Harmonic Analysis

Jacob Denson

February 25, 2020

Table Of Contents

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 periodic functions representable as a trigonometric series

$$f(t) = A + \sum_{m=1}^{\infty} B_n \cos(2\pi m t) + C_n \sin(2\pi m t).$$

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 it's paired statement on the real line, that a function f on the real line can be written as

$$f(t) = \int_{-\infty}^{\infty} A(\xi)\cos(2\pi\xi t) + B(\xi)\sin(2\pi\xi t)\, d\xi.$$

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

Introduction

The classic oscillatory functions are the trigonometric ones, given by

$$f(t) = A\cos(st) + B\sin(st) = C\cos(st + \phi).$$

The value ϕ is the *phase* of the oscillation, C is the *amplitude*, and $s/2\pi$ is the *frequency* of the oscillation. Fourier analysis is the topic devoted to studying the representation of a function as an analytical combination of these functions. In the discrete, periodic setting, we fix a periodic function $f : \mathbf{R} \to \mathbf{C}$ (unless otherwise noted, by periodic we always mean 1 periodic, i.e. a function such that $f(x+1) = f(x)$ for all $x \in \mathbf{R}$), and try and find coefficients $\{A_m\}$, $\{B_m\}$, and C such that

$$f(t) \sim C + \sum_{m=1}^{\infty} A_m \cos(2\pi mt) + B_m \sin(2\pi mt).$$

In the continuous setting, we fix a function $f : \mathbf{R} \to \mathbf{C}$, trying to find values $A(s)$, $B(s)$, and C such that

$$f(t) \sim C + \int_0^{\infty} A(s)\cos(2\pi st) + B(s)\sin(2\pi st)\, ds.$$

The main contribution of Fourier was a method to formally find a reliable choice of $A(s)$ and $B(s)$, A_m and B_m, which represents f. This choice is given by the *Fourier transform* of f in the continuous case, and the *Fourier series* in the discrete case.

1.1 Obtaining the Fourier Coefficients

A *formal trigonometric series* is a formal sum of the form

$$C + \sum_{m=1}^{\infty} A_m \cos(2\pi m t) + B_m \sin(2\pi m t).$$

Our goal is to find $\{A_m\}$, $\{B_m\}$, and C which 'represents' a given function f. In particular, we say a periodic function f *admits a trigonometric expansion* if there is a series such that for each $t \in \mathbf{R}$,

$$f(t) = C + \sum_{m=1}^{\infty} A_m \cos(2\pi m t) + B_m \sin(2\pi m t).$$

It is a *very difficult question* to characterize which functions f admit a trigonometric expansion. Nonetheless, using the method of Fourier, we can associate a formal trigonometric series with every function, known as the *Fourier series*. Later, we will show all differentiable periodic functions exhibit a trigonometric expansion in this Fourier series.

1.2 Orthogonality

The key technique Fourier realized could be used to come up with a canonical trigonometric series for a function is *orthogonality*. Note that the various frequencies of sine functions are orthogonal to one another, in the sense that

$$\int_0^1 \sin(2\pi m t)\sin(2\pi n t) = \int_0^1 \cos(2\pi m t)\cos(2\pi n t) = \begin{cases} 0 & : m \neq n, \\ 1/2 & : m = n, \end{cases}$$

and for any $m, n \in \mathbf{Z}$,

$$\int_0^1 \sin(2\pi m t)\cos(2\pi n t) = 0.$$

This means that for a finite trigonometric sum

$$f(t) = C + \sum_{m=1}^{N} A_m \cos(2\pi m t) + B_m \sin(2\pi m t),$$

we have

$$C = \int_0^1 f(t)\,dt,$$

$$A_m = 2\int_0^1 f(t)\cos(2\pi mt)\,dt, \quad \text{and} \quad B_m = 2\int_{-\pi}^{\pi} f(t)\sin(2\pi mt)\,dt.$$

We note that these values may still be defined even if f is not a trigonometric polynomial. Thus given *any* periodic function f, a reasonable candidate for the coefficients is given by the values A_m, B_m, and C above.

There is an additional choice of oscillatory functions, which replaces the sine and cosine with a single family of trigonometric functions. For $\xi \in \mathbf{R}$, we let $e_\xi(t) = e^{2\pi\xi it}$. For each integer m, e_m is periodic with period 1. Applying orthogonality again, we find

$$\int_0^1 e_n(t)\overline{e_m(t)}\,dt = \int_0^1 e_{n-m}(t) = \begin{cases} 0 & : m \neq n, \\ 1 & : m = n. \end{cases}$$

Thus a natural choice of an expansion

$$f(t) \sim \sum_{m\in\mathbf{Z}} C_m e^{2\pi mit},$$

is given by

$$C_m = \int_0^1 f(t)\overline{e_m(t)}\,dt.$$

Euler's formula $e^{mit} = \cos(mt) + i\sin(mt)$ shows this is the same as the Fourier expansion in sines and cosines. Thus the values $\{A_m, B_m, C : m \geqslant 0\}$ can be recovered from the values of $\{C_m : m \in \mathbf{Z}\}$. Because of it's elegance, unifying the three families of coefficients, the expansion by complex exponentials is the most standard in modern Fourier analysis.

To summarize, we have shown a periodic function $f : \mathbf{R} \to \mathbf{C}$ gives rise to a formal trigonometric series

$$\sum_{m\in\mathbf{Z}} C_m e_m(t).$$

This is the *Fourier series* of f. Because we will be concentrating on the Fourier series of a function, it is worth reserving a particular notation for them. Given a periodic function f, and an integer $m \in \mathbf{Z}$, we set

$$\widehat{f}(m) = \int_0^1 f(t)\overline{e_m(t)}\, dt.$$

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 notation is just a simple way to represent sums of sines and cosines.

1.3 The Fourier Transform

For a general function $f : \mathbf{R} \to \mathbf{C}$, we cannot rely *just* on orthogonality, because the functions $\sin(2\pi m x)$ are not integrable on the entirety of $\mathbf{R}$, and therefore cannot be integrated against one another. Nonetheless, we can consider the functions $g_N : [0,1] \to \mathbf{C}$ by setting $g_N(s) = f(N(s-1/2))$. Then for $|t| \leqslant N/2$, we can apply the usual Fourier series to conclude

$$\begin{aligned} f(t) &= g_N(t/N + 1/2) \\ &\sim \sum_{m \in \mathbf{Z}} \widehat{g_N}(m) e^{2\pi m i (t/N + 1/2)} \\ &= \sum_{m \in \mathbf{Z}} (-1)^m \left(\int_0^1 f(N(s-1/2)) e^{-2\pi m i s}\, ds \right) e^{2\pi (m/N) i t} \\ &= \sum_{m \in \mathbf{Z}} \frac{1}{N} \left(\int_{-N/2}^{N/2} f(s) e^{-2\pi (m/N) i s}\, ds \right) e^{(m/N) i t}. \end{aligned}$$

If we take $N \to \infty$, the exterior sum operates like a Riemann sum, so we might expect

$$f(t) \sim \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(s) e^{-2\pi \xi i s}\, ds \right) e^{2\pi \xi i t}\, d\xi.$$

The interior integral defines the *Fourier transform* of the function f, given for each $\xi \in \mathbf{R}$ as

$$\widehat{f}(\xi) = \int_{-\infty}^{\infty} f(s) e^{-2\pi \xi i s}\, ds.$$

Thus the resultant *Fourier inversion formula* takes the form

$$f(t) \sim \int_{-\infty}^{\infty} \widehat{f}(\xi) e^{2\pi \xi i s}\, ds.$$

As the *limit* of a discrete series defined in terms of orthogonality, the Fourier transform possesses many of the same properties at the Fourier series. But the limit does cause technical issues which are not present in the case of Fourier series, and so we begin by concentrating on the Fourier series.

1.4 Basic Properties of Fourier Series

One of the most important properties of the Fourier series is that the coefficients are controlled by reasonable transformations. A basic, but unappreciated property of the Fourier transform is *linearity*: For any two functions f and g,

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

Linearity is *essential* to most methods in this book, and much remains unknown about nonlinear transforms. Fourier series are also stable under various other transformations which occur in analysis, which makes the transform useful. We summarize these properties here:

- For each periodic function f, define $f^*(x) = \overline{f(x)}$. Then for each $n \in \mathbf{Z}$

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

  As a corollary, we find that if f is real-valued, then $\widehat{f}(-n) = \overline{\widehat{f}(n)}$.

- Define the translation and frequency modulation operators, for each $s \in \mathbf{R}$, $m \in \mathbf{Z}$, and 2π periodic f, by setting

  $$(T_s f)(t) = f(t-s) \quad \text{and} \quad (M_m f)(s) = e_m(s) f(s).$$

  Similarily, for a function $C : \mathbf{Z} \to \mathbf{C}$, and for each $m \in \mathbf{Z}$ and $\xi \in \mathbf{R}$, define

  $$(T_m C)(n) = C(n-m) \quad \text{and} \quad (M_\xi C)(n) = e_\xi(n) C(n)$$

  Then $\widehat{T_s f} = M_{-s} \widehat{f}$ and $\widehat{M_m f} = T_m \widehat{f}$, for each $s \in \mathbf{R}$ and $m \in \mathbf{Z}$.

- If f is odd, then $\widehat{f}$ is odd, and if f is even, $\widehat{f}$ is even.
- If f has a derivative f', then $\widehat{f'}(n) = 2\pi in\widehat{f}(n)$.

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

$$\int_0^1 |f(x)|\, dx < \infty.$$

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

$$\|f\|_{L^1(\mathbf{T})} = \int_0^1 |f(x)|\, dx.$$

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 $L^1(\mathbf{T})$, i.e. for $f \in W^{1,1}(\mathbf{T})$.

Remark. We note that if f is even, then $\widehat{f}$ is even, so formally

$$f(t) \sim \widehat{f}(0) + \sum_{m=1}^{\infty} \widehat{f}(m)[e_m(t) + e_{-m}(t)] = \widehat{f}(0) + 2\sum_{m=1}^{\infty} \widehat{f}(m)\cos(mt).$$

Moreover,

$$\widehat{f}(m) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\cos(mt)\, dt$$

If f is an odd function, then the fact that $\widehat{f}$ is odd implies formally that

$$f(t) \sim \sum_{m=1}^{\infty} \widehat{f}(m)[e_m(t) - e_{-m}(t)] = 2i\sum_{m=1}^{\infty} \widehat{f}(m)\sin(mt).$$

Thus we get a sine expansion, and moreover,

$$\widehat{f}(m) = \frac{1}{2\pi i}\int_{-\pi}^{\pi} f(t)\sin(mt)\, dt.$$

This is one way to reduce the study of complex exponentials back to the study of sines and cosines, since every function can be written as a sum of an even and an odd function.

Chapter 2

Fourier Series

Let's focus in on the problem we introduced in the last chapter. We write $\mathbf{T} = \mathbf{R}/\mathbf{Z}$, so that a function $f : \mathbf{T} \to \mathbf{C}$ is a complex-valued periodic function on the real line. We then have a metric on $\mathbf{T}$ given by setting $d(t,s) = |t-s|$, where $|t| = \min_{n \in \mathbf{Z}} |t + 2\pi n|$. The Lebesgue measure on $\mathbf{R}$ induces a natural Borel measure on $\mathbf{T}$, such that for any periodic function $f : \mathbf{T} \to \mathbf{C}$,

$$\int_{\mathbf{T}} f(t)\, dt = \int_0^1 f(t)\, dt.$$

For each integrable $f : \mathbf{T} \to \mathbf{C}$, we then have a *formal trigonometric series*

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

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 of Fourier series to the function from which they are defined, we consider the partial sums

$$S_N f = \sum_{|n| \leqslant N} \widehat{f}(n) e_n.$$

If f is real-valued, then the complex parts of $\widehat{f}(n)$ and $\widehat{f}(-n)$ cancel out, so $S_N f$ is a real-valued sum of cosines and sines.

An initial hope is that the Fourier series of a function converges pointwise, i.e. that for every t, $\lim_{N \to \infty} S_N(f)(t) = f(t)$. This is true if f is periodic and differentiable everywhere. But even if we try and look at continuous periodic functions, then $S_N f$ can diverge. Thus we look for more exotic forms of convergence, and different quantitative descriptions which determine how the Fourier series represents the function f.

2.1 Unique Representation of a Function?

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

Theorem 2.1. *If $\widehat{f}(n) = 0$ for all $n \in \mathbf{Z}$, then f vanishes at all it's continuity points.*

Proof. We shall prove, without loss of generality, that if f is continuous at the origin, then $f(0) = 0$. We treat the real-valued case first. For every trigonometric polynomial $g(x) = \sum a_n e_n(-x)$, we have

$$\int_{\mathbf{T}} f(x) g(x) dx = \sum a_n \widehat{f}(n) = 0.$$

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

$$g(x) = \varepsilon + \cos(2\pi x) = \varepsilon + \frac{e_1(x) + e_1(-x)}{2}.$$

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

$$\left| \int_{\mathbf{T}} g_n(x) f(x) dx \right| \geqslant \int_{|x| < \delta} g_n(x) f(x) dx - \left| \int_{\delta \leqslant |x|} g_n(x) f(x) dx \right|.$$

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

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

On the other hand,

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

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

$$0 = \left| \int_0^1 g_n(x) f(x) dx \right| \gtrsim A^n - o(1).$$

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. □

Corollary 2.2. *If $f, g \in C(\mathbf{T})$ and $\widehat{f} = \widehat{g}$, then $f = g$.*

Proof. Then $f - g$ is continuous with vanishing Fourier coefficients. □

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

Proof. If $\sum |\widehat{f}(n)| < \infty$, then the triangle inequality implies

$$|(S_{N+M} f)(x) - (S_N f)(x)| \leqslant \sum_{N < |k| \leqslant M} |\widehat{f}(k)|.$$

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

$$\widehat{Sf}(n) = \lim_{N \to \infty} \int_{\mathbf{T}} S_N(f)(t) e_n(-t) = \widehat{f}(n).$$

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 $\widehat{f}(n) = O(1/|n|^m)$. In particular, if $m \geqslant 2$, then $S_N f \to f$ uniformly. Moreover, if $f \in C^\infty(\mathbf{T})$, then this shows the k'th derivatives $(S_N f)^{(k)}$ converge uniformly to $f^{(k)}$ for each k, and $\widehat{f}(m) = O_m(|n|^{-m})$ for each $m \geqslant 1$. Conversely, if $\{a_m\}$ is a sequence with $|a_m| = O_m(|n|^{-m})$ for each m, then the infinite sum $\sum a_m e^{mix}$ and all it's derivatives converge uniformly to an infinitely differentiable function $f \in C^\infty(\mathbf{T})$, and $\widehat{f}(n) = a_n$. Thus there is a perfect duality between infinitely differentiable functions and arbitrarily fast decaying sequences of integers. 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

In practical contexts, most functions we deal with are arbitrarily smooth, so the picture established in the last section seems rather complete. However, a deeper understanding of the Fourier series involves studying more quantitative questions. For instance, does the Fourier series of a function which is uniformly small converge faster than a function which is 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*. Similarily, how does the Fourier series of a function f change under a small pertubation in a particular norm space. 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 our arguments, initially invented to study smooth functions, also work naturally with arbitrarily integrable functions. On the other hand, the density of regular functions in these norm spaces indicates that the behaviour of Fourier summation on an infinitely differentiable space with respect to a norm is at least as bad as the behaviour of Fourier summation on an integrable function.

We note these quantitative problems are still interesting even if we knew everything there was to know about the pointwise convergence of Fourier series, because a series of functions may converge pointwise, whereas none of the individual functions may 'look' like the function they converge to. So we may want to look at quantitative measures of how globally similar two functions are, and this leads to norm space estimates.

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. Thus a piano with 88 keys can only approximate square waves slightly better than a keyboard with 20 keys. 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 $\|\widehat{f}\|_{L^q(\mathbf{Z})} \lesssim \|f\|_{L^p(\mathbf{T})}$ 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 pertubations in the L^p norm. The first inequality we give is trivial, but is certainly tight, e.g. for $f(t) = e_n(t)$.

Theorem 2.4. *For any f, $\|\widehat{f}\|_{L^\infty(\mathbf{T})} \leqslant \|f\|_{L^1(\mathbf{T})}$.*

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

$$|\widehat{f}(n)| = \left| \int_{\mathbf{T}} f(t) e_n(-t) \right| \leqslant \int_{\mathbf{T}} |f(t)| = \|f\|_{L^1(\mathbf{T})},$$

which was the required bound. $\square$

This proof doesn't really take any deep features of the Fourier coefficients. The same bound holds for any integral

$$\int_{\mathbf{T}} f(t)K(t)\,dt,$$

where $|K(t)| \leqslant 1$ for all t. But the bound is still tight, which might be explained by the fact that the Fourier series gives oscillatory information which is not immediately present in the L^p norms of the phase spaces, other than by taking a naive absolute bound into 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 technique; this should be expected since orthogonality was used to motivate the definition of the Fourier series.

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

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 $\widehat{f}(n) = (f, e_n)$, we apply Bessel's inequality to conclude

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

The exponentials $\{e_n\}$ are actually an orthonormal basis for $L^2(\mathbf{T})$; This can be seen from the Stone Weirstrass theorem, since trigonometric polynomials separate points, or by results we prove independently, later on in these notes. Thus Parsevel's equality tells us $\|\widehat{f}\|_{L^2(\mathbf{Z})} = \|f\|_{L^2(\mathbf{T})}$. □

This equality 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. *If $1 \leqslant p \leqslant 2$, $\|\widehat{f}\|_{L^q(\mathbf{Z})} \leqslant \|f\|_{L^p(\mathbf{T})}$.*

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

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

Thus we can control the magnitude of the Fourier coefficients in terms of the width of the original function, but we are limited in our ability to control the width of the Fourier coefficients in terms of the magnitudes 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 the Fourier transform of 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 $T_n : X \to Y$ 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. Suppose that 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 $x \in X$,

$$\|T_n(x) - T(x)\| \leqslant \|T_n(x) - T_n(x')\| + \|T_n(x') - T(x')\| + \|T(x') - T(x)\|.$$

If we choose x' such that $\|x - x'\| \leqslant \varepsilon$, then for n large enough we find that $\|T_n(x) - T(x)\| \lesssim \varepsilon$. Since ε was arbitrary, this means that $T_n(x) \to T(x)$ as $n \to \infty$. 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 $T(x)$. Here is a simple application.

Lemma 2.7 (Riemann-Lebesgue)**.** *If $f \in L^1(\mathbf{T})$, then $\widehat{f}(n) \to 0$ as $|n| \to \infty$.*

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

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

By linearity of the integral, the Fourier transform of any step function vanishes at ∞. But if $\Lambda_n(f) = \widehat{f}(n)$, then $|\Lambda_n f| \leqslant \|\widehat{f}\|_{L^\infty(\mathbf{T})} \leqslant \|f\|_{L^1(\mathbf{T})}$, which shows that the sequence of functionals $\{\Lambda_n\}$ are uniformly bounded as maps from $L^1(\mathbf{T})$ to $L^\infty(\mathbf{Z})$. Since $\lim_{|n|\to\infty} \Lambda_n(f) = 0$ for any step function f, and the step functions are dense in $L^1(\mathbf{T})$, we conclude that $\lim_{|n|\to\infty} \Lambda_n(f) = 0$ 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, for any sequence of non-negative numbers $\{\varepsilon_m\}$ with $\varepsilon_m \to 0$ as $|m| \to \infty$, there exists a continuous function f such that $|\widehat{f}(n)| \geqslant \varepsilon_n$ for infinitely many values n. One simply chooses a subsequence ε_{m_k} with $\sum \varepsilon_{m_k} < \infty$, and consider the trigonometric series attached to this subsequence. This is precisely the penalty for using a soft type analytical argument. Nonetheless, for smooth functions, we can obtain a uniform 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 $|\widehat{f}(n)| \leqslant |n|^{-m} \|f^{(m)}\|_{L^1(\mathbf{T})}$.*

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

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

If $0 < \alpha < 1$, we say a function f is *Hölder continuous* of order α if there exists a constant A such that $|f(x+h) - f(x)| \leqslant A|h|^\alpha$ for all $x, h \in \mathbf{T}$. We define

$$\|f\|_{C^{0,\alpha}(\mathbf{T})} = \sup_{x,h \in \mathbf{T}} \frac{|f(x+h) - f(x)|}{|h|^\alpha}$$

Then the space $C^{0,\alpha}(\mathbf{T})$ of all functions satisfying a Hölder condition of order α forms a Banach space.

Theorem 2.9. *If $f \in C^{0,\alpha}(\mathbf{T})$, then $|\widehat{f}(n)| \lesssim \|f\|_{C^{0,\alpha}(\mathbf{T})} |n|^{-\alpha}$.*

Proof. We calculate that by periodicity,

$$\widehat{f}(n) = -\int_{\mathbf{T}} f(x + 1/n) e_n(-x)\, dx,$$

so

$$\widehat{f}(n) = \frac{1}{2} \int_{\mathbf{T}} [f(x) - f(x+1/n)] e_n(-x)\, dx.$$

Thus taking in absolute values and applying Hölder continuity gives

$$|\widehat{f}(n)| \leqslant \frac{\|f\|_{C^{0,\alpha}(\mathbf{T})}}{2|n|^\alpha}. \qquad \square$$

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

$$\widehat{\mu}(n) = \int_0^1 e_n(-x) d\mu(x).$$

If μ is absolutely continuous with respect to the normalized Lebesgue measure on $\mathbf{T}$, and $d\mu = f dx$, then $\widehat{\mu} = \widehat{f}$, so this is an extension of the Fourier series from integrable functions to measures with finite variation. One can verify that

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

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,

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

Thus the Fourier series of δ has no decay at all. Once 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

$$(f * g)(t) = \int_{\mathbf{T}} f(s) g(t-s)\, ds.$$

Thus we smear the values of g with respect to a density function f.

Lemma 2.10. *For any $1 \leqslant p < \infty$, and $f \in L^p(\mathbf{T})$, $\lim_{h \to 0} T_h f = f$ 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 $\|T_h f - f\|_{L^p(\mathbf{T})} \leqslant \|T_h f - f\|_{L^\infty(\mathbf{T})} \lesssim h$, and so $T_h f \to f$ 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 $1 \leqslant p < \infty$, we conclude that $\lim_{h\to 0} T_h f = f$ for all $f \in L^p(\mathbf{T})$. ☐

Theorem 2.11. *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 $1 \leqslant r \leqslant \infty$, then $f * g$ is well-defined by the integral formula almost everywhere, and*

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

  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

$$\begin{aligned} |(f * g)(t-h) - (f * g)(t)| &\leqslant \int_{\mathbf{T}} |f(t-h-s) - f(t-s)||g(s)|\, ds \\ &\leqslant \|f_h - f\|_{L^p(\mathbf{T})} \|g\|_{L^q(\mathbf{T})}. \end{aligned}$$

The right hand side is a bound independant of t and converges to zero as $h \to 0$, so $f * g$ is uniformly continuous. Applying Hölder's inequality again gives that $\|f * g\|_{L^\infty(\mathbf{T})} \leqslant \|f\|_{L^p(\mathbf{T})} \|g\|_{L^q(\mathbf{T})}$. If $f \in L^p(\mathbf{T})$, and $g \in L^1(\mathbf{T})$, we use Minkowski's inequality to conclude that

$$\begin{aligned}
\|f * g\|_{L^p(\mathbf{T})} &= \left(\int_{\mathbf{T}} \left| \int_{\mathbf{T}} f(t-s) g(s)\; ds \right|^p\; dt \right)^{1/p} \\
&\leqslant \int_{\mathbf{T}} \left(\int_{\mathbf{T}} |f(t-s) g(s)|^p\; dt \right)^{1/p}\; ds \\
&= \int_{\mathbf{T}} g(s) \|f\|_{L^p(\mathbf{T})}\; ds = \|f\|_{L^p(\mathbf{T})} \|g\|_{L^1(\mathbf{T})}.
\end{aligned}$$

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

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

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 $f, g \in L^1(\mathbf{T})$. Similarily, one can apply Fubini's theorem to obtain associativity for $f, g, h \in L^1(\mathbf{T})$. To obtain the product identity for the Fourier series, we can apply Fubini's theorem to write

$$\begin{aligned}
\widehat{f * g}(n) &= \int_{\mathbf{T}} (f * g)(t) e_n(-t)\; dt \\
&= \int_{\mathbf{T}} \int_{\mathbf{T}} f(s) g(t-s) e_n(-t)\; ds\; dt \\
&= \int_{\mathbf{T}} f(s) \int_{\mathbf{T}} (L_{-s} g)(t) e_n(-t)\; dt\; ds \\
&= \int_{\mathbf{T}} f(s) e_n(-s) \widehat{g}(n)\; ds \\
&= \widehat{f}(n) \widehat{g}(n),
\end{aligned}$$

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

$$\begin{aligned}
\int_{\mathbf{T}} (f' * g)(t)\phi(t)\, dt &= \int_{\mathbf{T}} \int_{\mathbf{T}} f'(t-s)g(s)\phi(t)\, ds\, dt \\
&= \int_{\mathbf{T}} g(s) \int_{\mathbf{T}} f'(t-s)\phi(t)\, dt\, ds \\
&= -\int_{\mathbf{T}} g(s) \int_{\mathbf{T}} f(t-s)\phi'(t)\, dt\, ds \\
&= -\int_{\mathbf{T}} \left(\int_{\mathbf{T}} g(s)f(t-s)\, ds \right) \phi'(t)\, dt \\
&= -\int_{\mathbf{T}} (f * g)(t)\phi'(t)\, dt.
\end{aligned}$$

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

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 $\sum a_n \widehat{f}(n)e_n$ might converge for a suitably fast decaying sequence a_n. But if a_n is close to one, this sequence might still reflect properties of the original sequence.

To make rigorous the idea of approximating the Fourier series of a function, we introduce families of *good kernels*. A good kernel is a sequence of integrable functions $\{K_n\}$ on $\mathbf{T}$ bounded in L^1 norm, for which

$$\int_{\mathbf{T}} K_n(t) = 1.$$

so that integration against K_n operates essentially like an average, and for any $\delta > 0$,

$$\lim_{n \to \infty} \int_{|t| > \delta} |K_n(t)| \to 0.$$

Thus the functions $\{K_n\}$ become concentrated at the origin as $n \to \infty$. If in addition, we have an estimate $\|K_n\|_{L^\infty(\mathbf{T})} \lesssim n$, we say it is an **approximation to the identity**.

Theorem 2.12. *Let $\{K_n\}$ be a good kernel. Then*

- $(K_n * f)(t) \to f(t)$ *for any continuity point t of f.*
- $(K_n * f) \to f$ *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 $1 \leqslant p < \infty$.*
- *If K_n is an approximation to the identity, $(K_n * f)(t) \to f(t)$ for all t 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 $(K_n * f)(t) \to f(t)$ at each point t 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 $K_n * f \to f$ in $L^p(\mathbf{T})$ for each $f \in L^p(\mathbf{T})$, for $1 \leqslant p < \infty$. To obtain pointwise convergence, we calculate

$$|(K_n * f)(t) - f(t)| \leqslant \int_{\mathbf{T}} |f(t-s) - f(t)||K_n(s)|\, ds.$$

Let $A(\delta) = \delta^{-1} \int_{|s|<\delta} |f(t-s) - f(t)|$. Then as $\delta \to 0$, $A(\delta) \to 0$ because t is in the Lebesgue set of f. And we find that for each k, since $|K_n(s)| \lesssim n$,

$$\int_{2^k/n < |t| < 2^{k+1}/n} |f(t-s) - f(t)||K_n(s)| \lesssim \frac{A(2^{k+1}/n)}{2^{k+1}}.$$

Thus we have a bound

$$|(K_n * f)(t) - f(t)| \lesssim \sum_{k=0}^{\infty} \frac{A(2^k/n)}{2^k}.$$

Because f is integrable, A is continuous, and hence bounded. This means that for each m,

$$|(K_n * f)(t) - f(t)| \lesssim \sum_{k=0}^{m} \frac{A(2^k/n)}{2^k} + \|A\|_\infty \sum_{k=m}^{\infty} \frac{1}{2^k} = \sum_{k=0}^{m} \frac{A(2^k/n)}{2^k} + O(1/2^m).$$

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

2.5 The Dirichlet Kernel

We calculate that

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

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 convolution with this kernel gives results about the sums of Fourier series.

Theorem 2.13. *For any integer N and $t \in \mathbf{R}$,*

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

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

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

Thus as $N \to \infty$, 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

$\int D_N(t) = 1$. However, we find

$$\begin{aligned}\int_0^1 |D_N(t)| &= \int_0^1 \left|\frac{\sin(2\pi(N+1/2)t)}{\sin(\pi t)}\right|\\ &\gtrsim \int_0^1 \frac{|\sin(2\pi(N+1/2)t)|}{\sin(\pi t)}\\ &\gtrsim \int_0^1 \frac{|\sin(2\pi(N+1/2)t)|}{t}\,dt\\ &= \int_0^{2\pi N+\pi} \frac{|\sin(t)|}{t}\\ &\gtrsim \sum_{n=0}^{N} \frac{1}{t} \gtrsim \log(N).\end{aligned}$$

Thus 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 2.14. *There exists $f \in C(\mathbf{T})$ such that $(S_N f)(0)$ diverges as $N \to \infty$.*

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 $\|\Lambda_N\| \to \infty$ as $N \to \infty$. 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 $N \to \infty$. $\square$

The situation is even worse than this for general integrable functions. In 1927, Andrey Kolmogorov constructed an integrable function whose Fourier series diverges everywhere. But there is some hope. In 1928, Marcel Riesz showed, using methods we will develop in these notes, that if $1 < p < \infty$, and $f \in L^p(\mathbf{R})$, that $S_N f$ converges in the L^p norm to f, by showing the Hilbert transform was bounded from $L^p(\mathbf{T})$ to $L^p(\mathbf{T})$. And after a half century of the development of techniques in harmonic analysis, in 1966, Carleson proved that for each $f \in L^p(\mathbf{T})$, for $p > 1$, the Fourier series of f converges almost everywhere to f.

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

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

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* $1-1+1-1+\dots = 1/2$, *which reflects the fact that the partials sums do 'converge', but to two different numbers* 0 *and* 1, *which the series oscillates between, and the Cesaro means average these two points of convergence out to give a single method of convergence.*

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

Example. *In the Abel sense, we have* $1 - 2 + 3 - 4 + 5 - \dots = 1/4$, *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 $f \in L^1(\mathbf{T})$,*

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

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

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

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

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

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 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

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

Thus, if we define the *Poisson kernel*

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

which is defined by a uniformly convergent series over $\mathbf{T}$, we calculate that $A_r(f) = P_r * f$. Thankfully, we find P_r is a good kernel. To see this, we can apply an infinite geometric series summation to obtain that

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

As $r \to 1$, the function concentrates at the origin, because as $r \to 1$, if $\delta \leqslant |t| \leqslant \pi$, then $1 - \cos 2\pi t$ is bounded away from the origin, so

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

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

Theorem 2.17. *For any $f \in L^1(\mathbf{T})$,*

- *$(A_r f)(t) \to f(t)$ for all x in the Lebesgue set of f.*
- *$A_r f \to f$ uniformly if $f \in C(\mathbf{T})$.*
- *$A_r f \to f$ in the L^p norm for $1 \leqslant p < \infty$, if $f \in L^p(\mathbf{T})$.*

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

$$\frac{1+re^{2\pi it}}{1-re^{2\pi it}},$$

and therefore links the study of trigonometric series and the theory of analytic functions.

Theorem 2.18. *$u(re^{2\pi it}) = (A_r f)(t)$ is $C^\infty(\mathbf{T})$, and harmonic for $r < 1$. Moreover, the function u is the* unique *$C^2(\mathbf{T})$ harmonic function such that as $r \to 1$, $u(re^{2\pi it}) \to f$ in the L^1 norm.*

Proof. The function u is infinitely differentiable, because of the rapid convergence of the series defining the Poisson kernel. In particular, we note that

$$\frac{\partial^2 u}{\partial \theta^2} = -\sum_{n=-\infty}^{\infty} \widehat{f}(n)|n|^2 r^{|n|} e_n(t),$$

$$\frac{\partial u}{\partial r} = \sum_{n=-\infty}^{\infty} \widehat{f}(n)|n| r^{|n|-1} e_n(t),$$

and

$$\frac{\partial^2 u}{\partial r^2} = \sum_{n=-\infty}^{\infty} \widehat{f}(n)|n|(|n|-1) r^{|n|-2} e_n(t).$$

But in polar coordinates, we have

$$\begin{aligned}
\Delta u &= \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{u}{\theta^2} \\
&= \sum_{|n|\geqslant 2} \widehat{f}(n)|n|(|n|-1) r^{|n|-2} e_n(t) \\
&\quad + \sum_{n\neq 0} \widehat{f}(n)|n| r^{|n|-2} e_n(t) - \sum_{n=-\infty}^{\infty} \widehat{f}(n)|n|^2 r^{|n|-2} e_n(t) = 0.
\end{aligned}$$

Thus u is harmonic.

Conversely, suppose $u \in C^2(\mathbf{T})$ is harmonic. Then we can find $a_n(t)$ such that

$$u(re^{it}) = \sum_{n=-\infty}^{\infty} a_n(r) e_n(t),$$

where

$$a_n(r) = \int_{\mathbf{T}} u(re^{it}) e_n(-t)\, dt.$$

Then

$$\int_{\mathbf{T}} \frac{\partial^2 u}{\partial \theta^2}(re^{it}) e_n(-t)\, dt = -n^2 a_n(r),$$

and so $a_n''(r) + (1/r)a_n'(r) - (n^2/r^2)a_n(r) = 0$. This is an ordinary differential equation, whose only bounded solutions are given by $a_n(r) = A_n r^{|n|}$. If $u(re^{it}) \to f$ in the L^1 norm as $r \to 1$, then we conclude

$$A_n = \lim_{r \to 1} \int_{\mathbf{T}} u(re^{it}) e_n(-t)\, dt = \int_{\mathbf{T}} f(t) e_n(-t) = \widehat{f}(n),$$

so

$$u(re^{it}) = \sum \widehat{f}(n) r^{|n|} e_n(t) = g(re^{it}). \qquad \square$$

Thus we can represent any function on $\mathbf{T}$ as a harmonic function on the interior of the unit disk.

2.9 The De la Valleé 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 Velleé 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 2.19. *If $\widehat{f}(n) = O(|n|^{-1})$, then for any integers N and k, if*

$$kN \leqslant M < (k+1)N,$$

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 $S_M f$,

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

Corollary 2.20. *If f is a function with $\widehat{f}(n) = O(|n|^{-1})$,*

- *$S_N f$ converges to f in the L^p norm for $1 \leqslant p < \infty$.*
- *$S_N f$ converges uniformly to f if $f \in C(\mathbf{T})$.*
- *$(S_N f)(x) \to f(x)$ 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 $kN \leqslant M < k(N+1)$,

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

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

2.10 Pointwise Convergence

One way around around the blowup in the L^1 norm of D_N is to consider only functions f which provide a suitable dampening condition on the oscillation of D_N near the origin. This is provided by smoothness of f, manifested in various ways. The first thing we note is that the convergence of $(S_N f)(t)$ for a *fixed* x_0 depends only *locally* on the function f.

Lemma 2.21 (Riemann Localization Principle)**.** *If f_0 and f_1 agree in an interval around t_0, then*

$$(S_N f_0)(t_0) = (S_N f_1)(t_0) + o(1).$$

Proof. Let

$$X = \{f \in L^1(\mathbf{T}) : f(x) = 0 \text{ for almost every } x \in (t_0 - \varepsilon, t_0 + \varepsilon)\}.$$

Then X is a closed subset of $L^1(\mathbf{T})$. Note that for all $x \in [-\pi, \pi]$,

$$\sin(t/2) \gtrsim t \quad \text{and} \quad \sin((N+1/2)t) \leqslant 1.$$

Thus if $|t| \geqslant \varepsilon$,

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

In particular, by Hölder's inequality, the functionals $T_N f = (S_N f)(t_0)$ are uniformly bounded on X, i.e. $\|T_N\| \lesssim 1/\varepsilon$. If f is smooth, and vanishes on $(t_0 - \varepsilon, t_0 + \varepsilon)$, then $T_N f \to 0$ as $N \to \infty$. But the space of such functions is dense in X, which implies that $T_N f \to 0$ for *any* $f \in X$. Thus if f_0, f_1 are two functions that agree in $(t_0 - \varepsilon, t_0 + \varepsilon)$, then $f_0 - f_1 \in X$, so $(S_N f_0)(t_0) = (S_N f_1)(t_0) + o(1)$. In particular, the pointwise convergence properties of f_0 and f_1 are equivalent at t_0. $\square$

Thus any result about the pointwise convergence of Fourier series must depend on the local properties of a function f. Here, we give two of the main criteria, which corresponds to the smoothness of a function about a point x: either f is in a sense, 'locally Lipschitz', or 'locally of bounded variation'.

Theorem 2.22 (Dini's Criterion)**.** *If there exists δ such that*

$$\int_{|t|<\delta} \left| \frac{f(x+t) - f(x)}{t} \right| dt < \infty,$$

then $(S_N f)(x) \to f(x)$.

Proof. Assume without loss of generality that $x = 0$ and $f(x) = 0$. Fix $\varepsilon > 0$, and pick δ_0 such that

$$\int_{|t|<\delta_0} \left| \frac{f(t)}{t} \right| dt < \varepsilon.$$

We have

$$|(S_N f)(0)| = \left| \left(\int_{|t|<\delta_0} + \int_{|t|\geqslant\delta_0} \right) f(t) D_N(t)\, dt \right|.$$

Now

$$\int_{|t|\geqslant\delta_0} f(t) D_N(t)\, dt = (D_N * \left(\mathbf{I}_{|t|\geqslant\delta_0} f\right))(0) = S_N(\mathbf{I}_{|t|\geqslant\delta_0} f)(0) = o(1)$$

since $f\mathbf{I}_{|t|\geqslant\delta_0}$ vanishes in a neighbourhood of the origin. On the other hand, we note that $t/\sin(\pi t)$ is a bounded function on $\mathbf{T}$, so

$$\begin{aligned} \int_{|t|<\delta_0} f(t) D_N(t)\, dt &= \int_{|t|<\delta_0} \left(\sin(2\pi(N+1/2)t) \frac{f(t)}{t} \right) \left(\frac{t}{\sin(\pi t)} \right) dt \\ &\lesssim \|f(t)/t\|_{L^1[-\delta_0,\delta_0]} \leqslant \varepsilon. \end{aligned}$$

Thus, for suitably large N, $|(S_N f)(0)| \lesssim \varepsilon$. Since ε was arbitrary, the proof is complete. $\square$

This proof applies, in particular, if f is locally Lipschitz at x. Note the application of the Riemann Lebesgue lemma to show that to analyze the pointwise convergence of $(S_N f)(x)$, it suffices to analyze

$$\lim_{N\to\infty} \int_{|t|<\delta} f(x+t) D_N(t)\, dt$$

for any fixed $\delta > 0$.

Lemma 2.23 (Jordan's Criterion)**.** *If $f \in L^1(\mathbf{T})$ locally has bounded variation about x, then*

$$(S_N f)(x) \to \frac{f(x^+) + f(x^-)}{2}.$$

Proof. By Riemann's localization principle, we may assume f has bounded variation everywhere. Then without loss of generality, we may assume f is an increasing function, since a bounded variation function is the difference of two monotonic functions. Since

$$\int_{-1/2}^{1/2} D_N(t)\,dt = \int_0^{1/2} [f(x+t)+f(x-t)]D_N(t),$$

it suffices without loss of generality to show that

$$\lim_{N\to\infty} \int_0^{1/2} f(x+t)D_N(t)\,dt = \frac{f(x+)}{2}.$$

Since $\int_0^{1/2} D_N(t) = 1/2$, this is equivalent to showing

$$\lim_{N\to\infty} \frac{1}{2\pi} \int_0^{\pi} [f(x+t)-f(x+)]D_N(t)\,dt = 0.$$

Because of this, we may assume without loss of generality that $x = 0$ and $f(x+) = 0$. Then by the mean value theorem for integrals (which only applies for monotonic functions), for each N, there exists $0 \leqslant \nu_N \leqslant 1/2$ such that

$$\int_0^{1/2} f(t)D_N(t)\,dt = \|f\|_\infty \int_{\nu_N}^{1/2} D_N(t)\,dt.$$

Now an integration by parts gives

$$\int_{\nu_N}^{1/2} D_N(t) \lesssim \int_{\nu_N}^{1/2} \frac{\sin((N+1/2)t)}{t}\,dt = \int_{\nu_N/(N+1/2)}^{1/2(N+1/2)} \frac{\sin(t)}{t}\,dt \lesssim \frac{1}{N+1/2}.$$

Thus

$$\int_0^{1/2} f(t)D_N(t) \lesssim \frac{1}{N+1/2} \to 0. \qquad \square$$

Remark. The calculations in this proof also show that if $f \in L^1(\mathbf{T})$ has bounded variation, then

$$\widehat{f}(n) = O(1/|n|).$$

We have seen that this implies $S_N f$ converges to f at every point on the Lebesgue set of f, $S_N f$ converges uniformly to f if $f \in C(\mathbf{T})$, and for any

$1 \leqslant p < \infty$, if $f \in L^p(\mathbf{T})$, $S_N f$ converges to f in $L^p(\mathbf{T})$. Dirichlet's theorem says that the Fourier series of a continuous function f with only finitely many maxima and minima converges uniformly to f everywhere. Such a function has bounded variation, and so Dirichlet's theorem is an easy consequence of our discussion.

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.

Example. *Consider the sawtooth function defined on $[-1/2, 1/2)$ by $s(t) = t$, and then made periodic on the entire real line. We can easily calculate the Fourier series here, obtaining that*

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

Thus for any $t \in (-1/2, 1/2)$,

$$\sum_{n=1}^{\infty} \frac{(-1)^n \sin(2\pi n t)}{n} = -t/2.$$

Theorem 2.24. *If $\widehat{f}(n) = O(|n|^{-1})$, and $f(t_0-)$ and $f(t_0+)$ exist, then*

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

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 $t_0 = 1/2$. Define $g(t) = f(t) + (f(1+) - f(1-))s(t)/2$ on $(-1/2, 1/2)$, where s is the sawtooth function. Then

$$\lim_{t \uparrow 1/2} g(t) = \lim_{t \downarrow -1/2} g(t) = \frac{f(1/2+) + f(1/2-)}{2}.$$

Thus g can be defined on $\mathbf{T}$ so it is continuous at t_0. Now we find $|\widehat{g}| \lesssim |\widehat{f}| + |\widehat{s}| = O(|n|^{-1})$, and so

$$(S_N g)(1/2) \to \frac{f(1/2+) + f(1/2-)}{2}.$$

We also have $(S_N s)(1/2) \to 0$. Thus

$$(S_N f)(1/2) = (S_N g)(1/2) - (S_N s)(1/2) \to \frac{f(1/2+) + f(1/2-)}{2}. \qquad \square$$

2.11 Pointwise Behaviour at Discontinuity Points

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, $S_N f \to f$ 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 overshot 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 occured in physical measurements where first thought to be a defect in the equipment used to take the measurements. Gibb's phenomenon is one instance where a series of functions $\{f_k\}$ converges pointwise to some function f, whereas qualitatively with respect to the L^∞ norm, $\{f_k\}$ does not converge to f.

Theorem 2.25. *Given f with finitely many discontinuity points and with $\widehat{f}(n) = O(|n|^{-1})$, in particular one at t_0, we find*

$$\lim_{N\to\infty} (S_N f)(t_0 \pm 1/N) = f(t_0\pm) \pm C \cdot \frac{f(t_0+) - f(t_0-)}{2},$$

where

$$C = 2\pi \int_0^\pi \frac{\sin x}{x} \approx 16.610.$$

Proof. First consider the jump function s, with $t_0 = 1/2$. Then

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

Here we're just taking averages of values of $\sin(x)/x$ at $x = 2\pi/N$, $x = 4\pi/N$, and so on and so forth up to $x = 2\pi$. Thus is a Riemann sum, so as $N \to \infty$, we get that

$$(S_N s)(\pi + 1/N) \to -2 \int_0^{2\pi} \frac{\sin x}{x}.$$

The same calculations give

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

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 $S_N h_j \to 0$ for all h_j uniformly in an interval outside of their discontinuity point. To see this, we note that an integration by parts gives

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

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

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

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. $\square$

Chapter 3

Applications of Fourier Series

3.1 Tchebychev Polynomials

If f is everywhere continuous, then for every ε, Fejér's theorem says that we can find N such that $\|\sigma_N(f) - f\| \leqslant \varepsilon$. But $\sigma_N f$ 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 3.1 (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 $\cos t \geqslant 0$, 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. $\square$

3.2 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 $I \subset \mathbf{T}$, $\#\{2\pi a_n \in I : n \leqslant N\} \sim N|I|$ as $N \to \infty$. By approximating continuous functions by step functions, this implies that if $f : \mathbf{T} \to \mathbf{C}$ is continuous, then

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

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

Proof. The condition in the theorem implies that for any trigonometric polynomial f, $S_N f \to Tf$. 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 $n \to \infty$. Since a_n is an integer, this means that $d(b_n, \mathbf{Z}) = d(\gamma^n, \mathbf{Z}) \to 0$. But this means that the average distribution of the γ^n modulo one is concentrated at the origin.

3.3 The Isoperimetric Inequality

TODO

3.4 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 means we have

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

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

Thus the temperature in the ground splits into a daily heating effect H_D, a seasonal heating effect H_Y, and a constant temperature H_C. 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 neglible, and only the seasonal temperature is important. Experimently, 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 which is cool during the summer months.

3.5 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 subtitute 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 height with respect to the rotation of the earth and the sun, and so on and so forth to more neglible 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 without hard work on the part of the navigator.

Theorem 3.3. *If $h(t) = \sum_{n=1}^N A_n \cos(\omega_n t)$, where $\omega_1, \dots, \omega_n$ are distinct, then for any S,*

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

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

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

We calculate that

$$\frac{1}{N}\sum_{n=1}^{N} h(S + t/\omega_n + 2\pi k/\omega_n) = A_n\cos(t) + O(1),$$

and so

$$\frac{2}{T}\int_S^{S+T} h(t)\cos(\omega_n t)\,dt = A_n + O(1/T).$$

and we then take $T \to \infty$. $\square$

Chapter 4

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

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

is the Fourier inversion formula

$$f(x) = \int_{-\infty}^{\infty} \widehat{f}(\xi) e(2\pi \xi x)\, d\xi,$$

where for each real number ξ, we define

$$\widehat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e(-2\pi \xi x)\, dx.$$

The function $\widehat{f}$ 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 $f : \mathbf{R}^n \to \mathbf{R}^n$, 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 $e(t) = \exp(2\pi i t)$ for any $t \in \mathbf{R}$. The basic theory is the same, though as n increases the transform certainly becomes more complicated. In particular, the issue of pointwise convergence becomes more difficult to understand.

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

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

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

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 $\widehat{\chi_I}(\xi) \to 0$ as $|\xi| \to \infty$. 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, current research cannot give a satisfactory description of the elements of $\mathbf{A}(\mathbf{R}^d)$.

The next lemma will be used to show $\mathbf{A}(\mathbf{R}^d) \neq C_0(\mathbf{R}^d)$.

Lemma 4.2. *For any $0 \leqslant a < b < \infty$, independantly of a and b,*

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

Proof. Since $\|\sin(x)/x\|_{L^\infty(\mathbf{R})} \leqslant 1$, we may assume $b > 1$, for otherwise we obtain a trivial bound. This also implies

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

An integration by parts then shows that

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

Theorem 4.3. $\mathbf{A}(\mathbf{R}) \neq C_0(\mathbf{R})$. *In particular,* $\mathbf{A}(\mathbf{R})$ *does not contain any odd functions g in $C_0(\mathbf{R})$ such that*

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

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

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

If $b \geqslant 1$, an application of Fubini's theorem shows that

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

But

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

Thus we obtain that

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

For instance, this implies that there is no $f \in L^1(\mathbf{R})$ such that

$$\widehat{f}(\xi) = \left| \frac{\sin(2\pi\xi)}{\log|\xi|} \right|$$

for all $\xi \in \mathbf{R}$. □

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

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

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

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

  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 $T : \mathbf{R}^d \to \mathbf{R}^d$ 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. Similarily, 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 is actually differentiable almost everywhere with derivative f_k. 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 $\delta_h f$ 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 the classical derivative may be a strong derivative. In particular, this is true of a Schwartz function, i.e. a function lying in the space

$$\mathcal{S}(\mathbf{R}^d) = \{f \in C^\infty(\mathbf{R}^d) : |(D_\alpha f)(x)| \lesssim_{\alpha,N} |x|^{-N} \text{ for all } N, \alpha, x\}$$

which is often a natural space to consider the relation of the Fourier transform to various analytical operations.

Theorem 4.4. *If $f \in L^1(\mathbf{R}^d)$, and $x_k f \in L^1(\mathbf{R}^d)$, then $\widehat{f}$ has a strong derivative in the L^∞ norm, and $\widehat{f}_k(\xi) = -2\pi i (x_k f)^\wedge(\xi)$.*

Proof. Note that a change of variables implies

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

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 $\Delta_h \widehat{f} = \widehat{g_h}$ 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, in all the L^p spaces.

Theorem 4.5. *If f has a strong derivative f_k in the L^1 norm, $\widehat{f_k}(\xi) = 2\pi i \xi_k \widehat{f}(\xi)$.*

Proof. It suffices to note that

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

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

4.1 Convergence Using Alternative Summation

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

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

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*

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

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

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

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

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

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. *Similarily, 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 $\delta \to 0$. For $f \in L^1(\mathbf{R}^d)$, we let

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

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

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

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

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

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 $t \to \infty$. Thus we study

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

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

Lemma 4.6. *Suppose $f \in L^1_{loc}(\mathbf{R}^d)$, that*

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

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

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

Proof. Let

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

For each $x \geqslant 0$, write

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

Then F is continuous, and $F(x) \to A$ as $x \to \infty$. 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 $s \to \infty$, using the fact that F is bounded so that $F(s)e^{-\delta s^2} \to 0$, 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 $|F(s) - A| \leqslant \varepsilon$ for $s \geqslant t$. Then

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

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

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

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

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

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

Abel summation is even more general than Gauss summation.

Lemma 4.7. *If f is Gauss summable, and $f(x)e^{-\delta|x|}$ is absolutely integrable for each $\delta > 0$, then f is Abel summable, and*

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

Proof. Let

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

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

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

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

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

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

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

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

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

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 $\delta \to 0$. Thus if $f \in L^1(\mathbf{R}^d)$, the dominated convergence theorem implies that f is Φ summable to it's usual integral. We now use these summability kernels to understand the Fourier summation formula.

Theorem 4.8 (The Multiplication Formula)**.** *If $f, g \in L^1(\mathbf{R}^n)$,*

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

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

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

If Φ is integrable, then the multiplication formula shows

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

Thus if we define $K_\delta^\Phi(x) = \delta^{-n}\widehat{\Phi}(-x/\delta)$, then $S_\delta(f,\Phi) = K_\delta^\Phi * f$. Thus we have expressed the 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 $\varepsilon > 0$,
$$\lim_{\delta\to 0}\int_{|\xi|\geqslant\varepsilon}|K_\delta(\xi)|\,d\xi \to 0.$$

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

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

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

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

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

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

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

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

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

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

If $\xi \geqslant 1$, the functions f_ξ and g_ξ are both negligible in the lower half plane, and have no poles in the lower half plane, so if we let γ denote the curve of length $2\pi n$ travelling anticlockwise along the lower semicircle with vertices $-n - iy$ and $n - iy$, then because $|z| \geqslant n$ on γ,

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

and so we conclude that

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

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

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

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

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

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

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

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

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

Taking $n \to \infty$, we conclude

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

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

Example. *In the next paragraph, we calculate that if $\Phi(x) = e^{-\pi|x|^2}$, then $\widehat{\Phi} = \Phi$. Thus if we define the* Weirstrass 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 $\widehat{\Phi}$ is differentiable, and

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

The uniqueness theorem for ordinary differential equations says that since

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

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

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

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

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

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

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

Thus the Abel mean is the Fourier inverse of the Poisson kernel on the upper half plane $\mathbf{H}^{n+1}$.

In order to conclude $\{P_\delta\}$ *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 $D_R \notin L^1(\mathbf{R})$. *Thus* D_R *is not a good kernel, which makes the convergence rates of* $S_R f$ *more subtle. Nonetheless,* D_R *does lie in* $L^p(\mathbf{R})$ *for all* $p \in (1,\infty]$, *and is* uniformly bounded *in* $L^p(\mathbf{R})$ *for all* $p \in (1,\infty)$, *a fact we will prove later. This is enough to conclude that for all* $p \in (1,\infty)$, $S_R f \to f$ *in* $L^p(\mathbf{R})$.

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,

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

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

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

Thus the inversion theorem holds pointwise almost everywhere.

Theorem 4.9. *If f and $\widehat{f}$, then for any x in the Lebesgue set of f,*

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

We define, for any integrable $f : \mathbf{R}^n \to \mathbf{R}$, 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 $\widehat{f}$ is integrable. In particular, this is true if f has strong derivatives in the L^1 norm for any multi-index $|\alpha| \leqslant n+1$, and so the Fourier inversion formula holds for sufficiently smooth functions.

Corollary 4.10. *If $f \in C(\mathbf{R})$ is integrable and $\widehat{f} \in L^1(\mathbf{R})$, $S_R f \to f$ uniformly.*

Proof. The dominated convergence theorem implies that for each $x \in \mathbf{R}$,

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

And

$$\int_{|x| \geqslant R} \widehat{f}(\xi) e(\xi \cdot x) \leqslant \|\widehat{f}\|_{L^1(\mathbf{R})}.$$

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

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

also converge uniformly to f.

Corollary 4.11. *The map $\mathcal{F} : L^1(\mathbf{R}^d) \to C_0(\mathbf{R}^d)$ is injective.*

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

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

Thus $f = 0$ almost everywhere. □

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 4.12. *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 $W_{\delta_1} * W_{\delta_2} = W_{\delta_1+\delta_2}$. Similarily, $P_{\delta_1+\delta_2}$ 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}. □

4.2 The L^2 Theory

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

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

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

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

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

Theorem 4.13. *If $f \in L^1(\mathbf{R}^n) \cap L^2(\mathbf{R}^n)$, then $\|\widehat{f}\|_{L^2(\mathbf{R}^d)} = \|f\|_{L^2(\mathbf{R}^d)}$.*

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

$$|\widehat{f}(\xi)| = \widehat{f}(\xi)\overline{\widehat{f}}(\xi),$$

and

$$\left(\overline{\widehat{f}}\right)^{\wedge}(\xi) = \overline{(f^{\wedge})^{\wedge}(-\xi)} = \overline{f(\xi)}.$$

This implies

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

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

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

where $2 \leqslant q \leqslant \infty$ 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 $1 \leqslant p \leqslant 2$. 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 $\mathcal{F} : L^p(\mathbf{R}^n) \to L^q(\mathbf{R}^n)$ 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

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

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

4.3 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 by applying symmetry arguments.

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*

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

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 calcluate 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 $t \to \infty$, we need $n/p - n/2 \leqslant 0$, so $p \leqslant 2$. As $t \to \infty$, $|f_s(x)| \sim t^{-n/2}e^{-\pi|x/t|^2}$, 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 $p \leqslant 2$ 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.

4.4 The Poisson Summation Formula

We now show a connection between the Fourier transform on $\mathbf{R}$, and the Fourier transform on $\mathbf{T}$. If $f \in \mathcal{S}(\mathbf{R})$, there are two ways of obtaining a 'periodic' version of f on $\mathbf{T}$. Firstly, we can define, for each $x \in \mathbf{T}$,

$$f_1(x) = \sum_{n=-\infty}^{\infty} f(x + 2\pi n),$$

which is a well defined element of $C^\infty(\mathbf{T})$. Secondly, we can define

$$f_2(x) = \sum_{n=-\infty}^{\infty} \widehat{f}(n) e_n(x),$$

The Poisson summation formula says that these two functions are the same function.

Theorem 4.15. *If $f \in \mathcal{S}(\mathbf{R})$, then*

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

4.5 Sums of Random Variables

TODO

We now switch to an application of harmonic analysis to studying sums of random variables probability theory. If X is a random vector, it's probabilistic information is given by it's 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 independant 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 independant 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 4.16. *Let X and Y be independant random variables. Then*

- $\varphi_X(0) = 1$, *and* $|\varphi_X(\xi)| \leqslant 1$ *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 4.17. *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. □

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

Theorem 4.18. *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*

$$\mathbf{P}(S_N \leqslant \sigma\sqrt{N}t) \to \Phi(t) = \frac{1}{\sqrt{2x}} \int_{-\infty}^{t} e^{-y^2/2}\, dy$$

Proof. We calculate that

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

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

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

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

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

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

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

By the dominated convergence theorem, as $\xi \to 0$, $\varphi_X(\xi) = 1 - \xi^2\sigma^2/2 + o(\xi^2)$. 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. $\square$

4.6 Transforms of Holomorphic Functions

TODO

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 it's Fourier transform, we will find that analyticity corresponds to an exponential decay. We let $S_a = \{z \in \mathbf{C} : |\text{Im}(z)| < a\}$ denote the horizontal strip.

Theorem 4.19. *Let f be holomorphic on S_a, integrable on each horizontal line contained in the strip, and such that $f(z) \to 0$ as $|Re(z)| \to \infty$. Then we find $|\widehat{f}(\xi)| \lesssim_b e^{-2\pi b|\xi|}$ 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 $R \to \infty$, the integral along the vertical lines of the rectangle tends to zero as $R \to \infty$, so we conclude that

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

where $f_b(x) = f(x - ib)$. Thus it suffices to show $|\widehat{f_b}(\xi) - \widehat{f}(\xi)| \lesssim_b 1$. 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 $\widehat{f}$ has exponential decay if f satisfies the hypothesis of the theorem. Thus we can always apply the inverse Fourier transform to conclude

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

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

Theorem 4.20. *A*

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

$$\int_{-\infty}^{\infty} \widehat{f}(\xi) e^2$$

$\square$

4.7 Characteristic Functions

Chapter 5

Finite Character Theory

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

- The functions are orthogonal to one another.
- A large family of functions can be approximated by linear combinations of the power maps.
- The power maps are multiplicative: $\phi_n(zw) = \phi_n(z)\phi_n(w)$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5.2 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 5.1. *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

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

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

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

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

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

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

since a finite abelian group is a finite product of cyclic groups, this proves the theorem. This seems almost like sweeping the algebra of the situation under the rug, however, so we will prove the statement only using elementary linear algebra. What's more, these linear algebraic techniques generalize to the theory of unitary representations in harmonic analysis over infinite groups.

Theorem 5.2. *Let $\{T_1, \dots, T_n\}$ be a family of commuting unitary matrices. Then there is a basis $v_1, \dots, v_m \in \mathbf{C}^m$ which are eigenvectors for each T_i.*

Proof. For $n = 1$, the theorem is the standard spectral theorem. For induction, suppose that the $T_1, \dots, T_{k-1}$ are simultaneously diagonalizable. Write

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

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 hypotheis 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 basic consisting of the $\chi_a(b) = N[a = b]$, for $a \in G$. In this basis, we comcpute $T_a \chi_b = \chi_{ba^{-1}}$, hence T_a is a permutation matrix with respect to this basis, hence unitary. The operators T_a commute, since $T_a T_b = T_{ab} = T_{ba} = T_b T_a$. Hence these operators can be simultaneously diagonalized. That is, there is a family $e_1, \dots, e_n \in V$ 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 5.3. *Let G be a finite abelian group. Then $\Gamma(G) \cong G$, and forms an orthonormal basis for the space of complex valued functions on G. For any function $f : G \to \mathbf{C}$,*

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

In this context, we also have Parseval's theorem

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

5.3 Convolutions

There is a version of convolutions for finite functions, which is analogous to the convolutions on $\mathbf{R}$. Given two functions f, g on G, we define a function $f * g$ on G by setting

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

The mapping $b \mapsto ab^{-1}$ is a bijection of G, and so we also have

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

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

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

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

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

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

$$D(a) = \sum_{e \in \Gamma(G)} e(a)$$

We claim that $D(a) = |G|$ if $a = 1$, and $D(a) = 0$ otherwise. Note that since $|G| = |\Gamma(G)|$, the character space of $\Gamma(G)$ is isomorphic to G. Indeed, for each $a \in G$, we have the maps $\widehat{a} : e \mapsto e(a)$, which is a character of $\Gamma(G)$. Suppose $e(a) = 1$ for all characters e. Then $e(a) = e(1)$ for all characters e, and for any function $f : G \to \mathbf{C}$, we have $f(a) = f(1)$, implying $a = 1$. Thus we obtain $|G|$ distinct maps $\widehat{a}$, 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.

5.4 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 $\widehat{f_n} \to \widehat{f}$ uniformly. The Fourier transform of f_n is the same as the Fourier transform of the corresponding function $k \mapsto a_k$ on $\mathbf{Z}_n$, and thus we can approximate the Fourier transform on $\mathbf{T}$ by a discrete computation on $\mathbf{Z}_n$. Looking at the formula in the definition of the discrete transform, we find that we can compute the Fourier coefficients of a function $f : \mathbf{Z}_n \to \mathbf{C}$ in $O(n^2)$ addition and multiplication operations. It turns out that there is a much better method of computation which employs a divide and conquer approach, which works when n is a power of 2, reducing the calculation to $O(n \log n)$ multiplications. 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 $f : \mathbf{Z}_{2n} \to \mathbf{C}$, define two functions $g, h : \mathbf{Z}_n \to \mathbf{C}$, defined by $g(k) = f(2k)$, and $h(k) = f(2k+1)$. Then g and h encode all the information in f, and if $\nu = e(\pi/n)$ is the canonical generator of $\mathbf{Z}_{2n}$, we have

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

Because

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

This is essentially a discrete analogue of the Poission summation formula, which we will generalize later when we study the harmonic analysis of abelian groups. If $H(m)$ is the number of operations needed to calculate the Fourier transform of a function on μ_{2^n} using the above recursive formula, then the above relation tells us $H(2m) = 2H(m) + 3(2m)$. 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}).

5.5 Dirichlet's Theorem

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

Theorem 5.4. *If m and n are relatively prime, then the set*

$$\{m + kn : k \in \mathbf{N}\}$$

contains infinitely many prime numbers.

An exploration of this requries the Riemann-Zeta function, defined by

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

The function is defined on $(1, \infty)$, since for $s > 1$ the map $t \mapsto 1/t^s$ is decreasing, and so

$$\sum_{n=1}^{\infty} \frac{1}{n^s} \leqslant 1 + \int_1^{\infty} \frac{1}{t^s} = 1 + \lim_{n \to \infty} \frac{1}{s-1}\left[1 - 1/n^{s-1}\right] = 1 + \frac{1}{s-1}$$

The series converges uniformly on $[1+\varepsilon, N]$ for any $\varepsilon > 0$, so ζ is continuous on $(1, \infty)$. As $t \to 1$, $\zeta(t) \to \infty$, because $n^s \to n$ for each n, and if for a fixed M we make s close enough to 1 such that $|n/n^s - 1| < 1/2$ for $1 \leqslant n \leqslant M$, then

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

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

The Riemann-Zeta function is very good at giving us information about the prime integers, because it encodes much of the information about the prime numbers.

Theorem 5.5. *For any $s > 1$,*

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

Proof. The general idea is this – we may write

$$\prod_{p \text{ prime}} \frac{1}{1 - p^s} = \prod_{p \text{ prime}} (1 + 1/p^s + 1/p^{2s} + \dots)$$

If we expand this product out formally, enumating the primes to be $p_1, p_2, \dots,$ we find

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

□

Chapter 6

Applications

6.1 The Wirtinger Inequality on an Interval

Theorem 6.1. *Given $f \in C^1[-\pi,\pi]$ with $\int_{-\pi}^{\pi} f(t)dt = 0$,*

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

Proof. Consider the fourier series

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

Then $a_0 = 0$, and so

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

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

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

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

Corollary 6.2. *Given $f \in C^1[a,b]$ with $\int_a^b f(t)\, dt = 0$,*

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

6.2 Energy Preservation in the String equation

Solutions to the string equation are

If $u(t,x)$

6.3 Harmonic Functions

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

Theorem 6.3. *If $f \in L^p(\mathbf{R}^n)$, for $1 \leqslant p \leqslant \infty$, and $u(x,y) = (f * P_y)(x)$, where*

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

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

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

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

$\square$

Part II

Euclidean Harmonic Analysis

Here, we try and describe the more modern approaches to real-variable harmonic analysis, as developed by the *Calderon-Zygmund school* of analysis as developed in the 1970s. Almost all of the problems we consider are to do with bounding operators on function spaces. Given some function f lying in a space V, we have an associated function Tf lying in some space W. The main goal of the techniques in this part of the book attempt to understand how quantitative control on certain properties of f imply quantitative control on properties of Tf. In particular, given some quantity $A(f)$ associated with each $f \in V$, and a quantity $B(g)$ defined for all $g \in W$. Our goal is to understand whether a general bound $B(Tf) \lesssim A(f)$ is possible for all functions $f \in V$, i.e. whether these exists a universal constant $C > 0$ such that $B(Tf) \leqslant C \cdot A(f)$ for all $f \in V$.

A core technique we employ here is the method of *decomposition*. We write $f = \sum_k f_k$, where the function f_k have particular properties, perhaps being concentrated in a particular region of space, or having a Fourier transform concentrated in a particular region. These concentration properties often simplify analysis, enabling us to obtain bounds $B(Tf_k) \lesssim A(f_k)$ for each n. Provided that the operator T, and the quantities A and B are 'stable under addition', we can then obtain the bound $B(Tf) \leqslant A(f)$ by 'summing' up the related quantities. The stability of A and B is often obtained by assuming these quantities are *norms* on their respective function spaces, i.e. that there exists norms $\|\cdot\|_V$ and $\|\cdot\|_W$ such that $A(f) = \|f\|_V$ for each $f \in V$ and $B(g) = \|g\|_W$ for each $g \in W$. The stability of T under addition is obtained by assuming linearity, or at least sub-linearity, in the sense that for each $f_1, f_2 \in V$,

$$\|T(f_1 + f_2)\|_W \leqslant \|Tf_1\|_W + \|Tf_2\|_W.$$

We can then use the triangle inequality to conclude that

$$\|Tf\|_W \leqslant \sum_k \|Tf_k\|_W \lesssim \sum_k \|f_k\|_V.$$

Thus if $\sum_k \|f_k\|_V \lesssim \|f\|_V$, our argument is complete. This will be true, for instance, if there exists $\varepsilon > 0$ such that $\|f_k\|_V \lesssim 2^{-\varepsilon k}\|f\|_V$. This can often be obtained if we employ a *dyadic decomposition technique*. For such decompositions, it is also possible to generalize are technique not only to norms, but also to *quasinorms*, i.e. maps $\|\cdot\|$ which are homogeneous and satisfy a *quasi-triangle inequality* $\|v+w\| \lesssim \|v\| + \|w\|$.

Lemma 6.4. *Suppose $\|\cdot\|_V$ is a quasi-norm on a vector space V, and under the topology induced by $\|\cdot\|_V$, we can write $f = \sum_{k=1}^\infty f_k$, where there is $\varepsilon > 0$ and $C > 0$ such that for each n, $\|f_k\| \leqslant C \cdot 2^{-\varepsilon k}$. Then $\|f\| \lesssim_\varepsilon C$.*

Remark. Thus if T is sublinear and we have $\|Tf_k\|_W \lesssim \|f_k\|_V$ and $\|f_k\|_V \lesssim 2^{-\varepsilon k}\|f\|_V$, we conclude $\|Tf_k\|_W \lesssim 2^{-\varepsilon k}\|f\|_V$, and then by sublinearity and the lemma applied to $\|\cdot\|_W$, we conclude

$$\|Tf\|_W \leqslant \|\sum_k Tf_k\|_W \lesssim_\varepsilon \|f\|_V.$$

A slight modification of the proof below even gives this claim provided T is *quasi sublinear*, in the sense that for all $f_1, f_2 \in V$, $\|T(f_1+f_2)\|_W \lesssim \|Tf_1\|_V + \|Tf_2\|_V$ for all $f_1, f_2 \in V$. However, such operators occur so rarely in practice that it isn't worth concentrating on them.

Proof. Pick $A > 0$ such that $\|f_1 + f_2\|_V \leqslant A \cdot (\|f_1\|_V + \|f_2\|_V)$ for all f_1 and f_2. If $A < 2^\varepsilon$, we can write apply the quasitriangle inequality iteratively to conclude

$$\|f\| \leqslant C \cdot \sum_{k=1}^\infty A^k \|f_k\|_V \leqslant C \cdot \left(\sum_{k=1}^\infty (A2^{-\varepsilon})^k \right) \leqslant C \cdot \left(\frac{1}{1 - A2^{-\varepsilon}} \right) \lesssim_\varepsilon C.$$

In general, fix N, and write $f = f^1 + \dots + f^N$, where $f^m = \sum_{k=0}^\infty f_{m+Nk}$. Then $\|f_{m+Nk}\|_V \leqslant C \cdot 2^{-N\varepsilon k}$, and if N is chosen large enough that $A < 2^{N\varepsilon}$, we can apply the previous case to conclude that $\|f^m\|_V \lesssim_\varepsilon C$. Then we can apply the quasi-triangle inequality to conclude that $\|f\| \lesssim_\varepsilon C$. $\square$

We can even apply the method of decomposition in the presence of suitably large polynomial decay.

Lemma 6.5. *Suppose $\|\cdot\|_V$ is a quasinorm on a function space V. Then there exists t such that for all $s > t$, if $f = \sum_{k=1}^\infty f_k$, and if $\|f_k\|_V \leqslant C \cdot k^{-s}$, for $s > t$, then $\|f\|_V \lesssim_s C$.*

Proof. As in the previous lemma, pick $A > 0$ such that $\|f_1 + f_2\|_V \leqslant A(\|f_1\|_V + \|f_2\|_V)$ for all $f_1, f_2 \in V$. We perform a decomposition of dyadic type, writing $f = \sum_{m=0}^\infty f^m$, where

$$f^m = \sum_{k=2^m}^{2^{m+1}-1} f_k.$$

By splitting up the sum into a binary tree, we can ensure that

$$\|f^m\|_V \lesssim A^{m+1} \sum_{k=2^m}^{2^{m+1}-1} \|f_k\|_V \leqslant C \cdot A^{m+1} \sum_{k=2^m}^{2^{m+1}-1} k^{-s} \lesssim C(A2^{1-s})^m.$$

If $s > 1 + \lg(A)$, we can thus apply the previous lemma to conclude that $\|f\|_V \lesssim C$. $\square$

In this part of the notes, we define the various classes of quasi-norms we will study, describe the general methods which make up the Calderon-Zygmund theory, and find applications to geometric measure theory, complex analysis, partial differential equations, and analytic number theory.

Chapter 7

Monotone Rearrangement Invariant Norms

In this chapter, we discuss common families of *monotone, rearrangement invariant quasinorms* that occur in harmonic analysis. The general framework is as follows. For each function f, we associate it's *distribution function* $F : [0, \infty) \to [0, \infty]$ given by $F(t) = |\{x : |f(x)| > t\}|$. A *rearrangement invariant space* is a subspace V of the collection of measurable complex-valued functions on some measure space X, equipped with a quasi-norm $\|\cdot\|$, satisfying the following two properties:

- *Monotonicity*: If $|f(x)| \leqslant |g(x)|$ for all $x \in X$, then $\|f\| \leqslant \|g\|$.
- *Rearrangement-Invariance*: If f and g have the same distribution function, then $\|f\| = \|g\|$.

A monotone rearrangement-invariant norm essentially provides a way of quantifying the height and width of functions on X. It has no interest in the 'shape' of the objects studied, because of the property of rearrangement invariance. In a particular problem, one picks the norm best emphasizing a particular family of features useful in the problem.

There are two very useful classes of functions useful for testing the behaviour of translation invariant norms:

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

The class of all simple functions forms a vector space, and for almost all the monotone rearrangement invariant norm we consider in this section, this vector space will form a dense subspace of the class of all functions. This means that when we want to study how an operator transforms the height and width of functions, the behaviour of the operator on simple functions often reflects the behaviour of an arbitrary function.

7.1 The L^p norms

For $p \in (0,\infty)$, we define the L^p norm on measurable function on a measure space X by

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

For $p = \infty$, we define

$$\|f\|_\infty = \min \{ t \geqslant 0 : |f(x)| \leqslant t \text{ almost surely} \}.$$

These are the most fundamental monotone, rearrangement invariant norms. The space of functions f with $\|f\|_p < \infty$ is denoted by $L^p(X)$. The most important spaces to consider here are the space $L^1(X)$, consisting of absolutely square integrable functions, $L^\infty(X)$, consisting of almost-everywhere bounded functions, and $L^2(X)$, consisting of square integrable functions. The main motivation for the introduction of the other L^p spaces is that much of the quantitative theory for $p \in \{1,\infty\}$ is rather trivial, in the sense that it is easy to see when certain operators are bounded on these spaces, or unbounded.

As p increases, the L^p norm of a particular function f gives more control over the height of the function f, and weaker control on values where f is particular small. At one extreme, $L^\infty(X)$ only has control over the height of a function, and no control over it's width. Conversely, one can think of $L^0(X)$ as being the space of functions with finite support, though no natural norm exists on this space of functions solely classifying width. After all, such a quantity couldn't be homogenous, since the width of f and αf are the same for each $\alpha \neq 0$. Thus the space $L^0(X)$ isn't so interesting to us from a quantitative perspective.

Example. *If $f(x) = |x|^{-s}$ for $x \in \mathbf{R}^d$ and $s > 0$, then integration by radial coordinates shows that*

$$\int_{\varepsilon \leqslant |x| \leqslant M} \frac{1}{|x|^{sp}}\, dx \approx \int_\varepsilon^M r^{d-1-ps}\, dr = \frac{M^{d-ps} - \varepsilon^{d-ps}}{d-ps}.$$

This quantity remains finite as $\varepsilon \to 0$ if and only if $d > ps$, and finite as we let $M \to \infty$ if and only if $d < ps$. Thus if $p < d/s$, f is locally *in L^p, in the sense that $f \in L^p(B)$ for every bounded $B \in \mathbf{R}^d$. The class of functions for which this condition holds is denoted $L^p_{loc}(X)$. Conversely, if $p > d/s$, then for every domain B separated from the origin, $f \in L^p(B)$. For $p = d/s$, the function f fails to be $L^p(\mathbf{R}^d)$, but only 'by a logarithm', in the sense that*

$$\int_{\varepsilon \leqslant |x| \leqslant M} \frac{1}{|x|^{sp}}\, dx \approx \int_\varepsilon^M \frac{dr}{r} = \log(M/\varepsilon).$$

We will later find 'weaker' versions of the L^p norm, and f will have finite version of these norms.

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

Example. *If $s = A\chi_E$, and we set $H = |A|$ and $W = |E|$, then $\|s\|_p = W^{1/p}H$. As $p \to \infty$, the value of $\|s\|_p$ depends more and more on H, and less on W, and in fact $\lim_{p\to\infty} \|s\|_p = H$. If $s = \sum A_n \chi_{E_n}$, 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 $p \to \infty$,*

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

This implies $\|s\|_p \to |A_m|$ as $p \to \infty$. But as $p \to 0$, $\lim_{p\to 0} \|f\|_p$ does not in general exist, even for step functions with finite support. Nonetheless, we can conclude that $\lim_{p\to 0} \|s\|_p^p = \sum |E_n|$, which is the measure of the support of s.

As $p \to \infty$, 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 $|f| \leqslant \|f\|_\infty$ 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 $\|s\|_p \to \|s\|_\infty$ if s is a simple function, and the density of such functions gives a general result.

Theorem 7.1. *Let $p \in (0, \infty)$. If $f \in L^p(X) \cap L^\infty(X)$, then*

$$\lim_{t \to \infty} \|f\|_t = \|f\|_\infty.$$

Proof. Without loss of generality, assume $p \geqslant 1$. Consider the norm $\|\cdot\|$ on $L^p(X) \cap L^\infty(X)$ given by

$$\|f\| = \|f\|_p + \|f\|_\infty.$$

Then $L^p(X) \cap L^\infty(X)$ is complete with respect to this metric. For each $t \in [p, \infty)$, define $T_t(f) = \|f\|_t$. Then the functions $\{T_t\}$ are uniformly bounded in the norm $\|\cdot\|$, since if $p = \theta t$, then

$$|T_t(f)| = \|f\|_t \leqslant \|f\|_p^\theta \|f\|_\infty^{1-\theta} \leqslant \|f\|^\theta \|f\|^{1-\theta} = \|f\|.$$

For any $\varepsilon > 0$, we can find a step function s with $\|s - f\|_p, \|s - f\|_\infty \leqslant \varepsilon$. This means that for all $t \in (p, \infty), \|s - f\|_t \leqslant \varepsilon$. And so

$$\Big| T_t(f) - \|f\|_\infty \Big| \leqslant |T_t(f) - T_t(s)| + |T_t(s) - \|s\|_\infty| + |\|s\|_\infty - \|f\|_\infty| \leqslant 2\varepsilon + o(1).$$

Taking $\varepsilon \to 0$ gives the result. □

Abusing notation, we define $\|f\|_0^0 = |\mathrm{supp} f| = |\{x : f(x) \neq 0\}|$, and let $L^0(X)$ be the space of functions with finite support. We know that for any simple function s, $\|s\|_p^p \to \|s\|_0^0$ as $p \to 0$. If $f \in L^0(X) \cap L^p(X)$ for some $p \in (0, \infty)$, then the monotone and dominated convergence theorems implies that

$$\|f\|_0^0 = \int \mathbf{I}(f(x) \neq 0) = \int \left(\lim_{t \to 0} |f(x)|^t \right) \, dx = \lim_{t \to 0} \int |f(x)|^t \, dx = \lim_{t \to 0} \|f\|_t^t.$$

Thus the space $L^0(X)$ lies at the opposite end of the spectrum to L^∞.

The fact that $\|f\|_0^0$ is a norm taken to the 'power of zero' implies that many nice norm properties of the L^p spaces fail to hold for $L^0(X)$. For instance, homogeneity no longer holds; in fact, for each $\alpha \neq 0$,

$$\|\alpha f\|_0^0 = \|f\|_0^0.$$

It does, however, satisfy the triangle inequality $\|f + g\|_0^0 \leqslant \|f\|_0^0 + \|g\|_0^0$, which follows from a union bound on the supports of the functions.

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 $\lambda \geqslant 0$,*

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

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 $A > 0$,

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

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 $\|f\|_r \leqslant 2\|f\|_p^\theta\|f\|_q^{1-\theta}$, 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 X^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

$$\|f\|_r = \|f^{\otimes n}\|_r^{1/n} \leqslant \left(2\|f^{\otimes n}\|_p^\theta\|f^{\otimes n}\|_q^{1-\theta}\right)^{1/n} = 2^{1/n}\|f\|_p^\theta\|g\|_q^{1-\theta}.$$

We can then take $n \to \infty$ to conclude that $\|f\|_r \leqslant \|f\|_p^\theta\|f\|_q^{1-\theta}$.

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. We now also show how to prove this inequality using convexity, which illustrates another core technique. In the next theorem, $1/\infty = 0$.

Theorem 7.2 (Hölder)**.** *If $0 < p,q \leqslant \infty$ and $1/p+1/q = 1/r$, $\|fg\|_r \leqslant \|f\|_p\|g\|_q$.*

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 $A, B \geqslant 0$ and $0 \leqslant \theta \leqslant 1$, $A^\theta B^{1-\theta} \leqslant \theta A + (1-\theta)B$. But since the logarithm is concave, we calculate

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

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

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

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

Remark. Note that $A^\theta B^{1-\theta} \leqslant \theta A + (1-\theta)B$ 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 7.3. *Given f,g, and $p \geqslant 1$, $\|f+g\|_p \leqslant \|f\|_p + \|g\|_p$.*

Proof. The inequality when $p = 1$ is obtained by integrating the inequality $|f(x)+g(x)| \leqslant |f(x)| + |g(x)|$, 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

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

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

Remark. Suppose $\|f+g\|_p = \|f\| + \|g\|_p$. 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 is 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.

If $p < 1$, then the proof of Corollary 7.3 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

$$\|f+g\|_p \geqslant \|f\|_p + \|g\|_p,$$

as proved in the next theorem.

Theorem 7.4. *If $p \geqslant 1$, then for any functions $f_1, \dots, f_N \geqslant 0$,*

$$(\|f_1\|_p^p + \dots + \|f_N\|_p^p)^{1/p} \leqslant \|f_1 + \dots + f_N\|_p \leqslant \|f_1\|_p + \dots + \|f_N\|_p. \tag{7.1}$$

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

$$\|f_1\|_p + \dots + \|f_N\|_p \leqslant \|f_1 + \dots + f_N\|_p \leqslant (\|f_1\|_p^p + \dots + \|f_N\|_p^p)^{1/p} \tag{7.2}$$

Proof. The upper bound in (7.1) is just obtained by applying the triangle inequality iteratively. To obtain the lower bound, we note that for $A_1, \dots, A_N \geqslant 0$,

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

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

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

Now assume $0 < p < 1$. We begin by proving the lower bound in 7.2. We can assume $N = 2$, and $\|f_1\|_p + \|f_2\|_p = 1$, and then it suffices to show $\|f_1 + f_2\|_p \geqslant 1$. For any $\theta \in (0,1)$, and $A, B \geqslant 0$, concavity implies

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

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

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

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

$$\|f_1 + \dots + f_N\|_p^p \leqslant \|f_1\|_p^p + \dots + \|f_N\|_p^p. \qquad \square$$

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 $p \geqslant 1$. One way to fix this is to use the theory of Hardy spaces. We note that for $p < 1$, we do have a *quasi* triangle inequality.

Theorem 7.5. *For $f_1, \dots, f_N \in L^p(X)$, with $0 < p < 1$,*

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

Proof. By Hölder's inequality applied to sums,

$$\|f_1 + \dots + f_N\|_p \leqslant (\|f\|_p^p + \dots + \|f_N\|_p^p)^{1/p} \leqslant N^{1/p-1}(\|f_1\|_p + \dots + \|f_N\|_p). \qquad \square$$

This result is sharp, i.e. if we take a disjoint family of sets $\{E_1, E_2, \dots\}$ with $|E_i| = 1$ for each i, and then set $f_i = \mathbf{I}_{E_i}$, then the inequality is sharp for each N.

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

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

lies in C, and

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

as $N \to \infty$, 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 $p \to \infty$, the L^p norm excludes functions with large peaks, or large height, and as $p \to 0$, 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 $0 < p \leqslant q \leqslant \infty$, $L^p(X) \subset L^q(X)$. Hölder's inequality implies $\|f\|_p = \|f\chi_X\|_p \leqslant \|f\|_q|X|^{1/p-1/q}$. Taking $q \to \infty$, we conclude $\|f\|_p \leqslant |X|^{1/p}\|f\|_\infty$. One can best remember the constants here by the formula*

$$\left(⨍ |f(x)|^p\right)^{1/p} \leqslant \left(⨍ |f(x)|^q\right)^{1/q}.$$

In particular, when X is a probability space, the L^p norms are increasing.

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 $|E| \geqslant \varepsilon$ for any measurable set E. Then $L^q(X) \subset L^p(X)$ for $0 < p \leqslant q \leqslant \infty$. First we check the $q = \infty$ case, which follows by the trivial estimate*

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

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

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

If $\varepsilon = 1$, which occurs if $X = \mathbf{Z}$, then the L^p norms are decreasing in p. This gives the best way to remember the constants involved, since the measure $\mu(E) = |E|/\varepsilon$ is one granular, and so

$$\left(\frac{1}{\varepsilon} \int |f(x)|^q \, dx \right)^{1/q} \leqslant \left(\frac{1}{\varepsilon} \int |f(x)|^p \, dx \right)^{1/p}.$$

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 $p \leqslant q$,

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

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.

Example. *We can obtain similar L^p bounds by controlling the functions f involved, rather than the measure space. For instance, if $|f(x)| \leqslant M$, and $p \leqslant q$, then then $\|f\|_q \leqslant \|f\|_p^{p/q} M^{1-p/q}$, which follows by log convexity. On the other hand, if $|f(x)| \geqslant M$ on the support of f, then $\|f\|_p \leqslant \|f\|_q^{q/p} M^{1-q/p}$.*

Theorem 7.6. *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,

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

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

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. $\square$

Remark. 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.

The last property of the L^p norms we want to focus on is the principle of *duality*. Given any values of p and q with $1/p + 1/q = 1$, Hölder's inequality implies that if $f \in L^p(X)$ and $g \in L^q(X)$, then $fg \in L^1(X)$. In particular, for each function $g \in L^q(X)$, the map

$$\lambda : f \mapsto \int f(x)g(x)\, dx$$

is a linear functional on $L^p(X)$. Hölder's inequality implies that $\|\lambda\| \leqslant \|g\|_q$. But this is actually an *equality*. In particular, if $1 < p < \infty$, one can show these are *all* linear functionals. For $p \in \{1, \infty\}$, the dual space of $L^p(X)$ is more subtle. But, since in harmonic analysis we concentrate on quantitative bounds, the following theorem often suffices as a replacement.

Theorem 7.7. *If $1 \leqslant p < \infty$, and $f \in L^p(X)$, then*

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

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

Proof. Suppose that $1 \leqslant p < \infty$. 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 $s_1 \leqslant s_2 \leqslant \dots \to |f|$. Each s_k lies in $L^p(X)$, but the monotone convergence theorem implies that $\|s_k\|_p \to \infty$. For each k, find a function $g_k \geqslant 0$ with $\|g_k\|_q = 1$, and $\int g_k(x)s_k(x) \geqslant \|s_k\|_p/2$. Then

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

this completes the proof in this case.

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

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

Taking $\varepsilon \to 0$ completes the claim. □

7.2 Decreasing Rearrangements

The properties of a functions distribution are best reflected quite simply in the *distribution function* of the function f, i.e. the function $F : [0,\infty) \to [0,\infty)$ given by $F(t) = |\{x : |f(x)| > t\}|$, and any rearrangement invariant norm on f should be a function of F. The function F is right-continuous and decreasing, but has a jump discontinuity whenever $\{x : |f(x)| = t\}$ is a set of positive measure. We denote distributions of functions g and h by G and H.

Lemma 7.8. *Given a function f and g, $\alpha \in \mathbf{C}$, and $t,s > 0$, then*

- *If $|g| \leqslant |f|$, then $G \leqslant F$.*

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

Proof. The first point follows because $\{x : |g(x)| > t\} \subset \{x : |f(x)| > t\}$, and the second because $\{x : |\alpha f(x)| > t\} = \{x : |f(x)| > t/|\alpha|\}$. The third point follows because if $|f(x) + g(x)| \geqslant t + s$, then either $|f(x)| \geqslant t$ or $|g(x)| \geqslant s$. Finally, if $|f(x)g(x)| \geqslant ts$, then $|f(x)| \geqslant t$ or $|g(x)| \geqslant s$. □

We can simplify the study of the distribution of f even more by defining the *decreasing rearrangement* of f, a decreasing function $f^* : [0, \infty) \to [0, \infty)$ such that $f^*(s)$ is the *smallest* number t such that $F(t) \leqslant s$. Effectively, $f^*(s)$ is the inverse of F:

- If there is a unique t with $F(t) = s$, then $f^*(s) = t$.
- If there are multiple values t with $F(t) = s$, let $f^*(s)$ be the *smallest* such value.
- If there are no values t with $F(t) = s$, then we pick the first value t with $F(t) < s$.

We find

$$\{s : f^*(s) > t\} = \{s : s < F(t)\} = [0, F(t)),$$

which has measure $F(t)$. This is the most important property of f^*; it is a decreasing function on the line which has the same distribution as the function $|f|$. It is also the unique such function which is right continuous. Thus our intuition when analyzing monotone, rearrangement invariant norms is not harmed if we focus on right continuous decreasing functions.

Theorem 7.9. *The function f^* is right continuous.*

Proof. We note that $F(t) > s$ if and only if $t < f^*(s)$. Since f^* is decreasing, for any $s \geqslant 0$, we automatically have $f^*(s^+) \leqslant f^*(s)$. If $f^*(s^+) < f^*(s)$, then

$$s < F\left(f^*(s^+)\right) \leqslant F(f^*(s)) \leqslant s,$$

which gives a contradiction, so $f^*(s) = f^*(s^+)$. □

Remark. We have a jump discontinuity at a point s wherever F is flat, and f^* is flat wherever F has a jump discontinuity.

In particular, when understanding intuition about monotone rearrangement invariant norms, one is allowed to focus on non-increasing, right continuous functions on $(0, \infty)$. For instance, this means that these norms do not care about the number of singularities that a function has, since all these singularities 'pile up' in the decreasing rearrangement.

7.3 Weak Norms

The weak L^p norms are obtained as a slight 'refinement' of the L^p norms.

Theorem 7.10. *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

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

As a special case we find

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

For this to be true, $F(t)$ must tend to zero 'logarithmically faster' than $1/t^p$. Indeed, we find

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

a fact known as *Chebyshev's inequality*. But a bound $F(t) \lesssim 1/t^p$ might be true even if $f \notin L^p(\mathbf{R}^d)$. This leads to the *weak* L^p *norm*, denoted by $\|f\|_{p,\infty}$, which is defined to be the smallest value A such that $F(t) \leqslant (A/t)^p$ for all t. We let $L^{p,\infty}(X)$ denote the space of all functions f for which $\|f\|_{p,\infty} < \infty$. By Chebyshev's inequality, $\|f\|_{p,\infty} \leqslant \|f\|_p$ for any function f. 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

$$G(t) = F(t/|\alpha|) \leqslant \left(\frac{A|\alpha|}{t} \right)^p,$$

so $\|\alpha f\|_{p,\infty} = |\alpha| \|f\|_p$. The weak norms do not satisfy a triangle inequality, but they do satisfy a quasitriangle inequality. This can be proven quite simply from the property that if $f = f_1 + \dots + f_N$, and $\alpha_1, \dots, \alpha_N \in [0,1]$ satisfy $\alpha_1 + \dots + \alpha_N = 1$, then

$$F(t) = F_1(\alpha_1 t) + \dots + F_N(\alpha_N t).$$

Thus if $f = g + h$, then

$$F(t) \leqslant G(t/2) + H(t/2) \leqslant \frac{\|g\|_{p,\infty}^p + \|h\|_{p,\infty}^p}{t^p} \lesssim_p \left(\frac{\|g\|_{p,\infty} + \|h\|_{p,\infty}}{t} \right)^p.$$

Thus $\|f+g\|_{p,\infty} \lesssim \|f\|_{p,\infty} + \|g\|_{p,\infty}$. We can measure the degree to which the weak L^p norm fails to be a norm by determining how much the triangle inequality fails for the sum of N functions, instead of just one function.

Theorem 7.11 (Stein-Weiss Inequality)**.** *Let $f_1, \dots, f_N$ be functions. If $p > 1$, then*

$$\|f_1 + \dots + f_N\|_{p,\infty} \lesssim_p \|f_1\|_{p,\infty} + \dots + \|f_N\|_{p,\infty}.$$

If $p = 1$, then

$$\|f_1 + \dots + f_N\|_{1,\infty} \lesssim \log N \left[\|f_1\|_{1,\infty} + \dots + \|f_N\|_{1,\infty} \right].$$

If $0 < p < 1$, then

$$\|f_1 + \dots + f_N\|_{p,\infty} \lesssim_p \left(\|f_1\|_{p,\infty}^p + \dots + \|f_N\|_{p,\infty}^{1/p} \right)^{1/p}$$

Proof. Begin with the case $p \geqslant 1$. Without loss of generality, assume $\|f_1\|_{p,\infty} + \cdots + \|f_N\|_{p,\infty} = 1$. Fix $t > 0$. For each $k \in [1, N]$, define

$$g_k(x) = \begin{cases} f_k(x) & : |f_k(x)| \geqslant t/2, \\ 0 & : \text{otherwise,} \end{cases}$$

and

$$h_k(x) = \begin{cases} f_k(x) & : |f_k(x)| \leqslant \|f_k\|_{p,\infty} \cdot (t/2), \\ 0 & : \text{otherwise.} \end{cases}$$

Also define $j_k = f_k - g_k - h_k$. Then write $f = f_1 + \cdots + f_N$, $g = g_1 + \cdots + g_N$, $h = h_1 + \cdots + h_N$, and $j = j_1 + \cdots + j_N$. Note that $\|h\|_\infty \leqslant t/2$, so

$$\{x : |f(x)| \geqslant t\} \subset \{x : |g(x)| \geqslant t/4\} \cup \{x : |j(x)| \geqslant t/4\}.$$

Each g_k is supported on a set of measure at most $\|f_k\|_{p,\infty}^p \cdot (2/t)^p$. We conclude that g is supported on a set of measure at most

$$(2/t)^p \sum_{k=1}^N \|f_k\|_{p,\infty}^p \leqslant (2/t)^p.$$

If $p > 1$, then the measure of $\{x : |j(x)| \geqslant t/4\}$ is bounded by

$$\begin{aligned} \frac{4}{t} \int |j(x)|\, dx &\leqslant \frac{4}{t} \sum_{k=1}^N \int |j_k(x)| \\ &= \frac{4}{t} \sum_{k=1}^N \int_{\|f_k\|_{p,\infty}(t/2)}^{t/2} \frac{\|j_k\|_{p,\infty}^p}{s^p}\, ds \\ &= \frac{2^{p+1}}{p-1} \frac{1}{t^p} \sum_{k=1}^N \|j_k\|_{p,\infty}^p \left(\frac{1}{\|f_k\|_{p,\infty}^{p-1}} - 1 \right) \\ &\leqslant \frac{2^{p+1}}{p-1} \frac{1}{t^p} \sum_{k=1}^N \|f_k\|_{p,\infty}^p \left(\frac{1}{\|f_k\|_{p,\infty}^{p-1}} - 1 \right) \\ &\leqslant \frac{2^{p+1}}{p-1} \frac{1}{t^p}. \end{aligned}$$

Thus in total, we conclude the measure of $\{x : |f(x)| \geqslant t\}$ is at most

$$\frac{2^p}{t^p} + \frac{2^{p+1}}{p-1} \frac{1}{t^p} \lesssim_p \frac{1}{t^p}.$$

If $p = 1$, then the measure of $\{x : |j(x)| \geqslant t/4\}$ is bounded

$$\begin{aligned}(4/t)\int |j(x)|\,dx &\leqslant (4/t)\sum_{k=1}^N \int |j_k(x)|\\ &= (4/t)\sum_{k=1}^N \int_{\|f_k\|_{1,\infty}(t/2)}^{t/2} \frac{\|j_k\|_{1,\infty}}{s}\,ds\\ &= (4/t)\sum_{k=1}^N \|f_k\|_{1,\infty}\log(1/\|f_k\|_{1,\infty}).\end{aligned}$$

Now the maximum of $x_1\log(1/x_1) + \cdots + x_N\log(1/x_N)$, subject to the constraint that $x_1 + \cdots + x_N = 1$, is maximized by taking $x_k = 1/N$ for all N, which gives a maximal bound of $\log(N)$. In particular, we find that

$$(2/t)\sum_{k=1}^N \|f_k\|_{1,\infty}\log(1/\|f_k\|_{1,\infty}) \leqslant (2\log N)/t.$$

Thus in total, we conclude the measure of $\{x : |f(x)| \geqslant t\}$ is at most

$$2(1 + \log N)/t \lesssim \log N/t.$$

If $p < 1$, we may assume without loss of generality that

$$\|f_1\|_{p,\infty}^p + \cdots + \|f_N\|_{p,\infty}^p = 1.$$

Then, we perform the same decomposition as before, with functions $\{g_k\}$, $\{h_k\}$, and $\{j_k\}$, defined the same as before, except that

$$h_k(x) = \begin{cases} f_k(x) & : |f_k(x)| \leqslant \|f_k\|_{p,\infty}^p \cdot (t/2),\\ 0 & : \text{otherwise.}\end{cases}$$

The function g_k has support at most $\|f_k\|_{p,\infty}^p \cdot (2/t)^p$, and thus g has total support

$$\sum \|f_k\|_{p,\infty}^p (2/t)^p = (2/t)^p.$$

The measure of $\{x : |j(x)| \geqslant t/4\}$ is bounded by

$$\begin{aligned}
\frac{4}{t}\int |j(x)|\,dx &\leqslant \frac{4}{t}\sum_{k=1}^{N}\int_{\|f_k\|_{p,\infty}^p(t/2)}^{t/2}\frac{\|f_k\|_{p,\infty}^p}{s^p}\,ds\\
&\leqslant \frac{2^{p+1}}{t^p}\frac{1}{1-p}\sum_{k=1}^{N}\|f_k\|_{p,\infty}^{p+p(1-p)}\\
&= \frac{2^{p+1}}{t^p}\frac{1}{1-p}\max\|f_k\|_{p,\infty}^{p(1-p)} \lesssim_p \frac{1}{t^p},
\end{aligned}$$

Combining the two bounds gives that $\|f_1 + \cdots + f_N\|_{p,\infty} \lesssim_p 1$. $\square$

Remark. For $p = 1$, compare this *logarithmic* failure to be a norm with the *polynomial* failure to be a norm found in the norms $\|\cdot\|_p$, when $p < 1$, in Theorem 7.5.

For $p = 1$, the Stein-Weiss inequality is asymptotically tight in N.

Example. *Let $X = \mathbf{R}$. For each k, let*

$$f_k(x) = \frac{1}{|x-k|}.$$

Then $\|f_k\|_{1,\infty} \lesssim 1$ is bounded independantly of k. If $|x| \leqslant N$, there are integers $k_1, \ldots, k_N > 0$ such that $|x - k_i| \leqslant 2i$, so

$$f(x) \geqslant \sum_{i=1}^{N}\frac{1}{|x-k_i|} \geqslant \sum_{i=1}^{N}\frac{1}{2i} \gtrsim \log(N).$$

Thus $\|f\|_{1,\infty} \gtrsim N\log N \gtrsim \log N \sum \|f_k\|_{1,\infty}$.

The weak L^p norms provide another monotone translation invariant norm, and it oftens 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

$$\|f\|_{p,\infty} = \left(\sup_{0\leqslant t<H} Wt^p\right)^{1/p} = W^{1/p}H^p = \|f\|_p.$$

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 $H_1 \leqslant H_2$, then

$$F(t) = \begin{cases} W_1 + W_2 & : t < H_1, \\ W_2 & : t < H_2, \\ 0 & : \textit{otherwise.} \end{cases}$$

Thus

$$\|f\|_{p,\infty} = \left(\max((W_1 + W_2)H_1^p, W_2 H_2^p)\right)^{1/p} = \max((W_1+W_2)^{1/p} H_1, W_2^{1/p} H_2).$$

Example. *The function $f(x) = 1/|x|^s$ does not lie in any $L^p(\mathbf{R}^d)$, but lies in $L^{p,\infty}$ precisely when $p = d/s$, since*

$$|\{1/|x|^{ps} > t\}| = \left|\left\{|x| \leqslant \frac{1}{t^{1/ps}}\right\}\right| \propto_d \frac{1}{t^{d/ps}}.$$

Before we move on, we consider a form of duality for the weak norm, at least when $p > 1$.

Theorem 7.12. *If $p > 1$, and X is σ finite, then*

$$\|f\|_{p,\infty} \sim_p \sup_{|E|<\infty} \frac{1}{|E|^{1-1/p}} \int_E |f(x)|\, dx$$

Proof. Suppose $\|f\|_{p,\infty} < \infty$. If we write $f = \sum f_k$, where $f_k = \mathbf{I}_{F_k} f$, and $F_k = \{x : 2^{k-1} < |f(x)| \leqslant 2^k\}$, then $|F_k| \leqslant \|f\|_{p,\infty}^p 2^{-kp}$. Thus

$$\left|\int_E |f_k(x)|\right| \leqslant 2^k \|f\|_{p,\infty}^p 2^{-kp} = \|f\|_{p,\infty}^p 2^{k(1-p)}.$$

Fix some integer n. Then

$$\begin{aligned} \int_E |f(x)|\, dx &\leqslant \sum_{k=-\infty}^{n-1} \int_E |f_k(x)|\, dx + \sum_{k=n}^{\infty} \int_E |f_k(x)|\, dx \\ &\leqslant |E| 2^{n-1} + \|f\|_{p,\infty}^p \sum_{k=n}^{\infty} 2^{k(1-p)} \\ &\lesssim_p |E| 2^n + \|f\|_{p,\infty}^p 2^{-k(1-p)}. \end{aligned}$$

If we let $2^n \sim \|f\|_{p,\infty} |E|^{1/p}$, then we conclude

$$\int_E |f(x)|\, dx \lesssim_p |E|^{1-1/p} \|f\|_{p,\infty}.$$

Conversely, write

$$A = \sup_{|E|<\infty} \frac{1}{|E|^{1-1/p}} \int_E |f(x)|\, dx/$$

If $G_t = \{x : |f(x)| \geqslant t\}$, then

$$|G_t| \leqslant \frac{1}{t} \int_{G_t} |f(x)|\, dx \leqslant \frac{A|G_t|^{1-1/p}}{t},$$

so

$$|G_t| \leqslant \frac{A^p}{t},$$

which gives $\|f\|_{p,\infty} \leqslant A$. □

For $p \leqslant 1$, the spaces $L^{p,\infty}(X)$ are not normable, as seen by the tightness of the Stein-Weiss inequality. Nonetheless, we still have a certain 'duality' property, that is often useful in the analysis of operators on these spaces. Most useful is it's application when $p = 1$.

Theorem 7.13. *Let $0 < p < \infty$, and let $f \in L^{p,\infty}(X)$, and let $\alpha \in (0,1)$. Then the following are equivalent:*

- $\|f\|_{p,\infty} \lesssim_{\alpha,p} A$.
- *For any set $E \subset X$ with finite measure, there is $E' \subset E$ with $|E'| \geqslant \alpha|E|$ such that*
$$\int_{E'} |f(x)|\, dx \lesssim_{\alpha,p} A|E'|^{1-1/p}.$$

Proof. By homogeneity, assume $\|f\|_{p,\infty} \leqslant 1$, so that if F is the distribution of f, $F(t) \leqslant 1/t^p$. If $|E| = (1-\alpha)^{-1}/t_0^p$, and we set

$$E' = \{x : |f(x)| \leqslant t_0\},$$

then

$$|E'| \geqslant |E| - F(t_0) = \frac{(1-\alpha)^{-1} - 1}{t_0^p} = \alpha|E|,$$

and

$$\int_{E'} |f(x)| \leqslant t_0 |E'| \lesssim_\alpha |E'|^{1-1/p}.$$

Conversely, suppose Property (2) holds. For each k, set

$$E_k = \{x : 2^k \leqslant |f(x)| < 2^{k+1}\}.$$

Then there exists E_k' with $|E_k'| \geqslant \alpha |E_k|$ and

$$\int_{E_k'} |f(x)|\, dx \leqslant |E_k'|^{1-1/p}$$

On the other hand,

$$\int_{E_k'} |f(x)|\, dx \geqslant 2^k |E_k'|.$$

Rearranging this equation gives $|E_k'| \leqslant 2^{-pk}$, and so $|E_k| \lesssim_\alpha 2^{-pk}$. But this means

$$F(2^N) = \sum_{k=N}^\infty |E_k| \lesssim_{\alpha,p} 2^{-Np},$$

and this implies $\|f\|_{p,\infty} \lesssim_{\alpha,p} 1$. □

7.4 Lorentz Spaces

Recall that we can write

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

Thus $F(t)t^p$ is integrable with respect to the Haar measure on $\mathbf{R}^+$. But if we change the integrality condition to the condition that $F(t)t^p \in L^q(\mathbf{R}^+)$ for some $0 < q \leqslant \infty$, we obtain a different integrability condition, giving rise to a monotone, translation-invariant norm. Thus leads us to the definition of the *Lorentz norms*. For each $0 < p, q < \infty$, we define the Lorentz norm

$$\|f\|_{p,q} = p^{1/q} \|tF^{1/p}\|_{L^q(\mathbf{R}^+)}$$

The *Lorentz space* $L^{p,q}(X)$ as the space of functions f with $\|f\|_{p,q} < \infty$. We can define the norm in terms of f^* as well.

Lemma 7.14. *For any measurable $f : X \to \mathbf{R}$, $\|f(t)\|_{p,q} = \|s^{1/p} f^*(s)\|_{L^q(\mathbf{R}^+)}$.*

Proof. First, assume f^* has non-vanishing derivative on $(0,\infty)$, and that f is bounded, with finite support. An integration by parts gives

$$\|f\|_{p,q} = p^{1/q} \left(\int_0^\infty t^{q-1} F(t)^{q/p}\, dt \right)^{1/q} = \left(\int_0^\infty t^q F(t)^{q/p-1} (-F'(t))\, dt \right)^{1/q}.$$

If we set $s = F(t)$, then $f^*(s) = t$, and $ds = F'(t)dt$, and so

$$\left(\int_0^\infty t^q F(t)^{q/p-1} F'(t)\, dt \right)^{1/q} = \left(\int_0^\infty f^*(s)^q s^{q/p-1} ds \right)^{1/q} = \|s^{1/p} f^*\|_{L^q(\mathbf{R}^+)}.$$

This gives the result in this case. The general result can then be obtained by applying the monotone convergence theorem to an arbitrary f^* with respect to a family of smooth functions. $\square$

The definition of the Lorentz space may seem confusing, but we really only require various special cases in most applications. Aside from the weak L^p norms $\|\cdot\|_{p,\infty}$ and the L^p norms $\|\cdot\|_p = \|\cdot\|_{p,p}$, the $L^{p,1}$ norms and $L^{p,2}$ norms also occur, the first, because of the connection with integrability, and the second because we may apply orthogonality techniques. As $q \to 0$, the norms $\|\cdot\|_{p,q}$ give stronger control over the function f.

Theorem 7.15. *For $q < r$, $\|f\|_{p,r} \lesssim_{p,q,r} \|f\|_{p,q}$.*

Proof. First we treat the case $r = \infty$. We have

$$\begin{aligned} s_0^{1/p} f^*(s_0) &= \left((p/q) \int_0^{s_0} [s^{1/p} f^*(s_0)]^q \frac{ds}{s} \right)^{1/q} \\ &\leqslant \left((p/q) \int_0^{s_0} [s^{1/p} f^*(s)]^q \frac{ds}{s} \right) \\ &\leqslant (p/q)^{1/q} \|f\|_{p,q}. \end{aligned}$$

When $r < \infty$, we can interpolate, calculating

$$\begin{aligned} \|f\|_{p,r} &= \left(\int_0^\infty [s^{1/p} f^*(s)]^r \frac{ds}{s} \right)^{1/r} \\ &\leqslant \|f\|_{p,\infty}^{1-q/r} \|f\|_{p,q}^{q/r} \leqslant (p/q)^{p(1/q-1/r)} \|f\|_{p,q}. \end{aligned}$$ $\square$

The fact that multiplying a function by a constant dilates the distribution implies that the Lorentz norm is homogeneous. We do not have a triangle inequality for the Lorentz norms, but we have a quasi triangle inequality.

Theorem 7.16. *For each $p, q > 0$, $\|f_1 + f_2\|_{p,q} \lesssim_{p,q} \|f_1\|_p + \|f_2\|_q$.*

Proof. We calculate that if $g = f_1 + f_2$,

$$\begin{aligned}
\|g\|_{p,q} &= \left(q \int_0^\infty \left[tG(t)^{1/p} \right]^q \frac{dt}{t} \right)^{1/q} \\
&\leqslant \left(q \int_0^\infty \left[t(F_1(t/2) + F_2(t/2))^{1/p} \right]^q \frac{dt}{t} \right)^{1/q} \\
&\lesssim \left(q \int_0^\infty \left[t\left(F_1(t) + F_2(t)\right)^{1/p} \right]^q \frac{dt}{t} \right)^{1/q} \\
&\lesssim_p \left(q \int_0^\infty t^q \left(F_1(t)^{q/p} + F_2(t)^{q/p} \right) \frac{dt}{t} \right)^{1/q} \\
&\lesssim_q \left(q \int_0^\infty t^q F_1(t)^{q/p} \frac{dt}{t} \right)^{1/q} + \left(q \int_0^\infty t^q F_2(t)^{q/p} \frac{dt}{t} \right)^{1/q} \\
&= \|f_1\|_{p,q} + \|f_2\|_{p,q}.
\end{aligned}$$

□

7.5 Dyadic Layer Cake Decompositions

An important trick to utilizing Lorentz norms is by utilizing a dyadic 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. We say f is a *sub step function* with height H and width W if f is supported on a set E with $|E| \leqslant W$, and $|f(x)| \leqslant H$. A *quasi step function* with height H and width W if f is supported on a set E with $|E| \sim W$ and on E, $|f(x)| \sim H$.

Remark. It might seem that sub step functions of height H and width W can take on a great many different behaviours, rather than that of a step function with height H and width W. However, from the point of view of monotone, translation invariant norms, this isn't so. This is because using the binary expansion of real numbers, for every sub-step function f of height H and width W, we can find sets $\{E_k\}$ such that

$$f(x) = H \sum_{k=1}^{\infty} 2^{-k} \mathbf{I}_{E_k},$$

where $|E_k| = 1$. Thus bounds on step functions that are stable under addition tend to automatically imply bounds on substep functions.

We start by discussing the *vertical dyadic layer cake decomposition*. We define, for each $k \in \mathbf{Z}$,

$$f_k(x) = f(x)\mathbf{I}(2^{k-1} < |f(x)| \leqslant 2^k)$$

Then we set $f = \sum f_k$. Each f_k is a quasi step function with height 2^k and width $F(2^{k-1}) - F(2^k)$. We can also perform a *horizontal layer cake decomposition*. If we define $H_k = f^*(2^k)$, and set

$$f_k(x) = f(x)\mathbf{I}(H_{k-1} < |f(x)| \leqslant H_k),$$

then f_k is a substep function with height H_k and width 2^k. These decompositions are best visualized with respect to the representation f^* of f, in which case the decomposition occurs over particular intervals.

Theorem 7.17. *The following values $A_1, \dots, A_4$ are all comparable up to absolute constant depending only on p and q:*

1. $\|f\|_{p,q} \leqslant A_1$.

2. *We can write $f = \sum_{k \in \mathbf{Z}} f_k$, where f_k is a quasi-step function with height 2^k and width W_k, and*

$$\left(\sum_{k \in \mathbf{Z}} \left[2^k W_k^{1/p} \right]^q \right)^{1/q} \leqslant A_2.$$

3. *We can write $f = \sum_{k \in \mathbf{Z}} f_k$, where f_k is a sub-step function with height 2^k and width W_k, and*

$$\left(\sum_{k \in \mathbf{Z}} \left[2^k W_k^{1/p} \right]^q \right)^{1/q} \leqslant A_3.$$

4. *We can write $f(x) = \sum_{k \in \mathbf{Z}} f_k$, where f_k is a sub-step function with width 2^k and height H_k, where $\{H_k\}$ is a decreasing family of functions, and*

$$\left(\sum_{k \in \mathbf{Z}} \left[H_k 2^{k/p} \right]^q \right)^{1/q} \leqslant A_4.$$

Proof. It is obvious that we can always select $A_3 \leqslant A_2$. 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 $W_k = F(2^{k-1}) - F(2^k) \leqslant F(2^{k-1})$, and so

$$\begin{aligned} \sum_{k \in \mathbf{Z}} [2^k W_k^{1/p}]^q &\leqslant \sum_{k \in \mathbf{Z}} [2^k F(2^{k-1})^{1/p}]^q \\ &\lesssim_q \sum_{k \in \mathbf{Z}} [2^{k-1} F(2^k)^{1/p}]^q \\ &\lesssim \sum_{k \in \mathbf{Z}} \int_{2^{k-1}}^{2^k} [tF(t)^{1/p}]^q \frac{dt}{t} \lesssim_q \|f\|_{p,q}^q \leqslant A_1^q. \end{aligned}$$

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 $W_k \leqslant 2^k$, and $H_{k+1} \leqslant |f_k(x)| \leqslant H_k$. Then a telescoping sum shows

$$\begin{aligned} H_k 2^{k/p} &= \left(\sum_{m=0}^{\infty} (H_{k+m}^q - H_{k+m+1}^q) 2^{kq/p} \right)^{1/q} \\ &\lesssim_q \left(\sum_{m=0}^{\infty} \int_{H_{k+m+1}}^{H_{k+m}} [t 2^{k/p}]^q \frac{dt}{t} \right)^{1/q} \\ &\leqslant \left(\sum_{m=0}^{\infty} \int_{H_{k+m+1}}^{H_{k+m}} [tF(t)^{1/p}]^q \frac{dt}{t} \right)^{1/q} \end{aligned}$$

Thus

$$\left(\sum_{k \in \mathbf{Z}} [H_k 2^{k/p}]^q \right)^{1/q} \leqslant \left(\int_0^{\infty} [tF(t)^{1/p}]^q \frac{dt}{t} \right)^{1/q} \lesssim_q A_1.$$

Thus $A_4 \lesssim_q A_1$. It remains to bound A_1 by A_4 and A_3. Given A_3, we can write $|f(x)| \leqslant \sum 2^k \mathbf{I}_{E_k}$, where $|E_k| \leqslant W_k$. We then find

$$F(2^k) \leqslant \sum_{m=1}^{\infty} W_{k+m}.$$

Thus

$$\int_{2^{k-1}}^{2^k} [tF(t)^{1/p}]^q \frac{dt}{t} \lesssim \left[2^k \left(\sum_{m=0}^{\infty} W_k \right)^{1/p} \right]^q.$$

Thus if $q \leqslant p$,

$$\begin{aligned} \|f\|_{p,q} &\lesssim_q \left(\sum_{k \in \mathbf{Z}} \left[2^k \left(\sum_{m=0}^{\infty} W_{k+m} \right)^{1/p} \right]^q \right)^{1/q} \\ &\leqslant \left(\sum_{k \in \mathbf{Z}} \sum_{m=0}^{\infty} \left[2^k W_{k+m}^{1/p} \right]^q \right)^{1/q} \\ &\leqslant \left(\sum_{m=0}^{\infty} 2^{-qm} \sum_{k \in \mathbf{Z}} \left[2^{k+m} W_{k+m}^{1/p} \right]^q \right)^{1/q} \\ &\leqslant \left(A_3^q \sum_{m=0}^{\infty} 2^{-mq} \right)^{1/q} \lesssim_q A_3. \end{aligned}$$

If $q \geqslant p$, we can employ the triangle inequality for $l^{q/p}$ to write

$$\begin{aligned} \|f\|_{p,q} &\lesssim_q \left(\sum_{k \in \mathbf{Z}} \left[2^k \left(\sum_{m=0}^{\infty} W_{k+m} \right)^{1/p} \right]^q \right)^{1/q} \\ &\leqslant \left(\sum_{m=0}^{\infty} \left(\sum_{k \in \mathbf{Z}} 2^{kq} W_{k+m}^{q/p} \right)^{p/q} \right)^{1/p} \\ &\leqslant \left(A_3^p \sum_{m=0}^{\infty} 2^{-mq} \right)^{1/p} \lesssim_{p,q} A_3. \end{aligned}$$

The bound of A_1 in terms of A_4 involves the same ‘shifting’ technique, and is left to the reader. $\square$

Remark. Heuristically, the theorem above says that if $f = \sum_{k \in \mathbf{Z}} f_k$, where f_k is a quasi-step function with width H_k and width W_k, and if either $\{H_k\}$ and $\{W_k\}$ grow faster than powers of two, then

$$\|f\|_{p,q} \sim_{p,q} \left(\sum_{k \in \mathbf{Z}} \left[H_k W_k^{1/p} \right]^q \right)^{1/q}.$$

Thus the $L^{p,q}$ norm has little interaction between elements of the sum when the sum occurs over dyadically different heights or width. This is one reason why we view the q parameter as a 'logarithmic' correction of the L^p norm. In particular, if we can write $f = f_1 + \dots + f_N$, and $q_1 < q_2$, then the last equation, combined with a l^{q_1} to l^{q_2} norm bound, gives

$$\left(\sum_{k \in \mathbf{Z}} \left[H_k W_k^{1/p} \right]^{q_1} \right)^{1/q_1} \leqslant N^{1/q_1 - 1/q_2} \left(\sum_{k \in \mathbf{Z}} \left[H_k W_k^{1/p} \right]^{q_2} \right)^{1/q_2}$$

This implies

$$\|f\|_{p,q_2} \lesssim_{p,q_1,q_2} \|f\|_{p,q_1} \lesssim_{p,q_1,q_2} N^{1/q_1 - 1/q_2} \|f\|_{p,q_2}.$$

In particular, this occurs if there exists a constant C such that $C \leqslant |f(x)| \leqslant C \cdot 2^N$ for all x. On the other hand, if we vary the p parameter, we find that for $p_1 < p_2$,

$$\left(\sum_{k \in \mathbf{Z}} \left[H_k W_k^{1/p_1} \right]^q \right)^{1/q} \leqslant \max(W_k)^{1/p_1 - 1/p_2} \left(\sum_{k \in \mathbf{Z}} \left[H_k W_k^{1/p_2} \right]^q \right)^{1/q},$$

$$\left(\sum_{k \in \mathbf{Z}} \left[H_k W_k^{1/p_2} \right]^q \right)^{1/q} \leqslant \left(\frac{1}{\min(W_k)} \right)^{1/p_1 - 1/p_2} \left(\sum_{k \in \mathbf{Z}} \left[H_k W_k^{1/p_2} \right]^q \right)^{1/q}.$$

which gives

$$\min(W_k)^{1/p_1 - 1/p_2} \|f\|_{p_2,q} \lesssim_{p_1,p_2,q} \|f\|_{p_1,q} \lesssim_{p_1,p_2,q} \max(W_k)^{1/p_1 - 1/p_2} \|f\|_{p_2,q}.$$

Both of these inequalities can be tight. Because of the dyadic decomposition of f, we find $\max(W_k) \geqslant 2^N \min(W_k)$, so these two norms can differ by at least $2^{N(1/p_1 - 1/p_2)}$, and at *most* if the f_k occur over consecutive dyadic values, which is *exponential* in N. Conversely, if the heights change dyadically, we find that

$$\left(\sum_{k \in \mathbf{Z}} \left[H_k W_k^{1/p_2} \right]^q \right)^{1/q} \leqslant \left(\sum_{k \in \mathbf{Z}} \left[H_k W_k^{1/p_2} \right]^{qp_2/p_1} \right)^{(p_1/p_2)/q}$$

$$\leqslant \max(H_k)^{1-p_1/p_2} \left(\sum_{k \in \mathbf{Z}} \left[H_k W_k^{1/p_1} \right]^q \right)^{(p_1/p_2)/q}$$

$$\left(\sum_{k \in \mathbf{Z}} \left[H_k W_k^{1/p_1} \right]^q \right)^{1/q} \lessapprox \left(\sum_{k \in \mathbf{Z}} \left[H_k W_k^{1/p_1} \right]^{qp_1/p_2} \right)^{(p_2/p_1)/q}$$

$$\leqslant \left(\frac{1}{\min(H_k)} \right)^{p_2/p_1 - 1} \left(\sum_{k \in \mathbf{Z}} \left[H_k W_k^{1/p_2} \right]^q \right)^{(p_2/p_1)/q}$$

where $\lessapprox$ denotes a factor ignoring polynomial powers of N occuring from the estimate. Thus

$$\min(H_k)^{p_2-p_1} \|f\|_{p_1,q}^{p_1} \lessapprox_{p_1,p_2,q} \|f\|_{p_2,q}^{p_2} \lesssim_{p_1,p_2,q} \max(H_k)^{p_2-p_1} \|f\|_{p_1,q}^{p_1}$$

again, these inequalities can be both tight, and $\max(H_k) \geqslant 2^N \min(H_k)$, with equality if the quasi step functions from which f is composed occur consecutively dyadically.

Example. *Consider the function $f(x) = |x|^{-s}$. For each k, let*

$$E_k = \{x : 2^{-(k+1)/s} \leqslant |x| < 2^{-k/s}\}$$

and define $f_k = f \mathbf{I}_{E_k}$. Then f_k is a quasi-step function with height 2^k, and width $1/2^{dk/s}$. We conclude that if $p = d/s$, and $q < \infty$,

$$\|f\|_{p,q} \sim_{p,q,d} \left(\sum_{k=-\infty}^{\infty} 2^{qk(1-d/ps)} \right)^{1/q} = \infty.$$

Thus the function f lies exclusively in $L^{p,\infty}(\mathbf{R}^d)$.

A simple consequence of the layer cake decomposition is Hölder's inequality for Lorentz spaces.

Theorem 7.18. *If* $0 < p_1, p_2, p < \infty$ *and* $0 < q_1, q_2, q < \infty$ *with*

$$1/p = 1/p_1 + 1/p_2 \quad \textit{and} \quad 1/q = 1/q_1 + 1/q_2,$$

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 $|f| \leqslant \sum_{k\in\mathbf{Z}} H_k \mathbf{I}_{E_k}$ and $|g| \leqslant \sum_{k\in\mathbf{Z}} H'_k \mathbf{I}_{F_k}$, where $|E_k|, |F_k| \leqslant 2^k$. Then

$$|fg| \leqslant \sum_{k,k'\in\mathbf{Z}} H_k H'_k \mathbf{I}_{E_k\cap F_{k'}}$$

For each fixed k, $|E_{k+m} \cap F_m| \leqslant 2^m$, and so

$$\begin{aligned}
\left\| \sum_{m\in\mathbf{Z}} H_{k+m} H'_m \mathbf{I}_{E_{k+m}\cap F_m} \right\|_{p,q} &\lesssim_{p,q} \left(\sum_{m\in\mathbf{Z}} [H_{k+m} H'_m 2^{m/p}]^q \right)^{1/q} \\
&= \left(\sum_{m\in\mathbf{Z}} \left[(H_{k+m} 2^{m/p_1})(H_m 2^{m/p_2}) \right]^q \right)^{1/q} \\
&\leqslant \left(\sum_{m\in\mathbf{Z}} [H_{k+m} 2^{m/p_1}]^{q_1} \right)^{1/q_1} \left(\sum_{m\in\mathbf{Z}} [H'_m 2^{m/p_2}]^{q_2} \right)^{1/q_2} \\
&\lesssim_{p,q,p_1,q_1,p_2,q_2} 2^{-k/p_1}
\end{aligned}$$

Summing over $k > 0$ gives that

$$\left\| \sum_{k\geqslant 0} \sum_{m\in\mathbf{Z}} H_{k+m} H'_m \mathbf{I}_{E_{k+m}\cap F_m} \right\| \lesssim_{p,q,p_1,q_1,p_2,q_2} 1$$

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 similarily, but using the bound $|E_{k+m} \cap F_m| \leqslant 2^{k+m}$ instead of the other bound. ☐

Corollary 7.19. *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

$$\|f\|_{L^1(E)} = \|\mathbf{I}_E f\|_{L^1(X)} \lesssim_{p,q} |E|^{1-1/p}\|f\|_{p,q} < \infty. \qquad \square$$

Finally, we consider the duality of the $L^{p,q}$ norms. If $1 < p < \infty$, and $1 < q < \infty$, then $L^{p,q}(X)^* = L^{p',q'}(X)$. When $q = 1$ or $q = \infty$, things are more complex, but the following theorem often suffices. When $p = 1$, things get more tricky, so we leave this case out.

Theorem 7.20. *Let $1 < p < \infty$ and $1 \leqslant q < \infty$. Then if $f \in L^{p,q}(X)$,*

$$\|f\|_{p,q} \sim \sup\left\{ \int fg : \|g\|_{p',q'} \leqslant 1 \right\}.$$

Proof. Without loss of generality, we may assume $\|f\|_{p,q} = 1$. We may perform a vertical layer cake decomposition, writing $f = \sum_{k \in \mathbf{Z}} f_k$, where $2^{k-1} \leqslant |f_k(x)| \leqslant 2^k$, 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 $g = \sum_{k \in \mathbf{Z}} g_k$, where $g_k(x) = a_k^{q-p}\operatorname{sgn}(f_k(x))|f_k(x)|^{p-1}$. Then

$$\begin{aligned} \int f(x)g(x) &= \sum_{k \in \mathbf{Z}} \int f_k(x) g_k(x) = \sum_{k \in \mathbf{Z}} a_k^{q-p} \int |f_k(x)|^p \\ &\gtrsim_p \sum_{k \in \mathbf{Z}} a_k^{q-p} W_k 2^{kp} = \sum_{k \in \mathbf{Z}} a_k^q \gtrsim_{p,q} 1. \end{aligned}$$

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. The 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

$$H_k = \sup_{l \geqslant 0} a_{k-l}^{q-p} 2^{kp} 2^{-lp/2}.$$

Then $|g_k(x)| \lesssim_{p,q} H_k$, and $H_{k+1} \geqslant 2^{p/2} H_k$. In particular, if we pick m such that $2^{mp/2} \geqslant 1$, then for any $l \leqslant m$, 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

$$\begin{aligned}
\|g\|_{p',q'} &\lesssim_{m,p,q} \left(\sum [H_k W_k^{1/p'}]^{q'} \right)^{1/q'} \\
&\lesssim \left(\sum_{k \in \mathbf{Z}} \left[\left(\sup_{l \geqslant 0} a_{k-l}^{q-p} 2^{kp} 2^{-lp/2} \right) (a_k 2^{-k})^{p-1} \right]^{q'} \right)^{1/q'} \\
&\lesssim_p \left(\sum_{k \in \mathbf{Z}} \left[a_k^{p-1} \sum_{l=0}^{\infty} a_{k-l}^{q-p} 2^{-lp/2} \right]^{q'} \right)^{1/q'} \\
&\lesssim \sum_{l=0}^{\infty} 2^{-lp/2} \left(\sum_{k \in \mathbf{Z}} \left[a_k^{p-1} a_{k-l}^{q-p} \right]^{q'} \right)^{1/q'}.
\end{aligned}$$

Applying's Hölder's inequality shows

$$\begin{aligned}
\left(\sum_{k \in \mathbf{Z}} \left[a_k^{p-1} a_{k-l}^{q-p} \right]^{q'} \right)^{1/q'} &\leqslant \left(\sum_{k \in \mathbf{Z}} a_k^q \right)^{(p-1)/q} \left(\sum_{k \in \mathbf{Z}} a_{k-l}^q \right)^{(q-p)/q} \\
&\lesssim_{p,q} \|f\|_{p,q}^{q-1} \lesssim_{p,q} 1. \qquad \square
\end{aligned}$$

Remark. This technique shows that if $f = \sum f_k$, where f_k is a quasi-step function with measure W_k and height 2^{ck}, then we can find m such that $cm > 1$, and then consider the m functions $f^1, \dots, f^m$, where $f_i = \sum f_{km+i}$. Then the functions f_{km+i} have heights which are separated by powers of two, and so the quasi-triangle inequality implies

$$\begin{aligned}
\|f\|_{p,q} &\lesssim_m \sum_{i=1}^{m} \|f^i\|_{p,q} \\
&\lesssim_{p,q} \sum_{i=1}^{m} \left(\sum \left[H_{km+i} W_{km+i}^{1/p} \right]^q \right)^{1/q} \\
&\lesssim_m \left(\sum \left[H_k W_k^{1/p} \right]^q \right)^{1/q}
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\|f\|_{p,q} &\gtrsim \max_{1\leqslant i\leqslant m} \|f^i\|_{p,q} \\
&\sim \max_{1\leqslant i\leqslant m} \left(\sum \left[H_{km+i} W_{km+i}^{1/p}\right]^q\right)^{1/q} \\
&\gtrsim_m \left(\sum \left[H_k W_k^{1/p}\right]^q\right)^{1/q}.
\end{aligned}$$

Thus the dyadic layer cake decomposition still works in this setting.

We remark that if $1 < p < \infty$ and $1 \leqslant q \leqslant \infty$, then for each $f \in L^{p,q}$, the value

$$\sup\left\{\int f g : \|g\|_{p',q'} \leqslant 1\right\}$$

gives a norm on $L^{p,q}(X)$ which is comparable with the $L^{p,q}$ norm. In particular, this implies that for $p > 1$ and $q \geqslant 1$,

$$\|f_1 + \cdots + f_N\|_{p,q} \lesssim_{p,q} \|f_1\|_{p,q} + \cdots + \|f_N\|_{p,q},$$

so that the triangle inequality has constants independent of N. We can also use a layer cake decomposition to get a version of the Stein-Weiss inequality for Lorentz norms.

Theorem 7.21. *For each $1 < q < \infty$, there is $\alpha(q) > 0$ such that for any functions $f_1, \dots, f_N$,*

$$\|f_1 + \cdots + f_N\|_{1,q} \lesssim (\log N)^{\alpha(q)} \left(\|f_1\|_{1,q} + \cdots + \|f_N\|_{1,q}\right).$$

Proof. For values A and B in this argument, we write $A \lessapprox B$ if there exists α such that $A \lesssim (\log N)^\alpha B$. Given $f_1, \dots, f_N$, write $f_i = \sum_{j=-\infty}^{\infty} f_{ij}$, where f_{ij} has width W_{ij} and height 2^j. If we assume, without loss of generality, that $\|f_1\|_{1,q} + \cdots + \|f_N\|_{1,q} = 1$, then

$$\sum_{i=1}^{N} \left(\sum_{j=-\infty}^{\infty} (2^j W_{ij})^q\right)^{1/q} \lesssim_q 1$$

Thus we want to show $\|f_1 + \cdots + f_N\|_{1,q} \lessapprox_q 1$. Our first goal is to upper bound the measure of the set

$$E = \{x : 2^{k-1} < |f_1(x) + \cdots + f_N(x)| \leqslant 2^k\}$$

The measure of the set E is upper bounded by the measure of the set

$$E' = \left\{ x : 2^{k-2} < \left| \sum_{j=k-\lg(N)}^{k} f_{1j}(x) + \dots + f_{Nj}(x) \right| \leqslant 2^{k+1} \right\}$$

Applying the usual Stein-Weiss inequality, we have

$$\left\| \sum_{i=1}^{N} \sum_{j=k-\lg N}^{k} f_{ij} \right\|_{1,\infty} \lessapprox \sum_{i=1}^{N} \sum_{j=k-\lg N}^{k} \|f_{ij}\|_{1,\infty} \lesssim \sum_{i=1}^{N} \sum_{j=k-\lg N}^{k} \|f_{ij}\|_{1,\infty} \lesssim_q \sum_{i=1}^{N} \sum_{j=k-\lg N}^{k} W_{ij} 2^j$$

Thus we conclude

$$|E'| \lessapprox_q 2^{-k} \sum_{i=1}^{N} \sum_{j=k-\lg N}^{k} W_{ij} 2^j$$

This implies that

$$\|f_1 + \dots + f_N\|_{1,q} \lessapprox_q \left(\sum_{k=-\infty}^{\infty} \left(\sum_{i=1}^{N} \sum_{j=k-\lg N}^{k} W_{ij} 2^j \right)^q \right)^{1/q}.$$

Applying Minkowski's inequality, we conclude

$$\begin{aligned}
\left(\sum_{k=-\infty}^{\infty} \left(\sum_{i=1}^{N} \sum_{j=k-\lg N}^{k} W_{ij} 2^j \right)^q \right)^{1/q} &\lesssim \sum_{i=1}^{N} \left(\sum_{k=-\infty}^{\infty} \left(\sum_{j=k-\lg N}^{k} W_{ij2^j} \right)^q \right)^{1/q} \\
&\lessapprox \sum_{i=1}^{N} \left(\sum_{k=-\infty}^{\infty} \sum_{j=k-\lg N}^{k} W_{ij}^q 2^{qj} \right)^{1/q} \\
&\lessapprox \sum_{i=1}^{N} \left(\sum_{j=-\infty}^{\infty} W_{ij}^q 2^{qj} \right)^{1/q} \lesssim 1. \qquad \square
\end{aligned}$$

7.6 Mixed Norm Spaces

Given two measure spaces X and Y, we can form the product measure space $X \times Y$. If we have a norm space V of functions on X, with norm $\|\cdot\|_V$

and a norm space W of functions on Y, with norm $\|\cdot\|_W$, we can consider a 'product norm'; for each function f on $X \times Y$, we can consider the function $y \mapsto \|f(\cdot,y)\|_V$, and take the norm of this function over Y, i.e. $\|\|f(\cdot,y)\|_V\|_W$. The most important case of this process is where we fix $0 < p,q \leqslant \infty$, and consider

$$\|f\|_{L^p(X)L^q(Y)} = \left(\int \left(\int |f(x,y)|^p \; dx \right)^{q/p} \; dy \right)^{1/q}.$$

Similarly, we can define $\|f\|_{L^q(Y)L^p(X)}$. Of course, in the case where Fubini's theorem can apply, if $p = q$ and does not equal ∞, we have

$$\|f\|_{L^p(X)L^p(Y)} = \|f\|_{L^p(Y)L^p(X)} = \|f\|_{L^p(X \times Y)}.$$

If $p = q = \infty$, then

More generally, the biggest norm is always obtained with the largest exponents on the inside.

Theorem 7.22. *If $q \geqslant p$, $\|f\|_{L^p(X)L^q(Y)} \leqslant \|f\|_{L^q(Y)L^p(X)}$.*

Proof. We apply complex interpolation. □

7.7 Orlicz Spaces

To develop the class of Orlicz spaces, we note that if $\|f\|_p \leqslant 1$, and we set $\Phi(t) = t^p$, then

$$\int \Phi(|f(x)|) \; dx = 1.$$

More generally, given any function $\Phi : [0,\infty) \to [0,\infty)$, we might ask if we can define a norm $\|\cdot\|_\Phi$ such that if $\|f\|_\Phi \leqslant 1$, then

$$\int \Phi(|f(x)|) \; dx = 1.$$

Since a norm would be homogenous, this would imply that if $\|f\|_\Phi \leqslant A$, then

$$\int \Phi\left(\frac{|f(x)|}{A}\right) \; dx \leqslant 1.$$

If we want these norms to be monotone, we might ask that if $A < B$, then

$$\int \Phi\left(\frac{|f(x)|}{B}\right) \; dx \leqslant \int \Phi\left(\frac{|f(x)|}{A}\right),$$

and the standard way to ensure this is to ask the Φ is an increasing function. To deal with the property that $\|0\| = 0$, we set $\Phi(0) = 0$. In order for $\|\cdot\|_\Phi$ to be a norm, the set of functions $\{f : \|f\|_\Phi \leqslant 1\}$ needs to be convex, and the standard way to obtain this is to assume that Φ is convex.

In short, we consider an increasing, convex function Φ with $\Phi(0) = 0$. We then define

$$\|f\|_\Phi = \inf\left\{A > 0 : \int \Phi\left(\frac{|f(x)|}{A}\right)\, dx \leqslant 1\right\}.$$

This function is a norm on the space of all f with $\|f\|_\Phi < \infty$. It is easy to verify that $\|f\|_\Phi = 0$ if and only if $f = 0$ almost everywhere, and that $\|\alpha f\|_\Phi = |\alpha|\|f\|_\Phi$. To justify the triangle inequality, we note that if

$$\int \Phi\left(\frac{|f(x)|}{A}\right) \leqslant 1 \quad \text{and} \quad \int \Phi\left(\frac{|f(x)|}{B}\right) \leqslant 1,$$

then applying convexity gives

$$\begin{aligned}\int \Phi\left(\frac{|f(x)+g(x)|}{A+B}\right) &\leqslant \int \Phi\left(\frac{|f(x)|+|g(x)|}{A+B}\right)\\ &\leqslant \int \left(\frac{A}{A+B}\right)\Phi\left(\frac{|f(x)|}{A}\right) + \left(\frac{B}{A+B}\right)\Phi\left(\frac{|g(x)|}{B}\right) \leqslant 1.\end{aligned}$$

Thus we obtain the triangle inequality.

The spaces $L^p(X)$ for $p \in [1,\infty)$ are Orlicz spaces with $\Phi(t) = t^p$. The space $L^\infty(X)$ is not really an Orlicz space, but it can be considered as the Orlicz function with respect to the 'convex' function

$$\Phi(t) = \begin{cases} \infty & t > 1, \\ t & t \leqslant 1. \end{cases}$$

More interesting examples of Orlicz spaces include

- $L \log L$, given by the Orlicz norm induced by $\Phi(t) = t\log(2+t)$.
- e^L, defined with respect to $\Phi(t) = e^t - 1$.
- e^{L^2}, defined with respect to $\Phi(t) = e^{t^2} - 1$.

One should not think too hard about the constants in the functions defined above, which are included to make $\Phi(0) = 0$. When we are dealing with a finite measure space, they are irrelevant.

Lemma 7.23. *If $\Phi(x) \lesssim \Psi(x)$ for all x, then $\|f\|_{\Phi(L)} \lesssim \|f\|_{\Psi(L)}$. If X is finite, and $\Phi(x) \lesssim \Psi(x)$ for sufficiently large x, then $\|f\|_{\Phi(L)} \lesssim \|f\|_{\Psi(L)}$.*

Proof. The first proposition is easy, and we now deal with the finite case. We note that the condition implies that for each $\varepsilon > 0$, there exists C_ε such that $\Phi(x) \leqslant C_\varepsilon \Psi(x)$ if $|x| \geqslant \varepsilon$. Assume that $\|f\|_{\Psi(L)} \leqslant 1$, so that

$$\int \Psi(|f(x)|)\, dx \leqslant 1.$$

Then convexity implies that for each $A > 0$,

$$\int \Psi\left(\frac{|f(x)|}{A}\right) \leqslant \frac{1}{A}.$$

Thus

$$\begin{aligned} \int \Phi\left(\frac{|f(x)|}{A}\right) dx &\leqslant \Phi(\varepsilon)|X| + C_\varepsilon \int \Psi\left(\frac{|f(x)|}{A}\right) \\ &\lesssim \Phi(\varepsilon)|X| + \frac{C_\varepsilon}{A}. \end{aligned}$$

If $\Phi(\varepsilon) \leqslant 2/|X|$, and $A \geqslant 2C_\varepsilon$, then we conclude that

$$\int \Phi\left(\frac{|f(x)|}{A}\right) dx \leqslant 1.$$

Thus $\|f\|_{\Phi(L)} \lesssim 1$. $\square$

The Orlicz spaces satisfy an interesting duality relation. Given a function Φ, which we assume is *superlinear*, in the sense that $\Phi(x)/x \to \infty$ as $x \to \infty$, define it's *Young dual*, for each $y \in [0,\infty)$, by

$$\Psi(y) = \sup\{xy - \Phi(x) : x \in [0,\infty)\}.$$

Then Ψ is the smallest function such that $\Phi(x) + \Psi(y) \geqslant xy$ for each x, y. This quantity is finite for each y because Φ is superlinear; for each $y \geqslant 0$, there exists $x(y)$ such that $\Phi(x(y)) \geqslant xy$, and thus the maximum of

$xy - \Phi(x)$ is attained for $x \leqslant x(y)$. In particular, since Φ is continuous, the supremum is actually attained. Conversely, for each $x_0 \in [0,\infty)$, convexity implies there exists a largest y such that the line $y(x - x_0) + f(x_0) \leqslant f(x)$ for all $x \in [0,\infty)$. This means that $\Psi(y) = x_0 y - x_0$.

We note also that $\Psi(0) = 0$, and Ψ is increasing. Most importantly, the function is convex. Given any $y, z \in [0,\infty)$, and any $x \in [0,\infty)$,

$$\begin{aligned} x(\alpha y + (1-\alpha)z) - \Phi(x) &\leqslant \alpha(xy - \Phi(x)) + (1-\alpha)(xz - \Phi(x)) \\ &\leqslant \alpha\Psi(y) + (1-\alpha)\Psi(z). \end{aligned}$$

Taking infimum over all x gives convexity. The function Ψ is also superlinear, since for any $x \in [0,\infty)$,

$$\lim_{y\to\infty} \frac{\Psi(y)}{y} \geqslant \lim_{y\to\infty} \frac{xy - \Phi(x)}{y} = x.$$

In particular, we can consider the Young dual of Ψ.

Lemma 7.24. *If Ψ is the Young dual of Φ, then Φ is the Young dual of Ψ.*

Proof. Π is the smallest function such that $\Pi(x) + \Psi(y) \geqslant xy$. Since $\Phi(x) + \Psi(y) \geqslant xy$ for each x and y, we conclude that $\Pi(x) \leqslant \Phi(x)$ for each x. For each x, there exists y such that $\Psi(y) = yx - \Phi(x)$. But this means that $\Phi(x) = yx - \Psi(y) \leqslant \Pi(x)$. $\square$

Given the Orlicz space $\Phi(L)$ for superlinear Φ, we can consider the Orlicz space $\Psi(L)$, where Ψ is the Young dual of Φ. The inequality $xy \leqslant \Phi(x) + \Psi(y)$, then

$$|f(x)g(x)| \leqslant \Phi(|f(x)|) + \Psi(|g(x)|),$$

so if $\|f\|_{\Phi(L)}, \|g\|_{\Psi(L)} \leqslant 1$, then

$$\left| \int f(x)g(x) \right| \leqslant \int |f(x)||g(x)| \leqslant \int \Phi(|f(x)|) + \int \Psi(|g(x)|) \leqslant 2.$$

Thus in general, we have

$$\left| \int f(x)g(x) \right| \leqslant 2\|f\|_{\Phi(L)} \|g\|_{\Psi(L)},$$

a form of Hölder's inequality. The duality between convex functions extends to a duality between the Orlicz spaces.

Theorem 7.25. *For any superlinear Φ with Young dual Ψ,*

$$\|f\|_{\Phi(L)} \sim \sup \left\{ \int fg : \|g\|_{\Psi(L)} \leqslant 1 \right\}.$$

Proof. Without loss of generality, assume $\|f\|_{\Phi(L)} = 1$. The version of Hölder's inequality proved above shows that

$$\|f\|_{\Phi(L)} \lesssim 1.$$

Conversely, for each x, we can find $g(x)$ such that $f(x)g(x) = \Phi(|f(x)|) + \Psi(|g(x)|$. Provided $\|g\|_{\Psi(L)} < \infty$, we have

$$\int fg = \int \Phi(|f(x)|) + \int \Psi(|g(x)|) \geqslant 1 + \|g\|_{\Psi(L)}.$$

Assuming $f \in L^\infty(X)$, we may choose $g \in L^\infty(X)$. For such a choice of function, $\|g\|_{\psi(L)} < \infty$, which implies the result. Taking an approximation argument then gives the result in general. $\square$

Let us now consider some examples of duality.

Example. *If $\Phi(x) = x^p$, for $p \geqslant 1$, and $1 = 1/p + 1/q$, then it's Young dual Ψ satisfies*

$$\Psi(y) = \sup_{x \geqslant 0} xy - x^p = y^{1+q/p}/p^{q/p} - y^q/p^q = y^q[p^{-q/p} - p^{-q}].$$

Thus the Young dual corresponds, up to a constant, to the conjugate dual in the L^p spaces.

Example. *Suppose X has finite measure. If $\Phi(t) = e^t - 1$, then it's dual satisfies, for large y,*

$$\begin{aligned}\Psi(y) &= \sup_{x \geqslant 0} xy - (e^x - 1)\\ &= y \log y - (y-1) \sim y \log y.\end{aligned}$$

This is comparable to $y \log(y+2)$ for large y. Thus $L \log L$ is dual to e^L.

Example. *Suppose X has finite measure. If $\Phi(x) = e^{x^2} - 1$, then for $y \geqslant 2$,*

$$\Psi(y) = \sup_{x \geqslant 0} xy - (e^{x^2} - 1) \sim y \log(y/2)^{1/2}.$$

Thus the dual of e^{L^2} is the space $L(\log L)^{1/2}$.

There is a generalization of both the Lorentz spaces and the Orlicz spaces, known as the Lorentz-Orlicz spaces, but these come up so rarely in analysis that we do not dwell on these norms.

Chapter 8

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. As should be expected, our main focus will be on the *interpolation of operators*.

8.1 Convex Interpolation

The most basic way to interpolate is using the notion of convexity. Given two inequalities $A_0 \leqslant B_0$ and $A_1 \leqslant B_1$, for any parameter $0 \leqslant \theta \leqslant 1$, if we define the additive weighted averages $A_\theta = (1-\theta)A_0 + \theta A_1$ and $B_\theta = (1-\theta)B_0 + \theta B_1$, then we conclude $A_\theta \leqslant B_\theta$ for all θ. Similarily, we can consider the weighted multiplicative averages $A_\theta = A_0^{1-\theta} A_1^\theta$ and $B_\theta = B_0^{1-\theta} B_1^\theta$, in which case we still have $A_\theta \leqslant B_\theta$. Note that the additive averages are obtained by taking the unique linear function between two values, and the multiplicative averages are obtained by taking the unique log-linear function between two values. In particular, if A_θ is defined to be any convex function, then $A_\theta \leqslant (1-\theta)A_0 + \theta A_1$, and if B_θ is logarithmically convex, so that $\log B_\theta$ is convex, then $B_\theta \leqslant B_0^{1-\theta} B_1^\theta$. Thus convexity provides us with a more general way of interpolating estimates, which is what makes this property so useful in analysis, enabling us to simplify estimates.

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

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

says

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

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

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

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

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

8.2 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 8.1 (The Three Lines Lemma). *If f is a holomorphic function on the strip $S = \{z : Re(z) \in [a,b]\}$ and there exists constants $A, B, \delta > 0$ such that for all $z \in S$,*

$$|f(z)| \leqslant Ae^{Be^{(\pi-\delta)|z|}}.$$

Then the function $M : [a,b] \to [0,\infty]$ given by

$$M(s) = \sup_{s \in \mathbf{R}} |f(s+it)|$$

is log convex on $[a,b]$.

Proof. By a change of variables, we can assume that $a = 0$, and $b = 1$, and we need only show that if there are $A_0, A_1 > 0$ such that

$$|f(it)| \leqslant A_0 \quad \text{and} \quad |f(1+it)| \leqslant A_1 \quad \text{for all } t \in \mathbf{R},$$

then for any $s \in [a,b]$ and $t \in \mathbf{R}$,

$$|f(s+it)| \leqslant A_0^{1-s} A_1^s.$$

By replacing $f(z)$ with the function $A_0^{1-z} A_1^z f(z)$, we may assume without loss of generality that $A_0 = A_1 = 1$, and we must show that $\|f\|_{L^\infty(S)} \leqslant 1$. If $|f(s+it)| \to 0$ as $|t| \to \infty$, then for large N, we can conclude that $|f(s+it)| \leqslant 1$ for $s \in [a,b]$ and $|t| \geqslant N$. But then the maximum principle entails that $|f(s+it)| \leqslant 1$ for $s \in [a,b]$ and $|t| \leqslant N$, which completes the proof in this case. In the general case, for each $\varepsilon > 0$, define

$$u_\varepsilon(z) = \exp(-2\varepsilon \sin((\pi-\varepsilon)z + \varepsilon/2)).$$

Then if $z = s + it$,

$$|u_\varepsilon(z)| = \exp(-\varepsilon[e^{(\pi-\varepsilon)t} + e^{-(\pi-\varepsilon)t}] \sin((\pi-\varepsilon)s + \varepsilon/2)),$$

So, in particular, $|u_\varepsilon(z)| \leqslant 1$, and there exists a constant C such that if $z \in S$,

$$|u_\varepsilon(z)| \leqslant e^{-C\varepsilon^2 e^{(\pi-\varepsilon)|z|}}$$

Note that if $\varepsilon < \delta$, then as $|\mathrm{Im}(z)| \to \infty$,

$$|f(z)u_\varepsilon(z)| \leqslant Ae^{Be^{(\pi-\delta)|z|} - C\varepsilon^2 e^{(\pi-\varepsilon)|z|}} \to 0.$$

Applying the previous case to the function $|f(z)u_\varepsilon(z)|$, we conclude that for any $\varepsilon > 0$,

$$|f(z)| \leqslant \frac{1}{|u_\varepsilon(z)|}.$$

Thus

$$|f(z)| \leqslant \lim_{\varepsilon \to 0} \frac{1}{|u_\varepsilon(z)|} = 1,$$

which completes the proof. □

Remark. The function $e^{-ie^{\pi is}}$ shows that the assumption of the three lines lemma is essentially tight. In particular, this means there is no family of holomorphic functions g_ε which decays faster than double exponentially, and pointwise approximates the identity as $\varepsilon \to 0$.

Remark. Similar variants can be used to show that if f is a holomorphic function on an annulus, then the supremum over circles centered around the origin is log convex in the radius of the circle (a result often referred to as the three circles lemma).

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

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

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

Example. *Let f be a holmomorphic function on a strip $S = \{z : Re(z) \in [a,b]\}$, such that if $z = a + it$, or $z = b + it$, for some $t \in \mathbf{R}$,*

$$|f(z)| \leqslant C_1(1 + |z|)^\alpha.$$

Then there exists a constant C' such that for any $z \in S$,

$$|f(z)| \leqslant C_2(1 + |z|)^\alpha.$$

Proof. The function

$$g(z) = \frac{f(z)}{(1+z)^\alpha}$$

is holomorphic on S, and if $z = a + it$ or $z = b + it$,

$$|g(z)| \leqslant \frac{C_1(1+|z|)^\alpha}{|1+z|^\alpha} \lesssim 1.$$

Thus the three lines lemma implies that $|g(z)| \lesssim 1$ for all $z \in S$, so

$$|f(z)| \lesssim |1+z|^\alpha \lesssim (1+|z|)^\alpha. \qquad \square$$

8.3 Interpolation of Operators

A major part of modern harmonic analysis is the study of operators, i.e. maps from function spaces to other function spaces. We are primarily interested in studying *linear operators*, i.e. operators T such that $T(f+g) = T(f) + T(g)$, and $T(\alpha f) = \alpha T(f)$, and also *sublinear operators*, such that $|T(\alpha f)| = |\alpha||T(f)|$ and $|T(f+g)| \leqslant |Tf| + |Tg|$. Even if we focus on linear operators, it is still of interest to study sublinear operators because one can study the *uniform boundedness* of a family of operators $\{T_k\}$ by means of the function $T^*(f)(x) = \max(T_k f)(x)$. This is the method of *maximal functions*. Another important example are the l^p sums

$$(S^p f)(x) = \left(\sum |T_k(x)|^p\right).$$

These two examples are specific examples where we have a family of operators $\{T_y\}$, indexed by a measure space Y, and we define an operator S by taking Sf to be the norm of $\{T_y f\}$ in the variable y.

Here we address the most basic case of operator interpolation. As we vary p, the L^p norms provide different ways of measuring the height and width of functions. Let us consider a simple example. Suppose that for an operator T, we have a bound

$$\|Tf\|_{L^1(Y)} \leqslant \|f\|_{L^1(X)} \quad \text{and} \quad \|Tf\|_{L^\infty(Y)} \leqslant \|f\|_{L^\infty(X)}.$$

The first inequality shows that the width of Tf is controlled by the width of f, and the second inequality says the height of Tf is controlled by the height of f. If we take a function $f \in L^p(X)$, for some $p \in (1,\infty)$, then we have some control over the height of f, and some control of the width. In particular, this means we might expect some control over the width and height of Tf, i.e. for each p, a bound

$$\|Tf\|_{L^p(Y)} \leqslant \|f\|_{L^p(X)}.$$

This is the idea of interpolation on the $L^p(X)$ spaces.

8.4 Complex Interpolation of Operators

The first theorem we give is the Riesz-Thorin theorem, which utilizes complex interpolation to give such a result. In the next theorem, we work with a linear operator T which maps simple functions f on a measure space X to functions on a measure space Y. For the purposes of applying duality, we make the mild assumption that for each simple function g,

$$\int |(Tf)(y)||g(y)|\, dy < \infty.$$

Our goal is to obtain L^p bounds on the function T. The Hahn-Banach theorem then guarantees that T has a unique extension to a map defined on all L^p functions.

Theorem 8.2 (Riesz-Thorin). *Let $p_0, p_1 \in (0,\infty]$ and $q_0, q_1 \in [1,\infty]$. Suppose that*

$$\|Tf\|_{L^{q_0}(Y)} \leqslant A_0 \|f\|_{L^{p_0}(X)} \quad \text{and} \|Tf\|_{L^{q_1}(Y)} \leqslant A_1 \|f\|_{L^{p_1}(X)}.$$

Then for any $\theta \in (0,1)$, if

$$1/p_\theta = (1-\theta)/p_0 + \theta/p_1 \quad \text{and} \quad 1/q_\theta = (1-\theta)/q_0 + \theta/q_1,$$

then

$$\|Tf\|_{L^{q_\theta}(Y)} \leqslant A_\theta \|f\|_{L^{p_\theta}(X)},$$

where $A_\theta = A_0^{1-\theta} A_1^\theta$.

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

$A_0 = A_1 = 1$. By duality and homogeneity, it suffices to show that for any two simple functions f and g such that $\|f\|_{q_\theta} = \|g\|_{q_\theta^*} = 1$,

$$\left| \int_Y (Tf) g \, dy \right| \leqslant 1.$$

Our challenge is to make this inequality complex analytic so we can apply the three lines lemma. We write $f = F_0^{1-\theta} F_1^{\theta} a$, where F_0 and F_1 are non-negative simple functions with $\|F_0\|_{L^{p_0}(X)} = \|F_1\|_{L^{p_1}(X)} = 1$, and a is a simple function with $|a(x)| = 1$. Similarily, we can write $g = G_0^{1-\theta} G_1^{\theta} b$. We now write

$$H(s) = \int_Y T(F_0^{1-s} F_1^s a) G_0^{1-s} G_1^s b \, dy.$$

Since all functions involved here are simple, $H(s)$ is a linear combination of positive numbers taken to the power of $1-s$ or s, and is therefore obviously an entire function in s. Now for all $t \in \mathbf{R}$, we have

$$\|F_0^{1-it} F_1^{it} a\|_{L^{p_0}(X)} = \|F_0\|_{L^{p_0}(X)} = 1,$$

$$\|G_0^{1-it} G_1^{it} b\|_{L^{q_0}(Y)} = \|G_0\|_{L^{q_0}(X)} = 1.$$

Therefore

$$|H(it)| = \left| \int T(F_0^{1-it} F_1^{it} a) G_0^{1-it} G_1^{it} b \, dy \right| \leqslant 1.$$

Similarily, $|H(1+it)| \leqslant 1$ for all $t \in \mathbf{R}$. An application of Lindelöf's theorem implies $|H(s)| \leqslant 1$ for all s. Setting $s = \theta$ completes the argument. □

If, for each p, q, we let $F(1/p, 1/q)$ to be the operator norm of a linear operator T viewed as a map from $L^p(X)$ to $L^q(Y)$, then the Riesz-Thorin theorem says that F is a log-convex function. In particular, the set of $(1/p, 1/q)$ such that T is bounded as a map from $L^p(X)$ to $L^q(Y)$ forms a convex set. If this is true, we often say T is of *strong type* (p, q).

Example. *For any two integrable functions $f, g \in L^1(\mathbf{R}^d)$, we can define an integrable function $f * g \in L^1(\mathbf{R}^d)$ almost everywhere by the integral formula*

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

If $f \in L^1(\mathbf{R}^d)$ and $g \in L^p(\mathbf{R}^d) \cap L^1(\mathbf{R}^d)$, for some $p \geqslant 1$, then Minkowski's integral inequality implies

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

Hölder's inequality implies that if $f \in L^p(\mathbf{R}^d)$ and $g \in L^q(\mathbf{R}^d)$, where p and q are conjugates of one another, then

$$\left| \int f(y) g(x-y) \; dy \right| \leqslant \int |f(y-x)||g(x)| \leqslant \|f\|_{L^p(\mathbf{R}^d)} \|g\|_{L^q(\mathbf{R}^d)}.$$

Thus we have the bound

$$\|f * g\|_{L^\infty(\mathbf{R}^d)} \leqslant \|f\|_{L^p(\mathbf{R}^d)} \|g\|_{L^q(\mathbf{R}^d)}.$$

Now that these mostly trivial results have been proved, we can apply convolution. For each $f \in L^1(\mathbf{R}^d) \cap L^p(\mathbf{R}^d)$, we have a convolution operator $T : L^1(\mathbf{R}^d) \to L^1(\mathbf{R}^d)$ defined by $Tg = f * g$. We know that T is of strong type $(1, p)$, and of type (q, ∞), where q is the harmonic conjugate of p, and T has operator norm 1 with respect to each of these types. But the Riesz Thorin theorem then implies that if $1/r = \theta + (1-\theta)/q$, then T is bounded as a map from $L^r(\mathbf{R}^d)$ to $L^{p/\theta}(\mathbf{R}^d)$ with operator norm one. Reparameterizing gives Young's convolution inequality. *Note that we never really used anything about $\mathbf{R}^d$ here other than it's translational structure, and as such Young's inequality continues to apply in the theory of any modular locally compact group. In particular, the Haar measure μ on such a group is only defined up to a scalar multiple, and if we swap μ with $\alpha\mu$, for some $\alpha > 0$, then Young's inequality for this measure implies*

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

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

Example. *Let X be a measure space with σ algebra Σ_0, and let $\Sigma \subset \Sigma_0$ be a σ finite sub σ algebra. Then $L^2(X, \Sigma)$ is a closed subspace of $L^2(X, \Sigma_0)$, and so there is an orthogonal projection operator $\mathbf{E}(\cdot|\Sigma) : L^2(X, \Sigma_0) \to L^2(X, \Sigma)$,* *which we call the* conditional expectation operator. *The properties of the projection operator imply that for any $f, g \in L^2(X, \Sigma_0)$,*

$$\int \mathbf{E}(f|\Sigma)\overline{g} = \int f\overline{g} = \int \mathbf{E}(f|\Sigma)\overline{\mathbf{E}(g|\Sigma)}.$$

If $g \in L^2(X, \Sigma)$, then

$$\int \mathbf{E}(f|\Sigma)\overline{g} = \int f\overline{g}.$$

This gives a full description of $\mathbf{E}(f|\Sigma)$. In particular, if $u \in L^\infty(X, \Sigma_0)$, then for each $g \in L^2(X, \Sigma)$

$$\int \mathbf{E}(uf|\Sigma)\overline{g} = \int f[u\overline{g}] = \int u\,\mathbf{E}(f|\Sigma)\overline{g}.$$

Since this is true for all $g \in L^2(X, \Sigma)$, we find $\mathbf{E}(uf|\Sigma) = u\,\mathbf{E}(f|\Sigma)$. Moreover, if $0 \leqslant f \leqslant g$, then $\mathbf{E}(f|\Sigma) \leqslant \mathbf{E}(g|\Sigma)$. This is easy to see because if $f \geqslant 0$, and $F = \{x : \mathbf{E}(f|\Sigma) < 0\}$, then if $|F| \neq 0$,

$$0 > \int \mathbf{E}(f|\Sigma)\mathbf{I}_F = \int f\mathbf{I}_F \geqslant 0.$$

Thus $|F| = 0$, and so $\mathbf{E}(f|\Sigma) \geqslant 0$ almost everywhere.

Like all other orthogonal projection operators, conditional expectation is a contraction in the L^2 norm, i.e. $\|\mathbf{E}(f|\Sigma)\|_{L^2(X)} \leqslant \|f\|_{L^2(X)}$. We now use interpolation to show that conditional expectation is strong (p, p), for all $1 \leqslant p \leqslant \infty$. It suffices to prove the operator is strong $(1, 1)$ and strong (∞, ∞). So suppose $f \in L^2(X, \Sigma_0) \cap L^\infty(X, \Sigma_0)$. If $|E| < \infty$, then $\mathbf{I}_E \in L^2(X)$, so

$$|\mathbf{E}(f|\Sigma)|\mathbf{I}_E = |\mathbf{E}(\mathbf{I}_E f|\Sigma)| \leqslant \mathbf{E}(\mathbf{I}_E|f||\Sigma) \leqslant \|f\|_\infty \mathbf{E}(\mathbf{I}_E|\Sigma) = \|f\|_\infty \mathbf{I}_E.$$

Since Σ is a sigma finite sigma algebra, we can take $E \to \infty$ to conclude $\|\mathbf{E}(f|\Sigma)\|_\infty \leqslant \|f\|_\infty$. The case $(1, 1)$ can be obtained by duality, since conditional expectation is self adjoint, or directly, since if $f \in L^1(X, \Sigma_0) \cap L^2(X, \Sigma_0)$, then for any set $E \in \Sigma$ with $|E| < \infty$,

$$\int |\mathbf{E}(f|\Sigma)|\mathbf{I}_E \leqslant \int \mathbf{E}(|f||\Sigma)\mathbf{I}_E = \int_E |f|\mathbf{I}_E \leqslant \|f\|_1.$$

Since Σ is σ finite, we can take $E \to \infty$ to conclude $\|\mathbf{E}(f|\Sigma)\|_1 \leqslant \|f\|_1$. Thus the Riesz interpolation theorem implies that for each $1 \leqslant p \leqslant \infty$, $\|\mathbf{E}(f|\Sigma)\|_p \leqslant \|f\|_p$.

Since $L^2(X,\Sigma_0)$ is dense in $L^p(X,\Sigma_0)$ for all $1 \leqslant p < \infty$, there is a unique extension of the conditional expectation operator from $L^p(X,\Sigma_0)$ to $L^p(X,\Sigma_0)$. For $p = \infty$, there are infinitely many extensions of the conditional expectation operator from $L^\infty(X,\Sigma_0)$ to $L^\infty(X,\Sigma_0)$. However, there is a unique *extension such that for each $f \in L^2(\Sigma_0)$ and $g \in L^\infty(\Sigma)$, $\mathbf{E}(fg|\Sigma) = g\,\mathbf{E}(f|\Sigma)$. This is because for any $E \in \Sigma$ with $|E| < \infty$, $\mathbf{E}(f\mathbf{I}_E|\Sigma) = \mathbf{I}_E\,\mathbf{E}(f|\Sigma)$ is uniquely defined since $f\mathbf{I}_E \in L^2(\Sigma_0)$, and taking $E \to \infty$ by σ finiteness.*

A simple consequence of the uniform boundedness of these operators on the various L^p spaces is that if $\Sigma_1, \Sigma_2, \dots$ are a family of σ algebras, and Σ_∞ is the smallest σ algebra containing all sets in $\bigcup_{i=1}^\infty \Sigma_i$, then for each $1 \leqslant p < \infty$, and for each $f \in L^p(\Sigma_0)$, $\lim_{i\to\infty} \mathbf{E}(f|\Sigma_i) = \mathbf{E}(f|\Sigma_\infty)$. This is because the operators $\{\mathbf{E}(\cdot|\Sigma_i)\}$ are uniformly bounded. The limit equation holds for any simple function f composed of sets in $\bigcup_{i=1}^\infty \Sigma_i$, and a σ algebra argument can then be used to show this family of simple functions is dense in $L^p(\Sigma_0)$.

It was an important observation of Elias-Stein that complex interpolation can be used not only with a single operator T, but with an 'analytic family' of operators $\{T_s\}$, one for each s, such that for each pair of simple functions f and g, the function

$$\int (T_s f)(y) g(y)$$

is analytic in s. Thus bounds on T_{0+it} and T_{1+it} imply intermediary bounds on all other operators, provided that we still have at most doubly exponential growth. The next theorem gives an example application.

Theorem 8.3 (Stein-Weiss Interpolation Theorem)**.** *Let T be a linear operator, and let $w_0, w_1 : X \to [0,\infty)$ and $v_0, v_1 : Y \to [0,\infty)$ be weights which are integrable on every finite-measure set. Suppose that*

$$\|Tf\|_{L^{q_0}(X,v_0)} \leqslant A_0 \|f\|_{L^{p_0}(X,w_0)} \quad \textit{and} \quad \|Tf\|_{L^{q_1}(X,v_1)} \leqslant A_1 \|f\|_{L^{p_1}(X,w_0)}.$$

Then for any $\theta \in (0,1)$,

$$\|Tf\|_{L^{q_\theta}(X,v_\theta)} \leqslant A_\theta \|f\|_{L^{p_\theta}(X,w_\theta)},$$

where $w_\theta = w_0^{1-\theta} w_\theta$ and $v_\theta = v_0^{1-\theta} v_1^\theta$.

Proof. Fix a simple function f with $\|f\|_{L^{p_\theta}(X,w_\theta)}$. We begin with some simplifying assumptions. A monotone convergence argument, replacing $w_i(t)$ with

$$w_i'(y) = \begin{cases} w_i(y) & : \varepsilon \leqslant w_i(t) \leqslant 1/\varepsilon, \\ 0 & : \text{otherwise,} \end{cases}$$

and then taking $\varepsilon \to 0$, enables us to assume without loss of generality that w_0 and w_1 are both bounded from below and bounded from above. Truncating the support of Tf enables us to assume that Y has finite measure. Since f has finite support, we may also assume without loss of generality that X has finite support, and by applying the dominated convergence theorem we may replace the weights v_i with

$$v_i'(x) = \begin{cases} v_i(x) & : \varepsilon \leqslant v_i(x) \leqslant 1/\varepsilon, \\ 0 & : \text{otherwise,} \end{cases}$$

and then take $\varepsilon \to 0$. Thus we can assume that the v_i are bounded from above and below. Restricting to the support of X, we can also assume X has finite measure.

For each s, consider the operator T_s defined by

$$T_s f = w_0^{\frac{1-s}{q_0}} w_1^{\frac{s}{q_1}} T\left(f v_0^{-\frac{1-s}{p_0}} v_1^{-\frac{s}{p_1}} \right).$$

The fact that all functions involved are simple means that the family of operators $\{T_s\}$ is analytic. Now for all $t \in \mathbf{R}$

$$\|T_{it} f\|_{L^{q_0}(Y)} = \|Tf\|_{L^{q_0}(Y,w_0)} \leqslant A_0 \|f v_0^{-1/p_0}\|_{L^{p_0}(X,v_0)} = A_0 \|f\|_{L^{p_0}(X)}.$$

For similar reasons, $\|T_{1+it} f\|_{L^{q_1}(Y)} \leqslant A_1 \|f\|_{L^{p_0}(X,v_0)}$. Thus the Stein variant of the Riesz-Thorin theorem implies that

$$\|T_\theta f\|_{L^{q_\theta}(Y)} \leqslant A_\theta \|f\|_{L^{p_\theta}(X)}.$$

But this, of course, is equivalent to the bound we set out to prove. □

8.5 Real Interpolation of Operators

Now we consider the case of real interpolation. One advantage of real interpolation is that it can be applied to sublinear as well as linear operators, and requires weaker endpoint estimates that the complex case. A disadvantage is that, usually, the operator under study cannot vary, and we lose out on obtaining explicit bounds.

A strong advantage to using real interpolation is that the criteria for showing boundedness at the endpoints can be reduced considerably. Let us give names for the boundedness we will want to understand for a particular operator T.

- We say T is *strong type* (p,q) if $\|Tf\|_{L^q(Y)} \lesssim \|f\|_{L^p(X)}$.
- We say T is *weak type* (p,q) if $\|Tf\|_{L^{q,\infty}(Y)} \lesssim \|f\|_{L^p(X)}$.
- We say T is *restricted strong type* (p,q) if we have a bound

  $$\|Tf\|_{L^q(Y)} \lesssim HW^{1/p}$$

  for any sub-step functions with height H and width W. Equivalently, for any set E,

  $$\|T(\mathbf{I}_E)\|_{L^q(Y)} \lesssim |E|^{1/p}.$$

- We say T is *restricted weak type* (p,q) if we have a bound

  $$\|Tf\|_{L^{q,\infty}(Y)} \lesssim HW^{1/p}$$

  for all sub-step functions with height H and width W. Equivalently, for any set E,

  $$\|T(\mathbf{I}_E)\|_{L^{q,\infty}(Y)} \lesssim |E|^{1/p}.$$

An important tool for us will be to utilize duality to make our interpolation argument 'bilinear'. Let us summarize this tool in a lemma. Proving the lemma is a simple application of Theorem 7.13.

Lemma 8.4. *Let $0 < p < \infty$ and $0 < q < \infty$. Then an operator T is restricted weak-type (p,q) if and only if for any finite measure sets $E \subset X$ and $F \subset Y$, there is $F' \subset Y$ with $|F'| \geqslant \alpha|F|$ such that*

$$\int_{F'} |T(\mathbf{I}_E)| \lesssim_\alpha |E|^{1/p}|F|^{1-1/q}.$$

Scalar interpoation leads to a simple version of real interpolation, which we employ as a subroutine to obtain a much more powerful real interpolation principle.

Lemma 8.5. *Let $0 < p_0, p_1 < \infty$, $0 < q_0, q_1 < \infty$. If T is restricted weak type (p_0, q_0) and (p_1, q_1), then T is restricted weak type (p_θ, q_θ) for all $\theta \in (0,1)$.*

Proof. By assumption, if $E \subset X$ and $F \subset Y$, then there is $F_0, F_1 \subset Y$ with $|F_i| \geqslant (3/4)|F|$ such that

$$\int_{F_i} |T(\mathbf{I}_E)| \lesssim |E|^{1/p_i} |F_i|^{1-1/q_i}.$$

If we let $F_\theta = F_0 \cap F_1$, then $|F_\theta| \geqslant |F|/2$, and for each i,

$$\int_{F_\theta} |T(\mathbf{I}_E)| \lesssim |E|^{1/p_i} |F_\theta|^{1-1/q_i}.$$

Scalar interpolation implies

$$\int_{F_\theta} |T(\mathbf{I}_E)| \lesssim |E|^{1/p_\theta} |F_\theta|^{1-1/q_\theta},$$

and thus we have shown

$$\|T(\mathbf{I}_E)\|_{q_\theta,\infty} \lesssim |E|^{1/p_\theta}.$$

This is sufficient to show T is restricted weak type (p_θ, q_θ). □

Theorem 8.6 (Marcinkiewicz Interpolation Theorem)**.** *Let $0 < p_0, p_1 < \infty$, $0 < q_0, q_1 < \infty$, and suppose T is restricted weak type (p_i, q_i), with constant A_i, for each i. Then, for any $\theta \in (0,1)$, if $q_\theta > 1$, then for any $0 < r < \infty$, then*

$$\|Tf\|_{L^{q_\theta,r}(Y)} \lesssim A_\theta \|f\|_{L^{p_\theta,r}(X)},$$

with implicit constants depending on p_0, p_1, q_0, and q_1.

Proof. By scaling T, and the measures on X and Y, we may assume that $\|f\|_{L^{p_\theta,r}(X)} \leqslant 1$, and that T is restricted type (p_i, q_i) with constant 1, so that for any step function f with height H and width W,

$$\|Tf\|_{L^{q_i,\infty}(Y)} \leqslant \|f\|_{L^{p_i}(X)}.$$

By duality, using the fact that $q_\theta > 1$, it suffices to show that for any simple function g with $\|g\|_{L^{q_\theta',r'}(Y)} = 1$,

$$\int |Tf||g| \leqslant 1.$$

Using the previous lemma, we can 'adjust' the values q_0, q_1 so that we can assume $q_0, q_1 > 1$. We can perform a horizontal layer decomposition, writing

$$f = \sum_{i=-\infty}^{\infty} f_i, \quad \text{and} \quad g = \sum_{i=-\infty}^{\infty} g_i,$$

where f_i and g_i are sub-step functions with width 2^i and heights H_i and H_i' respectively, and if we write $A_i = H_i 2^{i/p_\theta}$, and $B_i = H_i' 2^{i/q_\theta}$, then

$$\|A\|_{l^r(\mathbf{Z})}, \|B\|_{l^{r'}(\mathbf{Z})} \lesssim 1.$$

Applying the restricted weak type inequalities, we know for each i and j,

$$\int |Tf_i||g_j| \lesssim H_i H_j \min_{k \in \{0,1\}} \left[2^{i/p_k + j(1-1/q_k)} \right].$$

Applying sublinearity (noting that really, the decomposition of f and g is finite, since both functions are simple). Thus

$$\begin{aligned}
\int |Tf||g| &\leqslant \sum_{i,j} \int |Tf_i||g_j| \\
&\lesssim \sum_{i,j} H_i H_j' \min_{k \in \{0,1\}} \left[2^{i/p_k + j(1-1/q_k)} \right] \\
&\lesssim \sum_{i,j} A_i B_j \min_{k \in \{0,1\}} \left[2^{i(1/p_k - 1/p_\theta) + j(1/q_\theta - 1/q_k)} \right].
\end{aligned}$$

If $i(1/p_k - 1/p_\theta) + j(1/q_\theta - 1/q_k) = \varepsilon(i + \lambda j)$, where $\varepsilon = (1/p_k - 1/p_\theta)$. We then have

$$\sum_{i,j} A_i B_j \min_{k \in \{0,1\}} \left[2^{i(1/p_k - 1/p_\theta) + j(1/q_\theta - 1/q_k)} \right] \sim \sum_{k=-\infty}^{\infty} \min(2^{\varepsilon k}, 2^{-\varepsilon k}) \sum_i A_i B_{k - \lfloor i/\lambda \rfloor}.$$

Applying Hölder's inequality,

$$\begin{aligned}
\sum_i A_i B_{k-\lfloor i/\lambda \rfloor} &\leqslant \|A\|_{l^r(\mathbf{Z})} \left(\sum_i |B_{k - \lfloor i/\lambda \rfloor}|^{r'} \right)^{1/r'} \\
&\lesssim \lambda^{1/r'} \|A\|_{l^r(\mathbf{Z})} \|B\|_{l^{r'}(\mathbf{Z})} \lesssim 1.
\end{aligned}$$

Thus we conclude that

$$\sum_{k=-\infty}^{\infty} \min(2^{\varepsilon k}, 2^{-\varepsilon k}) \sum_i A_i B_{k-\lfloor i/\lambda \rfloor} \lesssim \sum_{k=-\infty}^{\infty} \min(2^{\varepsilon k}, 2^{-\varepsilon k}) \lesssim_\varepsilon 1. \qquad \square$$

There are many variants of the real interpolation method, but the general technique almost always remains the same: incorporate duality, decompose inputs, often dyadically, bound these decompositions, and then sum up.

Chapter 9

The Theory of Distributions

The theory of distributions is a tool which enables us to justify formal manipulations which occur in harmonic analysis, without the technical issues which occur from having to interpret such manipulations analytically. For instance, the ordinary integral formulation of the Fourier transform is only defined for L^1 functions, whereas the theory of tempered distributions enables us to define the Fourier transform of essentially any function we would ever want to take the Fourier transform of. Similarily, we can only classically differentiate a particular class of functions, but the theory of distributions enables us to define the derivative of almost any function that occurs in analysis. These reasons make distribution theory a cornerstone in the formulation of many problems in modern harmonic analysis and partial differential equations.

The power of measure theory is that it enables us to study a very general class of functionsome more. The problem is that as we study more general classes of functions, the operations we can perform on this class become more and more restricted. Nonetheless, $C_c^\infty(\mathbf{R}^d)$ is dense in every function space we consider, and we can apply all the fundamental analytical operations in this region, obtaining a general result by a density argument. The theory of distributions provides an alternate viewpoint.

From the perspective of set theory, functions $f : X \to Y$ are a way of assigning values in Y to each point in X. However, in analysis this is not often the way we view functions. For instance, in measure theory, we are used to identifying functions which are equal almost everywhere, so that functions in this setting are only defined 'almost everywhere'. In distribution theory, we view functions as 'integrands', whose properties are understand by integration against a family of 'test functions'. For instance, recall that for $1 \leqslant p < \infty$, the dual space of $L^p(\mathbf{R}^d)$ is $L^q(\mathbf{R}^d)$. Thus we can think of elements $f \in L^q(\mathbf{R}^d)$ as 'integrands', whose properties can be understood by integration against elements of $L^p(\mathbf{R}^d)$, i.e. through the linear functional $\Lambda[f]$ defined for each $\psi \in L^p(\mathbf{R}^d)$ by setting

$$\Lambda[f](\psi) = \int_{\mathbf{R}^d} f(x)\psi(x)\, dx.$$

Similarily, the dual space of $C(K)$, where K is a compact topological space, is the space $M(K)$ of finite Borel measures on K. Thus we can think of measures as a family of 'generalized functions'. For each measure $\mu \in M(K)$, we consider the linear functional $\Lambda[\mu]$, defined for each $\psi \in C(K)$, we set

$$\Lambda[\mu](\psi) = \int_K \psi(x) d\mu(x).$$

Notice that as we shrink the family of test functions, the resultant family of 'generalized functions' becomes larger and larger, and so elements can behave more and more erratically. A distribution is a 'generalized function' tested against functions in $C_c^\infty(\mathbf{R}^d)$. Since most operations in analysis can be applied to elements of $C_c^\infty(\mathbf{R}^d)$, most importantly, differentiation, we can use duality to extend these operations to distributions. Moreover, since $C_c^\infty(\mathbf{R}^d)$ is contained in most of the other function spaces, distributions are one of the largest family of generalized functions.

Remark. From the perspective of physics, viewing functions as integrands is completely natural. Points in space are idealizations which do not correspond to real world phenomena. One can never measure the exact value of some quantity of a function at a point, but rather only understand the function by looking at it's averages over a small region around that point. Thus a 'function' can be understood by the averages with respect to a family of integrands, known as test functions, since they test the value of the function over a region. As we make the family of integrands smaller and smaller, a 'function' can be behave more and more erratically. A distribution is an 'integrand' with respect to the space $C_c^\infty(\mathbf{R}^d)$ of infinitely differentiable functions with compact support. Since these functions are incredibly analytically nice, distributions are allowed to behave incredibly erratically, but we can still extend the operations of differentiation and integration to them.

9.1 The Space of Test Functions

Remark. This section is technical, and requires a strong knowledge of functional analysis, in particular the theory of locally convex topological spaces. It can be skipped without too much confusion.

We fix an open subset Ω of $\mathbf{R}^n$, and let $C_c^\infty(\Omega)$ denote the family of all smooth functions on Ω with compact support. Our goal is to equip $C_c^\infty(\Omega)$ with a complete locally convex topology, so that we can consider the dual space $C_c^\infty(\Omega)^*$ of *d*istributions on Ω. We could equip $C_c^\infty(\Omega)$ with a locally convex, metrizable topology with respect to the seminorms

$$\|f\|_{C^n(\Omega)} = \max_{|\alpha| \leqslant n} \|D^\alpha f\|_{L^\infty(\Omega)}$$

However, the resultant topology on $C_c^\infty(\Omega)$ isn't always complete.

Example. *Let $\Omega = \mathbf{R}$, pick a bump function $\phi \in C_c^\infty(\mathbf{R})$ supported on $[0,1]$ with $\phi > 0$ on $(0,1)$, and define*

$$\psi_m(x) = \phi(x-1) + \frac{\phi(x-2)}{2} + \dots + \frac{\phi(x-m)}{m}$$

Then ψ_m is compactly supported on $[1,m]$, and Cauchy, since for $m_1 \geqslant m_0$,

$$\|\psi_{m_0} - \psi_{m_1}\|_{C^n(\mathbf{R})} = \frac{\max_{r \leqslant n} \|D^r \phi\|_{L^\infty(\mathbf{R}^d)}}{m_0 + 1} \lesssim_n 1/m_0.$$

However, the sequence $\{\psi_m\}$ does not converge to any element of $C_c^\infty(\mathbf{R})$, since the sequence converges uniformly to the function

$$\psi(x) = \sum_{n=1}^\infty \psi(x-n)$$

an element of $C^\infty(\mathbf{R})$ which is not compactly supported.

We can assign $C_c^\infty(\Omega)$ a slightly stronger locally convex topology which makes is complete, but no longer metrizable. The process here is quite general. For each compact set $K \subset \Omega$, the subspace $C_c^\infty(K)$ of smooth functions compactly supported on K is a complete metric space, since limits of functions cannot 'escape' the set. Since $C_c^\infty(\Omega) = \bigcup_K C_c^\infty(K)$, we might be able to give $C_c^\infty(\Omega)$ a complete metric space structure by strengthening the topology by forcing limits to lie in a particular set of compact support. We now declare a convex topology, by considering the family of all sets $\{\phi + W\}$ as a basis, where ϕ ranges over all elements of $C_c^\infty(\Omega)$, and W ranges over all convex, balanced subsets of $C_c^\infty(\Omega)$ such that $W \cap C_c^\infty(K)$ is open for each compact set $K \subset \Omega$.

Theorem 9.1. *This gives a basis of a Hausdorff topology on $C_c^\infty(\Omega)$.*

Proof. If $\phi_1 + W_1$ and $\phi_2 + W_2$ both contain ϕ, then $\phi - \phi_1 \in W_1$ and $\phi - \phi_2 \in W_2$. The functions ϕ, ϕ_1, and ϕ_2 are all supported on some compact set K. By continuity of multiplication on $C_c^\infty(K)$, and the fact that $W_n \cap C_c^\infty(K)$ is open, there is a small constant δ such that $\phi - \phi_n \in (1-\delta)W_n$ for each $n \in \{1,2\}$. The convexity of the W_n implies that $\phi - \phi_n + \delta W_n \subset W_n$. But then $\phi + \delta W_n \subset \phi_n + W_n$, and so $\phi + \delta(W_1 \cap W_2) \subset (\phi_1 + W_1) \cap (\phi_2 + W_2)$. Thus we have verified the family of sets specified above is a basis. Now we show $C_c^\infty(\Omega)$ is Hausdorff under this topology. Suppose ϕ is in every open neighbourhood of the origin, then in particular, for each $\varepsilon > 0$, ϕ lies in the set $W_\varepsilon = \{f \in C_c^\infty(\Omega) : \|f\|_{L^\infty(\Omega)} < \varepsilon\}$, and it is easy to see these sets are open. Since $\bigcap_{\varepsilon > 0} W_\varepsilon = \{0\}$, this means $\phi = 0$. ☐

Remark. This technique can be formulated more abstractly to give a locally convex topological structure to the direct limit of locally convex spaces. From this perspective, we also see why our metrization doesn't work; if $X = \lim X_n$, with each X_n a locally convex metrizable space, then we cannot give X a complete metrizable topology such that each X_n is an embedding and has empty interior in X, because this would contradict the Baire category theorem. In particular, this means that the topology we have given to $C_c(\Omega)$ cannot be metrizable, and therefore the space cannot be first countable. Later we will see a more explicit proof of this.

Theorem 9.2. *$C_c^\infty(\Omega)$ is a locally convex space.*

Proof. Fix ϕ and ψ, and consider any neighbourhood W of the origin. By convexity, we have $(\phi + W/2) + (\psi + W/2) \subset (\phi + \psi) + W$. This shows addition is continuous. To show multiplication is continuous, fix λ, ϕ, and a neighbourhood W of the origin. Then ϕ is supported on some compact set K, and $W \cap C_c^\infty(K)$ is open, in particular absorbing, so there is $\varepsilon > 0$ such that if $|\alpha| < \varepsilon$, $\alpha\phi \in W/2$. Then if $|\gamma - \lambda| < \varepsilon$, then because W is balanced and convex,

$$\gamma\left(\phi+\frac{W}{2(|\lambda|+\varepsilon)}\right)=\lambda\phi+(\gamma-\lambda)\phi+\frac{\gamma}{2(|\lambda|+\varepsilon)}W$$
$$\subset\lambda\phi+W/2+W/2\subset\lambda\phi+W$$

so multiplication is continuous. □

Theorem 9.3. *For each compact set $K\subset\Omega$, the canonical embedding of $C_c^\infty(K)$ in $C_c^\infty(\Omega)$ is continuous.*

Proof. We shall prove a convex, balanced neighbourhood V is open in $C_c^\infty(\Omega)$ if and only if $C_c^\infty(K)\cap V$ is open in $C_c^\infty(K)$ for each K. Since V is open, V is the union of convex, balanced sets W_α with $W_\alpha\cap C_c^\infty(K)$ open in $C_c^\infty(K)$ for each K. But then $V\cap C_c^\infty(K)=(\bigcup W_\alpha)\cap C_c^\infty(K)$ is open in $C_c^\infty(K)$. The converse is true by definition of the topology. But this statement means exactly that the map $C_c^\infty(K)\to C_c^\infty(\Omega)$ is an embedding, because it is certainly continuous, and if W is a convex neighbourhood of the origin equal to the set of ϕ supported on K with $\|\phi\|_{C^n(K)}\leqslant\varepsilon$ for some n, then the image is the intersection of $C_c^\infty(K)$ with the set of all ϕ supported on Ω satisfying the inequality, which is open. This shows that the map is open onto its image, hence an embedding. □

It is difficult to see from the definition above why the topology is much stronger than the previous one given. We can see this more numerically by introducing the topology in terms of seminorms. The topology we have given $C_c^\infty(\Omega)$ is the same as the locally convex topology introduced by all norms $\|\cdot\|$ on the space which are continuous when restricted to each $C_c^\infty(K)$. As an example, if we choose an increasing family $U_1,U_2,\dots$ of precompact open sets whose closure is contained in Ω, then any compact set K is contained in some U_N for large enough N, and for any increasing sequence $\alpha_1,\alpha_2,\dots$ of positive constants and increasing sequence $k_1,k_2,\dots$ of positive integers the norm

$$\|f\|=\min_{\operatorname{supp}(f)\subset U_n}\alpha_n\|f\|_{C^{k_n}(U_n)}$$

is well defined on $C_c^\infty(\Omega)$ and continuous. But this means that if $f_1,f_2,\dots\to 0$, then $\|f\|\to 0$ for any choice of constants α_n and k_n, so asymptotically as we approach the boundary of Ω (or ∞, if Ω is unbounded), the L^∞ norms of the f_n and all of their derivatives must converge arbitrarily fast outside certain compact sets. The next theorem shows that this implies that the union of the domains f_n must actually be precompact. It is this 'uniform compactness' that gives us completeness.

Theorem 9.4. *E is a bounded subset of $C_c^\infty(\Omega)$ if and only if E is contained in $C_c^\infty(K)$ for some compact set K, and there is a sequence of constants $\{M_n\}$ such that $\|\phi\|_{C^n(\Omega)} \leqslant M_n$ for all $\phi \in E$.*

Proof. We shall now prove that if E is not contained in some $C_c^\infty(K)$ for any compact set $K \subset \Omega$, then E is not bounded. If our assumption is true, we can find functions $\phi_n \in E$ and a set of points $x_n \in X$ with no limit point such that $\phi_n(x_n) \neq 0$. For each n, set

$$W_n = \left\{ \psi \in C_c^\infty(\mathbf{R}^d) : |\psi(x_n)| < n^{-1} |\phi_n(x_n)| \right\}.$$

Certainly W_n is convex and balanced, and for each compact set K, if $\psi \in W_n \cap C_c^\infty(K)$, then there is $\varepsilon > 0$ such that $|\psi(x_n)| < n^{-1}|\phi_n(x_n)| - \varepsilon$. Thus if $\eta \in C_c^\infty(K)$ satisfies $\|\eta\|_{L^\infty(\mathbf{R}^d)} < \varepsilon$, then $\psi + \eta \in W_n$. In particular, this means $W_n \cap C_c^\infty(K)$ is open in $C_c^\infty(K)$ for each K, so W_n is open.

Now we claim $W = \bigcap_{n=1}^\infty W_n$ is open. Certainly this set is convex and balanced. Moreover, each compact set K contains finitely many of the points $\{x_n\}$, so $W \cap C_c^\infty(K)$ can be replaced by a finite intersection of the W_n, and is therefore open. Since $\phi_n \notin nW$ for all n, this implies that E is not bounded. The fact that $\|\cdot\|_{C^n(\Omega)}$ specifies the topological structure of $C_c^\infty(K)$ for each compact K now shows that if E is bounded, there exists constants $\{M_n\}$ such that $\|\phi\|_{C^n(\Omega)} \leqslant M_n$ for all $\phi \in E$. The converse property follows because $C_c^\infty(K)$ is embedded in $C_c^\infty(\Omega)$. □

Corollary 9.5. *$C_c^\infty(\Omega)$ has the Heine Borel property.*

Proof. This follows because if E is bounded and closed, it is a closed and bounded subset of some $C_c^\infty(K)$ for some K, hence E is compact since $C_c^\infty(K)$ satisfies the Heine-Borel property (this can be proved by a technical application of the Arzela-Ascoli theorem). □

Corollary 9.6. *$C_c^\infty(\Omega)$ is quasicomplete.*

Proof. If $\phi_1, \phi_2, \dots$ is a Cauchy sequence in $C_c^\infty(\Omega)$, then the sequence is bounded, hence contained in some common $C_c^\infty(K)$. Since the sequence is Cauchy, they converge in $C_c^\infty(K)$ to some ϕ, since $C_c^\infty(K)$ is complete, and thus the ϕ_n converge to ϕ in $C_c^\infty(\Omega)$. □

It is often useful to use the fact that we can perform a 'separation of variables' to a smooth function. This is done formally in the following manner. Say $f \in C_c^\infty(\mathbf{R}^d)$ is a *tensor function* if there are $f_1, \dots, f_n \in C_c^\infty(\mathbf{R})$ such that $f(x) = f_1(x_1)\dots f_n(x_n)$. We write $f = f_1 \otimes \dots \otimes f_n$. Since the product of two tensor functions is a tensor function, the family of all finite sums of tensor functions forms an algebra.

Theorem 9.7. *Finite sums of tensor functions are dense in $C_c^\infty(\mathbf{R}^d)$.*

Proof. Recall from the theory of multiple Fourier series that if $f \in C^\infty(\mathbf{R}^d)$ is N periodic, in the sense that $f(x+n) = f(x)$ for all $x \in \mathbf{R}^d$ and $n \in (N\mathbf{Z})^d$, then there are coefficients a_m for each $m \in \mathbf{Z}^n$ such that $f = \lim_{M \to \infty} S_M f$, where the convergence is dominated by the sminorms $\|\cdot\|_{C^n(\mathbf{R}^d)}$, for all $n > 0$, and

$$(S_M f)(x) = \sum_{\substack{m \in \mathbf{Z}^d \\ |m| \leqslant M}} a_m e^{\frac{2\pi i m \cdot x}{N}}.$$

Note that since

$$e^{\frac{2\pi i m \cdot x}{N}} = \prod_{k=1}^d e^{2\pi i m_i x_i / N}$$

is a tensor product, $S_M f$ is a finite sum of tensor functions. If $\phi \in C_c^\infty(\mathbf{R}^d)$ is compactly supported on $[-N,N]^d$, we let f be a $10N$ periodic function which is equal to ϕ on $[-N,N]^d$. We then find coefficients $\{a_m\}$ such that $S_M f$ converges to f. If $\psi : \mathbf{R} \to \mathbf{R}$ is a compactly supported bump function equal to one on $[-N,N]^d$, and vanishing outside of $[-2N,2N]^d$, then $\psi^{\otimes d} S_M f$ converges to ψ as $M \to \infty$, and each is a finite sum of tensor functions. $\square$

Because $C_c^\infty(\Omega)$ is the limit of metrizable spaces, it's linear operators still have many of the same properties as metrizable spaces.

Theorem 9.8. *If $T : C_c^\infty(\Omega) \to X$ is a map from $C_c^\infty(\Omega)$ to some locally convex space X, then the following are equivalent:*

(1) T is continuous.

(2) T is bounded.

(3) If $\{\phi_n\}$ converges to zero, then $\{T\phi_n\}$ converges to zero.

(4) *For each compact set $K \subset \Omega$, T is continuous restricted to $C_c^\infty(K)$.*

Proof. We already known that (1) implies (2). If T is bounded, and we have a sequence $\{\phi_n\}$ converging to zero, then the sequence is bounded, hence contained in some $C_c^\infty(K)$. Thend T is bounded as a map from $C_c^\infty(K)$ to X, hence $\{T\phi_n\} \to 0$. (3) implies (4) holds because each $C_c^\infty(K)$ is metrizable, and any convergent sequence is contained in some common $C_c^\infty(K)$. To prove that (4) implies (1), we let V be a convex, balanced, open subset of X. Then $T^{-1}(V) \cap C_c^\infty(K)$ is open for each K, and $T^{-1}(V)$ is convex and balanced, so $T^{-1}(V)$ is an open set. □

Because convergence is so strict in $C_c^\infty(\Omega)$, almost every operation we want to perform on smooth functions is continuous in this space.

- Since $f \mapsto D^\alpha f$ is a continuous operator from $C_c^\infty(K)$ to itself, it is therefore continuous on the entire space $C_c^\infty(\Omega)$. More generally, any linear differential operator with coefficients in $C_c^\infty(\Omega)$ is a continuous operator from $C_c^\infty(\Omega)$ to itself.

- The inclusion $C_c^\infty(\Omega) \to L^p(\Omega)$ is continuous. To prove this, it suffices to prove for each compact K, the inclusion $C_c^\infty(K) \to L^p(\Omega)$ is continuous, and this follows because $\|f\|_{L^p(\Omega)} \leqslant |K|^{1/p}\|f\|_\infty$.

- If $f \in L^1(\mathbf{R}^d)$ is compactly supported, then for any $g \in C_c^\infty(\mathbf{R}^d)$, $f * g \in C_c^\infty(\mathbf{R}^d)$. This is because $f * g$ is continuous since $g \in L^\infty(\mathbf{R}^n)$, and it's support is contained in the algebraic sums of the support of f and g, as well as the identity $D^\alpha(f * g) = f * (D^\alpha g)$. In fact, the map $g \mapsto f * g$ is a continuous operator on $C_c^\infty(\mathbf{R}^n)$. This is because if we restrict our attention to $C_c^\infty(K)$, and f has supported on K', then our convolution operator maps into the compact set $K + K'$, and since

  $$\|D^\alpha(g * f)\|_{L^\infty(K+K')} = \|D^\alpha g * f\|_{L^\infty(K+K')} \leqslant \|D^\alpha g\|_{L^\infty(K)}\|f\|_{L^1(K')},$$

  we conclude

  $$\|g * f\|_{C^n(K+K')} \leqslant \|g\|_{C^n(K)}\|f\|_{L^1(K')},$$

  which gives continuity of the operator as a map from $C_c^\infty(K)$ to $C_c^\infty(K+K')$. Since the latter space embeds in $C_c^\infty(\mathbf{R}^n)$, we obtain continuity of the operator on $C_c^\infty(\mathbf{R}^n)$.

Theorem 9.9. *If a map $T : C_c^\infty(K) \to C_c^\infty(\mathbf{R}^n)$ is continuous, then the image of $C_c^\infty(K)$ is actually $C_c^\infty(K')$ for some compact set K'.*

Proof. Suppose there is a sequence $x_1, x_2, \dots$ with no limit point and smooth functions $f_1, f_2, \dots$ compactly supported on $C_c^\infty(K)$ such that $(Tf_n)(x_n) \neq 0$. But then for any sequence α_n whatsoever, we cannot have $\alpha_n T f_n$ converging to zero, hence $\alpha_n f_n$ cannot converge to zero. But this is clearly not true, for if we let

$$\alpha_n = \frac{1}{2^n \|f_n\|_{C^n}}$$

Then for any fixed m, $\|f_n\|_{C^m}$ is eventually bounded above by $1/2^n$ and therefore converges to zero. Thus such a sequence x_n cannot exist, and therefore the image of T is supported on some compact set K'. $\square$

Thus the topology on the space $C_c^\infty(\mathbf{R}^d)$ is as strict as can be. As a consequence, we shall see that the weak $*$ topology on $C_c^\infty(\mathbf{R}^d)^*$ is essentially the weakest notion of convergence available in analysis, which makes it surprising that we still be able to recover the continuity of many operators on the dual space.

9.2 The Space of Distributions

We now have the tools to explain the idea of a distribution. If f is a locally integrable function defined on Ω, then the map

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

is a *continuous* linear functional defined for each $\phi \in C_c^\infty(\Omega)$. The functional $\Lambda[f]$ determines f up to a set of measure zero, and so we can safely identify $\Lambda[f]$ with f, and think of f 'distributionally'. The idea of the theory of distributions is to treat any continuous linear functional Λ on $C_c^\infty(\Omega)$ as if it were given by integration against a function as nice as possible. Using the properties of integration for these integration, we can usually cheat out a definition of operations usually only applicable to functions that works for all distributions. Thus the operations of analysis generalize to an incredibly large family of objects. As an example, if f was continuously differentiable, then we would find

$$\int f'(x)\phi(x)\, dx = -\int f(x)\phi'(x)\, dx$$

Since the right hand side is defined independantly of how nice the function $f(x)$ is, we could define the *derivative* of a continuous linear functional Λ as

$$\Lambda'(\phi) = -\Lambda(\phi')$$

and more generally, for a linear functional on n dimensional space, we could define $(D^\alpha \Lambda)(\phi) = (-1)^{|\alpha|}\Lambda(D^\alpha \phi)$.

Example. *Let $H(x) = \mathbf{I}(x > 0)$ denote the Heaviside step function. Then H is locally integrable, and so for any test function ϕ, we calculate*

$$\int H'(x)\phi(x)\,dx = -\int H(x)\phi'(x) = -\int_0^\infty \phi'(x) = \phi(0)$$

Thus the 'derivative' of the Heaviside step function is the Dirac delta. It is not a function, but if we were to think of it as a 'generalized function', it would be zero everywhere except at the origin, where it is infinitely peaked.

Example. *Consider the Dirac delta function at the origin, which is the distribution $\delta(f) = f(0)$. Then*

$$\delta'(f) = -\delta(f') = -f'(0)$$

which is a distribution which doesn't arise from integration with respect to a locally integrable function nor a Radon measure. Thus the 'derivative' of an infinitely peaked function at the origin is the negation of a derivative.

In general, we define a **distribution** to be a continuous linear functional on the space of test functions $C_c^\infty(\Omega)$. In the last section, our exploration of continuous linear transformations on $C_c^\infty(\Omega)$ guarantees that a linear functional Λ on $C_c^\infty(\Omega)$ is continuous if and only if for every compact $K \subset X$ there is an integer n_k such that $|\Lambda\phi| \lesssim_K \|\phi\|_{C^{n_k}}$ for $\phi \in C_c^\infty(K)$. If one integer n works for all K, and n is the smallest integer with such a property, we say that Λ is a distribution of *order n*. If such an n doesn't exist, we say the distribution has infinite order. If such an n doesn't exist, we say the distribution has infinite order.

Example. *If μ is a locally finite Borel measure, or a finite complex valued measure, then we can define*

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

Thus $\Lambda[\mu]$ is a distribution, since if ϕ is supported on K, then

$$|\Lambda[\mu](\phi)| \leqslant \mu(K)\|\phi\|_{L^\infty(K)}$$

Thus $\Lambda[\mu]$ is a distribution of order zero.

Example. *Not all distributions arise from functions or measures. For instance, if $\phi \in C_c^\infty(\mathbf{R})$, then*

$$p.v \int \frac{\phi(x)}{x} = \lim_{\varepsilon \to 0} \int_{|x|>\varepsilon} \frac{\phi(x)}{x}\, dx$$

exists. This is because the values of $1/x$ on either side of the real axis cancel each other out in the integral. More rigorously, if the support of ϕ is contained in $[-N,N]$, then

$$\left| \int_{|x|>\varepsilon} \frac{\phi(x)}{x} \right| = \left| \int_{\varepsilon<|x|<N} \frac{\phi(x)-\phi(0)}{x} \right|$$

The mean value theorem implies $\phi(x)-\phi(0) = x\phi'(y)$ for some y between zero and x, so

$$\left| \int_{|x|\leqslant\varepsilon} \frac{\phi(x)-\phi(0)}{x} \right| \leqslant 2\varepsilon\|\phi\|_{C^1(\mathbf{R})}$$

From this, we can conclude that these values are Cauchy as $\varepsilon \to 0$, hence the principal value exists, and moreover, if ϕ is compactly supported on $[-N,N]$,

$$\left| p.v. \int \frac{\phi(x)}{x} \right| \lesssim_N \|\phi\|_{C^1(\mathbf{R}^d)}$$

Thus this linear functional is actually continuous, hence a distribution, and it is not induced by any function or measure. This distribution is the derivative of the distribution induced by the locally integrable function $\log|x|$, since an integration by parts shows that for each $\phi \in C_c^\infty(\mathbf{R}^d)$,

$$\begin{aligned}
-\int \log|x|\phi'(x) &= \lim_{\varepsilon\to 0} \int_{|x|\geqslant\varepsilon} \log|x|\phi'(x) \\
&= \lim_{\varepsilon\to 0} \left(\log(\varepsilon)\cdot(\phi(x)-\phi(-x)) + \int_{|x|\geqslant\varepsilon} \frac{\phi(x)}{x} \right) \\
&= p.v. \int \frac{\phi(x)}{x}\, dx.
\end{aligned}$$

As we stated before, given any distribution Λ, we can define it's *derivative* $D^\alpha \Lambda$ to be the distribution

$$D^\alpha \Lambda(\phi) = (-1)^{|\alpha|} \Lambda(D^\alpha \phi)$$

which is continuous since the derivative operation is continuous on $C_c^\infty(\Omega)$. Just as the partial derivatives commutes on $C_c^\infty(\Omega)$, the partial differentiation operation commutes on the the space of distributions, i.e. $D^\alpha D^\beta \Lambda = D^\beta D^\alpha \Lambda$, and we take the common value to be $D^{\alpha+\beta}\Lambda$. If $D^\alpha f$ is continuous, then we already know an integration by parts gives $D^\alpha \Lambda[f] = \Lambda[D^\alpha f]$, so we can think of the distribution derivative as a true generalization of the usual derivative. On the other hand, in general the distribution derivative may disagree with the usual derivative if the function is less well behaved. If P is a polynomial, we have

$$P(D)(\Lambda)(\phi) = \Lambda(P(-D)(\phi))$$

if we understand the polynomial applications of derivatives linearly.

Example. *Let f be a left continuous function on the real line with bounded variation and with $f(-\infty) = 0$. Then f' exists almost everywhere in the classical sense, and $f' \in L^1(\mathbf{R})$. By Fubini's theorem, if we let μ be the measure defined by $\mu([a,b)) = f(b) - f(a)$, then for any $\phi \in C_c^\infty(\mathbf{R})$,*

$$\begin{aligned}\int_{-\infty}^\infty \phi(x) d\mu(x) &= -\int_{-\infty}^\infty \int_x^\infty \phi'(y)\ dy\ d\mu(x)\\ &= -\int_{-\infty}^\infty \phi'(y) \int_{-\infty}^y d\mu(x)\ dy\\ &= -\int_{-\infty}^\infty \phi'(y) f(y) dy\end{aligned}$$

and we know $f(-\infty) = 0$. Thus we find $\Lambda[f'] = \Lambda[\mu]$. In particular, we only have $\Lambda[f]' = \Lambda[f']$ if $f' dx = \mu$, which only holds if f is absolutely continuous.

If $f \in L^1_{\text{loc}}(\mathbf{R}^n)$, and g is C^∞, then fg is locally integrable. The identity

$$\int (f(x)g(x))\phi(x)\ dx = \int f(x)(g(x)\phi(x))\ dx$$

enables us to define the product of a $C^\infty(\Omega)$ function with a distribution. Given any distribution Λ, we define $(f\Lambda)(\phi) = \Lambda(f\phi)$. To see why $f\Lambda$ is a distribution, fix a compact set K, and pick A and N such that for any $\phi \in C_c^\infty(K)$, $|\Lambda(f)| \leqslant A\|f\|_{N,K}$. The Leibniz rule tells us that

$$D^\alpha(f\phi) = \sum_{\lambda+\gamma=\alpha} C_{\lambda\gamma} D^\lambda f D^\gamma \phi$$

and so

$$|\Lambda(f\phi)| \leqslant A\|f\phi\|_{N,K} \leqslant NA\left(\max|C_{\lambda\gamma}|\|D^\lambda f\|_{L^\infty(K)}\right)\|\phi\|_{C^N(K)}$$

so $f\Lambda$ is a distribution with the same order as Λ. It is important that f is a smooth function, so that fg is smooth for each function g.

Since $C_c^\infty(X)^*$ is the dual space of a topological vector space, we can give it a natural topology, the weak $*$ topology. Thus a net of distributions Λ_α converges to Λ if and only if $\Lambda_\alpha(\phi) \to \Lambda(\phi)$ for all test functions ϕ. This gives a further topology on the space of measures and functions, and we often write $f_\alpha \to f$ 'in the distribution sense' if we have a convergence $\Lambda[f_\alpha] \to \Lambda[f]$ for the corresponding distributions. Since the convergence in $C_c^\infty(\Omega)$ is incredibly strict, convergence of distributions is incredibly weak. The following is thus quite a surprising result.

Theorem 9.10. *Suppose that $\Lambda_1, \Lambda_2, \dots$ are a sequence of distributions such that for a fixed test function ϕ,*

$$\Lambda\phi = \lim \Lambda_n \phi$$

exists. Then Λ is a distribution, and $D^\alpha\Lambda_n \to D^\alpha\Lambda$ for each α.

Proof. Fix a compact set K. Then the Banach Steinhaus theorem guarantees that Λ restricted to $C_c^\infty(K)$ is a continuous functional, and we know this implies Λ is continuous in general. The fact that $D^\alpha\Lambda_n \to D^\alpha\Lambda$ is trivial, because for a fixed ϕ, $D^\alpha\phi \in C_c^\infty(X)$, so

$$D^\alpha\Lambda(\phi) = (-1)^{|\alpha|}\Lambda(D^\alpha\phi) = (-1)^{|\alpha|}\lim \Lambda_n(D^\alpha\phi) = \lim D^\alpha\Lambda_n(\phi)$$

Thus the sequence weakly converges to $D^\alpha\Lambda$. □

A similar application of the Banach Steinhaus theorem guarantees that if $g_n \to g$ in $C^\infty(\mathbf{R}^n)$, and $\Lambda_n \to \Lambda$ in the distributional sense, then $g_n\Lambda_n$ converges to $g\Lambda$ in the distributional sense. Thus the space of distributions is a topological $C^\infty(\mathbf{R}^n)$ module.

9.3 Localization of Distribuitions

Just as we can consider the local behaviour of functions around a point, we can consider the local behaviour of a distribution around points, and this local behaviour contains most of the information of the distribution. For instance, given an open subset U of X, we say two distributions Λ and Ψ are equal on U if $\Lambda\phi = \Psi\phi$ for every test function ϕ compactly supported in U. We recall the notion of a partition of unity, which, for each open cover U_α of Euclidean space, gives a family of C^∞ functions ψ_α which are positive, *locally finite*, in the sense that only finitely many functions are positive on each compact set, and satisfy $\sum \psi_\alpha = 1$ on the union of the U_α.

Theorem 9.11. *If X is covered by a family of open sets U_α, and Λ and Ψ are locally equal on each U_α, then $\Lambda = \Psi$. If we have a family of distributions Λ_α which agree with one another on $U_\alpha \cap U_\beta$, then there is a unique distribution Λ locally equal to each Λ_α.*

Proof. Since we can find a C^∞ partition of unity ψ_α compactly supported on the U_α, upon which we find if ϕ is supported on K, then finitely many of the ψ_α are non-zero on K, and so

$$\Lambda(\phi) = \sum \Lambda(\psi_\alpha \phi) = \sum \Psi(\psi_\alpha \phi) = \Psi(\phi)$$

Thus $\Lambda = \Psi$. Conversely, if we have a family of distributions Λ_α like in the hypothesis, then we can find a partition of unity $\psi_{\alpha\beta}$ subordinate to $U_\alpha \cap U_\beta$, and we can define

$$\Lambda(\phi) = \sum \Lambda_\alpha(\psi_{\alpha\beta}\phi) = \sum \Lambda_\beta(\psi_{\alpha\beta}\phi)$$

The continuity is verified by fixing a compact K, from which there are only finitely many nonzero $\psi_{\alpha\beta}$ on K, and the fact that this definition is independant of the partition of unity follows from the first part of the theorem. □

In the language of modern commutative algebra, the association of $C_c^\infty(U)^*$ to each open subset U of Ω gives a sheaf structure to Ω. Given a distribution Λ, we might have $\Lambda(\phi) = 0$ for every ϕ supported on some open set U. The complement of the largest open set U for which this is true is called the **support** of Λ. If f vanishes on a neighbourhood of the support of Λ, then by definition of the support, $\Lambda f = 0$. The neighbourhood condition is important – δ' is supported on $\{0\}$, since δ is, but it certainly doesn't vanish on f if $f(0) = 0$.

Theorem 9.12. *If a distribution has precompact support, the distribution has finite order, and extends uniquely to a continuous linear functional on $C^\infty(X)$.*

Proof. Let Λ be a distribution supported on a compact set. If ψ is a function with compact support with $\psi(x) = 1$ on the support of Λ, then $\psi\Lambda = \Lambda$, because for any ϕ, $\phi - \phi\psi$ is supported on a set disjoint from the support of Λ. But if ψ is supported on K, then there is N such that for any $\phi \in C_c^\infty(K)$,

$$|\Lambda(\phi)| \lesssim \|\phi\|_{N,K}$$

and so for any other compact set K,

$$|\Lambda(\phi)| = |\Lambda(\phi\psi)| \lesssim \|\phi\psi\|_{N,K} \lesssim \|\psi\|_{C^N(K)}\|\phi\|_{C^N(K)}$$

which shows Λ has order N. We have shown that Λ is continuous with respect to the seminorm $\|\cdot\|_{C^N(K)}$ on $C^\infty(X)$, and so by the Hahn Banach theorem, Λ extends uniquely to a continuous functional on $C^\infty(X)$. $\square$

Example. *If $\Lambda(\phi) = \sum_{|\alpha| \leqslant N} \lambda_\alpha D^\alpha \phi(x)$, then Λ is supported on x. Conversely, every distribution Λ supported on x is of this form. We know Λ must have finite order N, and consider ϕ with $D^\alpha \phi(x) = 0$ for all $|\alpha| \leqslant N$. We claim $\Lambda(\phi) = 0$. Fix $\varepsilon > 0$, and choose a compact neighbourhood K of the origin with $|D^\alpha \phi(x)| < \varepsilon$ on K for all $|\alpha| = N$. Then for $|\alpha| < N$, the mean value theorem implies that, by induction,*

$$|D^\alpha \phi(x)| \leqslant \varepsilon n^{N-|\alpha|}|x|^{N-|\alpha|}$$

Find A such that for functions ϕ supported on K,

$$|\Lambda(\phi)| \leqslant A\|\phi\|_{C^N(K)}$$

Fix a bump function ψ with support on the ball of radius one and $\psi(x) = 1$ in a neighbourhood of the origin, and define $\psi_\delta(x) = \psi(x/\delta)$. If δ is small enough, *then ψ is supported on K, and because Λ is supported on x,*

$$\begin{aligned}
|\Lambda(\phi)| = |\Lambda(\phi\psi_\delta)| &\leqslant A\|\phi\psi_\delta\|_{C^N(K)} \\
&\leqslant A \sum_{|\alpha+\beta|=N} |c_{\alpha\beta}| \|D^\alpha \phi\|_\infty \|D^\beta \psi_\delta\| \\
&\leqslant A\|\psi\|_{C^N} \sum_{|\alpha+\beta|=N} |c_{\alpha\beta}| \delta^{|\beta|-|\alpha|} \|D^\beta \phi\|_{L^\infty(K)} \\
&\leqslant \varepsilon A \left(\sum_{|\alpha+\beta|=N} |c_{\alpha\beta}| n^{N-|\beta|} \right)
\end{aligned}$$

We can then let $\varepsilon \to 0$ to conclude $\Lambda(\phi) = 0$. But this means that $\Lambda(\phi)$ is a linear function of the partial derivatives of ϕ with order $\leqslant N$, completing the proof.

Example. *If δ is the Dirac delta distribution, then $f\delta = f(0)\delta$ for any smooth function f. Thus, in particular, $x\delta = 0$. Conversely, if Λ is any distribution with $x\Lambda = 0$, then Λ is a multiple of the Dirac delta distribution. To see this, we note that this would imply $\Lambda(f) = 0$ for all functions f such that f/x is also smooth and compactly supported. In particular, this is true if the support of f does not contain the origin. Thus Λ is supported on the origin, hence there are constants a_n such that*

$$\Lambda f = \sum_{n=0}^{N} a_n f^{(n)}(0)$$

But $(xf)^{(n)}(0) = nf^{(n-1)}(0)$ only vanishes for all f when $n = 0$, so Λ is a multiple of the Dirac delta distribution. A more simple way to see this is that if f is compactly supported on $[-N, N]$, the function

$$g(x) = \frac{f(x) - f(0)}{x} = \int_0^1 f'(tx)\, dt$$

is smooth, and $f = f(0) + xg$. Since Λ and $x\Lambda$ have bounded support, they extend uniquely to $C^\infty(\Omega)$, and so $\Lambda f = f(0)\Lambda 1 + \Lambda(xg) = f(0)\Lambda 1$.

In many other ways, distributions act like functions. For instance, any distribution Λ can be uniquely written as $\Lambda_1 + i\Lambda_2$ for two distributions

Λ_1, Λ_2 that are real valued for any real-valued smooth continuous function. However, we cannot write a real-valued distribution as the difference of two positive distributions, i.e. those which are non-negative when evaluated at any non-negative functional. Given a non-negative functional Λ, we define Λf for a compactly supported continuous function $f \geqslant 0$ as

$$\Lambda f = \sup\{\Lambda g : g \in C_c^\infty(\mathbf{R}^n), g \leqslant f\}$$

and then in general define $\Lambda(f^+ - f^-) = \Lambda f^+ - \Lambda f^-$. Then Λ is obviously a positive extension of Λ to all continuous functions, and is linear. But then the Riesz representation theorem implies that there is a positive Radon measure such that $\Lambda = \Lambda_\mu$, completing the proof.

9.4 Derivatives of Continuous Functions

One of the main reasons to consider the theory of distributions is so that we can take the derivative of any function we want. We now show that, at least locally, every distribution is the derivative of some continuous function, which means the theory of distributions is essentially the minimal such class of objects which enable us to take derivatives of continuous functions.

Theorem 9.13. *If Λ is a distribution on Ω, and K is a compact set, then there is a continuous function f and α such that for every ϕ,*

$$\Lambda\phi = (-1)^{|\alpha|} \int_\Omega f(x)(D^\alpha \phi)(x)\, dx$$

Proof. TODO $\square$

Theorem 9.14. *If K is compact, contained in some open subset V, which in turn is a subset of Ω, and Λ has order N, then there exists finitely many continuous functions $f_\beta \in C(\Omega)$ supported on V, for each $|\beta| \leqslant N+2$, with supports on V, and with $\Lambda = \sum D^\beta f_\beta$.*

Theorem 9.15. *If Λ is a distribution on Ω, then there exists continuous functions g_α on Ω such that each compact set K intersects the supports of finitely many of the g_α, and $\Lambda = \sum D^\alpha g_\alpha$. If Λ has finite order, then only finitely many of the g_α are nonzero.*

9.5 Convolutions of Distributions

Using the convolution of two functions as inspiration, we will not define the convolution of a distribution Λ with a test function ϕ, and under certain conditions, the convolution of two distributions. Recall that if $f, g \in L^1(\mathbf{R}^n)$, then their convolution is the function in $L^1(\mathbf{R}^n)$ defined by

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

If we define the translation operators $(T_y g)(x) = g(x-y)$, then $(f * g)(x) = \int f(y)(T_x g^*)(y)\, dy$, where g^* is the function defined by $g^*(x) = g(-x)$. Thus, if Λ is any distribution on $\mathbf{R}^n$, and ϕ is a test function on $\mathbf{R}^n$, we can define a function $\Lambda * \phi$ by setting $(\Lambda * \phi)(x) = \Lambda(T_x \phi^*)$. Notice that since

$$\begin{aligned}\int (T_x f)(y) g(y)\, dy &= \int f(y-x) g(y)\, dy = \int f(y) g(x+y)\, dy \\ &= \int f(y)(T_{-x} g)(y)\, dy,\end{aligned}$$

so we can also define the translation operators on distributions by setting $(T_x \Lambda)(\phi) = \Lambda(T_{-x}\phi)$. One mechanically verifies that convolution commutes with translations, i.e. $T_x(\Lambda * \phi) = (T_x \Lambda) * \phi = \Lambda * (T_x \phi)$.

Theorem 9.16. *$\Lambda * \phi$ is C^∞, and $D^\alpha(\Lambda * \phi) = (D^\alpha \Lambda) * \phi = \Lambda * (D^\alpha \phi)$.*

Proof. It is easy to calculate that

$$\begin{aligned}(D^\alpha \Lambda * \phi)(x) &= (D^\alpha \Lambda)(\phi_x^*) = (-1)^{|\alpha|} \Lambda(D^\alpha (T_x \phi^*)) \\ &= \Lambda(T_x (D^\alpha \phi)^*) = (\Lambda * D^\alpha \phi)(x)\end{aligned}$$

If e is a unit vector, and we set $\Delta_h = h^{-1}(1 - T_{he})$, then $\Delta_h \phi$ converges to $D_e \phi$ in $C_c^\infty(\mathbf{R}^n)$, and as such,

$$\begin{aligned}\Delta_h(\Lambda * \phi)(x) &= \frac{(\Lambda * \phi)(x) - (\Lambda * \phi)(x - he)}{h} \\ &= \frac{\Lambda(T_x \phi^* - T_{-he}(T_x \phi^*)}{h} = \Lambda(T_x(\Delta_h \phi^*))\end{aligned}$$

and this converges to $\Lambda(D_e \phi^*) = (\Lambda * D_e \phi)(x)$ as $h \to 0$. Iteration of this fact gives the general result. $\square$

Theorem 9.17. *If $\phi, \psi \in C_c^\infty(\mathbf{R}^n)$, then $\Lambda * (\phi * \psi) = (\Lambda * \phi) * \psi$.*

Proof. Let ϕ and ψ be supported on K. We calculate

$$(\phi * \psi)^*(x) = \int \phi^*(x+y)\psi(y)\,dy = \int (T_y\phi)^*(x)\psi(y)\,dy$$

since the map $y \mapsto (T_y\phi)^*\psi(y)$ is continuous, and vanishes out of the compact set K, so that we have a $C_c^\infty(K)$ valued integral

$$(\phi * \psi)^* = \int_K \psi^*(y) T_y \phi^*\,ds$$

This means precisely that

$$\begin{aligned}(\Lambda * (\phi * \psi))(0) = \Lambda((\phi * \psi)^*) &= \int_K \psi^*(y)\Lambda(T_y\phi^*)\,dy\\ &= \int_K \psi^*(y)(\Lambda * \phi)(y)\,dy = ((\Lambda * \phi) * \psi)(0)\end{aligned}$$

The commutativity in general results from applying the commutativity of the translation operators. ☐

A net ϕ_α is known as an *approximate identity* in the space of distributions if $\Lambda * \phi_\alpha \to \Lambda$ as distributions for every distribution ϕ, and an approximate identity in the space of test functions if $\psi * \phi_\alpha \to \psi$ in $C_c^\infty(\mathbf{R}^n)$.

Theorem 9.18. *If ϕ_α is a family of non-negative functions in $C_c^\infty(\mathbf{R}^n)$ which are eventually supported on every neighbourhood of the origin, and integrate to one, then ϕ_α is an approximation to the identity in the space of test functions and in the space of distributions.*

Proof. It is easy to verify that if f is a continuous function, then $f * \phi_\delta$ converges locally uniformly to f as $\delta \to 0$. But now we calculate that if $f \in C_c^\infty(\mathbf{R}^n)$, then $D^\alpha(f * \phi_\delta) = (D^\alpha f) * \phi_\delta$ converges locally uniformly to $D^\alpha \phi$, which gives that $f * \phi$ converges to f in $C_c^\infty(\mathbf{R}^n)$. Now if Λ is a distribution, and ψ is a test function, then continuity gives

$$\begin{aligned}\Lambda(\psi^*) &= \lim_{\delta\to 0}\Lambda(\phi_\delta * \psi) = \lim_{\delta \to 0}(\Lambda * (\phi_\delta * \psi))(0)\\ &= \lim_{\delta\to 0}((\Lambda * \phi_\delta) * \psi)(0) = \lim_{\delta\to 0}(\Lambda * \phi_\delta)(\psi^*)\end{aligned}$$

and ψ was arbitrary. ☐

If Λ is a distribution on $\mathbf{R}^n$, then the map $\phi \mapsto \Lambda * \phi$ is a linear transformation from $C_c^\infty(\mathbf{R}^n)$ into $C^\infty(\mathbf{R}^n)$, which commutes with translations. It is also continuous. To see this, we consider a fixed compact K, and consider the map from $C_c^\infty(K)$ to $C^\infty(\mathbf{R}^n)$. We can apply the closed graph theorem to prove continuity, so we assume the existence of $\phi_1, \phi_2, \dots$ converging to ϕ in $C_c^\infty(K)$ and $\Lambda * \phi_1, \Lambda * \phi_2, \dots$ converges to f. It suffices to show $f = \Lambda * \phi$. But we calculate

$$f(x) = \lim (\Lambda * \phi_n)(x) = \lim \Lambda(T_x \phi_n^*) = \Lambda(\lim T_x \phi_n^*) = \Lambda(T_x \phi^*) = (\Lambda * \phi)(x)$$

where we have used the fact that $T_x \phi_n^*$ converges to $T_x \phi^*$ in $C_c^\infty(\mathbf{R}^n)$. Suprisingly, the converse is true.

Theorem 9.19. *If $L : C_c^\infty(\mathbf{R}^n) \to C^\infty(\mathbf{R}^n)$ and commutes with translations, then there is a distribution Λ such that $L(\phi) = \Lambda * \phi$.*

Proof. If $L(\phi) = \Lambda * \phi$, then we would have

$$\Lambda(\phi) = (\Lambda * \phi^*)(0) = L(\phi^*)(0)$$

and we take this as the definition of Λ. Λ is continuous because all the operations here are continuous, and because L commutes with translations, we conclude

$$(\Lambda * \phi)(x) = \Lambda(T_x \phi^*) = L(T_{-x} \phi)(0) = L(\phi)(x)$$

which gives the theorem. □

We now move onto the case where a distribution Λ has compact support. Then Λ extends to a continuous functional on $C^\infty(\mathbf{R}^n)$, and we can define the convolution $\Lambda * \phi$ if $\phi \in C^\infty(\mathbf{R}^n)$. The same techniques as before verify that translations and derivatives are carried into the convolution.

Theorem 9.20. *If ϕ and Λ have compact support, then $\Lambda * \phi$ has compact support.*

Proof. Let ϕ and Λ be supported on K. Then $(\Lambda * \phi)(x) = \Lambda(T_x \phi^*)$. Since $T_x \phi^*$ is supported on $x - K$, for x large enough $x - K$ is disjoint from K, and so $\Lambda * \phi$ vanishes outside of $K + K$. □

Theorem 9.21. *If Λ and ψ have compact support, and $\phi \in C^\infty(\mathbf{R}^n)$, then*

$$\Lambda * (\phi * \psi) = (\Lambda * \phi) * \psi = (\Lambda * \psi) * \phi$$

Proof. Let Λ and ψ be supported on some balanced compact set K. Let V be a bounded, balanced open set containing K. If ϕ_0 is a function with compact support equal to ϕ on $V+K$, then for $x \in V$,

$$(\phi * \psi)(x) = \int \phi(x-y)\psi(y)\,dy = \int \phi_0(x-y)\psi(y)\,dy = (\phi_0 * \psi)(x)$$

Thus

$$(\Lambda * (\phi * \psi))(0) = (\Lambda * (\phi_0 * \psi))(0) = ((\Lambda * \psi) * \phi_0)(0)$$

But $\Lambda * \psi$ is supported on $K+K$, so $((\Lambda * \psi) * \phi_0)(0) = ((\Lambda * \psi) * \phi)(0)$. Now we also calculate

$$(\Lambda * (\phi * \psi))(0) = ((\Lambda * \phi_0) * \psi)(0) = ((\Lambda * \phi) * \psi)(0) \int (\Lambda * \phi_0)(-y)\psi(y)$$

where the last fact follows because $\Lambda * \phi_0$ agrees with $\Lambda * \phi$ on K. The general fact follows by applying the translation operators. □

Now we come to the grand finale, defining the convolution of two distributions. Given two distributions Λ and Ψ, one of which has compact support, we define the linear operator

$$L(\phi) = \Lambda * (\Psi * \phi)$$

Then L commutes with translations, and is continuous, because if we have $\phi_1, \phi_2, \dots$ converging to ϕ in $C_c^\infty(K)$, then $\Psi * \phi_n$ converges to $\Psi * \phi$ in $C^\infty(\mathbf{R}^n)$. If Ψ is supported on a compact support C, then the $\Psi * \phi_n$ have common compact support $C + K$, and actually converge in $C_c^\infty(C + K)$, hence $\Lambda * (\Psi * \phi_n)$ converges to $\Lambda * (\Psi * \phi)$. Conversely, if Λ has compact support, then $\Psi * \phi_n$ converges in $C^\infty(\mathbf{R}^n)$, which implies $\Lambda * (\Psi * \phi_n)$ converges to $\Lambda * (\Psi * \phi)$ in $C^\infty(\mathbf{R}^n)$. Thus L corresponds to a distribution, and we define this distribution to be $\Lambda * \Psi$.

Theorem 9.22. *If Λ and Ψ are distributions, one of which has compact support, then $\Lambda * \Psi = \Psi * \Lambda$. Let S_Λ and S_Ψ, and $S_{\Lambda * \Psi}$ denote the supports of Λ, Ψ, and $\Lambda * \Psi$. Then $\Lambda * \Psi = \Psi * \Lambda$, and $S_{\Lambda * \Psi} \subset S_\Lambda + S_\Psi$.*

Proof. We calculate that for any two test functions ϕ and ψ,

$$(\Lambda * \Psi) * (\phi * \psi) = \Lambda * (\Psi * (\phi * \psi)) = \Lambda * ((\Psi * \phi) * \psi)$$

If Λ has compact support, then

$$\Lambda * ((\Psi * \phi) * \psi) = (\Lambda * \psi) * (\Psi * \phi)$$

Conversely, if Ψ has compact support, then

$$\Lambda * ((\Psi * \phi) * \psi) = \Lambda * (\psi * (\Psi * \phi)) = (\Lambda * \psi) * (\Psi * \phi)$$

We also calculate

$$\begin{aligned}\Psi * ((\Lambda * \phi) * \psi) &= \Psi * (\Lambda * (\phi * \psi)) = \Psi * (\Lambda * (\psi * \phi))\\ &= \Psi * ((\Lambda * \psi) * \phi) = (\Psi * \phi) * (\Lambda * \psi)\end{aligned}$$

But since convolution is commutative, we have

$$((\Lambda * (\Psi * \phi)) * \psi) = \Lambda * ((\Psi * \phi) * \psi) = \Psi * ((\Lambda * \phi) * \psi) = (\Psi * (\Lambda * \phi)) * \psi$$

Since ψ was arbitrary, we conclude

$$(\Lambda * \Psi) * \phi = \Lambda * (\Psi * \phi) = \Psi * (\Lambda * \phi) = (\Psi * \Lambda) * \phi$$

and now since ϕ was arbitrary, we conclude $\Lambda * \Psi = \Psi * \Lambda$. Now we know convolution is commuatative, we may assume S_Ψ is compact. The support of $\Psi * \phi^*$ lies in $S_\Psi - S_\phi$. But this means that if $S_\phi - S_\Psi$ is disjoint from S_Λ, which means exactly that S_ϕ is disjoint from $S_\Lambda + S_\Psi$, then

$$(\Lambda * \Psi)(\phi) = (\Lambda * (\Psi * \phi))(0) = 0$$

and this gives the support of $\Lambda * \Psi$. $\square$

This means that the convolution of two distributions with compact support also has compact support. This means that if we have three distributions Λ, Ψ, and Φ, two of which have compact support, then the distributions $\Lambda * (\Psi * \Phi)$ and $(\Lambda * \Psi) * \Phi$ are well defined, so convolution is associative and commutative. We calculate that for any test function ϕ,

$$(\Lambda * (\Psi * \Phi)) * \phi = \Lambda * (\Psi * (\Phi * \phi))$$

$$((\Lambda * \Psi) * \Phi) * \phi = (\Lambda * \Psi) * (\Phi * \phi)$$

If Φ has compact support, then $\Phi * \phi$ has compact support, and so we can move $(\Lambda * \Psi)$ into the equation to prove equality. If Φ does not have compact support, then Λ and Ψ have compact support, and

$$\Lambda * (\Psi * \Phi) = \Lambda * (\Phi * \Psi)$$

and we can apply the previous case to obtain that this is equal to $(\Lambda * \Phi) * \Psi$. Repeatedly applying the previous case brings this to what we want.

Theorem 9.23. *If Λ and Ψ are distributions, then*

$$D^\alpha(\Lambda * \Psi) = (D^\alpha \Lambda) * \Psi = \Lambda * (D^\alpha \Psi)$$

Proof. The Dirac delta function δ satisfies

$$(\delta * \phi)(x) = \int \phi(y)\delta(x-y)\, dy = \phi(x)$$

so $\delta * \phi = \phi$. Now $D^\alpha \delta$ is also supported at x, since

$$(D^\alpha \delta)(\phi) = (-1)^{|\alpha|} \int \delta(x)(D^\alpha \phi)(x)\, dx = (-1)^{|\alpha|}(D^\alpha \phi)(0)$$

which means that for any distribution Λ, then $(D^\alpha \delta) * \Lambda$ has compact support,

$$(((D^\alpha \delta) * \Lambda) * \phi)(0) = (D^\alpha \delta)((\Lambda * \phi)^*) = (-1)^{|\alpha|} D^\alpha (\Lambda * \phi)^* = ((D^\alpha \Lambda) * \phi)(0)$$

which verifies that $(D^\alpha \delta) * \Lambda = \delta * (D^\alpha \Lambda)$. But now we find

$$D^\alpha(\Lambda * \Psi) = (D^\alpha \delta) * \Lambda * \Psi = ((D^\alpha \delta) * \Lambda) * \Psi = D^\alpha \Lambda * \Psi$$

$$D^\alpha(\Lambda * \Psi) = D^\alpha(\Psi * \Lambda) = (D^\alpha \Psi) * \Lambda = \Lambda * (D^\alpha \Psi)$$

which verifies the theorem in general. □

9.6 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

$$|f_\alpha(x)| \leqslant \frac{C_{\alpha,m}}{1 + |x|^m}.$$

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

$$\|f\|_{n,m} = \sup_{|\beta| \leqslant n} \|(1+|x|)^m f_\beta\|_{L^\infty(\mathbf{R}^d)}.$$

We then consider

$$\mathcal{S}(\mathbf{R}^d) = \left\{ f : \mathbf{R}^d \to \mathbf{R} : \text{for all } n, m, \|f\|_{n,m} < \infty \right\}.$$

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 $\|f\|_{\alpha,\beta} = \|x^\alpha f_\beta\|_{L^\infty(\mathbf{R}^d)}$, 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 9.24. *$\mathcal{S}(\mathbf{R}^d)$ 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 $\|f_k - f\|_{n,m} \to 0$ 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 $f_k \to f \in \mathcal{S}(\mathbf{R}^d)$. $\square$

Example. *The Gaussian function $\phi : \mathbf{R}^d \to \mathbf{R}$ 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*

$$|(1+|x|)^m \phi_\alpha| \leqslant (1+|x|)^m |P_\alpha \phi| \lesssim \frac{1+|x|^{|\alpha|+m}}{1+|x|^{|\alpha|+m}} = 1.$$

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 $\Lambda : \mathcal{S}(\mathbf{R}^d) \to \mathbf{R}$, 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 9.25. *If g is a function, with g and all it's 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 $|\beta| \leqslant n$,

$$|g_\beta(x)| \leqslant A \cdot (1+|x|)^M.$$

Consider $|\alpha| \leqslant n$. Then the Leibnitz formula implies that

$$\begin{aligned}(1+|x|)^m |(gf)_\alpha| &\leqslant 2^{|\alpha|} \sum_{\beta \leqslant \alpha} (1+|x|)^m |g_\beta f_{\alpha-\beta}| \\ &\leqslant A \cdot 2^n \sum_{\beta \leqslant \alpha} (1+|x|)^{m+M} |f_{\alpha-\beta}| \\ &\leqslant A \cdot 4^n \|f\|_{n,m+M}.\end{aligned}$$

Thus $\|gf\|_{n,m} \lesssim_{n,m} \|f\|_{n,m+M}$, which implies the operator is bounded. □

If f and g are Schwartz functions, and $|\alpha| \leqslant n$, then the Leibnitz formula again implies that if $k + k' = m$, then

$$(1+|x|)^m (gf)_\alpha \lesssim_n \sum_{\beta \leqslant \alpha} (1+|x|)^m g_\beta f_{\alpha-\beta} \lesssim_n \|f\|_{n,k} \|g\|_{n,k'}$$

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 9.26. *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 $(M_\xi f)(x) = e(\xi \cdot x) f(x)$.*
- *The L^p norms $\|f\|_{L^p(\mathbf{R}^n)}$, for $1 \leqslant p \leqslant \infty$.*
- *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 $|\alpha| \leqslant m$, $|\beta| \leqslant n$, then we can use the Leibnitz formula to conclude that

$$\begin{aligned} |\xi^\alpha \mathcal{F}(f)_\beta| &\lesssim_{\alpha,\beta} \mathcal{F}((x^\beta f)_\alpha) \\ &\lesssim_{\alpha,\beta} \max_{\gamma \leqslant \alpha \wedge \beta} |\mathcal{F}(x^{\beta-\gamma} f_{\alpha-\gamma})| \\ &\lesssim_{\alpha,\beta} \max_{\gamma \leqslant \alpha \wedge \beta} \|x^{\beta-\gamma} f_\gamma\|_{L^1(\mathbf{R}^d)} \\ &\leqslant \max_{\gamma \leqslant \alpha \wedge \beta} \|(1+|x|)^{|\beta|} f_\gamma\|_{L^1(\mathbf{R}^d)} \lesssim \|f\|_{|\alpha|,|\beta|+d+1}. \end{aligned}$$

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. □

Corollary 9.27. *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. □

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 $C_c^\infty(\mathbf{R}^d) \to \mathcal{S}(\mathbf{R}^d)$ 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 $\Lambda[f]$ to be the distribution*

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

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

$$|\Lambda[f](\phi)| \leqslant \|f\|_{L^p(\mathbf{R}^d)} \|\phi\|_{L^q(\mathbf{R}^d)}.$$

Since $\|\cdot\|_{L^q(\mathbf{R}^d)}$ is a continuous norm on $\mathcal{S}(\mathbf{R}^d)$, this shows $\Lambda[f]$ 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 $\Lambda[f]$ 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*

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

But this distribution is not always tempered. If $|\mu|$ is finite, the inequality $\|\Lambda[\mu](\phi)\| \leqslant \|\mu\| \|\phi\|_{L^\infty(\mathbf{R}^d)}$ 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*

$$|\Lambda[\mu](\phi)| \leqslant \|\mu(x)/(1+|x|^\alpha)\| \|\phi\|_{L^\infty(\mathbf{R}^d, 1+|x|^\alpha)}.$$

Example. *Suppose Λ is a distribution supported 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**

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

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

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

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 $\varepsilon \geqslant 0$. But this means that

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

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

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

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 9.28. *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

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

so $\widehat{1}$ is the Dirac delta distribution δ. Similarily,

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

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 9.29. *If μ is a finite measure, $\widehat{\mu}$ is a uniformly continuous bounded function with $\|\widehat{\mu}\|_{L^\infty(\mathbf{R}^d)} \leqslant \|\mu\|$, and*

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

The function $\widehat{\mu}$ 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

$$\widehat{\mu} \cdot \phi$$

If f is integrable, we can apply Fubini's theorem to conclude

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

This gives that the distribution $\widehat{\mu}$ is given by integration with respect the required function. It is easy to check that $\|\widehat{\mu}\|_{L^\infty(\mathbf{R}^d)} \leqslant \|\mu\|$. Moreover, $\widehat{\mu}$ is continuous, since by the dominated convergence theorem,

$$\widehat{\mu}(\xi + \eta) - \widehat{\mu}(\xi) = \int e(-\xi \cdot x)(e(-\eta \cdot x) - 1) d\mu(x)$$

as $h \to 0$, 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 $\widehat{\mu}$ is smooth if it has all moments, we calculate that if $|\eta| = 1$, and $t \in \mathbf{R}$, then

$$\frac{\widehat{\mu}(\xi + t\eta) - \widehat{\mu}(\xi)}{t} = \int e(-\xi \cdot x)(e(-t\eta \cdot x) - 1) \, d\mu(x).$$

□

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,

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

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

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

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

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 $\alpha \geqslant d$, we need to work with principal values TODO:

9.7 Convolution Operators

It is know that if $T : D(\mathbf{R}^d) \to C^\infty(\mathbf{R}^d)$ 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, with rely on a regularity result on the differentiation of functions in L^p spaces.

Lemma 9.30. *If $f \in L^p(\mathbf{R}^d)$, has derivatives in the L^p norm of all orders $\leqslant d+1$, then f is almost everywhere equal to a continuous function g such that*

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

where the hidden constant depends only on n and p.

Proof. s □

Theorem 9.31. *If $T : L^p(\mathbf{R}^d) \to L^q(\mathbf{R}^d)$ 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

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

The map $\phi \mapsto g_\phi(0)$ is therefore continuous on $\mathcal{S}(\mathbf{R}^d)$, 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 $T : L^p(\mathbf{R}^d) \to L^q(\mathbf{R}^d)$ 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 it's derivatives.

For each p and q, we will let (L^p, L^q) denote the space of tempered distributions which define a continuous linear map from $L^p(\mathbf{R}^d)$ 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* $(\Lambda * \Phi)^\wedge = \Phi \Lambda^\wedge$ *is an element of* $L^2(\mathbf{R}^d)$. *Thus we can think of* $\widehat{\Lambda} = e^{\pi|x|^2}(\Lambda * \Phi)^\wedge$ *as a function* f. *Plancherel's theorem implies that for any Schwarz function* ϕ,

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

But this means that $f \in L^\infty(\mathbf{R}^d)$, *for if there is a set* E *of positive measure where* $|f| \geqslant M$, *we can find* $\widehat{\phi}$ *with* $\widehat{\phi} = 1$ *on* E *and with* $\|\widehat{\phi}\|_2 = |E| + \varepsilon$, *and then* $\|f\widehat{\phi}\|_2 \geqslant M\|\widehat{\phi}\|_2$. *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* $\|\mu * \phi\|_1 \leqslant \|\mu\|_1\|\phi\|_1$. *Conversely, if* $\Lambda \in (L^1, L^1)$, *and* Φ_δ *is the Gauss kernel, then we set* $\Lambda_\delta = \Lambda * \Phi_\delta$. *By assumption,* Λ_δ *is an* L^1 *function, and so* Λ, $\|\Lambda_\delta\|_1 \lesssim \|\Phi_\delta\|_1 = 1$. *This implies that the* Λ_δ *are uniformly bounded in* L^1, *so by the Banach Alaoglu theorem, since* $L^1(\mathbf{R}^d)$ *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 9.32. *For any two (p, q), $(L^p, L^q) = (L^{q^*}, L^{p^*})$.*

Chapter 10

Sobolev Spaces

Let Ω be an open subset of $\mathbf{R}^d$. A natural problem when studying smooth functions $\phi \in C_c^\infty(\Omega)$ is to obtain estimates on the partial derivatives of ϕ. For instance, one can consider the norms

$$\|\phi\|_{C^n(\Omega)} = \max_{|\alpha| \leqslant n} \|D^\alpha f\|_{L^\infty(\Omega)}.$$

The space $C_c^\infty(\Omega)$ is not complete with respect to this norm, but it's completion is the space $C_b^n(\Omega)$ of n times bounded continuously differentiable functions on Ω, which still consists of regular functions. Unfortunately, such estimates are only encountered in the most trivial situations. As in the non-smooth case, one can often get much better estimates using the L^p norms of the derivatives, i.e. considering the norms

$$\|\phi\|_{W^{n,p}(\Omega)} = \left(\sum_{|\alpha| \leqslant p} \|D^\alpha \phi\|_{L^p(\Omega)}^p \right)^{1/p}.$$

As might be expected, $C_c^\infty(\Omega)$ is not complete with respect to the $W^{n,p}(\Omega)$ norm. However, it's completion cannot be identified with a family of n times differentiable functions. Instead, to obtain a satisfactory picture of the compoetion under this norm, a Banach space we will denote by $W^{n,p}(\Omega)$, we must take a distribution approach.

For each multi-index α, if f and f_α are locally integrable functions on Ω, we say f_α is a weak derivative for f if for any $\phi \in C_c^\infty(\Omega)$,

$$\int_\Omega f_\alpha(x)\phi(x)\, dx = (-1)^{|\alpha|} \int_\Omega f(x)\phi_\alpha(x)\, dx.$$

In other words, this is the same as the derivative of f viewed as a distribution on Ω. We define $W^{n,p}$ to be the space of all functions $f \in L^p(\Omega)$ such that for each $|\alpha| \leqslant n$, a weak derivative f_α exists and is an element of $L^p(\Omega)$. We then define

$$\|f\|_{W^{n,p}(\Omega)} = \left(\sum_{|\alpha| \leqslant n} \|f_\alpha\|_{L^p(\Omega)} \right)^{1/p}.$$

Where this sum is treated as a maximum in the case $p = \infty$. Later on we will be able to show this space is a complete Banach space.

Example. *Let B be the open unit ball in $\mathbf{R}^d$, and let $u(x) = |x|^{-s}$, where $s < n-1$. For which p is $u \in W^{1,p}(B)$? We calculate by an integration by parts that if $\phi \in C_c^\infty(B)$, we fix $\varepsilon > 0$ and write*

$$\int_B \phi_i(x) u(x)\, dx = \int_{|x| \leqslant \varepsilon} \phi_i(x) u(x) + \int_{\varepsilon < |x| \leqslant 1} \phi_i(x) u(x).$$

The integral on the ε ball is neglible since $s < n$. Since u is smooth away from the origin, it's distributional derivative agrees with it's standard derivative, which is

$$u_i(x) = \frac{-\alpha x_i}{|x|^{s+2}}.$$

Thus $|u_i| \lesssim 1/|x|^{s+1}$. An integration by parts gives

$$\int_{\varepsilon < |x| \leqslant 1} \phi_i(x) u(x) = \int_{|x| = \varepsilon} \phi(x) u(x) \nu_i \, dS + \int_{\varepsilon < |x| \leqslant 1} \frac{s\phi(x) x_i}{|x|^{s+2}}\, dx,$$

where ν_i is the normal vector to the sphere pointing inward. Since $s < n-1$, the surface integral tends to zero as $\varepsilon \to 0$. Thus the weak derivative of u is equal to the standard derivative. Consequently, $u \in W^{1,p}(B)$ if $s < n/p - 1$.

Example. *If $\{r_k\}$ is a countable, dense subset of B, then we can define*

$$u(x) = \sum_{k=1}^\infty \frac{|x - r_k|^{-s}}{2^k}$$

Then $u \in W^{1,p}(B)$ if $0 < \alpha < n/p - 1$, yet u has a dense family of singularities, and thus does not behave like any differentiable function we would think of.

Theorem 10.1. *For each $k \in \mathbf{N}$ and $1 \leqslant p \leqslant \infty$, $W^{k,p}(\Omega)$ is a Banach space.*

Proof. It is easy to verify that $\|\cdot\|_{W^{k,p}}$ is a norm on $W^{k,p}(\Omega)$. Let $\{u_n\}$ be a Cauchy sequence in $W^{k,p}(\Omega)$. In particular, this means that $\{D^\alpha u_n\}$ is a Cauchy sequence in $L^p(\Omega)$ for each multi-index α with $|\alpha| \leqslant k$. In particular, these are functions v_α such that $D^\alpha u_n$ converges to v_α in the L^p norm for each α. Thus it suffices to prove that if $v = \lim u_n$, then $D^\alpha v = v_\alpha$ for each α. But this follows because the Hölder inequality implies that for each fixed $\phi \in C_c^\infty(\Omega)$,

$$\begin{aligned}(-1)^{|\alpha|} \int \phi_\alpha(x) v(x)\, dx &= \lim_{n \to \infty} (-1)^{|\alpha|} \phi_\alpha u_n(x)\, dx \\ &= \lim_{n \to \infty} \int \phi(x)(D^\alpha u_n)(x)\, dx \\ &= \int \phi(x) v_\alpha(x)\, dx.\end{aligned}$$

Thus $W^{k,p}(\Omega)$ is complete. □

10.1 Smoothing

It is often useful to be able to approximate elements of $W^{k,p}(\Omega)$ by elements of $C^\infty(\Omega)$. This is mostly possible. If $u \in W^{k,p}(\Omega)$, and $\{\eta_\varepsilon\}$ is a family of smooth mollifiers, then, viewing u as a function on $\mathbf{R}^n$ supported on Ω, we can consider the convolution $u^\varepsilon = u * \eta_\varepsilon$, i.e. the function defined by setting

$$u^\varepsilon(x) = \int_\Omega u(x-y)\eta_\varepsilon(y)\, dy.$$

This is just normal convolution, where we identify the function u with the function $u\mathbf{I}_\Omega$ on $\mathbf{R}^d$. Then u^ε is a smooth function on $\mathbf{R}^d$ supported on a ε thickening of Ω. However, u^ε does not necessarily converge to u in $W^{k,p}(\Omega)$ as $\varepsilon \to 0$, since the behaviour of the convolution can cause issues at the boundary of Ω, where the distributional derivative $D^\alpha(u\mathbf{I}_\Omega)$ does not behave like a locally integrable function. This is the only problem, however.

Theorem 10.2. *If $U \Subset \Omega$, then $\lim_{\varepsilon \to 0} \|u^\varepsilon - u\|_{L^p(U)} = 0$.*

Proof. For each $\varepsilon > 0$, let $U^\varepsilon = \{x \in \Omega : d(x, \partial\Omega) > \varepsilon\}$. If $x \in \Omega^\varepsilon$, then

$$((D^\alpha u) * \eta_\varepsilon)(x) = (u_\alpha \mathbf{I}_\Omega * \eta_\varepsilon)(x),$$

since the convolution only depends on the behaviour of $D^\alpha u$ on a ε ball around x, which is contained in the interior of Ω. We can apply standard results about mollifiers to conclude that $u_\alpha \mathbf{I}_\Omega * \eta_\varepsilon$ converges to $u_\alpha \mathbf{I}_\Omega$ in $L^p(\mathbf{R}^d)$ as $\varepsilon \to 0$. Since $U \Subset \Omega$, we have $U \subset U^\varepsilon$ for small enough ε, and so $(D^\alpha u) * \eta_\varepsilon$ converges to u_α in $L^p(U)$ as $\varepsilon \to 0$. Since this is true for each α with $|\alpha| \leqslant k$, we obtain the result. □

If we are a little more careful, then we can fully approximate elements of $W^{k,p}(\Omega)$ by smooth functions on U.

Theorem 10.3. *$C_c^\infty(\Omega) \cap W^{k,p}(\Omega)$ is dense in $W^{k,p}(\Omega)$.*

Proof. Consider a family of open sets $\{V_n\}$ such that $V_n \Subset \Omega$ for each n, and $U = \bigcup V_n$. Then we can consider a smooth partition of unity $\{\xi_n\}$ subordinate to the cover $\{V_n\}$. For each $u \in W^{k,p}(\Omega)$, we can write $u = \sum_n u\xi_n$. In particular, this means that for each $\varepsilon > 0$, there is N such that $\|\sum_{n=N+1}^\infty u\xi_n\|_{W^{k,p}(\Omega)} \leqslant \varepsilon$. For each $n \in \{1, \dots, N\}$, we can find δ_n small enough that the δ_n thickening of V_n is compactly contained in Ω. If ε_n is small enough, we find $(u\xi_n)^{\varepsilon_n}$ is supported on the δ_n thickening of V_n, and $\|(u\xi_n)^{\varepsilon_n} - u\xi_n\|_{W^{k,p}(V_n)} \leqslant \varepsilon/N$. But we then find

$$\|u - \sum_{n=1}^N (u\xi_n)^{\varepsilon_n}\|_{W^{k,p}(\Omega)} \leqslant \varepsilon + \sum_{n=1}^N \|u\xi_n - (u\xi_n)^{\varepsilon_n}\|_{W^{k,p}(\Omega)} \leqslant 2\varepsilon.$$

Thus $C_c^\infty(\Omega)$ is dense in $W^{k,p}(\Omega)$. □

Approximation by elements of $C^\infty(\overline{\Omega})$ requires some more care, and additional assumptions on the behaviour of $\partial\Omega$.

Chapter 11

Basics of Kernel Operators

We now consider a very general family of operators, which can be seen as the infinite dimensional analogue of matrix multiplication. We fix two measure spaces X and Y, and consider a function $K : X \times Y \to \mathbf{C}$, which we call a *kernel*. From this kernel, we obtain an induced operator T_K taking functions on X to functions on Y, given by the integral formula

$$(T_K f)(y) = \int_X K(x,y) f(x)\, dx.$$

Our goal is to understand what properties of K imply boundedness of the operator T_K.

Example. *If $X = Y = \mathbf{R}^d$, equipped with the Lebesgue measure. If we set $K(x,\xi) = e^{2\pi i \xi \cdot x}$, then we can use this function as a kernel to obtain an integral operator*

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

Thus the Fourier transform is a kernel operator.

Example. *Let $X = \{1,\dots,N\}$ and $Y = \{1,\dots,M\}$, each equipped with the counting measure. Then each kernel K corresponds to an $M \times N$ matrix A, with $A_{ij} = K(j,i)$, and then*

$$(T_K f)(m) = \sum_{n=1}^N f(n) K(n,m) = \sum_{n=1}^N A_{mn} f(n),$$

so with respect to the standard basis, T_K is just given by matrix multiplication by A.

In the case where we are mapping *from* $L^1(X)$, or *into* $L^\infty(Y)$, the conditions on K which determine boundedness are rather trivial. This can be taken as a motivation for introducing the L^p norms, since these norm allow us to study the boundedness properties of K with more detail. Since it is not even obvious that T_K is well defined even for simple functions, in the very weak case where K is measurable, we introduce the sublinear analogue S_K, defined by setting

$$(S_K f)(y) = \int |f(x)||K(x,y)|\, dx.$$

Due to the flexibility of non-negative integrals, this operator is defined for any measurable f, though the output may be infinite for various values of y.

Theorem 11.1. *Fix $q \geqslant 1$. If $\|K\|_{L^q(Y)L^\infty(X)} < \infty$, then S_K is bounded as an operator from $L^1(X)$ to $L^q(Y)$, with operator norm bounded above by $\|K\|_{L^q(Y)L^\infty(X)}$, with equality if X and Y are σ finite. Correspondingly, for each $f \in L^1(X)$, we have*

$$\int K(x,y) f(x)\, dx < \infty \text{ for almost every } y,$$

and $\|T_K f\|_{L^q(Y)} \leqslant \|K\|_{L^q(Y)L^\infty(X)} \|f\|_{L^1(X)}$.

Proof. Applying Minkowski's inequality, we conclude that

$$\begin{aligned}
\|S_K f\|_{L^q(Y)} &= \left(\left(\int |f(x)||K(x,y)|\, dx \right)^q \right)^{1/q} \\
&\leqslant \int \left(\int |f(x)|^q |K(x,y)|^q\, dy \right)^{1/q} dx \\
&\leqslant \int |f(x)| \|K\|_{L^q(Y)}(x)\, dx \\
&\leqslant \|f\|_{L^1(X)} \|K\|_{L^q(Y)L^\infty(X)}.
\end{aligned}$$

To show tightness, consider the first case where K can be written as

$$\sum_{i=1}^N \sum_{j=1}^M a_{ij} \mathbf{I}_{E_i \times F_j},$$

where $E_1,\dots,E_N$ are disjoint, finite measure sets in X, and $F_1,\dots,F_M$ are disjoint, finite measure sets in Y. Then there exists $i \in \{1,\dots,N\}$ such that for each $x \in E_i$,

$$\left(\int |K(x,y)|^q \, dy\right)^{1/q} = \left(\sum_{j=1}^{M} |a_{ij}|^q |F_j|\right)^{1/q} = \|K\|_{L^q(Y)L^\infty(X)}.$$

If $f = \mathbf{I}_{E_i}$, then $\|f\|_{L^1(X)} = |E_i|$, and

$$\begin{aligned}\left(\left(\int |K(x,y)f(x)|\, dx\right)^q dy\right)^{1/q} &= \left(\sum_{j=1}^{M} |F_j||a_{ij}|^q|E_i|^q\right)^{1/q}\\ &= \|f\|_{L^1(X)}\|K\|_{L^q(Y)L^\infty(X)}.\end{aligned}$$

Thus f is an extremizer for S_K.

To show this inequality is tight. Let us first consider the case where $q < \infty$. By a monotone convergence result if X and Y are σ finite, we may assume that X and Y have finite measure. It then follows that for each $\varepsilon > 0$, there are functions $u_1,\dots,u_n \in L^1(X)$ and $v_1,\dots,v_n \in L^1(Y)$ such that $\|K - u_1 \otimes v_1 - \dots - u_n \otimes v_n\|_{L^1(X\times Y)} < \varepsilon$. □

Lemma 11.2. *BLAH*

Proof. Let Π be the family of all sets $E \times F \subset X \times Y$, where E is a measurable subset of X, and F is a measurable subset of Y. Then Π is a π system, in the sense that if $E_1 \times F_1, E_2 \times F_2 \in \Pi$, then $(E_1 \times F_1) \cap (E_2 \times F_2) = (E_1 \cap E_2) \times (F_1 \cap F_2) \in \Pi$. Now let

$$\Delta = \left\{ G \subset X \times Y : \begin{pmatrix} \text{for all } \varepsilon > 0\text{, there are simple } u_1,\dots,u_n \\ \text{on } X \text{ and } v_1,\dots,v_n \text{ on } Y \text{ such that} \\ \|\mathbf{I}_G - \sum u_i \otimes v_i\|_{L^q(Y)L^\infty(X)} < \varepsilon \end{pmatrix} \right\}.$$

If $G \in \Delta$, then $G^c \in \Delta$, since $1 = \mathbf{I}_X \otimes \mathbf{I}_Y$. Thus if $\|\mathbf{I}_G - \sum u_i \otimes v_i\|_{L^q(Y)L^\infty(X)} < \varepsilon$, then

$$\mathbf{I}_{G^c} - (\mathbf{I}_X \otimes \mathbf{I}_Y - \sum u_i \otimes v_i) = (1 - \mathbf{I}_G) - (1 - \sum u_i \otimes y_i) = \sum u_i \otimes y_i - \mathbf{I}_G,$$

and so $\|\mathbf{I}_{G^c} - (\mathbf{I}_X \otimes \mathbf{I}_Y - \sum u_i \otimes v_i)\|_{L^q(Y)L^\infty(X)} < \varepsilon$. If $G_1, G_2, \dots$ are a disjoint family of sets in Δ, then for each $\varepsilon > 0$, and for each k we can find $u_{k1},\dots,u_{kN_k}$ and $v_{k1},\dots,v_{kN_k}$ such that

$$\|\mathbf{I}_{G_k} - \textstyle\sum_i u_{ki} \otimes v_{ki}\|_{L^q(Y)L^\infty(X)} < \varepsilon/2^k.$$

By monotone convergence, if $G = \bigcup G_k$, then for each fixed x,

$$\lim_{N\to\infty} \int \mathbf{I}_G(x,y) - \sum_{k=1}^{N} \mathbf{I}_{G_k}(x,y)\, dx$$

$\square$

Chapter 12

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 12.1. *For any sequence u_n and set E of finite measure,*

$$\lim_{n\to\infty}\int_E \cos^2(nx+u_n)\,dx = |E|/2$$

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 $\int\chi_E(x)\sin(2nx)\to 0$ as $n\to\infty$, 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. ∎

Theorem 12.2 (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* $a_n, b_n \to 0$.

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 $r_n \to 0$. If this is not true, then we must have $\cos(nx + c_n) \to 0$ for every $x \in E$. In particular, the dominated convergence theorem implies that

$$\lim_{n\to\infty} \int_E \cos(nx + c_n)^2 \, dx = 0$$

Yet we know this tends to $|E|/2$ as $n \to \infty$, which is a contradiction. $\square$

TODO: EXPAND ON THIS FACT.

12.1 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 $f = \sum_{|n| \leqslant N} c_n e_n$ as $Tf = \sum_{|n| \leqslant N} \lambda_n c_n e_n$. 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 $\lambda_n = [|n| \leqslant N]$. The truncation is continuous, since for any integrable function f, the Fourier coefficients are uniformly bounded by $\|f\|_1$, so $\|S_N f\|_1 \leqslant N\|f\|_1$. Similarily, the Féjer truncation σ_N associated to the multipliers $\lambda_N = [|n| \leqslant N](1 - |n|/N)$ 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*

$$|A_r f| = \left|\sum r^{|n|}\widehat{f}(n)e_n(t)\right| \leqslant \|f\|_1 \sum r^{|n|} = \|f\|_1 \left(1 + \frac{2}{1-r}\right)$$

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 coeficients

$$\lambda_n = \frac{\text{sgn}(n)}{i} = \begin{cases} +1/i & n > 0 \\ 0 & n = 0 \\ -1/i & n < 0 \end{cases}$$

Because

$$[|n| \leqslant N] = \frac{\text{sgn}(n+N) - \text{sgn}(n-N)}{2} + \frac{[n = N] + [n = -N]}{2}$$

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 $L^1(\mathbf{T})$, we conclude that the operators S_n are uniformly bounded as endomorphisms on $L^p(\mathbf{T})$ provided that H is bounded as an operator from $L^p(\mathbf{T})$ to $L^q(\mathbf{T})$. 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 $S_N f \to f$ 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 their is an integrable function g whose Fourier coefficients corresponds to the multipliers of an operator T, then $f * g = Tf$ 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 $\Lambda * f = Tf$ 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 12.3. *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,

$$\|Tf\|_2^2 = \sum \left|\widehat{Tf}(n)\right|^2 = \sum |\lambda_n|^2 \left|\widehat{f}(n)\right|^2$$

If the λ_n are bounded, then we can obtain from this formula the bound

$$\|Tf\|_2^2 \leqslant \max |\lambda_n| \|f\|_2^2$$

Conversely, if Tf is bounded, then

$$|\lambda_n^2| = \|T(e_n)\|_2^2 \leqslant \|T\|^2$$

so the λ_n are bounded. $\square$

Corollary 12.4. *The Hilbert transform is a bounded endomorphism on $L^2(\mathbf{T})$. Note that we already know that $S_N f \to f$ 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 $n \to \infty$. Nonetheless, we can use Abel summation to treat the Hilbert transform as convolution with an appropriate operator. For $0 < r < 1$, consider, for $z = e^{it}$,

$$K_r(z) = \sum_{n \in \mathbf{Z}} \frac{\mathrm{sgn}(n)}{i} r^{|n|} z^n = K * P_r$$

Since we know the Hilbert transform is continuous in $L^2(\mathbf{T})$, we can conclude that, in particular, for any C^∞ function f,

$$Hf = \lim_{r \to 1} K * (P_r * f) = \lim_{r \to 1} (K * P_r) * f = \lim_{r \to 1} K_r * f$$

So it suffices to determine the limit of the K_r. We find that

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{(rz)^n - (r\overline{z})^n}{i} &= \frac{r}{i}\left(\frac{1}{\overline{z} - r} - \frac{1}{z - r}\right) = \frac{r}{i} \frac{z - \overline{z}}{|z|^2 - 2r\mathrm{Re}(z) + r^2} \\
&= \frac{2r\sin(t)}{1 - 2r\cos(t) + r^2} = \frac{4r\sin(t/2)\cos(t/2)}{(1-r)^2 + 4r\sin^2(t/2)} \\
&= \cot(t/2) + O\left(\frac{(1-r)^2}{t^3}\right)
\end{aligned}$$

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

$$\left| \int_{|t| \geqslant \varepsilon} [K_r(t) - \cot(t/2)] f(t) \right| \lesssim (1-r)^2 \|f\|_\infty \int_{|t| \geqslant \varepsilon} \frac{dt}{|t|^3} \lesssim \frac{(1-r)^2 \|f\|_\infty}{\varepsilon^2}$$

$$\begin{aligned}
\left| \int_{|t| < \varepsilon} K_r(t) f(t)\, dt \right| &\leqslant \int_0^\varepsilon |K_r(t)||f(t) - f(-t)| \\
&\lesssim \int_0^\varepsilon |tK_r(t)||f'(0)| \lesssim \frac{|f'(0)|}{(1-r)^2} \int_0^\varepsilon t^2 \lesssim \|f'\|_\infty \frac{\varepsilon^3}{(1-r)^2}
\end{aligned}$$

$$\left| \int_{|t|<\varepsilon} \cot(t/2) f(t)\, dt \right| \lesssim \int_0^\varepsilon \frac{|f(t) - f(-t)|}{t} \lesssim \varepsilon f'(0)$$

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,

$$(Hf)(x) = \lim_{r \to 1} \int \cot(t/2) f(x-t)\, dt$$

12.2 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

$$\sum_{n=1}^{\infty} \frac{e_n(t)}{n} \qquad \sum_{n=-\infty}^{-1} \frac{e_n(t)}{n}$$

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

$A_r(f)(t) =$

12.3 Conjugate Fourier Series

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

$$\sum_{n=-\infty}^{\infty} \hat{f}(n) e_n(t) = \mathrm{Re}\left(\hat{f}(0) + 2\sum_{n=1}^{\infty} \hat{f}(n) e_n(t)\right)$$

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

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

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 13

Oscillatory Integrals

The goal of the theory of oscillatory integrals is to obtain estimates of integrals with highly oscillatory integrals, where standard techniques such as taking in absollute values, or various spatial decomposition strategies, fail completely to give tight estimates. A typical oscillatory integral is of the form

$$I(\lambda) = \int e(\lambda \Phi(x)) \psi(x)\, dx$$

where Φ is the *phase*, ψ is the *amplitude*, and λ is a parameter measuring the degree of oscillation, often allowed to increase to large values. Obtaining bounds on these integrals for large λ requires techniques to control cancellation properties of integrals, so decomposition techniques like those found in the theory of singular integrals completely fail here.

Example. *The most basic example of an oscillatory integral is the Fourier transform, where for each function $f \in L^1(\mathbf{R})$, and each $\xi \in \mathbf{R}$, we consider the quantity*

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

This f plays the role of the amplitude, we have a phase function $\Phi(x) = x$, and ξ takes the role of λ. In the basic theory of the Fourier transform, we showed that one has qualitative decay in this oscillatory integral for any $f \in L^1(\mathbf{R})$, a consequence of the Riemann-Lebesuge lemma. If f was appropriately smooth, i.e. Schwartz, then $\widehat{f}$ also exhibited polynomial decay properties. One of the main applications of the theory of oscillatory integrals is to show that Fourier *transforms of measures have fast decay properties, which has applications in many different areas of analysis.*

There are two modern tools to estimate oscillatory integrals. The first, the principle of stationary phase, states that if Φ is smooth, and $\nabla\Phi$ has an isolated family of zeroes, then the oscillatory integral asymptotics can be localized to regions around the values x_0 with $\nabla\Phi(x_0) = 0$. Heuristically, each zero x_0 contributes $\psi(x_0)e(\lambda\Phi(x_0))$, times the volume of the region around x_0 where Φ deviates by $O(1/\lambda)$ to the overall asymptotics. The second method, known as the method of steepest descent, uses complex analysis to shift domains of integration to a domain where less oscillation occurs, and standard dyadic decomposition strategies can be employed. However, this method seems to only work for single-variable oscillatory integrals, and therefore has limited applicability in many problems.

13.1 The Principle of Stationary Phase

The basic principle of the theory of stationary phase is that the asymptotics of the integral are determined by the points where $\nabla\Phi$ vanishes. The fact that we can reduce our analysis to critical points is because an oscillatory principle whose phase has no critical points exhibits rapid decay. Let us consider a simple example of an oscillatory integral, i.e.

$$I(\lambda) = \int_0^1 e^{i\lambda\phi(x)}\, dx,$$

where $\phi : [0,1] \to \mathbf{R}$ is Borel measurable. Taking in absolute values shows that $|I(\lambda)| \leqslant 1$ for all λ. If ϕ is constant, then $I(\lambda) = e^{i\lambda\phi}$ has magnitude one for all λ, so this inequality is sharp. But if ϕ varies, we expect I to decay as $\lambda \to \infty$, and so we hope to determine what properties guarantee that $|I(\lambda)| \ll 1$. For instance, the Esseén concentration inequality relates the lack of concentration of the distribution of ϕ to the *average* decay of I.

Theorem 13.1 (Esseén Concentration Inequality)**.** *Let $\phi : [0,1] \to \mathbf{R}$ be Borel measurable, and for each $\lambda \in \mathbf{R}$, set*

$$I(\lambda) = \int_0^1 e^{i\lambda\phi(x)}\, dx.$$

Then for any $\varepsilon > 0$,

$$\sup_{\phi_0 \in \mathbf{R}} |\{x \in [0,1] : |\phi(x) - \phi_0| \leqslant \varepsilon\}| \lesssim \varepsilon \int_0^{1/\varepsilon} |I(\lambda)| \, d\lambda.$$

Proof. For any choice of ϕ_0, we may replace ϕ with $\phi - \phi_0$, reducing the analysis to the case where $\phi_0 = 0$. Similarily, replacing ϕ with ϕ/ε reduces us to the situation where $\varepsilon = 1$. Thus we must show

$$|\{x \in [0,1] : |\phi(x)| \leqslant 1\}| \lesssim \int_0^1 |I(\lambda)| \, d\lambda.$$

Now let ψ be an integrable function supported on $[0,1]$. By Fubini's theorem,

$$\int_0^1 \psi(\lambda) I(\lambda) \, d\lambda = \int_0^1 \widehat{\psi}(-\phi(x)/2\pi) \, dx.$$

In particular, this means that

$$\left| \int_0^1 \widehat{\psi}(-\phi(x)/2\pi) \, dx \right| \leqslant \|\psi\|_{L^\infty[0,1]} \int_0^1 |I(\lambda)| \, d\lambda.$$

If we choose a bounded function ψ such that $\widehat{\psi}$ is non-negative, and bounded below on $[-2\pi, 2\pi]$, then

$$\left| \int_0^1 \widehat{\psi}(-\phi(x)/2\pi) \, dx \right| \gtrsim |\{x \in [0,1] : |\phi(x)| \leqslant 1\}|,$$

and so the claim follows easily. □

Thus if large cancellation happens in $I(\lambda)$ for the average λ, this automatically implies that ϕ cannot be concentrated around any particular point. Conversely, we want to show that if ϕ is significantly varying, then I exhibits some cancellation properties. The condition that ϕ' is bounded below is not sufficient to guarantee cancellation, as the next example shows, if ϕ has oscillation at wavelength $1/\lambda$.

Example. *Fix $\lambda_0 \in \mathbf{Z}$, and let $\phi(x) = 2\pi x + f(\lambda_0 x)/\lambda_0$, where f is smooth and 1-periodic, $\|f'\|_{L^\infty(\mathbf{R}/\mathbf{Z})} \leqslant 1$, and*

$$\int_0^1 e^{2\pi i x + i f(x)} \, dx \neq 0.$$

For instance, we could take $f(x) = x/2$. Then $|\phi'(x)|$ is bounded below independantly of λ_0. Since $\phi(x + 1/\lambda_0) = \phi(x) + 2\pi/\lambda_0$, we find $e^{i\lambda_0\phi(x)}$ is $1/\lambda_0$ periodic. In particular, this means

$$I(\lambda_0) = \int_0^1 e^{\lambda_0 i\phi(x)} = \int_0^1 e^{2\pi ix + if(x)}\, dx.$$

which is comparable to 1, independantly of λ_0.

One way to get around this condition is by controlling ϕ'' in addition to controlling ϕ'.

Theorem 13.2. *Let $\phi : \mathbf{R} \to \mathbf{R}$ be smooth, with $|\phi'(x)| \geqslant A$ and $|\phi''(x)| \leqslant B$ for all $x \in [0,1]$, where $A, B > 0$. Then for all $\lambda > 0$, we find $|I(\lambda)| \lesssim s$*

Theorem 13.3. *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

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

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

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

Taking absolute values in the last integral gives that

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

which gives the required bounds. □

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 $n \leqslant N$, $\Phi^{(n)}(a) = \Phi^{(n)}(b)$ and $\psi^{(n)}(a) = \psi^{(n)}(b)$, then we can still get $1/|\lambda|^N$ decay using the same proof as above, except now the integration by parts must take into account the endpoints of the integral, which now cancel to a suitably high degree.

13.2 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 $|\phi^{(n)}(x)| \geqslant 1$ on (a,b), then

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

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

Theorem 13.4. *There exists a constant C_n such that if ϕ is smooth in (a,b) with $|\phi^{(n)} \geqslant 1|$ for all $x \in (a,b)$, then*

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

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

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

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

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

$$\begin{aligned}\left| \int_a^b e(\lambda\phi) D^*(1)\, dx \right| &\leqslant |\lambda|^{-1} \int_a^b |(1/\phi')'|\, dx \leqslant |\lambda|^{-1} \left| \int_a^b (1/\phi')'\, dx \right| \\ &= |\lambda|^{-1} \left| \frac{1}{\phi'(b)} - \frac{1}{\phi'(a)} \right| \leqslant |\lambda|^{-1}\end{aligned}$$

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) $\phi^{(n+1)}(x) \geqslant 1$ on $[a,b]$. Let x_0 be the point at which $|\phi^{(n)}(x)|$ is minimized. Without loss of generality, we may assume that $|\phi^{(n)}(x)$ TODO: FINISH THIS ARGUMENT, GIVES $C_n = 5 \cdot 2^{n-1} - 2$. ☐

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

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

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

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

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

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

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

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

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 13.5. *If an amplitude ψ is present, then*

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

Proof. Integrating by parts, if $J(x) = \int_a^x e(\lambda\Phi(u))\,du$, 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 $|J(x)| \leqslant 8\lambda^{-1/2}$, the proposition follows. TODO: ADDRESS MULTIDIMENSIONAL CASE. □

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 $\cos(x) \geqslant 1/\sqrt{2}$, the second where $\sin(x) \geqslant 1/\sqrt{2}$. On the first part, we may apply the corollary above to obtain a $O(r^{-1/2})$ bound, and on the second, we can apply the first to obtain a $O(r^{-1})$ bound. Summing these two bounds up gives the theorem.

13.3 Surface Carried Measures

Theorem 13.6. *If a hypersurface Σ has non-vanishing Gauss curvature at each point in the support of a surface carried measure μ, then*

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

s

13.4 Restriction Theorems

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

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

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

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

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

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

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

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

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

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

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

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

Part III

Abstract Harmonic Analysis

The main property of spaces where Fourier analysis applies is symmetry – for a function $\mathbf{R}$, 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 14

Topological Groups

In abstract harmonic analysis, the main subject matter is the **topological group**, a group G equipped with a topology which makes the operation of multiplication and inversion continuous. In the mid 20th century, it was realized that basic Fourier analysis could be generalized to a large class of groups. The nicest generalization occurs over the locally compact groups, which simplifies the theory considerably.

Example. *There are a few groups we should keep in mind for intuition in the general topological group.*

- *The classical groups* $\mathbf{R}^n$ *and* $\mathbf{T}^n$*, from which Fourier analysis originated.*
- *The group* μ *of roots of unity, rational numbers* $\mathbf{Q}$*, and cyclic groups* $\mathbf{Z}_n$*.*
- *The matrix subgroups of the general linear group* $GL(n)$*.*
- *The product* $\mathbf{T}^\omega$ *of Torii, occurring in the study of Dirichlet series.*
- *The product* $\mathbf{Z}_2^\omega$*, which occurs in probability theory, and other contexts.*
- *The field of p-adic numbers* $\mathbf{Q}_p$*, which are the completion of* $\mathbf{Q}$ *with respect to the absolute value* $|p^{-m}q|_p = p^m$*.*

14.1 Basic Results

The topological structure of a topological group naturally possesses large amounts of symmetry, simplifying the spatial structure. For any topological group, the maps

$$x \mapsto gx \qquad x \mapsto xg \qquad x \mapsto x^{-1}$$

are homeomorphisms. Thus if U is a neighbourhood of x, then gU is a neighbourhood of gx, Ug a neighbourhood of xg, and U^{-1} a neighbourhood of x^{-1}, and as we vary U through all neighbourhoods of x, we obtain all neighbourhoods of the other points. Understanding the topological structure at any point reduces to studying the neighbourhoods of the identity element of the group.

In topological group theory it is even more important than in basic group theory to discuss set multiplication. If U and V are subsets of a group, then we define

$$U^{-1} = \{x^{-1} : x \in U\} \qquad UV = \{xy : x \in U, y \in V\}$$

We let $V^2 = VV$, $V^3 = VVV$, and so on.

Theorem 14.1. *Let U and V be subsets of a topological group.*

(i) If U is open, then UV is open.

(ii) If U is compact, and V closed, then UV is closed.

(iii) If U and V are connected, UV is connected.

(iv) If U and V are compact, then UV is compact.

Proof. To see that (i) holds, we see that

$$UV = \bigcup_{x \in V} Ux$$

and each Ux is open. To see (ii), suppose $u_i v_i \to x$. Since U is compact, there is a subnet u_{i_k} converging to y. Then $y \in U$, and we find

$$v_{i_k} = u_{i_k}^{-1}(u_{i_k} v_{i_k}) \to y^{-1}x$$

Thus $y^{-1}x \in V$, and so $x = yy^{-1}x \in UV$. (iii) follows immediately from the continuity of multiplication, and the fact that $U \times V$ is connected, and (iv) follows from similar reasoning. $\square$

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 14.2. *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 14.3. *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 14.4. *Every Kolmogorov topological group is Hausdorff.*

Theorem 14.5. *For any set $A \subset G$,*

$$\overline{A} = \bigcap_V AV$$

Where V ranges over the set of neighbourhoods of the origin.

Proof. If $x \notin \overline{A}$, then the last theorem guarantees that there is V for which $\overline{A}V$ and Ax are disjoint. We conclude $\bigcap AV \subset \overline{A}$. Conversely, any neighbourhood contains a closed neighbourhood, so that $\overline{A} \subset AV$ for a fixed V, and hence $\overline{A} \subset \bigcap AV$. ☐

Theorem 14.6. *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. ☐

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.

14.2 Quotient Groups

If G is a topological group, and H is a subgroup, then G/H can be given a topological structure in the obvious way. The quotient map is open, because VH is open in G for any open set V, and if H is normal, G/H is also a topological group, because multiplication is just induced from the quotient map of $G \times G$ to $G/H \times G/H$, and inversion from G to G/H. We should think the quotient structure is pleasant, but if no conditions on H are given, then G/H can have pathological structure. One particular example is the quotient $\mathbf{T}/\mu_\infty$ of the torus modulo the roots of unity, where the quotient is lumpy.

Theorem 14.7. *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 14.8. *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.

14.3 Uniform Continuity

An advantage of the real line $\mathbf{R}$ is that continuity can be explained in a *uniform sense*, because we can transport any topological questions about a certain point x to questions about topological structure near the origin via the map $g \mapsto x^{-1}g$. We can then define a uniformly continuous function $f : \mathbf{R} \to \mathbf{R}$ to be a function possessing, for every $\varepsilon > 0$, a $\delta > 0$ such that if $|y| < \delta$, $|f(x+y) - f(x)| < \varepsilon$. Instead of having to specify a δ for every point on the domain, the δ works uniformly everywhere. The group structure is all we need to talk about these questions.

We say a function $f : G \to H$ between topological groups is (left) uniformly continuous if, for any open neighbourhood U of the origin in H, there is a neighbourhood V of the origin in G such that for each x, $f(xV) \subset f(x)U$. Right continuity requires $f(Vx) \subset Uf(x)$. The requirement of distinguishing between left and right uniformity is important when we study non-commutative groups, for there are certainly left uniform maps which are not right uniform in these groups. If $f : G \to \mathbf{C}$, then left uniform continuity is equivalent to the fact that $\|L_x f - f\|_\infty \to 0$ as $x \to 1$, where $(L_x f)(y) = f(xy)$. Right uniform continuity requires $\|R_x f - f\|_\infty \to 0$, where $(R_x f)(y) = f(yx)$. R_x is a homomorphism, but L_x is what is called an antihomomorphism.

Example. *Let G be any Hausdorff non-commutative topological group, with sequences x_i and y_i for which $x_i y_i \to e$, $y_i x_i \to z \neq e$. Then the uniform structures on G are not equivalent.*

It is hopeless to express uniform continuity in terms of a new topology on G, because the topology only gives a local description of continuity, which prevents us from describing things uniformly across the whole group. However, we can express uniform continuity in terms of a new topology on $G \times G$. If $U \subset G$ is an open neighbourhood of the origin, let

$$L_U = \{(x,y) : yx^{-1} \in U\} \quad R_U = \{(x,y) : x^{-1}y \in U\}$$

The family of all L_U (resp. R_U) is known as the left (right) uniform structure on G, denoted $LU(G)$ and $RU(G)$. Fix a map $f : G \to H$, and consider the map

$$g(x,y) = (f(x), f(y))$$

from G^2 to H^2. Then f is left (right) uniformly continuous if and only if g is continuous with respect to $LU(G)$ and $LU(H)$ ($RU(G)$ and $RU(H)$). $LU(G)$ and $RU(G)$ are weaker than the product topologies on G and H, which reflects the fact that uniform continuity is a strong condition than normal continuity. We can also consider uniform maps with respect to $LU(G)$ and $RU(H)$, and so on and so forth. We can also consider uniform continuity on functions defined on an open subset of a group.

Example. *Here are a few examples of easily verified continuous maps.*

- *If the identity map on G is left-right uniformly continuous, then $LU(G) = RU(G)$, and so uniform continuity is invariant of the uniform structure chosen.*
- *Translation maps $x \mapsto axb$, for $a, b \in G$, are left and right uniform.*
- *Inversion is uniformly continuous.*

Theorem 14.9. *All continuous maps on compact subsets of topological groups are uniformly continuous.*

Proof. Let K be a compact subset of a group G, and let $f : K \to H$ be a continuous map into a topological group. We claim that f is then uniformly continuous. Fix an open neighbourhood V of the origin, and let V' be a symmetric neighbourhood such that $V'^2 \subset V$. For any x, there is U_x such that

$$f(x)^{-1} f(xU_x) \subset V'$$

Choose U'_x such that $U'^2_x \subset U_x$. The xU'_x cover K, so there is a finite subcover corresponding to sets $U'_{x_1}, \dots, U'_{x_n}$. 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. □

Corollary 14.10. *All maps with compact support are uniformly continuous.*

Corollary 14.11. *Uniform continuity on compact groups is invariant of the uniform structure chosen.*

14.4 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

$$a^{-1} = a^{-1} b b^{-1} > a^{-1} a b^{-1} = b^{-1}$$

This implies that the inverse image of an interval (a, b) under inversion is (b^{-1}, a^{-1}), hence inversion is continuous.

Now let $e < b < a$. We claim that there is then $e < c$ such that $c^2 < a$. This follows because if $b^2 \geqslant a$, then $b \geqslant ab^{-1}$ and so

$$(ab^{-1})^2 = ab^{-1}ab^{-1} \leqslant ab^{-1}b = a$$

Now suppose $a < e < b$. If $\inf\{y : y > e\} = x > e$, then $(x^{-1}, x) = \{e\}$, and the topology on G is discrete, hence the continuity of operations is obvious. Otherwise, we may always find c such that $c^2 < b$, $a < c^{-2}$, and then if $c^{-1} < g, h < c$, then

$$a < c^{-2} < gh < c^2 < b$$

so multiplication is continuous at every pair (x, x^{-1}). In the general case, if $a < gh < b$, then $g^{-1}ah^{-1} < e < g^{-1}bh^{-1}$, so there is c such that if $c^{-1} < g', h' < c$, then $g^{-1}ah^{-1} < g'h' < g^{-1}bh^{-1}$, so $a < gg'h'h < b$. The set of gg', where $c^{-1} < g' < c$, is really just the set of $gc^{-1} < x < gc$, and the set of $h'h$ is really just the set of $c^{-1}h < x < ch$. Thus multiplication is continuous everywhere.

Example (Dieudonne). *For any well ordered set S, the dictionary ordering on $\mathbf{R}^S$ induces a linear ordering inducing a topological group structure on the set of maps from S to $\mathbf{R}$.*

Let us study Dieudonne's topological group in more detail. If S is a finite set, or more generally possesses a maximal element w, then the topology on $\mathbf{R}^S$ can be defined such that $f_i \to f$ if eventually $f_i(s) = f(s)$ for all $s < w$ simultaneously, and $f_i(w) \to f(w)$. Thus $\mathbf{R}^S$ is isomorphic (topologically) to a discrete union of a certain number of copies of $\mathbf{R}$, one for each tuple in $S - \{w\}$.

If S has a countable cofinal subset $\{s_i\}$, the topology is no longer so simple, but $\mathbf{R}^S$ is still first countable, because the sets

$$U_i = \{f : (\forall w < s_i : f(w) = 0)\}$$

provide a countable neighbourhood basis of the origin.

The strangest properties of $\mathbf{R}^S$ occur when S has no countable cofinal set. Suppose that $f_i \to f$. We claim that it follows that $f_i = f$ eventually. To prove by contradiction, we assume without loss of generality (by thinning the sequence) that no f_i is equal to f. For each f_i, find the largest $w_i \in S$ such that for $s < w_i$, $f_i(s) = f(s)$ (since S is well ordered, the set of elements for which $f_i(s) \neq f(s)$ has a minimal element). Then the w_i form a countable cofinal set, because if $v \in S$ is arbitrary, the f_i eventually satisfy $f_i(s) = f(s)$ for $s < v$, hence the corresponding w_i is greater than v_i. Hence, if $f_i \to f$ in $\mathbf{R}^S$, where S does not have a countable cofinal subset, then eventually $f_i = f$. We conclude all countable sets in $\mathbf{R}^S$ are closed, and this proof easily generalises to show that if S does not have a cofinal set of cardinality $\mathfrak{a}$, then every set of cardinality $\leqslant \mathfrak{a}$ is closed.

The simple corollary to this proof is that compact subsets are finite. Let $X = f_1, f_2, \dots$ 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

$$f_1 \leqslant f_2 \leqslant \dots < h < \dots \leqslant g_2 \leqslant g_1$$

Suppose that $f_i \leqslant k < h < k' \leqslant g_j$. Then the intersection of the (f_i, g_i) contains an interval (k, k') around h, so that the intersection is open near h. The only other possiblity is that $f_i \to h$ or $g_i \to h$, which can only occur if $f_i = h$ or $g_i = h$ eventually, in which case we cannot have $f_i < h$, $h < g_i$. We conclude the intersection of countably many intervals is open, because we can always adjust any intersection to an intersection of this form without changing the resulting intersecting set (except if the set is empty, in which case the claim is trivial). The general case results from noting that any open set in an ordered group is a union of intervals.

14.5 Topological Groups arising from Normal subgroups

Let G be a group, and $\mathcal{N}$ a family of normal subgroups closed under intersection. If we interpret $\mathcal{N}$ as a neighbourhood base at the origin, the resulting topology gives G the structure of a totally disconnected topological group, which is Hausdorff if and only if $\bigcap \mathcal{N} = \{e\}$. First note that $g_i \to g$ if g_i is eventually in gN, for every $N \in \mathcal{N}$, which implies

$g_i^{-1} \in Ng^{-1} = g^{-1}N$, hence inversion is continuous. Furthermore, if h_i is eventually in hN, then $g_ih_i \in gNhN = ghN$, so multiplication is continuous. Finally note that $N^c = \bigcup_{g \neq e} gN$ is open, so that every open set is closed.

Example. *Consider $\mathcal{N} = \{\mathbf{Z}, 2\mathbf{Z}, 3\mathbf{Z}, \dots\}$. Then $\mathcal{N}$ induces a Hausdorff topology on $\mathbf{Z}$, such that $g_i \to g$, if and only if g_i is eventually in $g + n\mathbf{Z}$ for all n. In this topology, the series $1, 2, 3, \dots$ converges to zero!*

This example gives us a novel proof, due to Furstenburg, that there are infinitely many primes. Suppose that there were only finitely many, $\{p_1, p_2, \dots, p_n\}$. By the fundamental theorem of arithmetic,

$$\{-1, 1\} = (\mathbf{Z}p_1)^c \cap \cdots \cap (\mathbf{Z}p_n)^c$$

and is therefore an open set. But this is clearly not the case as open sets must contain infinite sequences.

Chapter 15

The Haar Measure

One of the reasons that we isolate locally compact groups to study is that they possess an incredibly useful object allowing us to understand functions on the group, and thus the group itself. A **left (right) Haar measure** for a group G is a Radon measure μ for which $\mu(xE) = \mu(E)$ for any $x \in G$ and measurable E ($\mu(Ex) = \mu(E)$ for all x and E). For commutative groups, all left Haar measures are right Haar measures, but in non-commutative groups this need not hold. However, if μ is a right Haar measure, then $\nu(E) = \mu(E^{-1})$ is a left Haar measure, so there is no loss of generality in focusing our study on left Haar measures.

Example. *The example of a Haar measure that everyone knows is the Lebesgue measure on $\mathbf{R}$ (or $\mathbf{R}^n$). It commutes with translations because it is the measure induced by the linear functional corresponding to Riemann integration on $C_c^+(\mathbf{R}^n)$. A similar theory of Darboux integration can be applied to linearly ordered groups, leading to the construction of a Haar measure on such a group.*

Example. *If G is a Lie group, consider a 2-tensor $g_e \in T_e^2(G)$ inducing an inner product at the origin. Then the diffeomorphism $f : a \mapsto b^{-1}a$ allows us to consider $g_b = f^*\lambda \in T_b^2(G)$, and this is easily verified to be an inner product, hence we have a Riemannian metric. The associated Riemannian volume element can be integrated, producing a Haar measure on G.*

Example. *If G and H have Haar measures μ and ν, then $G \times H$ has a Haar measure $\mu \times \nu$, so that the class of topological groups with Haar measures is closed under the product operation. We can even allow infinite products, provided that the groups involved are compact, and the Haar measures are normalized* *to probability measures. This gives us measures on F_2^ω and $\mathbf{T}^\omega$, which models the probability of an infinite sequence of coin flips.*

Example. *dx/x is a Haar measure for the multiplicative group of positive real numbers, since*

$$\int_a^b \frac{1}{x} = \log(b) - \log(a) = \log(cb) - \log(ca) = \int_{ca}^{cb} \frac{1}{x}$$

If we take the multiplicative group of all non-negative real numbers, the Haar measure becomes $dx/|x|$.

Example. *$dxdy/(x^2+y^2)$ is a Haar measure for the multiplicative group of complex numbers, since we have a basis of 'arcs' around the origin, and by a change of variables to polar coordinates, we verify the integral is changed by multiplication. Another way to obtain this measure is by noticing that $\mathbf{C}^\times$ is topologically isomorphic to the product of the circle group and the multiplicative group of real numbers, and hence the measure obtained should be the product of these measures. Since*

$$\frac{dxdy}{x^2+y^2} = \frac{drd\theta}{r}$$

We see that this is just the product of the Haar measure on $\mathbf{R}^+$, dr/r, and the Haar measure on $\mathbf{T}$, $d\theta$.

Example. *The space $M_n(\mathbf{R})$ of all n by n real matrices under addition has a Haar measure dM, which is essentially the Lebesgue measure on $\mathbf{R}^{n^2}$. If we consider the measure on $GL_n(\mathbf{R})$, defined by*

$$\frac{dM}{det(M)^n}$$

To see this, note the determinant of the map $M \mapsto NM$ on $M_n(\mathbf{R})$ is $det(N)^n$, because we can view $M_n(\mathbf{R})$ as the product of $\mathbf{R}^n$ n times, multiplication operates on the space componentwise, and the volume of the image of the unit paralelliped in each $\mathbf{R}^n$ is $det(N)$. Since the multiplicative group of complex numbers $z = x + iy$ can be identified with the group of matrices of the form

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix}$$

and the measure on $\mathbf{C} - \{0\}$ *then takes the form* $dM/det(M)$*. More generally, if* G *is an open subset of* $\mathbf{R}^n$*, and left multiplication acts affinely,* $xy = A(x)y + b(x)$*, then* $dx/|det(A(x))|$ *is a left Haar measure on* G*, where* dx *is Lebesgue measure.*

It turns out that there is a Haar measure on any locally compact group, and what's more, it is unique up to scaling. The construction of the measure involves constructing a positive linear functional $\phi : C_c(G) \to \mathbf{R}$ such that $\phi(L_x f) = \phi(f)$ for all x. The Riesz representation theorem then guarantees the existence of a Radon measure μ which represents this linear functional, and one then immediately verifies that this measure is a Haar measure.

Theorem 15.1. *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 infinum of $\sum c_i$, over all choices of $c_i > 0$ and $\int \phi \geqslant 1$, for which $f \leqslant \sum c_i L_{x_i} \phi$. Without the integral, we cannot measure the size of the functions ϕ, so we have to normalize by a different factor. We define $(f : \phi)$ to be the infinum of the sums $\sum c_i$, where $f \leqslant \sum c_i L_{x_i} \phi$ for some $x_i \in G$. We would then have

$$\int f d\mu \leqslant (f : \phi) \int \phi d\mu$$

If k is fixed with $\int k = 1$, then we would have

$$\int f d\mu \leqslant (f : \phi)(\phi : k)$$

We cannot change k if we wish to provide a limiting result in ϕ, so we notice that $(f:g)(g:h) \leqslant (f:h)$, which allows us to write

$$\int f \, d\mu \leqslant \frac{(f:\phi)}{(k:\phi)}$$

Taking the support of ϕ to be smaller and smaller, this value should approximate the integral perfectly accurately.

Define the linear functional

$$I_\phi(f) = \frac{(f:\phi)}{(k:\phi)}$$

Then I_ϕ is a sublinear, monotone, function with a functional bound

$$(k:f)^{-1} \leqslant I_\phi(f) \leqslant (f:k)$$

Which effectively says that, regardless of how badly we choose ϕ, the approximation factor $(f:\phi)$ is normalized by the approximation factor $(k:\phi)$ so that the integral is bounded. Now we need only prove that I_ϕ approximates a linear functional well enough that we can perform a limiting process to obtain a Haar integral. If $\varepsilon > 0$, and $g \in C_c^+(G)$ with $g = 1$ on $\text{supp}(f_1 + f_2)$, then the functions

$$h = f_1 + f_2 + \varepsilon g$$

$$h_1 = f_1/h \quad\quad h_2 = f_2/h$$

are in $C_0^+(G)$, if we define $h_i(x) = 0$ if $f_i(x) = 0$. This implies that there is a neighbourhood V of e such that if $x \in V$, and y is arbitrary, then

$$|h_1(xy) - h_1(y)| \leqslant \varepsilon \quad\quad |h_2(xy) - h_2(y)| < \varepsilon$$

If $\text{supp}(\phi) \subset V$, and $h \leqslant \sum c_i L_{x_i} \phi$, then

$$f_j(x) = h(x)h_j(x) \leqslant \sum c_i \phi(x_i x) h_j(x) \leqslant \sum c_i \phi(x_i x) \left[h_j(x_i^{-1}) + \varepsilon \right]$$

since we may assume that $x_i x \in \text{supp}(\phi) \subset V$. Then, because $h_1 + h_2 \leqslant 1$,

$$(f_1:\phi) + (f_2:\phi) \leqslant \sum c_j [h_1(x_j^{-1}) + \varepsilon] + \sum c_j [h_2(x_j^{-1}) + \varepsilon] \leqslant \sum c_j [1 + 2\varepsilon]$$

Now we find, by taking infinums, that

$$I_\phi(f_1) + I_\phi(f_2) \leqslant I_\phi(h)(1+2\varepsilon) \leqslant [I_\phi(f_1+f_2) + \varepsilon I_\phi(g)][1+2\varepsilon]$$

Since g is fixed, and we have a bound $I_\phi(g) \leqslant (g:k)$, we may always find a neighbourhood V (dependant on f_1, f_2) for any $\varepsilon > 0$ such that

$$I_\phi(f_1) + I_\phi(f_2) \leqslant I_\phi(f_1+f_2) + \varepsilon$$

if $\text{supp}(\phi) \subset V$.

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 $F : C_c^+(G) \to \mathbf{R}$ such that $(k:f)^{-1} \leqslant F(f) \leqslant (f:k)$. For each neighbourhood V of the identity, let $K(V)$ be the closure of the set of I_ϕ such that $\text{supp}(\phi) \subset V$. Then the set of all $K(V)$ has the finite intersection property, so we conclude there is some $I : C_c^+(G) \to \mathbf{R}$ 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

$$|I(f_1+f_2) - (I(f_1) - I(f_2))| \leqslant 2\varepsilon + |I_\phi(f_1+f_2) - (I_\phi(f_1) + I_\phi(f_2))| < 3\varepsilon$$

Taking $\varepsilon \to 0$, we conclude I is linear. Similar limiting arguments show that I is homogenous of degree 1, and commutes with all left translations. We conclude the extension of I to a linear functional on $C_0(G)$ is well defined, and the Radon measure obtained by the Riesz representation theorem is a Haar measure. □

We shall prove that the Haar measure is unique, but first we show an incredibly useful regularity property.

Proposition 15.2. *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 $x_1, \dots, x_n \in G$,

$$\mu\left(\bigcup_{i=1}^n x_i U\right) \leqslant \sum_{i=1}^n \mu(x_i U) = 0$$

If K is compact, then K can be covered by finitely many translates of U, so $\mu(K) = 0$. But then $\mu = 0$ by regularity, a contradiction. $\square$

Theorem 15.3. *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

$$A = \text{supp}(f)V \cup V\text{supp}(f) \qquad B = \text{supp}(g)V \cup V\text{supp}(g)$$

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,

$$\left|\int h d\mu \int f d\lambda - \int h d\lambda \int f d\mu\right| \leqslant \int h(y)|F_y(x)| d\mu(y) d\lambda(x)$$
$$\leqslant \varepsilon\lambda(A) \int h d\mu$$

In the same way, we find this is also true when f is swapped with g, and A with B. Dividing this inequalities by $\int h d\mu \int f d\mu$, we find

$$\left|\frac{\int f d\lambda}{\int f d\mu} - \frac{\int h d\lambda}{\int h d\mu}\right| \leqslant \frac{\varepsilon\lambda(A)}{\int f d\mu}$$

and this inequality holds with f swapped out with g, A with B. We then combine these inequalities to conclude

$$\left|\frac{\int f d\lambda}{\int f d\mu} - \frac{\int g d\lambda}{\int g d\mu}\right| \leqslant \varepsilon\left[\frac{\lambda(A)}{\int f d\mu} + \frac{\lambda(B)}{\int g d\mu}\right]$$

Taking ε to zero, we find $\lambda(A), \lambda(B)$ remain bounded, and hence

$$\frac{\int f d\lambda}{\int f d\mu} = \frac{\int g d\lambda}{\int g d\mu}$$

Thus there is a cosntant $c > 0$ such that $\int f d\lambda = c \int f d\mu$ for any function $f \in C_c^+(G)$, and we conclude that $\lambda = c\mu$. □

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 $\mu(A) = \nu(A)$.

15.1 Fubini, Radon Nikodym, and Duality

Before we continue, we briefly mention that integration theory is particularly nice over locally compact groups, even if we do not have σ finiteness. This essentially follows because the component of the identity in G is σ compact (take a compact neighbourhood and its iterated multiples), hence all components in G are σ compact. The three theorems that break down outside of the σ compact domain are Fubini's theorem, the Radon Nikodym theory, and the duality between $L^1(X)$ and $L^\infty(X)$. We show here that all three hold if X is a locally compact topological group.

First, suppose that $f \in L^1(G \times G)$. Then the essential support of f is contained within countably many components of $G \times G$ (which are simply products of components in G). Thus f is supported on a σ compact subset of $G \times G$ (as a locally compact topological group, each component of $G \times G$ is σ compact), and we may apply Fubini's theorem on the countably many components (the countable union of σ compact sets is σ compact). The functions in $L^p(G)$, for $1 \leqslant p < \infty$, also vanish outside of a σ compact subset (for if $f \in L^p(G)$, $|f|^p \in L^1(G)$ and thus vanishes outside of a σ compact set). What's more, all finite sums and products of functions from these sets (in either variable) vanish outside of σ compact subsets, so we almost never need to explicitly check the conditions for satisfying Fubini's theorem, and from now on we apply it wantonly.s

Now suppose μ and ν are both Radon measures, with $\nu \ll \mu$, and ν is σ-finite. By inner regularity, the support of ν is a σ compact set E. By inner regularity, μ restricted to E is σ finite, and so we may find a Radon

Nikodym derivative on E. This derivative can be extended to all of G because ν vanishes on G.

Finally, we note that $L^\infty(X) = L^1(X)^*$ can be made to hold if X is not σ finite, but locally compact and Hausdorff, provided we are integrating with respect to a Radon measure μ, and we modify $L^\infty(G)$ slightly. Call a set $E \subset X$ **locally Borel** if $E \cap F$ is Borel whenever F is Borel and $\mu(F) < \infty$. A locally Borel set is **locally null** if $\mu(E \cap F) = 0$ whenever $\mu(F) < \infty$ and F is Borel. We say a property holds **locally almost everywhere** if it is true except on a locally null set. $f : X \to \mathbf{C}$ is **locally measurable** if $f^{-1}(U)$ is locally Borel for every borel set $U \subset \mathbf{C}$. We now define $L^\infty(X)$ to be the space of all functions bounded except on a locally null set, modulo functions that are locally zero. That is, we define a norm

$$\|f\|_\infty = \inf\{c : |f(x)| \leqslant c \text{ locally almost everywhere}\}$$

and then $L^\infty(X)$ consists of the functions that have finite norm. It then follows that if $f \in L^\infty(X)$ and $g \in L^1(X)$, then g vanishes outside of a σ-finite set Y, so $fg \in L^1(X)$, and if we let $Y_1 \subset Y_2 \subset \dots \to Y$ be an increasing subsequence such that $\mu(Y_i) < \infty$, then $|f(x)| \leqslant \|f\|_\infty$ almost everywhere for $x \in Y_i$, and so by the monotone convergence theorem

$$\int |fg| d\mu = \lim_{Y_i \to \infty} \int_{Y_i} |fg| d\mu \leqslant \|f\|_\infty \int_{Y_i} |g| d\mu \leqslant \|f\|_\infty \|g\|_1$$

Thus the map $g \mapsto \int fg d\mu$ is a well defined, continuous linear functional with norm $\|f\|_\infty$. That $L^1(X)^* = L^\infty(X)$ follows from the decomposibility of the Carathéodory extension of μ, a fact we leave to the general measure theorists.

15.2 Unimodularity

We have thus defined a left invariant measure, but make sure to note that such a function is not right invariant. We call a group who's left Haar measure is also right invariant **unimodular**. Obviously all abelian groups are unimodular.

Given a fixed y, the measure $\mu_y(A) = \mu(Ay)$ is a new Haar measure on the space, hence there is a constant $\Delta(y) > 0$ depending only on y such that $\mu(Ay) = \Delta(y)\mu(A)$ for all measurable A. Since $\mu(Axy) = \Delta(y)\mu(Ay) =$

$\Delta(x)\Delta(y)\mu(A)$, we find that $\Delta(xy) = \Delta(x)\Delta(y)$, so Δ is a homomorphism from G to the multiplicative group of real numbers. For any $f \in L^1(\mu)$, we have

$$\int f(xy)d\mu(x) = \Delta(y^{-1})\int f(x)d\mu(x)$$

If $y_i \to e$, and $f \in C_c(G)$, then $\|R_{y_i}f - f\|_\infty \to 0$, so

$$\Delta(y_i^{-1})\int f(x)d\mu = \int f(xy_i)d\mu \to \int f(x)d\mu$$

Hence $\Delta(y_i^{-1}) \to 1$. This implies Δ, known as the unimodular function, is a continuous homomorphism from G to the real numbers. Note that Δ is trivial if and only if G is unimodular.

Theorem 15.4. *Any compact group is unimodular.*

Proof. $\Delta : G \to \mathbf{R}^*$ is a continuous homomorphism, hence $\Delta(G)$ is compact. But the only compact subgroup of $\mathbf{R}$ is trivial, hence Δ is trivial. □

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 15.5. *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. □

The modular function relates right multiplication to left multiplcation in the group. In particular, if $d\mu$ is a Left Haar measure, then $\Delta^{-1}d\mu$ is a right Haar measure. Hence any right Haar measure is a constant multiple of $\Delta^{-1}d\mu$. Hence the measure $\nu(A) = \mu(A^{-1})$ has a value c such that for any function f,

$$\int \frac{f(x)}{\Delta(x)}d\mu(x) = c\int f(x)d\nu(x) = c\int f(x^{-1})d\mu$$

If $c \neq 1$, pick a symmetric neighbourhood U such that for $x \in U$, $|\Delta(x) - 1| \leqslant \varepsilon|c-1|$. Then if $f > 0$

$$|c-1|\mu(U) = |c\mu(U^{-1}) - \mu(U)| = \left|\int_U [\Delta(x^{-1}) - 1]d\mu(x)\right| \leqslant \varepsilon\mu(U)|c-1|$$

A contradiction if $\varepsilon < 1$. Thus we have

$$\int f(x^{-1}) d\mu(x) = \int \frac{f(x)}{\Delta(x)} d\mu(x)$$

A useful integration trick. When Δ is unbounded, then it follows that $L^p(\mu)$ and $L^p(\nu)$ do not consist of the same functions. There are two ways of mapping the sets isomorphically onto one another – the map $f(x) \mapsto f(x^{-1})$, and the map $f(x) \mapsto \Delta(x)^{1/p} f(x)$.

From now on, we assume a left invariant Haar measure is fixed over an entire group. Since a Haar measure is uniquely determined up to a constant, this is no loss of generality, and we might as well denote our integration factors $d\mu(x)$ and $d\mu(y)$ as dx and dy, where it is assumed that this integration is over the Lebesgue measure.

15.3 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

$$\begin{aligned} \left| \int \phi(xy) d\mu(x) d\nu(y) \right| &\leqslant \int_G \int_G \phi(xy) d|\mu|(x) d|\nu|(y) \\ &\leqslant \|\mu\| \|\nu\| \|\phi\|_\infty \end{aligned}$$

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 $\|\mu * \nu\| \leqslant \|\mu\| \|\nu\|$, because of the bound we placed on the integrals above. $M(G)$ is therefore an involutive Banach algebra, which has a unit, the dirac delta measure at the identity.

As a remark, we note that involutive Banach algebras have nowhere as near a nice of a theory than that of C^* algebras. $M(G)$ cannot be renormed to be a C^* algebra, since every weakly convergent Cauchy sequence converges, which is impossible in a C^* algebra, except in the finite dimensional case.

A **discrete measure** on G is a measure in $M(G)$ which vanishes outside a countable set of points, and the set of all such measures is denoted $M_d(G)$. A **continuous measure** on G is a measure μ such that $\mu(\{x\}) = 0$ for all $x \in G$. We then have a decomposition $M(G) = M_d(G) \oplus M_c(G)$, for if μ is any measure, then $\mu(\{x\}) \neq 0$ for at most countably many points x, for

$$\|\mu\| \geqslant \sum_{x \in G} |\mu|(x)$$

This gives rise to a discrete measure ν, and $\mu - \nu$ is continuous. If we had another decomposition, $\mu = \psi + \phi$, then $\mu(\{x\}) = \psi(\{x\}) = \nu(\{x\})$, so $\psi = \nu$ by discreteness, and we then conclude $\phi = \mu - \nu$. $M_c(G)$ is actually a closed subspace of $M(G)$, since if $\mu_i \to \mu$, and $\mu_i \in M_c(G)$, and $\|\mu_i - \mu\| < \varepsilon$, then for any $x \in G$,

$$\varepsilon > \|\mu - \mu_i\| \geqslant |(\mu_i - \mu)(\{x\})| = |\mu(\{x\})|$$

Letting $\varepsilon \to 0$ shows continuity.

The convolution on $M(G)$ gives rise to a convolution on $L^1(G)$, where

$$(f * g)(x) = \int f(y)g(y^{-1}x)dy$$

which satisfies $\|f * g\|_1 \leqslant \|f\|_1 \|g\|_1$. This is induced by the identification of f with $f(x)dx$, because then

$$\begin{aligned}
\int \phi(f(x)dx * g(x)dx) &= \int\int \phi(yx)f(y)g(x)dydx \\
&= \int \phi(y)\left(\int f(y)g(y^{-1}x)dx\right)dy
\end{aligned}$$

Hence $fd\mu * gd\mu = (f * g)d\mu$. What's more,

$$\|f\|_1 = \|fd\mu\|$$

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 15.6. *$L^1(G)$ and $M_c(G)$ are closed ideals in $M(G)$, and $M_d(G)$ is a closed subalgebra.*

Proof. If $\mu_i \to \mu$, and each μ_i is discrete, the μ is discrete, because there is a countable set K such that all μ_i are equal to zero outside of K, so μ must also vanish outside of K (here we have used the fact that $M(G)$ is a Banach space, so that we need only consider sequences). Thus $M_d(G)$ is closed, and is easily verified to be subalgebra, essentially because $\delta_x * \delta_y = \delta_{xy}$. If $\mu_i \to \mu$, then $\mu_i(\{x\}) \to \mu(\{x\})$, so that $M_c(G)$ is closed in $M(G)$. If ν is an arbitrary measure, and μ is continuous, then

$$(\mu * \nu)(\{x\}) = \int_G \mu(\{y\}) d\nu(y^{-1}x) = 0$$

$$(\nu * \mu)(\{x\}) = \int_G \mu(\{y\}) d\nu(xy^{-1}) = 0$$

so $M_c(G)$ is an ideal. Finally, we verify $L^1(G)$ is closed, because it is complete, and if $\nu \in M(G)$ is arbitrary, and if U has null Haar measure, then

$$(f dx * \nu)(U) = \int \chi_U(xy) f(x) dx\, d\nu(y) = \int_G \int_{y^{-1}U} f(x) dx d\nu(y) = 0$$

$$(\nu * f dx)(U) = \int \chi_U(xy) d\nu(x) f(y) dy = \int_G \int_{Ux^{-1}} f(y) dy d\nu(x) = 0$$

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 $L_z(f*g) = (L_zf)*g$, and $R_z(f*g) = f*(R_zg)$, or from the fact that

$$(f*g)(zx) = \int f(y)g(y^{-1}zx)dy = \int f(zy)g(y^{-1}x)dy = [(L_zf)*g](x)$$

$$(f*g)(xz) = \int f(y)g(y^{-1}xz)dy = [f*(R_zg)](x)$$

Convolution can also be applied to the other L^p spaces, but we have to be a bit more careful with our integration.

Theorem 15.7. *If $f \in L^1(G)$ and $g \in L^p(G)$, then $f*g$ is defined for almost all x, $f*g \in L^p(G)$, and $\|f*g\|_p \leqslant \|f\|\|g\|_p$. If G is unimodular, then the same results hold for $g*f$, or if G is not unimodular and f has compact support.*

Proof. We use Minkowski's inequality to find

$$\begin{aligned}
\|f*g\|_p &= \left(\int\left|\int f(y)|g(y^{-1}x)dy\right|^p dx\right)^{1/p}\\
&\leqslant \int |f(y)|\left(\int |g(y^{-1}x)|^p dx\right)^{1/p} dy\\
&= \|f\|_1\|g\|_p
\end{aligned}$$

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

$$\begin{aligned}
\|g*f\|_p &= \left(\int\left|\int \frac{g(xy^{-1})f(y)}{\Delta(y)}dy\right|^p dx\right)^{1/p}\\
&\leqslant \int\left(\int\left|\frac{g(xy^{-1})f(y)}{\Delta(y)}\right|^p dx\right)^{1/p} dy\\
&= \|g\|_p \int_K \frac{|f(y)|}{\Delta(y)} d\mu(y)\\
&\leqslant M\|g\|_p\|f\|_1
\end{aligned}$$

which shows that $g * f$ is defined almost everywhere. $\square$

Theorem 15.8. *If G is unimodular, $f \in L^p(G)$, $g \in L^q(G)$, and $p = q^*$, then $f * g \in C_0(G)$ and $\|f * g\|_\infty \leqslant \|f\|_p \|g\|_q$.*

Proof. First, note that

$$\begin{aligned}
|(f * g)(x)| &\leqslant \int |f(y)||g(y^{-1}x)| dy \\
&\leqslant \|f\|_p \left(\int |g(y^{-1}x)|^q dy \right)^{1/q} \\
&= \|f\|_p \|g\|_q
\end{aligned}$$

For each x and y, applying Hölder's inequality, we find

$$\begin{aligned}
|(f * g)(x) - (f * g)(y)| &\leqslant \int |f(z)||g(z^{-1}x) - g(z^{-1}y)| dz \\
&\leqslant \|f\|_p \left(\int |g(z^{-1}x) - g(z^{-1}y)|^q dz \right)^{1/q} \\
&= \|f\|_p \left(\int |g(z) - g(zx^{-1}y)|^q dz \right)^{1/q} \\
&= \|f\|_p \|g - R_{x^{-1}y} g\|_q
\end{aligned}$$

Thus to prove continuity (and in fact uniform continuity), we need only prove that $\|g - R_x g\|_q \to 0$ for $q \neq \infty$ as $x \to \infty$ or $x \to 0$. This is the content of the next lemma. $\square$

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 15.9. *If $p \neq \infty$, then $\|g - R_x g\|_p \to 0$ and $\|g - L_x g\|_p \to 0$ as $x \to 0$.*

Proof. If $g \in C_c(G)$, then one verifies the theorem by using left and right uniform continuity. In general, we let $g_i \in C_c(G)$ be a sequence of functions converging to g in the L_p norm, and we then find

$$\|g - L_x g\|_p \leqslant \|g - g_i\|_p + \|g_i - L_x g_i\|_p + \|L_x(g_i - g)\|_p = 2\|g - g_i\|_p + \|g_i - L_x g_i\|_p$$

Taking i large enough, x small enough, we find $\|g - L_x g\|_p \to 0$. The only problem for right translation is the appearance of the modular function

$$\|R_x(g - g_i)\|_p = \frac{\|g - g_i\|_p}{\Delta(x)^{1/p}}$$

If we assume our x values range only over a compact neighbourhood K of the origin, we find that $\Delta(x)$ is bounded below, and hence $\|R_x(g-g_i)\|_p \to 0$, which effectively removes the problems in the proof. $\square$

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 15.10. *If $f \in L^1(G)$ and $g \in L^\infty(G)$, then $f * g$ is left uniformly continuous, and $g * f$ is right uniformly continuous.*

Proof. We have

$$\|L_z(f * g) - (f * g)\|_\infty = \|(L_z f - f) * g\|_\infty \leqslant \|L_z f - f\|_1 \|g\|_\infty$$

$$\|R_z(g * f) - (g * f)\|_\infty = \|g * (R_z f - f)\|_\infty \leqslant \|g\|_\infty \|R_z f - f\|_1$$

and both integrals converge to zero as $z \to 1$. $\square$

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 15.11. *For each neighbourhood U of the origin, pick a function $f_U \in (L^1)^+(G)$, with $\int \phi_U = 1$, supp$(f_U) \subset U$. Then if g is any function in $L^p(G)$,*

$$\|f_U * g - g\|_p \to 0$$

where we assume g is left uniformly continuous if $p = \infty$, and if f_U is viewed as a net with neighbourhoods ordered by inclusion. If in addition $f_U(x) = f_U(x^{-1})$, then $\|g * f_U - g\|_p \to 0$, where g is right uniformly continuous for $p = \infty$.

Proof. Let us first prove the theorem for $p \neq \infty$. If $g \in C_c(G)$ is supported on a compact K, and if U is small enough that $|g(y^{-1}x) - g(x)| < \varepsilon$ for $y \in U$, then because $\int_U f_U(y) = 1$, and by applying Minkowski's inequality, we find

$$\begin{aligned}
\|f_U * g - g\|_p &= \left(\int \left| \int f_U(y)[g(y^{-1}x) - g(x)] dy \right|^p dx \right)^{1/p} \\
&\leqslant \int f_U(y) \left(\int |g(y^{-1}x) - g(x)|^p dx \right)^{1/p} dy \\
&\leqslant 2\mu(K)\varepsilon \int f_U(y) dy \leqslant 2\mu(K)\varepsilon
\end{aligned}$$

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

$$|(f_U * g - g)(x)| = \left| \int f_U(y)[g(y^{-1}x) - g(x)] \right| \leqslant \varepsilon$$

For right convolution, we find that for $g \in C_c(G)$, where $|g(xy) - g(x)| < \varepsilon$ for $y \in U$, then

$$\begin{aligned}
\|g * f_U - g\|_p &= \left(\int \left| \int g(y) f_U(y^{-1}x) - g(x) dy \right|^p dx \right)^{1/p} \\
&= \left(\int \left| \int [g(xy) - g(x)] f_U(y) dy \right|^p dx \right)^{1/p} \\
&\leqslant \int \left(\int |g(xy) - g(x)|^p dx \right)^{1/p} f_U(y) dy \\
&\leqslant \mu(K)\varepsilon \int f_U(y)(1 + \Delta(y)) dy \\
&= \mu(K)\varepsilon + \mu(K)\varepsilon \int f_U(y)\Delta(y) dy
\end{aligned}$$

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

$$|(g * f_U - g)(x)| = \left| \int [g(xy) - g(x)] f_U(y) dy \right| \leqslant \varepsilon$$

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 15.12. *If V is a closed subspace of $L^1(G)$, then V is a left ideal if and only if it is closed under left translations, and a right ideal if and only if it is closed under right translations.*

Proof. If V is a closed left ideal, and f_U is an approximate identity at the origin, then for any g,

$$\|(L_z f_U) * g - L_z g\|_1 = \|L_z(f_U * g - g)\|_1 = \|f_U * g - g\| \to 0$$

so $L_z g \in V$. Conversely, if V is closed under left translations, $g \in L^1(G)$, and $f \in V$, then

$$g * f = \int g(y) L_{y^{-1}} f\, dy$$

which is in the closed linear space of the translates of f. Right translation is verified very similarily. $\square$

15.4 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 15.13. *For any $0 < \theta < 1$, and $0 < p, q \leqslant \infty$. If we define*

$$1/r_\theta = (1-\theta)/p + \theta/q$$

to be the inverse interpolation of the two numbers. Then

$$\|f\|_{r_\theta} \leqslant \|f\|_p^{1-\theta} \|f\|_q^\theta$$

Proof. We apply Hölder's inequality to find

$$\|f\|_{r_\theta} \leqslant \|f\|_{p/(1-\theta)} \|f\|_{q/\theta} = \left(\int |f|^{p/(1-\theta)} \right)^{(1-\theta)/p} \left(\int |f|^{q/\theta} \right)^{\theta/q}$$

so it suffices to prove $\|f\|_{p/(1-\theta)} \leqslant \|f\|_p^{1-\theta}$, $\|f\|_{q/\theta} \leqslant \|f\|_q^\theta$.

The map $x \mapsto x^p$ is concave for $0 < p < 1$, so we may apply Jensen's inequality in reverse to conclude

$$\left(\int |f|^{p/(1-\theta)} \right)^{(1-\theta)/p} \leqslant \left(\int |f|^p \right)^{1/p}$$

□

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.

15.5 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 15.14. *If a σ compact topological group G has a transitive topological action on X, and $x \in X$, then the continuous bijection from G/G_x to X is a homeomorphism.*

Proof. It suffices to show that the map $\phi : G \to X$ is open, and we need only verify this for the neighbourhood basis of compact neighbourhoods V of the origin by properties of the action. G is covered by countably many translates $y_1 V, y_2 V, \dots$, 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 $\phi(y V y_0^{-1}) \subset \phi(V V y_0^{-1})$, 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 if we switch to a new basepoint y, and c maps x to y, then $ax = x$ holds if and only if $cac^{-1}y = y$. The question here is to determine whether we have a G-invariant measure on X. This is certainly not always possible. If we had a measure on $\mathbf{R}$ invariant under the affine maps $ax + b$, then it would be equal to the Haar measure by uniqueness, but the Haar measure is not invariant under dilation $x \mapsto ax$.

Let G and H have left Haar measures μ and ν respectively, denote the projection of G onto G/H as $\pi : G \to G/H$, and let Δ_G and Δ_H be the respective modular functions. Define a map $P : C_c(G) \to C_c(G/H)$ by

$$(Pf)(Hx) = \int_H f(xy) d\nu(y) = \int_H$$

this is well defined by the invariance properties of ν. Pf is obviously continuous, and $\text{supp}(Pf) \subset \pi(\text{supp}(f))$. 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 15.15. *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. □

Lemma 15.16. *A compact $F \subset G/H$ gives rise to a function $f \geqslant 0$ in $C_c(G)$ such that $Pf = 1$ on E.*

Proof. Let E be a compact neighbourhood containing F, and if $\pi(K) = E$, there is a function $g \in C_c(G)$ with $g > 0$ on K, and $\phi \in C_c(G/H)$ is supported on E and $\phi(x) = 1$ for $x \in F$, let

$$f = \frac{\phi \circ \pi}{Pg \circ \pi} g$$

Hence

$$Pf = \frac{\phi}{Pg} Pg = \phi$$

□

Lemma 15.17. *If $\phi \in C_c(G/H)$, there is $f \in C_c(G)$ with $Pf = \phi$, and $\pi(suppf) = supp(\phi)$, and also $f \geqslant 0$ if $\phi \geqslant 0$.*

Proof. There exists $g \geqslant 0$ in $C_c(G/H)$ with $Pg = 1$ on $\text{supp}(\phi)$, and then $f = (\phi \circ \pi)g$ satisfies the properties of the theorem. □

We can now provide conditions on the existence of a measure on G/H.

Theorem 15.18. *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. □

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.

15.6 Function Spaces In Harmonic Analysis

There are a couple other function spaces that are interesting in Harmonic analysis. We define $\text{AP}(G)$ 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 $\text{WAP}(G)$ to be the space of weakly almost periodic functions (the translates are relatively compact in the weak topology). It is a deep fact that $\text{WAP}(G)$ contains $C_0(G)$, but $\text{AP}(G)$ can be quite small. The reason these function spaces are almost periodic is that in the real dimensional case, $\text{AP}(\mathbf{R})$ is just the closure of the set of all trigonometric polynomials.

Chapter 16

The Character Space

Let G be a locally compact group. A character on G is a *continuous* homomorphism from G to $\mathbf{T}$. The space of all characters of a group will be denoted $\Gamma(G)$.

Example. *Determining the characters of* $\mathbf{T}$ *involves much of classical Fourier analysis. Let* $f : \mathbf{T} \to \mathbf{T}$ *be an arbitrary continuous character. For each* $w \in \mathbf{T}$, *consider the function* $g(z) = f(zw) = f(z)f(w)$. *We know the Fourier series acts nicely under translation, telling us that*

$$\hat{g}(n) = w^n \hat{f}(n)$$

Conversely, since $g(z) = f(z)f(w)$,

$$\hat{g}(n) = f(w)\hat{f}(n)$$

Thus $(w^n - f(w))\hat{f}(n) = 0$ *for all* $w \in \mathbf{T}$, $n \in \mathbf{Z}$. *Fixing* n, *we either have* $f(w) = w^n$ *for all* w, *or* $\hat{f}(n) = 0$. *This implies that if* $f \neq 0$, *then* f *is just a power map for some* $n \in \mathbf{Z}$.

Example. *The characters of* $\mathbf{R}$ *are of the form* $t \mapsto e(t\xi)$, *for* $\xi \in \mathbf{R}$. *To see this, let* $e : \mathbf{R} \to \mathbf{T}$ *be an arbitrary character. Define*

$$F(x) = \int_0^x e(t)dt$$

Then $F'(x) = e(x)$. *Since* $e(0) = 1$, *for suitably small* δ *we have*

$$F(\delta) = \int_0^\delta e(t)dt = c > 0$$

and then it follows that

$$F(x+\delta)-F(x)=\int_x^{x+\delta} e(t)dt=\int_0^\delta e(x+t)dt=ce(x)$$

As a function of x, F is differentiable, and by the fundamental theorem of calculus,

$$\frac{dF(x+\delta)-F(x)}{dt}=F'(x+\delta)-F'(x)=e(x+\delta)-e(x)$$

This implies the right side of the above equation is differentiable, and so

$$ce'(x)=e(x+\delta)-e(x)=e(x)[e(\delta)-1]$$

Implying $e'(x)=Ae(x)$ for some $A\in\mathbf{C}$, so $e(x)=e^{Ax}$. We require that $e(x)\in\mathbf{T}$ for all x, so $A=\xi i$ for some $\xi\in\mathbf{R}$.

Example. *Consider the group $\mathbf{R}^+$ of positive real numbers under multiplication. The map $x\mapsto\log x$ is an isomorphism from $\mathbf{R}^+$ and $\mathbf{R}$, so that every character on $\mathbf{R}^+$ is of the form $e(s\log x)=x^{is}$, 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 16.1. *For any character $\phi: G\to\mathbf{C}$, the map*

$$\varphi(f)=\int\frac{f(x)}{\phi(x)}dx$$

is a non-zero character on the convolution algebra $L^1(G)$, and all characters arise this way.

Proof. The induced map is certainly linear, and

$$\begin{aligned}\varphi(f*g)&=\int\int\frac{f(y)g(y^{-1}x)}{\phi(x)}dydx\\&=\int\int\frac{f(y)g(x)}{\phi(y)\phi(x)}dydx\\&=\int\frac{f(y)}{\phi(y)}dy\int\frac{g(x)}{\phi(x)}dx\end{aligned}$$

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

$$\begin{aligned}\int\int f(y)g(x)\phi(yx)dydx &= \int\int f(y)g(y^{-1}x)\phi(x)dydx\\ &= \int f(x)\phi(x)dx \int g(y)\phi(y)dy\\ &= \int f(x)g(y)\phi(x)\phi(y)dxdy\end{aligned}$$

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 $\mathbf{T}$, 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 $\phi_i \to \phi$ if and only if

$$\int \frac{f(x)}{\phi_i(x)} dx \to \int \frac{f(x)}{\phi(x)} dx$$

for all functions $f \in L^1(G)$. This actually makes the map

$$(f, \phi) \mapsto \int \frac{f(x)}{\phi(x)} dx$$

a jointly continuous map, because as we verified in the proof above,

$$\widehat{f}(\phi)\phi(y) = \widehat{L_y f}(\phi)$$

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

$$\{\gamma : \|1 - \gamma(x)\| < \varepsilon \text{ for all } x \in K\} \qquad \{x : \|1 - \gamma(x)\| < \varepsilon \text{ for all } \gamma \in C\}$$

And these sets actually form a base for the topology on $\Gamma(G)$.

Theorem 16.2. *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

$$\frac{1}{\phi(y)}\widehat{1}(\phi) = \int \frac{dx}{\phi(yx)} = \int \frac{dx}{\phi(x)} = \hat{1}(\phi)$$

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. ☐

Given a function $f \in L^1(G)$, we may take the Gelfand transform, obtaining a function on $C_0(\Gamma(L^1(G)))$. The identification then gives us a function on $C_0(\Gamma(G))$, if we give $\Gamma(G)$ the topology induced by the correspondence (which also makes $\Gamma(G)$ into a topological group). The formula is

$$\widehat{f}(\phi) = \phi(f) = \int \frac{f(x)}{\phi(x)}$$

This gives us the classical correspondence between $L^1(\mathbf{T})$ and $C_0(\mathbf{Z})$, and $L^1(\mathbf{R})$ and $C_0(\mathbf{R})$, which is just the Fourier transform. Thus we see the Gelfand representation as a natural generalization of the Fourier transform. We shall also denote the Fourier transform by $\mathcal{F}$, especially when we try and understand it's properties as an operator. Gelfand's theory (and some basic computation) tells us instantly that

- $\widehat{f * g} = \widehat{f}\widehat{g}$ (The transform is a homomorphism).
- $\mathcal{F}$ is norm decreasing and therefore continuous: $\|\widehat{f}\|_\infty \leqslant \|f\|_1$.
- If G is unimodular, and $\gamma \in \Gamma(G)$, then $(f * \gamma)(x) = \gamma(x)\widehat{f}(\gamma)$.

Whenever we integrate a function with respect to the Haar measure, there is a natural generalization of the concept to the space of all measures on G. Thus, for $\mu \in M(G)$, we define

$$\widehat{\mu}(\phi) = \int \frac{dx}{\phi(x)}$$

which we call the **Fourier-Stieltjes transform** on G. It is essentially an extension of the Gelfand representation on $L^1(G)$ to $M(G)$. Each $\widehat{\mu}$ is a bounded, uniformly continuous function on $\Gamma(G)$, because the transform is still contracting, i.e.

$$\left| \int \frac{d\mu(x)}{\phi(x)} dx \right| \leqslant \|\mu\|$$

It is uniformly continuous, because

$$(L_\nu \widehat{\mu} - \widehat{\mu})(\phi) = \int \frac{1 - \nu(x)}{\nu(x)\phi(x)} d\mu(x)$$

The regularity of μ implies that there is a compact set K such that $|\mu|(K^c) < \varepsilon$. If $\nu_i \to 0$, then eventually we must have $|\nu_i(x) - 1| < \varepsilon$ for all $x \in K$, and then

$$|(L_\nu \widehat{\mu} - \widehat{\mu})(\phi)| \leqslant 2|\mu|(K^c) + \varepsilon\|\mu\| \leqslant \varepsilon(2 + \|\mu\|)$$

Which implies uniform continuity.

Let us consider why it is natural to generalize operators on $L^1(G)$ to $M(G)$. The first reason is due to the intuition of physicists; most of classical Fourier analysis emerged from physical considerations, and it is in this field that $L^1(G)$ is often confused with $M(G)$. Take, for instance, the determination of the electric charge at a point in space. To determine this experimentally, we take the ratio of the charge over some region in space to the volume of the region, and then we limit the size of the region to zero. This is the historical way to obtain the density of a measure with respect to the Lebesgue measure, so that the function we obtain can be integrated to find the charge over a region. However, it is more natural to avoid taking limits, and to just think of charge as an element of $M(\mathbf{R}^3)$. If we consider a finite number of discrete charges, then we obtain a discrete measure, whose density with respect to the Lebesgue measure does not exist. This doesn't prevent physicists from trying, so they think of the density obtained as shooting off to infinity at points. Essentially, we obtain the Dirac Delta function as a 'generalized function'. This is fine for intuition, but things seem to get less intuitive when we consider the charge on a subsurface of $\mathbf{R}^3$, where the 'density' is 'dirac'-esque near the function, where as measure theoretically we just obtain a density with respect to the two-dimensional Hausdorff measure on the surface. Thus, when physicists discuss quantities as functions, they are really thinking of measures, and trying to take densities, where really they may not exist.

There is a more austere explanation, which results from the fact that, with respect to integration, $L^1(G)$ is essentially equivalent to $M(G)$. Notice that if $\mu_i \to \mu$ in the weak-$*$ topology, then $\widehat{\mu_i} \to \widehat{\mu}$ pointwise, because

$$\int \frac{d\mu_i(x)}{\phi(x)} \to \int \frac{d\mu(x)}{\phi(x)}$$

(This makes sense, because weak-$*$ convergence is essentially pointwise convergence in $M(G)$). Thus the Fourier-Stietjes transform is continuous with respect to these topologies. It is the unique continuous extension of the Fourier transform, because

Theorem 16.3. *$L^1(G)$ is weak-$*$ dense in $M(G)$.*

Proof. First, note that the Dirac delta function can be weak-$*$ approximated by elements of $L^1(G)$, since we have an approximate identity in the space.

First, note that if $\mu_i \to \mu$, then $\mu_i * \nu \to \mu * \nu$, because

$$\int f d(\mu_i * \nu) = \int \int f(xy) d\mu_i(x) d\nu(y)$$

The functions $y \mapsto \int f(xy) d\mu_i(x)$ converge pointwise to $\int f(xy) d\mu(y)$. Since

$$\left| \int f(xy) d\mu_i(x) \right| \leqslant \|f\|_1 \|\mu_i\|$$

If i is taken large enough that $\square$

If $\phi_\alpha \to \phi$, in the sense that $\phi_\alpha(x) \to \phi(x)$ for all $x \in G$, then, because $\|\phi_\alpha(x)\| = 1$ for all x, we can apply the dominated convergence theorem on any compact subset K of G to conclude

$$\int_K \frac{d\mu(x)}{\phi_\alpha(x)} \to \int_K \frac{d\mu(x)}{\phi(x)}$$

It is immediately verified to be a map into $L^1(\Gamma(G))$, because

$$\int \left| \int \frac{d\mu(x)}{\phi(x)} \right| d\phi \leqslant \int \int \|\mu\|$$

The formula above immediately suggests a generalization to a transform on $M(G)$. For $\nu \in M(G)$, we define

$$\mathcal{F}(\nu)(\phi) = \int \frac{d\nu}{\phi}$$

If $\mathcal{G} : L^1(G) \to C_0(\Gamma(G))$ is the Gelfand transform, then the transform induces a map $\mathcal{G}^* : M(\Gamma(G)) \to L^\infty(G)$.

The duality in class-ical Fourier analysis is shown through the inversion formulas. That is, we have inversion functions

$$\mathcal{F}^{-1}(\{a_k\}) = \sum a_k e_k(t) \qquad \mathcal{F}^{-1}(f)(x) = \int f(t) e(xt)$$

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 unimdoular, then it is closed under conjugation. If we let $g(x) = \overline{f(x^{-1})}$, we find

$$\mathcal{F}(g)(\phi) = \int \frac{g(x)}{\phi(x)} dx = \overline{\int \frac{f(x^{-1})}{\phi(x^{-1})} dx} = \int \frac{f(x)}{\phi(x)} dx = \overline{\mathcal{F}(f)(\phi)}$$

so that by the Stone Weirstrass theorem $A(G)$ is dense in $C_0(\Gamma(L^1(G)))$.

Chapter 17

Banach Algebra Techniques

In the mid 20th century, it was realized that much of the analytic information about a topological group can be captured in various C^* algebras related to the group. For instance, consider the Gelfand space of $L^1(\mathbf{Z})$ is $\mathbf{T}$, which represents the fact that one can represent functions over $\mathbf{T}$ as sequences of numbers. Similarily, we find the characters of $L^1(\mathbf{R})$ are the maps $f \mapsto \widehat{f}(x)$, so that the Gelfand space of $\mathbf{R}$ is $\mathbf{R}$, and the Gelfand transform is the Fourier transform on this space. For a general G, we may hope to find a generalized Fourier transform by understanding the Gelfand transform on $L^1(G)$. We can also generalize results by extending our understanding to the class $M(G)$ of regular, Borel measures on G.

Chapter 18

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 $\mathbf{K} \times \mathbf{K} \to \mathbf{C}$, and therefore we can define $x^* \in \mathbf{K}^*$ by $x^*(y) = \langle x, y \rangle$. If $x^*(y) = xy = 0$ for all y, then in particular $x^*(1) = x$, so $x = 0$. This means that the homomorphism $\mathbf{K} \to \mathbf{K}^*$ is injective.

Chapter 19

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 $\widehat{T(f)} = \widehat{f}\widehat{K}$. The function $\widehat{K}$ is known as a **Fourier multiplier**, because it operates by multiplication on the frequencies of the function f. We say $\widehat{K}$ is a **Fourier multiplier on** $L^p(\mathbf{R}^n)$ if T is a bounded map from $S(\mathbf{R}^n)$ to $L^p(\mathbf{R}^n)$, under the L^p norms. Such maps clearly extend uniquely to maps from $L^p(\mathbf{R}^n)$ to $L^p(\mathbf{R}^n)$, and so we can think of T as operating by convolution on the space of L^p functions. We will denote the space of all Fourier multipliers on L^p by M_p. We define the L^p norm on these distributions K, denoted $\|K\|_p$, to be the operator norm of the associated operator T.

Example. *Consider the space M_∞. If K is a distribution in M_∞, then $\|K\|_\infty < \infty$, and since convolution commutes with translations, in the sense that $f_h * K = (f * K)_h$, then*

$$\|K\|_\infty = \sup_{f \in L^\infty(\mathbf{R}^n)} \frac{|(f * K)(0)|}{\|f\|_\infty}$$

But then the map $f \mapsto (f * K)(0)$ is a bounded operator on the space of bounded continuous functions, and so the Riesz representation says there is a bounded Radon measure μ such that

$$(f * K)(0) = \int f(-y)\, d\mu(y)$$

But now we know

$$(f * K)(x) = (f_{-x} * K)(0) = \int f(x-y) d\mu(y) = (f * \mu)(x)$$

Thus M_∞ is really just the space of all bounded Radon measures, and

$$\|K\|_\infty = \sup_{f \in L^\infty(\mathbf{R}^n)} \frac{\left|\int f(y)\, d\mu(y)\right|}{\|f\|_\infty} = \|\mu\|_1$$

so M_∞ even has the same norm as the space of all bounded Radon measures. Note that it becomes a Banach algebra under convolution of distributions, since the convolution of two bounded Radon measures is a bounded Radon measure.

Theorem 19.1. *For any $1 \leqslant p \leqslant \infty$, and $q = p^*$, then $M_p = M_q$.*

Proof. Let $f \in L^p$, and $g \in L^q$, then Hölder's inequality gives

$$|(K * f * g)(0)| \leqslant \|K * f\|_p \|g\|_q \leqslant \|K\|_p \|f\|_p \|g\|_p$$

Thus $K * g \in L_q$, and that $K \in M_q$ with $\|K\|_q \leqslant \|K\|_p$. By symmetry, we find $\|K\|_p = \|K\|_q$. ☐

Example. *Consider M_2. If K is a distribution with $\|f * K\|_2 \leqslant A\|f\|_2$, then Parsevel's inequality implies that*

$$\|\widehat{f}\widehat{K}\|_2 = \|f * K\|_2 \leqslant A\|f\|_2 = A\|\widehat{f}\|_2$$

so for each $\widehat{f}$, TODO: PROVE THAT THIS IS REALLY JUST THE SPACE $L^\infty(\mathbf{R}^n)$, with the supremum norm. Note that this is also a Banach algebra under pointwise multiplication.

Using the Riesz-Thorin interpolation theorem, we find that if $1/p = (1-\theta)/p_0 + \theta/p_1$, then $\|K\|_p \leqslant \|K\|_{p_0}^{1-\theta} \|K\|_{p_1}^{\theta}$, when K lies in the three spaces. In particular, $\|K\|_p$ is a decreasing function of p for $1 \leqslant p \leqslant 2$, so we find $M_1 \subset M_p \subset M_q \subset M_2$ for $1 \leqslant p < q \leqslant 2$. In particular, all Fourier multipliers can be viewed as Fourier multipliers with respect to bounded, measurable functions on L^∞. Riesz interpolation shows that each M_p is a Banach algebra under multiplication in the frequency domain, or convolution in the spatial domain.

Theorem 19.2. *Let $T : \mathbf{R}^n \to \mathbf{R}^m$ be a surjective affine transformation. Then the endomorphism T^* on $M_p(\mathbf{R}^n)$ defined by $(T^* f)(\xi) = f(T(\xi))$ is an isometry, and if T is a bijection, so too is T^*.*

Proof. TODO ☐

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 19.3. *There exists a Schwartz function φ on $\mathbf{R}^n$ which is supported on the annulus*

$$\{\xi : 1/2 \leqslant |\xi| \leqslant 2\}$$

is positive for $1/2 < |\xi| < 2$, and satisfies

$$\sum_{k=-\infty}^{\infty} \varphi(2^{-k}\xi) = 1$$

for all $\xi \neq 0$.

Lemma 19.4 (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*

$$\sigma < \frac{1}{|I_n|} \int_{I_n} |f(x)|\, dx \leqslant 2^n \sigma$$

and with $|f(x)| \leqslant \sigma$ for almost all x outside these cubes.

Theorem 19.5 (The Mihlin Multiplier Theorem)**.** *Let m be a bounded function on $\mathbf{R}^n$ which is smooth except possibly at the origin, such that*

$$\sup_{\substack{\xi \in \mathbf{R}^n \\ |\alpha| \leqslant L}} |\xi|^{|\alpha|} |(D^\alpha m)(x)| < \infty$$

Then m is an L^p Fourier multiplier for $1 < p < \infty$.

19.1 Besov Spaces

Recall the Schwarz function φ used to prove the Mihlin multiplier theorem. We now define functions φ_k such that

$$\widehat{\varphi_n}(\xi) = \varphi(2^{-n}\xi) \qquad \widehat{\psi}(\xi) = 1 - \sum_{n=1}^{\infty} \varphi(2^{-n}\xi)$$

Thus φ_n essentially covers the annulus $2^{n-1} \leqslant |\xi| \leqslant 2^{n+1}$, and the function ψ covers the remaining low frequency parts covered in the frequency ball of radius 2. We have

$$\varphi_n(\xi) = \widetilde{\varphi_{2^{-n}}}(\xi) = 2^{dn}\check{\varphi}(2^n\xi)$$

Given $s \in \mathbf{R}$, and $1 \leqslant p, q \leqslant \infty$, we write

$$\|f\|^s_{pq} = \|\psi * f\|_p + \left(\sum_{n=1}^{\infty} (2^{sn}\|\varphi_k * f\|_p)^q \right)^{1/q}$$

The convolution $\varphi_n * f$ essentially captures the portion of f whose frequencies lie in the annulus $2^{n-1} \leqslant |\xi| \leqslant 2^{n+1}$

19.2 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 20

The Cap Set Problem

The cap set problem comes out of additive combinatorics, whose goal is to understand additive structure in some abelian group, typically the integers. For instance, we can think of a set A as being roughly closed under addition if $|A + A| = O(|A|)$. Over rings, we can study the interplay between additive and multiplicative structure. For instance, one conjecture of Erdös and Szemerédi says that if A is a finite subset of real numbers, then $\max(|A + A|, |A \cdot A|) \gtrsim |A|^{1+c}$ for some positive $c \in (0, 1)$. The best known c so far is $c \sim 1/3$, though it is conjectured that we can take c arbitrarily close to 1. This can be seen as a discrete version of the results of Bourgain and Edgar-Miller on the Hausdorff dimensions of Borel subrings.

Theorem 20.1 (Van Der Waerden - 1927)**.** *For any positive integes r and k, there is N such that if the integers in $[1, N]$ are given an r coloring, then there is a monochromatic k term arithmetic progression.*

The coloring itself is not so important, more just the partitioning. We just pidgeonhole, using the density of k term arithmetic progressions. This problem suggests the Ramsey type problem of determining the largest set A of the integers $[1, N]$ which does not contain k term arithmetic progressions. Behrend's theorem says we can choose A to be on the order of $N \exp(-c\sqrt{\log N})$.

Theorem 20.2 (Roth - 1956)**.** *If A is a set of integers in $[1, N]$ which is free of three term arithmetic progressions, then $|A| = O(N / \log \log N)$.*

Szemerédi proved that if A is free of k term arithmetic progressions, $|A| = o(N)$. If Erdös Turan, if $\sum_{x \in X} 1/x$ diverges, then X contains arbitrarily long arithmetic progressions. For now, we'll restrict our attention to three term arithmetic progressions. Heath and Brown showed that three term arithmetic progresisons are $O(N/(\log N)^c)$ for some constant c. In 2016, the best known bound was given by Bloom, given $O(N(\log\log N)^4/\log N)$.

One way we can simplify our problem is to note that avoiding three term arithmetic progressions is a local issue, so we can embed $[1, N]$ in $\mathbf{Z}/M\mathbf{Z}$ for suitably large M, and we lose none of the problems we had over the integers. A heuristic is that it is easier to solve these kind of problems in $\mathbf{F}_p^n$, where p is small and n is large, which should behave like $\{1, \dots, p^n\}$. This leads naturally to the cap set problem.

Theorem 20.3 (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 20.4 (Meschulam - 1995)**.** *Let $A \subset \mathbf{F}_3^n$ be a cap set. Then $|A| = O(3^n/n)$. This is analogous to a $N/\log N$ case over the integers, giving evidence that the finite field case is easier.*

In 2012, Bateman and Katz showed $|A| = O(3^n/n^{1+\varepsilon})$ for some $c > 0$. This was a difficult proof. In 2016, there was a more significant breakthrough, which gave an easy proof using the polynomial method of an exponentially small bound of c^n, where $c < 4$, over $\mathbf{Z}/4\mathbf{Z}$, and a week later Ellenberg-Gijswijt used this argument in the $\mathbf{F}_3$ case to prove that if A is a capset in $\mathbf{F}_3$,then $|A| = O(c^n)$, for $c = 2.7551\dots$.

The idea of the polynomial method is to take combinatorial information about some set, encode it as some algebraic structura information, and then apply the theory of polynomials to encode this algebraic information and use it to limit and enable certain properties to occur.

If V is the space of polynomials of degree d vanishing on a set A, then we know $\dim V \geqslant \dim \mathcal{P}_d - |A|$. This gives a lower bound on the size of A, whereas we want a lower bound. To get an upper bound, we take $|A|^c$ instead, which shows

$$\dim V \geqslant \dim \mathcal{P}_d + |A| - 3^n$$

whichs gives $|A| \leqslant 3^n + \dim V - \dim \mathcal{P}_d$. Now using linear algebra, we can find a polynomial P vanishing on A^c with support of cardinality greater than or equal to $\dim V$, hence

$$|A| \leqslant 3^n - \dim \mathcal{P}_d + \max |\mathrm{supp}(P)|$$

It follows that A is a cap set if and only if $x + y = 2z$, or $x + y + z = 0$ holds if and only if $x = y = z$. This is an algebraic property which says directly that A has no nontrivial three term arithmetic progressions. Thus for any $a_1, \dots, a_m \in A$, $P(-a_i - a_j) = 0$ 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 $x, y \in \mathbf{F}_3^n$,

Part IV

Restriction and Decoupling

Decoupling Theory is an in depth study of how 'interference patterns' can show up when combined waves with frequency supports in disjoint regions of space. The geometry of these regions effects how much constructive interference can happen. Of course decoupling theory is essential to studying many dispersive partial differential equations, but also has surprising applications in number theory as well, as well as other areas of harmonic analysis, such as restriction theory.

Chapter 21

The General Framework

In any norm space X, given $x_1,\dots,x_N \in X$, one can apply the Cauchy-Schwartz inequality to obtain the estimate

$$\|x_1 + \dots + x_N\|_X \leqslant \|x_1\|_X + \dots + \|x_N\|_X \leqslant N^{1/2} \left(\|x_1\|_X^2 + \dots + \|x_N\|_X^2 \right)^{1/2}.$$

Such a result is often sharp for general $x_1,\dots,x_N$. For instance, when $X = L^1(\mathbf{R}^d)$, and the $x_1,\dots,x_N$ are functions with disjoint supports, but with equal L^1 norm. However, if the $x_1,\dots,x_n$ are 'uncorrelated', then one can often expect this result to be substantially improved. For instance, if X is a Hilbert space, and if $x_1,\dots,x_N$ are pairwise orthogonal, Bessel's inequality allows us to conclude that

$$\|x_1 + \dots + x_N\|_X \leqslant \left(\|x_1\|_X^2 + \dots + \|x_N\|_X^2 \right)^{1/2}.$$

Thus we obtain a significant 'square root cancellation' in N. For instance, in $L^2(\mathbf{R}^d)$, this occurs if $x_1,\dots,x_N$ have disjoint supports, or more interestingly, if their Fourier transforms have disjoint supports.

We are interested in determining what causes 'square root cancellation' in general norm spaces. The theory of *almost orthogonality* studies this phenomena in Hilbert spaces, but we are interested in this phenomenon in other norm spaces. Informally, we say $x_1,\dots,x_N$ satisfies a *decoupling inequality* in a norm space X if for all $\varepsilon > 0$, we have

$$\|x_1 + \dots + x_N\|_X \lesssim_\varepsilon N^\varepsilon \left(\|x_1\|_X^2 + \dots + \|x_N\|_X^2 \right)^{1/2}.$$

Thus decoupling theory is the study of when correlation occurs in various norm spaces. Of particular importance in harmonic analysis will be to determine what properties of the Fourier transform of a function enable us to obtain decoupling phenomena.

Remark. We are interested in studying decoupling in $L^p(\Omega)$. However, the fact that we are obtaining estimates on the l^2 sum implies that we can only obtain such results when $p \geqslant 2$. To see why, note that if $p < 2$ and $f_1, \dots, f_N \in L^p(\Omega)$ have no interference, i.e. they have disjoint support, then

$$\|f_1 + \cdots + f_N\|_{L^p(\Omega)} = \left(\|f_1\|_{L^p(\Omega)}^p + \cdots + \|f_N\|_{L^p(\Omega)}^p\right)^{1/p},$$

This l^p sum can exceed the l^2 sum by a factor of $N^{1/p-1/2}$.

There are certain cases where we can obtain decoupling in $L^p(\Omega)$ for $p > 2$. For instance, we say $f_1, \dots, f_N \in L^4(\Omega)$ are *biorthogonal* if $\{f_i f_j : i < j\}$ forms an orthogonal family in $L^2(\Omega)$.

Theorem 21.1. *If $f_1, \dots, f_N$ are biorthogonal, then*

$$\|f_1 + \cdots + f_N\|_{L^4(\Omega)} \lesssim \left(\|f_1\|_{L^4(\Omega)}^2 + \cdots + \|f_N\|_{L^4(\Omega)}^2\right)^{1/2}.$$

Proof. First, we rearrange

$$\begin{aligned}
\|f_1 + \cdots + f_N\|_{L^4(\Omega)}^2 &= \|(f_1 + \cdots + f_N)^2\|_{L^2(\Omega)} \\
&= \left\| \sum_{1 \leqslant i,j \leqslant N} f_i f_j \right\|_{L^2(\Omega)} \lesssim \sum_{i=1}^N \|f_i^2\|_{L^2(\Omega)} + \left\| \sum_{1 \leqslant i < j \leqslant N} f_i f_j \right\|_{L^2(\Omega)}
\end{aligned}$$

Applying Bessel's inequality, we conclude that

$$\begin{aligned}
\left\| \sum_{1 \leqslant i < j \leqslant N} f_i f_j \right\|_{L^2(\Omega)} &= \left(\sum_{1 \leqslant i < j \leqslant N} \|f_i f_j\|_{L^2(\Omega)}^2 \right)^{1/2} \\
&= \left\| \sum_{i=1}^N |f_i|^2 \right\|_{L^2(\Omega)} \lesssim \sum_{i=1}^N \|f_i^2\|_{L^2(\Omega)}.
\end{aligned}$$

Combining these calculations, noticing that $\|f_i^2\|_{L^2(\Omega)} = \|f_i\|_{L^4(\Omega)}^2$, and taking square roots completes the claim. $\square$

Remark. If $\{x_1, \dots, x_N\}$ are elements of a Hilbert space X, and each x_i is orthogonal to all but at most $M \geqslant 1$ vectors x_j, then one can establish an 'almost Bessel inequality'

$$\|x_1 + \cdots + x_N\|_X^2 \lesssim M \left(\|x_1\|_X^2 + \cdots + \|x_N\|_X^2\right).$$

The idea is to reduce to rearrange the vectors such that $\|x_1\|_X \geqslant \cdots \geqslant \|x_N\|_X$, upper bound $\|x_1 + \cdots + x_N\|_X^2$ by $\sum_{i \leqslant j} (x_i, x_j)$, and then apply Cauchy-Schwartz. In particular, this implies that if each element of $\{f_i f_j : i < j\}$ is orthogonal to all but at most $O_\varepsilon(N^\varepsilon)$ elements of the family, then we still have a decoupling inequality.

Remark. Similarily, if $f_1, \dots, f_N \in L^6(\Omega)$ are chosen to be *triorthogonal*, in the sense that $\{f_i f_j f_k\}$ are mostly orthogonal to one another, one can obtain a decoupling inequality in the L^6 norm.

We will be most interested in studying families of functions with disjoint Fourier supports in $L^p(\mathbf{R}^d)$, where $p \geqslant 2$. Just because functions have disjoint Fourier supports does not mean that decoupling automatically happens however; constructive interference can still occur. In general, the best result we can obtain in the L^p norm for $p > 2$ involves a polynomial dependence on N, and we require additional geometric features like that in the corollary to guarantee a genuine decoupling inequality.

Theorem 21.2. *If $f_1, \dots, f_N$ are Schwartz functions on $\mathbf{R}^d$ with disjoint Fourier support, and $2 \leqslant p \leqslant \infty$, then*

$$\|f_1 + \cdots + f_N\|_{L^p(\mathbf{R}^d)} \leqslant N^{1/2 - 1/p} \left(\|f_1\|_{L^p(\mathbf{R}^d)}^2 + \cdots + \|f_N\|_{L^p(\mathbf{R}^d)}^2\right)^{1/2}.$$

Proof. If $f_1, \dots, f_N$ have disjoint Fourier support, then by orthogonality, we have

$$\|f_1 + \cdots + f_N\|_{L^2(\mathbf{R}^d)} \leqslant \left(\|f_1\|_{L^2(\mathbf{R}^d)}^2 + \cdots + \|f_N\|_{L^2(\mathbf{R}^d)}^2\right)^{1/2}.$$

We also have the trivial inequality

$$\begin{aligned}\|f_1 + \cdots + f_N\|_{L^\infty(\mathbf{R}^d)} &\leqslant \|f_1\|_{L^\infty(\mathbf{R}^d)} + \cdots + \|f_N\|_{L^\infty(\mathbf{R}^d)} \\ &\leqslant N^{1/2} \left(\|f_1\|_{L^\infty(\mathbf{R}^d)}^2 + \cdots + \|f_N\|_{L^\infty(\mathbf{R}^d)}^2\right)^{1/2}.\end{aligned}$$

Interpolation then gives the result. □

In general, this result is optimal.

Example. *Let u be a Schwartz function on $\mathbf{R}$ with $u(0) = 1$, and with Fourier support in $[0,1]$. For each $k \in \{1,\dots,N\}$, define $f_k = e^{4\pi kix}u$. Then f_k has Fourier support in $[2k, 2k+1]$. If $|x| \lesssim 1/N$, we have $|f_k(x)-1| \leqslant c < 1$ for each k, where c is independant of N. But this means that the values $f_1(x),\dots,f_N(x)$ have positive real part bounded below by a universal constant, and so if $|x| \lesssim 1/N$, we find $|f_1(x) + \dots + f_N(x)| \gtrsim N$. Thus*

$$\|f_1 + \dots + f_N\|_{L^p(\mathbf{R})} \gtrsim N^{1-1/p}.$$

On the other hand, we have

$$\left(\|f_1\|_{L^p(\mathbf{R})}^2 + \dots + \|f_N\|_{L^p(\mathbf{R})}^2\right)^{1/2} \lesssim N^{1/2},$$

where the implicit constant here depends only on the L^p norm of u. Thus

$$\|f_1 + \dots + f_N\|_{L^p(\mathbf{R})} \gtrsim N^{1/2-1/p} \left(\|f_1\|_{L^p(\mathbf{R})}^2 + \dots + \|f_N\|_{L^p(\mathbf{R})}^2\right)^{1/2},$$

which shows our result is tight up to constants.

To restate our desire, we are interested in knowing, for a given family $\mathcal{S}$ of disjoint sets in $\mathbf{R}^d$, whether it is true that if $f_1,\dots,f_N$ have Fourier support on distinct regions $S_1,\dots,S_N \in \mathcal{S}$, we have

$$\|f_1 + \dots + f_N\|_{L^p(\mathbf{R}^d)} \lesssim_\varepsilon N^\varepsilon \left(\|f_1\|_{L^p(\mathbf{R}^d)}^2 + \dots + \|f_N\|_{L^p(\mathbf{R}^d)}^2\right)^{1/2}.$$

Such a result depends significantly on the geometric structure of the regions in $\mathcal{S}$. The techniques we will use (e.g. induction on scales) imply the need for the 'ε loss' given by the N^ε factor. Below is a positive result for a particular family $\mathcal{S}$, easily proved using the biorthogonality arguments established above.

Theorem 21.3. *If $\mathcal{S}$ is a family of sets in $\mathbf{R}^d$ such that for $S_1, S_2, S_3, S_4 \in \mathcal{S}$, then $S_1 + S_2$ is disjoint from $S_3 + S_4$ except in trivial circumstances. Then if distinct sets $S_1,\dots,S_N \in \mathcal{S}$ are selected from $\mathcal{S}$, and $f_1,\dots,f_N$ are a family of Schwartz functions in $\mathbf{R}^d$ such that f_i has Fourier support in S_i for each i, then*

$$\|f_1 + \dots + f_N\|_{L^4(\Omega)} \lesssim \left(\|f_1\|_{L^4(\Omega)}^2 + \dots + \|f_N\|_{L^4(\Omega)}^2\right)^{1/2}.$$

Remark. We say a set of integers $A \subset \{0, \dots N-1\}$ is a *Sidon set* if there does not exist a nontrivial solution to the equation $a_1 + a_2 = a_3 + a_4$. If A is Sidon, then $\mathcal{S} = \{[2k, 2k+1] : k \in A\}$ satisfies the constraints of the corollary, and so we can obtain a decoupling result that if $\{f_k : k \in A\}$ are a family of Schwartz functions such that f_k has Fourier support in $[2k, 2k+1]$, then

$$\| \sum_{k \in A} f_k \|_{L^4(\mathbf{R})} \lesssim \left(\sum_{k \in A} \|f_k\|_{L_4(\mathbf{R})}^2 \right)^{1/2}.$$

On the other hand, a variant of the example above shows that for any Sidon set A, there is a family of functions $\{f_k : k \in A\}$ with f_k having Fourier support on $[2k, 2k+1]$, and with

$$\left\| \sum_{k \in A} f_k \right\|_{L^4(\mathbf{R})} \gtrsim \frac{\#(A)^{1/2}}{N^{1/4}} \left(\sum_{k \in A} \|f_k\|_{L^4(\mathbf{R})}^2 \right)^{1/2}.$$

Combining this inequality with the decoupling inequality, we obtain the surprising number theoretic result that any Sidon set A must satisfy $\#(A) \lesssim N^{1/2}$. We can extend this result to show that any set $A \subset \{0, \dots, N-1\}$ having no nontrivial solutions to the equation $a_1 + \dots + a_m = a_1' + \dots + a_m'$ should satisfy $\#(A) \lesssim N^{1/m}$.

Another example is obtained using Littlewood-Paley theory.

Theorem 21.4. *Let $\mathcal{S}$ be the collection of all boxes in $\mathbf{R}^d$ of the form $I_1 \times \dots \times I_d$, such that there are integers $(k_1, \dots, k_d) \in \mathbf{Z}^d$ such that $I_i = [2^{k_i}, 2^{k_i+1}]$ or $I_i = [-2^{k_i}, -2^{k_i+1}]$. Littlewood-Paley theory implies that if $S_1, \dots, S_N \in \mathcal{S}$ and $f_1, \dots, f_N$ are Schwartz functions with f_i having Fourier support on S_i for each i, then for each $1 < p < \infty$,*

$$\|f_1 + \dots + f_N\|_{L^p(\mathbf{R}^d)} \sim_{p,d} \left\| (|f_1|^2 + \dots + |f_N|^2)^{1/2} \right\|_{L^p(\mathbf{R}^d)}.$$

A norm interchange then implies that if $p \geqslant 2$,

$$\left\| (|f_1|^2 + \dots + |f_N|^2)^{1/2} \right\|_{L^p(\mathbf{R}^d)} \leqslant \left(\|f_1\|_{L^p(\mathbf{R}^d)}^2 + \dots + \|f_N\|_{L^p(\mathbf{R}^d)}^2 \right)^{1/2}.$$

Thus we get a decoupling inequality.

21.1 Localized Estimates

Suppose $f_1, \dots, f_N$ are Schwartz functions in $\mathbf{R}^d$ with disjoint Fourier supports, and $\Omega \subset \mathbf{R}^d$. A natural question to ask is when one should expect

$$\|f_1 + \cdots + f_N\|_{L^2(\Omega)}^2 \lesssim \|f_1\|_{L^2(\Omega)}^2 + \cdots + \|f_N\|_{L^2(\Omega)}^2.$$

If we consider the bump function counterexample constructed from earlier, and let $\Omega = \{x \in \mathbf{R} : |x| \lesssim 1/N\}$, then $\|f_1 + \cdots + f_N\|_{L^2(\Omega)} \gtrsim N$, whereas $\|f_k\|_{L^2(\Omega)}^2 \lesssim 1/N$ so $\|f_1\|_{L^2(\Omega)}^2 + \cdots + \|f_N\|_{L^2(\Omega)}^2 \lesssim 1$, which means such a result cannot be obtained. However, we shall find that such a result holds if Ω is large enough, depending on the supports of $f_1, \dots, f_N$, and if we allow weighted estimates.

Let us begin with the case in one dimension. Given an interval I with centre x_0, and length R, we consider the weight function

$$w_I(x) = \left(1 + \frac{|x - x_0|}{R}\right)^{-M}$$

It is a useful heuristic that if f has Fourier support in I, then f is 'locally constant' on intervals of length $1/|I|$.

In $\mathbf{R}^d$, given a ball B with centre x_0 and radius R, we consider the weight function

$$w_B(x) = \left(1 + \frac{|x - x_0|}{R}\right)^{-M},$$

where M is a large integer. Then

$$\int w_B(x)\, dx$$

TODO FINISH THIS

21.2 Local Orthogonality