tf)g \right| \leq \|f\|_{p_0} \|g\|_{q'_0} \|f\|_{p_1} \|g\|_{q'_1}$$

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

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

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

$$F(s) = \int_Y T(|f|^{(1-s)p_\theta/p_0 + sp_\theta/p_1} \operatorname{sgn}(f)) |g|^{(1-s)q'_\theta/q'_0 + sq'_\theta/q'_1} \operatorname{sgn}(g)$$

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 that $\|Tf\|_{q_\theta} \leq \|f\|_{p_\theta}$, which was what was required to be proven. \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

$$\begin{aligned} \|f * g\|_q &= \left(\int |(f * g)(x)|^q dx \right)^{1/q} \leq \int \left(\int |f(y)g(x-y)|^q dx \right)^{1/q} dy \\ &= \int |f(y)| \|g\|_q dy = \|f\|_1 \|g\|_q \end{aligned}$$

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 its 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

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

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

Chapter 10

Oscillatory Integrals

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

$$I(\lambda) = \int e(\lambda\Phi(x))\psi(x) dx$$

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 $\lambda \rightarrow \infty$ 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*

$$\hat{f}(\xi) = \int e(-\xi \cdot x)f(x) dx$$

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

$$\hat{\mu}(\xi) = \int e(-\xi \cdot x)d\mu(x) dx$$

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.

10.1 The Principle of Stationary Phase

10.1.1 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 10.1. *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 $D(e(\lambda\Phi)) = e(\lambda\Phi)$, so $D^N(e(\lambda\Phi)) = e(\lambda\Phi)$ for all integers N , and so

$$I(\lambda) = \int D^N(e(\lambda\Phi))\psi = \int e(\lambda\Phi)(D^*)^N(\psi)$$

Taking absolute values in the last integral gives that

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

which gives the required bounds. \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

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

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 $n \leq N$, $\Phi^{(n)}(a) = \Phi^{(n)}(b)$ 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.

10.2 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

$$\int e(\lambda\phi(x))\psi(x)$$

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 $|\phi^{(n)}(x)| \geq 1$ on (a, b) , then

$$\left| \int_a^b e(\lambda\phi(x)) dx \right| \leq C_n/|\lambda|^\alpha$$

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.

Theorem 10.2. *There exists a constant C_n such that if ϕ is smooth in (a, b) with $|\phi^{(n)}| \geq 1$ for all $x \in (a, b)$, then*

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

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

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

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

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

$$\begin{aligned} \left| \int_a^b e(\lambda \phi) D^*(1) dx \right| &\leq |\lambda|^{-1} \int_a^b |(1/\phi')'| dx \leq |\lambda|^{-1} \left| \int_a^b (1/\phi')' dx \right| \\ &= |\lambda|^{-1} \left| \frac{1}{\phi'(b)} - \frac{1}{\phi'(a)} \right| \leq |\lambda|^{-1} \end{aligned}$$

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) $\phi^{(n+1)}(x) \geq 1$ on $[a, b]$. Let x_0 be the point at which $|\phi^{(n)}(x)|$ is minimized. Without loss of generality, we may assume that $|\phi^{(n)}(x)|$ is increasing on $[x_0, b]$. **TODO: FINISH THIS ARGUMENT, GIVES $C_n = 5 \cdot 2^{n-1} - 2$.** \square

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

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 $\int_0^{2\pi N} \sin(\phi(x)) dx$ is unbounded as we let $N \rightarrow \infty$, 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

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

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

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

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

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

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

Corollary 10.3. *If an amplitude ψ is present, then*

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

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

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

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

Example. *The Bessel functions*

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

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 $\psi(x) = e_m(-x)/2\pi$. Split $[0, 2\pi]$ into two intervals, the first upon which $\cos(x) \geq 1/\sqrt{2}$, the second where $\sin(x) \geq 1/\sqrt{2}$. 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.

10.3 Surface Carried Measures

Theorem 10.4. *If a hypersurface Σ has non-vanishing Gauss curvature at each point in the support of a surface carried measure μ , then*

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

s

10.4 Restriction Theorems

If $f \in L^p(\mathbf{R}^n)$, then the Hausdorff Young theorem says that \hat{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 \hat{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 \hat{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

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

for Schwartz functions f . If this is true, we can apply a density argument to show that the restriction operator $R(f) = \hat{f}|_M$ 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

$$\begin{aligned} \int_M (Rf)(\xi) \overline{g(\xi)} d\mu(\xi) &= \int_M \left(\int_{\mathbf{R}^n} f(x) e(-\xi \cdot x) dx \right) \overline{g(\xi)} d\mu(\xi) \\ &= \int_{\mathbf{R}^n} f(x) \overline{\int_M g(\xi) e(\xi \cdot x) d\mu(\xi)} dx \end{aligned}$$

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

$$(R^*f)(x) = \int_M e(\xi \cdot x) f(\xi) d\mu(\xi)$$

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

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

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

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

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

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

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

We shall verify this using Stein's interpolation theorem. Consider the family of kernels k_s , where $k_s = \widetilde{K_s}$, 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 \mathbf{R} , we can translate and negate. On \mathbf{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 11

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}^ω of Torii, occurring in the study of Dirichlet series.*
- *The product \mathbf{Z}_2^ω , 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 $|p^{-m}q|_p = p^m$.*

11.1 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

$$x \mapsto gx \quad x \mapsto xg \quad x \mapsto x^{-1}$$

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

$$U^{-1} = \{x^{-1} : x \in U\} \quad UV = \{xy : x \in U, y \in V\}$$

We let $V^2 = VV$, $V^3 = VVV$, and so on.

Theorem 11.1. *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 $u_i v_i \rightarrow x$. Since U is compact, there is a subnet u_{i_k} converging to y . Then $y \in U$, and we find

$$v_{i_k} = u_{i_k}^{-1}(u_{i_k} v_{i_k}) \rightarrow y^{-1}x$$

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

Example. If U is merely closed, then (ii) need not hold. For instance, in \mathbf{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 \mathbf{R} , and is hence not closed.

There are useful ways we can construct neighbourhoods under the group operations, which we list below.

Lemma 11.2. *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 11.3. *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 11.4. *Every Kolmogorov topological group is Hausdorff.*

Theorem 11.5. *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 $\bigcap_V AV \subset \overline{A}$. Conversely, any neighbourhood contains a closed neighbourhood, so that $\overline{A} \subset AV$ for a fixed V , and hence $\overline{A} \subset \bigcap_V AV$. \square

Theorem 11.6. *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. \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, \dots 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.

11.2 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}/μ_{∞} of the torus modulo the roots of unity, where the quotient is lumpy.

Theorem 11.7. *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 . \square

Theorem 11.8. *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 $U_iH \subset V$. \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.

11.3 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 $f : \mathbf{R} \rightarrow \mathbf{R}$ 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 $f : G \rightarrow H$ 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 $f : G \rightarrow \mathbf{C}$, then left uniform continuity is equivalent to the fact that $\|L_x f - f\|_\infty \rightarrow 0$ as $x \rightarrow 1$, where $(L_x f)(y) = f(xy)$. Right uniform continuity requires $\|R_x f - f\|_\infty \rightarrow 0$,

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 $x_i y_i \rightarrow e$, $y_i x_i \rightarrow z \neq e$. 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 $f : G \rightarrow H$, 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 11.9. 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 $f : K \rightarrow H$ 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 $U'_{x_1}, \dots, U'_{x_n}$. Let $U = U'_{x_1} \cap \dots \cap U'_{x_n}$. 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. \square

Corollary 11.10. *All maps with compact support are uniformly continuous.*

Corollary 11.11. *Uniform continuity on compact groups is invariant of the uniform structure chosen.*

11.4 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 $b^2 \geq a$, then $b \geq ab^{-1}$ and so

$$(ab^{-1})^2 = ab^{-1}ab^{-1} \leq ab^{-1}b = a$$

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 $f_i \rightarrow f$ if eventually $f_i(s) = f(s)$ for all $s < w$ simultaneously, and $f_i(w) \rightarrow f(w)$. 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 $f_i \rightarrow f$. 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 $f_i \rightarrow f$ 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 $\leq \mathfrak{a}$ is closed.

The simple corollary to this proof is that compact subsets are finite. Let $X = f_1, f_2, \dots$ be a denumerable, compact set. Since all subsets of X are compact, we may assume $f_1 < f_2 < \dots$ (or $f_1 > f_2 > \dots$, 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 $(g, f_2), (f_1, f_3), (f_2, f_4), \dots$ 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

$$f_1 \leq f_2 \leq \dots < h < \dots \leq g_2 \leq g_1$$

Suppose that $f_i \leq k < h < k' \leq g_j$. 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 possibility is that $f_i \rightarrow h$ or $g_i \rightarrow h$, 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.

11.5 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 $g_i \rightarrow g$ 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 $g_i \rightarrow g$, 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,

$$\{-1, 1\} = (\mathbf{Z}p_1)^c \cap \dots \cap (\mathbf{Z}p_n)^c$$

and is therefore an open set. But this is clearly not the case as open sets must contain infinite sequences.

Chapter 12

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}^ω , 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 parallelliped 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

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix}$$

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 $\phi : C_c(G) \rightarrow \mathbf{R}$ 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 12.1. *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 $\int \phi \geq 1$, for which $f \leq \sum c_i L_{x_i}\phi$. 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 $f \leq \sum c_i L_{x_i}\phi$ for some $x_i \in G$. We would then have

$$\int f d\mu \leq (f : \phi) \int \phi d\mu$$

If k is fixed with $\int k = 1$, then we would have

$$\int f d\mu \leq (f : \phi)(\phi : k)$$

We cannot change k if we wish to provide a limiting result in ϕ , so we notice that $(f : g)(g : h) \leq (f : h)$, which allows us to write

$$\int f d\mu \leq \frac{(f : \phi)}{(k : \phi)}$$

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

$$(k : f)^{-1} \leq I_\phi(f) \leq (f : k)$$

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 $\text{supp}(f_1 + f_2)$, 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

$$|h_1(xy) - h_1(y)| \leq \varepsilon \quad |h_2(xy) - h_2(y)| < \varepsilon$$

If $\text{supp}(\phi) \subset V$, and $h \leq \sum c_i L_{x_i} \phi$, then

$$f_j(x) = h(x)h_j(x) \leq \sum c_i \phi(x_i x) h_j(x) \leq \sum c_i \phi(x_i x) [h_j(x_i^{-1}) + \varepsilon]$$

since we may assume that $x_i x \in \text{supp}(\phi) \subset V$. Then, because $h_1 + h_2 \leq 1$,

$$(f_1 : \phi) + (f_2 : \phi) \leq \sum c_j [h_1(x_j^{-1}) + \varepsilon] + \sum c_j [h_2(x_j^{-1}) + \varepsilon] \leq \sum c_j [1 + 2\varepsilon]$$

Now we find, by taking infimums, that

$$I_\phi(f_1) + I_\phi(f_2) \leq I_\phi(h)(1 + 2\varepsilon) \leq [I_\phi(f_1 + f_2) + \varepsilon I_\phi(g)][1 + 2\varepsilon]$$

Since g is fixed, and we have a bound $I_\phi(g) \leq (g : k)$, we may always find a neighbourhood V (dependant on f_1, f_2) for any $\varepsilon > 0$ such that

$$I_\phi(f_1) + I_\phi(f_2) \leq I_\phi(f_1 + f_2) + \varepsilon$$

if $\text{supp}(\phi) \subset V$.

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 $F : C_c^+(G) \rightarrow \mathbf{R}$ such that $(k : f)^{-1} \leq F(f) \leq (f : k)$. For each neighbourhood V of the identity, let $K(V)$ be the closure of the set of I_ϕ such that $\text{supp}(\phi) \subset V$. Then the set of all $K(V)$ has the finite intersection property, so we conclude there is some $I : C_c^+(G) \rightarrow \mathbf{R}$ contained in $\bigcap K(V)$. This means that every neighbourhood of I contains I_ϕ with $\text{supp}(\phi) \subset V$, 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 $\text{supp}(\phi) \subset V$, 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

$$|I(f_1 + f_2) - (I(f_1) + I(f_2))| \leq 2\varepsilon + |I_\phi(f_1 + f_2) - (I_\phi(f_1) + I_\phi(f_2))| < 3\varepsilon$$

Taking $\varepsilon \rightarrow 0$, 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. \square

We shall prove that the Haar measure is unique, but first we show an incredibly useful regularity property.

Proposition 12.2. *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 $x_1, \dots, x_n \in G$,

$$\mu\left(\bigcup_{i=1}^n x_i U\right) \leq \sum_{i=1}^n \mu(x_i U) = 0$$

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 12.3. *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

$$A = \text{supp}(f)V \cup V\text{supp}(f) \quad B = \text{supp}(g)V \cup V\text{supp}(g)$$

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 $\text{supp}(h) \subset W$ (take $h(x) = k(x)k(x^{-1})$ for some function $k \in C_c^+(G)$ with $\text{supp}(k) \subset W$, and $k = 1$ on a symmetric neighbourhood of the origin). Then

$$\begin{aligned} \left(\int h d\mu\right) \left(\int f d\lambda\right) &= \int h(y)f(x) d\mu(y) d\lambda(x) \\ &= \int h(y)f(yx) d\mu(y) d\lambda(x) \end{aligned}$$

and

$$\begin{aligned} \left(\int h d\lambda\right) \left(\int f d\mu\right) &= \int h(x)f(y) d\mu(y) d\lambda(x) \\ &= \int h(y^{-1}x)f(y) d\mu(y) d\lambda(x) \\ &= \int h(x^{-1}y)f(y) d\mu(y) d\lambda(x) \\ &= \int h(y)f(xy) d\mu(y) d\lambda(x) \end{aligned}$$

Hence, applying Fubini's theorem,

$$\begin{aligned} \left| \int h d\mu \int f d\lambda - \int h d\lambda \int f d\mu \right| &\leq \int h(y) |F_y(x)| d\mu(y) d\lambda(x) \\ &\leq \varepsilon \lambda(A) \int h d\mu \end{aligned}$$

In the same way, we find this is also true when f is swapped with g , and A with B . Dividing these 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 constant $c > 0$ such that $\int f d\lambda = c \int f d\mu$ for any function $f \in C_c^+(G)$, and we conclude that $\lambda = c\mu$. \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 ,

$$\mu(A) = \int_A f(y) d\nu(y) = f(x) \nu(A) = f(x) \mu(A) + f(x) \nu(A)$$

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 $\mu(A) = \nu(A)$.

12.1 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 $1 \leq p < \infty$, 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.

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

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. $f : X \rightarrow \mathbf{C}$ 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

$$\|f\|_\infty = \inf\{c : |f(x)| \leq c \text{ locally almost everywhere}\}$$

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 fg vanishes outside of a σ -finite set Y , so $fg \in L^1(X)$, and if we let $Y_1 \subset Y_2 \subset \dots \rightarrow Y$ be an increasing subsequence such that $\mu(Y_i) < \infty$, then $|f(x)| \leq \|f\|_\infty$ almost everywhere for $x \in Y_i$, and so by the monotone convergence theorem

$$\int |fg| d\mu = \lim_{Y_i \rightarrow \infty} \int_{Y_i} |fg| d\mu \leq \|f\|_\infty \int_{Y_i} |g| d\mu \leq \|f\|_\infty \|g\|_1$$

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 decomposability of the Carathéodory extension of μ , a fact we leave to the general measure theorists.

12.2 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 whose 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 $y_i \rightarrow e$, and $f \in C_c(G)$, then $\|R_{y_i}f - f\|_\infty \rightarrow 0$, so

$$\Delta(y_i^{-1}) \int f(x)d\mu = \int f(xy_i)d\mu \rightarrow \int f(x)d\mu$$

Hence $\Delta(y_i^{-1}) \rightarrow 1$. 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 12.4. *Any compact group is unimodular.*

Proof. $\Delta : G \rightarrow \mathbf{R}^*$ 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 12.5. *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 multiplication 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 $x \in U$, $|\Delta(x) - 1| \leq \varepsilon|c - 1|$. Then if $f > 0$

$$|c - 1|\mu(U) = |c\mu(U^{-1}) - \mu(U)| = \left| \int_U [\Delta(x^{-1}) - 1]d\mu(x) \right| \leq \varepsilon\mu(U)|c - 1|$$

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.

12.3 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

$$\begin{aligned} \left| \int \phi(xy)d\mu(x)d\nu(y) \right| &\leq \int_G \int_G \phi(xy)d|\mu|(x)d|\nu|(y) \\ &\leq \|\mu\| \|\nu\| \|\phi\|_\infty \end{aligned}$$

This defines an operation on $M(G)$ which is associative, since, by applying the associativity of G and Fubini's theorem.

$$\begin{aligned}
\int \phi d((\mu * \nu) * \lambda) &= \int \int \phi(xz) d(\mu * \nu)(x) d\lambda(z) \\
&= \int \int \int \phi((xy)z) d\mu(x) d\nu(y) d\lambda(z) \\
&= \int \int \int \phi(x(yz)) d\mu(x) d\nu(y) d\lambda(z) \\
&= \int \int \phi(xz) d\mu(x) d(\nu * \lambda)(z) \\
&= \int \phi d(\mu * (\nu * \lambda))
\end{aligned}$$

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 $\|\mu * \nu\| \leq \|\mu\| \|\nu\|$, 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

$$\|\mu\| \geq \sum_{x \in G} |\mu|(x)$$

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 $\mu_i \rightarrow \mu$, and $\mu_i \in M_c(G)$, and $\|\mu_i - \mu\| < \varepsilon$, then for any $x \in G$,

$$\varepsilon > \|\mu - \mu_i\| \geq |(\mu_i - \mu)(\{x\})| = |\mu(\{x\})|$$

Letting $\varepsilon \rightarrow 0$ 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 $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$. This is induced by the identification of f with $f(x)dx$, because then

$$\begin{aligned} \int \phi(f(x)dx * g(x)dx) &= \int \int \phi(yx)f(y)g(x)dydx \\ &= \int \phi(y) \left(\int f(y)g(y^{-1}x)dx \right) dy \end{aligned}$$

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 12.6. $L^1(G)$ and $M_c(G)$ are closed ideals in $M(G)$, and $M_d(G)$ is a closed subalgebra.

Proof. If $\mu_i \rightarrow \mu$, 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 $\mu_i \rightarrow \mu$, then $\mu_i(\{x\}) \rightarrow \mu(\{x\})$, 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

$$(f dx * \nu)(U) = \int \chi_U(xy) f(x) dx d\nu(y) = \int_G \int_{y^{-1}U} f(x) dx d\nu(y) = 0$$

$$(\nu * f dx)(U) = \int \chi_U(xy) d\nu(x) f(y) dy = \int_G \int_{Ux^{-1}} f(y) dy d\nu(x) = 0$$

So $L^1(G)$ is a two-sided ideal. \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 $\phi \in L^\infty(\mu)$,

$$\int (f * g)(x) \phi(x) dx = \int f(y) g(y^{-1}x) \phi(x) dx dy$$

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

$$(f * g)(zx) = \int f(y)g(y^{-1}zx)dy = \int f(zv)g(y^{-1}x)dy = [(L_z f) * g](x)$$

$$(f * g)(xz) = \int f(y)g(y^{-1}xz)dy = [f * (R_z g)](x)$$

Convolution can also be applied to the other L^p spaces, but we have to be a bit more careful with our integration.

Theorem 12.7. *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 $\|f * g\|_p \leq \|f\|_1 \|g\|_p$. 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

$$\begin{aligned} \|f * g\|_p &= \left(\int \left| \int f(y)g(y^{-1}x)dy \right|^p dx \right)^{1/p} \\ &\leq \int |f(y)| \left(\int |g(y^{-1}x)|^p dx \right)^{1/p} dy \\ &= \|f\|_1 \|g\|_p \end{aligned}$$

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

$$\begin{aligned} \|g * f\|_p &= \left(\int \left| \int \frac{g(xy^{-1})f(y)}{\Delta(y)} dy \right|^p dx \right)^{1/p} \\ &\leq \int \left(\int \left| \frac{g(xy^{-1})f(y)}{\Delta(y)} \right|^p dx \right)^{1/p} dy \\ &= \|g\|_p \int_K \frac{|f(y)|}{\Delta(y)} d\mu(y) \\ &\leq M \|g\|_p \|f\|_1 \end{aligned}$$

which shows that $g * f$ is defined almost everywhere. \square

Theorem 12.8. *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 $\|f * g\|_\infty \leq \|f\|_p \|g\|_q$.*

Proof. First, note that

$$\begin{aligned} |(f * g)(x)| &\leq \int |f(y)| |g(y^{-1}x)| dy \\ &\leq \|f\|_p \left(\int |g(y^{-1}x)|^q dy \right)^{1/q} \\ &= \|f\|_p \|g\|_q \end{aligned}$$

For each x and y , applying Hölder's inequality, we find

$$\begin{aligned} |(f * g)(x) - (f * g)(y)| &\leq \int |f(z)| |g(z^{-1}x) - g(z^{-1}y)| dz \\ &\leq \|f\|_p \left(\int |g(z^{-1}x) - g(z^{-1}y)|^q dz \right)^{1/q} \\ &= \|f\|_p \left(\int |g(z) - g(zx^{-1}y)|^q dz \right)^{1/q} \\ &= \|f\|_p \|g - R_{x^{-1}y}g\|_q \end{aligned}$$

Thus to prove continuity (and in fact uniform continuity), we need only prove that $\|g - R_x g\|_q \rightarrow 0$ for $q \neq \infty$ as $x \rightarrow \infty$ or $x \rightarrow 0$. This is the content of the next lemma. \square

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 12.9. *If $p \neq \infty$, then $\|g - R_x g\|_p \rightarrow 0$ and $\|g - L_x g\|_p \rightarrow 0$ as $x \rightarrow 0$.*

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

$$\|g - L_x g\|_p \leq \|g - g_i\|_p + \|g_i - L_x g_i\|_p + \|L_x(g_i - g)\|_p = 2\|g - g_i\|_p + \|g_i - L_x g_i\|_p$$

Taking i large enough, x small enough, we find $\|g - L_x g\|_p \rightarrow 0$. 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 $\|R_x(g - g_i)\|_p \rightarrow 0$, which effectively removes the problems in the proof. \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 12.10. *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

$$\|L_z(f * g) - (f * g)\|_\infty = \|(L_z f - f) * g\|_\infty \leq \|L_z f - f\|_1 \|g\|_\infty$$

$$\|R_z(g * f) - (g * f)\|_\infty = \|g * (R_z f - f)\|_\infty \leq \|g\|_\infty \|R_z f - f\|_1$$

and both integrals converge to zero as $z \rightarrow 1$. \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 12.11. *For each neighbourhood U of the origin, pick a function $f_U \in (L^1)^+(G)$, with $\int \phi_U = 1$, $\text{supp}(f_U) \subset U$. Then if g is any function in $L^p(G)$,*

$$\|f_U * g - g\|_p \rightarrow 0$$

*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 $\|g * f_U - g\|_p \rightarrow 0$, 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

$$\begin{aligned} \|f_U * g - g\|_p &= \left(\int \left| \int f_U(y) [g(y^{-1}x) - g(x)] dy \right|^p dx \right)^{1/p} \\ &\leq \int f_U(y) \left(\int |g(y^{-1}x) - g(x)|^p dx \right)^{1/p} dy \\ &\leq 2\mu(K)\varepsilon \int f_U(y) dy \leq 2\mu(K)\varepsilon \end{aligned}$$

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

$$|(f_U * g - g)(x)| = \left| \int f_U(y) [g(y^{-1}x) - g(x)] dy \right| \leq \varepsilon$$

For right convolution, we find that for $g \in C_c(G)$, where $|g(xy) - g(x)| < \varepsilon$ for $y \in U$, then

$$\begin{aligned} \|g * f_U - g\|_p &= \left(\int \left| \int g(y) f_U(y^{-1}x) - g(x) dy \right|^p dx \right)^{1/p} \\ &= \left(\int \left| \int [g(xy) - g(x)] f_U(y) dy \right|^p dx \right)^{1/p} \\ &\leq \int \left(\int |g(xy) - g(x)|^p dx \right)^{1/p} f_U(y) dy \\ &\leq \mu(K)\varepsilon \int f_U(y) (1 + \Delta(y)) dy \\ &= \mu(K)\varepsilon + \mu(K)\varepsilon \int f_U(y) \Delta(y) dy \end{aligned}$$

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

$$|(g * f_U - g)(x)| = \left| \int [g(xy) - g(x)] f_U(y) dy \right| \leq \varepsilon$$

We will always assume from hereon out that the approximate identities in $L^1(G)$ are of this form. \square

We have already obtained enough information to characterize the closed ideals of $L^1(G)$.

Theorem 12.12. *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 ,

$$\|(L_z f_U) * g - L_z g\|_1 = \|L_z(f_U * g - g)\|_1 = \|f_U * g - g\| \rightarrow 0$$

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 similarly. \square

12.4 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 12.13. *For any $0 < \theta < 1$, and $0 < p, q \leq \infty$. If we define*

$$1/r_\theta = (1 - \theta)/p + \theta/q$$

to be the inverse interpolation of the two numbers. Then

$$\|f\|_{r_\theta} \leq \|f\|_p^{1-\theta} \|f\|_q^\theta$$

Proof. We apply Hölder's inequality to find

$$\|f\|_{r_\theta} \leq \|f\|_{p/(1-\theta)} \|f\|_{q/\theta} = \left(\int |f|^{p/(1-\theta)} \right)^{(1-\theta)/p} \left(\int |f|^{q/\theta} \right)^{\theta/q}$$

so it suffices to prove $\|f\|_{p/(1-\theta)} \leq \|f\|_p^{1-\theta}, \|f\|_{q/\theta} \leq \|f\|_q^\theta$.

The map $x \mapsto x^p$ is concave for $0 < p < 1$, so we may apply Jensen's inequality in reverse to conclude

$$\left(\int |f|^{p/(1-\theta)} \right)^{(1-\theta)/p} \leq \left(\int |f|^p \right)^{1/p}$$

□

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.

12.5 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 12.14. *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 $\phi : G \rightarrow X$ 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 $y_1 V, y_2 V, \dots$, 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 $\phi(y V y_0^{-1}) \subset \phi(V y_0^{-1})$, so if we fix a compact U , and find V with $V^3 \subset U$, we have shown $\phi(U)$ is open in X . □

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 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 $\pi : G \rightarrow G/H$, and let Δ_G and Δ_H be the respective modular functions. Define a map $P : C_c(G) \rightarrow C_c(G/H)$ 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 $\text{supp}(Pf) \subset \pi(\text{supp}(f))$. 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 12.15. *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 12.16. *A compact $F \subset G/H$ gives rise to a function $f \geq 0$ 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

$$f = \frac{\phi \circ \pi}{Pg \circ \pi} g$$

Hence

$$Pf = \frac{\phi}{Pg}Pg = \phi$$

□

Lemma 12.17. *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 $f \geq 0$ if $\phi \geq 0$.*

Proof. There exists $g \geq 0$ 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. □

We can now provide conditions on the existence of a measure on G/H .

Theorem 12.18. *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 $y \in H$,

$$\begin{aligned} \Delta_G(y) \int f(x) d\mu(x) &= \int f(xy^{-1}) d\mu(x) \\ &= \int_{G/H} \int_H f(xzy^{-1}) d\nu(z) d\psi(xH) \\ &= \Delta_H(y) \int_{G/H} \int_H f(xz) d\nu(z) d\psi(xH) \\ &= \Delta_H(y) \int f(x) d\mu(x) \end{aligned}$$

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

$$0 = Pf(xH) = \int_H f(xy) d\nu(y) = \Delta_G(y^{-1}) \int_H f(xy^{-1}) d\nu(y)$$

so

$$\begin{aligned}
0 &= \int_G \int_H \Delta_G(y^{-1}) \phi(x) f(xy^{-1}) d\nu(y) d\mu(x) \\
&= \int_H \int_G \phi(xy) f(x) d\mu(x) d\nu(y) \\
&= \int_G P\phi(xH) f(x) d\mu(x) \\
&= \int_G f(x) d\mu(x)
\end{aligned}$$

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 .

12.6 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 13

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 $f : \mathbf{T} \rightarrow \mathbf{T}$ 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 $t \mapsto e(t\xi)$, for $\xi \in \mathbf{R}$. To see this, let $e : \mathbf{R} \rightarrow \mathbf{T}$ 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 $e(s \log x) = x^{is}$, 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 13.1. For any character $\phi : G \rightarrow \mathbf{C}$, 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

$$\begin{aligned} \varphi(f * g) &= \int \int \frac{f(y)g(y^{-1}x)}{\phi(x)} dy dx \\ &= \int \int \frac{f(y)g(x)}{\phi(y)\phi(x)} dy dx \\ &= \int \frac{f(y)}{\phi(y)} dy \int \frac{g(x)}{\phi(x)} dx \end{aligned}$$

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

$$\begin{aligned} \int \int f(y)g(x)\phi(yx)dydx &= \int \int f(y)g(y^{-1}x)\phi(x)dydx \\ &= \int f(x)\phi(x)dx \int g(y)\phi(y)dy \\ &= \int f(x)g(y)\phi(x)\phi(y)dx dy \end{aligned}$$

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

$$\begin{aligned} \int \varphi(f)g(y)\phi(y)dy &= \varphi(f * g) \\ &= \int \int g(y)f(y^{-1}x)\phi(x)dydx \\ &= \int \int (L_{y^{-1}}f)(x)g(y)\phi(x)dydx \\ &= \int \varphi(L_{y^{-1}}f)g(y)dy \end{aligned}$$

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 \mathbb{T} , for if $\|\phi(x)\| < 1$ for any particular x , $\|\phi(x^{-1})\| > 1$. \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 $\phi_i \rightarrow \phi$ if and only if

$$\int \frac{f(x)}{\phi_i(x)} dx \rightarrow \int \frac{f(x)}{\phi(x)} dx$$

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,

$$\widehat{f}(\phi)\phi(y) = \widehat{L_y f}(\phi)$$

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 13.2. *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

$$\widehat{1}(\phi) = \int \frac{dx}{\phi(x)}$$

And

$$\frac{1}{\phi(y)} \widehat{1}(\phi) = \int \frac{dx}{\phi(yx)} = \int \frac{dx}{\phi(x)} = \widehat{1}(\phi)$$

So either $\phi(y) = 1$ for all y , and it is then verified by calculation that $\widehat{1}(\phi) = 1$, or $\widehat{1}(\phi) = 0$. Since $\widehat{1}$ is continuous, the trivial character must be an open set by itself, and hence $\Gamma(G)$ is discrete. \square

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 its 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: $\|\widehat{f}\|_\infty \leq \|f\|_1$.
- 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

$$\widehat{\mu}(\phi) = \int \frac{d\mu}{\phi(x)}$$

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.

$$\left| \int \frac{d\mu(x)}{\phi(x)} \right| \leq \|\mu\|$$

It is uniformly continuous, because

$$(L_\nu \widehat{\mu} - \widehat{\mu})(\phi) = \int \frac{1 - \nu(x)}{\nu(x)\phi(x)} d\mu(x)$$

The regularity of μ implies that there is a compact set K such that $|\mu|(K^c) < \varepsilon$. If $v_i \rightarrow 0$, then eventually we must have $|v_i(x) - 1| < \varepsilon$ for all $x \in K$, and then

$$|(L_\nu \hat{\mu} - \hat{\mu})(\phi)| \leq 2|\mu|(K^c) + \varepsilon\|\mu\| \leq \varepsilon(2 + \|\mu\|)$$

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 $\mu_i \rightarrow \mu$ in the weak-* topology, then $\hat{\mu}_i \rightarrow \hat{\mu}$ pointwise, because

$$\int \frac{d\mu_i(x)}{\phi(x)} \rightarrow \int \frac{d\mu(x)}{\phi(x)}$$

(This makes sense, because weak-* convergence is essentially pointwise convergence in $M(G)$). Thus the Fourier-Stieltjes transform is continuous with respect to these topologies. It is the unique continuous extension of the Fourier transform, because

Theorem 13.3. $L^1(G)$ 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 $\mu_i \rightarrow \mu$, then $\mu_i * \nu \rightarrow \mu * \nu$, 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

$$\left| \int f(xy) d\mu_i(x) \right| \leq \|f\|_1 \|\mu_i\|$$

If i is taken large enough that □

If $\phi_\alpha \rightarrow \phi$, in the sense that $\phi_\alpha(x) \rightarrow \phi(x)$ 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

$$\int_K \frac{d\mu(x)}{\phi_\alpha(x)} \rightarrow \int_K \frac{d\mu(x)}{\phi(x)}$$

It is immediately verified to be a map into $L^1(\Gamma(G))$, because

$$\int \left| \int \frac{d\mu(x)}{\phi(x)} \right| d\phi \leq \int \|\mu\|$$

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 $\mathcal{G} : L^1(G) \rightarrow C_0(\Gamma(G))$ is the Gelfand transform, then the transform induces a map $\mathcal{G}^* : M(\Gamma(G)) \rightarrow L^\infty(G)$.

The duality in classical Fourier analysis is shown through the inversion formulas. That is, we have inversion functions

$$\mathcal{F}^{-1}(\{a_k\}) = \sum a_k e_k(t) \quad \mathcal{F}^{-1}(f)(x) = \int f(t) e(xt)$$

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 unimodular, then it is closed under conjugation. If we let $g(x) = \overline{f(x^{-1})}$, we find

$$\mathcal{F}(g)(\phi) = \int \frac{g(x)}{\phi(x)} dx = \overline{\int \frac{f(x^{-1})}{\phi(x^{-1})} dx} = \int \frac{f(x)}{\phi(x)} dx = \overline{\mathcal{F}(f)(\phi)}$$

so that by the Stone Weirstrass theorem $A(G)$ is dense in $C_0(\Gamma(L^1(G)))$.

Chapter 14

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. Similarly, we find the characters of $L^1(\mathbf{R})$ are the maps $f \mapsto \hat{f}(x)$, 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 15

Vector Spaces

If \mathbf{K} is a closed, multiplicative subgroup of the complex numbers, then \mathbf{K} is also a locally compact abelian group, and we can therefore understand \mathbf{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 \mathbf{K} is a subfield of the complex numbers, then \mathbf{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 $\mathbf{K} \times \mathbf{K} \rightarrow \mathbf{C}$, 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 $\mathbf{K} \rightarrow \mathbf{K}^*$ is injective.

Chapter 16

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

$$\|K\|_\infty = \sup_{f \in L^\infty(\mathbf{R}^n)} \frac{|\int f(y) d\mu(y)|}{\|f\|_\infty} = \|\mu\|_1$$

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 16.1. For any $1 \leq p \leq \infty$, and $q = p^*$, then $M_p = M_q$.

Proof. Let $f \in L^p$, and $g \in L^q$, then Hölder's inequality gives

$$|(K * f * g)(0)| \leq \|K * f\|_p \|g\|_q \leq \|K\|_p \|f\|_p \|g\|_p$$

Thus $K * g \in L_q$, and that $K \in M_q$ with $\|K\|_q \leq \|K\|_p$. By symmetry, we find $\|K\|_p = \|K\|_q$. \square

Example. Consider M_2 . If K is a distribution with $\|f * K\|_2 \leq A\|f\|_2$, then Parseval's inequality implies that

$$\|\hat{f}\hat{K}\|_2 = \|f * K\|_2 \leq A\|f\|_2 = A\|\hat{f}\|_2$$

so for each \hat{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 $\|K\|_p \leq \|K\|_{p_0}^{1-\theta} \|K\|_{p_1}^\theta$, when K lies in the three spaces. In particular, $\|K\|_p$ is a decreasing function of p for $1 \leq p \leq 2$, so we find $M_1 \subset M_p \subset M_q \subset M_2$ for $1 \leq p < q \leq 2$. 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 16.2. *Let $T : \mathbf{R}^n \rightarrow \mathbf{R}^m$ 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 □

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 16.3. *There exists a Schwartz function φ on \mathbf{R}^n which is supported on the annulus*

$$\{\xi : 1/2 \leq |\xi| \leq 2\}$$

is positive for $1/2 < |\xi| < 2$, and satisfies

$$\sum_{k=-\infty}^{\infty} \varphi(2^{-k}\xi) = 1$$

for all $\xi \neq 0$.

Lemma 16.4 (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, \dots with edges parallel to the coordinate axis and*

$$\sigma < \frac{1}{|I_n|} \int_{I_n} |f(x)| dx \leq 2^n \sigma$$

and with $|f(x)| \leq \sigma$ for almost all x outside these cubes.

Theorem 16.5 (The Mihlin Multiplier Theorem). *Let m be a bounded function on \mathbf{R}^n which is smooth except possibly at the origin, such that*

$$\sup_{\substack{\xi \in \mathbf{R}^n \\ |\alpha| \leq L}} |\xi|^{|\alpha|} |(D^\alpha m)(x)| < \infty$$

Then m is an L^p Fourier multiplier for $1 < p < \infty$.

16.1 Besov Spaces

Recall the Schwarz function φ used to prove the Mihlin multiplier theorem. We now define functions φ_k such that

$$\widehat{\varphi_n}(\xi) = \varphi(2^{-n}\xi) \quad \widehat{\psi}(\xi) = 1 - \sum_{n=1}^{\infty} \varphi(2^{-n}\xi)$$

Thus φ_n essentially covers the annulus $2^{n-1} \leq |\xi| \leq 2^{n+1}$, and the function ψ covers the remaining low frequency parts covered in the frequency ball of radius 2. We have

$$\varphi_n(\xi) = \widetilde{\varphi_{2^{-n}}}(\xi) = 2^{dn} \check{\varphi}(2^n \xi)$$

Given $s \in \mathbf{R}$, and $1 \leq p, q \leq \infty$, we write

$$\|f\|_{pq}^s = \|\psi * f\|_p + \left(\sum_{n=1}^{\infty} (2^{sn} \|\varphi_n * f\|_p)^q \right)^{1/q}$$

The convolution $\varphi_n * f$ essentially captures the portion of f whose frequencies lie in the annulus $2^{n-1} \leq |\xi| \leq 2^{n+1}$

16.2 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 $\nu * K_\rho$

Chapter 17

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 Erdős 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 17.1 (Van Der Waerden - 1927). *For any positive integers 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 pigeonhole, 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 17.2 (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 Erdős 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 progressions 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 $\{1, \dots, p^n\}$. This leads naturally to the cap set problem.

Theorem 17.3 (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 17.4 (Mészáros - 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 $c = 2.7551\dots$.

The idea of the polynomial method is to take combinatorial information about some set, encode it as some algebraic structural 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 $\dim V \geq \dim \mathcal{P}_d - |A|$. This gives a lower bound on the size of A , whereas we want an upper bound. To get an upper bound, we take $|A|^c$ instead, which shows

$$\dim V \geq \dim \mathcal{P}_d - |A| - 3^n$$

whichs gives $|A| \leq 3^n + \dim V - \dim \mathcal{P}_d$. 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

$$|A| \leq 3^n - \dim \mathcal{P}_d + \max |\text{supp}(P)|$$

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 $a_1, \dots, a_m \in A$, $P(-a_i - a_j) = 0$ 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 $x, y \in \mathbb{F}_3^n$,

Chapter 18

Appendix: Interpolation Theory